AN ACOUSTIC ANALYSIS OF VOWEL SEQUENCES IN JAPANESE

By

Isao Hara

Thesis submitted in partial fulfilment for the requirements of the degree of Doctor of Philosophy

School of Education, Communication and Language Sciences

Newcastle University

2015

Declaration

I certify that all the material submitted in this work which is not my own work has been identified and that no material is included which has been submitted for any other award or qualification.

Signed: Isao Hara

Date: 19 May 2015

Acknowledgements

The successful completion of this dissertation is the result of immense encouragement and help of many people. I would like to express my gratitude for their kindness.

I am deeply grateful to Professor Gerry Docherty for giving me the confidence to explore my research interests and the guidance to avoid getting lost in my exploration. Gerry patiently listened to what I was interested in, guided me from the start of the course to the completion. I also appreciate the help of Dr. Ghada Khattab, Dr. Katya Samoylova, Dr. Jalal Al-Tamimi, Dr. Sophie Salffner and other members of the Phonetics research group in Newcastle University. They helped me to develop an idea, gave valuable advice and supported me both openly and behind the scenes.

My academic work at Newcastle University had been supported by the ambassadorial scholarship of Rotary International. I would like to show my gratitude to my host Rotarian, Mr. Peter Calvert and his wife Shirley. They provided a great assistance to settle in an “unfamiliar” environment. I also appreciate very much the support from the Rotary Club of Seaburn, the Rotary Club of Hokuto and my sponsor Rotarian, Mr. Chiyozou Shimizu.

Ms. Tomoko Yoshino and Dr. Martin Gore kindly sent copies of important research papers from Japan. Without their help, I could not collect wide range of literature.

I have been thought about how I could repay what they had done, and I came to the conclusion that there was nothing better for me than to give back to the community something positive.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

For many years, researchers on phonetics have been trying to characterise a sound of languages from various points of view and employing a variety of approaches and methods. Some researchers have focused on finding some common features across languages while others have concentrated on the phonetic details of a particular language. This thesis aims to investigate, using acoustic analysis methods, the properties of non-identical vowel sequences of a particular language, Japanese, thereby enhancing our knowledge of the phonetic characteristics of this particular language and in the process shedding light on phonetic categories which are routinely used across languages.

It is uncontroversial that Japanese has five distinctive vowels but no diphthong as a phoneme. These monophthong vowels can occur in sequences and each vowel can be followed by each of the other. Where two vowels do occur successively, they are typically analysed as constituting two syllables and two morae. Therefore, Japanese vowel sequences are not diphthongs because they violate the syllabicity condition from the phonological point of view. However, despite the phonological condition, a syllabicity of a diphthong, there is some debate in the literature about whether Japanese VV sequences can be regarded as diphthongs in certain environments.

Traditionally there are three key defining characteristics of diphthongs. Firstly, from a phonological point of view, diphthongs are analysed as a single phoneme, even if they are transcribed phonetically by a digraph consisting of two vocalic symbols. Secondly, they are inherently dynamic vowel sounds, realised as a continuous glide from one zone of the vowel space to another. And thirdly, from a phonological point of view, they are analysed as constituting a single syllable nucleus.

The central question of my study is to consider the extent to which the vowels sequences in Japanese which apparently don’t have the defining characteristics of diphthongs noted above, nevertheless, from an acoustic point of view, may share many features, and in doing so to contribute to something of a debate on this matter of Phonetics and Phonology of Japanese as well as a better understanding of a notion of diphthongs.

Experimental materials are VV sequences produced by native speakers of Japanese in different boundary conditions and speech style conditions. All the possible VV sequences with possible accent contour for the sequence are prepared so that the results are expected to fill the gap between the previous works on describing the vowel sequences of Japanese and to extend the knowledge to the notion of diphthongs based on a comprehensive data from Japanese.

Chapter 2 summarises the previous works on the definition of diphthongs, the phonetic and phonological account of diphthongs and vowel sequences, the acoustic analysis of diphthongs and why it is problematic in Japanese. Research questions are subsequently formulated based on the literature review.

Chapter 3 explains the experimental method for obtaining the data. A detailed explanation of the experimental condition and the data to be obtained is given. The methods for the acoustic measurement and the statistical analysis are also described.

Chapter 4 reports the results of the time-domain analysis of Japanese VV sequences. The total duration for the vowel sequence, the ratios of the subsection intervals are described and analysed statistically in respect of various independent variables. The correlation between the total duration of VV sequences and the ratios for transition is also shown.

Chapter 5 reports the results of the spectral domain analysis. F1 and F2 of the vowels in VV sequences are shown and compared across a variety of elements. The degree that the vowel in VV sequences is reduced from the corresponding singletons is represented by the Euclidian distance and is also compared across a variety of factors. The transition speed, calculated by dividing the Euclidian distances between V1 and V2 by the transition durations, is also reported.

Chapter 6 discusses the acoustic properties of VV sequences in Japanese and the extent to which they have diphthong-like features. Describing phonetic properties of VV sequences of Japanese, the association between the timing properties and the spectral properties are investigated. The summary of the findings is then compared with the existing literature on Japanese VV sequences and leads to the discussion on whether Japanese VV sequences can be classified as diphthongs. These discussions lead to a conclusion in Chapter 7.

Chapter 2. Literature Review

2.1 Introduction

Phonetic research mainly aims to investigate the articulatory movement of speech sounds of a particular language or compare the phonetic characteristics of speech sounds between languages. There has been a great deal of research on vowels, focusing on their phonetic characteristics in one language or cross-linguistically as well as their phonological features. Figure11 1 1below illustrates the phonetic characteristics of two types of single vowels, which are generally called (1) monophthongs and (2) diphthongs and one type of sound pattern, (3) vowel sequences. The articulatory position of monophthongs is generally described by the vertical/horizontal position of the tongue in a vowel space, whereas diphthongs are described as a movement and vowel sequences involve a movement from one target to another.

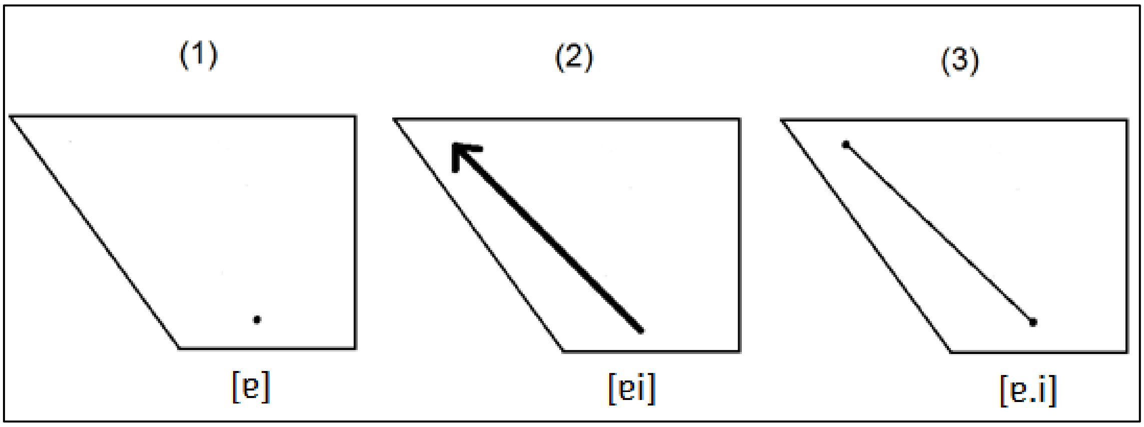


Figure 1 An illustration of monophthongs, diphthongs and vowel sequences

In general, many vowels can be given a unitary description because they are described as single-target vowels (see (1) in Figure 1) and their place of articulation is given in terms of the vertical and horizontal positions in a vowel space as well as lip-rounding. This type of vowel is called monophthong (Catford, 1988). However, some languages including English have diphthongs, which have been described as gliding sounds (Catford, 1988) and their articulation begins from a particular position and moves to another in a vowel space (see (2) in Figure 1). The detailed explanation of diphthongs will be given in 2.2.

There has been much less analysis and discussion characterising the features of diphthongs, although these are abundantly present in languages of many different types and it is reported that they occur in about one third of the world’s languages (Lindau, Norlin and Svantesson, 1990). In addition, the nature of their phonetic characterisation is so varied that it is necessary to examine how the concept of a diphthong is considered. Some languages lack diphthongs in their phonological inventory but observe a phenomenon that a monophthong vowel is followed by another monophthong vowel. The articulation of such vowel sequences seems to be similar to that of diphthongs because both of them are vocalic sounds which involve some changes in their vocalic qualities. But these vowel sequences consist of two single-target vowels; therefore, their articulatory movement starts from the former vowel and moves to the latter (see (3) in Figure 1).

Although there have been a number of studies describing the phonetic properties of vowels themselves, few studies attempted to analyse the phonetic properties of two-sequential vowels and compare with the properties of diphthongs. The aim of this chapter is to suggest that diphthongs leave much room for discussion, and to focus especially on the differentiation on diphthongs and vowel sequences, a matter which has become a point of debate within the analysis of one language, Japanese.

2.2 Definition of the diphthong

The term ‘diphthong’ originates from the Greek word ‘δίφθογγος’, which is combined with ‘δι’ (twice/doubly) and ‘φθόγγος’ (voice/sound) (Oxford English Dictionary, 2004). Even if the original Greek word means ‘having two sounds’, this term is not necessary defined as such in the phonetics and phonology literature. One definition states that a diphthong consists of two vowels in a syllable such as 'a sequence of two perceptually different vowel sounds within one and the same syllable' (Catford, 1977, p. 215). Cruttenden (2001) also describes it as 'the sequences of vocalic elements which form a glide within one syllable' (p. 140). Another definition explains it as a single vowel with changing quality such as 'an independent vowel-glide not containing within itself either a "peak" or a "trough" of prominence' (Jones, 1972) and 'a change from one vowel quality to another' (O'Connor, 1973, p. 154). The latter definition can be found in various sources of literature (Abercrombie, 1967; Denes and Pinson, 1993; Ashby and Maidment, 2005; Ladefoged, 2006).These definitions at least imply that a diphthong seems to be a vowel with changing quality, but it is not still clear whether the two different vowel sounds are in the one and the same syllable or whether one vowel occupies one syllable. In addition, the internal structure of a diphthong is also explained in different ways. Although Catford (1977) and O'Connor (1973) note that a diphthong involves a change of its vowel quality, although it is unclear that the change being argued is partial or overall in total diphthong duration, Laver (1994) implies that the formant movement of a diphthong can be divided into several phases. He describes a diphthong as 'a vocoid[1] in which the medial phase explicitly consists of an articulatory trajectory across the vocoid space' (p. 284). He does not explain further

---

[1] Laver (1994) explains that the term 'vocoid' corresponds straightforwardly…to the traditional phonetic term 'vowel'(p. 270). He uses different terms to account for the phonetics and phonology of 'resonants'.

but his definition seems to mean that a diphthong can have initial, medial and final phases and the medial phase involves a change of the vocoid quality. Clark and Yallop (1995) also note that a substantial portion of a diphthong is occupied by an articulatory movement. Their explanation implies that the articulation of a diphthong involves some phases other than the articulatory movement. This viewpoint is also given in Ladefoged (2006). Kent and Read (2002) describe diphthongs as overall glides, ‘dynamic sounds in which the articulatory shape slowly change during the sound's production’. Therefore, overall, diphthongs can at least be characterised as vowels whose quality changes, although it still remains unclear whether a diphthong is a sequence of two vowels (Catford, 1977) or one vowel (O’Connor, 1973), and whether its quality changes throughout its duration (Catford, 1977; Kent and Read, 2002) or whether there are intervals during which the vowel quality keeps steady (Laver, 1994; Clark and Yallop, 1995). However, none of the above literature gives any acoustic evidence to support their argument. Some research on acoustic analysis of English diphthongs implies that the vowel qualities keep steady for both onset and offset of diphthongs and their analyses seem to support the definition of Laver (1994); these will be discussed in section 2.3. Therefore, there is some agreement that the common phonetic feature of a diphthong may be a transitional and gliding vowel sound, but there is less consensus on characterising the course of the changing quality and few studies investigated the possible presence of relatively steady formant values during the production of a diphthong.

Phonologically, however, there seems to be a clear agreement that a diphthong occurs within and consists of one and the same syllable (Jones, 1972; Catford, 1977; Denes and Pinson, 1993; Catford, 2001; Ashby and Maidment, 2005; Ladefoged, 2006). Figure 2 summarises this view. From a phonological point of view, for languages which have such vowels, a diphthong is analysed as one vowel with changing quality and occupies one syllable. On the other hand, as just described from a phonetic point of view, there is a question about whether a diphthong is realised as (1) a changing quality throughout its duration (Kent and Read, 2002), or whether (2) the quality changes from one steady target position to another (Catford, 1977). The changing acoustic quality of a diphthong can be represented as the movement of formant frequencies on the spectrogram. A schematic view of these two possibilities of formant movement is also shown in Figure 2.

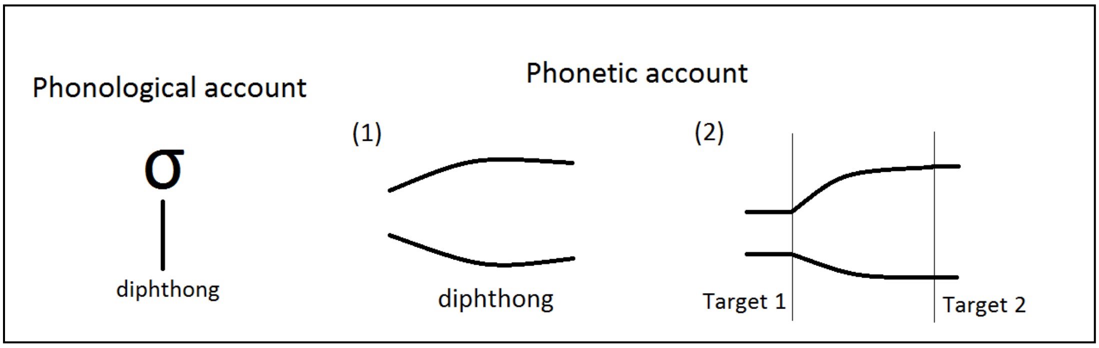


Figure 2 Phonological and phonetic account of a diphthong

Considering the definition of a diphthong, a sequence of two non-identical vowels is an interesting phenomenon because it has the potential to share the phonetic features of a diphthong but violate the phonological ones. When two vowels are produced sequentially, without any gaps between them, the vowel quality changes from the former vowel to the latter. Based on the general feature of articulation, it is quite unlikely that the quality of the latter vowel in a sequence begins just immediately after the end of the former vowel. Although it is apparent that producing two non-identical vowels without gaps between them involves a shift of articulation from the former vowel to the latter, it is not clear how gradual the vowel quality changes. In other words, the question is whether two non-identical vowels have the characteristics of two vowels or whether the vowel quality gradually changes throughout its duration.

From the phonological description of the two non-identical vowels, it is apparent that each vowel occupies one syllable because a monophthong generally takes up a syllable nucleus, not an onset or coda and is analysed as one syllable when it is produced independently. Figure 3 represents the phonological and phonetic account of the sequence of vowels. A sequence of two vowels can be analysed as two syllables because each monophthong vowel occupies one syllable. However, the production of the vowel sequence can be either (1) a gradual change throughout its duration or (2) starting with the steady state of the first vowel, then changing towards the steady state of the second vowel as long as these two different vowels are produced without a gap between them (Kawakami, 1977).

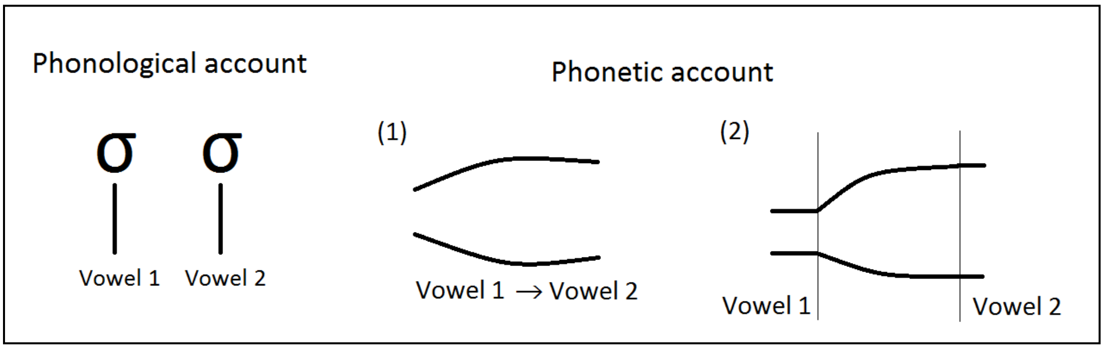


Figure 3 Phonological and phonetic account of a vowel sequence

Therefore, the difference between a diphthong and a two-vowel sequence lies in the phonological fact that a diphthong occupies one syllable whereas a vowel sequence takes up two. However, there seems to be less obvious difference in the phonetic features between a diphthong and a vowel sequence, despite the fact that a diphthong consists of two ‘targets’ and a vowel sequence consists of two single vowels.

Catford (2001) acknowledges that a sequence of two different vowels can occur in two successive syllables but such a sequence is not given a particular term. Abercrombie (1967) clearly denies that a vowel sequence occupying two syllables is a diphthong and Clark and Yallop (1995) note that these two vowels constitute separate syllabic peaks. However, vowel sequences, which Catford (2001) does not characterise with a particular term, frequently occur in many languages and it is worth comparing their acoustic properties to those of a diphthong so that the definition of a diphthong becomes clearer and more accurate based on the acoustic analysis.

2.3 Acoustic properties of diphthongs

As is noted in the previous section, a diphthong is commonly recognised as having a dynamic nature which is represented acoustically as the movement of formants, although there is a disagreement on the internal structure of a diphthong. The main methods of characterising the acoustic properties of diphthongs are formant analysis to clarify the characteristics of the formant trajectory and duration analysis to examine the duration of transition and any subsections which are identified. However, the measurement techniques are so varied among researchers that the acoustic properties of diphthongs are described in different ways as a result. There are mainly two different ways of characterising and describing diphthongs i.e., measuring formant frequencies at some sampling points on the one hand, and segmenting the total diphthong duration by the formant movement on the other.

The former method was adopted by Holbrook and Fairbanks (1962). They attempted to describe the whole figure of the formant movement by measuring the formant frequencies and amplitudes of diphthongs at five sampling points that were equally located throughout the total diphthong duration, which Simpson (1998) later called the ‘five-point method’. This method was also used in Shimoda, Narahara, and Okamoto (1973) for analysing English and German diphthongs and Simpson (1998) for German diphthongs. This ‘five-point method’ seems to be valid on the condition that the formants of a diphthong are thought to change gradually throughout its duration.

Although this method may reflect the formant movement more accurately depending on the number of the sampling points, it may also have problems of shortage of data because the sampling points do not necessary coincide with the start or the end of the transition if it occurs in the medial part of the total diphthong duration. Clermont (1992) conducted a detailed sampling of formant frequencies throughout the total diphthong period, measuring formants every 10 msec to capture the transition. Although he admitted that this was not economical, he insisted that his method gave more insights into the nature of the transition of diphthongs. This representation has the merit of capturing the formant movement in detail, although the volume of the data storage should be taken into account. In addition, the analysis should be conducted by the automatic formant tracking system. It would be unnecessarily complicated if this analysis is done manually.

Furthermore, some researchers set their sampling points at relative locations of the total diphthong duration. Weil, Fitch and Wolfe (2000) measured the formant frequencies at 0% (onset of a diphthong), 25%, 50%, 75% and 100% (offset) of the diphthong period, Roengpitya (2002) and van Heuven (2002) measured at 25%, 50% and 75% and Tsukada (2008) at 20%, 50% and 80% of the total diphthong duration. Their ‘relative location’ method may take into account the imbalance of the duration of diphthongs produced by different individuals. Also, it enables the capturing of formant trajectories by normalising the other factors of variation such as speech rate. However, this method also has the issue of the number of the samples, as with the ‘five-point method’.

The latter method, which can be called the segmenting method, is based on the assumption that diphthongs have two targets and each target is realised as a steady state of the formant. The earliest research for adopting this method was Lehiste and Peterson (1961). They divided the total diphthong durations into five parts; Onglide, Target 1, Glide, Target 2 and Offglide. The onglide and offglide parts are transition from the previous or the following consonants in a word or a sentence. It is true that formant frequencies are affected by the neighbouring consonants, which can be regarded as parts of targets when the Consonant-Diphthong-Consonant sequence is analysed phonemically. This ‘five-period method’ was adopted by Yu, Li and Wang (2004), Cox (2006) and Alam, Habib and Khan (2008).

On the other hand, Gay (1968) divided diphthongs into three periods, such as the onset steady states, glides and offset steady states for diphthongs, although he did not describe the analysis method in detail. However, his research has been widely cited and his ‘three-period’ approach has been applied in many studies of diphthongs. Table 1 below summarises the two major types of methods with two different ways of analysis of diphthongs. It shows that the segmentation method has been widely used among researchers and segmenting the total diphthong duration into three periods has been recognised as one of the standard way of analysing diphthongs.

Table 1 Summary of methods for analysing diphthongs

Trajectory capturing method		**Segmentation method**	
Regular intervals	Relative locations	Onglide, Target1, Transition, Target2 & Offglide	Target1, Transition & Target2
Holbrook and Fairbanks (1962) Shimoda et al. (1973) Clermont (1992) Simpson (1998)	Weil et al. (2000) Tsukada (2008)	Lehiste and Peterson (1961) Yu et al. (2004) Asher et al. (2005) Cox (2006) Alam et al. (2008)	Gay (1968) Hirasaka and Kamata (1981) Svantesson (1984) Jha (1985) Kasuya and Sato (1990) Lindau et al. (1990) Penny (1992) Harrington et al. (1994) Roengpitya (2002) Gore (2006) Alam et al. (2007) Cheung (2007) Misono and Hirasaka (2008) Tasko et al. (2010) Keerio et al. (2011)

2.4 Acoustic studies of diphthongs in languages

Despite the variety of methodologies, the objective of the acoustic analysis is to describe the movement of the formant frequencies either as a part of the study of the vowel system and variation of a particular language or as the cross-linguistic study of diphthongs.

Holbrook and Fairbanks (1962) is one of the oldest papers on the acoustic analysis of diphthongs of General American English. They compared the intervals of F1, F2 and F3 of 6 American diphthongs in carrier sentences produced by 20 General

American male speakers and showed that the total duration of /ei/, /ou/ and /ju/[2] were relatively longer than that of /ai/, /oi/ and /au/. They also described the whole trajectory of those diphthongs and compared them with the frequencies of monophthongs produced by the same subject group. Their formant analysis showed that the /_i/ diphthongs ended with the similar positions in the vowel space. In addition, the F1/F2 vowel spaces of both initial and final parts of diphthongs extensively overlap with each other but they considered these parts to be mutually exclusive as far as their median values were considered.

Lehiste and Peterson (1961) characterised the formant movements of diphthongs as well as consonant transitions and glides by dividing the total duration into subsections and discussed whether they have either one or two target positions. They concluded that diphthongs had two steady states which they call targets and the transitions between targets were longer than both targets.

Gay (1968) analysed five American English diphthongs (/ɔɪ/, /ɑɪ/, /ɑʊ/, /eɪ/, /oʊ/) produced by 5 American male speakers. 10 real words containing each diphthong were prepared and subjects were asked to produce each word ten times in three different speech rates, which were normal conversation rate, fast and slow. He segmented the total diphthong period into three parts i.e. onset-steady period, gliding period and the offset-steady period. The durations of each period were measured and F1 and F2 were

---

[2] They called /ju/ as the ‘diphthong-like’ sound because this sound can be characterised as having the formant movement as with other diphthongs.

taken at steady state periods. His duration analysis showed that the total diphthong durations became shorter along with the increase in speech rate and it was also reflected either by the short duration of steady states or the missing of them. He only reported the mean diphthong durations and not the ratios of each internal period to the total duration of the diphthongs. However, the calculations of percentages by the author showed that the percentages of steady states to the total diphthong durations became larger by approximately 20% by the increase of the speech rate (Slow: 69.7%, Moderate: 73.3%, Fast: 89.5%). Also, the percentages of both onset and offset steady states were found to be smaller as the speech rate increases. The percentages calculated by the author are found in Table 2 (p. 23) for the purpose of comparing the results from various studies. His formant analysis showed that the formant frequencies of the offset steady periods were different among different speech rates but the formants of the onset steady periods were not affected by the speech rate.

The influence of speaking style on the acoustic realisation of American English diphthongs was also investigated by Tasko and Greilick (2010). They hypothesized that the clarity of speech would influence the duration and formant frequencies of diphthongs. 20 male and 29 female American English speakers were asked to produce a word containing /ai/ by reading sentences and speaking in a conversational style. Although the detailed data was not presented in their article, they argued that all diphthongs were longer in the clear speech condition but there was no gender effect nor interaction between gender and the speech clarity condition on both total diphthong duration and durations of onset, transition and offset of diphthongs. Their formant analysis showed that there were effects on speech styles for F1 of onset and F2 of offset, although the detailed data was not given.

The attempts for describing diphthongs in other varieties of English were reported by several researchers. Penny (1992) conducted a formant analysis of 8 Australian English diphthongs (/eɪ/, /ɑɪ/, /ɔɪ/, /oʊ/, /ɑʊ/, /ɪə/, /eə/ and /uə/) in a /h_d/ environment, produced by 48 female speakers. She gave detailed descriptions of formant movements of each diphthong but no statistics were conducted. Cox (2006) analysed the diphthongs produced by the Australian teenagers. He collected real word productions containing five diphthongs (/ɑe/, /æɪ/, /oɪ/, /æɔ/, /əʉ/) and nine monophthongs by 60 male and 60 female Australian teenagers. His segmentation method was dividing total diphthong duration into five periods as is shown in Table 1 above. His duration analysis showed that the onglide and offglide of diphthongs were very short in all diphthongs and the transition was the major part of the total diphthong duration. His formant analysis compared the F1 and F2 of ‘target’ periods (i.e. steady states) of diphthongs with those of monophthongs. She concluded that the starting and ending points of diphthongs were different from monophthongs in the F1/F2 vowel space.

Deterding (1996) compared two diphthongs (/ei/ and /ou/) produced by eight Singaporean and three British English speakers. He examined the rate of change of the formant frequencies of diphthongs and concluded that Singaporean English had a smaller change in the formant movement than British English. Maxwell (2010) examined the formant frequencies of six diphthongs (/aɪ/,/ɔɪ/, /aʊ/, /ɪə/, /eə/, /ʊə/) of Indian English. seven male speakers of Punjabi and Hindi were asked to produce isolated words containing diphthongs and F1 and F2 were measured at the onset and the offset. He found that the position of [a] was the same for all speakers and some diphthongs were produced as long monophthongs in Indian English.

Mono-linguistic studies of diphthongs have also been conducted for other languages. Jha (1985) compared the formant frequencies of the two Maithili (a modern Indo-Aryan language) diphthongs /əi/ and/əu/ with equivalent monophthongs to argue whether these diphthongs are pairs of monophthongs. Although the number of participants was not described in the article, two diphthongs were produced in isolation and in words with three different speech rates. He concluded that the F1 and F2 of the onset of both diphthongs and the rate of change in F2 were stable in all speech rate conditions, whereas the F1 and F2 of the offset position were variable and affected by the speech rate. In addition, onset durations were shorter than offset durations and the transition became shorter in faster speech. Therefore, it can be assumed that the Maithili diphthongs are not the combinations of two monophthongs because the offset target of the diphthongs are variable.

Roengpitya (2002) revealed the acoustic differences between short and long diphthongs in Thai. Thai is said to have either six diphthongs (three short and three long) or three phonemic diphthongs but they are realised as long and short depending on the phonological environment. He found that the steady states of the first element occupied a large proportion of the total diphthong duration and diphthongs in open syllables were 1.7 times as long as diphthongs in closed syllables. However, there was no significant difference between short and long diphthongs in terms of formant frequencies. His later paper (Roengpitya, 2007) attempted to mark a boundary between the first and second vocalic elements of diphthongs by using the zero-crossings. The results largely followed his previous paper and thus the first element was longer than the second one and the total duration in open syllable was longer.

Diphthongs in several dialects of Chinese have also been analysed acoustically. Svantesson (1984) made an acoustic analysis of Mandarin Chinese comprehensively and reported the formant frequencies of monophthongs, diphthongs and triphthongs as well as the durations of onset, transition and offset of diphthongs. Yu, Li and Wang (2004) compared nine diphthongs of Mandarin and Shanghai dialects and argued that the formant pattern and the vowel space of diphthongal elements were different between these dialects. Diphthongs in Fuzhou dialect were examined by Peng (2007), in the Meixian Hakka dialect by Cheung (2007) and in the Beijing dialect by Lee (2010).

Lindau (1985) described the vowel system of Hausa, which has two diphthongs, /ai/ and /au/. She measured the durations of steady states and transition as well as the formant frequencies of both steady states. Asher and Keane (2005) investigated the colloquial form of Tamil diphthong /ai/ in different conditions such as formal vs. informal speeches, word length, syllable position and lexical vs. grammatical morphemes.

Cross-linguistic comparison of diphthongs has also been conducted. Shimoda, Narahara, and Okamoto (1973) analysed three German diphthongs /ai/, /au/ and /oy/ and compared with the results of English diphthongs in Holbrook and Fairbanks (1962). The major acoustic difference between German and English diphthongs appeared in overall formant trajectory and the formant frequencies of onset and offset of diphthongs compared with those of monophthongs. Lindau, Norlin and Svantesson (1990) compared the duration and formant frequencies of /ai/ and /au/ diphthongs of Arabic, Hausa, Chinese and English. They showed that the ratio of transition by the total diphthong duration was significantly different among them.

Therefore, the common acoustic characteristics of diphthongs have been described as changing quality of sonorant sound with/without the steady states as onset, offset or both. However, the method of analysing those subsections in diphthongs was not necessarily well described in those papers. Some studies suggest that the steady states seem not to be the period whose formant frequencies are kept steady. This is because some authors might determine the onset of the transition as the point at which there is a clear and remarkable change in formant frequencies and periods of whose ratio of change is smaller than others might be defined as ‘steady’ states.

Table 2 below shows the comparison of onset, transition and offset ratios cited from different articles. Results are obtained from those papers dealing with the productions by native speakers of the language. Some studies just described the durations of each part and the ratios were not calculated and presented in the original paper and some papers just gave information of transition ratio and did not mention anything about the other subsections. Moreover, the actual method of dividing subsections is slightly different among them. The target diphthongs are different from each other and the ratios given in this table were combined data of all diphthongs in each paper. However, this table may help to understand the acoustic characteristics of diphthongs in different studies or different languages. Notably, the ratios of the transition are different between languages and also widely range among studies, especially of English. It also suggests that diphthongs have something other than a trajectory period. In general, the acoustic properties of monophthong vowels are largely steady in the production. The question of whether there is an acoustic difference between diphthongs and pairs of monophthongs arises from research on diphthongs.

Table 2 Percentages of Onset, Transition and Offset of previous works

Language	**Reference**	**Target**	**Condition (if specified)**	**Onset**	**Transition**	**Offset**
English	Lehiste and Peterson (1960)*[3]	ai, au, oi		35.8%	40.9%	23.3%
	Lehiste and Peterson (1961)*	ai, au, oi		36.4%	42.1%	21.5%
	Gay (1968)*	ai, oi ,au, ei, ou	Slow	18.5%	69.7%	11.8%
			Moderate	15.5%	73.3%	11.2%
			Fast	8.5%	89.5%	1.9%
			Total	14.2%	77.5%	8.3%
	Hirasaka and Kamata (1981)	ai, oi		39.3%	51.7%	9%
	Kasuya and Sato (1990)	ai		41%	38%	21%
	Lindau et al. (1990)	ai, au		-	66.5%	-
	Cox (2006)*	ɑe, æɪ, oɪ, æɔ, əʉ		26.9%	51.0%	22.1%
	Gore (2006)	ai		31.7%	39.5%	28.8%
	Misono and Hirasaka (2008)	ai, au, oi		37.4%	44.4%	18.2%
Chinese	Svantesson (1984)*	iu, ui, ia, ai, ua, ao, ie, ei, uo, ou, üe		29.2%	24.0%	37.3%
	Lindau et al. (1990)	ai, au		-	40-50%	-
Maithili	Jha (1985)*	əi, əu	Slow	21.2%	46.0%	32.8%
			Moderate	16.7%	57.5%	25.8%
			Fast	11.4%	68.6%	19.9%
			Total	16.4%	57.4%	26.2%
Tamil	Asher and Keane (2005)	ai		-	50.7%	-
Arabic	Lindau et al. (1990)	ai, au		-	16-20%	-
Hausa				-	16-20%	-

Although the diphthongs used in those experiments are different, especially in English, the ratios of those subsections are quite different from each other. The

---

[3] The subsection ratios of studies with asterisks were calculated by the author based on the durations reported in the original articles.

transition of English diphthongs seems to be the largest part in most study, but Kasuya and Sato (1990) argued that the onset of /ai/ is the largest part of the diphthong. Comparing the ratio of the transition, Gay (1968), Hirasaka and Kamata (1981) and Lindau et al. (1990) showed that it occupied more than half of the total diphthong duration, whereas other works presented that the transition took up less than half of the total diphthong duration. It can also be pointed out that the onset ratio is commonly larger than the offset ratio in English, whereas the offset ratio is larger than onset ratio in Maithili.

However, many studies seem to focus on a particular diphthong of a language or limited types of diphthongs. Further research still needs to be done and there is a room for further investigation on the acoustic analysis of diphthongs. The current work does not deal with the analysis of diphthongs.

2.5 Research on diphthongs and vowel sequences

Differences between diphthongs and vowel sequences are a matter of debate in some languages. Spanish is one of the languages in which the distinction between diphthong and vowel sequence has been studied from a variety of points of view. Vowel sequences are called “hiatus” in some studies on Spanish vowel systems (Aguilar, 1999). He uses this term as “a combination of two phonological vowels” (p. 58). To review studies, the term “hiatus” refers to a vowel sequence. Although these studies have some common phonetic and phonological features that distinguish between them, these features become problematic when one attempts to apply them to Japanese. This section reviews discussions on the diphthong/hiatus distinction, points out their phonetic and phonological accounts and addresses the difficulties of them if we try to apply those accounts for Japanese.

In Spanish, it is reported that two adjacent vowels can be pronounced as a diphthong or a hiatus, that is, vowel sequences are diphthongised by the stress dislocation or by alteration of the vowel quality of one of the elements in the sequence (Garrido, 2007). In addition, the phonetic difference between diphthongs and hiatuses are that the high vowels are realised as the semivowels (Carlyle, 1987). Spanish is said to have a five-vowel system. Manrique (1979) reported that the combinations of a close vowel (/i/ or /u/) and an open vowel (/e/, /o/ or /a/) or of two close vowels whose stress was placed on the open vowels were treated as diphthongs. It is also noted that the vowel sequences which consist of open vowels or whose stress is on the close vowel should be distinguished from diphthongs and are termed as a hiatus in many relevant literature (Aguilar, 1999; Hualde and Prieto, 2002; Cabre´ and Prieto, 2007; Chitoran and Hualde, 2007). For example, the vowel sequence of /oi/ is treated as a diphthong when the stress is on /o/ such as ‘hoy (today)’ but as a hiatus if it is pronounced as [oˈi] such as ‘oi (I heard)’ (Manrique, 1979). However, there seems to be dialectal variation and vowel sequences of mid/low vowels such as /eo/ tend to be diphthongs in Latin

America (Garrido, 2007). This phonetic and phonological phenomenon is called ‘diphthongisation’ and Spanish diphthongs are created by this process. Mauder and van Heuven (1996) gave a detailed explanation of diphthongised vowels. They argued that high vowels of the initial position were consonants, those after a consonant were semi-consonants and those in a final position were semi-vowels. Although the acoustic analysis to show the diphthong/hiatus distinction has been conducted in many studies, none of them present that the high vowels in diphthongs are realised not as vowels but as semi-vowels, semi-consonants or consonants. Garrido (2007) also noted that the hiatus had longer duration than diphthongs.

The major acoustic analysis of Spanish diphthongs and hiatuses involves measuring the total durations and formant frequencies. In terms of duration, it is reported that the total diphthong duration is shorter than the total hiatus duration (Aguilar, 1999; Hualde and Prieto, 2002; Face and Alvord, 2004; Garrido, 2007). Aguilar (1999) analysed 12 Spanish diphthongs and 12 hiatuses by asking 16 male speakers to do a map task to obtain productions in a dialog and to read out carrier sentences to obtain read data. In both speech styles, the total diphthong durations were significantly shorter than the total hiatus durations. He also noted that the diphthong duration was longer in stressed contexts than unstressed context and hiatuses became longer when /i/ and /u/ are stressed. There was also a significant effect of the speech style and the total durations from the map task were reduced by 20% for diphthongs and 15% for hiatuses to the reading task. Hualde and Prieto (2002) also investigated the acoustic contrast between Spanish diphthong /ja/ and hiatus /ia/. They prepared 20 words containing ten diphthongs and ten hiatuses and asked six Spanish speakers to pronounce these words in carrier sentences. They concluded that the diphthong/hiatus contrast appeared in the total duration and the diphthong duration was significantly shorter than the hiatus duration. But they also mentioned that there was an overlap in the range of duration between diphthongs and hiatuses. Face and Alvord (2004) also compared Spanish diphthong /ja/ and hiatus /ia/ by asking five participants to repeat 20 words in a carrier sentence containing target diphthongs and hiatuses three times and found that the overall duration of diphthongs were significantly shorter than hiatuses. Garrido (2007) compared the Spanish diphthong /io/ and Spanish hiatus /eo/ of both reading task and story-telling task produced by nine Caribbean and eight Andean Spanish speakers. Although his major findings on duration was the same as Aguilar (1999), such as the total diphthong duration being shorter than the total hiatus duration and the duration from the story-telling task being shorter than the reading task, the significance lies in the fact that there was a dialectal difference in the diphthong/hiatus contrast and its contrast in duration disappeared in informal speech by Andean Spanish speakers.

The formant analysis of Spanish diphthongs and hiatuses varies among researchers. Manrique (1976) compared F1, F2 and F3 of high vowels /i/ and /u/ in 1) /h_e/, 2) /C_e/, and 3) /Ce_/ environments and those vowels of monophthongs produced by 4 speakers. It was shown that the F1 of /i/ in the /hie/ was lowered compared to that of monophthongs and that F1 and F2 of /u/ in /hue/ were remarkably lowered, representing that those vowels in /h_e/ environment, which was called ‘absolute initial position’, might be described as semi-vowels. It was also shown that the rate of change of F2 values was larger in /h_e/ environment than others. Aguilar (1999) investigated the degree of formant curvature for the total diphthong duration by extracting F1 and F2 every 10 msec anticipating that the degree of curvature of formants would be a parameter to distinguish between diphthong and hiatus. It was reported that the F2 trajectory showed a greater curvature in hiatus than in diphthong, which he concluded was a feature of the hiatus/diphthong distinction. The effect of speech style was recognised mainly in F2. Garrido (2007) measured the F1 value only of /i/ and /e/ in the /io/ diphthong and the /eo/ hiatus produced in reading and narrating tasks. This is due to the idea that the lowering of F1 value would associate with a higher degree of diphthongisation. The result showed that F1 values of /e/ produced by Caribbean Spanish speakers were lower than those of Andean Spanish speakers. Speech style was also associated with the distinction and there was no difference between Caribbean and Andean speakers in the narration task but in the reading task, Caribbean speakers showed lower F1, which indicated that there was a contrast between diphthongs and hiatuses.

The diphthong/hiatus contrast studies on Spanish indicate that the shorter overall duration, the greater difference on formant frequencies from monophthongs, the small trajectory curvature and the low F1 can be the features of hiatuses being diphthongised.

However, these studies do not mention the internal structural difference between diphthongs and hiatuses. None of them indicate the ratios of each vowel in hiatuses and the transition from the first vowel to the second. It should be of interest whether there is a difference in those ratios when hiatuses are diphthongised. Until present, there are only two studies which segment both diphthongs and hiatuses into onset steady state, transition and offset steady state, which is a common method in studies on diphthongs to indicate the difference between them. Manrique (1979) investigated 10 Spanish diphthongs and those equivalent vowel sequences produced by 2 male speakers of standard Buenos Aires Spanish. Durations of steady states and transition were measured and compared between diphthongs and vowel sequences. F1, F2 and F2 at the mid-point of steady states were measured and compared with the formants of monophthongs. Gore (2006) compared the English diphthong /ai/ and the vowel sequence, in which /a/ and /i/ occur across word boundary (e.g. ‘Pap**a e**ats bread’). He asked 10 English speakers to produce 40 words containing the English diphthong /ai/ and three sentences containing 4 /a#i/ sequences across the word boundary at slow and fast speed. Durations of steady states and transition were measured and compared between monophthongs and vowel sequences. F1 and F2 at the mid-point of steady states were also measured and compared with each other. However, it should be noted that the segmentation methods of the two studies are slightly different with each other. Manrique (1979) segmented the onset, transition and offset by the F2 trajectory but Gore (2006) identified the steady states as the period that the stable quality of /a/ and /i/ could be heard. Also, Manrique

(1979) did not report the ratios of onset, transition and offset by the total diphthong/hiatus duration. For the purpose of comparing the results from Manrique (1979) and Gore (2006), the subsection ratios of Manrique (1979) were calculated and integrated based on the duration measurement. The terms ‘onset’ and ‘offset’ used in diphthong analysis are changed to ‘V1’ and ‘V2’. Their results are presented in Table 3 below.

Table 3 Comparisons of subsection percentages between diphthongs and VV sequences

Language	Reference	Target	Condition	V1	Transition	V2
Spanish	Manrique (1979)	io, oi, ie, ei, au ua, ia, ai, ue, eu	Diphthong	30.4%	38.0%	31.5%
			VV Sequence	31.0%	35.7%	33.3%
English	Gore (2006)	ai	Diphthong	31.7%	39.5%	28.8%
			VV Sequence	27.6%	21.0%	51.5%

In Spanish, there seems to be less difference between diphthongs and vowel sequences, but the transition ratio is slightly smaller in vowel sequences than in diphthongs. Instead, the ratios of V1 and V2 become consistently larger in vowel sequences than in diphthongs. On the contrary, English shows a greater difference between a diphthong and a vowel sequence. The ratio of V2 occupies more than half of the total vowel sequence and the transition ratio becomes the smallest among those subsections. However, his result was taken from only one diphthong and one sequence.

The actual characteristics of diphthongs and vowel sequences in English may not be described until the comprehensive analysis is conducted.

However, their formant analysis presented a different outcome. Manrique (1979) reported that all the vowels in Spanish diphthongs were centralised compared to the monophthongs in V1/V2 vowel space and the vowel areas of /e-i/ and /o-a/ overlapped respectively. However, Gore (2006) presented that F1s of both /a/ and /i/ of English vowel sequence /a#i/ were lower than those of diphthongs. It can be assumed that /i/ in a vowel sequence was located at more centralised position but /a/ in a diphthong was more centralised. However, he did not compare those values with monophthongs. Although there was a significant difference between a diphthong and a vowel sequence in F1, no significant difference was found in F2.

2.6 Studies on diphthong/hiatus distinction

The argument that a vowel sequence is regarded as the diphthong is also presented in other languages. Carlyle (1987) gave a metrical phonological account of the diphthongisation of Breton, a Celtic language spoken in France. Breton is said to have seven monophthong vowels and those monophthongs can be combined freely. Although no acoustic evidence of the distinction between diphthongs and vowel sequences was given, it was argued that the second vowels of the vowel sequence were realised as semi-vowels when there was a large sonority difference between the first and the second vowels thus the vowel sequences were diphthongised. Falling sequences, whose first vowel is more sonorous than the second one, are said to be realised as diphthongs invariably. On the contrary, rising sequences, whose second vowel is more sonorous, are realised as diphthongs in faster and casual speech but as vowel sequences in slow and careful speech. Level sequences, whose vowels in a sequence have an equal sonority, have various realisations between diphthongs and vowel sequences. He concluded that the sonority differences between the first and the second vowels in vowel sequences was the crucial factor of differentiating between diphthongs and vowel sequences and the sonority differences between vowels are language-specific. Therefore, different sonority scales were given for Breton and Spanish.

Urdu, a language spoken in Pakistan and India, is also said to have diphthongisation phenomenon. Sarwar, Ahmed and Tarar (2004) described the acoustic properties of Urdu diphthongs. However, they mentioned that Urdu had seven long vowels, three short vowels and six nasalised long vowels as phonemes but there were no phonemic diphthongs. Even though Urdu syllable structure does not permit two vowel sequences, diphthongs are possible when the intervening consonants between two vowels are deleted. For example, the vowel sequence /ao/ in the word /ʤao/ is possible when [ʔ] in /ʤaʔo/ is deleted. They note that they are diphthongs but there is still a room for discussion of whether they can be called diphthongs both phonetically and phonologically. In addition to the perceptual test of whether a variety of vowel sequences in Urdu can be perceived as diphthongs by native speakers, the acoustic analysis of Urdu diphthongs was conducted. They asked three male and three female Urdu speakers to produce words containing Urdu diphthongs in carrier sentences three times and Formant frequencies of the steady states of the first and the second vowels. Although they segmented into steady states and transition, no information was presented for the duration. They presented the speaker variation and the gender difference but there was no general information of diphthongs in Urdu.

Asu, Lippus, Niit and Türk (2012) presented the acoustic properties of diphthongs in the Kihnu variety of Estonian, which was formed by combining the long vowels and mid vowels. Kihnu Estonian is said to belong to the Insular dialects of the north Estonian dialect group and be spoken in the Island of Kihnu. There is no phonemic diphthong but diphthongs can be formed by diphthongisation. They asked 6 female Kihnu speakers to read 153 sentences containing 165 test vowels including both monophthongs and diphthongs to compare the properties of diphthongs with those of monophthongs. They measured the formant frequencies at 10 points that were equally distributed throughout the vowel period. It was shown that the formant frequencies of vowels in diphthongs were close to those of monophthongs. They also examined whether the Euclidean distance between vowels in diphthongs and the distance between the adjacent points used for formant analysis were associated with the phonological vowel quantity of Estonian[4]. There was a significant effect of vowel quantity and vowel type for the Euclidean distance between vowels, showing that the distance of 'long' diphthong was longer than 'overlong' diphthongs. They also showed that the first element of the diphthong was longer in 'overlong' diphthongs by examining the distance of the adjacent measurement points.

2.7 Factors on diphthong/hiatus differences

By reviewing works on diphthong/hiatus distinction, it is assumed that the major acoustic realisation of diphthongised hiatus is that the vowel in a sequence is realised as semi-vowels, supported by the findings from Spanish studies that the vowel quality is reduced. However, the duration of those semi-vowel realisations are not measured and the study on Kihnu Estonian shows an opposite outcome. The factors on differentiating

---

[4] Estonian diphthongs are classified to Long or overlong quantity degrees.

diphthongs and hiatuses vary between researchers or between languages. Sonority difference between vowels and the stress position are major factors of diphthongisation in Spanish but only sonority difference is applied in Breton. These factors are not examined in Urdu and Kihnu Estonian. It is not clear whether there is a universal factor in the diphthong/hiatus distinction or if there is a language-specific factor for this distinction. Therefore, it is necessary to test all those factors discussed in these studies when arguing the diphthong/hiatus distinction in other languages. However, some of the factors become problematic concerning the target language of the present study, Japanese. For example, Japanese is a pitch-accent language and the fall/rise/flat distinction is not applied unlike Spanish or Breton. Moreover, the semi-vowels in Japanese are considered to be consonants from the phonological point of view. It is possible to combine semi-vowels and monophthongs to make one syllable but semi-vowels always occupy the onset position. The matter of interest in the present study is in the two-vowel sequences. The following sections will discuss the phonological and phonetic properties of Japanese vowels and how the sequences of two monophthong vowels are to be analysed.

2.6 Japanese vowel inventory

According to Lindau, Nolin and Svantesson (1990), Japanese can be classified into one of two thirds of the world’s languages that lack phonemic diphthongs. Japanese is generally believed to have five distinct monophthong vowels (Kawakami, 1977; Vance, 1987; Saito, 1997; Kashima, 2002; Inozuka and Inozuka, 2003; Tsujimura, 2007; Vance, 2008). Although Shibatani (1989) points out that there are variations in some dialects, many researchers agree on the number of vowels in modern standard Japanese, therefore, there is no diphthong as a phoneme in Japanese. Table 4 presents the vowel phonemes of Japanese and their defining characteristics.

Table 4 Japanese Vowel Inventories

phoneme	**roundness**	**height**	**backness**
/i/	unrounded	high	front
/e/	unrounded	mid	front
/a/	unrounded	low	central
/o/	rounded	mid	back
/ ɯ /	unrounded	high	back

Phonetically, these monophthong vowels are short (Vance, 1987; Saito, 1997; Vance, 2008) and characterised as high front, high back, mid front, mid back and low central (Tsujimura, 2007). Therefore it can be assumed that Japanese has a five vowel system and there is no diphthong in the inventory of Japanese vowel phonemes. However, according to the phonological rule of Japanese, these monophthong vowels can be combined freely with one another to construct a vowel sequence. Table 5 below illustrates all the possible combinations of Japanese vowels. It is observed that each vowel can be combined with other vowels as well as the identical vowel. Five vowels can be followed by five vowels and thus there are 25 patterns in total. However, the sequence of identical vowels is produced as a prolonged vowel, in which the vowel quality is kept steady for the length of two rhythmic units (mora). In addition, two non-identical vowel sequences, /ei/ and /oɯ/, can be pronounced as [eː] and [oː] respectively. This is due to the discrepancy between the syllabic characters representing pronunciation and the actual pronunciation Japanese when both sequences are written in *Hiragana*, a standard syllabic character of Japanese. However, Inozuka and Inozuka (2003) point out that their phonetic realisations vary depending on the speech style and Kawakami argues that the sequential pronunciation appears in careful and formal speech. For example, the 4 moraic word meaning “tax” is transcribed as 「ぜいきん」 and its phonemic transcription is /ze.i.ki.n/. Each mora is represented by one character. However, the phonetic realisation of this word is either [ʣeːkɪɴ] or [ʣeɪkɪɴ] depending on the speech style (Kawakami, 1977).The Hiragana letter 「い」 represents the vowel phoneme /i/ and its realisation is [ɪ] in general but when it follows /e/, the vowel quality of /e/ is kept for one mora and the letter represents the different pronunciation. The same phenomenon occurs in the case of /ou/. Thus, in total there are 20 sequences of non-identical vowels in Japanese but two of them can be pronounced as prolonged vowels in an informal speech.

Table 5 Possible Japanese Vowel Sequences

		First Vowel				
		a	i	u	e	o
Second Vowel	a	aa*	ia	ua	ea	oa
	i	ai	ii*	ui	ei**	oi
	u	au	iu	uu*	eu	ou***
	e	ae	ie	ue	ee*	oe
	o	ao	io	uo	eo	oo*

* two identical vowels are produced as long vowels
** /ei/ is often produced as [e:]
*** /ou/ is often produced as [o:]

It should also be noted that there is a morphological condition on the occurrence of those sequences. Takayama (2003) points out that a limited number of vowel sequences can occur within a morpheme. These sequences are shown in Table 6 below. There is no restriction on occurrence if the two vowels occur across a morpheme boundary. Therefore, any combination is possible within compound words where the vowel sequence straddles two morphemes.

Table 6 Vowel Sequences appearing within a morpheme

		First Vowel				
		a	i	u	e	o
Second Vowel	a					
	i	ai		ui		oi
	u					
	e	ae	ie	ue		oe
	o	ao	io	uo		

Phonologically, each monophthong vowel in Japanese occupies one mora, therefore vowel sequences should be analysed as occupying two morae. Figure 4 shows all the possible mora structures in Japanese (Inozuka and Inozuka, 2003). In Japanese, there are six types of mora structures and a vowel sequence can be created when a vowel follows the mora types of 1, 2 and 3 in the figure 3 to make VV, CVV and CjVV (See Figure 5 below).

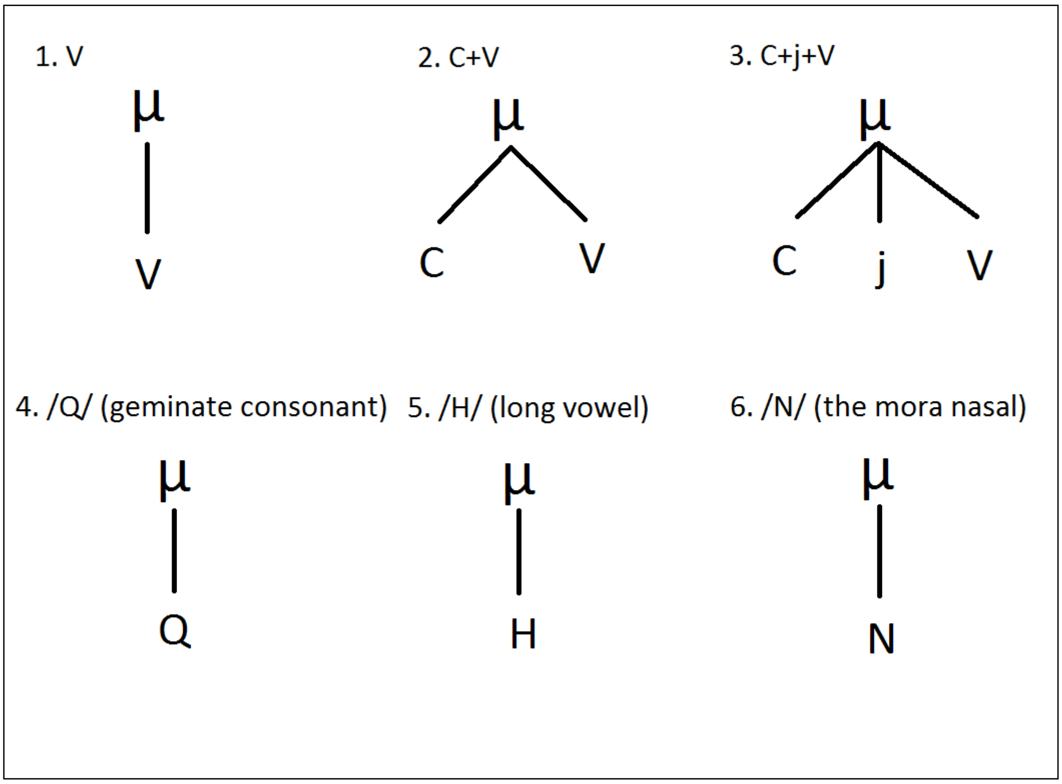


Figure 4 Mora structures of Japanese

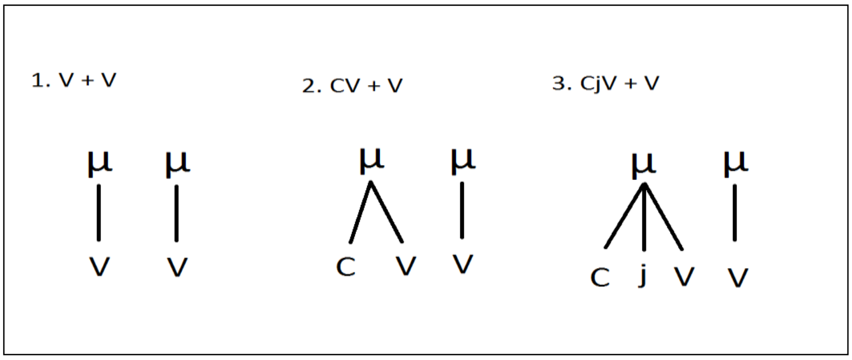


Figure 5 Mora structures of Japanese vowel sequences

As is discussed in the studies on diphthongisation of vowel sequences in Spanish, the phonetic outcome of diphthongised vowel sequences is that one of the vowels in a sequence is realised as a semivowel. In Japanese, two semivowels /j/ and /w/ are considered to be phonemes. Despite the fact that their acoustic realisation is similar to vowels, they are considered to be consonants because they occur only at the consonant position in a syllable or a mora such as /ja/ or /wa/. Moreover, a semivowel /j/ also appears in CjV pattern. This pattern is considered to be a syllable initial consonant cluster (Vance, 2008) or the palatalisation of the consonant (Kashima, 2002; Inozuka and Inozuka, 2003) and the acoustic characteristics of the semivowel /j/ is similar to those of vowels, but its duration is shorter than that of a vowel. Kashima (2002) points out that /j/ is realised as an extremely short [ɪ]. It is also instructed, for example, that [kja] can be produced by trying to shorten the [ɪ] sound in [kɪa] (Kashima 2002, pp. 71-72). Therefore, [kɪokɯ] (memory) is different from [kjokɯ] (bureau). For these reasons, these syllables are not treated as vowel sequences in this study and are not taken into consideration in the research design.

It has commonly been understood that each monophthong vowel occupies one mora and therefore each vowel in two consecutive vowels will occupy one mora (Inozuka and Inozuka, 2003; Koizumi, 2003). However, there is a controversy when they are analysed in terms of the syllable structure. Figure 6 below presents the possible syllable structures in Japanese. The main difference from the mora structure is the status of /H/, /N/ and /Q/. These three morae are called the special mora (Vance, 2008) because they occupy one mora although they do not take up one syllable. /H/ refers to a part of a lengthened vowel. The phonological description of /paH/ will be realised as [paː], which can also be described as [paa]. Because the lengthened part, which is a realisation of /H/, occupies almost the same amount of time in production, the realisation of /H/ is analysed as one mora. But the realisation of /H/ depends on the previous vowels, it is thought as a part of a syllable.

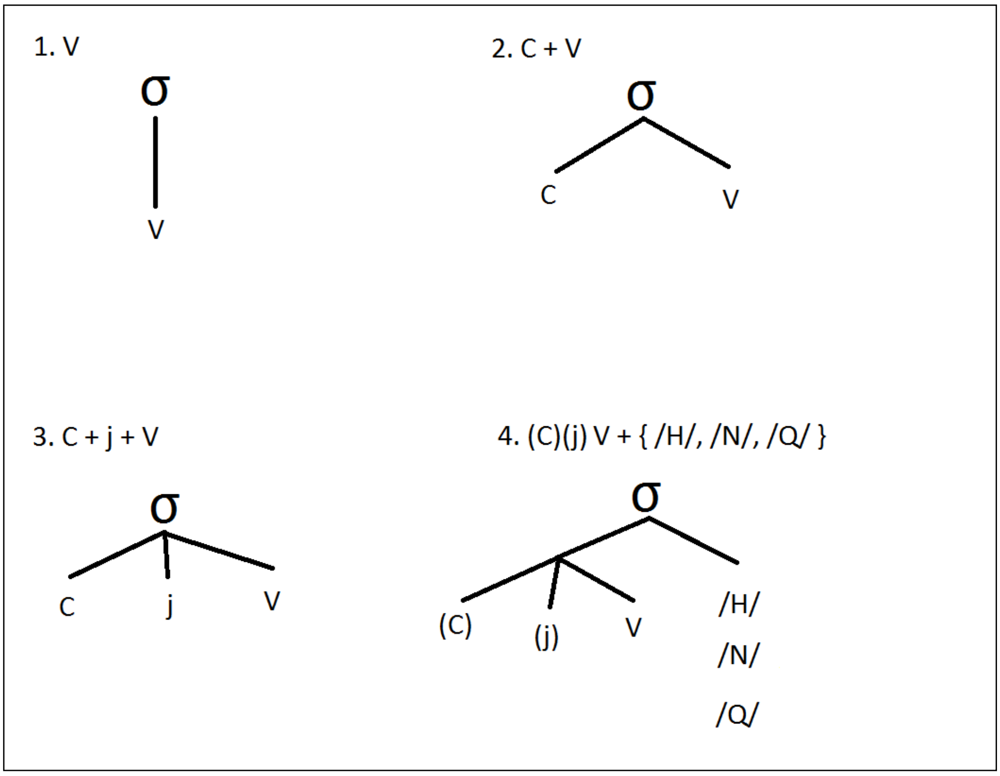


Figure 6 Syllable structures of Japanese

Inozuka and Inozuka (2003) treat the vowel sequence as two syllables and the syllable structure of the vowel sequence can be represented as follows (Figure 7). As is the same with mora structure, two vowels are not permitted to appear in one and the same syllable.

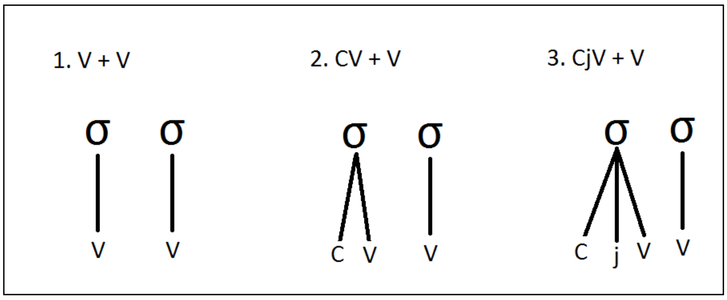


Figure 7 Syllabification of vowel sequences in Japanese

However, Vance (2008) analyses the Japanese vowel sequence as a single syllable (Figure 8). He admits that monophthong vowels can occupy one mora but allows two vowels occurring in one and the same syllable.

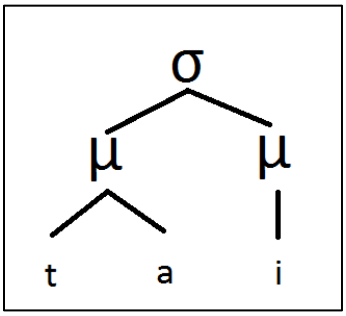


Figure 8 The syllable structure of the Japanese Vowel sequence (cited from Vance, 2008 p. 118)

He also argues that it is not always easy to decide whether the two vowels are in the same syllable or in separate syllables. He notes that it seems safe to say there is a syllable division between the two vowels as long as the second vowel is not /i/ or /u/.

Kubozono (1999) argues that the latter vowel in a two vowel sequence is analysed as one of the special mora (See 4 in Figure 6).

However, it is apparent that five monophthong vowels in Japanese can be produced independently and therefore a pause can be inserted in a very careful speech. For example, the word /ai/ “love” can be pronounced as [a.ı] separately rather than connectively [ai] when speakers intend to speak it more clearly. On the contrary, it is impossible to insert a pause during the production of the English diphthong /ai/ even if speakers would like to emphasise the word ‘I’ in a sentence. The analysis in Kubozono (1999) states that a vowel sequence in Japanese may belong to one and the same syllable but it would raise a problem of the independence of the syllable because any detailed syllable structure of the vowel sequence is not given. It should be a matter of question whether two monophthongs belong to the nucleus position or the former vowel belongs to the nucleus and the latter to the coda.

2.7 Two-vowel sequences in Japanese

In addition to the phonological controversy over the syllable structure of vowel sequences of Japanese, there is controversy over whether two-vowel sequences in Japanese can be classified as diphthongs from the phonetic point of view.

Kawakami (1977) and Saito (1997) argue that some vowel sequences can be produced as diphthongs in normal speech and Takubo, Maekawa, Kubozono, Honda, Shirai and Nakagawa (1998) state that there are falling diphthongs such as /ai/ and /oi/ in Japanese because the first vowel is always accented when the vowel sequence is in the same morpheme. However, Koizumi (2003) explains that the two-vowel sequences and diphthongs are different phonetic phenomenon and gives different terms, such as *ren-boin* (successive vowels) for the former and *nijyu-boin* (diphthongs) for the latter. He states that the two-vowel sequences in Japanese should be analysed as successive vowels rather than diphthongs because each vowel in a sequence keeps its vowel quality.

Labrune (2012) clearly disagree with the view that Japanese vowel sequences are regarded as diphthongs. It is stated that they are a succession of two distinct vocalic nuclei. As evidence, three phonetic and phonological facts are offered. Firstly, the vowel quality of each vowel in a sequence does not differ from isolated production, although no acoustic evidence is given. Secondly, a pause or a glottal stop can be inserted. Thirdly, a vowel sequence in Japanese is not created by the diphthongisation of a monophthong. These statements seem to be true in some sense, but the difference in phonetic features between diphthongs and vowel sequences are not stated nor are the overall phonetic characteristics of either diphthongs or vowel sequences. Inozuka and Inozuka (2003) deny that two-vowel sequences in Japanese are diphthongs because each vowel makes up a different syllable or mora and each vowel can be produced separately, although they admit that two-vowel sequences are likely to be diphthongs when the second element of a sequence is /i/ and state that this vowel sequence will be counted as two morae but one syllable. Kashima (2002) states that Japanese vowel sequences are hiatuses. He also refers to the fact that some researchers claim that the vowel sequences in Japanese are diphthongs.

Therefore, the controversy derives from the phonetic, phonological and morphological interpretation of two-vowel sequences.

From a phonetic point of view, Saito (1997) notes that the movement of articulation from one vowel to the following vowel is quick in the case of vowel sequences. Kawakami (1977) describes a sequence of Japanese vowels as a sequence in which the quality of each vowel is kept steady for a while and the power of each vowel is equal. There is no further explanation of ‘power’ in his literature but it can be interpreted that he means the intensity of the two vowels in a sequence is the same. Inozuka and Inozuka (2003) refer to the possibility that each vowel in a sequence in Japanese can be produced separately.

Phonologically, a sequence of two vowels is to be analysed as two syllables and two morae because a single vowel occupies the nucleus of a syllable or a mora in Japanese (Inozuka et. al 2003) as is discussed in the previous section. However, Kawakami (1977) argues that a sequence of two vowels belongs to the same syllable in natural speech. Inozuka et al. (2003) also say that a sequence of two vowels is likely to be one syllable when the second element is /i/. In addition, Saito (1997) argues that a sequence of vowels is unlikely to become a diphthong when there is a semantic boundary between two vowels. A similar argument is found in Kawakami (1977) but he notes that a diphthong occurs in careful speech. Furthermore, Takayama (2003) points out that the pattern of vowel sequences is limited within a morpheme or a simple word but all possible patterns occur in compounds. However, none of the arguments give acoustic evidence.

The controversial argument of whether vowel sequences in Japanese become diphthongs might have been discussed from the points of view of theoretical backgrounds other than phonetics. Even if there is no word boundary between vowels, more explanation is necessary to characterise the difference between a diphthong and a vowel sequence. Furthermore, when vowel sequences are produced in natural or careful speech and the second vowel of a sequence is /i/, it is not clear what sort of acoustic characteristics distinguishes between a diphthong and a vowel sequence.

2.8 Acoustic properties of two-vowel sequences in Japanese

Some researchers have worked on analysing the acoustic properties of vowel sequences in Japanese. Among the earliest works on the acoustic analysis of vowel sequences in Japanese is the work by Fujisaki and Higuchi (1979). Although they did not intend to address the question of whether Japanese vowel sequences were diphthongs, their work showed that the transition duration between steady state of the first vowel in sequences varied among different pairs of vowels. They collected 100 utterances of Japanese real and nonsense words containing all possible vowel sequences with the 'flat'[5] word accent by one male speaker of the Tokyo dialect of Japanese and measured the duration of the steady state of the first vowel and transition. By comparing the sequence of different orders (e.g. /ai/ and /ia/), the steady states for the first vowels were found to be significantly larger than the second formant trajectory produced by an opening movement of the jaw in e.g. sequence and in those by a lip rounding in e.g. sequence except for /au/ vs. /ua/. This finding implies that the ratios of steady states for the first vowel will be the major part of the vowel sequence and the opening movement will be an important factor for the result. They also mentioned that the onset duration was negatively correlated with the transition duration. However, they did not mention the steady states for the second vowels. Although they conducted a formant analysis, they used it just for calculating the transition distance; formant characteristics of vowels in vowel sequences were not discussed.

To the present, Fujisaki et al. (1979) seems to be the only research dealing with solely the acoustic properties of vowel sequences in Japanese. Other studies mainly focus on comparing acoustic characteristics of vowel sequences in Japanese and those of diphthongs which are thought to be considered to be equivalent, such as Japanese /ai/ and English /ai/.

---

[5] Flat word accent means that the first mora is low and all the following morae are high. 'Flat' accent means that the word does not have an accent kernel, which is realised as the falling of pitch.

Hirasaka and Kamata (1981) compared Japanese /a+i/and /o+i/ produced by 2 Japanese speakers of Tokyo dialect and English /ai/ and /oi/ produced by three General American speakers. They measured the duration of steady states of the first and the second elements of both Japanese vowel sequences and English diphthongs[6]. The measurement of the total diphthong or vowel-sequence duration and the F1, F2 and F3 of both elements were also conducted. In addition, they compared formant frequencies of both elements with those of singleton productions. They concluded that the characteristics of the time and the spectral domains are different between Japanese vowel sequences and English diphthongs. They noted that the duration of the first element is longer than that of the second element in Japanese vowel sequences and the second element of Japanese vowel sequences reached the target area of corresponding monophthongs. On the other hand, the second elements of English diphthongs were very short and their formant frequencies were different from those of equivalent monophthongs. Their research suggests that the durations of elements in vowel sequences of Japanese differ and the formant frequency of each element in a sequence may be similar to that of monophthongs. However, their number of samples is small considering the possible pair of vowels in the Japanese vowel inventory. In addition, they did not refer to the duration of transition between elements and the ratio by the total duration.

---

[6] They divided English diphthongs into the first and the second elements and each section is satisfied with the equation that $df(t)/dt = 0$ using the spectrum analyser YHP-3582A.

Kasuya and Sato (1990) compared the internal structure of Japanese /ai/ and English /ai/ [7]and concluded that they were completely different. Japanese /ai/ was analysed as having 15% steady state of /a/, 40% transition and 45% steady state of /i/ whereas English /ai/ had 41%, 38%, 21% respectively. However, the method of analysis is not clearly described, the accent of /ai/ in their stimuli was not specified and the sample is very small. They did not mention anything about the formant frequencies.

Gore (2006) compared English /ai/ and Japanese /ai/ in different morphological and speech-rate conditions and measured the durations, formant frequencies and amplitudes of subsections. Five male speakers of English (two North East American and three British) were asked to produce the word ‘eye’ four times and Five Japanese speakers[8] (three Kagoshima and two Tokyo) were asked to pronounce /ai/ (‘love’ in Japanese) four times. Morphologically conditioned data were obtained from 10 English speakers and 10 Japanese speakers. Japanese speakers were asked to read out three names of Japanese people which contain vowel sequences in their given names (morpheme-internal) and three names which two vowels across their surnames and given names both slowly and quickly. English speakers were asked to read out three sentences which contain both /ai/ and /a#i/. He reported that Japanese /ai/ in different morphological conditions showed shorter transition than steady states. Transition across

---

[7] Their data were produced by two speakers of Tokyo dialect. Details of English speakers were not specified in the article.

[8] Their genders were not specified.

boundaries is shorter than transition within word and the second element /i/ is significantly longer in the across-boundary condition. Comparing formant frequencies, he reported that F1 was significantly lower in the across-boundary condition whereas F2 showed no significant difference. His research suggests that the morphological condition of vowel sequences influences the duration and the formant frequencies. However, he did not take into consideration the accent pattern of stimuli. In addition, the measurement method of dividing into subsections seemed to be subjective, because his method of dividing vowel sequences into subsections was based on the auditory impression of each vowel in sequences.

Misono and Hirasaka (2008) measured the duration and formant frequencies of three English closing diphthongs /ai/, /oi/ and /au/ and the equivalent Japanese vowel sequences. Three male speakers of Japanese were asked to produce Japanese real words containing those sequences and English real words of ‘eye’, ‘out’, ‘toy’ and ‘hoy’. Three male speakers of General American were asked to pronounce real words ‘eye’, ‘high’, ‘oil’, ‘hoy’, ‘toy’, ‘out’ and ‘house’[9]. They noted that the first and second elements were almost the same in duration or the second element was longer than the first element in Japanese vowel sequences. For English diphthongs produced by Japanese speakers, the first element was longer than the second element. English diphthongs produced by English speakers were found to have longer first elements, but one speaker had longer second elements in ‘hoy’, ‘toy’, ‘out’ and ‘house’. Their formant

---

[9] They do not mention the reason why Japanese and English speakers had different words.

analysis showed that the second element of English diphthongs was centralised in both English and Japanese speakers whereas the second element of Japanese vowel sequences did not show centralisation. Although they concluded that Japanese vowel sequences have some features of diphthongs, more detailed analysis may be necessary.

Table 7 below is the summary of their findings on formant trajectory of the vowel sequences in Japanese. Although each study divides a sequence in three parts, the method of the measurement is slightly different in each case. Hirasaka and Kamata (1981) and Misono and Hirasaka (2008) described only the actual durations of subsections, so the ratios of each subsection to the total sequence duration were calculated based on their data for the purpose of comparison.

Table 7 Comparisons of subsection percentages of vowel sequences of Japanese

references	target	V1	Transition	V2
Hirasaka and Kamata (1981)	ai oi	17.7%	46.3%	36.0%
Kasuya and Sato (1990)	ai	15.0%	40.0%	45.0%
Gore (2006)	ai	33.1%	23.0%	43.8%
Misono and Hirasaka (2008)	ai oi	24.2%	37.2%	38.5%

It is apparent from the table that the number of the target vowel sequences studied is very small considering the number of possible vowel sequences of Japanese. There is less agreement on the ratio of subsections of vowel sequences compared to the results of diphthongs. Transitions seems to be a large part but not necessary the largest part. Comparing V1 and V2, V2 is consistently larger than V1.

It is not still clear what sort of acoustic characteristics diphthongs have, what sort of acoustic characteristics divide between diphthongs and vowel sequences, whether Japanese vowel sequences can be classify into diphthongs from the literature cited above. The same method of analysis seems to be applied to both diphthongs and vowel sequences and both diphthongs and vowel sequences may begin with steady states (onset or V1), transitions and other steady states (offset or V2), although the ratios of those subsections are different among languages, speech rate, speech styles or even between different research. Moreover, only a few diphthongs or vowel sequences have been investigated. The overall characteristics of vowel sequences in Japanese cannot be discussed without the comprehensive analysis to include all the possible sequence patterns with different conditions which has been tested in previous literature.

2.9 Research question

So far, the research on acoustic properties of diphthongs and vowel sequences have been reviewed. There seems to be a clear distinction between them in Spanish and Breton but there is room for further research into this issue in Japanese. Apart from the debate on the possibility of regarding Japanese vowel sequences as diphthongs, phonetic properties of vowel sequences in Japanese have not been investigated comprehensively. Although some researchers state that some vowel sequences can be identified as diphthongs, acoustic evidence is given only in some studies. If some vowel sequences can be regarded as diphthongs but others cannot, some difference should appear from the acoustic analysis. It also leads to find distinctive features between them. Therefore, the research questions can be summarised as follows. Also, not all studies take into consideration the speech style difference of the vowel sequence.

1) What are the acoustic properties of two-vowel sequences in Japanese?
2) To what extent do these vary as a function of the degree of cohesion between the two vowels in the sequence and as a function of different speech styles?
3) How do the findings of this study contribute to the debate in the literature regarding whether Japanese VV sequences can be classified as diphthongs?

To answer these research questions, Japanese VV sequences across various conditions and speech styles were collected from native speakers of the standard Japanese. Chapter 3 explains how the data was collected and analysed.

Chapter 3 Methodology

This Chapter explains the objectives and the details of the experiments conducted, including the range of materials obtained and the participants involved. Also, the method of data analysis is described.

3.1 Introduction

The experimental work described below focuses on the acoustic properties of all possible Japanese vowel sequences across a range of morphological conditions and speech styles based on the analysis of audio recordings of Japanese vowel sequences produced by native speakers. By analysing the vowel quality and the duration of VV sequences and the different elements which make up these sequences a full picture is presented of the phonetic properties of such sequences. Further analysis compares the properties of VV sequences with the relevant monophthongs and with what is known of vowel sequences and diphthongs in other languages in order to address the validity of the claims made by some investigators that some Japanese VV sequences in some environments have diphthong-like properties.

3.2 Data collection

3.2.1. *Description of the data obtained*

The experimental work was designed to generate a data-set permitting analysis of the vowel sequences in Japanese across a range of conditions produced by native Japanese speakers. In addition, the same vowels produced as singletons were recorded for the purpose of comparison. This section explains the conditions applying to the VV sequences, details of the participants and recording procedure. The two sections of the experimental works for the morphological condition and the speech style condition were conducted separately and at the different times. As a result, the volume of data and participants are different between the experiments. 16 native speakers of Japanese (six male and ten female speakers) participated in the recordings designed to address the effect of the morphological condition and 32 subjects (15 male and 17 female speakers) participated in the recordings designed to investigate the influence of different the speech styles. Vowels in singletons were produced by all the participants as a baseline data. The detailed information of participants is described in 3.2.3 below.

3.2.1.1 Baseline data: Vowels produced as singletons

This part of the recording procedure was designed to investigate the vowel qualities of all Japanese vowels (/a/, /i/, /u/, /e/ and /o/) produced as singletons. This provides a baseline for comparison with the realisation of the same vowels in VV sequences. All the participants for both morphological and speech style condition recording sessions were asked to produce each Japanese vowel in isolation three times. Each vowel was printed on a card in Hiragana character. This character represents the phonetic properties for each vowel and is used readily interpretable as such by Japanese speakers. In the recording, participants were presented each vowel symbol on a card and asked to read each one out three times each with their normal voice. The order of the cards was randomised. Each participant was asked to produce 15 singleton tokens (five vowels, 3 times); therefore, 720 singleton vowels (144 tokens for each vowel) were obtained. Among them, 315 tokens were from 21 male speakers and 405 tokens were from 27 female speakers.

3.2.1.2 VV sequences as a function of morphological condition

Based on the previous literature, the morphological condition (i.e. the nature of the boundary between the two vowels which make up a VV sequence) is hypothesised to be one of the factors which could be associated with variation in the phonetic realisation of VV sequences in Japanese. The present study investigated three morphological environments for the VV sequences.

The first morphological condition is VV sequences produced in /hVV/ nonsense environments. This condition was designed to investigate Japanese VV sequences in a CVV environment. A typical syllable structure of Japanese is either V or CV therefore VV sequences can be composed of two single vowels or CV syllable followed by a single vowel. As the consonant in the CV syllable, the voiceless glottal fricative /h/ was used as it is thought to have no effect on formant structures of the following vowel (Peterson and Barney, 1952). Therefore, the advantage of inserting VVs in the /h__/ environment is to reduce the risk of the consonant-induced formant transitions in the onset of the following vowels thereby providing maximum clarity regarding the properties of the VV sequence. However, it needs to be borne in mind in interpreting the findings from this experimental condition that there is a potential drawback in identifying the starting point of the VV sequence in this context because of the likelihood that the production of the articulatory gesture for the vowels is prepared while the /h/ sound is being produced. 16 Participants (six male and ten female speakers) were asked to produce all[10] possible vowel sequences in /hVV/ nonsense words with both High-Low and Low-High accent contour three times, therefore each participant produced 120 sequences (20 sequences × two accent contours, three times). In total, 1920 tokens were obtained of which 1690 tokens were used in the analysis. Tokens were omitted from the analysis where background noise rendered the analysis difficult to interpret or where speakers had produced an audible gap between the vowels.

---

[10] 20 non-identical vowel sequences are possible by five vowels of Japanese. See Table 5 above.

Also, two sequences /ei/ and /ou/ were omitted for analysis because they were produced more as long singleton vowels rather than two non-identical sequences. /hVV/ nonsense words are transcribed in two Hiragana characters, one for the former /hV/ syllables and the other for the following /V/ syllable. Accent patterns were specified by defining a circled character as High-accented syllable.

The second morphological condition is VV sequences within real words. This condition was designed to investigate whether the acoustic characteristics of Japanese VV sequences were different if both vowels are contained within a single word. 16 Participants were asked to read out three times real words of Japanese printed on a card, written in Hiragana and Kanji characters so that participants could recognise the word meaning and produce them in normal accent type. There were 80 words to be produced for each participants and the word list is shown in Appendix 1. The real words were chosen on the condition that the vowel sequences in these words were produced with one of four different accent contours (High -High, High-Low, Low-High and Low-Low). This was achieved by choosing words to be produced which consisted of greater than three morae, thereby generating the full range of accent contours. Saito (1996) explains that the number of accent pattern of words in Japanese, which is realised by the assignment of High or Low tone for each mora, is the same as the number of morae in a word. For example, two moraic words can be produced with two patterns (High-Low or Low High) but three moraic words can be produced with three patterns (High-Low-Low, Low-High-High and Low-High-Low). Therefore, 20 vowel sequences were combined with 4 accent contours and each participant was asked to produce 240 tokens (20 sequences × four accent contours, three times). In total, 3840 tokens were obtained but 3426 tokens were used for the analysis due to the background noise on the recording and omitting two sequences.

The third morphological condition considered VV sequences across word boundaries, that is, where one vowel is located at the end of the word and the following vowel is positioned at the beginning of the following word. This condition was designed to investigate the case in which the two vowels in a VV sequence belong to different words. 16 participants were asked to produce two words in a connected sequence. The stimuli were Japanese names of people in which a vowel comes at the end of the surname and another vowel at the beginning of the first name so that vowel sequences were across word boundaries and the full name of a person is as such that the family name first followed by the first name, which is the ordinary way of representing names in Japan. This type of stimuli was adopted in Gore (2006). In the present study, five family names ending with each vowel and five first names starting with each vowel (See Appendix 2) were prepared and 20 full names with non-identical vowels at the boundary were combined. Also, family and given names were selected on the condition that the target vowels are produced with High tone so that both vowels were to be pronounced High accent to obtain High-High accent contour for VV sequences. All the names were written in Hiragana characters and participants were asked to read out 20 full names of people. Therefore, each participant was asked to produce 60 tokens (20 sequences, three times). In total, 960 tokens were obtained but 840 tokens were used for the analysis because of the background noise and the omission of two sequences.

3.2.1.3 VV sequences across the different speech styles

In addition to the degree of cohesion between the vowels in a VV sequence as indexed by the morphological conditions investigated in this study, speech style is also hypothesised to be one of factors to vary the phonetic realisation of VV sequences in

Japanese (Kawakami, 1977; Saito, 1997). The specific aim of this part of the experimental work reported here was to investigate VV sequences in a more ‘natural’ condition and to compare them with those produced in more ‘careful’ speech styles, such as read speech. The approach to obtaining speech produced in different styles was informed by Llisterri’s (1992) critical overview of methods for sampling different styles. This set of recordings was undertaken at a different time from the recordings designed to investigate the influence of different morphological conditions. Participants were asked to produce vowels in singletons and then VV sequences in three different speech styles. Therefore, the number of participants is different from the morphological condition.

The first speech style condition is VV sequences produced within words which are in responses to specific questions. Participants are expected to state a word containing a target VV sequence for the response to the questions they are addressed. In the present study, 20 words which contain one VV sequence and the questions designed to elicit those words were prepared so that all possible sequences were covered (see Appendix 3). All the words were chosen so that VV sequences in those words were produced with the High-High accent contour. In the recording session, 32 participants

(15 male and 17 female speakers) were asked to answer each question. If the answer was not as expected, they received hints to guide them to the answer. Some participants produced the word more than twice and all VV productions were included in the analysis. In total, 658 tokens were obtained but 595 tokens were used for the analysis because of the omitting of two sequences.

The second speech style investigated is where VV sequences are produced within words generated in a picture-description task. 32 Participants were asked to describe the pictures of certain objects (See Appendix 4) the labels for which contain 8 target VV sequences. Target words were designed to investigate whether a sub-set of VV sequences which have been argued to be diphthong-like in previous research actually do have features of diphthongs. Tamaoka and Makioka (2004) argued that two continuous vowels that change from an open vowel to a close vowel were to be called diphthongs and those Japanese ‘diphthongs’ were found to be 84.48 % of the total number of VV sequences in the lexical corpus created by Amano and Kondo (2000), using the Asahi Newspaper published in Japan. The sequences which change from open to close vowels are the following eight patterns; /ai/, /au/, /ae/, /oi/, /ou/, /oe/, /ei/ and /eu/. However, it should be noted that /ei/ and /ou/ tend to be long monophthong vowels ([e:] and [o:]) rather than changing from the first vowel to the second one especially in natural speech (Tamaoka and Makioka, 2004). For this reason, /ei/ and /ou/ were omitted for the analysis. Other researchers suggest that [ui] and [iu] can also be diphthongs (Hattori, 1967; Saito, 1997; Kubozono, 2001). Therefore, the following 8 sequences, [ai], [au], [ae], [iu], [ui], [eu], [oi] and [oe] are thought to cover the majority of Japanese VV sequences that are likely to be produced as diphthong-like. The stimuli were arranged so that VV sequences were produced with High-High accent contour. Each participant produced each VV sequence at least once for one picture but the number of VV tokens varied among participants. In total, 405 tokens were obtained but 328 tokens were used for the analysis because of the omission of the two sequences.

The third speech style condition investigated is VV sequences produced within words in natural unscripted conversation. This condition was designed to investigate VV sequences in an uncontrolled casual conversation. There were approximately ten minute breaks between recordings for each of the conditions described above. The conversations between participants and the experimenter at those times were recorded to investigate whether VV sequences in natural conversation are different from VVs in experimental environment (speakers had given their consent to this in advance). There was no particular target sequence but participants were sometimes asked questions the answer to which might contain VV sequences. The VV sequences to be analysed are those within words. In total, there were 1690 tokens but 1675 tokens were used for the analysis because of the omission to two VV sequences.

3.2.2 *Recordings*

For both morphological and speech style conditions, recordings were made in a quiet room, but not under studio conditions, using an EDIROL R-09; a 24-bit, 44.1 kHz WAVE recorder and a built-in microphone. Only participant and the experimenter were in the recording rooms.

3.2.3 *Participants*

The recording relating to the morphological conditions (3.2.1.2) and the speech style conditions (3.2.1.3) were conducted at different times and with different subjects.

The material relating to the different morphological conditions (3.2.1.2) was produced by 6 male and 10 female speakers of Japanese who were born and brought up in areas where the standard Japanese is spoken. The material relating to the different speech styles (3.2.1.3) were produced by 15 male and 17 female speakers of Japanese from the same areas. Singleton data (3.2.1.1) was obtained from both groups; therefore, it was obtained from 21 male and 27 female speakers. The area where Standard (Tokyo) Japanese is spoken is shown in Figure 9.

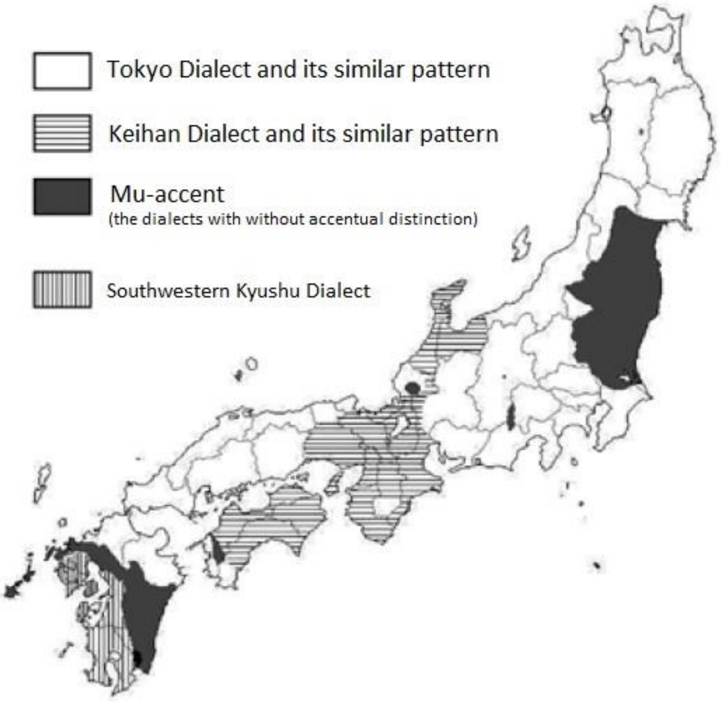


Figure 9 Dialect map of Japan. Tokyo Dialect is generally spoken in the white area (cited from Takayama, 2010)

Participants were asked to sign a consent form (See Appendix 5) after the explanation of the detail of the experiment. It was explained to them that they had the right to withdraw from the study at any times and that all the recording data were used for the research purpose only and would be anonymised. Also, they were given a memento from the UK for their kind participation.

3.3 Data analysis

3.3.1 *Segmenting procedure*

All the recorded data were clipped to individual sound files by the digital audio software, Cool Edit version 96, and all tokens were processed by means of *Praat* version 5.0.05, using a spectrogram display set to a 10 msec window length for male speakers' productions and a 5 msec window length for female speakers' productions and a 40 dB dynamic range as a default setting. But due to variability in the background noise present in some of the recordings, the dynamic range setting was occasionally adjusted in order to obtain a clear visual representation.

Firstly, the start and the end of the VV sequences were determined to measure the total VV sequence duration. In the present study, the start of the vowel sequences is determined as the leftmost point where formant structure of V1 is recognised on the spectrogram. Identifying the end of a vowel sequence is sometimes problematic because it is not often the case that F1, F2 and F3 end simultaneously. Therefore, the end of the vowel sequence is determined by the point by which F2 for V2 reduces in amplitude and the regularity of the waveform of vowels ceases. An example of identifying the starting and ending points of a VV sequence is shown in Figure 10. In the tier below the spectrogram window, those points are presented.

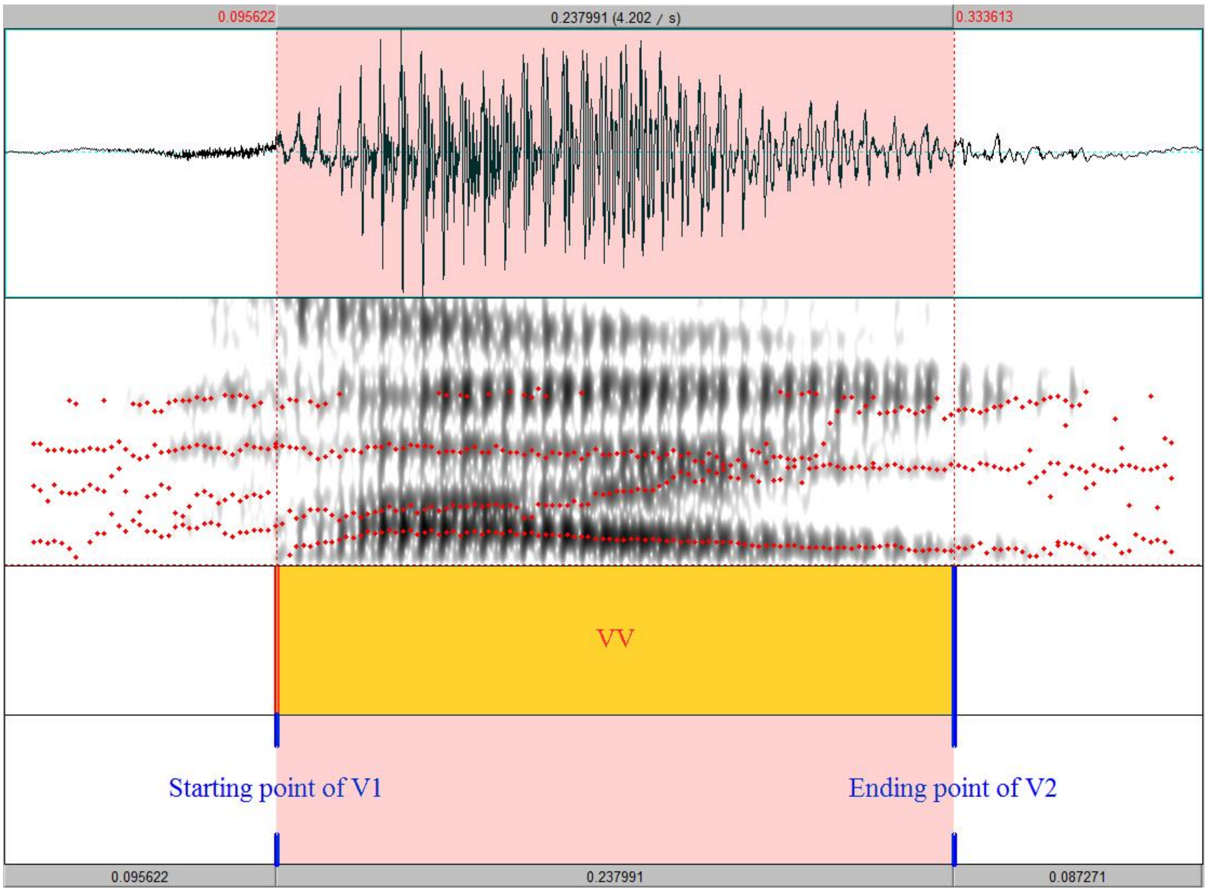


Figure 10 An example of identifying the starting and ending points ([ai] in /hai/ by a male speaker)

Having identified the starting and the ending points for the VV sequences, the start and the end of the formant movement associated with producing two vowels continuously were identified. Following Gay (1968) for his segmenting method for analysing the diphthong interval, three intervals within the VV sequences were identified, two steady states and a transition between the two. As is described in the literature review chapter, this method of segmenting is widely used among researchers dealing with the acoustic properties of both diphthongs and VV sequences. In the case of analysing VV sequences, that is, two monophthongs produced sequentially, the questions which arise are whether there is evidence for one or both vowels being produced with a steady state target, and what is the nature and duration of the transition between the two vowels. The main aim of the analysis is to identify any steady state intervals for V1 and V2 and the transition from any V1 to V2 steady states. It is generally said that the tongue height relates to F1 and tongue forwardness to F2, therefore either F1 or F2 or both gradually change as the tongue moves from the first vowel to the second vowel. In the present study, the list of F2 values which are extracted by Praat at regular intervals was used to capture the F2 trajectory throughout the total VV sequence intervals and to identify the start and the end of the transition intervals from the list of F2 values throughout the VV sequence intervals. The reason that the movement of F2 value was used is that for majority of cases the values extracted at the sampling points change more greatly than F1 for the transition intervals. F1 movement was alternatively used to define the transition interval only when there was no obvious movement on F2 trajectory. The start of the transition was defined as a point where the formant value started to change continuously in the extracted formant list. As is the nature of sounds, the formant value is not exactly steady throughout the identical vowel production. But the changes of formant in ‘steady’ periods are irregular and not continuous. The period where the formant frequencies change continuously in a particular dimension is defined as a transition. Therefore, the end of the transition is defined as the point where the continuous change in formant frequencies ceases in the list of extracted formants. The formant track display of Praat was also referred to for confirmation. As an example, Figure 11 presents the start and the end of the transition. These points were identified by the list of F2 values extracted by Praat. From the starting point of V1, F2 values keeps almost steady at 1320Hz until 198 msec. Then the value gradually increases until 256 msec and keeps almost steady at 2200 Hz. Therefore the period between 198 msec and 256 msec was identified as the transition.

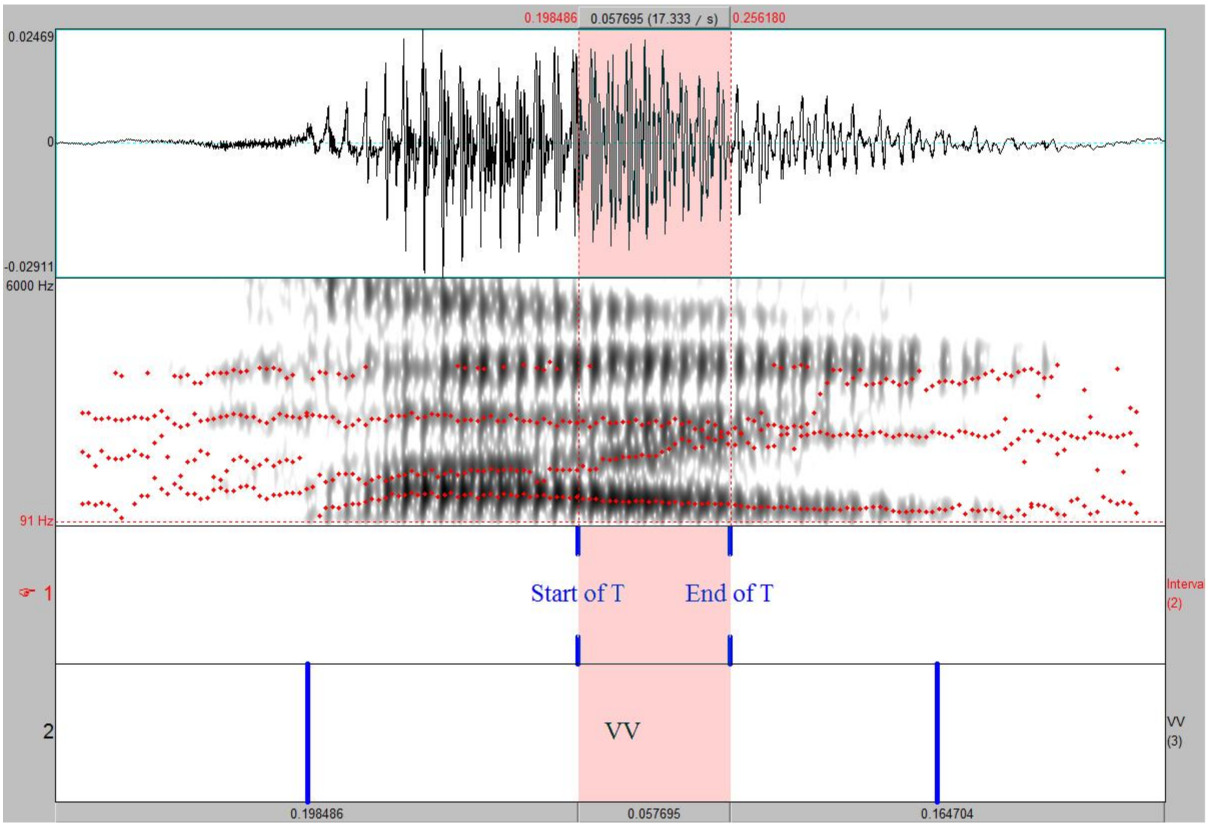


Figure 11 An example of identifying transition ([ai] in /hai/ by a male speaker)

By identifying the start and the end of the VV sequences and the start and the end of the transition intervals, the durations of three intervals can be measured. The interval between the start of the VV sequences and the start of the transition is the steady states for the first vowel (V1), the interval between the start and end of the transition is the transition from the first to the second vowels and the interval between the end of the transition and the end of the VV sequences is the steady states for the second vowel (V2). In order to control for the fact that speech rate was not regulated at the recording sessions; the ratios of these three intervals to the overall duration for VV sequences were calculated and these results from the bulk of the time domain results presented in Chapter 4.

Then, F1 and F2 for the vowels in VV sequences were measured at the mid points of the steady states. Also, F1 and F2 for the vowels in singletons were measured at the mid points of the vowel intervals. The results for the spectral domain analysis are reported in Chapter 5.

3.3.2 *Statistical analysis*

Formant frequencies measured at the mid points of steady states of the vowel sequences and the duration of the first steady state, transition and the second steady states were across the various conditions built in to the design of the experiment.

A Multivariate analysis of variance (MANOVA) was conducted for the time-domain analysis to investigate whether certain factors (morphological and speech-style conditions) have a significant effect on the ratios of the V1, the transition and the V2 intervals to the total VV sequence durations. These subsection ratios are associated with each other and therefore it is thought to be appropriate to treat these proportions as dependent variables within a MANOVA. The independent variables are the morphological condition with three levels (/hVV/ nonsense words, word-internal, across-word-boundary), the accent patterns of the vowel sequence with four (HH, HL, LH, LL), the position in a sequence (V1, V2), the sonority scale (falling, rising, level), the degree of opening/closing, the degree of fronting/backing as well as the vowel identity. These latter two variables are described in greater detail in Chapter 4.

For the formant frequencies, F1 and F2 were analysed separately because F1 is thought to be associated to the tongue height and F2 to the tongue advancement. An analysis of Variance (ANOVA) was conducted to investigate the extent to which these parameters were affected by the same experimental conditions as referred to above for the durational measures.

To find a significant association between the classes for the independent variables, the Scheffe post-hoc test was used.

It has been common to report effect sizes for statics rather than just reporting the significance. The Omega squared (ω^2) estimator is used to present the strength of the association because Grissom and Kim (2005) argue that it is nearly unbiased compared with eta-squared (η^2) estimator. Therefore, ω^2 estimator is used as an effect size for the MANOVA and the AVOVA in this study. For the estimation of the post-hoc tests, Cohen's d is used because it can be calculated by the mean values, the standard deviations and the number of cases of two comparing groups. The effect size was evaluated by following the interpreting scheme given in Kotrlik and Williams (2003). For the omega squared estimator .010 is defined as a small effect size, .059 is as medium and .138 is as large., For Cohen's d, .20 is defined as a small effect size, .50 is as medium and .80 is as large.

The following two chapters report the results of the analysis, whose method for the measurement is explained in this chapter. Chapter 4 explains the results of the time-domain analysis, the total VV duration, the ratios of the V1, the transition and the V2 intervals to the total VV duration and the correlation between the total VV duration and the subsection ratios. Chapter 5 explains the results of the spectral-domain analysis, formant frequencies of the vowels in VV sequences, their variation across different conditions and the transition speed calculated by dividing the Euclidian distance between the start and the end of the transition by the transition duration.

3.3.3 Reliability test

In order to test the reliability of the measurements, the formant and durational measurements were repeated for 200 tokens sampling randomly across speakers and conditions. To get an indication of the consistency with which the segmentation criteria were applied, the pooled measures were subject to a correlation analysis with the following results. For formant measures (F1/2 for each of V1 and V2), there was an overall correlation of the original and remeasured values of .97 indicating a very high level of agreement. For the duration measurements (VV, V1, V2) the overall correlation was a slightly lower at .95. Further inspection of the duration data suggested that V2 in HVV sequences proved to be less consistently measured than other durational parameters due it would appear to uncertainty in identifying the end of V2 in a gradual shift from the vowel nucleus to silence. As a result, the HVV durational data needs to be interpreted with a degree of caution. Since the transition duration measurements are derived from the durations of V1 and V2, they are unavoidably subject to the influence of both of these primary measures with the result that their variance is somewhat higher at .48. This is perhaps not surprising given that, for example, a small re-measurement difference in opposing directions on each of V1 and V2 can readily add up to a larger difference in the transition duration measurement. So, here too the findings needs to be interpreted with caution.

Chapter 4 Time Domain Analysis

4.1 Introduction

This chapter presents the results of the time domain analysis of vowel sequences in Japanese. The total duration of the vowel sequences and the ratios of V1, Transition and V2 to the total VV durations are the principal focus of the presentation. Previous studies have presented the total vowel interval durations and the durations of subsections of the total vowel interval (V1/onset steady state, transition and V2/offset steady state) to describe the characteristics of diphthongs or hiatuses and they have argued that the differences between diphthongs and hiatuses are evident in these characteristics. As was mentioned in the methodology chapter, in the present study the speech rate of the speakers was not controlled. We recognise that the speech rate may influence on the timing features of vowel sequences and the duration should be normalised enable analysis of the features common to participants. For this reason, the results for the subsections of VV sequences are presented in terms of the relative proportions of the overall VV interval occupied by V1, Transition and V2.

4. 2 Total vowel sequence duration

This section reports the differences in durations of VV sequences depending on the types of VV sequences and speech styles. As the speech rate was not controlled during the data collection, the duration of the VV sequences varied in different conditions and different subjects.

4.2.1 *Morphological condition*

As was shown in chapter 2, the total vowel sequence duration is one of the acoustic characteristics highlighted in previous work as distinguishing diphthongs and hiatuses. For example, it has been observed that diphthongs are shorter than hiatuses in studies of Breton and Spanish (Aguilar, 1999; Garrido, 2007; Face et al., 2004). This section presents the total vowel sequence duration in Japanese and investigates whether differences in duration across different morphological conditions are consistent with the view that in some conditions Japanese vowel sequences may be more diphthong-like. As is explained in the methodology, the present study investigates VV sequences in three different morphological conditions to see whether the morphological conditions of the sequences affect the duration. From now on, the nonsense words which vowel sequences are inserted in /h__/ environment will be referred to as /hVV/, the vowel sequences in real words will be word-internal (WI), and the vowel sequences in which V1 is at the end of a word V2 is at the beginning of the following words will be referred to as vowel sequences across a word boundary (AWB). The summary of total duration measurements is presented in Table 8 below in terms of mean durations, standard deviations and ranges. For the purpose of capturing the general properties of the duration of the vowel sequences, male and female data are pooled. The durations are represented in milliseconds (msec).

Table 8 Durational measurement on vowel sequences in /hVV/, WI and AWB (in msec)

	/hVV/	**WI**	**AWB**
Numbers of tokens	1690	3426	840
Mean	374.80	269.82	210.99
S. D	108.69	72.38	54.64
Range	127.37-682.73	88.38-556.00	82.65-365.89

As discussed in Chapter 2, morphological condition has been identified in previous studies as one of the factors that may lead vowel sequences in Japanese to have features of diphthongs (Kawakami 1977, Saito 1997). These investigators argue that VV sequences appear to be more like diphthongs when there is no 'semantic boundary' between two consecutive vowels. Thus, the present results permit a test of how the unity of the VV sequences, that is, the morphology of VV sequences affect their temporal characteristics. However, the present results show that VV sequences which have a 'semantic boundary' between the vowels are arguably more 'diphthong-like' than other morphological conditions because the vowel sequences across a word boundary were shorter than VV sequences of other morphological conditions. The mean duration of VVs in the WI condition was 71% of that for the /hVV/ condition and that of the AWB condition was 56% of the /hVV/ tokens. A One-way ANOVA with grouping factor 'morphological condition' (/hVV/, WI, AWB) revealed significant differences (*F* (2, 5955) = 1383.08, $p < .001$). However, its effect size represented by eta squared was small ($\omega^2 = .274$). Scheffe post-hoc tests revealed that the total duration of each morphological condition was significantly different from each other. Duration of /hVV/ nonsense words was significantly longer than WI ($p < .001$, $d = 1.24$) and AWB (p

< .001, d = 1.81) and duration of VVs within words was significantly longer than VVs across word boundary (p < .001, d = 0.85). The difference between the /hVV/ condition and the other two may well be a result of the fact that the /hVV/ tokens were nonsense words while the words in other conditions were real words, leading the latter to be produced more fluently and/or the former to be produced more slowly overall. Comparing VV sequences in WI and AWB conditions, unlike previous studies which have argued that VV sequences without semantic boundaries between vowels have diphthong-like features, the current results show that the diphthong-like feature of relatively shorter total VV duration appears more reliably in VVs across a word boundary than in VVs within words.

Secondly, the accent associated with a VV sequence may also be a factor associated with Japanese vowel sequences being more or less as diphthong-like. Misono and Hirasaka (2008) pointed out that the accent difference of VV sequences should be taken into account because VVs produced with a High-Low accent are expected to be more like ‘diphthongs’ than VVs with other accent types. They also mentioned that none of the research they could find referred to the accent type differences of VV sequences. Kawakami (1977) did not directly refer to the accent type but he implied that VVs with a High-Low accent type become more diphthong-like than VVs with a Low-High accent. Therefore, the following results test whether the accent type is a significant factor in respect of variability in the relative durations of VV sequence subsections in the present dataset. The mean VV durations of VV sequences produced with HH, HL, LH and LL accent types are shown in Table 9 below.

Table 9 Durational measurement on vowel sequences with four accent patterns (in msec)

	HH	**HL**	**LH**	**LL**
Number of tokens	1698	1688	1703	869
Mean	244.99	309.75	337.79	254.52
S.D	74.06	96.83	110	70.92

VV sequences bearing the same tone across both Vowels are shorter than those with different tones. Comparing the vowel sequences with HL and LH accent types, those with HL accent were shorter than those with LH accent types and this result supports Kawakami (1977) and Misono and Hirasaka (2008). However, the vowel sequences with the same tones (HH and LL) were much shorter than VVs with different tones. A One-way AVOVA with grouping factor ‘accent’ (HH, HL, LH, LL) revealed significant differences (*F* (3, 5954) = 359.49, p < .001), although its effect size was small (ω^2= .145). Scheffe post-hoc tests showed that all the pairwise comparisons were significantly different (p < .001) other than HH /LL and HL/LH comparisons. Effect sizes represented by Cohen’s d were large in HH/LH (d = -1.01) and LH/LL (d = 0.86) comparisons and medium in HH/HL (d = -0.76) and HL/LL (d = 0.63) comparisons. It was found to be true that VV sequences bearing the HL accent were more diphthong-like than VVs with the LH accent, but VV sequences with the HH and LL accent types were much more diphthong-like. The tonal change from V1 to V2 seems to be associated with the long duration and the tonal change from Low to High makes the total duration much longer. The effort to rise and fall pitch may relate to the difference in duration.

Thirdly, previous work has suggested that the sonority scale between the two vowels is a factor which might distinguish between hiatus and diphthongal realisation. Sonority is said to be related to the vowel height and, based on his work on Breton and Spanish, Carlyle (1987) suggested that vowel sequences can be categorised into the following groups: 1) If the first vowel is more sonorous than the second, then it is a falling sequence. 2) If the first vowel is less sonorous than the second, then it is a rising sequence. 3) If the sonority of the first and the second vowel is equal, then it is a level sequence. However, as is noted in the data analysis section, VV sequences that can be realised as long monophthongs were omitted. Therefore, to apply this categorisation to Japanese VV sequences, the 18 sequences in Japanese can be grouped into as follows.

1) Falling sequences: ai, au, ae, ao, eu, oi
2) Rising sequences: ia, ie, io, ua, ue, uo, ea, oa
3) Level sequences: iu, ui, eo, oe

However, a One-way ANOVA showed no significant difference in terms of sonority (*F* (2, 5955) = 0.144, p = .866 > .05). Thus Carlyle's findings for diphthongs and hiatuses in Breton and Spanish do not appear to apply to Japanese vowel sequences.

However, he also pointed out that the sonority scale may vary across languages. The mean duration of each sonority scale is presented in Table 10 below.

Table 10 Durational measurement on vowel sequences with three degrees of the sonority scale (in msec)

	Falling	**Rising**	**Level**
Number of tokens	2014	2575	1369
Mean	290.31	291.58	292.03
S.D	98.89	98.66	102.65

In parallel with the sonority scale analysis, the extent of the opening /closing movement between the two vowels of VV sequences can also be used to classify vowel sequences in Japanese. Kubozono (1999) states that the vowel sequences (V1V2) in which V1 is an open vowel and moves to a close vowel are diphthongs and that V1V2 sequences which moves from a close vowel to an open vowel are vowel sequences. However, in his later paper he gave /ui/ and /iu/ as an example of VV sequences having diphthong-like features (Kubozono, 2001). In addition, some researchers referred to particular VV sequences in Japanese which are likely to be realised as more diphthong-like (Hattori, 1967; Saito, 1997; Kubozono, 2001) but it is difficult to describe what kind of characteristics these VVs have in common. Table 11 below presents the list of VV sequences which are considered to be more diphthong-like by different researchers.

Table 11 List of VV sequences which is more likely to be diphthong-like

Hattori (1967)	ai	ui	oi	ei	-	-	-	-
Saito (1997)	ai	ui	oi	-	ae	au	-	-
Kubozono (2001)	ai	ui	oi	ei	-	au	iu	ou

In the explanation of Kubozono (1999), /oi/ and /ei/ are not considered to be diphthong-like because the first vowel is not an open vowel, but these vowels are thought to be diphthong-like by other researchers. Also, many of them can be characterised as ‘closing’ sequences because the first vowel is more open than the second ones. However, not all VV sequences having ‘closing’ features are considered to be diphthong-like. For example, /e/ is more open than /u/ but /eu/ is not considered to be diphthong-like. Also, both /i/ and /u/ are close vowels but /ui/ is considered to be diphthong-like. Therefore, the articulatory movement from V1 and V2 needs to be systematically classified to explain which VV sequences are more likely to be diphthong-like. In the present study, 18 VV sequences of Japanese are classified to five levels in terms of their opening/closing characteristics. Depending on the degrees of opening/closing movements between two vowels, 18 sequences can be grouped into five classes as follows. Degree 2 sequences move between open to close position in the vowel triangle in Japanese (See Literature review Chapter) and Degree 1 sequences move from or end with the mid position.

1) Degree 2 closing (Close 2): ai, au
2) Degree 1 closing (Close 1): ae, ao, eu, oi

3) No difference (N.D): iu, ui, eo, oe

4) Degree 1 opening (Open 1): ie, io, ue, uo, ea, oa

5) Degree 2 opening (Open 2): ia, ua

However, a one-way ANOVA showed no difference in total VV sequence duration of the open/close category (*F* (4, 5953) = 1.277. p = .276 > .05). Durations of total VV sequence were not different from each other in this category. The mean durations of each level of the category is shown in Table 12 below.

Table 12 Durational measurement on vowel sequences with fife degrees of the open/close dimension (in msec)

	Close 2	**Close 1**	**N.D**	**Open 1**	**Open 2**
Number of tokens	676	1338	1369	1918	657
Mean	292.29	289.32	292.03	293.99	284.52
S.D	95.51	100.58	102.65	96.46	104.57

In addition, the Front/Back dimension was investigated. The 18 possible VV sequences were grouped depending on the degrees of fronting/backing movements between two vowels. Degree 2 sequences move between the front and back positions in the vowel triangle in Japanese and Degree 1 sequences move from or end with the mid position.

1) Degree 2 fronting (Front 2): ui, ue, oi, oe

2) Degree 1 fronting (Front 1): ai, ae, ua, oa

3) No difference (N.D): ie, uo

4) Degree 1 backing (Back 1): au, ao, ia, ea

5) Degree 2 backing (Back 2): iu, io, eu, eo

A one-way ANOVA with front/back grouping factor revealed that there was a significant difference in total sequence duration (F (4, 5953) = 4.599, p =.001), although its effect size was very small ($\omega^2 = .002$). A Scheffe post-hoc test showed that Degree 1 backing sequences were significantly longer than Degree 2 fronting sequences (p = 0.003, d = -0.16). No other pairwise comparison showed significant difference. The mean durations of each level of the category is shown in Table 13 below.

Table 13 Durational measurement on vowel sequences with five degrees of Front/Back dimension (in msec)

	Front 2	**Front 1**	**N.D**	**Back 1**	**Back 2**
Number of tokens	1326	1343	619	1308	1362
Mean	282.8	288.9	292.4	298.28	294.54
S.D	93.23	98.3	99.93	99.84	105.99

Figure 12 below summarises the differences in total vowel sequence durations of each experimental condition described above. Despite the fact that the effect size of the statistics is small, there was a significant difference in total vowel sequence duration for morphological condition, accent types and the front/back dimension. If we adopt the position that the shorter overall duration equates to more diphthong-like vowel sequences, the VV sequences across word boundary, the VV sequences of the same tone and the VV sequences with backward movement may present some diphthong-like features not found in other sequences because the total duration of the VV sequences in those conditions were significantly shorter than other conditions.

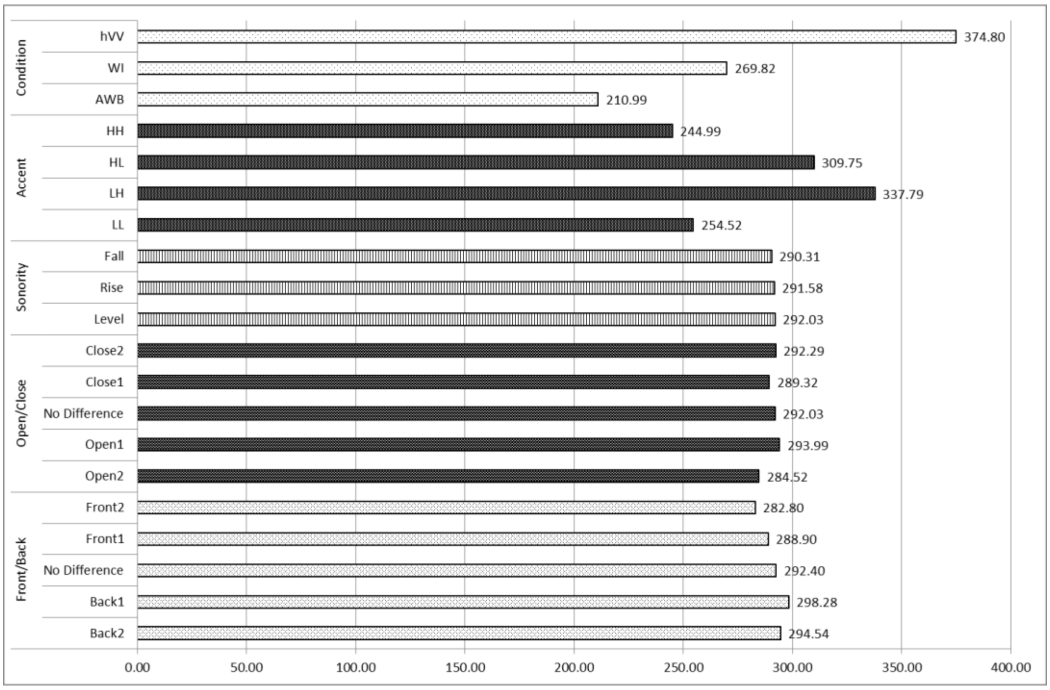


Figure 12 Duration means compared in five categories (in msec)

4.2.2 *Speech style condition*

In addition to the morphological condition, it has also been argued that differences in speech style are associated with vowel sequences having diphthong-like features. Kawakami (1977) argued that VV sequences in general become diphthongs in natural speech but not in careful speech. Saito (1997) stated that the realisation of VV sequences as diphthongs is speech style-related and diphthong realisations appear when VV sequences are pronounced ‘smoothly’ but not when they are produced carefully. Thus, this section presents the result of the time-domain analysis of VV sequences produced in different speech styles. As explained in the methodology, for the present purposes, those VV sequences produced in the within-words morphological condition described above, are here re-categorised as a ‘read’ speech, and compared with the vowel sequences within words produced in more ‘casual’ or ‘spontaneous’ speech styles. As was also noted in the methodology chapter, the participants for morphological and speech style conditions were different because these recordings were conducted at different times. It is recognised therefore that in the presentation of the speech style results, like is not always being compared to like, and this needs to be borne in mind in our interpretation of the findings.

Initially, the summary of total duration measurements is presented in Table 14 below in terms of mean durations, standard deviations and ranges. For the purpose of capturing the general timing feature, male and female data were pooled.

Table 14 Durational measurement on vowel sequences within words in 4 speech styles (in msec)

	Read	**Response**	**Description**	**Conversation**
Number of tokens	3431	595	328	1675
Mean	269.82	197.64	185.62	154.91
S.D	72.38	53.83	58.75	57.68
Range	88.38-556.00	61.20-385.24	80.99-457.03	42.75-597.66

In comparison with read speech, the total duration was reduced by 38% for the response style, 51% for the description speech and 74% for the conversation style. Kawakami (1977) and Saito (1997) argued that more ‘natural, and ‘spontaneous’ speeches had VV sequences in Japanese more diphthong-like. The present result shows that the total VV sequence duration in the response, description and conversation styles are shorter than VV sequences of read speech. This finding implies that the VV sequences in more ‘natural’ and ‘spontaneous’ speeches are realised as more diphthong-like compared to the read speech. A One-way ANOVA with the speech style grouping factor revealed that there was a significant difference between them (*F* (3, 6025) = 1222.736, $p < .001$), although the effect size was small ($\omega^2 = .336$). Scheffe post-hoc tests revealed that there was a significant difference in all the pairwise comparisons other than response/description comparison. The effect size for the pairwise comparison was large in read/response ($d = 1.04$), read/description ($d = 1.18$), read/conversation ($d = 1.70$) comparisons and medium in response/conversation ($d = 0.75$) and description/conversation ($d = 0.53$) comparisons. Figure 13 below summarises the differences in VV sequence duration in different speech styles.

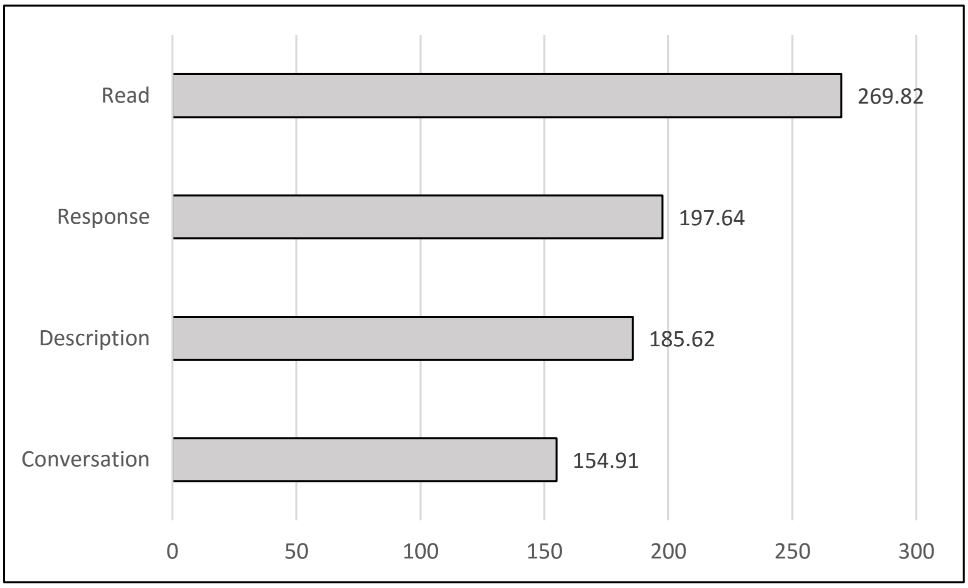


Figure 13 Duration means compared in different speech styles (in msec)

4.3 Ratios of V1, Transition and V2 to total VV sequence duration

4.3.1 *Introduction*

Having discussed the total VV durations, this chapter now focuses on the durations of the intervals which can be identified within a VV sequence; the steady states for the first vowel (V1), the transition and the steady states for the second vowel (V2). Because the speech rate was not controlled across or within speakers, the durations of V1, transition and V2 were presented as their proportions to the total VV duration. In addition, the male and female data was pooled to obtain overall results among Japanese participants. Table **エラー! 参照元が見つかりません。** below presents the mean proportions, standard deviations and ranges of V1, transition and V2 for overall data set of VV sequences across morphological conditions.

Table 15 Interval proportions to the total VV duration

	V1	**Transition**	**V2**
Mean	32.1%	28.4%	39.5%
S.D	10.7%	10.7%	10.6%

We notice that the vowel sequences in Japanese are characterised by a large amount of steady states for both vowels; the steady state for V1 is 32.1% and V2, 39.5%. The vowel quality for the first vowel is kept for a certain period, then the articulators move towards the second vowel and finally the vowel quality for the second vowel is maintained. The transition proportion occupies on average 28.4% of the total duration and the smallest part of the VV sequences. Compared with previous studies, the current result is close to the ratios that Gore (2006) reported. Although his segmentation method is slightly different, the vowel /ai/ in Japanese was analysed to have its V1 for 33.1%, transition for 23.0%, and V2 for 43.8%. However, other studies reported that the transition ratios for VV sequences in Japanese were 46.3% (Hirasaka and Kamata, 1981), 40.0% (Kasuya and Sato, 1990) and 37.2% (Misono and Hirasaka, 2008). Compared with the studies on diphthongs, the transition ratio is smaller than what those studies reported. Svantesson(1984) reported that the Chinese diphthongs were analysed that the onset was 29.2%, the transition was 24.0% and the offset was 37.3%. Therefore, the transition ratio for the current study is assumed to be small among the studies on VV sequences in Japanese as well as the studies on diphthongs.

We also notice that the V2 ratio is larger than the V1 ratio on average. This is consistent with other studies on VV sequences in Japanese but the difference between V1 and V2 ratios are smaller than what their studies report. Most studies on diphthongs reported that the offset ratio was smaller than the onset but Jha (1985) reported that the offset ratios for diphthongs in Maithili were larger than the onset ratios. We investigated the correlation between the V1 and V2 ratios of the current results and found that there was a medium-negative correlation between the V1 and V2 proportions (r (5958) = -0.50, p < .001). There were also negative correlations between the transition and V1 steady state (r (5958) = -0.51, p < .001) and between the transition and V2 steady state (r (5958) = -0.49, p < .001). Therefore, the increase and the decrease of one interval affect other intervals. The following sections present how these ratios are influenced by the morphological conditions and other phonetic characteristics of the VV sequences.

4.3.2 *Morphological condition*

4.3.2.1 General remarks

Firstly, we look at the effect of morphological conditions on the proportions of V1, transition and V2 in the VV sequences. Table 16 below presents the mean proportions (Mean), standard deviation (S.D) and ranges across the morphological conditions.

Table 16 Proportions of V1, transition and V2 to the total VV durations in /hVV/, WI and AWB

	V1			Transition			V2		
	Mean	S.D	Range	Mean	S.D	Range	Mean	S.D	Range
hVV	36.0%	9.7%	7.4-65.6%	22.8%	8.2%	4.3-71.0%	41.2%	9.6%	13.2-74.6%
WI	30.2%	10.4%	2.7-69.9%	30.4%	10.4%	5.3-77.0%	39.4%	10.7%	9.4-78.3%
AWB	31.7%	11.9%	5.8-77.6%	31.7%	11.6%	6.6-72.2%	36.6%	11.5%	5.6-75.4%

Compared to the overall ratios presented in Table 15, the VV sequences in /hVV/ nonsense words had larger V1 and V2 steady state proportions and smaller transition ratios. For VVs of WI and AWB, the transition ratios were larger and the V1 ratios were smaller than the /hVV/ condition. Comparing steady states, V2 ratio was consistently larger than V1 ratio in all conditions. The difference between V1 and V2 ratios was 9.2% for WI, whereas it was about 5% for other conditions.

It is also noticed that the range of variability in the relative duration of subsections is quite wide as is shown in Table 16 above. Thus, each subsection can occupy a large part of the vowel sequence depending on some factors. The distribution of each subsection proportions will be illustrated in the following figures. It is also evident that all Japanese vowel sequences had steady states for the first and the second target vowels on every occasion, and the steady states may occupy the large part of the total VV duration. Within the VV sequences, the steady states occupy on average 77.5% for the /hVV/ nonsense words, 69.7% for the vowel sequences within words (WI) and 68.7% for the vowel sequences across word boundary (AWB). In addition, the ratios for the second target vowels (V2) are larger overall than those for the first target vowels (V1), which is consistent with the findings in the overall result. Figures 14, 15 and 16 below show the distributions of V1, Transition and V2 proportions of three morphological conditions. The number of tokens for each condition was different and the ratio of number of tokens within a data range to the total number of data was calculated to compare the difference between conditions. The vertical axis shows the ratios of token numbers in that range to the total number and the horizontal axis represents the data range. Three conditions were represented by the pattern fill of the bar.

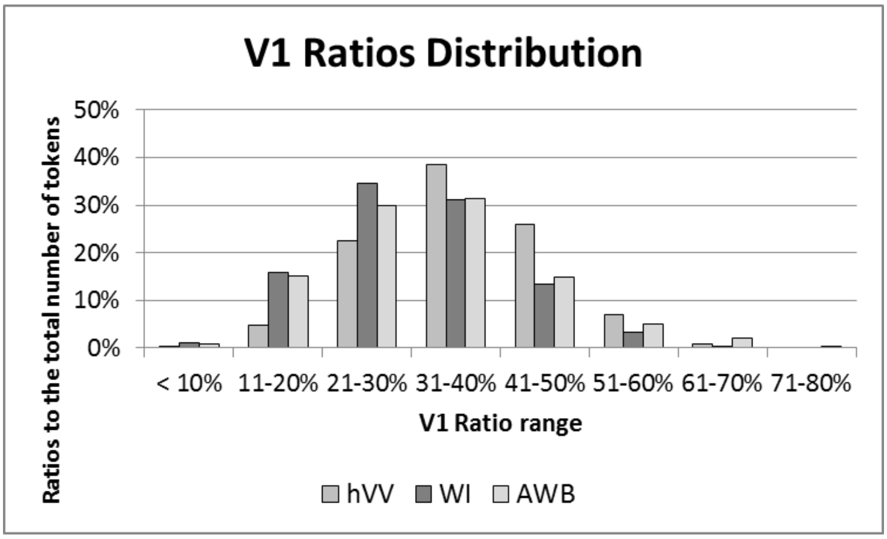


Figure 14 Distribution of V1 proportions

In all conditions, V1 proportions have a central peak between 20 and 40 % and range from 10 % to 80 %. WI and AWB conditions have a wide distribution with both short (less than 10 %) and long (more than 70%) ratios, but the /hVV/ nonsense word condition has smaller range of distribution. 3 tokens of /hVV/ nonsense word condition, 37 tokens of WI and 8 tokens of AWB had V1 of less than 10%, and 3 tokens of WI and 4 tokens of AWB had V1 of more than 70 %.

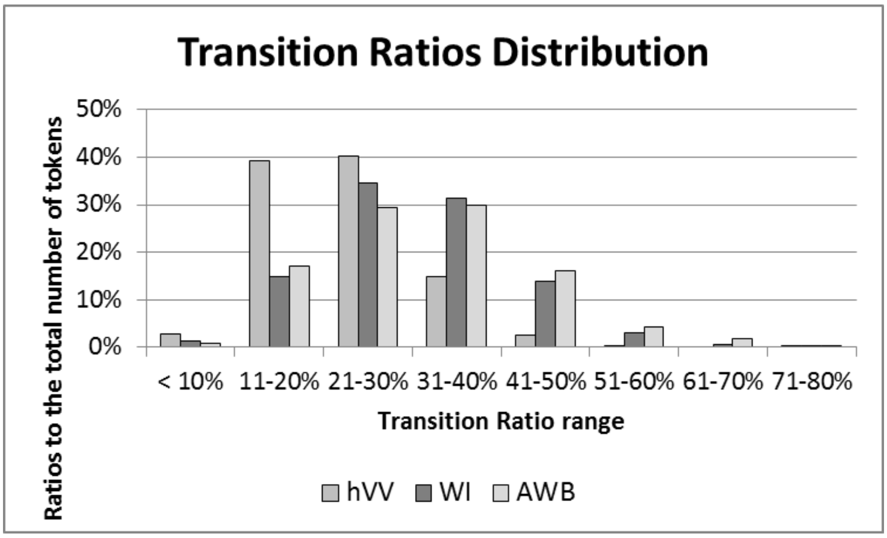


Figure 15 Distribution of Transition proportions

The distribution of transition proportions varies among conditions. About 80 % of transitions of /hVV/ nonsense words were concentrated in a range of 10-30 %, but the distribution of WI and AWB conditions extended to 70 %. 20 tokens of WI condition and 18 tokens of AWB condition had Transition of more than 60%.

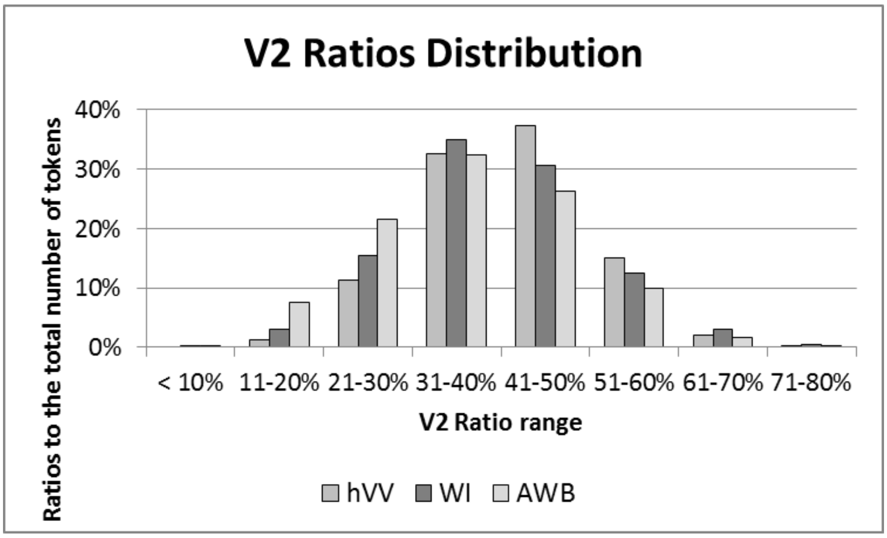


Figure 16 Distribution of V2 proportions

The distribution of V2 proportions had a similar pattern across the morphological conditions, with their central peak around 30-50%.Their ranges are between 10 % and 70 %.

The results of the calculation for the ratios of V1, transition and V2 are now investigated further to study whether there are effects arising from the morphological conditions, accent types of the vowel sequences, the sonority scale, the open/close dimension and the front/back dimension investigated above for the total VV duration results. The subsection ratios presented in Table 12 above are investigated to see whether there is a significant effect of the three morphological conditions. A Multivariate analysis of variance (MANOVA) was conducted on the three dependent variables (V1, Transition and V2) with the morphological conditions (hVV, WI and

AWB) as independent variable. Although the effect size represented by the omega square was small ($\omega^2 = .123$), there was a significant difference among the three morphological conditions on the dependent measures (Wilks' Lambda = 0.87, F (6, 11908) = 131.947, $p < .001$). The ratios of V1, transition and V2 were all found to be significantly different for the morphological conditions. (V1: F (2, 5955) = 175.884, $p < .001$, $\omega^2 = .054$, transition: F (2, 5955) = 381.105, $p < .001$, $\omega^2 = .107$, V2: F (2, 5955) = 55.176, $p < .001$, $\omega^2 = .018$). The MANOVA showed that the morphological conditions were found to be one of the factors of affecting each proportion. Scheffe post-hoc tests were conducted to reveal if one of the morphological conditions is significantly different from the others. Table 17 below presents all the pairwise comparisons for the post-hoc tests.

Table 17 Post-hoc comparisons for the interval ratios for the morphological conditions

	Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
V1	hVV>WI**	0.57	M	hVV>AWB**	0.41	S	WI<AWB*	-0.14	W
Transition	hVV<WI**	-0.79	M	hVV<AWB**	-0.96	L	WI<AWB*	-0.12	W
V2	hVV>WI**	0.18	W	hVV>AWB**	0.45	S	WI>AWB**	0.26	S

*: $0.001 < p < 0.05$, **: $p < 0.001$

W: Weak, S: Small, M: Medium, L: Large

With regard to the V1 ratios, the ascending order was WI < AWB < /hVV/. Although there was a significant difference between WI and AWB, the effect size for the pairwise comparison was less than a small effect. There was a significant difference between the /hVV/ condition and other morphological conditions. Thus, the VV sequences in the /hVV/ condition had significantly larger V1 ratio. For the transition ratios, the ascending order was /hVV/ < WI < AWB. There was a significant difference between the /hVV/ condition and other morphological conditions. WI and AWB are significantly different from each other but the effect size was less than the small effect. Thus, VV sequences within the real words and across the word boundary had significantly larger transition ratios than the /hVV/ condition. As for the V2 ratios, the ascending order was AWB < WI < /hVV/. There was a significant difference between AWB and WI. Although a significant difference was found between WI and /hVV/, the effect size was less than the small effect size. Thus, the VV sequences across word boundaries had significantly smaller V2 ratio than others. Therefore, the morphological conditioning applied to the vowel sequence affected the relative durations of the different elements of the sequence, VV sequences in nonsense words (hVV) had longer steady states and shorter transition than real word (WI) or real phrase (AWB) conditions. The difference between WI and AWB was significant and the V1 and transition of WI condition were longer and V2 was shorter than AWB condition.

Although the effect size was small, the finding is not consistent with the claims of Kawakami (1977) and Saito (1997) that the Japanese vowel sequences having a ‘semantic boundary’ between them can be regarded as diphthongs. In the present study, a word boundary within a compound can presumably be regarded as a ‘semantic boundary’ because each word has a separate meaning but the definition of ‘semantic boundary' in their argument is not clearly defined. Also, the vowel sequences within real words do not have 'semantic' boundary between them because two vowels are in a meaningful unit. This morphological difference was referred to as a cue that the vowel sequences in Japanese have a diphthong-like feature but the term 'semantic boundary' was not clearly defined. The difference between VV sequence within a morpheme and across morphological boundary would also be considered to be a different 'semantic boundary'. Also, VV sequences in the /hVV/ condition are assumed to be without 'semantic boundary' between vowels because the /hVV/ produced with a particular accent and transcribed in Hiragana character will not convey any meanings. Therefore, the vowel sequences without 'semantic boundary' between them are less likely to be regarded as having a diphthong-like feature compared to other conditions. In addition, the word boundary can be regarded as a 'semantic boundary' and the VV sequences across word boundary are more diphthong-like than VVs within words because the transition ratio becomes larger.

4.3.2.2 Accent contour

Having looked at the effect of morphological condition, we now consider the possible effects of properties of the vowels themselves on the realisation of Japanese VV sequences. As is discussed in the literature review chapter, the mora containing vowels are produced with either High or Low tones and consequently the words are produced with a particular pitch accent. The difference of the tone of the vowel, that is, the accent contour of the vowel sequence, was also pointed out to be a factor associated with Japanese VV sequences have diphthong-like feature (Takubo et al., 1998). Therefore, the ratios of V1, transition and V2 across different accent contour of VV sequences are presented and investigated to gauge whether the tonal difference of the vowel (the accent difference of the vowel sequence) has a significant effect for the ratios of V1, transition and V2. Table 18 below presents the mean proportions, standard deviations and ranges of V1, transition and V2. VV sequences with HH and LL accent have slightly larger transition ratios and slightly smaller V2 ratios. The feature that V2 ratio is consistently larger than V1 ratio is also found in the accent type difference.

However, as is noted in the methodology chapter, it was only possible to look at the effect of the variation of the accent contour in the /hVV/ and WI conditions. Furthermore, in the /hVV/ condition the effect of accent is limited to the difference between HL and LH. The interaction between the morphological conditions can also be obtained in HL and LH accent. Table 18 below shows the subsection proportions of different morphological conditions sorted by the accent types.

Table 18 Proportions of V1, Transition and V2 of each accent type of morphological conditions

		HH		**HL**		**LH**		**LL**	
		Mean	S.D	Mean	S.D	Mean	S.D	Mean	S.D
hVV	V1			36.6%	9.7%	35.5%	9.7%		
	Transition			23.5%	8.5%	22.1%	7.8%		
	V2			40.0%	9.5%	42.5%	9.6%		
WI	V1	30.7%	10.4%	30.4%	10.3%	27.8%	9.4%	32.0%	10.8%
	Transition	30.5%	10.6%	30.5%	10.9%	30.2%	10.5%	30.5%	9.8%
	V2	38.8%	10.9%	39.2%	10.4%	42.0%	10.0%	37.5%	10.8%
AWB	V1	31.7%	11.9%						
	Transition	31.7%	11.6%						
	V2	36.6%	11.5%						

Firstly, the effect of accent in the /hVV/ condition is tested. An ANOVA was conducted for subsection ratios for the /hVV/ condition with the accent type (HL and LH) as an independent variable. It showed a significant difference in all subsection proportions (V1, transition and V2) on the accent type (V1: *F* (1, 1685) = 5.313, p = .021, ω^2 = .003), transition: *F* (1, 1685) = 12.315, p < .001, ω^2 = .007 and V2: *F* (1, 1685) = 28.558, p < .001, ω^2 = .017). In the case of the /hVV/ condition, the accent had a significant effect on subsection ratios and VVs with HL accent had larger V1, larger transition and smaller V2 ratios than LH accent. Although the effect size is small, the finding is consistent with the claim of Takubo et al. (1998).

Secondly, the effect of accent in the WI condition is tested. An ANOVA was conducted for subsection ratios for the WI condition with the accent type (HH, HL, LH and LL) as an independent variable. But in the case of the WI condition, the significant effect was found in V1 (*F* (3, 3427) = 24.692, $p < .001$, $\omega^2 = .003$) and V2 (*F* (3, 3427) = 27.411, $p < .001$, $\omega^2 = .022$) ratios but not in the transition ratio (*F* (3, 3427) = 0.181, $p = .91$). Table 19 below shows all the pairwise comparisons and their effect. For the V1 ratio, the ascending order was LH < HL < HH <LL. There was a significant difference between LH and others. There was also a difference between HL and LL but the effect size was less than the small effect. For the V2 ratio, the ascending order was LL < HH < HL <LH. There was a significant difference between LH and others. There was a significant difference between LL and HL but the effect size was less than the small effect. Therefore, the VV sequences of LH accent contour had significantly smaller V1 and larger V2 ratios than VVs of other accent contours, although the effect size was small.

Table 19 Post-hoc comparisons for the interval ratios for the WI condition

	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
V1				HH>LH**	0.29	S			
	HL>LH**	0.26	S	HL<LL*	-0.15	W	LH<LL**	-0.41	S
V2				HH<LH**	-0.30	S			
	HL<LH**	-0.28	S	HL>LL*	0.15	W	LH>LL**	0.43	S

*: 0.001 < p < 0.05, **: p < 0.001

W: Weak, S: Small, M: Medium, L: Large[11]

The interaction between the morphological condition and the accent type was tested between the /hVV/ and WI condition on HL and LH accent. A MANOVA was conducted to test the interaction, but no interaction was found (Wilks' Lambda = 0.998, F (3, 3385) = 1.79, p = .147). A weak interaction was found in V1 ratio (F (1, 3387) = 4.733, p = .030, ω^2 = .001), whose effect size was quite small. The effect of the accent in the /hVV/ condition and WI condition was consistent in V1 and V2 ratios. VVs with HL accent had significantly larger ratios than LH accent for V1 and VVs with LH accent had significantly larger ratios than LH accent for V2. However, there was a significant difference for the transition ratios of the /hVV/ condition but no difference was found in the WI condition.

---

[11] Cohen's d is evaluated as weak (less than 0.20), small (0.20-0.50), medium (0.50-0.80) and large (above 0.80). See 3.3.2

4.3.2.3 Vowel sequence types

4.3.2.3.1 Sonority

In addition to the tonal differences in VV sequences, the sonority differences between vowels in VV sequences were also investigated. As is mentioned in 4.3.2.1 above, Carlyle (1987) suggested that vowel sequences in Breton are realised as diphthongs if the first vowel is more sonorous than the second, although did not provide any phonetic analysis of this. The aim of the present investigation is to test whether the sonority difference within a VV sequence is a relevant factor in accounting for the realisation of those sequences in Japanese. Table 20 below presents the mean proportions, standard deviations and ranges of V1, transition and V2 as a function of the sonority contour within the sequence of three morphological conditions.

Table 20 Proportions of V1, transition and V2 to the total VV durations of falling, rising and level sequences

		Falling		Rising		Level	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
V1	hVV	36.9%	10.3%	35.9%	9.3%	35.0%	9.6%
	WI	32.1%	10.7%	29.5%	10.1%	28.7%	9.9%
	AWB	35.1%	11.6%	29.7%	12.3%	30.5%	10.8%
	Overall	33.9%	10.9%	31.3%	10.6%	30.9%	10.3%
Transition	hVV	23.5%	8.5%	21.6%	7.6%	23.7%	8.4%
	WI	30.4%	10.4%	29.3%	10.6%	32.5%	9.8%
	AWB	32.1%	12.4%	30.0%	10.5%	34.4%	12.0%
	Overall	28.7%	10.7%	27.3%	10.5%	30.1%	10.6%
V2	hVV	39.6%	9.5%	42.5%	9.6%	41.4%	9.4%
	WI	37.5%	9.4%	41.2%	11.3%	38.7%	10.6%
	AWB	32.8%	10.3%	40.3%	11.5%	35.1%	11.0%
	Overall	37.4%	9.8%	41.4%	10.9%	39.0%	10.5%

V2 ratios are still larger than V1 ratios in all conditions, which is consistent with the other conditions investigated so far. The ratio between V1 and V2 is bigger in the rising and the level sequences than in the falling sequences. The difference of transition ratios across the sonority scale condition is within 3%, therefore the sonority scale difference were not assumed to have a major effect on the steady state ratios.

Firstly, A MANOVA was conducted to investigate whether there is a significant relation between the morphological condition and the sonority. Although its effect size was small, it showed a significant difference (Wilks' Lambda = .993, *F* (6, 15735) = 3.563, $p < .001$, $\omega^2 = .005$). Also, the proportions of V1 and V2 were found to be significantly different (V1: *F* (4, 5949) = 4.862, $p = .001$, $\omega^2 < 0.001$ and V2: *F* (4,

5949) = 6.515, $p < .001$, $\omega^2 < 0.027$). But the transition proportion was not found to be significant (F (4, 5949) = 2.219, $p = .064$).

Secondly, the interaction between the subsection proportion and the sonority in each morphological condition was investigated. An ANOVA was conducted for the subsection ratios of each morphological condition with the sonority scale as an independent variable. It showed significant differences in all subsection proportions of all morphological conditions, although the effect size was small. In a condition off hVV nonsense words, an ANOVA showed significant differences in all subsection proportions (V1: F (2, 1686) = 4.812, $p = .008$, $\omega^2 = 0.005$, transition: F (2, 1686) = 12.11, $p < .001$, $\omega^2 = 0.013$, V2: F (2, 1686) = 15.20, $p < .001$, $\omega^2 = 0.016$). In the case of the WI condition, an ANOVA also showed significant differences in all subsection proportions (V1: F (2, 3430) = 31.481, $p < .001$, $\omega^2 = 0.017$, transition: F (2, 3430) = 24.347, $p < .001$, $\omega^2 = 0.013$, V2: F (2, 3430) = 42.683, $p < .001$, $\omega^2 = 0.023$). In the AWB condition, an ANOVA showed significant differences in all subsection proportions (V1: F (2, 839) = 18.150, $p < .001$, $\omega^2 = 0.038$, transition: F (2, 839) = 9.338, $p < .001$, $\omega^2 = 0.019$, V2: F (2, 839) = 35.378, $p < .001$, $\omega^2 = 0.080$).

Table 21 below presents all the pairwise comparisons for the post-hoc tests.

Table 21 Post-hoc comparisons for the interval ratios for the sonority scale

		Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
hVV	**V1**				F>L*	0.19	W			
	Transition	F>R**	0.24	S				R<L**	0.26	S
	V2	F<R**	0.31	S	F<L*	0.19	W			
WI	**V1**	F>R**	0.25	S	F>L**	0.33	S			
	Transition	F>R*	0.11	W	F<L**	0.21	S	R<L**	0.31	S
	V2	F<R**	0.36	S	F<L*	0.13	W	R<L**	0.22	S
AWB	**V1**	F>R**	0.45	S	F>L**	0.41	S			
	Transition							R<L**	0.40	S
	V2	F<R**	0.68	M				R>L**	0.46	S

*: 0.001 < p < 0.05, **: p < 0.001

W: Weak, S: Small, M: Medium, L: Large

As for the V1 ratios, the ascending order was Level < Rising < Falling. There was a significant difference between the fall sequences and level sequences in all morphological conditions, although the effect size was weak or small. There was also a significant difference between the falling and the rising sequences in the cases of WI and AWB conditions. Thus, the level sequences are assumed to have smaller V1 ratios than falling sequences. With regard to the transition ratios, the ascending order was Rising < Falling < Level. There was a significant difference for all the pairwise comparisons in the WI condition but the falling sequences and level sequences did not show any significance in the hVV condition and there was no significant difference between the falling sequences and the rising sequences in the AWB condition. Thus, the level sequences are supposed to have larger transition ratios than the rising sequences.

For the V2 ratios, the ascending order was Falling < Level < Rising. There was a significant difference between the falling and the level sequences in the hVV and WI conditions but the effect size was less than the small effect. The difference between the level and the rising sequences was significant in the WI and AWB conditions, although the effect size was small. Thus, the rising sequences are assumed to have larger V2 ratios.

It should be noted that the WI condition seems to contribute to the significant effect on the proportion differences among three categories of sonority in VV sequences because the significant differences in subsection proportion was found in all cases apart from the V1 subsection between the rising and level sequences. Therefore, there is no positive evidence to support the claim by Carlyle (1987) but the level sequences are assumed to have smaller steady state ratios and larger transition ratio, although the effect size was small.

4.3.2.3.2 Closing-Opening dimension

In addition to the sonority difference in VV sequences, we look at the articulatory characteristics of VV sequences; the closing /opening of the mouth and the fronting/backing of the tongue. Table 22 below presents the mean proportions, standard deviations and ranges of V1, transition and V2 across the different degrees of opening and closing movement for pronouncing VV sequences in three morphological conditions. As is explained in 4.2.1, Close 2 refers to the movement from /a/ to either /i/ or /u/, Close 2 refers to the movement from/to the mid position. ‘N.D.’ means that the both vowels are in the same height, such as /iu/.

Table 22 Proportions of V1, transition and V2 to the total VV durations of open/close dimensions

		Close2		**Close1**		**N.D.**		**Open1**		**Open2**	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
V1	hVV	36.1%	10.7%	37.3%	10.0%	35.0%	9.6%	35.5%	9.3%	36.8%	9.2%
	WI	35.4%	10.4%	30.4%	10.5%	28.7%	9.9%	30.5%	10.4%	26.4%	8.7%
	AWB	34.4%	10.7%	35.4%	12.0%	30.5%	10.8%	30.2%	12.3%	28.4%	12.1%
	Overall	35.5%	10.5%	33.1%	11.0%	30.9%	10.3%	31.8%	10.6%	29.7%	10.4%
Transition	hVV	23.4%	9.7%	23.6%	7.8%	23.7%	8.4%	21.3%	7.2%	22.4%	8.5%
	WI	28.2%	9.9%	31.6%	10.5%	32.5%	9.8%	29.0%	10.2%	30.2%	11.6%
	AWB	33.4%	11.9%	31.4%	12.6%	34.4%	12.0%	30.2%	10.8%	29.5%	9.7%
	Overall	27.6%	10.6%	29.3%	10.8%	30.1%	10.6%	27.1%	10.2%	27.9%	11.1%
V2	hVV	40.4%	9.7%	39.1%	9.3%	41.4%	9.4%	43.1%	9.5%	40.8%	9.9%
	WI	36.4%	9.1%	38.0%	9.6%	38.7%	10.6%	40.5%	10.8%	43.3%	12.4%
	AWB	32.2%	9.9%	33.2%	10.6%	35.1%	11.0%	39.7%	11.5%	42.1%	11.2%
	Overall	36.9%	9.7%	37.6%	9.8%	39.0%	10.5%	41.1%	10.6%	42.4%	11.6%

Overall, V2 ratios are consistently larger than V1 ratios in all the degrees of opening/closing movement. The V1 to V2 ratios are slightly smaller in closing sequence (1.4% for Close 2 and 4.5% for Close 1) than in opening sequences (9.3% for Open 1 and 12.7% for Open 2). In addition, the difference in the transition ratios across the condition is 3.0%. This may suggest that the VV sequences produced through the opening movement need more effort to produce their second (more open) vowel than VVs produced through the closing movement. Also, the transition ratios are consistently the smallest among three subsections. Therefore, the opening and closing movements may not be associated with VV sequences having diphthong-like feature.

Firstly, A MANOVA was conducted to investigate whether there is a significant relation between the morphological condition and the different degree of opening and closing movement for VV sequences. Although its effect size was small, it showed a significant difference (Wilks' Lambda = .080, F (24, 17231) = 5.064, $p < .001$, $\omega^2 = .016$). In addition, the proportions of all subsection proportions were found to be significantly different (V1: F (8, 5943) = 10.554, $p < .001$, $\omega^2 = 0.001$, transition: F (8, 5943) = 3.815, $p < .001$, $\omega^2 < 0.001$ and V2: F (8, 5943) = 7.272, $p < .001$, $\omega^2 < 0.001$).

Secondly, the interaction between the subsection proportion and the Opening/Closing dimension in each morphological condition was investigated. An ANOVA was conducted for the subsection ratios of each morphological condition with the Open/Close dimension as an independent variable. It showed significant differences in all subsection proportions of all morphological conditions, although the effect size was small. In a condition off hVV nonsense words, an ANOVA showed significant differences in all subsection proportions (V1: F (4, 1286) = 3.498, $p = .007$, $\omega^2 = 0.006$, transition: F (4, 1286) = 6.689, $p < .001$, $\omega^2 = 0.013$, V2: F (4, 1286) = 10.429, $p < .001$, $\omega^2 = 0.022$). In the case of the WI condition, an ANOVA also showed significant differences in all subsection proportions (V1: F (4, 3430) = 43.481, $p < .001$, $\omega^2 = 0.047$, transition: F (4, 3430) = 20.288, $p < .001$, $\omega^2 = 0.022$, V2: F (4, 3430) = 28.222, $p < .001$, $\omega^2 = 0.031$). In the AWB condition, an ANOVA showed significant differences in all subsection proportions (V1: F (4, 839) = 9.576, $p < .001$, $\omega^2 = 0.039$, transition: F (4, 839) = 5.205, $p < .001$, $\omega^2 = 0.020$, V2: F (4, 839) = 20.738, $p < .001$, $\omega^2 = 0.084$).

Scheffe post-hoc tests were also conducted to test if the subsection proportions of VV sequences with different levels of opening and closing movements were significantly different with each other. Table 23 below presents all the pairwise comparisons for the post-hoc tests.

Table 23 Post-hoc comparisons for the interval ratios for the opening/closing contrast

		Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
hVV	V1									
					C1>ND*	0.24	S			
	Transition									
								C1>O1*	0.30	S
					ND>O1*	0.30	S			
	V2							C2<O1*	0.28	S
					C1<ND*	0.24	S	C1<O1**	0.43	S
WI	V1	C2>C1**	0.48	S	C2>ND**	0.66	M	C2>O1**	0.47	S
		C2>O2**	0.94	L	C1>ND**	0.17	W			
		C1>O2**	0.4	S	ND<O1*	0.18	W	ND>O2**	0.24	S
		O1>O2*	0.41	S						
	Transition	C2<C1*	0.33	S	C2<ND**	0.44	S			
								C1>O1**	0.25	S
					ND>O1**	0.35	S			
	V2				C2<ND*	0.23	S	C2<O1**	0.39	S
		C2<O2**	0.65	M				C1<O1**	0.24	S
		C1<O2**	0.51	M	ND<O1*	0.16	W	ND<O2**	0.41	S
		O1<O2**	0.26	S						
AWB	V1									
		C2>O2**	0.53	M	C1>ND**	0.25	S	C1>O1*	0.26	S
		C1>O2**	0.43	S						
	Transition									
					ND>O1*	0.37	S	ND>O2*	0.44	S
	V2							C2<O1**	0.68	M
		C2<O2**	0.94	L				C1<O1**	0.59	M
		C1<O2**	0.83	L	ND<O1*	0.40	S	ND<O2**	0.63	M

*: $0.001 < p < 0.05$, **: $p < 0.001$

W: Weak, S: Small, M: Medium, L: Large

With regard to the V1 ratios, the ascending order was Open 2 (O2) < N.D. < Open 1 (O1) < Close 1 (C1) < Close 2 (C2) but its significance is different among three morphological conditions. There was a significant difference between O1 and C1 in the WI and AWB conditions but the effect size was small. Also, C1 and C2 had a significant difference in the WI condition but the effect size was small. There was a large and medium effect size on the comparison between the smallest ratio (O2) and the largest ratio (C2) on the V1 ratio of the WI and AWB conditions. Thus, the closing sequences had slightly larger V1 ratios than others. Secondly, for the transition ratios, the ascending order was O1 < C2 < O2 < C1 < N.D. Significant differences were found between N.D and O1 in all morphological conditions, although the effect size was small. There was no critical boundary of significant difference for the transition ratio. Lastly for the V2 ratios, the ascending order was C2 < C1 < N.D. < O1 < O2. Significant differences were found between the Opening categories (O1 and O2) and other categories in most cases. Therefore, the closing sequences had smaller V2 ratio than the opening sequences and the V2 ratios of VVs consisted with the same height were between the opening and the closing sequences. The opening and closing movement is supposed to be associated with the V2 ratios.

4.3.2.3.3 Fronting-Backing dimension

Lastly we look at the fronting/backing of the tongue for producing VV sequences. Table 24 below presents the mean proportions, standard deviations and ranges of V1, transition and V2 across the different degrees of fronting and backing movement for pronouncing VV sequences. As is explained in 4. 2.1, Front 2 refers to the movement from back vowels to front vowels and Front 2 refers to the movement from/to /a/. ‘N.D.’ means that the both vowels are in the same front-back dimension, such as /ui/ and /ie/.

Table 24 Proportions of V1, transition and V2 to the total VV durations for Front/Back dimension

		Front2		**Front1**		**N.D.**		**Back1**		**Back2**	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
V1	hVV	31.9%	9.1%	36.3%	9.8%	36.7%	8.3%	37.6%	9.7%	37.8%	9.5%
	WI	25.7%	8.9%	30.1%	10.6%	32.6%	10.3%	31.1%	9.9%	32.8%	10.4%
	AWB	28.2%	10.8%	33.8%	12.0%	29.4%	13.8%	32.2%	12.3%	33.8%	10.8%
	Overall	27.8%	9.6%	32.3%	11.0%	33.1%	10.7%	33.1%	10.6%	34.5%	10.4%
Transition	hVV	23.9%	7.9%	23.2%	8.2%	19.0%	6.4%	22.6%	8.0%	22.9%	8.6%
	WI	31.9%	10.3%	29.6%	10.4%	27.3%	10.6%	30.7%	10.4%	31.0%	10.2%
	AWB	34.3%	13.0%	32.2%	12.1%	29.4%	11.7%	30.0%	11.1%	31.5%	9.6%
	Overall	30.0%	10.9%	28.1%	10.6%	25.6%	10.6%	28.3%	10.5%	28.6%	10.4%
V2	hVV	44.2%	10.1%	40.6%	9.1%	44.3%	8.6%	39.8%	9.5%	39.3%	9.1%
	WI	42.3%	10.6%	40.3%	11.2%	40.1%	10.8%	38.2%	9.7%	36.2%	10.0%
	AWB	37.6%	11.5%	34.1%	12.5%	41.2%	12.2%	37.8%	10.8%	34.7%	9.6%
	Overall	42.2%	10.8%	39.5%	11.1%	41.3%	10.7%	38.6%	9.8%	36.9%	9.8%

As is the same with the opening/closing dimension, V2 ratios are consistently larger than V1 ratios in all the levels of the front/back movement. The V1 to V2 ratios are larger in fronting sequence (14.4% for Front 2 and 7.2% for Front 1) than in backing sequences (5.5% for Back 1 and 2.4% for Back 2). The difference in transition ratios across the front/back levels is 4.4%, which is slightly larger than the open/back dimension. This suggests that the movement from the first and the second vowel of VV sequences may quicker in fronting sequences than backing sequences. As the difference in the steady state ratios for front and back sequences are small, the articulatory effort to produce both vowels of VV sequences are assumed to be affected by the front/back movement of the tongue.

Firstly, A MANOVA was conducted to investigate whether there is a significant relation between the morphological condition and the different degree of fronting and backing movement for VV sequences. Although its effect size was small, it showed a significant difference (Wilks' Lambda = .989, F (24, 17231) = 2.840, $p < .001$, $\omega^2 = .007$). Also, the proportions of all subsection proportions were found to be significantly different (V1: F (8, 5943) = 3.469, $p < .001$, $\omega^2 < 0.001$, transition: F (8, 5943) = 1.971, $p = .046$, $\omega^2 < 0.001$ and V2: F (8, 5943) = 6.664, $p < .001$, $\omega^2 < 0.001$).

Secondly, the interaction between the subsection proportion and the Fronting/Backing dimension in each morphological condition was investigated. An ANOVA was conducted for the subsection ratios of each morphological condition with the Front/Back dimension as an independent variable. It showed significant differences in all subsection proportions of all morphological conditions, although the effect size was small. In a condition off hVV nonsense words, an ANOVA showed significant differences in all subsection proportions (V1: F (4, 1686) = 24.197, $p < .001$, $\omega^2 = 0.051$, transition: F (4, 1686) = 10.441, $p < .001$, $\omega^2 = 0.022$, V2: F (4, 1686) = 20.511, $p < .001$, $\omega^2 = 0.044$). In the case of the WI condition, an ANOVA also showed significant differences in all subsection proportions (V1: F (4, 3430) = 58.637, $p < .001$, $\omega^2 = 0.062$, transition: F (4, 3430) = 14.310, $p < .001$, $\omega^2 = 0.015$, V2: F (4, 3430) = 37.579, $p < .001$, $\omega^2 = 0.040$). In the AWB condition, an ANOVA showed significant differences in all subsection proportions (V1: F (4, 839) = 8.088, $p < .001$, $\omega^2 = 0.032$, transition: F (4, 839) = 4.378, $p = .002$, $\omega^2 = 0.016$, V2: F (4, 839) = 8.303, $p < .001$, $\omega^2 = 0.033$).

Scheffe post-hoc tests were also conducted to test if the subsection proportions of VV sequences with different levels of fronting and backing movements were significantly different with each other. Table 25 below presents all the pairwise comparisons for the post-hoc tests.

Table 25 Post-hoc comparisons for the interval ratios for the opening/closing contrast

		Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
hVV	V1	F2<F1**	0.46	S	F2<ND**	0.54	M	F2<B1**	0.60	M
		F2<B2**	0.64	M						
	Transition				F2>ND**	0.66	M			
					F1>ND**	0.54	M			
					ND<B1**	0.48	S	ND<B2**	0.48	S
	V2	F2>F1**	0.38	S				F2>B1**	0.45	S
		F2>B2**	0.51	M	F1<ND*	0.42	S			
					ND>B1**	0.49	S	ND>B2**	0.56	M
WI	V1	F2<F1**	0.45	S	F2<ND**	0.73	M	F2<B1**	0.58	M
		F2<B2**	0.74	M						
		F1<B2**	0.26	S						
		B1<B2*	0.17	W						
	Transition	F2>F1**	0.23	S	F2>ND**	0.44	S			
					F1>ND**	0.22	S			
					ND<B1**	0.32	S	ND<B2**	0.35	S
	V2	F2>F1*	0.19	W	F2>ND*	0.21	S	F2>B1**	0.41	S
		F2>B2**	0.60	M				F1>B1*	0.20	S
		F1>B2**	0.39	S				ND>B2**	0.38	S
		B1>B2*	0.20	S						
AWB	V1	F2<F1**	0.49	S				F2<B1*	0.35	S
		F2<B2**	0.52	M						
	Transition				F2>ND*	0.39	S	F2>B1*	0.36	S
	V2									
					F1<ND*	0.57	M	F1<B1*	0.32	S
								ND>B2*	0.62	M

*: 0.001 < p < 0.05, **: p < 0.001

W: Weak, S: Small, M: Medium, L: Large

With regard to the V1 ratios, the ascending order was Front 2 (F2) < Front 1 (F1) < Back 1 (B1) <[12] N.D. < Back 2. There was a significant difference between F2 and F1 in all morphological conditions. V1 proportion in F2 is significantly smaller than other categories in the hVV and WI conditions but not in the AWB condition. For the transition proportions, the ascending order was N.D. < F1 < B1 < B2 < F2. Significant differences were found between N.D and F1 as well as B2 and F2 in the hVV and WI conditions, although the effect size is small. Judging just from the mean transition ratios, VV sequences of larger front /back movement had larger transition ratio than VVs with smaller movement. However, it is not supported by the statistics and its effect size. As for the V2 ratios, the ascending order was B2 < B1 < F1 < N.D. < F2. However, the results for the post-hoc tests vary among three morphological conditions. The effect sizes for the difference between F2 and B2 in the hVV and WI conditions were medium and those between N.D. and B2 in all morphological conditions were also medium.

Although the statistics showed that all of the factors we considered were meaningful, none of them showed the large effect size. Table 16 above and figures 14, 15 and 16 present the wide range of subsection ratios, whereas none of the factors does not account for it.

---

[12] The V2 ratio of B1 was 33.12% and N.D. was 33.13%.

4.3.3 *Speech style Conditions*

4.3.3.1 General remarks

Having observed the timing properties of VV sequences of different morphological conditions, we now look at the VV sequences within the real words produced in various speech styles. The word-internal data, which was presented in the previous section, is presented here as ‘read speech’ to compare with other speech styles. Table 26 below presents the overall summary of the ratios of V1, Transition and V2 to the total vowel sequence durations. This result is from the vowel sequences within real words, that is, their morphological condition is fixed, and the all the speech styles are pooled for the purpose of showing the effect of the speech style.

Table 26 Interval proportions to the total VV duration

	V1	**Transition**	**V2**
Mean	28.6%	33.4%	38.0%
S.D	11.2%	12.4%	11.7%

The pooled data for the word-internal VV sequences produced in different speech styles had slightly different properties from VV sequences across the morphological condition (See Table 15). Although the transition was the smallest part in VV sequences, the current data shows the V1 proportion occupies the smallest part of the sequences. The largest part of the VV sequences is V2, which is consistent with VVs across the morphological condition. Therefore the overall description of these VV sequences is they have large ratios of V2 steady state and small ratios of V1 steady states. Investigating the correlation among three interval ratios, there is a medium-negative correlation between V1 and V2 ratios (r (6029) = -0.43, p < .001). The decrease in the V1 ratio relates to the increase of the V2 ratio. The transition ratio correlates with both V1 (r (6029) = -0.51, p < .001) and V2 (r (6029) = -0.57, p < .001) ratios negatively. Therefore, three intervals are found to be significantly correlated with each other. In the following sections, we look at how these ratios are influenced by the speech style difference and other factors.

Firstly, the effect of speech styles is investigated. Real words containing VV sequences were produced by asking participants to read out words, asking questions to respond target words, asking to describe pictures and doing free conversation. Read speech is represented as ‘read‘ or ‘rd’, responses to questions is ‘resp’ or ‘rs’, describing pictures is ‘desc’ or ‘ds’ and free conversation is ‘conv’ or ‘cv’ from now on. Table 27 below presents the mean proportions (Mean), standard deviation (S.D) and ranges across the speech styles.

Table 27 Proportions of V1, transition and V2 to the total VV durations in four speech styles

	V1			**Transition**			**V2**		
	Mean	S.D	Range	Mean	S.D	Range	Mean	S.D	Range
Read	30.2%	10.4%	2.7-69.9%	30.4%	10.4%	5.3-77.0%	39.4%	10.7%	9.4-78.3%
Response	24.8%	11.1%	4.4-70.7%	36.1%	12.4%	7.8-77.0%	39.2%	12.9%	5.3-75.7%
Desc	22.2%	10.5%	5.6-69.3%	42.7%	14.5%	12.7-76.9%	35.1%	13.2%	7.8-68.2%
Conv	27.8%	12.1%	5.1-72.0%	36.8%	13.5%	6.0-80.2%	35.4%	12.4%	5.7-76.1%

From the table 27 above, we notice that the ranges of proportions are very wide as is the same with the morphological conditions. Each of the subsection may occupy either large or small proportion in VV sequences, ranging from 3% to 80%. It is also noticed that all Japanese vowel sequences produced with four different speech styles had steady states for the first and the second target vowels rather than overall glide from the onset to the offset, which is consistent with the observation for the morphological conditions. Also, steady states appeared to occupy the large part of the total VV duration. Within the vowel sequences, steady states occupied 69.6% for the read speech, 64.1% for responses to questions, 57.3% for picture descriptions and 63.2% for conversation. The tendency that the second target vowels ratios (V2) are larger than those for the first target vowels (V1) was also recognised in the speech style data set. Figures 17, 18 and 19 below illustrate the distributions of subsection proportions in different speech styles. Because of the unequal numbers of tokens in each speech style group, the ratio of number of tokens within a data range to the total number of data was calculated. Vertical axis shows the ratios of token numbers in that range to the total number and the horizontal axis represents the data range. Four speech style conditions are represented by the pattern fill of the bar.

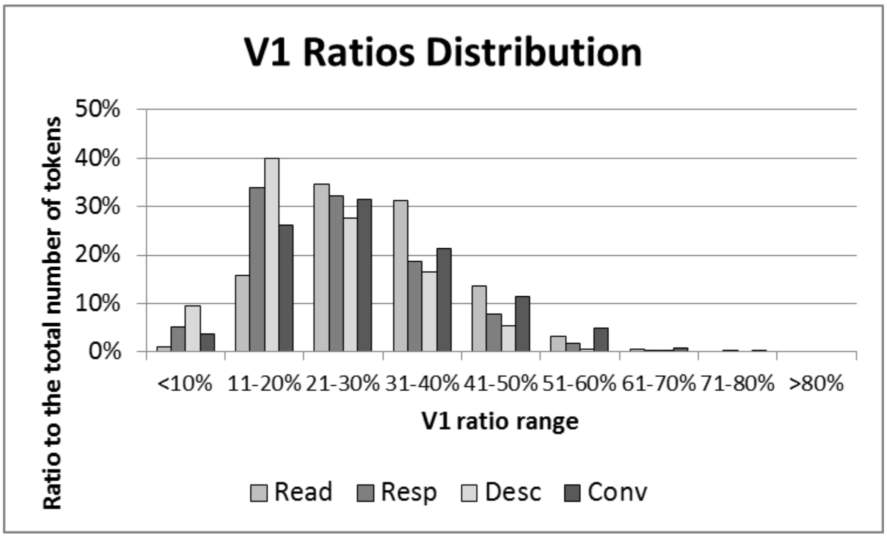


Figure 17 Distribution of V1 proportions

Figure 17 presents the distributions of V1 proportions. The central peak is between 10-30% in all speech styles and ranges to the larger proportions of more than 70%. V1 proportions of about 90% of tokens distribute between 10-50%. The numbers of sequences with smaller V1 ratios (less than 10%) are 37 for read speech, 30 for responses, 31 for description and 64 for conversation. The numbers of them with larger V1 ratios (more than 50%) are 171 for read speech, 16 for responses, 26 for description and 99 for conversation.

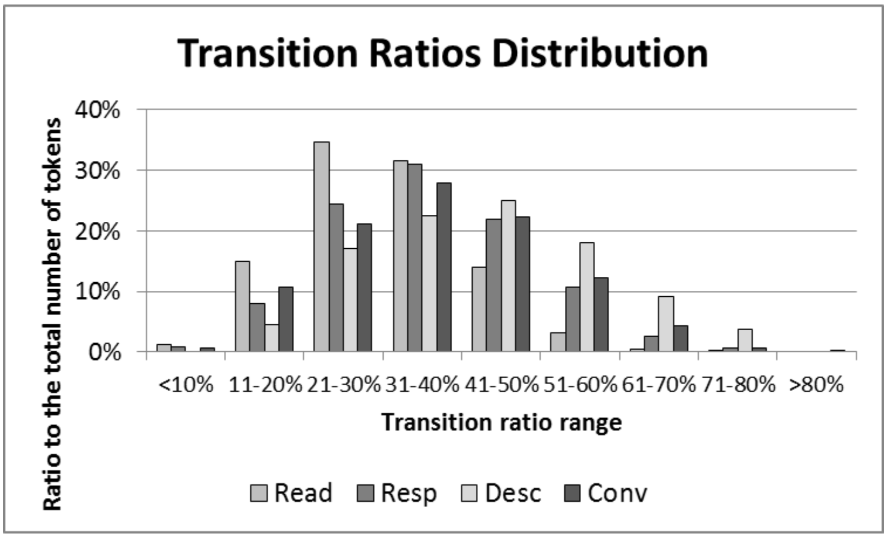


Figure 18 Distribution of transition proportions

Figure 18 shows the transition ratios among speech styles. The central peak for the read speech is in the 21-30% range, but the peak for the response and conversation is in 31-40% and the peak for the description is in 41-50%. The transition ratios of about 90% of the tokens analysed range from 10-60%. The numbers of shorter (less than 10%) transitions are 48 for read speech, 8 for response, 2 for description and 12 for conversation. The numbers of them with larger transition ratios (more than 60%) are 20 for read speech, 22 for response, 48 for description and 84 for conversation.

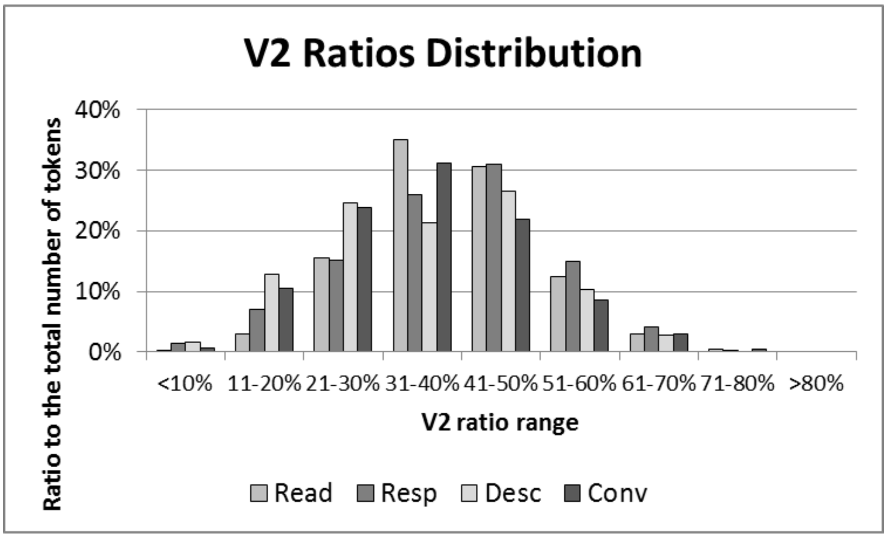


Figure 19 Distribution of V2 proportions

Figure 19 presents the V2 ratios among speech styles. The central peaks for the read speech and conversation are in 30-40%, the peak for the response is in 40-50% and the peak for the description has a wide range of 20-50%. V2 ratios of about 95% of the sequences are between 10-60%. The numbers of shorter (less than 10%) V2 are 4 for read speech, 12 for response, 10 for description and 10 for conversation. The numbers of them with larger V2 ratios (more than 60%) are 118 for read speech, 29 for response, 11 for description and 60 for conversation.

The ratios of V1, transition and V2 to the total vowel sequence durations are investigated further to investigate whether there are effects arising from the speech styles, accent types, the sonority scale, the open/close dimension and the front/back dimension as is the same with the total duration and morphological condition. A MANOVA was conducted was conducted on the three dependent variables (V1, Transition and V2 ratios) with the speech style conditions (read, response, description and conversation) as independent variables. There was a significant effect for the speech style conditions (Wilks' Lambda = .899, F (9, 14659) = 72.776, $p < .001$, ω^2 =. 100), although the effect size was small. Also there was a significant effect for V1 (F (3, 6025) = 88.87, $p < .001$, $\omega^2 = .041$), transition (F (3, 6025) = 197.91, $p < .001$, $\omega^2 = .087$) and V2 (F (3, 6025) = 54.47, $p < .001$, $\omega^2 = .026$) ratios. Despite the small effect size, the speech style was found to have an effect for the variation in the ratios of V1, transition and V2 of VV sequences in Japanese. Scheffe post-hoc tests were also conducted to reveal if there is a significant difference on the interval ratios among the speech styles. Table 28 presents all the pairwise comparisons for the Scheffe test.

Table 28 Post-hoc comparisons for the interval ratios for the speech style conditions

	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
V1	rd>rs**	0.52	M	rd>dc**	0.77	M	rd>cv**	0.22	S
	rs>dc*	0.23	S	rs<cv**	-0.26	S	dc<cv**	-0.47	S
Transition	rd<rs**	-0.52	M	rd<dc**	-1.14	L	rd<cv**	-0.56	M
	rs<dc**	-0.50	M				dc>cv**	0.43	S
V2				rd>dc**	0.39	S	rd>cv**	0.36	S
	rs>dc**	0.31	S	rs>cv**	0.30	S			

*: $0.001 < p < 0.05$, **: $p < 0.001$

S: Small, M: Medium, L: Large

Regarding V1 ratios, VVs of the description style had the smallest ratio and VVs of the read speech had the largest ratio. Ordering from smallest to largest, it was Desc < Resp < Conv < Read. As is seen in Table 17 above, the all the pairwise comparisons were significant. The read to response difference and the read to description difference had a medium effect size for the comparison and other pairwise comparisons had a small effect size. For transition ratios, VVs of the read speech had the smallest ratio and VVs of the description style had the largest ratio, which ascending order is Read < Resp < Conv < Desc. There was no significant difference between the response style and the conversation style, but all the other pairwise comparisons were significant. The difference between the description style and the conversation style had small effect size but the other comparisons had medium or large effect size. As for V2 ratios, VVs of the description style had the smallest ratio and VVs of the read speech had the largest ratio, its order was Desc < Conv < Resp < Read. Although there was no significant difference between the read speech and the response style as well as between the description style and the conversation. However, the effect size for the pairwise comparisons of V2 ratio was small. Thus, VVs of the read speech had the longest steady states and the shortest transition. Although it is still a matter of question that how other styles are spontaneous or natural, they had shorter steady states and longer transition than the read speech. The current result shows that VVs obtained by describing pictures had the shortest steady states and the longest transition. This finding corresponds to the argument that the vowel sequences in Japanese become diphthong-like in ‘natural’, ‘connected’ and ‘spontaneous’ speeches (Kawakami, 1977; Saito, 1997). This finding also agrees with the result for total VV duration. Although the transition parts occupy larger part of the vowel sequences compared to the findings of the result in the morphological condition, there are still steady states for the two target vowels.

4.3.3.2 Accent contour

Having looked at the speech style condition, we now consider the possible effects of properties of the vowels themselves on the realisation of Japanese VV sequences. Firstly the accent pattern of the VV sequences is investigated. However, it was only possible to look at the effect of the variation of the accent type in the read speech and the conversation because the target words for the other styles had HH accent only. Table 29 below presents the subsection ratios of different speech styles sorted by the accent types.

Table 29 Proportions of V1, transition and V2 to the total VV durations in four accent types

		HH		HL		LH		LL	
		Mean	S.D	Mean	S.D	Mean	S.D	Mean	S.D
Read	V1	30.7%	10.4%	30.4%	10.3%	27.8%	9.4%	32.0%	10.8%
	Transition	30.5%	10.6%	30.5%	10.9%	30.2%	10.5%	30.5%	9.8%
	V2	38.8%	10.9%	39.2%	10.4%	42.0%	10.0%	37.5%	10.8%
Resp	V1	24.8%	11.1%						
	Transition	36.1%	12.4%						
	V2	39.2%	12.9%						
Desc	V1	22.2%	10.5%						
	Transition	42.7%	14.5%						
	V2	35.1%	13.2%						
Conv	V1	26.4%	11.5%	28.5%	12.3%	27.9%	12.6%	28.6%	12.0%
	Transition	37.5%	13.5%	36.7%	13.5%	36.3%	13.8%	36.2%	12.7%
	V2	36.1%	13.2%	34.8%	12.1%	35.8%	12.5%	35.2%	11.4%

Because of the limitation in data collection, VV sequences with a HH accent type could be obtained. Therefore, we are going to compare the different speech style of the same accent type (HH) and the relation between two speech styles (read speech and conversation) and four accent types (HH, HL, LH and LL).

Firstly, the subsection proportions of VV sequences with HH accent types is examined. An ANOVA was conducted on the three dependent variables (V1, Transition and V2 ratios) with the four speech styles as independent variables. It showed a significant difference in all variables (V1: *F* (3, 2246) = 62.781, $p < .001$, $\omega^2 = 0.074$, transition: *F* (3, 2246) = 88.149, $p < .001$, $\omega^2 = 0.101$, V2: *F* (3, 2246) = 12.913, $p < .001$, $\omega^2 = 0.016$).

Secondly, A MANOVA was conducted on the three dependent variables (V1, Transition and V2 ratios) of the read speech and the conversation styles with the accent contour of VV sequences (HH, HL, LH and LL) as independent variables. The MANOVA showed a significant effect for the speech style difference between the read speech and the conversation (Wilks' Lambda = .995, *F* (9, 12402) = 2.872, p = .002, ω^2 = .003), although the effect size was small. However, there was an effect for the V1 ratio (*F* (3, 5098) = 6.771, p < .001, ω^2 = 0.0004) and V2 ratio (*F* (3, 5098) = 4.714, p = .003, ω^2 = 0.0002) but no difference was found for the transition ratio (*F* (3, 5098) = 0.475, p = .7). In addition, an ANOVA was conducted on the three dependent variables (V1, Transition and V2 ratios) of the conversation style with the accent contour of VV sequences (HH, HL, LH and LL) as independent variables. It showed a significant difference in the V1 ratio only (*F* (3, 1671) = 2.943, p = .032, ω^2 = 0.003) and there was no significant difference for transition and V2 (transition: *F* (3, 1671) = 0.610, p = .609, V2: *F* (3, 1671) = 1.148, p = .329). Scheffe post-hoc test was also conducted to reveal if there is a significant difference on the interval ratios among the accent contour. For V1, the ascending order was HH < LH < HL <LL and the significant difference was found only between HH and HL.

4.3.3.3 Vowel sequence types

4.3.3.3.1 Sonority

In addition to the tonal differences in VV sequences, the sonority differences between vowels in VV sequences were also investigated. The sonority scale was proposed in Carlyle (1987) to classify VV sequences in Breton based on the sonority differences between V1 and V2 in VV sequences. Under this framework, VV sequences are classified into three categories; falling, rising and level sequences. If V1 is more sonorous than V2, then this sequence is a falling sequence. If V2 is more sonorous than V1, then this sequence is a rising sequence. If the sonority of V1 and V2 is equal, then it is called a level sequences. Carlyle (1987) classified VV sequences of Breton and Spanish and discussed the conditions of the diphthongisation of VV sequences. The current research applied this framework to Japanese VV sequences. Table 30 below presents the mean ratios and the standard deviations of V1, transition and V2 in Japanese.

Table 30 Proportions of V1, transition and V2 to the total VV durations in different sonority scale

		Falling		**Rising**		**Level**	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
V1	Read	32.1%	10.7%	29.5%	10.1%	28.7%	9.9%
	Resp	23.8%	10.6%	25.9%	11.4%	24.4%	11.0%
	Desc	22.4%	10.8%			21.9%	10.2%
	Conv	28.2%	12.1%	26.7%	12.6%	26.3%	11.3%
	Overall	29.0%	11.8%	28.7%	10.7%	27.2%	10.5%
Transition	Read	30.4%	10.4%	29.3%	10.6%	32.5%	9.8%
	Resp	36.9%	12.8%	35.7%	12.7%	35.4%	11.4%
	Desc	43.4%	15.0%			41.6%	13.6%
	Conv	37.2%	13.6%	34.4%	12.7%	36.8%	13.7%
	Overall	34.9%	13.1%	30.7%	11.4%	34.4%	11.4%
V2	Read	37.5%	9.4%	41.2%	11.3%	38.7%	10.6%
	Resp	39.3%	12.0%	38.5%	13.9%	40.2%	12.3%
	Desc	34.2%	13.9%			36.5%	12.1%
	Conv	34.6%	12.3%	38.9%	12.7%	36.9%	12.2%
	Overall	36.1%	11.4%	40.6%	11.8%	38.4%	11.2%

Firstly, A MANOVA was conducted on the three dependent variables (V1, Transition and V2 ratios) with the sonority scale of VV sequences (rising, falling and level sequences) as independent variables. It revealed that there was a significant difference among the three sonority scale groups on the dependent measures (Wilks' Lambda = .993, *F* (15, 16608) = 2.807, $p < .001$, $\omega^2 = .005$). Also, there was a significant effect for V1 (*F* (5, 6018) = 4.586, $p < .001$, $\omega^2 = 0.0004$), the transition (*F* (5, 6018) = 74.525, $p = .004$, $\omega^2 = 0.0002$), and V2 (*F* (5, 6018) = 4.204, $p < .001$, $\omega^2 = 0.0002$). Although the effect size was very small, the sonority scale had an effect for the subsection ratios.

Secondly, the interaction between the subsection proportion and the sonority in each speech style condition was investigated. An ANOVA was conducted for the subsection ratios of each speech style condition with the sonority as an independent variable. In a condition of the responding style, an ANOVA did not show significant differences in V1 and V2 proportions (V1: *F* (2, 592) = 2.150, p = .117, transition: *F* (2, 592) = 0.793, p = .453, V2: *F* (2, 592) = 0.781, p = .458). In the case of the description style, an ANOVA also did not show significant differences in all subsection proportions (V1: *F* (1, 326) = 0.183, p = .669, transition: *F* (1, 326) = 1.253, p = .264, V2: *F* (1, 326) = 2.455, p = .118). In the conversation style, an ANOVA showed significant differences in the transition and the V2 proportion (transition: *F* (2, 1672) = 4.182, p = .015, $\omega^2 = 0.004$, V2: *F* (2, 1672) = 12.797, p < .001, $\omega^2 = 0.014$). There was no significant difference in the V1 proportion (*F* (2, 1672) = 2.694, p = .068). Table 31 below presents all the pairwise comparisons for the Scheffe test.

Table 31 Post-hoc comparisons for the interval ratios for the sonority scale

		Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
Read	**V1**	F>R**	0.25	S	F>L**	0.33	S			
	Transition	F>R*	0.11	W	F<L**	0.21	S	R<L**	0.31	S
	V2	F<R**	0.36	S	F<L*	0.13	W	R<L**	0.22	S
Resp	**V1**									
	Transition									
	V2									
Desc	**V1**									
	Transition									
	V2									
Conv	**V1**									
	Transition	F>R*	0.21	S						
	V2	F<R**	0.35	S						

Regarding the V1 ratio, the ascending order was Level < Rising < Falling. But most of its significance was contributed by the read speech. In the conversation style, the transition of the falling sequences was significantly larger than that of the rising sequences and the V2 of rising sequences was significantly larger than that of the falling sequences. The effect sizes for both comparisons were small.

4.3.3.3.2 Closing-Opening dimension

Secondly the articulatory characteristics of VV sequences are investigated. VV sequences can theoretically be produced by the opening/closing of the mouth and the fronting/backing of the tongue. Table 32 below presents the mean proportions, standard deviations and ranges of V1, transition and V2 of the Open/Close dimension. The movement of closing and opening was classified by the degree of the movement. Close 2 is the movement from the open vowel to the close and Close 1 is the closing movement from/to the mid vowels such as /e/ and /o/. N.D. refers to No Difference; both vowels in VV sequences are described as the same height such as /iu/ or /eo/.

Table 32 Proportions of V1, transition and V2 to the total VV durations in the Open/Close dimension

		Close2		**Close1**		**N.D.**		**Open1**		**Open2**	
		Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
V1	Read	35.4%	10.4%	30.4%	10.5%	28.7%	9.9%	30.5%	10.4%	26.4%	8.7%
	Resp	27.0%	10.9%	21.7%	9.8%	24.4%	11.0%	25.7%	11.8%	26.5%	9.8%
	Desc	23.2%	12.1%	21.9%	9.8%	21.9%	10.2%				
	Conv	28.7%	12.2%	25.5%	11.5%	26.3%	11.3%	26.6%	12.8%	29.2%	9.1%
	Overall	30.0%	12.2%	27.8%	11.1%	27.2%	10.5%	29.4%	11.1%	26.5%	8.8%
Transition	Read	28.2%	9.9%	31.6%	10.5%	32.5%	9.8%	29.0%	10.2%	30.2%	11.6%
	Resp	36.3%	13.1%	37.3%	12.6%	35.4%	11.4%	36.7%	12.3%	31.5%	13.4%
	Desc	44.8%	15.9%	42.5%	14.5%	41.6%	13.6%				
	Conv	36.9%	13.4%	38.6%	14.1%	36.8%	13.7%	34.1%	12.8%	39.2%	8.7%
	Overall	35.2%	13.5%	34.5%	12.5%	34.4%	11.4%	30.7%	11.3%	30.6%	11.8%
V2	Read	36.4%	9.1%	38.0%	9.6%	38.7%	10.6%	40.5%	10.8%	43.3%	12.4%
	Resp	36.7%	12.0%	40.9%	11.8%	40.2%	12.3%	37.6%	14.0%	42.1%	13.3%
	Desc	32.0%	14.8%	35.6%	13.1%	36.5%	12.1%				
	Conv	34.3%	11.9%	35.9%	13.7%	36.9%	12.2%	39.3%	12.9%	31.5%	6.3%
	Overall	34.8%	11.5%	37.7%	11.1%	38.4%	11.2%	39.9%	11.6%	42.9%	12.5%

Firstly, A MANOVA was conducted to investigate whether there is a significant relation between the speech style condition and the different degree of opening and closing movement for VV sequences. Although its effect size was small, it showed a significant difference (Wilks' Lambda = .986, *F* (30, 17638) = 2.873, $p < .001$, $\omega^2 = .009$). In addition, the proportions of all subsection proportions were found to be significantly different (V1: *F* (10, 6011) = 4.798, $p < .001$, $\omega^2 = 0.001$, transition: *F* (10, 6011) = 4.561, $p < .001$, $\omega^2 = 0.001$ and V2: *F* (10, 6011) = 3.140, $p < .001$, $\omega^2 = 0.0003$).

Secondly, the interaction between the subsection proportion and the Opening/Closing dimension in each speech style condition was investigated. An ANOVA was conducted for the subsection ratios of each speech style condition with the Open/Close dimension as an independent variable. In a condition of the responding style, an ANOVA showed significant differences in the V1 and V2 proportions (V1: *F* (4, 590) = 4.065, p = .003, $\omega^2 = 0.020$, V2: *F* (4, 590) = 2.935, p = .020. $\omega^2 = 0.013$). But there was no significant difference in the transition proportion. (*F* (4, 590) = 2.177, p = .070). In the case of the description style, an ANOVA showed a significant difference in the V2 proportions (*F* (2, 325) = 3.092, p = .047, $\omega^2 = 0.013$). It did not show significant differences in the V1 and the transition proportions (V1: *F* (2, 325) = 0.493, p = .611, transition: *F* (2, 325) = 1.228, p = .294). In the conversation style, an ANOVA showed significant differences in the all subsection proportions (V1: *F* (4, 1670) = 4.814, p = .001, $\omega^2 = 0.009$, transition: *F* (4, 1670) = 3.163, p = .013, $\omega^2 = 0.005$, V2: *F* (4, 1670) = 8.119, $p < .001$, $\omega^2 = 0.017$). Table 33 below presents all the pairwise comparisons for the Scheffe test.

Table 33 Post-hoc comparisons for the interval ratios for the opening/closing dimension

		Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
Read	V1	C2>C1**	0.48	S	C2>ND**	0.66	M	C2>O1**	0.47	S
		C2>O2**	0.94	L	C1>ND**	0.17	W			
		C1>O2**	0.4	S	ND<O1*	0.18	W	ND>O2**	0.24	S
		O1>O2*	0.41	S						
	Transition	C2<C1*	0.33	S	C2<ND**	0.44	S			
								C1>O1**	0.25	S
					ND>O1**	0.35	S			
	V2				C2<ND*	0.23	S	C2<O1**	0.39	S
		C2<O2**	0.65	M				C1<O1**	0.24	S
		C1<O2**	0.51	M	ND<O1*	0.16	W	ND<O2**	0.41	S
		O1<O2**	0.26	S						
Resp	V1	C2>C1*	0.52	M						
								C1<O1*	0.36	S
Conv	V1	C2>C1*	0.27	S						
	Transition									
								C1<O1*	0.33	S
	V2							C2<O1**	0.23	S

*: 0.001 < p < 0.05, **: p < 0.001

W: Weak, S: Small, M: Medium, L: Large

The pairwise comparison for the description style showed no significant difference. Regarding the V1 ratios, the ascending order was Open 2 < N.D < Close 1 < Open 1 < Close 2. There was a significant difference between Close 1 and Close 2 in the response and the conversation style, although the effect size was small. On the transition ratios, the ascending order was Open 2 < Open 1 < N.D < Close 1 < Close 2. But the significant differences were mainly due to the read speech. A Significant difference was found between Close 1 and Open 1 for the conversation style. As for the V2 ratios, the ascending order was Close 2 < Close 1 < N.D < Open 1 < Open 2, though the significant differences were mainly due to the read speech. A Significant difference was found between Close 2 and Open 1 for the conversation style. Therefore, the opening and the closing dimensions is not a major factor of changing subsection proportions in different speech styles.

4.3.3.3.3 Fronting-Backing dimension

Then the fronting and backing movement is considered. Table 34 below presents the mean proportions, standard deviations and ranges of V1, transition and V2 of the Front/Back dimension. The movement of fronting and backing was classified by the degree of the movement. Front 2 is the movement from the back vowel to the front and Front 1 is the fronting movement from/to the central vowels such as /a/. N.D. refers to No Difference; both vowels in VV sequences are described as the same tongue advancement such as /uo/ or /ie/.

Table 34 Proportions of V1, transition and V2 to the total VV durations in the Front/Back dimension

		Front2		**Front1**		**N.D.**		**Back1**		**Back2**	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
V1	Read	25.7%	8.9%	30.1%	10.6%	32.6%	10.3%	31.1%	9.9%	32.8%	10.4%
	Resp	21.7%	8.9%	24.7%	10.5%	31.2%	13.4%	24.5%	11.0%	25.0%	11.1%
	Desc	20.3%	9.5%	18.8%	8.5%			27.5%	13.9%	25.3%	10.1%
	Conv	26.6%	11.6%	28.4%	12.1%	37.8%	12.9%	27.3%	10.6%	19.8%	8.0%
	Overall	25.0%	9.8%	28.3%	11.6%	33.0%	11.2%	29.9%	10.5%	29.9%	11.2%
Transition	Read	31.9%	10.3%	29.6%	10.4%	27.3%	10.6%	30.7%	10.4%	31.0%	10.2%
	Resp	37.7%	11.9%	38.5%	13.0%	33.8%	13.6%	32.8%	11.4%	35.7%	11.9%
	Desc	48.9%	13.5%	48.4%	14.7%			35.1%	10.4%	33.7%	10.2%
	Conv	37.8%	13.5%	37.3%	13.6%	27.7%	12.6%	32.4%	12.4%	37.1%	11.4%
	Overall	35.2%	12.5%	35.1%	13.4%	28.2%	11.5%	31.2%	10.7%	32.4%	10.8%
V2	Read	42.3%	10.6%	40.3%	11.2%	40.1%	10.8%	38.2%	9.7%	36.2%	10.0%
	Resp	40.6%	12.6%	36.8%	13.7%	35.1%	15.5%	42.7%	11.6%	39.3%	11.1%
	Desc	30.8%	12.4%	32.8%	14.4%			37.4%	13.2%	40.9%	10.8%
	Conv	35.6%	12.4%	34.3%	12.0%	34.6%	12.1%	40.3%	15.1%	43.0%	11.2%
	Overall	39.8%	12.0%	36.6%	12.3%	38.8%	11.9%	38.8%	10.6%	37.7%	10.6%

Firstly, A MANOVA was conducted to investigate whether there is a significant relation between the speech style condition and the different degree of fronting and backing movement for VV sequences. Even though its effect size was small, it showed a significant difference (Wilks' Lambda = .940, *F* (33, 17701) = 11.386, $p < .001$, $\omega^2 = .055$). In addition, the proportions of all subsection proportions were found to be significantly different (V1: *F* (11, 6010) = 15.599, $p < .001$, $\omega^2 = 0.003$, transition: *F* (11, 6010) = 13.630, $p < .001$, $\omega^2 = 0.002$ and V2: *F* (11, 6010) = 21.581, $p < .001$, $\omega^2 = 0.003$).

Secondly, the interaction between the subsection proportion and the Fronting/Backing dimension in each speech style condition was investigated. An

ANOVA was conducted for the subsection ratios of each speech style condition with the Front/Back dimension as an independent variable. In a condition of the responding style, an ANOVA showed significant differences in the all subsection proportions (V1: $F\ (4, 590) = 8.493$, $p < .001$, $\omega^2 = 0.047$, transition: $F\ (4, 590) = 4.634$, $p = .001$, $\omega^2 = 0.024$, V2: $F\ (4, 590) = 5.547$, $p < .001$, $\omega^2 = 0.030$). In the case of the description style, an ANOVA also showed significant differences in the all subsection proportions (V1: $F\ (3, 324) = 11.348$, $p < .001$, $\omega^2 = 0.084$, transition: $F\ (3, 324) = 35.161$, $p < .001$, $\omega^2 = 0.221$, V2: $F\ (3, 324) = 12.250$, $p < .001$, $\omega^2 = 0.090$). In the conversation style, an ANOVA showed significant differences in the all subsection proportions (V1: $F\ (4, 1670) = 25.004$, $p < .001$, $\omega^2 = 0.053$, transition: $F\ (4, 1670) = 9.876$, $p < .001$, $\omega^2 = 0.021$, V2: $F\ (4, 1670) = 15.848$, $p < .001$, $\omega^2 = 0.034$). Table 35 below presents all the pairwise comparisons for the Scheffe test.

Table 35 Post-hoc comparisons for the interval ratios for the fronting/backing dimension

		Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
Read	V1	F2<F1**	0.45	S	F2<ND**	0.73	M	F2<B1**	0.58	M
		F2<B2**	0.74	M						
		F1<B2**	0.26	S						
		B1<B2*	0.17	W						
	Transition	F2>F1**	0.23	S	F2>ND**	0.44	S			
					F1>ND**	0.22	S			
					ND<B1**	0.32	S	ND<B2**	0.35	S
	V2	F2>F1*	0.19	W	F2>ND*	0.21	S	F2>B1**	0.41	S
		F2>B2**	0.60	M				F1>B1*	0.20	S
		F1>B2**	0.39	S				ND>B2**	0.38	S
		B1>B2*	0.20	S						
Resp	V1				F2>ND**	0.91	L			
					F1>ND*	0.57	M			
					ND>B1*	0.56	M	ND>B2*	0.52	M
	Transition							F2>B1*	0.42	S
								F1>B1*	0.46	S
	V2									
								F1<B1*	0.46	S
					ND<B1*	0.58	M			
Desc	V1							F2<B1*	0.68	M
		F2<B2*	0.52	M				F1<B1**	0.86	L
		F1<B2**	0.7	M						
	Transition							F2>B1**	1.09	L
		F2>B2**	1.27	L				F1>B1**	0.99	L
		F1>B2**	1.18	L						
	V2									
		F2<B2**	0.87	L						
		F1<B2**	0.65	M						
Conv	V1				F2<ND**	0.94	L			
		F2>B2**	0.64	M	F1<ND**	0.77	M			
		F1>B2**	0.72	M	ND>B1**	0.89	L			
		B1>B2	0.83	L						

	Transition				F2>ND**	0.76	M			
					F1>ND**	0.71	M			
								ND<B2*	0.80	M
	V2									
		F2<B2**	0.61	M						
		F1<B2**	0.63	M				ND<B2*	0.74	M

*: 0.001 < p < 0.05, **: p < 0.001

W: Weak, S: Small, M: Medium, L: Large

Regarding the V1 ratios, the ascending order was Front 2 < Front 1 < Back 1 = Back 2 < N.D. It should be noted that the effect sizes for pairwise comparisons in the responding, describing and conversation styles were large. Significant differences were found between Front 2 and Front 1, Front 1 and Back 1, Front 1 and Back 2, Back 1 and N.D and Back 2 and N.D. Thus, backing sequences had larger V1 proportions than fronting sequences. On the transition ratios, the ascending order was N.D < Back 1 < Back 2 < Front 1 < Front 2. The effect sizes for the pairwise comparisons of the describing style is large. Significant differences were found between N.D and Back 1 and Back 2 and Front 1. Therefore, the fronting sequences had larger transition ratio than the backing sequences. As for the V2 ratios, the ascending order was Front 1 < Back 2 < N.D = Back 1 = Front 2. There was a significant difference between Front 1 and Front 2 with a small effect size but no other pairwise comparisons showed more than minimum values for a small effect size. Therefore, the V1 ratios and the transition ratios were related to the movement of fronting and backing of the tongue.

4.4 Correlation between the total VV duration and subsection proportions

So far, we examined the total duration of the vowel sequences and the proportions of V1, transition and V2 to the total duration of vowel sequences. In terms of the total duration, various factors such as the morphological condition, the speech style condition and the accent type contributed to the significant difference. However, there was only a slight difference in terms of the subsection proportions. It was true that the range of subsection proportions was large and the proportions of subsections ranged from a few per cents to more than 70 % of the total VV duration, the mean proportions were not so different with each other. The purpose of this section is to investigate whether there is a correlation between the total duration of vowel sequences and the proportions of subsections in vowel sequences.

4.4.1 *Morphological condition*

The scatter plot of Figure 20 shows how the total duration of VV sequences and the ratios of V1 steady states are correlated. There was a weak positive correlation between them (*r* (5956) = 0.21, p < .001). The longer the total VV duration, the larger the V1 ratio becomes. It should be noted that VV sequences with shorter duration would have larger V1 proportions; it was unlikely that VV sequences with longer duration had smaller V1 proportions.

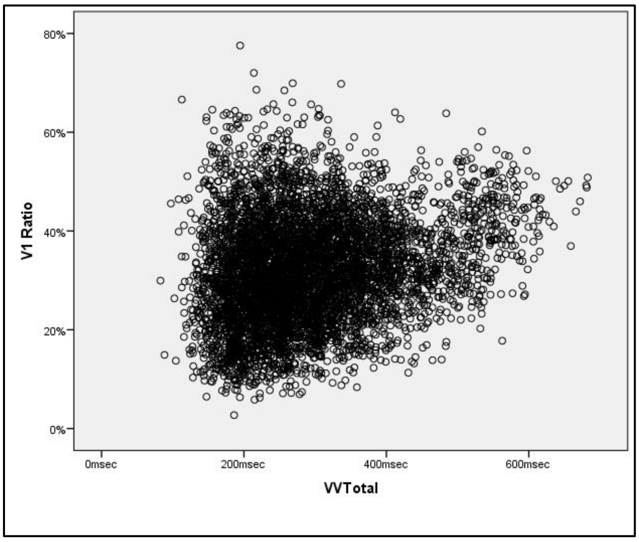


Figure 20 Correlation between total VV duration and V1 Ratio

Figure 21 below presents the correlation between the total VV duration and the transition ratio. There was a moderate negative correlation between them (r (5956) = -0.32, p < .001). The longer the VV duration is, the smaller the transition proportion becomes. Although VV sequences with shorter duration had smaller transition proportions in many cases, it was less likely that VV sequences with longer duration had larger transition proportions.

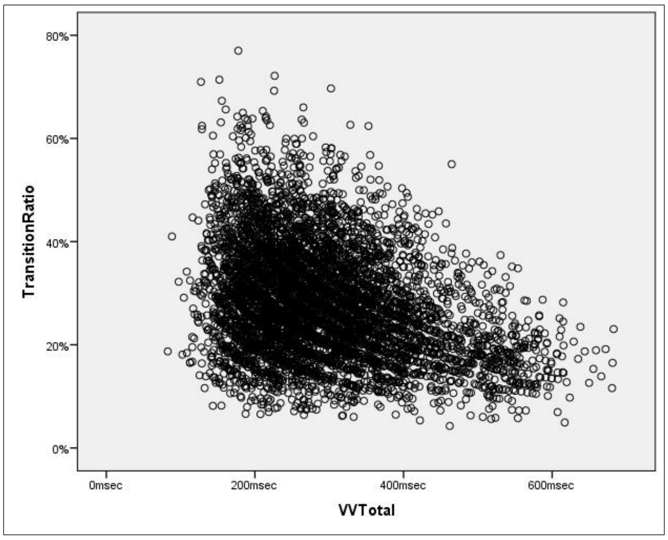


Figure 21 Correlation between total VV duration and Transition Ratio

Figure 22 below shows the correlation between the total VV duration and the V2 ratio. There was a weak positive correlation between them. The longer the total duration is, the larger the V2 ratio becomes. As is the same as the V1 steady state, VV sequences with shorter duration had larger V2 proportion in many cases but VVs with longer duration were less likely to have smaller V2 proportion.

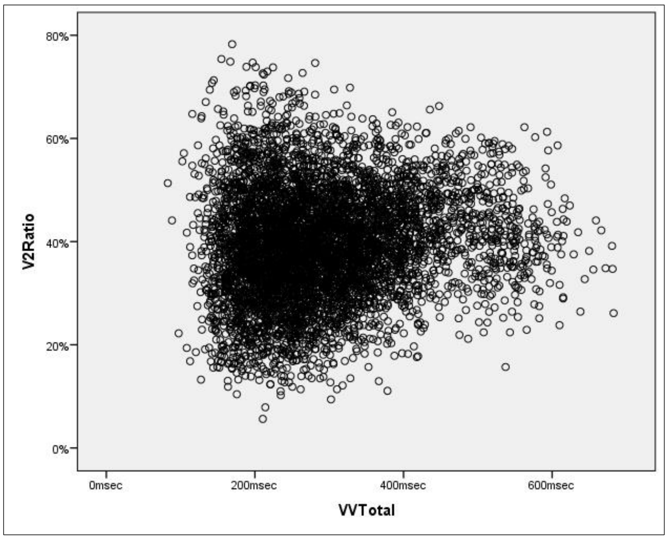


Figure 22 Correlation between total VV duration and V2 Ratio

Therefore, the short duration, which is said to be one of the features of diphthongs, was compatible with the characteristics of internal structure of the VV sequences. When the duration of the vowel sequences were longer, the proportions of steady states became larger and the transition proportion became smaller. This can be described as one of the acoustic features of diphthongs. Diphthongs are characterised as a gradual formant movement throughout the duration and the current result showed that the shorter duration of VV sequences had larger transition proportions.

4.4.2 *Speech Style Condition*

The scatter plot of Figure 23 presents the correlation of word-internal vowel sequences with different speech styles. There was a weak positive correlation between the total VV duration and the V1 ratio (r (6027) = 0.12, p < .001). VVs with longer duration were likely to have larger V1 proportion. Although VVs with shorter duration had larger V1 proportion in many cases, VVs with longer duration were not likely to have smaller V1 proportion.

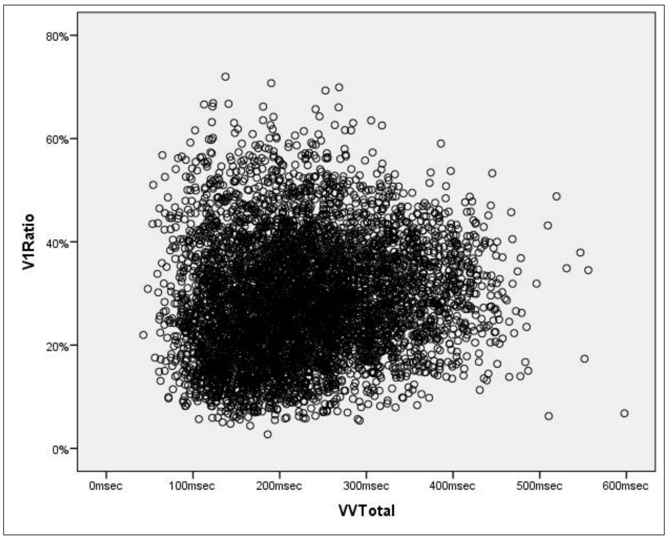


Figure 23 Correlation between total VV duration and V1 Ratio

Figure 24 shows the correlation between the total VV duration and the transition ratio. There was a weak negative correlation between them (r (6027) = -0.23, p < .001). VVs with longer transition were likely to have smaller transition proportion. There were many VV sequences with shorter duration and smaller transition proportion, but it was less likely that VV sequences with longer transition had larger transition proportion.

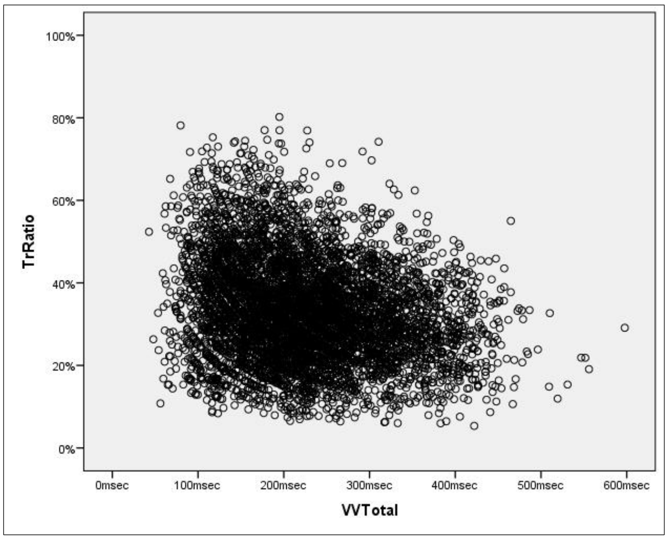


Figure 24 Correlation between total VV duration and Transition Ratio

Figure 25 presents the correlation between the total VV duration and the V2 ratio. There was a weak positive correlation between them (r (6027) = 0.13, p < .001). VVs with longer duration were likely to have larger V2 proportion. There were many VV sequences having shorter duration and larger V2 proportion, but it was unlikely that VVs with longer duration had smaller V2 ratio.

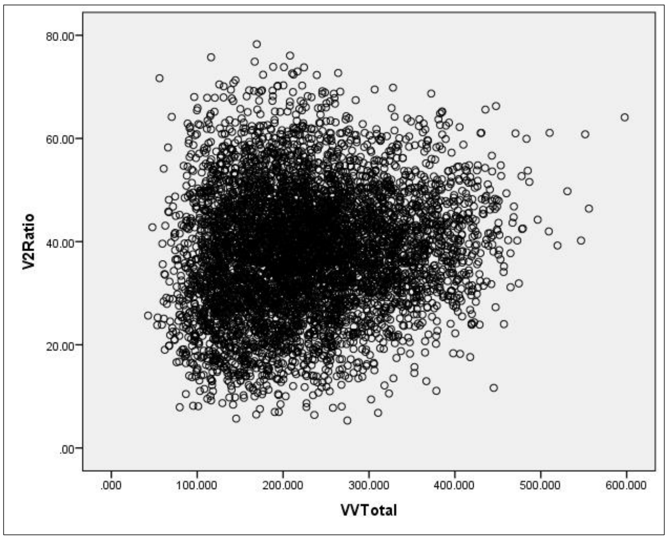


Figure 25 Correlation between total VV duration and V2 Ratio

The correlation also coincided with the argument that diphthongised VV sequences had shorter duration than hiatuses. When the total VV duration is shorter, the proportions of steady states become larger and the transition ratio becomes smaller. The two factors that VV sequences may have diphthong-like features were found to coincide with each other and this finding was recognised in both morphological and speech-style conditions. Shorter VV duration is said to be one of the factors that VV sequences are diphthongised but VV sequences with shorter duration were found to have shorter steady states and longer transition.

4.5 Summary

In this chapter, the overall durations for VV sequences across the morphological conditions and the speech styles were firstly shown. The VV sequences across the word boundary were found to have shorter duration on average than other morphological conditions and VVs in the conversation style were also found to be shorter than other speech style conditions.

Secondly, all VV sequences in Japanese were found to have the steady states for V1, the transition and the steady states for V2. The ratios of those intervals varied across the morphological conditions and the speech style conditions.

The correlations between the overall durations and the ratios of the V1, the transition, and the V2 intervals were also analysed. For both conditions, V1 and V2 ratios had negative correlations but the transition ratios had positive correlations.

Chapter 5 Spectral-domain Analysis

5.1 Introduction

5.1.1 *General remarks*

This chapter reports the spectral domain analysis of Japanese VV sequences. Representative values of F1 and F2 of the first and the second target vowels of a VV sequences were measured and compared in terms of different morphological conditions and speech styles to investigate whether these factors are significant in respect of the realisation of VV sequences. While it is undeniable that vowel sequences in Japanese are phonologically a sequence of monophthong vowels rather than diphthongs, as pointed out in Chapter 2, some researchers have argued that certain VV sequences can be classified as diphthongs. However, it is not still clear whether the vowel qualities of VV sequences are similar to the combination of the two identical singleton productions. It is well known that connected speech influences the properties of articulation and, as a result, vowel quality may be reduced. Also, the nature of the articulatory movement from the first vowel in a VV sequence to the second may result in a reduction of or modification the target vowel quality. The reduction of vowel quality compared with singleton production is related to the features of diphthongisation which have been highlighted by previous studies in the literature. For example, the formant frequencies at the onset and offset of a diphthong, which by convention are transcribed with a digraph consisting of two vowel symbols which are otherwise individually used for singletons are different from those of the corresponding monophthongs, transcribed by the identical symbols (Ladefoged, 2006). Even if a diphthong is transcribed by two monophthong symbols, the quality of the onset and the offset parts transcribed by each symbol is different from monophthongs. According to Jones (1972), the two phonetic symbols representing one diphthong represent the furthest limit of the tongue movement which could be associated with that diphthong but that normally the articulation of diphthongs do not reach these extremes. Therefore, it is to be expected that the articulatory movement of diphthongs do not reach the locations in vowel space that the symbols represents. Based on this assumption, in the present study the formant frequencies of each vowel in VV sequences and the qualities of vowels in singleton production were measured to investigate whether the vowel qualities of the vowel sequences are significantly different from the singleton productions or the same as the singletons, thereby addressing the issue of whether Japanese VV sequences possess diphthong-like characteristics.

In addition to the comparison of formant frequencies across a range of different conditions, the Euclidean distance between the singleton and the target vowels of sequences was calculated and compared in terms of various factors. This measure shows the extent and nature of any differences between vowels in vowel sequences in Japanese and the corresponding singleton vowels.

Moreover, it has been discussed in the literature (See section 2.3) that the speed of formant movement would also characterise diphthongs/hiatuses. Simpson (1998) reported that the increase of the total duration of three German diphthongs is associated with the wider articulatory movement from the onset to the offset and the appearance of the steady states at the onset and the offset. This means that the articulatory movement becomes greater as the total duration of diphthongs becomes longer. Also, the steady states for onset and offset tend to appear in diphthongs of longer duration. In the present study, the Euclidean distance between the first and the second target vowels in Japanese VV sequences are calculated and regarded as the distance of articulatory movement, and then the speed of the movement is calculated by dividing by the transition durations. The results of male and female speakers are presented separately because it can be assumed that the length of the vocal tract is different between gender, therefore meaning that it is not possible to average across the formant values of male and female speakers.

5.1.2 *The Euclidian distance*

The Euclidian distance between singletons and the vowels in VV sequences was calculated by computing the differences in F1 and F2. Regarding the line between V1 and V2 as a hypotenuse of a right triangle, its distance was calculated by Pythagorean theorem; extracting the square root of the sum of the squares of the other two sides, which is equal to the difference of F1 and F2 between singletons and vowels in VV sequences (See figure 26 below). It should be noted that the scales of F1 and F2 are not represented as the scale for length and the distance calculated is not an actual distance in the oral cavity. But the scales for F1 and F2 are the same (in Hz); therefore, the measure calculated represents the distance on the F1-F2 diagram.

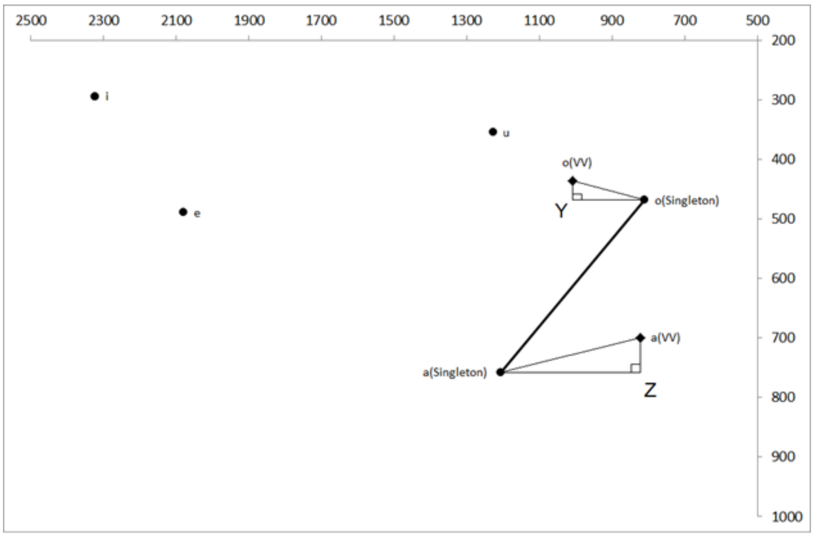


Figure 26 Euclidian distance between singleton and VV sequences

For VV sequences Figure 1 shows the calculation of the Euclidian Distance between the singleton targets (/a/ and /o/) and the realisation of those vowels in the /ao/ sequence. For example, two sides other than the hypotenuse are calculated by subtracting F1 of singleton from F1 of VV sequences and F2 of singleton from F2 of

VV sequences. The Pythagorean Theorem states that the square of the hypotenuse is equal to the sum of the squares of the other two sides. Thus, the relationship of three lines can be expressed as the following equation.

$$Distance^2 = (F1sgl - F1vv)^2 + (F2sgl - F2vv)^2$$

Therefore, the distance between singleton and the vowels in VV sequences is a square root of this value.

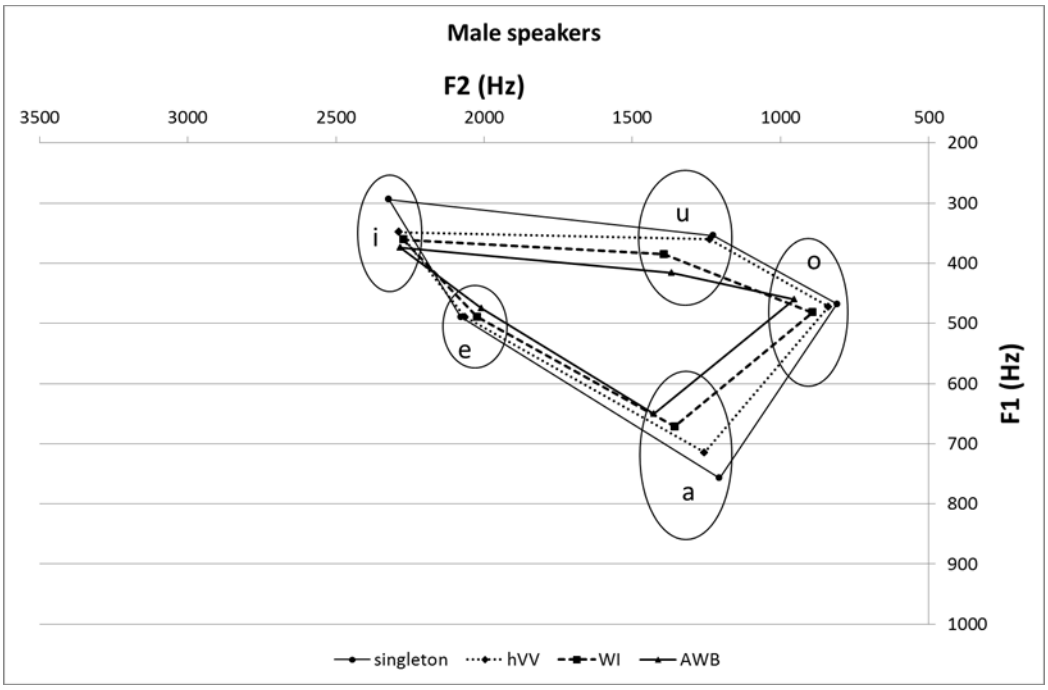
$$Distance = \sqrt{(F1sgl - F1vv)^2 + (F2sgl - F2vv)^2}$$

We now look at the differences in vowel qualities between vowels in VV sequences and vowels in singletons from various points of view. Firstly the difference in F1 and F2 across the different morphological conditions is reviewed and the differences in the Euclidian distance on various factors are considered. Secondly the difference in F1 and F2 across the speech style conditions is reported and the differences in Euclidian distance on the same factors are addressed.

5.2 Morphological Conditions

5.2.1 *General remarks*

The overall comparison of the realisation of all vowels across the different morphological conditions is presented in Figure 27 below. The vertical axis represents the F1 value and the horizontal axis the F2 value. Morphological conditions are presented by the line style and the shape of the marker. The vowels of the same identity are circled.



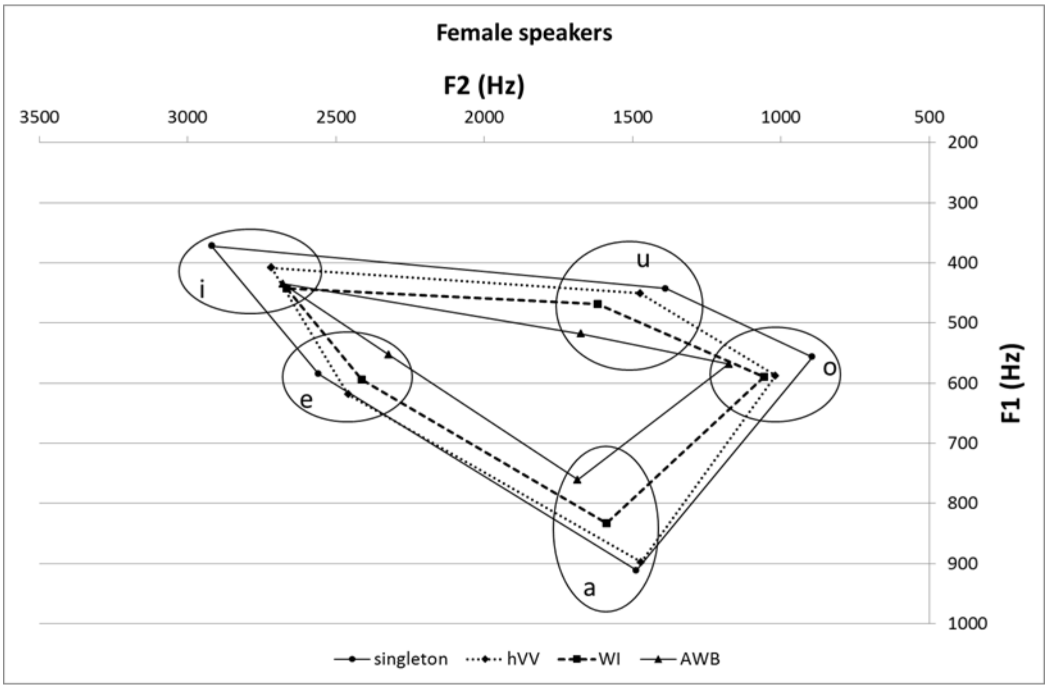


Figure 27 Comparison of vowel quality of five vowels in singleton, hVV, WI and AWB conditions by male (top) and female (bottom) speakers

It can be seen that all the vowels in VV sequences are somewhat ‘centralised ‘compared to singletons, that is, the overall shape of the vowel space of VV sequences are reduced from singletons. Observing the male data, /i/ in VV sequences across all morphological conditions points to the main difference being in F1, relating to the degree of aperture of the vocal tract. Also, the difference between morphological conditions is small for /i/ compared to some of the other vowels.

For female speakers, both F1 and F2 values of /i/ in VV sequences are different from the corresponding singleton. Among the morphological conditions, VVs in /hVV/ condition are found to have lower F1 and VVs in WI and AWB are very close to each other. Regarding /e/, the vowels in VV sequences and in singleton are very close to each other in the case of male speakers but F2 values of the vowels in VV sequences are lower than the singleton in the case of female speakers. In terms of F2 value for /e/, WI is lower than the /hVV/ condition and AWB is further lower than WI. For /a/, the vowels in VV sequences had lower F1 values and higher F2 values than singleton in both male and female speakers. For /o/, F2 values rather than F1 values are affected.

For both male and female speakers, F2 values of the vowels in VV sequences are higher than the vowel in singleton. Lastly for /u/ of male speakers, the /hVV/ condition is at similar position to singleton but the vowels in WI and AWB had higher F1 and higher F2 compared to singleton. For female speakers, /u/ of the vowels in VV sequences had higher F1 and higher F2 than singleton. Therefore, it can be seen that in general vowels in VV sequences are ‘reduced’ by comparison to the corresponding singletons. But the degree of reduction varies across conditions and between F1 and F2.

It is also noticeable that while the vowels are reduced, the five vowels nevertheless remain clearly separate each other.

5.2.2 *Effects of morphological conditions*

To investigate whether the vowels in VV sequences are significantly reduced from singletons and whether there is a significant difference in formant frequencies between the morphological conditions, ANOVAs were conducted. A two-way ANOVA was performed on the F1 and F2 values (in Hz) for each vowel, with the morphological condition (four levels; singleton, hVV, WI and AWB) and vowel identity (five levels; /a/, /i/, /u/, /e/ and /o/) as independent variables. In the ANOVA analysis, the effect of the each of the two independent variables and the interaction between those two factors were investigated. The purpose of the statistical analysis is to find out whether F1 or F2 values (or both) are more ‘centralised’ in F1-F2 vowel space depending on the morphological conditions of the VV sequences. Also, the interaction between the morphological condition and the vowel identity is investigated to see whether all vowels are centralised to the same degree or not.

With respect to F1 values, the two-way ANOVA did not show a significant effect for morphological conditions for male speakers [F (3, 4673) = 0.68, p = 0.56] but there was a main effect for morphological conditions for female speakers [F (3, 7925) =

15.89, $p < .001$, $\omega^2 = 0.002$], although the omega-squared effect size was small. The vowel identity showed an effect for both male [$F(4, 4673) = 2102.73$, $p < .001$, $\omega^2 = 0.41$] and female [$F(4, 7925) = 2042.86$, $p < .001$, $\omega^2 = 0.31$] speakers but it is presumably due to the fact that F1 of five different vowels were tested. There was a medium interaction between the morphological condition and the vowel identity for F1 values for both male and female speakers. The two-way interaction between the morphological condition and the vowel identity was significant for both male [$F(12, 4673) = 25.17$, $p < .001$, $\omega^2 = 0.016$] and female [$F(12, 7925) = 37.31$, $p < .001$, $\omega^2 = 0.018$] speakers, although the effect size was small in both cases. A significant interaction between the morphological condition and the vowel identity implies that the variability of F1 across morphological conditions is significantly different across different vowels. Based on the findings that there is an interaction between the morphological conditions and the vowel identity for F1, the degrees of difference between the vowels in VV sequences and in singletons and the direction of differences was considered in greater detail.

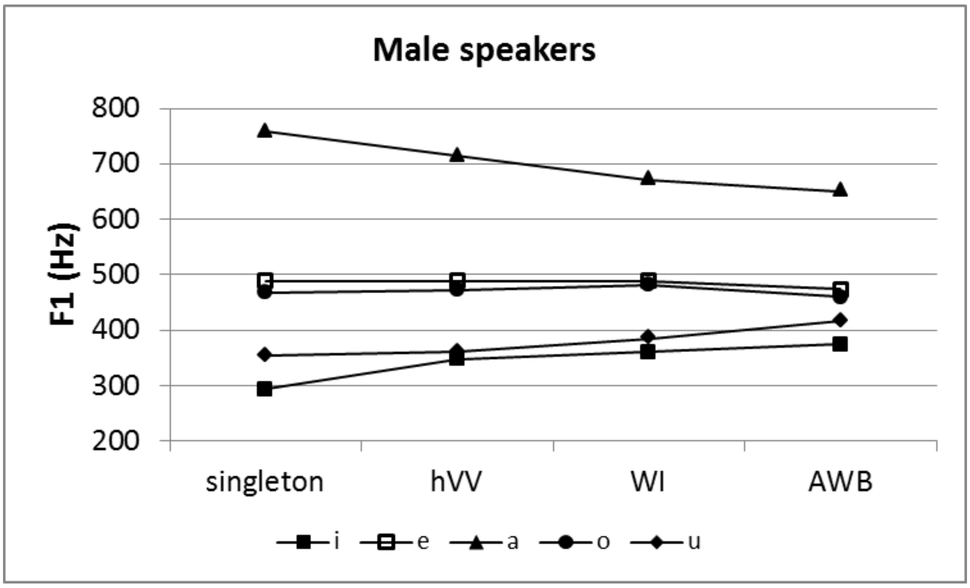
Table 36 presents the F1 of each vowel in singleton and in VV sequences across the morphological conditions. The F1 differences between the vowels in VV sequences across the morphological conditions and the corresponding singletons, that is, the degree of difference from the singleton target is also presented below the F1 values.

Table 36 Mean F1 values (in Hz) of the vowels in singleton and in VV sequences across the morphological conditions

		Male				Female			
		Singleton	hVV	WI	AWB	Singleton	hVV	WI	AWB
i	Frequency	293.9	348.6	360.5	374.3	371.9	408.2	442.6	435.1
	Difference		54.6	66.6	80.4		36.3	70.7	63.2
e	Frequency	489.0	489.3	489.5	474.3	585.1	618.7	594.5	552.6
	Difference		0.3	0.5	14.7		33.6	9.4	32.5
a	Frequency	757.4	714.5	671.8	650.6	911.5	898.0	833.3	760.9
	Difference		42.8	85.5	106.8		13.5	78.2	150.6
o	Frequency	467.8	472.4	482.1	460.3	556.4	587.7	590.1	568.5
	Difference		4.6	14.3	7.4		31.3	33.7	12.1
u	Frequency	354.1	360.4	385.2	416.5	442.7	450.9	468.8	518.1
	Difference		6.3	31.1	62.4		8.2	26.1	75.4

Although F1 for all the vowels in VV sequences are somewhat different from the singletons, the degrees of difference across the morphological conditions are not consistent across the vowel identities. WI is more reduced than the /hVV/ condition and AWB is further reduced from other conditions in /a/ and /u/ for both male and female speakers and in /i/ and /e/ for male speakers. But the degree of difference in WI is larger than AWB in /o/ for both male and female speakers and /i/ for female speakers, although the degrees of both conditions are larger than the /hVV/ condition. /e/ for female speakers presents the inconsistent degree of difference, /e/ in the /hVV/ environment showed the largest difference to singletons. Also, the degree of difference in /e/ and /o/ is smaller than other vowels for male speakers. Female speakers show a similar tendency but it is not consistent in the difference for the /hVV/ condition of /a/ and /u/. It is shown that F1 for vowels in VV sequences are different from the vowels in singletons to some extent, but the degree of difference in F1 values is different both in the morphological conditions and in the vowel identity.

Figure 28 presents the comparison of F1 values of each vowel across the morphological conditions, showing the direction of difference. It is noticed that the open vowel /a/ shows a decrease but the close vowels /i/ and /u/ show an increase in F1 values from singletons to VV sequences. F1 of mid vowels /e/ and /o/ do not show remarkable differences. The ranges of V1 values across the vowel identity are smaller in the vowels in VV sequences than in the vowels in singletons, that is, the vowel quality of the vowels in VV sequences in Japanese is not simply a combination of two corresponding singletons in terms of F1 values.



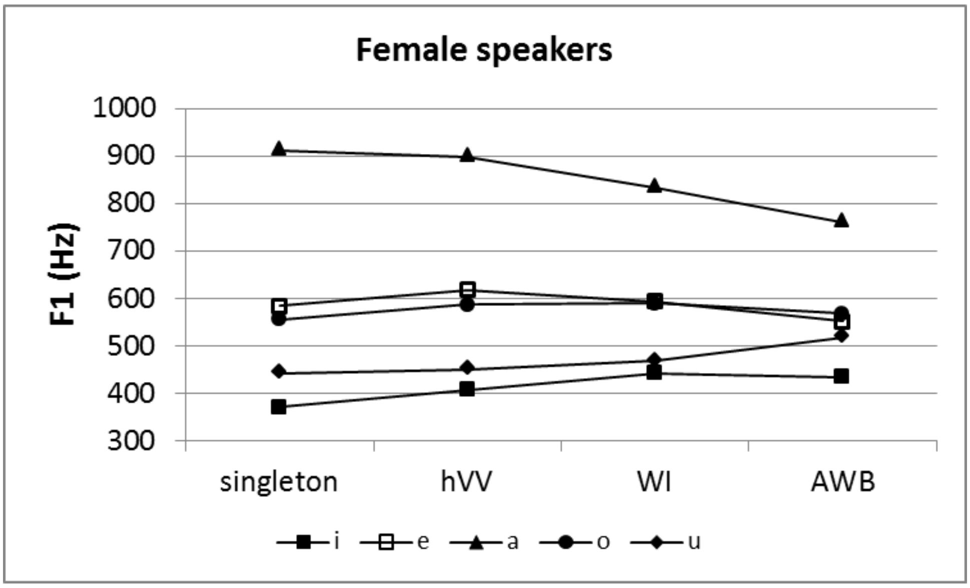


Figure 28 Mean F1 (Hz) of five vowels in singleton, hVV, WI and AWB conditions by male (top) and female (bottom) speakers

For F2, the morphological condition had a significant effect for both male [F (3, 4673) = 31.22, $p < .001$, $\omega^2 = 0.002$] and female [F (3, 7925) = 33.47, $p < .001$, $\omega^2 =$

0.001] speakers, although the effect size was small. The vowel identity showed an effect on F2 for both male [F (4, 4673) = 4827.94, p <.001, ω^2 = 0.46] and female [F (4, 7925) = 7989.28, p < .001, ω^2 = 0.42] speakers, but this is due to the fact that the F2 of five vowels are unsurprisingly very different from each other. The two-way interaction between the morphological condition and the vowel identity was significant for both male [F (12, 4673) = 16.26, p < .001, ω^2 = 0.005] and female [F (12, 7925) = 49.68, p < .001, ω^2 = 0.009] speakers. The effect size of both ANOVAs was medium. As is the same with F1, the variability of F2 across the morphological conditions is significantly different across the different vowels.

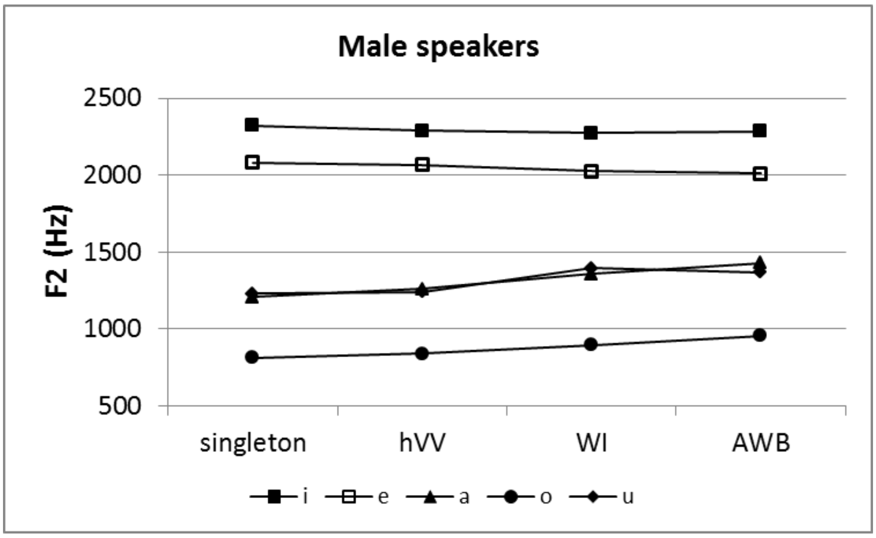
Table 37 presents the F2 of each vowel in singleton and in VV sequences across the morphological conditions. F2 differences between the vowels in VV sequences across the morphological conditions and the corresponding singletons that is, the degree of difference from singletons are also presented below the F2 values.

Table 37 F2 values (in Hz) of the vowels in singleton and in VV sequences across the morphological conditions

		Male				**Female**			
		Singleton	hVV	WI	AWB	Singleton	hVV	WI	AWB
i	Frequency	2321.8	2287.1	2272.1	2283.2	2918.6	2717	2665.3	2679
	Difference		34.7	49.7	38.6		201.7	253.4	239.7
e	Frequency	2079.4	2065.5	2023.8	2010.2	2559.3	2459.3	2412.9	2322.5
	Difference		13.9	55.7	69.2		100	146.4	236.8
a	Frequency	1205.7	1256.9	1357.1	1426.4	1486.8	1471	1587.7	1684.4
	Difference		51.1	151.4	220.7		15.9	100.8	197.5
o	Frequency	809.5	837.9	892.5	954.7	893.9	1018	1056.1	1176
	Difference		28.4	83	145.3		124.1	162.2	282.1
u	Frequency	1227.3	1239.5	1393.6	1367.2	1388.5	1475.1	1617.4	1673.3
	Difference		12.2	166.2	139.9		86.5	228.8	284.8

Regarding the degree of difference in F2 values across the morphological conditions, the difference in WI was larger than the /hVV/ condition and the difference in AWB is further larger than WI in /e/, /a/, /o/ for both male and female speakers and /u/ for female speakers. In cases of /e/ for both male and female speakers and /u/ for male speakers, the differences to singletons are larger in WI than AWB.

Figure 29 presents the comparison of F2 values of each vowel across morphological conditions. It can be seen that the front vowels /i/ and /e/ shows a decrease but the central and back vowels /a/, /u/ and /o/ show an increase in F2 values from singletons to VV sequences. The vowels in VV sequences are found to be more ‘centralised’ compared to singletons. The vowel quality of VV sequences in Japanese is not simply a combination of two corresponding singletons in terms of F2 values.



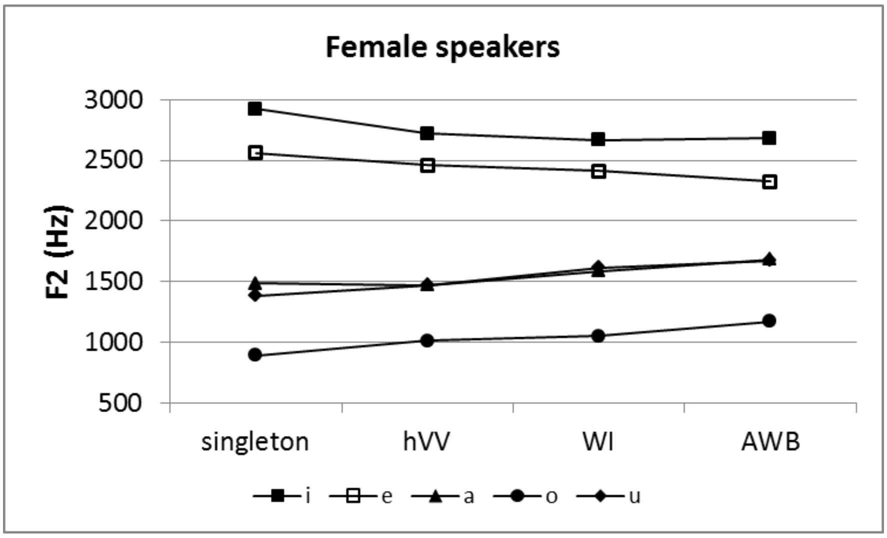


Figure 29 Mean F2 (Hz) of five vowels in singleton, hVV, WI and AWB conditions by male (top) and female (bottom) speakers

Therefore, we find that the morphological condition is one factor associated with realisations of the vowels in VV sequences being different from their corresponding singleton realisations. Also, there was an interaction between morphological condition and vowel identity. Now we need to investigate how formant frequencies of each vowel are different depending on the morphological conditions. That is, we look at how the formant frequencies of the vowels in VV sequences are different from the singletons and how the formant frequencies of vowels of those morphological conditions are different from each other.

The additional one-way ANOVAs on F1 separately for each vowel, with the morphological condition as independent variables (four levels; singleton, hVV, WI and AWB) were conducted to test whether the F1 of each vowel is significantly different across the morphological conditions. Scheffe post-hoc tests were also conducted to see how different it is between the conditions.

Table 38 below presents the summary of the one-way ANOVAs to see whether the qualities of vowels in VV sequences are significantly different from those of singleton productions. The results revealed that there was a significant difference in F1 of /a/, /i/, /u/ and /o/ for male speakers and F1 of all vowels for female speakers. Apart from /e/ of male speakers, V1 values are significantly different across the morphological conditions.

Table 38 ANOVA table for F1 differences between singletons and VVs

	Male					Female				
	df	error	F	p	ω^2	df	error	F	p	ω^2
a	3	1033	33.98	<.001	0.085	3	1754	69.33	<.001	0.101
i	3	911	29.65	<.001	0.083	3	1550	22.56	<.001	0.039
u	3	913	30.95	<.001	0.087	3	1542	30.43	<.001	0.053
e	3	918	1.49	0.216	0.002	3	1532	20.81	<.001	0.037
o	3	898	3.94	0.008	0.010	3	1547	4.49	0.004	0.007

The results of the Scheffe-post hoc tests show two issues, the significant difference of F1 values of VV sequences from singletons and the significant difference between the morphological conditions. Table 39 below shows the summary of the Scheffe-post hoc tests to investigate the F1 differences between singletons and VV sequences of different morphological conditions. For the post-hoc tests, the effect size is presented in Cohen's d. The blank in the table means there is no significant difference in the post-hoc tests.

Table 39 Post-hoc comparison for F1 between singletons and VVs

		Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect
M	a	SGL>hVV*	0.49	S	SGL>WI**	0.97	L	SGL>AWB**	1.22	L
	i	SGL<hVV**	-1.10	L	SGL<WI**	-1.11	L	SGL<AWB**	-1.43	L
	u				SGL<WI*	-0.51	M	SGL<AWB**	-1.13	L
	e									
	o									
F	a				SGL>WI**	0.63	M	SGL>AWB**	1.17	L
	i	SGL<hVV**	-0.41	S	SGL<WI*	-0.72	M	SGL<AWB*	-0.74	M
	u							SGL<AWB**	-0.88	L
	e									
	o									

*: 0.001 < p < 0.05, **: p < 0.001
S: Small, M: Medium, L: Large

Table 39 reveals that the vowels in peripheral positions of the vowel triangle in Japanese such as /a/, /i/ and /u/ show significant differences but there is no difference in /e/ and /o/.In general, F1 of close vowels (/i/ and /u/) in VV sequences becomes higher and F1 of open vowel (/a/) lower than the vowels in singletons, showing that overall the vowel quality becomes more centralised. Although the tendency for F1 centralisation is recognised in all morphological conditions, it becomes more obvious in the comparison between singletons and VVs across word boundary.

Then we look at the comparisons between the morphological conditions. Table 400 below presents the post-hoc comparisons for the different morphological conditions.

Table 40 Post-hoc comparison for F1 among morphological conditions

		Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect
M	a	hVV>WI**	0.47	S	hVV>AWB**	0.69	M			
	i				hVV>AWB*	-0.46	S			
	u	hVV<WI**	-0.43	S	hVV<AWB**	-1.10	L	WI<AWB**	-0.50	M
	e									
	o							WI>AWB**	0.30	S
F	a	hVV>WI**	0.50	M	hVV>AWB**	0.99	L	WI>AWB**	0.57	M
	i	hVV<WI**	-0.36	S	hVV<AWB*	-0.30	S			
	u	hVV<WI*	-0.21	S	hVV<AWB**	-0.71	M	WI<AWB**	-0.58	M
	e	hVV>WI*	0.24	S	hVV>AWB**	0.68	M	WI>AWB**	0.41	S
	o									

*: 0.001 < p < 0.05, **: p < 0.001
S: Small, M: Medium, L: Large

Comparing the /hVV/ condition and other morphological conditions, male speakers showed a significant difference in /a/ and /u/ and female speakers in /a/, /i/, /u/, /e/ for both conditions. /i/ for male speakers shows a significant difference between the /hVV/ condition and AWB but not between the /hVV/ and WI. /e/ and /o/ for male speakers and /o/ for male speakers showed no difference between the /hVV/ and others. Comparing WI and AWB, not all the vowels showed a significant difference and they are different between male and female speakers. For male speakers, /u/ and /o/ showed a significant difference. F1 in /u/ of AWB was significantly higher than WI and F1 of /o/ in AWB was significantly lower than WI. For female speakers, /a/, /u/ and /e/ showed a significant difference between WI and AWB. F1 of /u/ in AWB was significantly higher than WI and F1 of /a/ and /e/ in AWB were significantly lower than WI. These results show that the difference between the /hVV/ and other conditions was variable across the vowel identities and that the difference between WI and AWB was not consistent across male and female speakers. Kawakami (1977) and Saito (1997) argued that VV sequences without a ‘semantic boundary’ between the vowels became more diphthong-like than VV sequences across a ‘semantic boundary’. However, the present result showed that VV sequences in the WI and AWB conditions were more reduced in F1 than VV sequences of nonsense words.

As with F1, additional one-way ANOVAs on F2 separately for each vowel, with the morphological condition as independent variables (four levels; singleton, hVV, WI and AWB) were conducted to test whether the F2 of each vowel is significantly different across the morphological conditions. Scheffe post-hoc tests were also conducted to see how different individual vowels were across the conditions. Table 41 below presents the summary of the one-way ANOVAs to see whether the qualities of vowels in VV sequences are significantly different from those of singleton productions. The results revealed that there was a significant difference in F2 of /a/, /u/, /e/ and /o/ for male speakers and F1 of all vowels for female speakers. Apart from /i/ of male speakers, V2 values are significantly different across the morphological conditions.

Table 41 ANOVA table for F2 differences between singletons and VVs

	Male					**Female**				
	df	error	F	p	ω^2	df	error	F	p	ω^2
a	3	1033	52.21	<.001	0.124	3	1754	69.27	<.001	0.101
i	3	911	1.19	0.313	0.001	3	1550	26.20	<.001	0.046
u	3	913	21.29	<.001	0.061	3	1542	51.47	<.001	0.087
e	3	918	7.51	<.001	0.021	3	1532	42.97	<.001	0.074
o	3	898	21.25	<.001	0.062	3	1547	54.11	<.001	0.090

Table 42 below presents the results of the Scheffe post-hoc tests, showing the differences for F2 values between singletons and VV sequences of three types of morphological conditions.

Table 42 Post-hoc comparison for F2 between singletons and VVs

		Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect
M	a				SGL<WI**	-0.94	L	SGL<AWB**	-1.49	L
	i									
	u				SGL<WI**	-0.61	M	SGL<AWB*	-0.69	M
	e				SGL>WI*	0.41	S	SGL>AWB*	0.45	S
	o				SGL<WI*	-0.55	S	SGL<AWB**	-0.84	L
F	a				SGL<WI**	-0.49	S	SGL<AWB**	-0.99	L
	i	SGL>hVV**	0.71	M	SGL>WI**	1.01	L	SGL>AWB**	1.15	L
	u				SGL<WI**	-0.86	L	SGL<AWB**	-1.37	L
	e	SGL>hVV**	0.60	M	SGL>WI**	0.84	L	SGL>AWB**	1.23	L
	o	SGL<hVV**	-0.73	M	SGL<WI**	-0.96	L	SGL<AWB**	-1.34	L

*: $0.001 < p < 0.05$, **: $p < 0.001$
S: Small, M: Medium, L: Large

It is apparent that male and female speakers showed a different tendency. There was no significant difference between singleton and VVs in /hVV/ nonsense word for male speakers but a significant difference between singleton and /hVV/ was found in /i/, /e/ and /o/ for female speakers. For male speakers, the finding that F2 of vowels was not significantly different across singleton and nonsense words suggests that the vowels of nonsense words were not reduced from those singletons in terms of F2. Comparing singletons and real word conditions, all the vowels apart from /i/ for male speakers showed a significant difference. F2 of front vowels were lower in VVs in sequences than singletons and F2 of back vowels were higher in sequences than singletons, showing that the vowels in VV sequences were more centralised compared to singletons. The centralisation of back vowels is associated with the tongue advancement. The tongue advancement in VV sequences, which is associated with higher F2, also occurred in /a/, the central vowel. F2 of the central vowel had higher F2 in VV sequences than in singletons, resulting in tongue advancement compared to singletons. Female speakers showed a wider difference across the vowel identity and the morphological conditions. But the tendency of centralisation is identical between male and female speakers and F2 of real words was more reduced than singletons.

Table 43 below shows the pairwise post-hoc comparison among three morphological conditions. Apart from /i/, all the vowels showed that F2 for VV sequences within words and across word boundary were centralised by comparison to VVs in /hVV/ condition. The difference between word-internal and across-boundary sequences were found in /a/ and /o/ for both male and female and in /e/ for female speakers.

Table 43 Post-hoc comparison for F2 among morphological conditions

		Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect
M	a	hVV<WI**	-0.61	M	hVV<AWB**	-1.03	L	WI<AWB**	-0.42	S
	i									
	u	hVV<WI**	-0.54	M	hVV<AWB*	-0.49	S			
	e	hVV>WI*	0.29	S	hVV>AWB*	0.35	S			
	o	hVV<WI**	-0.38	S	hVV<AWB**	-0.79	M	WI<AWB*	-0.40	S
F	a	hVV<WI**	-0.57	M	hVV<AWB**	-1.05	L	WI<AWB**	-0.46	S
	i	hVV>WI*	0.20	S						
	u	hVV<WI**	-0.52	M	hVV<AWB**	-0.79	M			
	e	hVV>WI**	0.26	S	hVV>AWB**	0.73	M	WI>AWB**	0.48	S
	o	hVV<WI*	-0.21	S	hVV<AWB**	-0.76	M	WI<AWB**	-0.62	M

*: $0.001 < p < 0.05$, **: $p < 0.001$
S: Small, M: Medium, L: Large

These series of ANOVAs suggest that the morphological condition applying to a VV sequence appears to be a significant factor in respect of vowel qualities becoming more centralised, which is thought to be one of the features which is considered to render a VV sequence more diphthong-like. However, the present results differ from those reported by Kawakami (1977) and Saito (1997). They argued that a vowel sequences contained within a single morpheme can be identified as diphthong-like but that a vowel sequence cannot be identified as a diphthong if there is a ‘semantic boundary’ between vowels in a sequence. In the current study, VV sequences across a word boundary showed greatest centralisation in F1-F2 vowel space for both male and female speakers. If the semantic boundary which Kawakami (1977) and Saito (1997) referred to is assumed to be present as a consequence of the word boundary, WI should have shown a greater centralisation than AWB. However, the current result showed the opposite, although in both cases the VV targets were centralised compared to the corresponding singletons.

There is still a need to investigate what ‘semantic boundary’ means in their arguments but we find that the formant frequencies of the vowels in VV sequences are reduced from singletons and the formant frequencies of VVs in real words are reduced from VVs in nonsense words. In the next section, we investigate whether the phonetic properties of vowels are significantly different depending on the position of the vowels in VV sequences, that is, whether there is a difference between the first (onset) position or the second (offset) position of VV sequences.

5.2.3 ***Effects for the position in VV sequence***

In a VV sequence, a vowel can appear in either the first target vowel position (V1) or the second target vowel position (V2). This means that the vowel will be either the starting point or the ending point of a VV sequence. In 5.2.2, it is shown that the vowels in VV sequences are reduced from the corresponding singletons. However, it is not still clear whether there is a difference in reduction depending on the position of the vowels in VV sequences, that is, whether the qualities of vowels in V1 and V2 positions in VV sequences are equally reduced from the corresponding singletons. In the case of diphthongs, the quality of the start and the end of a diphthong has been said to be different from the monophthong vowel being transcribed by the identical phonetic symbol (Ladefoged, 2006). It means that the glide of a diphthong does not reach the position in vowel space which the identical monophthong symbol represents. Roach (1983), for example, notes that the closing diphthongs do not reach a position that could be called close. Therefore, it is assumed that the qualities of the offset of diphthongs are different from those of corresponding singletons. However, it is not clear whether the qualities of the onset of diphthongs are different from corresponding singletons. Therefore, the degree of difference in the vowel quality of V1 position in VV sequence, which is thought to be equivalent to the ‘start of a diphthong’ and the quality of V2 position, which can be thought of the ‘end’ of a diphthong, needs to be investigated.

For the purpose of presenting the overall distance of the vowels in VV sequences from singletons, the Euclidian distance was calculated as described in 5.1.2.

Table 44 below presents the mean Euclidian distance between singletons and VV sequences in V1 and V2 positions for all the vowels. Data was pooled across morphological conditions. In total, the distances in V2 position was larger than in V1 position for both male and female speakers. However, male speakers showed consistency in that all the vowels in V2 position were more reduced than in V1 position whereas /a/, /i/ and /e/ for female speakers showed the opposite; the distance in V1 position was larger than in V2 position.

Table 44 Mean Euclidian distances (Hz) for five vowels in V1 and V2 positions

	Male		**Female**	
	V1	V2	V1	V2
a	209.52	230.47	242.18	239.10
i	197.08	214.10	311.16	289.09
u	240.57	284.85	279.09	316.79
e	129.72	154.49	227.32	223.44
o	131.55	151.33	197.10	247.06
Total	187.19	202.76	255.69	258.82

A two-way AVOVA was conducted to test whether there is a significant difference for the position in VV sequences and the vowel identity and whether there is an interaction between them. The position had a significant effect on the Euclidian distance for both male ($F(1, 4368) = 43.17$, $p < .001$, $\omega^2 = 0.015$) and female (F (2,

7523) = 83.22, $p < .001$, $\omega^2 = 0.020$) speakers. Also, the vowel identity had a significant effect for both male (F (4, 4368) = 144.61, $p < .001$, $\omega^2 = 0.112$) and female (F (4, 7523) = 77.90, $p < .001$, $\omega^2 = 0.039$) speakers. However, an interaction between the position in sequences and the vowel identity was found only in female speakers (F (2, 7528) = 12.14, $p < .001$, $\omega^2 = 0.006$) and not for male speakers (F (2, 4368) = 1.57, $p = 0.18$). Therefore, the position in VV and the vowel identity is not necessary associated with each other in the degree of centralisation.

5.2.4 *Effects of the accent type*

As is discussed in the literature review chapter, vowels in VV sequences carry either High tone (H) or Low tone (L) associated with the meaning of words and different tones are related to the difference in prominence. Vowels with High tone are more prominent than vowels with Low tone. If the first vowel in a VV sequence carries a High tone and the second vowel carries Low tone, the first vowel is more prominent than the second one. In the case of diphthongs, the prominence of the onset and offset are said to be different. For example, Ladefoged (2006) noted that the first part of the diphthong is usually more prominent than the last, that is, the onset of the diphthong is more prominent than the offset. Japanese VV sequences are also expected to have a prominence shift depending on the accent types. For example, the prominence shift of a VV sequence with High-Low accent is assumed to be similar to the one of a falling diphthong because the first vowel is more prominent than the second vowel despite the fact that each vowel occupies independent syllables. This section reports how the prominence of Japanese vowel is associated with the vowel quality in VV sequences.

Table 45 below shows the mean Euclidian distance between singletons and VV sequences produced with High or Low tones for all the vowels. The vowels with Low had longer distance for male speakers but those with High tone had longer distance for female speakers. The mean distances for each vowel were inconsistent between male and female speakers.

Table 45 *Mean Euclidian distances (Hz) for five vowels with High or Low tones*

	Male				**Female**			
	High		Low		High		Low	
	V1	V2	V1	V2	V1	V2	V1	V2
a	203.60	241.66	217.38	215.44	244.93	238.77	250.30	223.65
i	196.41	197.95	214.22	213.94	316.41	304.07	264.54	320.31
u	227.82	257.36	289.35	278.82	267.29	294.94	336.20	291.40
e	132.69	125.91	142.63	169.84	245.01	204.14	221.99	225.36
o	156.03	100.03	126.05	185.53	234.93	145.68	238.32	258.73

Firstly, a univariate analysis was conducted to test whether there is a significant interaction among the tone (High or Low), the position in a sequence (V1 or V2) and the vowel identity (a, i, u, e and o). There was a significant interaction for both male speakers (F (4, 4358) = 13.848, $p < 0.001$, $\omega^2 = 0.003$) and female speakers (F (4, 7518) = 19.952, $p < 0.001$, $\omega^2 = 0.002$), although the effect size was small.

Secondly, a two-way AVOVA was conducted to test whether there is a significant difference for the tones and the vowel identity and whether there is an interaction between them. The tone did not have a significant effect on the Euclidian distance for male speakers (*F* (1, 4368) = 2.06, p = 0.15) but there was a significant effect for female speakers (*F* (1, 7528) = 5.17, p = 0.02, $\omega^2 = 0.001$). In terms of the vowel identity, there was a significant effect for both male (*F* (1, 4368) = 132.70, $p < 0.001$, $\omega^2 = 0.104$) and female (*F* (1, 7528) = 77.83, $p < 0.001$, $\omega^2 = 0.039$) speakers. Also, the interaction between the tones and the vowel identity was found in female speakers (*F* (2, 7528) = 3.53, p = 0.01, $\omega^2 = 0.001$) and not for male speakers (*F* (2, 4368) = 0.88, p = 0.47). Therefore, the degree of centralisation for male speakers is only affected by the vowel identity but the centralisation for female speakers was affected by both the tone of a vowel and the vowel identity. Therefore, it is not necessary that the accent contour and the vowel identity are associated with each other in the degree of centralisation.

5.2.5 ***The interaction between the V1/V2 position and the accent contour, the sonority scale and the articulatory movements***

This section compares the degrees of reduction of V1 and V2 in VV sequences from the corresponding singletons taking the characteristics of the VV sequences (the accent contour, the sonority scale, the open/close dimension and the front/back dimension) into consideration. Table 46 below presents the mean degrees of reduction of V1 and V2 of vowel sequences with different accent types. In total, VV sequences produced with HH and LL accent had longer distances (i.e. were more reduced) than VVs with HL and LH accent types for both male and female speakers.

Table 46 Mean Euclidian distances for V1 and V2 of vowel sequences with different accent types

	Male								**Female**							
	HH		HL		LH		LL		HH		HL		LH		LL	
	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
a	241.97	287.53	164.95	199.53	194.91	198.40	261.37	248.70	277.17	288.45	212.27	208.03	232.28	211.18	250.05	255.71
i	199.85	217.31	192.93	201.61	192.75	211.12	209.15	237.93	344.98	277.50	285.48	263.97	325.43	251.29	260.46	421.74
u	237.61	280.31	217.89	287.15	231.57	298.47	306.08	262.63	299.31	327.93	235.54	293.29	287.40	344.34	309.96	287.56
e	157.09	139.45	109.44	154.78	119.65	145.78	138.31	199.55	312.64	231.26	180.77	199.08	207.03	212.88	198.42	277.93
o	223.75	134.45	86.98	180.12	106.06	117.66	88.18	195.43	324.13	238.60	147.19	238.16	162.37	238.04	111.73	295.86

Firstly, a univariate analysis was conducted to test whether there is a significant interaction among the accent type (HH, HL, LH and LL), the position in a sequence (V1 or V2) and the vowel identity (a, i, u, e and o). There was a significant interaction for both male speakers (*F* (12, 4338) = 8.117, $p < 0.001$, $\omega^2 = 0.005$) and female speakers (*F* (12, 7498) = 12.803, $p < 0.001$, $\omega^2 = 0.005$), although the effect size was small.

Secondly, a two-way ANOVA was conducted to test whether there is a significant difference for the accent contours and the positions of the vowels in VV sequences and whether there is an interaction between them. The accent pattern of VV sequences had a significant effect on the Euclidian distance for both male (*F* (3, 4370) = 21.89, p < .001, ω^2 = 0.014) and female (*F* (3, 7530) = 55.45, p < .001, ω^2 = 0.021) speakers. Scheffe post-hoc tests revealed that the degrees of reduction in vowel sequences produced with HH and LL accent types were significantly greater than others for male speakers. All the pairwise comparisons for female speakers were significantly different from each other. Table 47 shows all the pairwise comparisons. Although the Post hoc comparison showed a significant difference, the effect size for some pairwise comparisons did not show an effect.

Table 47 Post-hoc comparison for Euclidian distances among accent types of VVs

	Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
M	HH>HL**	0.23	S	HH>LH**	0.22	S			
				HL<LL**	-0.28	S	LH<LL**	-0.27	S
F	HH>HL**	0.38	S	HH>LH**	0.25	S	HH>LL*	0.11	W
	HL<LH*	-0.13	W	HL<LL**	-0.26	S	LH<LL*	-0.13	W

*: 0.001< p < 0.05, **: p < 0.001
W: Weak, S: Small, M: Medium, L: Large

There was a significant interaction for the degree of reduction between the accent contour and the positions in VVs for both male (*F* (3, 4370) = 6.50, p < .001, ω^2 = 0.004) and female (*F* (3, 7530) = 29.45, p < .001, ω^2 = 0.011) speakers. The vowels of VVs with HH and LL accent were more reduced than the vowels of VVs with HL and LH accent, although the effect size is small.

Table 48 presents the mean degrees of reduction of vowels in V1 and V2 positions of vowel sequences of three types of sonority scale. The degree of reduction in Level sequences was larger than others and the rising sequences were less reduced for both male and female speakers.

Table 48 Mean Euclidian distances for V1 and V2 of vowel sequences with different sonority scale

		Fall	**Rise**	**Level**
Male	V1	180.07	192.61	187.88
	V2	214.44	186.39	214.48
	Total	197.26	189.50	201.18
Female	V1	231.85	267.30	268.77
	V2	287.56	225.81	280.81
	Total	259.71	246.56	274.79

A two-way AVOVA was conducted to test whether there is a significant difference for the sonority scale and the positions in VV sequences and whether there is an interaction between them. There was a significant difference for the degree of reduction for female speakers (*F* (2, 7532) = 15.85, $p < .001$, $\omega^2 = 0.004$) but no difference was found for male speakers (*F* (2, 4372) = 2.87, $p = 0.06$). Scheffe post-hoc tests revealed that all the pairwise comparisons were significantly different for female speakers but there was no difference for male speakers. Table 49 below shows the pairwise comparison for female speakers.

Table 49 Post-hoc comparison for Euclidian distances for female speakers of sonority scale

Comparison	Cohen's d	effect
Fall>Rise*	0.08	Weak
Fall<Level*	-0.09	Weak
Rise<Level**	-0.17	Weak

*: $0.001 < p < 0.05$, **: $p < 0.001$

Although the effect size was small, there was a significant interaction between the sonority scale of VVs and the position in VVs in the degrees of reduction for both male ($F(3, 4372) = 10.55$, $p < .001$, $\omega^2 = 0.004$) and female ($F(3, 7532) = 59.92$, $p < .001$, $\omega^2 = 0.015$) speakers. The degrees of reduction for vowels in V1 position were consistently larger than in V2 position in Falling and Level sequences and those in V2 position were larger in rising sequences.

Table 50 below presents the mean distance of V1 and V2 of vowel sequences of different degrees of the closing/opening dimension. In total, VV sequences with larger degree of opening or closing movement were more reduced than those with smaller degrees. VV sequences with no difference in the open/close dimension had longer distance compared to VV sequences with small degrees.

Table 50 Mean Euclidian distances for V1 and V2 of vowel sequences of the close/open dimension

		Close 2	**Close 1**	**No difference**	**Open 1**	**Open 2**
Male	V1	207.82	166.31	187.88	186.34	210.35
	V2	222.49	210.45	214.48	170.11	232.50
	Total	215.16	188.38	201.18	178.22	221.42
Female	V1	251.22	221.96	268.77	251.24	314.98
	V2	292.27	285.16	280.81	221.04	240.00
	Total	271.74	253.56	274.79	236.14	277.49

A two-way AVOVA was conducted to test whether there is a significant difference for the close/open dimension and the positions in VV sequences and whether there is an interaction between them. There was a significant difference for the close/open dimension for both male (F (4, 4368) = 13.99, $p < .001$, $\omega^2 = 0.012$) and female (F (4, 7528) = 19.02, $p < .001$, $\omega^2 = 0.009$) speakers. Scheffe post-hoc tests revealed that the degree of opening movement had a significant effect on distance, although the effect size was small. Table 51 below shows all the pairwise comparisons.

Table 51 Post-hoc comparison for Euclidian distances of the close/open dimension

	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
M	C2>C1*	0.20	S				C2>O1**	0.30	S
	C1<O2*	-0.25	S						
	O1<O2**	-0.36	S						
F							C2>O1**	0.21	S
				C1<ND*	-0.12	W	C1>O1*	0.11	W
	C1<O2*	-0.14	W	ND>O1*	0.23	S			
	O1<O2**	-0.25	S						

*: 0.001< p < 0.05, **: p < 0.001
W: Weak, S: Small, M: Medium, L: Large

There was a significant interaction between the distance and the close/open dimension for both male (*F* (4, 4368) = 8.17, p < .001, ω^2 = 0.006) and female (*F* (3, 7528) = 33.54, p < .001, ω^2 = 0.017) speakers. The distances of vowels in V1 position were consistently larger than in V2 position in VV sequences of Open 1 but in all the other close/open dimension had longer V2 distances than V1 distances.

Table 52 below presents the mean distance of V1 and V2 of vowel sequences of different degrees of the fronting/backing dimension. In total, VV sequences consisted with vowels of the same degree of the antero-posterior position had shorter distance than others.

Table 52 Mean Euclidian distances for V1 and V2 of vowel sequences of the front/back dimension

		Front2	**Front1**	**Same**	**Back1**	**Back2**
Male	V1	199.20	215.54	194.06	166.72	165.24
	V2	178.81	202.50	126.06	217.94	241.28
	Total	189.00	209.02	160.06	192.33	203.26
Female	V1	249.71	255.32	236.06	249.42	277.95
	V2	247.41	256.02	180.95	260.00	310.48
	Total	248.56	255.67	208.50	254.71	294.22

A two-way AVOVA was conducted to test whether there is a significant difference for the front/back dimension and the positions in VV sequences and whether there is an interaction between them. There was a significant difference for the front/back dimension for both male (F (4, 4368) = 11.50, $p < .001$, $\omega^2 = 0.009$) and female (F (4, 7528) = 38.64, $p < .001$, $\omega^2 = 0.019$) speakers. Scheffe post-hoc tests revealed that the fronting/backing movement of vowel sequences significantly affected the distance. Table 53 below shows all the pairwise comparisons.

Table 53 Post-hoc comparison for Euclidian distances of the front/back dimension

	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
M	F2<F1*	-0.15	W	F2>ND*	0.22	S			
				F1>ND**	0.40	S			
				ND<O1*	-0.28	S	ND<O2**	-0.30	S
F				F2>ND**	0.26	S			
	F2<B2**	-0.26	S	F1>ND**	0.29	S			
	F1<B2**	-0.21	S	ND<O1**	-0.30	S	ND<O2**	-0.51	M
	O1<O2**	-0.23	S						

*: 0.001< p < 0.05, **: p < 0.001

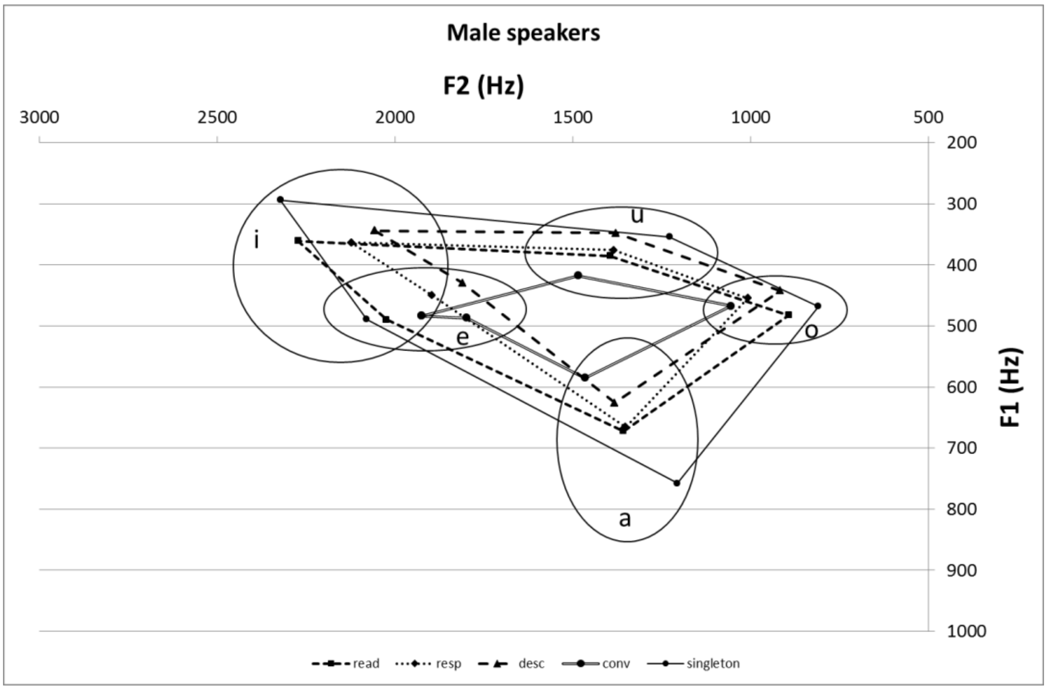
W: Weak, S: Small, M: Medium, L: Large

There was a significant interaction between the distance and the front/back dimension for both male (*F* (4, 4368) = 8.17, $p < .001$, $\omega^2 = 0.006$) and female (*F* (3, 7528) = 33.54, $p < .001$, $\omega^2 = 0.017$) speakers. Fronting sequences (Front 1 and Front 2) had larger V1 distances than V2 and Backing sequences (Back 1 and Back 2) had larger V2 distance than V1. Therefore, V1 was more reduced than V2 in fronting sequences, that is, the articulation started with a reduced position and end with a less reduced position while producing fronting sequences. Also, V2 was more reduced than V1 for backing sequences; that is, the articulation ended with a reduced position.

5.3. Speech Style Condition

5.3.1 *General remarks*

Having observed the formant properties of VV sequences of different morphological conditions, we now look at the properties of VV sequences of different speech styles. As is mentioned in the methodology chapter, VV sequences with various morphological conditions were obtained by asking participants to read out the material. Regarding this speech style as 'read' speech, the stimuli for the speech style condition are all real words containing VV sequences and they were produced in different speech styles. As is discussed in previous studies such as Kawakami (1977), differences in speech style are thought to be associated with vowel sequences having diphthong-like features. This section reports how F1 and F2 of vowels are different for vowels in VV sequences produced across four speech styles as well as from the corresponding singletons. The overall comparison of all vowels across the speech style condition is presented in Figure 30 below. The vertical axis represents the F1 value and the horizontal axis the F2 value. Speech styles are presented by the line style and the shape of the marker. The vowels of the same identity are circled.



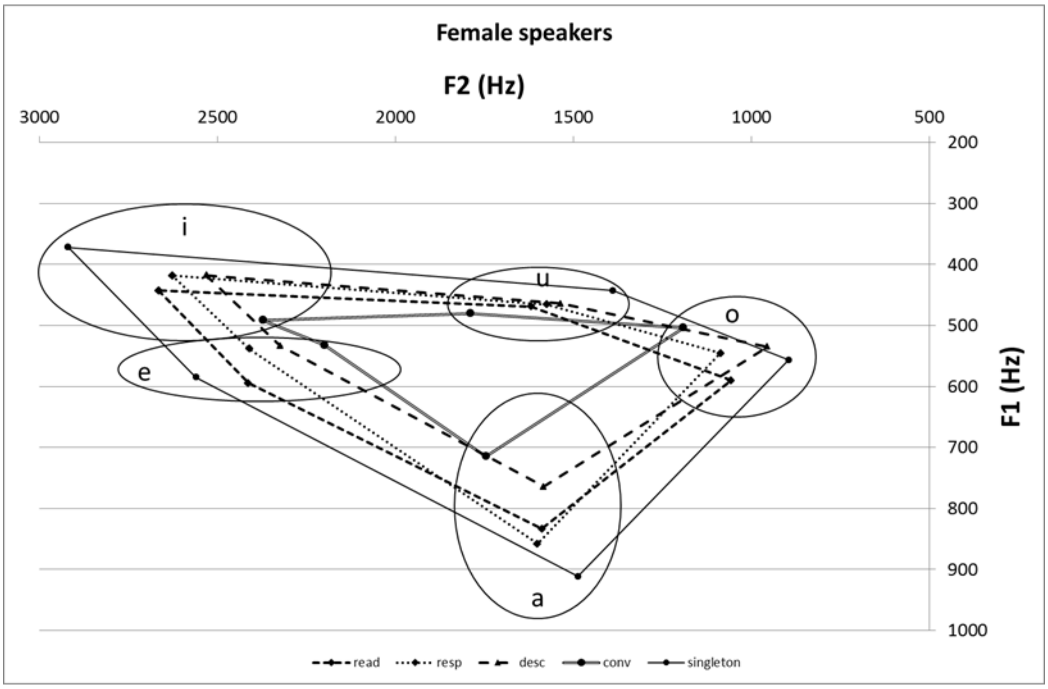


Figure 30 Comparison of vowel quality of five vowels of read, responses, description and conversation styles of VV sequences compared with singletons by male (top) and female (bottom) speakers

It is noticeable that most of the vowels in VV sequences are ‘centralised ‘compared to singletons, that is, the overall shape of the vowel space of VV sequences are reduced from singletons.

Observing the male data, F1 of /i/ is higher in sequences than in singletons, and F1 values for read, response and description styles are close to one another. F2 of /i/ is lower in sequences than singleton and the F2 value for /i/ in sequences varies depending on the speech style. F1 and F2 of /i/ for natural conversation are substantially reduced, having higher F1 and lower F2, the realisation of /e/ overlaps with that of /e/. F1 of /e/ in sequences is close to the corresponding singleton but somewhat lower than the singleton. F1 of /a/ is lower and F2 is higher in sequences than in the corresponding singleton. F1 and F2 of read and response styles are very close to each other and those of description and conversation style are much more reduced from those styles. F1 of /o/ in sequences is close to singleton and F2 in sequences is higher than singleton. F2 for /o/ is found to be varied depending on the speech style. F1 of /u/ for read speech is close to singleton but F1 for other speech styles are higher than singleton. F1 for response and description styles are close to each other and F1 for natural conversation is higher than any other styles.

For female speakers, F1 of /i/ is slightly higher in sequences than in singleton and F1 for read, response and description is close to each other. F2 of /i/ in sequences is lower than singleton and different with each other. F1 and F2 of /i/ for natural conversation are substantially reduced from singleton and other styles, having higher F1 and lower F2 values, although, unlike male speakers, it is not in the /e/ region. F1 of /e/ for read speech is close to the singleton value and is slightly lower in other speech styles. F2 of /e/ for sequences is lower than singleton and varies across speech styles. F1 of /a/ is lower in sequences than singleton and varies across speech styles. F2 of /a/ in sequences is higher than singleton but F2 for read, response and description styles are similar to each other. F2 of natural conversation is higher than singleton and other speech styles. F1 of /o/ is close to each other but F2 of /o/ is higher in sequences than singleton and varies across the speech styles. For /u/, F1 in VV sequences are slightly higher than the singleton and F2 in sequences are higher than singleton. Moreover, F2 for the natural conversation is higher than other speech styles and F2 for read, response and description styles are similar to each other.

The observation shows that the vowels in VV sequences produced in spontaneous and natural speeches are more reduced than singletons and read speech. Also, the vowels in VV sequences in the natural conversation are substantially reduced from singletons and other speech styles. For natural conversation, the vowel positions of /i/ and /e/ are close to each other. In 3.2, we investigate whether these differences are statistically significant.

5.3.2 *Effects of Speech style*

ANOVAs were conducted to investigate whether the vowels produced with different speech styles are reduced from singletons and whether there is a significant interaction between speech styles and the vowel identity. A two-way ANOVA was performed on the F1 and F2 values (in Hz) for each vowel, with the speech style (five levels; singleton, read, responses, description and conversation) and vowel identity (five levels; /a/, /i/, /u/, /e/ and /o/) as independent variables. As is the same with the analysis of morphological condition in 2.2, the purpose of the analysis is to find out whether F1 or F2 (or both) are more centralised in the vowel space depending on the speech style difference. In addition, the interaction between the speech styles and the vowel identity is investigated to see whether the all the vowels are reduced to the same degree depending on the speech styles or not.

With respect to F1 values, the two-way ANOVA showed a significant effect for speech styles for both male [F (4, 4860) = 18.89, $p < .001$, $\omega^2 = 0.005$] and female [F (4, 7870) = 38.89, $p < .001$, $\omega^2 = 0.008$] speakers, although the effect size was small. The vowel identity showed the expected effect on F1 for both male [F (4, 4860) = 1031.93, $p <.001$, $\omega^2 = 0.291$] and female [F (4, 7870) = 1303.63, $p < .001$, $\omega^2 = 0.250$] speakers reflecting the fact that F1 of five vowels are different from each other. There was a significant interaction between the speech style and the vowel identity for both male [F (16, 4860) = 77.25, $p < .001$, $\omega^2 = 0.092$] and female [F (16, 7870) = 49.10, $p < .001$, ω^2

= 0.039] speakers, although the effect size was small. An interaction between the speech style and the vowel identity presumably suggests that the variability on F1 is depending on the vowel identity. Figure 31 below illustrates the 2-way interaction of the speech style by the vowel identity on F1.

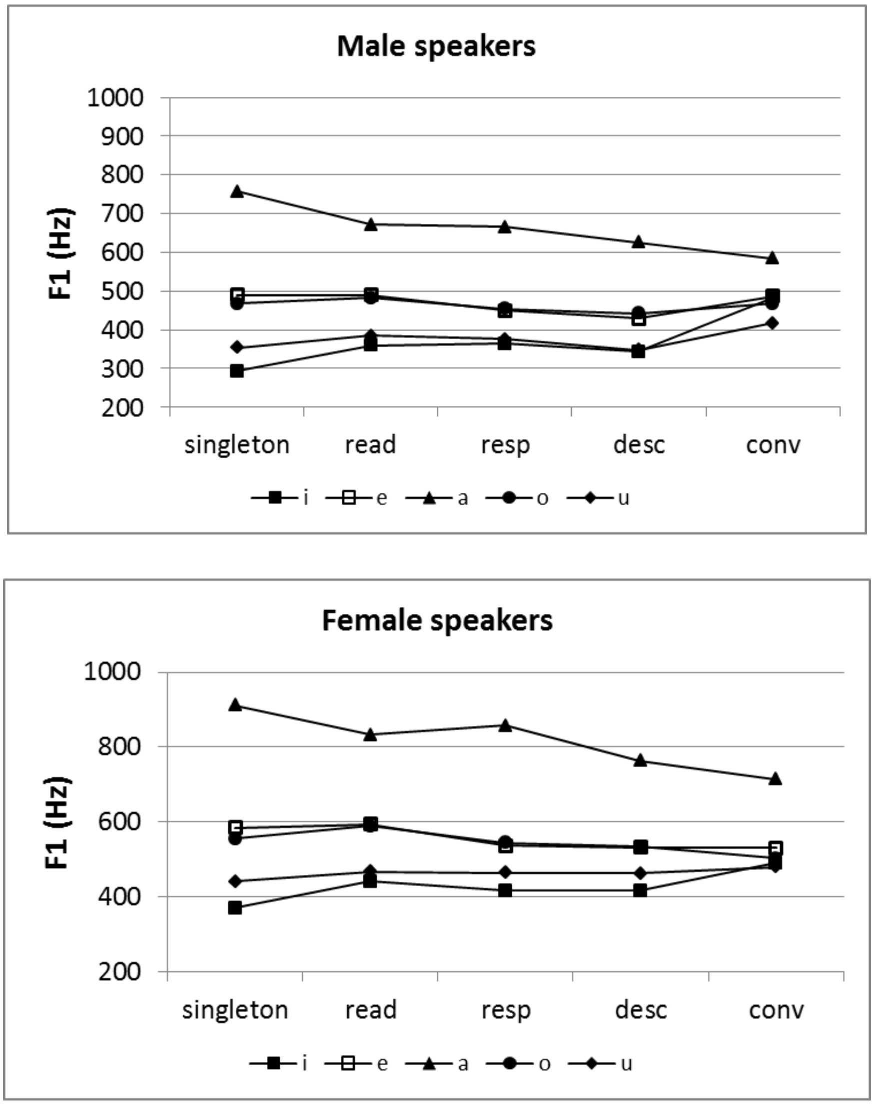


Figure 31 Mean F1 (Hz) of five vowels of read, responses, description and conversation styles compared with singletons by male (top) and female (bottom) speakers

Table 54 presents the mean F1 for each vowel in singleton and in VV sequences across the speech style conditions. F1 differences between the vowels in VV sequences across the speech style conditions and the corresponding singletons that is, the degree of difference from singletons are also presented below the F1 values.

Table 54 F1 values (in Hz) of the vowels in singleton and in VV sequences across the speech style conditions

		Male					**Female**				
		Singleton	Read	Resp	Desc	Conv	Singleton	Read	Resp	Desc	Conv
i	Frequency	293.9	360.5	363.5	343.7	483.7	371.9	442.6	417.9	417.7	491.2
	Difference		66.6	69.6	49.8	189.7		70.7	46	45.8	119.3
e	Frequency	489	489.5	449.4	429.6	486.6	585.1	594.5	537.9	532.9	532.1
	Difference		0.5	39.6	59.4	2.4		9.4	47.2	52.2	53
a	Frequency	757.4	671.8	665.4	626	585	911.5	833.3	858.2	764.2	714.6
	Difference		85.5	92	131.4	172.4		78.2	53.3	147.4	196.9
o	Frequency	467.8	482.1	454.7	442.1	467.1	556.4	590.1	545.6	534.9	503.4
	Difference		14.3	13.1	25.7	0.7		33.7	10.8	21.5	53
u	Frequency	354.1	385.1	375.5	347.6	417.8	442.7	468.8	465.6	463.7	479.9
	Difference		31	21.4	6.5	63.7		26.1	22.9	21	37.2

Regarding F1 values, the degrees of difference across the speech styles are different from the corresponding singletons. The degrees of difference are greater in the natural conversation than read speech except for /o/ of male speakers. However, the degree of difference in the response and description styles varies depending on the vowel identity. /i/ of the description style of male speakers showed a smaller degree than the read speech and /i/ of the response and the description styles of female speakers showed smaller degrees than the read speech. For /e/ of male speakers, the response and the description styles showed a greater degree than the read speech. For /a/ of male and female speakers, the response and the description styles were intermediate between the read and the conversation styles. For /o/ of male speakers, the degree in the response style was smaller than the read speech but the degree in the description was larger than the read speech. For /o/ of female speakers, the degrees of difference in the response and the description styles were smaller than the read speech. For /u/ of male and female speakers, the degrees in the response and the description styles were smaller than the read speech. Also, the overall degree of difference across the speech style is larger in /i/ and /a/ than others. In sum, the results show that F1 in the vowels in VV sequences are different from the vowels in singletons to some extent, but the degree of difference in F1 values is different both in the speech style conditions and in the vowel identity.

Figure 32 below illustrates the 2-way interaction of the speech style by the vowel identity on F2. With respect to F2 values, the two-way ANOVA showed a significant effect for the speech styles on F2 for both male [F (4, 4860) = 13.87, $p < .001$, $\omega^2 = 0.002$] and female [F (4, 7870) = 9.79, $p < .001$, $\omega^2 = 0.001$] speakers, although the effect size was small. The vowel identity showed an effect on F2 for both male [F (4, 4860) = 2045.18, $p <.001$, $\omega^2 = 0.313$] and female [F (4, 7870) = 4475.51, $p < .001$, $\omega^2 = 0.336$] speakers but, as with F1, this is because F2 values of five vowels are different from each other. There was a significant interaction between the speech style and the vowel identity for both male [F (16, 4860) = 58.00, $p < .001$, $\omega^2 = 0.038$] and female [F (16, 7870) = 89.03, p < .001, ω² = 0.029] speakers, although the effect size was small. Aside from the fact that the effect size was fairly small, the variability of F2 across the speech styles is different across the vowel identities.

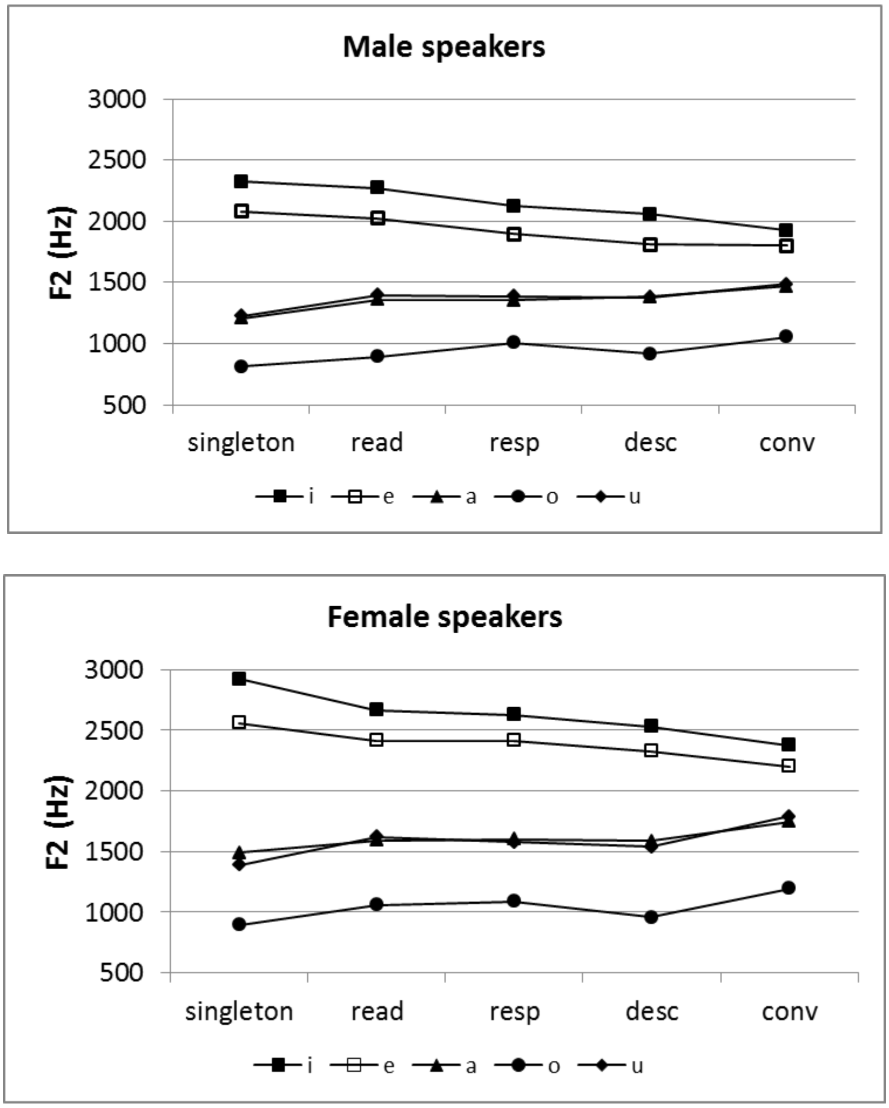


Figure 32 Mean F2 (Hz) of five vowels of read, responses, description and conversation styles compared with singletons by male (top) and female (bottom) speakers

Table 55 presents the F2 of each vowel in singleton and in VV sequences across the speech style conditions. F2 differences between the vowels in VV sequences across the speech style conditions and the corresponding singletons that is, the degree of difference from singletons are also presented below the F2 values.

Table 55 F2 values (in Hz) of the vowels in singleton and in VV sequences across the speech style conditions

		Male					**Female**				
		Singleton	Read	Resp	Desc	Conv	Singleton	Read	Resp	Desc	Conv
i	Frequency	2321.8	2272.1	2123.1	2056.4	1924.8	2918.6	2665.3	2626.8	2530.7	2371.9
	Difference		49.7	198.7	265.4	397		253.4	291.8	387.9	546.7
e	Frequency	2079.4	2023.8	1896.3	1810.1	1798.3	2559.3	2412.9	2410	2324.3	2199.3
	Difference		55.7	183.1	269.4	281.2		146.4	149.3	235.1	360.1
a	Frequency	1205.7	1357.1	1351.7	1382	1465.3	1486.8	1587.7	1600.3	1583.1	1745.1
	Difference		151.4	146	176.3	259.6		100.8	113.4	96.3	258.3
o	Frequency	809.5	892.5	1007.3	916.7	1054.1	893.9	1056.1	1085.7	954.5	1190.6
	Difference		83	197.8	107.3	244.6		162.2	191.8	60.5	296.7
u	Frequency	1227.3	1394.3	1384.9	1379	1484.2	1388.5	1617.4	1573.2	1536.4	1789.3
	Difference		167	157.6	151.7	256.9		228.8	184.7	147.9	400.8

Regarding F2 values, the degrees of difference of the vowels in the conversation style were consistently larger than the vowels of the read speech. /a/ and /o/ of the description style of female speakers were less reduced than the read speech and /u/ of the response and the description styles were also less reduced than the read speech. The degree of differences in response and the description styles of other vowels were between the read speech and the conversation.

Therefore, the speech style in which the VV sequences are produced can be regarded as one of the factors associated with the reduction in the quality of vowels in VV sequences, although the effect sizes for the ANOVAs were typically small. But there was a significant interaction between the speech styles and the vowel identity, thus, it is meaningful to investigate how the formant frequencies of each vowel are changed depending on the speech styles. To investigate further the effect of speech styles on formant properties, one-way ANOVAs on F1 and F2 separately for each vowel, with the speech style as independent variables (five levels; singleton, read, responses, description and conversation) were conducted to assess the statistical significance of factors influencing F1 and F2 measurements. Scheffe post-hoc tests were also conducted for the pairwise comparison.

Table 56 below presents the summary of the one-way ANOVAs for F1. It shows that there was a significant difference for F1 values for all vowel identities for male speakers and female speakers with the exception of /u/ for female speakers. In the case of the morphological conditions, significant differences were found in all the cases except for /e/ of male speakers. Thus, all the vowels are found to be significantly different across speech styles.

Table 56 ANOVA table for F1 differences between singletons and vowels in different speech styles

	Male					**Female**				
	df	error	F	p	ω^2	df	error	F	p	ω^2
a	4	1319	94.199	<.001	0.208	4	1960	106.78	<.001	0.170
i	4	1239	157.89	<.001	0.309	4	2052	42.10	<.001	0.073
u	4	775	10.787	<.001	0.047	4	1247	2.31	0.056	0.004
e	4	786	12.542	<.001	0.054	4	1278	22.20	<.001	0.061
o	4	741	5.26	<.001	0.022	4	1333	33.47	<.001	0.087

The results of the Scheffe-post hoc tests present two issues, firstly the significant difference of F1 values of VV sequences from singletons and the significant difference between the speech styles as with the morphological conditions reported in 5.2.2. Table 57 below shows the summary of the Scheffe-post hoc tests to present the differences of F1 values between singletons and the vowels in VV sequences produced in four different speech styles. For the post-hoc tests, the effect size is presented in Cohen’s d.

Table 57 Post-hoc comparison for F1 between singletons and vowels with different speech styles

	Male			**Female**		
	Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect
a	SGL>Rd**	0.97	L	SGL>Rd**	0.63	M
	SGL>Rs**	1.08	L			
	SGL>Ds**	1.36	L	SGL>Ds**	1.24	L
	SGL>Cv**	1.98	L	SGL>Cv**	1.38	L
i	SGL<Rd**	-1.11	L	SGL<Rd**	-0.72	M
	SGL<Rs**	-0.99	L			
	SGL<Cv**	-1.71	L	SGL<Cv**	-0.96	L
u	SGL<Rd**	-0.51	M			
	SGL<Cv**	-0.93	L			
e						
	SGL>Rs*	0.6	M	SGL>Rs*	0.53	M
	SGL>Ds*	1.03	L	SGL>Ds*	0.58	M
				SGL>Cv*	0.48	S
o						
				SGL>Cv*	0.48	S

*: 0.001< p < 0.05, **: p < 0.001
S: Small, M: Medium, L: Large

Read speech is presented here as one of the speech styles but it is the same as the word-internal condition, so we mainly look at other speech styles. Comparing the response, the description and the conversation styles, the degree of the difference on F1 from singletons varies across the speech styles and between male and female speakers. The centralisation in the response style is found in /a/, /i/ and /e/ for male speakers but only in /e/ for female speakers. In the case of the description style, /a/ and /e/ are found to have significantly lower F1 than singletons for both male and female speakers. For conversation style, /a/, /i/ and /u/ were centralised from singletons for male speakers but all the vowels except for /u/ for female speakers. Although we recognised that the vowel space for the vowels in sequences were largely located on the inside of the vowel space of singleton productions, the significant difference on F1 was not found in all vowels. /a/ for male speakers is centralised in all speech styles but the statistical significance was not found for vowels. Also, the effect size for centralisation is from nearly medium to large if there is a significant difference but it is not consistent within the speech style.

Then, we look at the inter-style differences. Table 58 below shows the pairwise post-hoc comparison among four speech style conditions. Some differences between read speech and natural conversation from other speech styles were very clear, but not consistent across the vowel identity and between male and female speakers. Among four speech styles, there was only one case in which the response style and the description style were significantly different, which is /a/ for female speakers. Although not all the pairwise comparisons were significantly different, the read speech was least reduced and the conversation style was most reduced.

Table 58 Post-hoc comparison for F1 among vowels with different speech styles

		Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect
M	a				Rd>Ds*	0.49	S	Rd>Cv**	0.97	L
					Rs>Cv**	0.90	L			
	i							Rd<Cv**	-1.34	L
					Rs<Cv**	-1.06	L	Ds<Cv**	-1.22	L
	u				Rd>Ds*	0.59	M	Rd<Cv*	-0.50	M
					Rs<Cv*	-0.50	M	Ds<Cv**	-0.82	L
	e	Rd>Rs**	0.53	M	Rd>Ds**	0.80	L			
					Rs<Cv*	-0.47	S	Ds<Cv*	-0.74	M
	o	Rd>Rs*	0.38	S						
F	a				Rd>Ds*	0.55	M	Rd>Cv**	0.88	L
		Rs>Ds**	0.69	M	Rs>Cv**	0.98	L			
	i							Rd<Cv**	-0.42	S
					Rs<Cv**	-0.58	M	Ds<Cv**	-0.59	M
	u									
	e	Rd>Rs**	0.56	M	Rd>Ds**	0.61	M	Rd>Cv**	0.59	M
	o	Rd>Rs*	0.42	S	Rd>Ds*	0.52	M	Rd>Cv**	0.80	L
					Rs>Cv*	0.37	S			

*: $0.001 < p < 0.05$, **: $p < 0.001$

S: Small, M: Medium, L: Large

Table 59 below presents the summary of the ANOVA. It shows that F2 values for all the vowels are significantly different between speech styles for both male and female speakers. It is revealed that F2 values for all the vowels are significantly different across the speech style.

Table 59 ANOVA table for F2 differences between singletons and vowels in different speech styles

	Male					**Female**				
	df	error	F	p	ω²	df	error	F	p	ω²
a	4	1319	40.937	<.001	0.105	4	1960	58.34	<.001	0.102
i	4	1239	126.81	<.001	0.269	4	2052	162.01	<.001	0.225
u	4	775	6.425	<.001	0.027	4	1247	25.16	<.001	0.070
e	4	786	59.02	<.001	0.215	4	1278	65.57	<.001	0.161
o	4	741	28.13	<.001	0.123	4	1333	46.50	<.001	0.116

Table 60 below summarises the pairwise comparisons for F2 values between singletons and VVs produced in the other speech styles. Vowels in response and conversation styles are significantly different from singletons across all the vowel identities. The significance is also supported by a larger effect size. Not all the vowels with the description style are reduced and the vowels are different between male and female speakers but the effect size is fairly large. Comparing with the F1 difference, F2 is found to be more apparently reduced from the singletons.

Table 60 Post-hoc comparison for F2 between singletons and vowels with different speech styles

	Male			**Female**		
	Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect
a	SGL<Rd**	-0.94	L	SGL<Rd**	-0.49	S
	SGL<Rs**	-0.81	L	SGL<Rs*	-0.53	M
	SGL<Ds**	-1.11	L			
	SGL<Cv**	-1.29	L	SGL<Cv**	-1.00	L
i				SGL>Rd**	1.01	L
	SGL>Rs**	0.86	L	SGL>Rs**	1.17	L
	SGL>Ds**	1.00	L	SGL>Ds**	1.37	L
	SGL>Cv**	1.4	L	SGL>Cv**	1.76	L
u	SGL<Rd*	-0.61	M	SGL<Rd**	-0.86	L
	SGL<Rs*	-0.66	M	SGL<Rs**	-0.71	M
				SGL<Ds*	-0.54	M
	SGL<Cv**	-1.14	L	SGL<Cv**	-1.49	L
e				SGL>Rd**	0.84	L
	SGL>Rs**	1.13	L	SGL>Rs**	0.86	L
	SGL>Ds**	1.54	L	SGL>Ds**	1.38	L
	SGL>Cv**	1.55	L	SGL>Cv**	1.82	L
o	SGL<Rd*	-0.55	M	SGL<Rd**	-0.96	L
	SGL<Rs**	-1.11	L	SGL<Rs**	-1.22	L
	SGL<Cv**	-1.29	L	SGL<Cv**	-1.37	L

*: $0.001 < p < 0.05$, **: $p < 0.001$
S: Small, M: Medium, L: Large

Table 61 below shows the pairwise post-hoc comparisons among speech styles. It reveals that F2 of VVs in free conversation was significantly reduced from VVs in other speech styles in most vowels. Vowels in free conversation were more centralised than vowels in read speech in all the vowels other than /u/ in the case of male speakers. For female speakers, the difference between the read speech and the free conversation was found in all the vowels. The difference between the response style and the description style was found in /e/ for male speakers and /e/ and /o/ for female speakers.

Although not all the pairwise comparisons showed significance, the vowels in free conversation were most reduced in all the speech styles and the response and the description styles were reduced from the read speech.

Table 61 Post-hoc comparison for F2 among vowels with different speech styles

	Male			**Female**		
	Comparison	Cohen's d	Effect	Comparison	Cohen's d	Effect
a						
	Rd<Cv**	-0.58	M	Rd<Cv**	-0.67	M
	Rs<Cv**	-0.54	M	Rs<Cv**	-0.54	M
				Ds<Cv**	-0.61	M
i	Rd>Rs**	0.69	M			
	Rd>Ds**	0.96	L	Rd>Ds**	0.53	M
	Rd>Cv**	1.37	L	Rd>Cv**	1.04	L
	Rs>Cv**	0.69	M	Rs>Cv**	0.84	L
	Ds>Cv*	0.44	S	Ds>Cv**	0.51	M
u						
				Rd<Cv**	-0.62	M
				Rs<Cv**	-0.71	M
				Ds<Cv**	-0.78	M
e	Rd>Rs**	0.88	L			
	Rd>Ds**	1.46	L	Rd>Ds*	0.49	S
	Rd>Cv**	1.51	L	Rd>Cv**	1.14	L
	Rs>Ds*	0.44	S	Rs>Ds*	0.42	S
	Rs>Cv*	0.49	S	Rs>Cv**	0.98	L
				Ds>Cv**	0.57	M
o	Rd<Rs**	-0.73	M			
				Rd>Ds*	0.58	M
	Rd<Cv**	-1.02	L	Rd<Cv**	-0.69	M
				Rs>Ds*	0.69	M
				Rs<Cv**	-0.44	S
	Ds<Cv*	-0.73	M	Ds<Cv**	-0.97	L

*: 0.001 < p < 0.05, **: p < 0.001

S: Small, M: Medium, L: Large

These statistical comparisons show that the speech style difference contributed to the reduction of the vowel qualities. Vowels in free conversation were reduced from singletons as well as other speech styles. Especially the VV sequences in free conversation are more greatly reduced from singletons. In Chapter 4, we looked at the timing properties of VV sequences and showed that VVs in conversation had relatively shorter duration than in other speech styles but that they still had steady states. The steady states in conversations were still long in some cases, that is, VV sequences in the conversation style may be more diphthong-like in one sense, but they still look like VV sequences. Here the reduction from the corresponding singletons was recognised, which seems to be diphthong-like. The question arises whether there is a case that the degrees of reduction are variable. It is also noted that the reduction in formant frequencies was more remarkable in F2 values than F1 values, judging from the pairwise comparisons from singletons and across speech styles. In morphological conditions, there were more significant pairwise comparisons in F2 than F1. Therefore, F2 values seem to be affected by the conditions of VV sequences.

5.3.3 *Effects of the position in a sequence*

Based on the findings of the ‘centralisation’ effect across the speech styles, the Euclidian distance of the vowels in VV sequences from singleton is considered. Firstly, the position of the vowels in VV sequences is considered. The purpose is to show whether the vowels in the first target position or the second target position are significantly different from singletons and whether there is a significant difference between the position in the sequences and the vowel identity. Pooled across speech style conditions, a two-way ANOVA was conducted using the position in VV sequence with two levels (singleton, V1 and V2) and vowel identity (/a/, /i/, /u/, /e/ and /o/) as independent variables. The dependent variables were the F1 and the F2 values (in Hz). Each dependent variable was analysed separately. In the ANOVA analysis, the effect of the each of the two independent variables and the interaction between those two factors were investigated. This allowed an evaluation of whether either F1 or F2 values (or both) are more ‘centralised’ in F1-F2 vowel space depending on the position in the VV sequence. Also, the interaction between the position and the vowel identity is investigated to whether centralisation is more prevalent in V1 or V2 position.

Table 62 below presents the distance between singletons and VV sequences in V1 and V2 positions for all the vowels. In general, the distances in V2 position were larger than in V1 position for both male and female speakers. However, the mean distance varies depending on the vowel identity for both male and female speakers. For /a/, V1 was more reduced than V2 but other vowels in V2 position were more reduced than in V1 position. This is consistent with the finding of the morphological conditions but the difference between the distances on V1 and V2 become larger in the speech style conditions.

Table 62 Mean Euclidian distances for five vowels in V1 and V2 positions

	Male		**Female**	
	V1	V2	V1	V2
a	300.91	252.35	327.53	253.52
i	233.12	405.51	373.18	483.91
u	267.75	313.57	308.93	334.16
e	153.77	203.06	208.48	279.43
o	156.27	185.30	196.32	284.07
Total	244.74	295.40	299.30	351.43

A two-way ANOVA was conducted to test whether there is a significant difference for the position in VV sequences and the vowel identity and whether there is an interaction between them. The position had a significant effect on the Euclidian distance for both male (F (1, 4560) = 84.52, $p < .001$, $\omega^2 = 0.015$) and female (F (2, 7478) = 75.02, $p < .001$, $\omega^2 = 0.008$) speakers. The vowel identity had an effect on the distance for both male (F (1, 4560) = 125.06, $p < .001$, $\omega^2 = 0.087$) and female (F (2, 7478) = 216.41, $p < .001$, $\omega^2 = 0.096$) speakers. Also, the interaction between the position in sequences and the vowel identity was found in both male (F (4, 4560) = 55.02, $p < .001$, $\omega^2 = 0.038$) and female (F (4, 7478) = 48.47, $p < .001$, $\omega^2 = 0.022$) speakers. Therefore, the degree of reduction of vowel qualities was found to be varied across the vowel identity. The interaction between the position and the vowel identity was found in male speakers for the morphological condition but both male and female speakers showed an interaction for the speech style.

5.3.4 *Effects of the accent types*

Secondly, the tone which a vowel carries is considered. A two-way ANOVA was conducted using the tone of a vowel in VV sequences (High and Low) and vowel identity (/a/, /i/, /u/, /e/ and /o/) as independent variables. The dependent variables were the F1 and the F2 values (in Hz). Each dependent variable was analysed separately. In the ANOVA analysis, the effect of the each of the two independent variables and the interaction between those two factors were investigated.

Table 63 below shows the distance between singletons and VV sequences produced with High or Low tones for all the vowels. The vowels with Low tone tend to have lower values than the vowels with high tone for male speakers but those with High tone had slightly higher values for female speakers. The mean distances for each vowel were inconsistent between male and female speakers. This difference is the same as the findings of the morphological condition. Investigating the difference in the vowel identity, /i/ for male speakers and /e/ for female speakers showed a greater reduction on vowels with Low tone but the vowels with High tone was more reduced for other vowels.

Table 63 Mean Euclidian distances for five vowels with High or Low tones

	Male				**Female**			
	High		Low		High		Low	
	V1	V2	V1	V2	V1	V2	V1	V2
a	304.49	268.40	292.82	229.32	338.37	261.77	304.03	242.13
i	249.00	383.70	208.70	424.97	415.38	464.70	305.41	501.11
u	245.90	339.63	303.63	263.09	295.59	360.77	330.94	287.02
e	162.98	208.04	139.79	195.05	209.68	261.29	206.78	301.18
o	182.26	176.54	106.11	198.71	214.94	292.57	160.21	269.34

Initially, a univariate analysis was conducted to test whether there is a significant interaction among the tone (High or Low), the position in a sequence (V1 or V2) and the vowel identity (a, i, u, e and o). There was a significant interaction for both male speakers (*F* (4, 4550) = 13.567, $p < 0.001$, $\omega^2 = 0.003$) and female speakers (*F* (4, 7468) = 16.467, $p < 0.001$, $\omega^2 = 0.002$), although the effect size was small.

Then, a two-way AVOVA was conducted to test whether there is a significant difference for the tones and the vowel identity and whether there is an interaction between them. The tone did not have a significant effect on the Euclidian distance for male speakers (F (1, 4560) = 1.34, p = 0.25) but there was a significant effect for female speakers (F (1, 7478) = 4.98, p = 0.03, $\omega^2 = 0.0005$), although the effect size was fairly small. The vowel identity had a significant effect on the distance for both male (F (1, 4560) = 163.81, $p < .001$, $\omega^2 = 0.120$) and female speakers (F (1, 7478) = 238.00, $p <.001$, $\omega^2 = 0.109$) However, the interaction between the position in sequences and the vowel identity was found for both male (F (4, 4560) = 5.48, $p < .001$, $\omega^2 = 0.003$) and female (F (4, 7478) = 5.13, $p < .001$, $\omega^2 = 0.002$) speakers, although the interaction was small. This result suggests that the degree of reduction varies depending on vowel identities.

5.3.5 *The Interaction between V1/V2 position and the accent contour, the sonority scale and the articulatory movements*

Having observed a significant interaction between the vowel identity and other factors, the interaction between the position and the accent is considered. Table 64 below presents the mean Euclidian distance of V1 and V2 of vowel sequences with different accent types of the vowel sequences. In total, VV sequences produced with HH and HL accent had greater distances than VVs with LH and LL accent types for both male and female speakers. Compared with the morphological conditions, the overall distance is larger suggesting that the VV sequences are more reduced in the speech style conditions. Also, VVs with HL accent become more reduced in the speech style condition whereas VVs with HL accent is not reduced compared with other accent types. The common characteristics of the reduction were the large reduction on HH accent and the less reduction on LH accent.

Table 64 Mean Euclidian distances for V1 and V2 of vowel sequences with different accent types

	Male								**Female**							
	HH		HL		LH		LL		HH		HL		LH		LL	
	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
a	286.40	283.16	322.14	195.86	266.82	241.63	324.90	265.14	330.13	270.54	346.27	229.10	282.91	247.23	324.86	255.28
i	273.07	388.67	195.36	440.96	200.56	374.64	218.41	375.93	452.18	487.33	324.10	513.79	336.91	412.24	260.46	465.99
u	227.41	321.93	275.21	263.52	299.09	389.87	308.43	262.63	247.79	347.09	351.19	286.50	341.80	396.10	319.45	287.56
e	180.79	235.01	125.78	190.07	128.56	146.61	150.63	201.60	216.29	274.08	197.69	313.42	212.87	240.75	200.76	277.93
o	187.05	195.70	174.17	204.07	122.52	141.30	88.18	193.73	210.83	309.54	220.73	246.99	194.30	261.53	119.61	289.33

Initially, a univariate analysis was conducted to test whether there is a significant interaction among the accent type (HH, HL, LH and LL), the position in a sequence (V1 or V2) and the vowel identity (a, i, u, e and o). There was a significant interaction for both male speakers (*F* (12, 4530) = 6.017, $p < 0.001$, $\omega^2 = 0.003$) and female speakers (*F* (12, 7448) = 9.230, $p < 0.001$, $\omega^2 = 0.004$), although the effect size was small.

Then, a two-way AVOVA was conducted to test whether there is a significant difference for the accent types and the positions in VV sequences and whether there is an interaction between them. The accent pattern of VV sequences had a significant effect on the Euclidian distance for both male (F (3, 4562) = 15.29, $p < .001$, $\omega^2 = 0.009$) and female (F (3, 7480) = 24.25, $p < .001$, $\omega^2 = 0.009$) speakers. Scheffe post-hoc tests revealed that distances in vowel sequences which the first vowel was produced with High (HH and HL) were significantly longer than others for both male and female speakers, although the effect size was fairly small. Table 65 shows all the pairwise comparisons.

Table 65 Post-hoc comparison for Euclidian distances among accent types of VVs

	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
M	HH<HL*	-0.14	W	HH>LH*	0.14	W			
	HL>LH**	0.27	S	HL>LH**	0.21	S			
F				HH>LH**	0.17	W	HH>LL**	0.17	W
	HL>LH**	0.24	S	HL>LL**	0.23	S			

*: $0.001 < p < 0.05$, **: $p < 0.001$

W: Weak, S: Small, M: Medium, L: Large

There was a significant interaction between the accent contour and the positions in VVs for both male ($F(3, 4370) = 6.50$, $p < .001$, $\omega^2 = 0.004$) and female ($F(3, 7530) = 29.45$, $p < .001$, $\omega^2 = 0.011$) speakers.

Table 66 below presents the mean distance of V1 and V2 of vowel sequences with three degrees of the sonority scale. In total, Falling sequences were found to be more reduced than rising and level sequences and rising sequences were not reduced compared to others for both male and female speakers.

Table 66 Mean Euclidian distances for V1 and V2 of vowel sequences with different sonority scale

		Falling	**Rising**	**Level**
M	V1	272.48	210.01	230.3
	V2	358.71	208.15	275.47
	Total	315.59	209.08	252.88
F	V1	305.58	295.12	291.56
	V2	425.28	256.67	336.73
	Total	365.43	275.9	314.15

A two-way ANOVA was conducted to test whether there is a significant difference for the sonority scale and the positions in VV sequences and whether there is an interaction between them. The sonority scale difference between V1 and V2 of VV sequences had a significant effect on the Euclidian distance for both male ($F(2, 4564) = 159.28$, $p < .001$, $\omega^2 = 0.061$) and female ($F(2, 7482) = 126.76$, $p < .001$, $\omega^2 = 0.031$) speakers. Scheffe post-hoc tests revealed that all the pairwise comparisons were significantly different from each other for both male and female speakers. Table 67 below presents all the pairwise comparisons.

Table 67 Post-hoc comparison for Euclidian distances of VVs with different sonority scale

Male			**Female**		
Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
Fall>Rise**	0.60	M	Fall>Rise**	0.40	S
Fall>Level**	0.31	S	Fall>Level**	0.22	S
Rise<Level**	-0.28	S	Rise<Level**	-0.19	W

*: 0.001< p < 0.05, **: p < 0.001

W: Weak, S: Small, M: Medium, L: Large

There was a significant interaction between the positions in VVs and the sonority scale of VVs for both male (F (2, 4564) = 26.52, p < .001, ω^2 = 0.010) and female (F (2, 7482) = 97.18, p < .001, ω^2 = 0.024) speakers. The distances of vowels in V2 position were consistently larger than in V1 position in Falling and Level sequences and those in V2 position were larger in rising sequences. This result is the opposite to the findings of morphological condition.

Table 68 below presents the mean distance of V1 and V2 of vowel sequences of different degrees of the closing/opening dimension. In total, VV sequences with larger degree of opening or closing movement had longer distance than those with smaller degrees. VV sequences with no difference in the open/close dimension had longer distance compared to VV sequences with small degrees.

Table 68 Mean Euclidian distances for V1 and V2 of vowel sequences of open/close dimension

		Close2	**Close1**	**Same**	**Open1**	**Open2**
M	V1	320.51	200.75	230.30	201.66	236.88
	V2	428.03	255.19	275.47	193.22	256.19
	Total	374.27	227.97	252.88	197.44	246.54
F	V1	354.05	246.87	291.56	283.83	337.58
	V2	492.19	344.25	336.73	255.71	260.27
	Total	423.12	295.56	314.15	269.77	298.93

A two-way ANOVA was conducted to test whether there is a significant difference for the open/close dimension and the positions in VV sequences and whether there is an interaction between them. The open/close dimension of VV sequences had a significant effect on the Euclidian distance for both male ($F(4, 4560) = 190.63$, $p < .001$, $\omega^2 = 0.133$) and female ($F(4, 7478) = 146.52$, $p < .001$, $\omega^2 = 0.068$) speakers. Scheffe post-hoc tests revealed that the degree of closing had a significant effect on distance and the vowels are more reduced as the degree of closing increases. Closing sequences were more reduced than Opening sequences comparing at the same degree. Table 69 below presents all the pairwise comparisons.

Table 69 Post-hoc comparison for Euclidian distances of the close/open dimension

	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
M	C2>C1**	0.76	M	C2>ND**	0.61	M	C2>O1**	1.02	L
	C2>O2**	0.67	M				C1>O1*	0.20	S
				ND>O1**	0.35	S			
	O1<O2**	-0.37	S						
F	C2>C2**	0.54	M	C2>ND**	0.45	S	C2>O1**	0.68	M
	C2>O2**	0.50	M				C1>O1*	0.13	W
				ND>O1**	0.22	S			

*: 0.001< p < 0.05, **: p < 0.001
W: Weak, S: Small, M: Medium, L: Large

There was a significant interaction between the positions in VVs and the close/open dimension for both male (F (4, 4560) = 18.15, p < .001, $\omega^2 = 0.012$) and female (F (4, 7478) = 54.13, p < .001, $\omega^2 = 0.025$) speakers. Vowels in V2 position were more reduced in Closing sequences and VVs with the same height. For Open 1, vowels in V1 position were more reduced. For Open 2, male speakers had more reduced V2 but female speakers had reduced V1.

Table 70 below presents the mean distance of V1 and V2 of vowel sequences of different degrees of the fronting/backing dimension. In total, VV sequences consisted with vowels of the same degree of the antero-posterior position had shorter distance than others.

Table 70 Mean Euclidian distances for V1 and V2 of vowel sequences of front/back dimension

		Front2	Front1	Same	Back1	Back2
M	V1	236.82	301.00	193.47	200.36	195.07
	V2	235.98	381.97	151.93	238.44	293.26
	Total	236.40	341.48	172.70	219.40	244.16
F	V1	275.18	334.13	239.54	248.77	331.54
	V2	329.55	440.08	214.90	263.63	349.10
	Total	302.37	387.11	227.22	256.20	340.32

A two-way AVOVA was conducted to test whether there is a significant difference for the front/back dimension and the positions in VV sequences and whether there is an interaction between them. The front/back dimension of VV sequences had a significant effect on the Euclidian distance for both male (F (4, 4560) = 127.71, $p < .001$, $\omega^2 = 0.094$) and female (F (4, 7478) = 124.45, $p < .001$, $\omega^2 = 0.060$) speakers. Scheffe post-hoc tests revealed that the degree of fronting had a negative effect on distance and the vowels are more reduced as the degree of fronting becomes less. For backing sequences, vowels were more reduced when the degree of backing increases. VV sequences which have the same front/back feature were not reduced compared with others. Table 71 below presents all the pairwise comparisons.

Table 71 Post-hoc comparison for Euclidian distances of the front/back dimension

	Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
M	F2<F1**	-0.54	M	F2>ND**	0.39	S			
				F1>ND**	0.90	L	F1>B1**	0.67	M
	F1>B2**	0.50	M	ND<B1*	-0.35	S	ND<B2**	-0.44	S
F	F2<F1**	-0.35	S	F2>ND**	0.38	S	F2>B1**	0.23	S
	F2<B2**	-0.18	W	F1>ND**	0.69	M	F1>B1**	0.58	M
	F1>B1**	0.20	S				ND<B2**	-0.59	M
	B1<B2**	-0.44	S						

*: 0.001< p < 0.05, **: p < 0.001

W: Weak, S: Small, M: Medium, L: Large

There was a significant interaction between the positions in VV sequences and the front/back dimension for both male (F (4, 4560) = 18.01, p < .001, ω² = 0.013) and female (F (4, 7478) = 19.96, p < .001, ω² = 0.009) speakers. Vowels in V2 position were more reduced in most of the front/back levels except for the degree 2 fronting for male speakers. VV sequences with the same degree of the front/back had more reduced V1.

5.4. Transition speed

Transition speed is a rate of change of formant movement with respect to time. As Gore (2003) argues, transition speed is one of the factors of distinguishing between diphthongs and VV sequences and the difference of transition speed may indicate whether VV sequences in certain condition may have diphthong-like features. In general, diphthongs in English, for example, see a gradual change of the formants but the transition of the vowel sequences in Japanese is rapid and their transition speed is supposed to be faster than English diphthongs. Therefore, slower transition is assumed to be associated with more diphthong-like features. This section reports whether the transition speed is significantly different in various factors. In the present study, transition is defined as the Euclidian distance between V1 and V2, which is calculated by the following equation.

$$Transition\ Distance\ (Hz) = \sqrt{(F1V1 - F1V2)^2 + (F2V1 - F2V2)^2}$$

This distance is divided by the transition duration measured (in msec) and the transition speed is computed.

5.4.1 *Morphological condition*

Table 72 below presents the transition speed of three morphological conditions for both male and female speakers. A one-way ANOVA showed a significant difference for the transition speed among morphological conditions for both male (F (2, 2186) = 16.10, p < .001, $\omega^2 = 0.014$) and female (F (2, 3766) = 9.32, p < .001, $\omega^2 = 0.004$) speakers but effect size for both tests were small. Scheffe post hoc tests revealed that the transition speed for /hVV/ condition was significantly faster than WI condition for male speakers (p < .001, d = 0.28) and that for AWB was significantly faster than WI for female speakers (p < .001, d = -0.20). No other pairwise comparison was found to be significant. It is interesting to note that the total duration for VV sequences and the ratios for transition in VV sequences were different among morphological conditions but the transition speed was not necessary greatly different from each other. The mean transition speed for AWB was faster than WI for both male and female speakers, although a significant difference was not apparent for male speakers. Assuming that the transition speed is one of the factors of diphthongisation, VV sequences within words are arguably more diphthong-like compared to other morphological conditions although the difference is a small one.

Table 72 Transition Speed of VV sequences in different morphological conditions (Hz/msec)

	hVV	**WI**	**AWB**
Male	11.59	9.64	10.52
Female	12.64	11.94	13.55

The accent types of the vowel sequences were not found to be a factor of changing the transition speed. There was no significant difference for the transition speed among four accent types of VV sequences for male (F (3, 2185) = 1.68, $p = 0.17$) and female (F (3, 3765) = 2.49, $p = 0.06$) speakers. Therefore the transition part of VV sequences was produced almost at the same speed despite the different pitch pattern for the accent. Table 73 shows the transition speed for these accent patterns.

Table 73 Transition Speed of VV sequences in different accent pattern (Hz/msec)

	HH	**HL**	**LH**	**LL**
Male	9.82	10.72	10.34	10.40
Female	12.52	12.77	11.84	12.34

However, the articulations of VV sequences contributed to variability in the transition speed. Table 74 below presents the transition speed as a function of different sonority contours between vowels in VV sequences. Significant differences for transition speed were found for both male (F (2, 2186) = 38.00, $p < .001$, $\omega^2 = 0.032$) and female (F (2, 3766) = 95.10, $p < .001$, $\omega^2 = 0.046$) speakers from a one-way ANOVA. Scheffe post-hoc tests revealed that the transition speed of the level sequences was faster than other sonority scale for both male (Falling: $p< .001$, $d = -0.23$, Rising: $p < .001$, $d = -0.54$) and female (Falling: $p< .001$, $d = -0.43$, Rising: $p < .001$, $d = -0.61$) speakers. Comparing the falling and the rising sequences, the transition speed for falling sequences was faster than rising sequences for male (p< .001, d = 0.22) and female (p< .001, d = 0.13) speakers. This result means that the rising sequences became more diphthong-like because they have slower transition speed. Therefore it is necessary to investigate the direction and the degree of articulatory movements in more detail.

Table 74 Transition Speed of VV sequences in different sonority scale (Hz/msec)

	Falling	**Rising**	**Level**
Male	10.52	8.98	12.28
Female	12.04	11.00	15.60

Table 75 shows the transition speed for VV sequences of different close/open dimension. As was shown in the analysis of sonority scale, VV sequences of the vowels with the same height had the fastest transitions overall and the opening and the closing movement for VV sequences contributed to the slow transition speed to some extent. A One-way ANOVA revealed that there was a significant difference of transition speed for the open/close dimension for both male ($F(4, 2184) = 21.79$, $p < .001$, $\omega^2 = 0.036$) and female ($F(4, 3764) = 52.10$, $p < .001$, $\omega^2 = 0.051$) speakers.

Table 75 Transition Speed of VV sequences in the open/close dimension (Hz/msec)

	Close2	**Close1**	**N.D**	**Open1**	**Open2**
Male	9.55	10.99	12.28	9.25	8.22
Female	10.74	12.70	15.60	10.98	11.08

Scheffe post-hoc tests revealed that some of the pairwise comparisons between closing and opening sequences showed a significant difference and opening sequences had slower transition speed in all pairs. Although the different degree of closing movement was found to be significantly different in female speakers, no other pairwise comparison between the same directions of movement had a significant difference. The degree of opening/closing movement was not a major factor for VV sequences and way, therefore, nor contribute to their being diphthong-like. Table 76 presents all the pairwise comparisons.

Table 76 Post-hoc comparison for the transition speed of the open/close dimension

	Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
M				C2<ND**	-0.39	S			
							C1>O1*	0.24	S
	C1>O2**	0.38	S	ND>O1**	0.47	S	ND>O2**	0.69	M
F	C2<C1*	-0.23	S	C2<ND**	-0.63	M			
				C1<ND*	-0.34	S	C1>O1**	0.21	S
	C1>O2**	0.19	W	ND>O1**	0.61	M	ND>O2**	0.59	M

*: 0.001< p < 0.05, **: p < 0.001
W: Weak, S: Small, M: Medium, L: Large

Table 77 below summarises the transition speed of VV sequences of different front/back dimension. In contrast to the open/close dimension, VV sequences of vowels with the same position in the front/back dimension had the slowest transition speed. As the degree of fronting and backing increases, the transition speed becomes faster. A One-way ANOVA revealed that there was a significant difference of transition speed for the open/close dimension for both male ($F(4, 2184) = 136.02$, $p < .001$, $\omega^2 = 0.189$) and female ($F(4, 3764) = 231.20$, $p < .001$, $\omega^2 = 0.058$) speakers.

Table 77 Transition Speed of VV sequences in the front/back dimension (Hz/msec)

	Front2	**Front1**	**Same**	**Back1**	**Back2**
Male	15.08	8.44	5.53	7.71	11.83
Female	16.78	10.06	6.88	9.63	15.74

Scheffe post-hoc tests confirmed that the VV sequences of vowels with the same position in the front/back dimension had the slowest transition speed. This means that the VV sequences which both vowels are either both front or both back could be viewed as being more diphthong-like. Comparing the same direction of the movement (fronting or backing), the larger the degree of the movement becomes, the faster the transition speed becomes. It suggests that the vowel sequences of less front/back movement have more diphthong-like features. Thus, the front/back movements between V1 and V2 are negatively associated with diphthong-like features. Table 78 presents all the pairwise comparisons.

Table 78 Post-hoc comparison for the transition speed of the front/back dimension

	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
M	F2>F1**	0.91	L	F2>ND**	1.37	L	F2>B1**	1.15	L
	F2>B2**	0.43	S	F1>ND**	0.55	M			
	F1<B2**	-0.53	M	ND<B1*	-0.56	M	ND<B2**	-1.10	L
	B1<B2**	-0.75	M						
F	F2>F1**	0.87	L	F2>ND**	1.36	L	F2>B1**	0.94	L
				F1>ND**	0.53	M			
	F1<B2**	-0.77	M	ND<B1**	-0.48	S	ND<B2**	-1.29	L
	B1<B2**	-0.84	L						

*: 0.001< $p < 0.05$, **: $p < 0.001$
S: Small, M: Medium, L: Large

With respect to the transition speed on VV sequences of different morphological condition, VV sequences within word were found to be more diphthong-like compared to other condition. The phonetic features which may be related to the transition speed difference was the sonority scale of VV sequences and rising sequences which V2 was more sonorant than V1 had faster transition. The degree of the closing and opening was also related to the transition speed and the movement of closing corresponds to the increase of the transition speed. However, there was no correspondence between the degree of opening and the transition speed. For the fronting/backing movement of VV sequences, VV sequences with the same position had faster transition than others.

5.4.2 *Speech style condition*

This section intends to investigate whether the findings of morphological condition is applied to the dataset which contains VV sequences within words produced in different speech styles. Table 79 below presents the transition speed of four speech style conditions for both male and female speakers. One-way ANOVA showed a significant difference for the transition speed among speech style conditions for both male ($F (3, 2281) = 4.74$, $p < .001$, $\omega^2 = 0.005$) and female ($F (3, 3740) = 16.51$, $p < .001$, $\omega^2 = 0.012$) speakers but effect size for both tests were small. However, Scheffe post hoc tests revealed that only a few pairwise comparisons were significantly different. For male speakers, the transition speed of VVs of free conversation style was significantly faster than those of read speech ($p = 0.014$, $d = -0.17$). No other pairwise comparisons were different with each other. For female speakers, the transition speed of VVs of free conversation were significantly faster than those of read speech ($p < .001$, $d = -0.25$), response ($p = 0.001$, $d = -0.24$), and description ($p = 0.022$, $d = -0.22$). Thus, the conversation style may contribute to the faster transition. For morphological conditions, VV sequences within words were found to be more diphthong-like because the transition speed was slower than other conditions. However, the present data shows less variation in the transition speed across the speech styles.

Table 79 Transition Speed of VV sequences in different speech style conditions (Hz/msec)

	read	resp	desc	conv
Male	9.63	10.98	10.86	10.98
Female	11.94	11.93	12.09	14.03

As is the case in the morphological conditions, the accent types of the vowel sequences were not found to be a factor in changing the transition speed of VV sequences produced in different speech styles. A one-way ANOVA revealed that there was no significant difference for the transition speed among four accent types of VV sequences for male (F (3, 2281) = 2.48, p = 0.06) and female (F (3, 3740) = 2.44, p = 0.06) speakers. Therefore, the tonal differences of VV sequences such as High-Low and Low-High did not affect the transition speed and the transition between the steady states of VV sequences was produced almost at the same speed despite the different pitch pattern for the accent. Table 80 shows the transition speed for these accent patterns.

Table 80 Transition Speed of VV sequences in different accent pattern (Hz/msec)

	HH	HL	LH	LL
Male	10.20	10.61	9.30	10.67
Female	12.26	13.11	12.28	12.49

However, the articulations of VV sequences contributed to the change of the transition speed of VV sequences in different speech styles, which also agrees with the results from the morphological conditions. Table 81 below presents the transition speed of different sonority scale between vowels in VV sequences. Significant differences for transition speed were found for both male ($F(2, 2282) = 36.65$, $p < .001$, $\omega^2 = 0.030$) and female ($F(2, 3741) = 54.11$, $p < .001$, $\omega^2 = 0.027$) speakers from a one-way ANOVA. Scheffe post-hoc tests showed the same tendency for the pairwise comparison and revealed that the transition speed of the level sequences was also faster than other sonority scale for both male (Falling: $p< .001$, $d = -0.25$, Rising: $p < .001$, $d = -0.62$) and female (Falling: $p< .001$, $d = -0.29$, Rising: $p < .001$, $d = -0.53$) speakers. Comparing the falling and the rising sequences, the transition speed for falling sequences was faster than rising sequences for male ($p< .001$, $d = 0.24$) and female ($p< .001$, $d = 0.18$) speakers. This result also means that the rising sequences, which are characterised as having more open V2, could be viewed as more diphthong-like because they have slower transition speed. Therefore it is necessary to investigate the direction and the degree of articulatory movements in more detail.

Table 81 Transition Speed of VV sequences in different sonority scale (Hz/msec)

	Falling	**Rising**	**Level**
Male	10.38	8.44	12.72
Female	12.56	11.06	14.94

Table 82 below shows the transition speed for VV sequences of different close/open dimension. As it was shown in the sonority scale, VV sequences of the vowels with the same height had the fastest transition than others and the opening and the closing movement for VV sequences contributed to the slow transition speed to some extent. This category is intended to discuss whether the degree of opening or closing movement is associated with the transition speed. A One-way ANOVA revealed that there was a significant difference of transition speed for the open/close dimension for both male ($F(4, 2280) = 19.91$, $p < .001$, $\omega^2 = 0.032$) and female ($F(4, 3739) = 30.88$, $p < .001$, $\omega^2 = 0.031$) speakers.

Table 82 Transition Speed of VV sequences in the open/close dimension (Hz/msec)

	Close2	**Close1**	**N.D**	**Open1**	**Open2**
Male	9.96	11.01	12.72	8.68	7.68
Female	11.99	13.26	14.94	11.30	10.15

Scheffe post-hoc tests revealed that some of the pairwise comparisons between closing and opening sequences showed a significant difference and opening sequences had slower transition speed in all pairs. This means that the opening movement are slightly related to a feature of diphthongs. However, there was no difference between the different degree of opening and closing movement such as Open 1 vs. Open 2 and Close 1 vs. Close 2. Therefore, the degree of opening/closing movement was not a major factor for VV sequences being diphthong-like but the opening movement for male speakers seems to be associated with the diphthongisation of VV sequences. Table 83 presents all the pairwise comparisons.

Table 83 Post-hoc comparison for the transition speed of the open/close dimension

	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect	Comparison	Cohen's d	effect
M				C2<ND**	-0.30	S			
	C2>O2*	0.26	S				C1>O1*	0.31	S
	C1>O2*	0.42	S	ND>O1**	0.56	M	ND>O2**	0.68	M
F				C2<ND**	-0.37	S			
	C2>O2*	0.23	S	C1<ND*	-0.21	S	C1>O1**	0.23	S
	C1>O2**	0.37	S	ND>O1**	0.49	S	ND>O2**	0.69	M

*: 0.001< p < 0.05, **: p < 0.001

S: Small, M: Medium, L: Large

Table 84 below presents the transition speed of VV sequences of different front/back dimension. The tendency for the transition speed was similar to the morphological conditions. VV sequences of vowels with the same position in the front/back dimension had the slowest transition speed. As the degree of fronting and backing increases, the transition speed becomes faster. A One-way ANOVA revealed that there was a significant difference of transition speed for the open/close dimension for both male (F (4, 2280) = 77.94, p < .001, ω^2 = 0.115) and female (F (4, 3739) = 164.53, p < .001, ω^2 = 0.044) speakers.

Table 84 Transition Speed of VV sequences in the front/back dimension (Hz/msec)

	Front2	Front1	N.D	Back1	Back2
Male	14.91	9.28	5.59	7.15	11.94
Female	16.35	11.43	6.90	9.09	15.61

Scheffe post-hoc tests confirmed that the VV sequences of vowels with the same position in the front/back dimension had the slowest transition speed, which agrees with the outcome of the morphological conditions. This means that the VV sequences which both vowels are either front or back could be viewed as more diphthong-like. Comparing the same direction of the movement, the larger the degree of the movement becomes, the faster the transition speed becomes. It means that the vowel sequences of less front/back movement could be said to be more diphthong-like. Table 85 presents all the pairwise comparisons.

Table 85 Post-hoc comparison for the transition speed of the front/back dimension

	Comparison	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**	**Comparison**	**Cohen's d**	**effect**
M	F2>F1**	0.62	M	F2>ND**	1.18	L	F2>B1**	1.07	L
	F2>B2**	0.34	S	F1>ND**	0.47	S	F1>B1*	0.29	S
	F1<B2**	-0.31	S				ND<B2**	-0.99	L
	B1<B2**	-0.78	M						
F	F2>F1**	0.61	M	F2>ND**	1.36	L	F2>B1**	1.01	L
				F1>ND**	0.62	M	F1>B1**	0.32	S
	F1<B2**	-0.51	M	ND<B1*	-0.41	S	ND<B2**	-1.24	L
	B1<B2**	-0.89	L						

*: 0.001< p < 0.05, **: p < 0.001

S: Small, M: Medium, L: Large

With respect to the transition speed on VV sequences within words of different speech styles, the difference in the speech style was not significantly associated with it, although the transition speed of VV sequences in the conversation by female speakers were slightly faster than others. But the fast speed in transition represents less diphthong-like feature. In terms of the sonority contour, the transition speed for rising sequences was slower. The directions of the tongue movements were partly associated with the transition speed. For the opening movement, VVs with the same height were less diphthong-like, whereas VVs of both front and both back could be viewed as more diphthong-like.

5.5 Summary

This chapter reported the formant analysis and the transition speed of VV sequences from various factors. Vowels in VV sequences were found to be reduced from the corresponding singletons but the degrees of reduction was varied across the conditions. The slow transition speed was thought to be a diphthong-like factor, but the results showed that the transition speed did not have a significant effect for the variation across factors.

Chapter 6 Discussion

6.1 Introduction

Based on the reviews of literature in Chapter 2, three research questions were addressed, that is, 1) what are the acoustic properties of two-vowel sequences in Japanese? 2) To what extent do these vary as a function of the degree of cohesion between the two vowels in the sequence and as a function of different speech styles? 3) How do the findings of this study contribute to the debate in the literature regarding whether Japanese VV sequences can be classified as diphthongs? This chapter answers these questions based on the findings reported in Chapter 4 and 5.

6.2 Acoustic properties of the VV sequences in Japanese

Summarising the findings in Chapter 4 and 5, the results from the duration measurement for the VV sequences in Japanese indicate that the VV sequences in Japanese consist of two steady states, one for each vowel and a transition between the vowels. But the total VV duration and the proportions of VV interval occupied by the steady states and the transition vary across the morphological conditions of the VV sequences and the speech style in which they are produced. The total durations for the

VVs in the /hVV/ nonsense words were longer than those in real words and the durations for the VVs across a word boundary were shorter than the VVs within a word. The ratios of steady states for VVs in the /hVV/ condition were larger and the ratios for the transition were smaller than those in real words. Also, the total durations for the VVs of read speech were longer than other spontaneous speech styles. Among the spontaneous style conditions, the total durations of the VVs in the natural conversation were shorted than those in the response and the description styles. The proportions of steady states for VVs in the spontaneous speech styles were smaller than in the read speech style and the ratios of the transition for VVs in the spontaneous speech were larger than in the read speech. In addition, the steady states for V2 are consistently longer than those for V1. Although the morphological conditions and the speech style conditions were found to have major effects on the total duration and the subsection ratios, relatively little effect was found for the accent contour, the sonority scale and the direction of tongue movement for VV sequences.

The results from the formant measurement for the VV sequences indicate that the formant frequencies in the VV sequences are reduced in comparison with the corresponding singletons. The degrees of reduction depended on the morphological condition and the speech style. Also, the extent to which F1 and F2 values are reduced from the corresponding singletons varied across the vowel identity as well as between male and female speakers. Despite the different degrees of reduction between F1 and F2, the vowels in VVs across a word boundary are more reduced than VVs within words and VVs in the /hVV/ condition. In terms of the speech styles of VV sequences, the vowels in VVs produced in the natural conversation are more reduced than the read speech as well as other spontaneous styles. However, little effect is found for the degree of reduction of the accent contour, the sonority scale and the directions of the tongue movement.

Of course, the results just described and presented in detail in Chapters 4 and 5 are based on an analysis of the timing and spectral properties separately, as indeed has generally been the case in the existing literature on Japanese vowel sequences (e.g. Misono and Hirasaka 2008). However, the question arises whether the results for the timing measures are congruent with the results for the spectral measures. For example, is it the case that greater reduction for V1 and V2 is found in VV sequences which are shorter in duration overall and which as a result provide less opportunity (time-wise) for the tongue to achieve a relatively peripheral target configuration.

Firstly, the correlation between the measurements of total VV durations and the degree of reduction across singletons and the vowels in VV sequences is presented. Figure 33 below shows the correlation between the total VV sequence duration and the distance (i.e. the degree of reduction as indexed by the Euclidean distance) between V1 of the morphological condition dataset and the corresponding singletons. A Pearson correlation coefficient was calculated to identify the relationship between these two variables and there was a weak negative correlation between the two variables ($r (5958) = -0.197$, $p < 0.001$). Despite the weak correlation coefficient, the scatter plot illustrated that VV sequences of shorter duration showed the full range of reduction for V1 distance but VV sequences with longer duration were unlikely to have degrees of V1 reduction at the high end of the observed range.

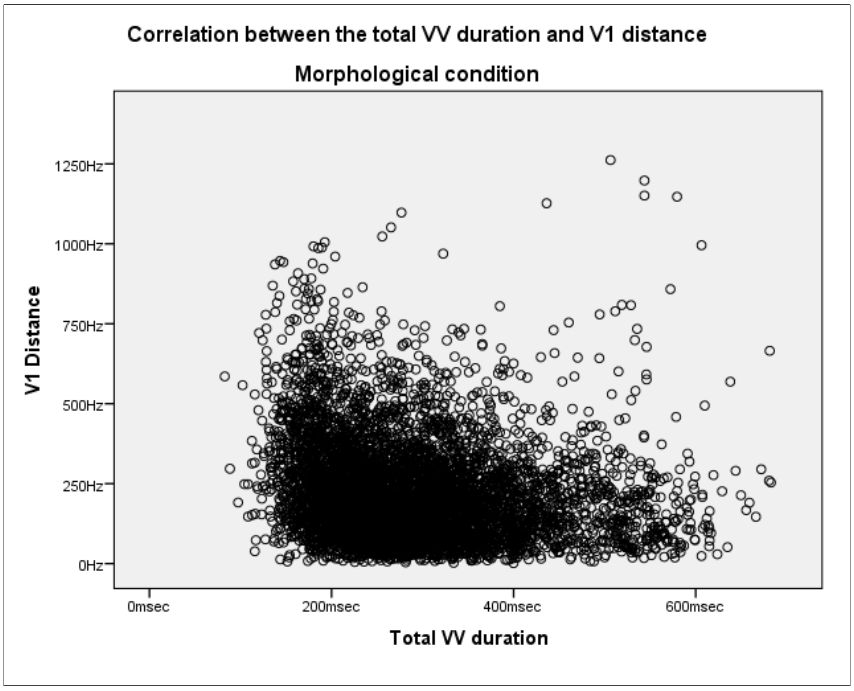


Figure 33 Correlation between the total VV duration and V1 distance of morphological condition dataset

Figure 34 below shows the correlation between the total VV sequence duration and the distance between V2 and the corresponding singletons of the morphological condition dataset. A Pearson correlation coefficient was computed to present the relationship between these two variables and there was a weak negative correlation between the two variables (r (5958) = -0.174, $p < 0.001$). As was the same with the V1 distance, VV sequences of shorter duration showed the full range of reduction for V2 distance but VV sequences with longer duration were unlikely to have degrees of V2 reduction at the high end of the observed range.

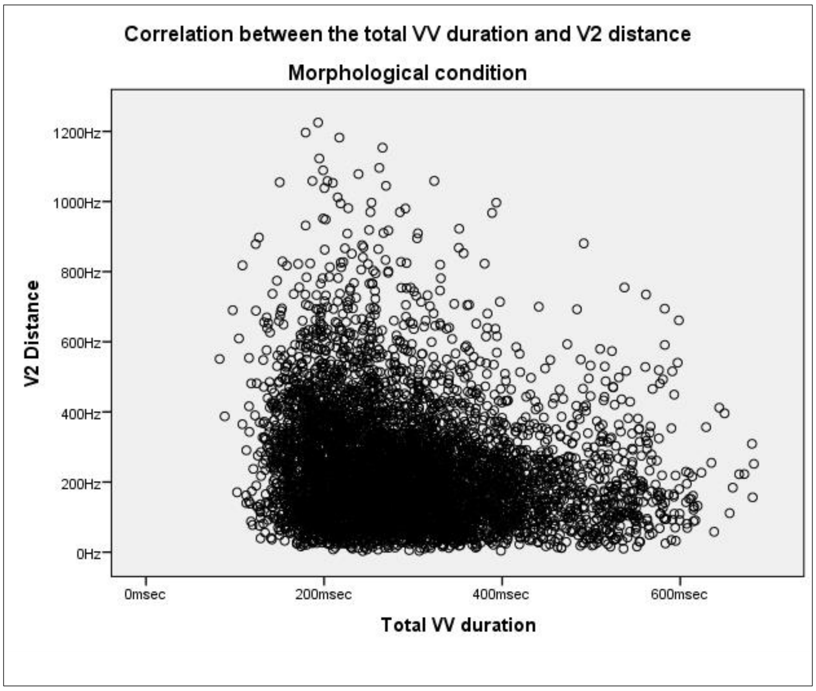


Figure 34 Correlation between the total VV duration and V2 distance of morphological condition dataset

Overall, the total VV duration and the degree of reduction of V1 and V2 from the corresponding singletons were associated with each other but the correlation was weak. The only reliable association resides in the fact that vowels in VV sequences of longer duration tend to undergo less reduction from the corresponding singletons.

Figure 35 below presents the correlation between the total VV sequence duration and the distance between V1 and the corresponding singletons of the speech style condition dataset. A Pearson correlation coefficient was calculated to investigate the relationship between these two variables and there was a weak negative correlation between the two variables (r (2598) = -0.235, p < 0.001). The scatter plot illustrates that VV sequences of shorter duration showed a wider range of reduction for V1 but VV sequences with longer duration were unlikely to have degrees of V1 reduction at the high end of the observed range. However, this tendency is not so apparent compared to the morphological condition data set.

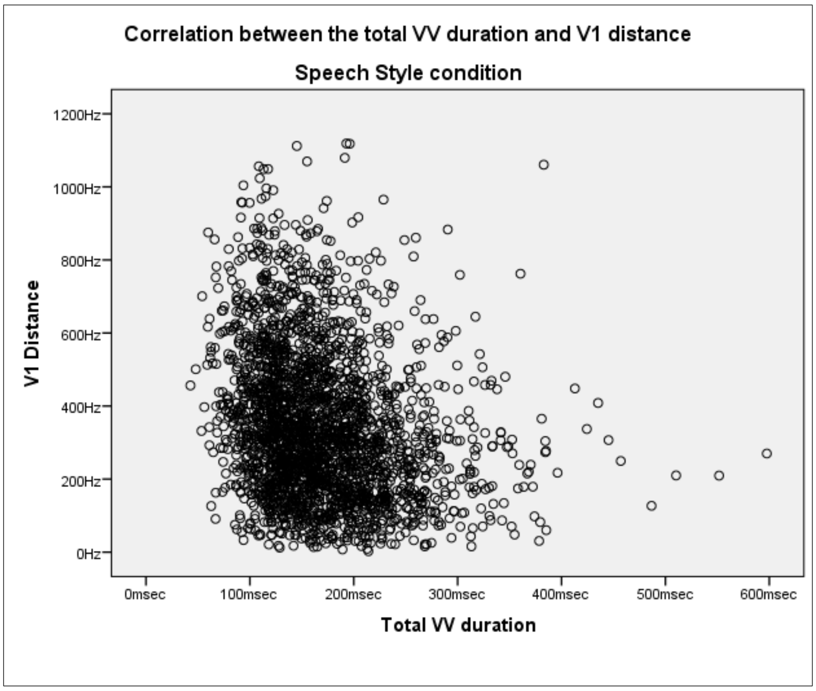


Figure 35 Correlation between the total VV duration and V1 distance of the speech style condition dataset

Figure 36 below illustrates the correlation between the total VV sequence duration and the distance between V2 and the corresponding singletons of the speech style condition dataset. A Pearson correlation coefficient was computed to identify the relationship between these two variables and there was a weak negative correlation between the two variables (r (2598) = -0.288, p < 0.001). It appeared to be a slightly clearer tendency that VV sequences with shorter duration showed the full range of reduction for V2 distance but VV sequences with longer duration were unlikely to have degrees of V2 reduction at the high end of the observed range.

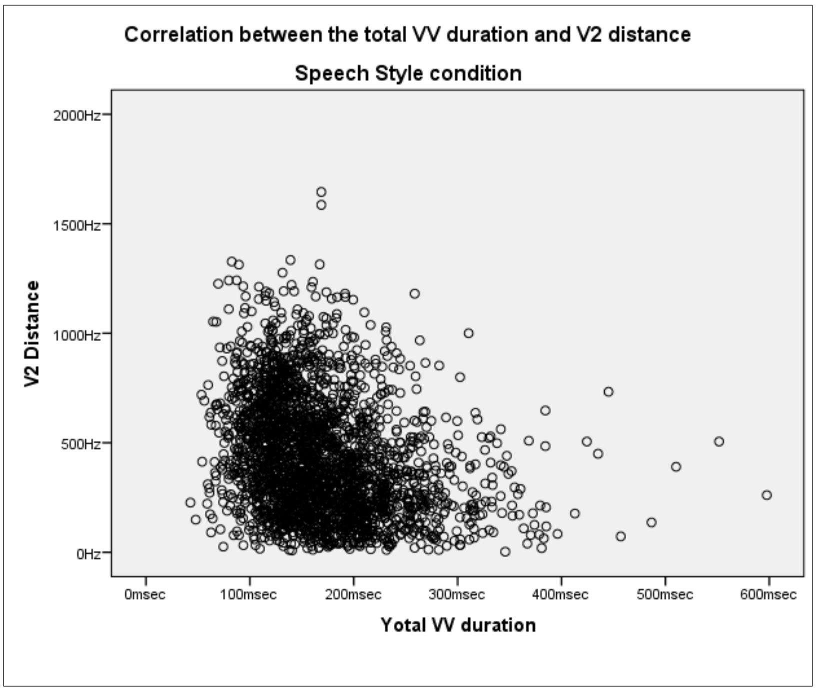


Figure 36 Correlation between the total VV duration and V2 distance of the speech style condition dataset

In sum, the vowels in VV sequences produced with various speech styles are found to be relatively less reduced from the corresponding singletons when the total VV sequence duration is relatively long. Therefore, the reduction feature and the duration feature are associated but the present study shows that overall this is a rather weak correlation. And the weakness of this correlation is highlighted by the fact that the vowels in VV sequences with shorter duration show the full range of reduction of quality.

In order to investigate this further, consideration is now given to the correlation between the degree of reduction of V1 and V2 in VV sequences from the corresponding singletons and the relative duration of V1 and V2. Figure 37 below shows the correlation between the V1 distance from the corresponding singletons and the V1 duration for the data acquired across all three morphological conditions. A Pearson correlation coefficient was computed to identify any relationship between these two variables and there was a very weak negative correlation between the two variables ($r (5958) = -0.071$, $p < 0.001$). The scatter plot suggests that the lower degree of reduction for V1 in VV sequences is associated with shorter V1 durations because tokens whose V1 reduction is greater are not as likely to show long V1 durations as those with lesser degrees of reduction.

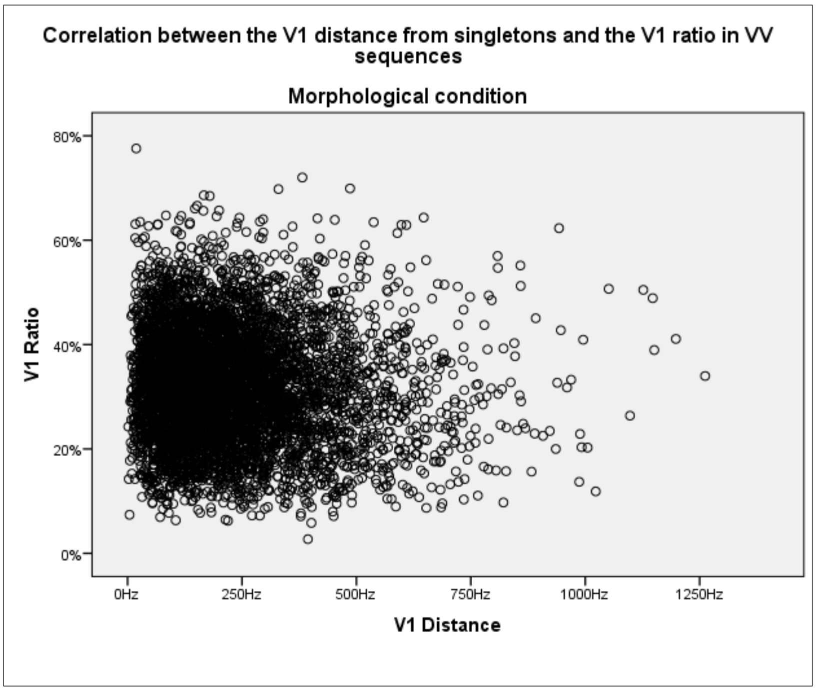


Figure 37 Correlation between the V1 distance and V1 ratios of VV sequences of three morphological conditions

Figure 38 below presents the correlation between the V2 distance from the corresponding singletons and the V2 durations across all three morphological conditions. A Pearson correlation coefficient was calculated to show the relationship between these two variables and yielded a very weak negative correlation between the two variables (r (5958) = -0.035, $p < 0.001$). A wide range of V2 duration can be seen irrespective of the degree of V2 reduction.

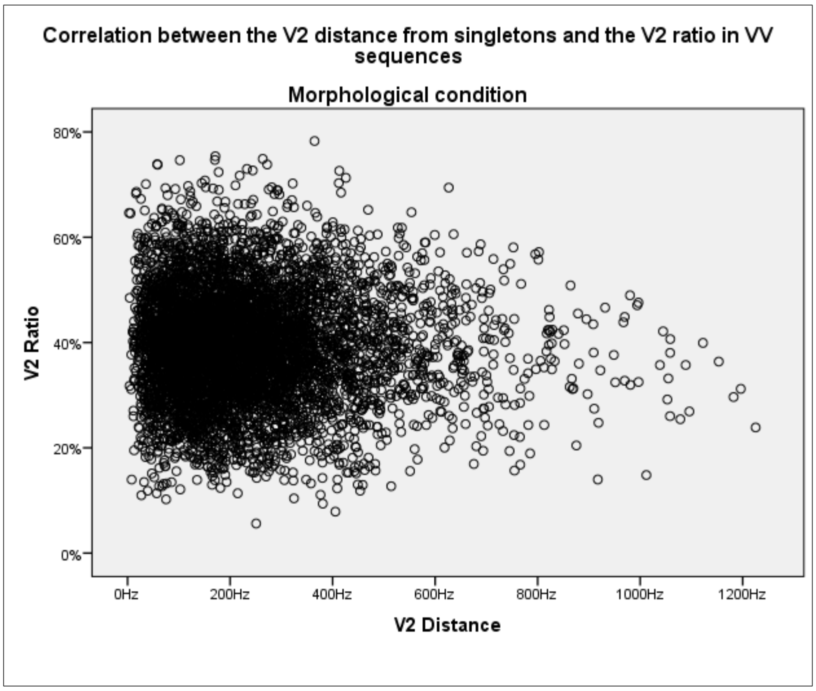


Figure 38 Correlation between the V2 distance and V2 ratios of VV sequences of three morphological conditions

Overall, the V1 and V2 distance and their ratios are found to be slightly associated with each other. Despite the fact that the ratios for V1 and V2 vary when the distances from the corresponding singletons are short, the V1 and V2 ratios become less variable as the V1 and V2 distances are getting longer. This finding suggests that the quality of the steady states for V1 and V2 varies irrespective of the degree of reduction.

Figure 39 below illustrates the correlation between the V1 distance from the corresponding singletons and the V1 ratios occupying the VV sequences produced with the response, description and conversation speech styles. A Pearson correlation coefficient was computed to represent the relationship between these two variables and there was a very weak positive correlation between the two variables (r (2598) = 0.085, p < 0.001). The scatter plot presents that V1 ratios vary irrespective of the degree of V1 reduction from the singletons.

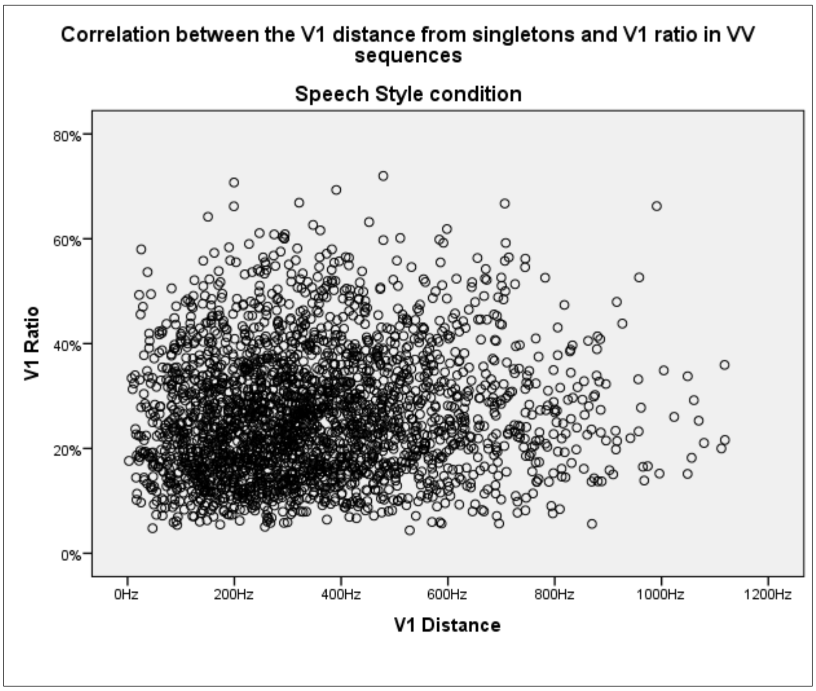


Figure 39 Correlation between the V1 distance and V1 ratios of VV sequences of three speech styles

Figure 40 below shows the correlation between the V2 distance from the corresponding singletons and the V2 ratios occupying the VV sequences produced with the response, description and conversation speech styles. A Pearson correlation coefficient was calculated to show the relationship between these two variables and there was a very weak negative correlation between the two variables (r (2598) = -0.059, p = 0.003). The scatter plot shows that the V2 ratios vary when the V2 distance is below

500Hz but the V2 ratios become slightly smaller when the V2 distance becomes longer. Thus, as is the same with V1, V2 ratios vary irrespective of the degree of V2 reduction from the singletons overall but V2 ratios seems to be slightly smaller when the degree of the V2 reduction is greater.

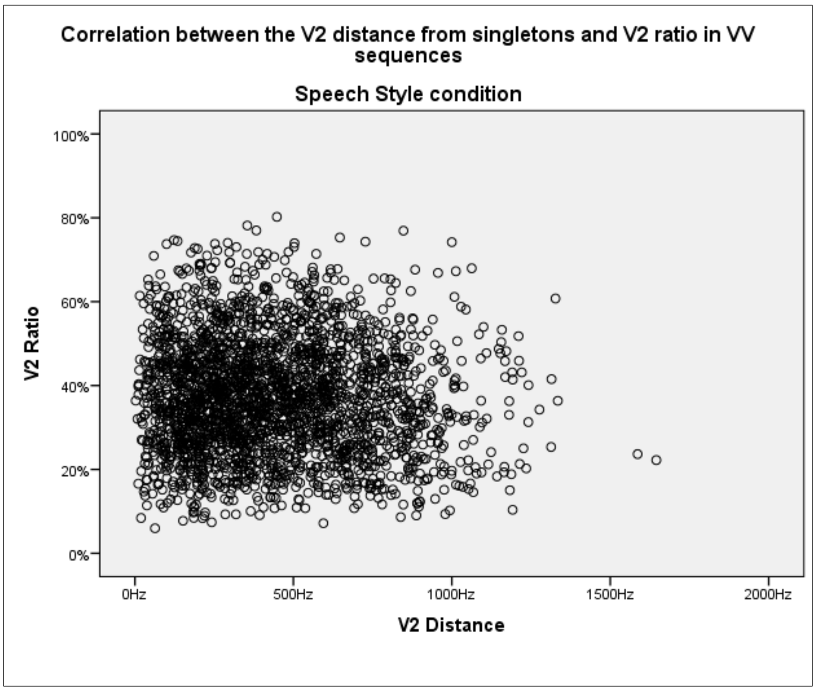


Figure 40 Correlation between the V2 distance and V2 ratios of VV sequences of three speech styles

In total, there was a very small association between the ratios of the steady states and the degree of reductions and the correlation coefficient is smaller in the case of the speech style condition. This implies that the steady states in VV sequences can be large when the degree of reduction is small and they can also be small when the quality is greatly reduced from singletons.

Therefore, the vowel quality of the steady states of VV sequences had a very weak association with their ratios in VV sequences. There are cases where a VV sequence has large steady state ratios with reduced qualities from the corresponding singletons and case where a VV sequence has small steady state ratios with the similar vowel qualities to the corresponding singletons. For VV sequences having longer duration, it is natural that the vowels are less likely to be reduced because the articulators have enough time to prepare for the target quality for V1 and move towards the target quality for V2, and the quality for those vowels are kept for certain amount of time to be perceived as such. However, the association between the durations and the degree of reduction is not straightforward in the case of VV sequences with shorter duration, although there is a weak correlation. The ranges of the reduction degree widely varied in VV sequences with shorter durations and the ranges of ratios for the steady states also varied in VV sequences with less reduced qualities. This means that some short VV sequences have short steady states with more reduced qualities but it is also possible that shorter VV sequences have longer steady states with less reduced qualities, that is, articulators hit the target positions for two vowels in a short time and the articulations for both vowels are kept steady for certain amount of time.

Looking back the relationship between the articulatory movement and the acoustic features of VV sequences reported in Chapter 4 and 5, the total VV duration was not associated with the opening/closing movements and was associated with the fronting/backing movements for the morphological conditions. For the speech style conditions, the closing sequences were found to have significantly shorter VV durations than the opening sequences and the fronting sequences were found to have significantly shorter VV durations than the backing sequences. However, the association between the articulatory movement and the ratios of transitions were weak for the morphological condition, although V1 ratios of closing sequences were larger than opening sequences and V2 ratios of closing sequences were smaller than opening sequences for VV sequences with different morphological condition. For the speech style condition, VV sequences without opening and closing movements had longer V1and a shorter transition. Therefore, the association between the articulatory movement and the ratios of subsections was variable. The association between the articulatory movement and the degrees of reduction was also weak and inconsistent between the morphological condition and the speech style. However, the transition speed was associated with the articulatory movement. The movement from V1 steady states and V2 steady states was faster when the vowels in VV sequences had the same height and there was a fronting or a backing movement for VV sequences.

Having summarised the main findings of the study, the question arises as to how consistent these are with what has previously been reported for Japanese VV sequences. This is the topic of the next section.

6.3 Comparison with the existing literature on Japanese VV sequences

6.3.1 *Papers dealing with VV sequences in Japanese*

Previous studies on the acoustic analysis on Japanese VV sequences have typically looked at only a subset of possible VV sequences. For example, Hirasaka and Kamata (1984) used two sequences /ai/ and /oi/, Gore (2006) used /ai/ for the experiment because /ai/ was thought to be one of the most common sequences in Japanese and frequently used in modern word endings and given names of Japanese people and Misono and Hirasaka (2008) used three Japanese VV sequences /ai/, /oi/, /au/ for the purpose of comparing with the realisations of English diphthong [ɑɪ], [ɔɪ] and [ɑʊ]. It is true that the /ai/ sequence represents Japanese VV sequence in a sense; it is just a part of the sequences out of 20 candidates despite the frequency. Therefore, the acoustic features of VV sequences previous studies have described do not reflect the full range of VV sequences in Japanese. The present study is based on the full range of VV sequences in Japanese thereby providing a more comprehensive account of the properties of VV sequences of Japanese. Although Fujisaki et al. (1979) covered all the possible Japanese VV sequences, they analysed the durations of the first steady states and the transitions but did not mention the second steady and the total durations. Mean durations of the transition interval were shown but what is not shown or could not be calculated is the ratios of the transition intervals occupying the total durations of VV sequences.

The fact that this study investigated the full set of VV sequences in Japanese enabled a number of factors to be considered which had not been investigated in previous studies of Japanese vowel sequences.

Firstly, the ratios for V1, transition and V2 ratios were within a similar range to that reported in the previous literature (See table 7, Chapter 2). Where the present results differ, however, is that they showed only a relatively small difference between V1 and V2, in contrast to findings such as those reported by Hirasaka and Kamata (1981) who reported ratios of 17.7% and 36.0% for V1 and V2 respectively. A similar picture emerged from the work of Kasuya and Sato (1990) who reported ratios of 15.0% and 45.0% respectively, and Misono and Hirasaka (2008) with 24.2% and 38.5% respectively. On the other hand, in Gore's (2006) study yielded ratios of 33.1% and 43.8% for V1 and V2 respectively, and was therefore a little more in line with the present results, 32.1% and 39.5% respectively for the morphological condition and 28.6% and 38.0% respectively for the speech style condition. But clearly, what the present have in common with these pre-existing ones is the finding that the V2 ratios are on average larger than the V1 ratio.

Secondly, Misono and Hirasaka (2008) argued that the vowel qualities of V2 in Japanese VV sequences were similar to those of monophthongs, although they did not conduct any statistical analysis. However, the present study showed that both V1 and

V2 in VV sequences were significantly reduced from the corresponding singletons for both morphological condition and the speech styles but the difference between V1 and V2 in the degree of reduction was varied across the vowel identity and genders. For example, V1 was more reduced than V2 in /a/ and /i/ for female speakers in the morphological condition, /a/ for both male and female speakers in the speech style condition.

Thirdly, in terms of the accent contours of VV sequences, the possible accent contours were not covered in the existing literature. Hirasaka and Kamata (1984) and Gore (2006) did not mention anything about the accent contour of the stimuli, Misono and Hirasaka (2008) compared HL and LH accent contours and Fujisaki et al. (1979) used the HH accent contour. The present study covered the full range of accent contours for VV sequences. Total durations for VV sequences produced with the flat accent contours (HH and LL) were significantly shorter than those with the changing accent contour (HL and LH) for the morphological conditions but it was not consistent with the VV sequences within words produced in different speech styles. In terms of the transition ratios, there was no effect across the four accent contours of VV sequences within words. For speech styles, there was a significant difference across the accent contour but the effect size was small. In terms of the degree of reduction, it varied between genders as well as the vowel identity. Also, the transition speed for VV sequences was not significantly different across the accent contours for both morphological condition and speech style. Takubo et al (1998) and Kubozono (2001) stated that accent contour of VV sequences forming diphthongs is High-Low (HL), however, there was no positive evidence to support their statement.

Finally, the speech style variation was not taken into consideration for the description of the VV sequences in Japanese in existing literature. The present study showed that the speech style influenced the total duration of VV sequences, the ratios of subsection intervals and the degrees of reduction. Although the transition ratio can be large in some cases, steady states for both V1 and V2 are found.

6.3.2 *Studies arguing that Japanese VV sequences have diphthong-like features*

Turning to the second research question, as discussed in Chapter 2, there are some researchers who argue that certain vowels can be classified as diphthongs in certain conditions. Despite the phonological analysis for the VV sequences, some vowel sequences in Japanese are said to have diphthong-like features if articulators continuously move during the VV sequence (Saito, 1997). It has also been pointed out that the conditions in which VV sequences in Japanese are likely to be diphthong-like include VV sequences without ‘a semantic boundary’ between the vowels (Kawakami, 1977). This statement may be related to the phonological definition of diphthongs, such as the vocalic elements which form a glide within one syllable (Cruttenden, 2001). It should not be possible to separate the elements in VV sequences if they are regarded as more diphthong-like. Even if it is undeniable that VV sequences in Japanese occupy two morae, VV sequences within a ‘unit’ may be assumed to be more diphthong-like in terms of their cohesion. The previous studies did not provide detailed acoustic evidence, but Gore (2006) reported that one Japanese VV sequence, /ai/ within a morpheme had significantly shorter transition than /ai/ across a morpheme boundary. His finding is consistent with the arguments addressed by Kawakami (1977) and Saito (1997). However, the present study showed that the VV sequences across a word boundary (AWB) had longer transitions but the effect size was small. The difference between the VV sequences in /hVV/ nonsense words and the VV sequences in real words (WI) as well as across word boundary (AWB) was large. The ‘semantic boundary’ is not well defined in Kawakami (1977) and Saito (1997), therefore, there is much room for discussion for the condition that Japanese VV sequences can be more diphthong-like. Different types of boundary can be tested in the future studies.

VV sequences in normal and natural speech are also said to be more diphthong-like (Kawakami, 1977; Saito, 1997). However, Kawakami and Saito did not provide any acoustic evidence. In the present study, VV sequences were produced in read, response, description and conversational styles. The transition ratios were 30.4% for the read speech, 36.1% for the response style, 42.7% for the description style and 36.8% for the conversation style. The read speech data was the same as the word-internal data in the morphological conditions, therefore, the ratios of transition of ‘spontaneous’ speech were much larger than the read speech. Also, the degrees of reduction were different across the speech styles and the VV sequences produced with natural conversation were greatly reduced from singletons. However, the degree of reduction was different among the vowel identity and the degree of reduction varied between F1 and F2. This finding supports the idea that the VV sequences in natural speech are more-diphthong-like in terms of the timing properties.

For the sonority difference between V1 and V2, there was less positive evidence on the statement of Kubozono (2001) that V1 of VV sequences in Japanese must be at least as sonorous as V2 to have diphthong-like features. For example, the total duration of VV sequences of different sonority scale showed a significant difference only in the speech style condition. Also, it was found that falling sequences had larger transition proportions than rising sequences but the effect size was fairly small or even weak. Furthermore, the transition speed for falling sequences was consistently faster than rising sequences for both morphological condition and speech style.

Now the question arises whether it is still appropriate to think of Japanese VV sequences as diphthongs as some researchers have argued. It can definitely be denied from the phonological point of view because each vowel in a VV sequence belongs to a different mora or syllable, therefore it is against the phonological condition of diphthongs that a diphthong must necessary consist of one syllable. However, from a phonetic point of view, there are many VV sequences which can be regarded as diphthong-like. For example, some tokens had longer transitions and the qualities of the vowels in VV sequences were significantly reduced from the corresponding singletons.

The findings of the present study are not unlike those reported for some languages which are uncontroversially analysed as having diphthongs. For example, the mean ratios for V1, transition and V2 across the morphological condition are 32.1%, 28.4% and 39.5% respectively. These ratios are similar to the findings that Svantesson (1984) reported for Chinese diphthongs, showing that the ratios of the onset, the transition and the offset are 29.2%, 24.0% and 37.3% respectively. However, the present result for the Japanese VV sequences is a bit different from the findings that Jha (1985) reported for Maithili diphthongs, showing that the ratios of the onset, the transition and the offset are 16.4%, 57.4% and 26.2% respectively. English diphthongs have been reported by many researchers and the results vary widely. Ratios of the onset, the transition and the offset for English diphthongs are 41%, 38% and 21% respectively reported by Kasuya and Sato (1990) whereas 14.2%, 77.5% and 8.3% respectively reported by Gay (1968) and 26.9%, 51.0% and 22.1% reported by Cox (2006). Therefore, a diphthong-like feature or even a diphthong feature depends on the point of reference. This suggests that the phonetic category of a diphthong is perhaps less clearly defined than is typically thought which makes it not surprising perhaps that debates break out, as has been the case with Japanese, about whether a language has diphthongs or not.

Notwithstanding the above, the present results show that there are conditions in which VV sequences are more diphthong-like. For example, the vowels in VV sequences in spontaneous speech reported in Chapter 5 are fairly reduced from the corresponding singletons. Another way of demonstrating this is to consider the patterns of distribution of Japanese VV tokens with relatively long transition ratios in order to see if there are particular environment or conditions which give rise to these (the idea being that tokens with longer transition ratios are more diphthong-like than those with shorter ones).

The number of VV sequences whose transition ratio was more than 50% is presented in Table 86 below. The overall number of VV sequences for morphological condition was 188 (3.2%). The VV sequences with long transition make up 0.4% of the total /hVV/ condition, 3.6% of WI and 6.7% of AWB. The overall results showed that VV sequences of AWB had larger transition ratios and this fact is also reflected in the ratios of the VV sequences of long transition.

Table 86 Number of VV sequences of the morphological condition having large (more than 50%) transition ratios

Condition	n	Total n	%
hVV	7	1687	0.4%
WI	125	3431	3.6%
AWB	56	840	6.7%
Total	188	5958	3.2%

On the other hand, the VV sequences of different speech styles had slightly larger number of the long-duration VV sequences, making up 18.2% of the total number of tokens (See Table 87). Among the speech styles, the description style had the largest ratio of the long-transition VVs, occupying 30.8%. Read speech is the same as the WI condition, therefore, the long-transition VV sequences are more likely to occur in spontaneous and natural speech. This finding supports the argument that Japanese VV sequences can have diphthong-like features in natural speech.

Table 87 Number of VV sequences of the speech style condition having large (more than 50%) transition ratios

Condition	n	Total n	%
resp	83	595	13.9%
desc	101	328	30.8%
conv	288	1675	17.2%
Total	472	2598	18.2%

Now we look at the relative frequency with which VV sequences with a long transition (> 50% of total VV sequence duration) are produced across some of the different conditions investigated in this study. Only that sub-set of conditions which helps refine the general picture emerging from Tables 86 and 87 are considered here.

The direction and extent of tongue movement was found to be associated with the variation of VV sequences with long-transition. Table 88 below shows the ratios of long-transition VVs across the five categories of the opening/closing dimension.

Table 88 Number of VV sequences having large (more than 50%) transition ratios across the opening/closing dimension

C/O	**Morphological**			**Speech Style**		
	n	Total n	%	n	Total n	%
Close 2	22	676	3.30%	229	1240	18.50%
Close 1	46	1338	3.40%	105	475	22.10%
N.D	55	1369	4.00%	86	421	20.40%
Open 1	42	1918	2.20%	46	404	11.40%
Open 2	23	657	3.50%	6	58	10.30%

For the morphological conditions, VV sequences consisting vowels of the same tongue anterior position (N.D.) had the largest proportion of the long-transition VVs and therefore, more likely to have diphthongised features, although the proportions in Close 1, Close 2 and Open 2 were just slightly lower than N.D. Open 2 was greatly lower than other categories. As for the speech style conditions, opening sequences (Open 1 and 2) were unlikely to have long transition compared with others. Comparing the directions of the tongue movement, a closing movement is more associated with longer transition than an opening movement. But there is a variation because VVs of Open 2 category had large amount of long transition VVs in morphological condition but the smallest amount in the speech style condition.

Table 89 below shows the ratios of long-transition VVs across the five categories of the fronting/backing dimension.

Table 89 Number of VV sequences having large (more than 50%) transition ratios across the fronting/backing dimension

F/B	Morphological			Speech Style		
	n	Total n	%	n	Total n	%
Front2	61	1326	4.60%	120	503	23.90%
Front1	48	1343	3.60%	290	1401	20.70%
N.D	11	619	1.80%	10	128	7.80%
Back1	34	1308	2.60%	13	221	5.90%
Back2	34	1362	2.50%	39	345	11.30%

For the morphological condition, VV sequences consisting vowels of the same tongue height (N.D) had the smallest proportion of the long-transition VVs. For the speech style conditions, these VVs did not have the lowest proportion but lower proportion among the categories. Comparing the fronting and the backing sequences, the fronting sequences were likely to have long transition for both morphological conditions and speech style conditions. Also, the degree of the fronting movement is associated with the larger ratios of VVs with long transition. Therefore, the fronting of the tongue in producing VV sequences is slow enough to have longer transitions. On the other hand, the backing movement is less likely to make long transition for both morphological and speech style conditions.

In addition to the conditions of the VV sequences and the characteristics of VV sequences, we also look at the inter-subject variance. The mean proportions of producing the VV sequence with long duration among the total number of productions were 3.2 % for the morphological condition and 18.6% for the speech style condition.

However, the proportions greatly vary between subjects. Table 90 below presents the data from 6 participants; the three who produced the smallest number of long-transition tokens, and the three who produced the greatest number of long-transition tokens.

Table 90 Inter-subject variance of VV sequences having large (more than 50%) transition ratios

Morphological				**Speech Style**			
Participants	n	Total n	%	Participants	n	Total n	%
1	19	378	5.0%	a	17	54	31.5%
2	18	377	4.8%	b	22	76	28.9%
3	16	376	4.3%	c	20	74	27.0%
4	6	378	1.6%	d	6	59	10.2%
5	7	376	1.9%	e	9	98	9.2%
6	2	376	0.5%	f	15	179	8.4%

The table shows whether there is a personal tendency to produce VV sequences with long transition or not. All the participants were found to produce VV sequences with long transition for both conditions. For the morphological condition, the total number of tokens is almost the same but 5% of the VVs produced by the participant 1 had longer transitions and the participant 6 was less likely to produce VV sequences with long transition. The number of tokens for the speech style is different between participants because the number of VV sequences in the natural conversation was greatly varied. 31.5 % of VV sequences produced by the participant had long transition whereas 8.4% of VVs produced by the participant had long transition. Although it is more likely that VV sequences with long transition are produced in more spontaneous or natural speech styles than read speech, there is a personal tendency to produce VV sequences with long transition. Therefore, certain people tend to produce VV sequences with more diphthong-like features but certain people do not.

Finally we look at the variability in individual sequence to see which VV sequence is more likely to have long transitions. Table 91 below shows the proportions of long-transition VVs across the VV sequences.

Table 91 Number of VV sequences having large (more than 50%) transition ratios across the VV sequences

VVs	Morphological			Speech Style		
	n	Total n	%	n	Total n	%
ai	17	345	4.90%	225	1150	19.60%
au	5	331	1.50%	4	90	4.40%
ae	20	335	6.00%	51	176	29.00%
ao	6	334	1.80%	2	61	3.30%
ia	21	324	6.50%	2	25	8.00%
iu	7	335	2.10%	9	94	9.60%
ie	5	332	1.50%	9	85	10.60%
io	14	323	4.30%	16	139	11.50%
ua	2	333	0.60%	4	33	12.10%
ui	12	333	3.60%	35	164	21.30%
ue	6	327	1.80%	5	50	10.00%
uo	6	323	1.90%	1	43	2.30%
ea	2	319	0.60%	5	45	11.10%
eu	5	336	1.50%	8	76	10.50%
eo	8	332	2.40%	6	36	16.70%
oa	9	330	2.70%	10	42	23.80%
oi	15	333	4.50%	44	162	27.20%
oe	28	333	8.40%	36	127	28.30%

In the data obtained from different speech styles, /ai/, /ae/, /ui/, /oa/, /oi/ and /oe/ stand out as being sequences in which long transitions are found with relatively high frequency. This is entirely consistent with the data presented in Tables 88 and 89 above which show that long transitions are associated with sequences with the same tongue height (N.D) and Front 2. What this points to is the presence of articulatory influence on the realisation of VV sequences; such sequences are more likely to be more diphthongal if they fall within certain articulatory parameters. This is not the first study to identify a sub-set of sequences as being more diphthong-like (Hattori (1967); Saito (1997); Kubozono (2001)), but what the present study has shown, which has not emerged from this previous work, is that this sub-set can be defined in articulatory terms.

Therefore, there is no simple answer to the question whether VV sequences in Japanese are diphthongs. The present findings point to a need for much further work to be done across-languages on the realisation of diphthongs and vowel sequences in order to explore these issues in a better-informed manner.

6.4 Application to the pronunciation teaching

Having discussed the factors that VV sequences in Japanese may have diphthong-like features, we now consider how it is related to the potential problems of speakers of languages with and without phonemic diphthongs learning each other’s languages, especially from the point of the production. It has been pointed out that Japanese speakers tend to make inaccurate pronunciation for diphthongs such as English diphthongs (Niwa, 1956; Hiraiwa 1981). However, few studies provided details of why they believe this to be the case. Yamaguchi (2007) argued that English diphthong [ou] can be replaced by Japanese [o] and [u]. Also, Nakamura et al. (2010) pointed out Japanese learners tend to produce English diphthongs as vowel sequences, especially making the latter part of diphthongs unnecessarily clearly. Therefore, the problem of speakers of Japanese, which lacks phonemic diphthongs, is that they are likely to replace diphthongs with the vowel sequences. However, it is still uncertain whether it is difficult for Japanese speakers to pronounce English diphthongs. Kitagawa (2012) reported how Japanese speakers produce English diphthongs and how their pronunciations of diphthongs are similar to those of native speakers. She measured the total duration, trajectory duration and the rate of change of English diphthongs [ei, oi, ai, ou, au] produced by 5 Japanese learners of English and compared them with those produced by American and British English speakers. She pointed out that the degree of formant movement for [ei] produced by Japanese learners was significantly smaller than others, but other diphthongs did not show significant difference. However, the formant frequencies were not reported in her study. Thus, the problem of producing diphthongs by Japanese speakers is that they do not make an articulatory movement from the onset and the offset of [ei]. This tendency is also reported in other researches. For example, Hansen et. al (1995) noted that Japanese learners did not change the position of the tongue and making glide in producing diphthongs. Several researchers argued that Japanese learners tended to make no distinction between [ɔː] and [ou] (Tejima 2011; Tsukuma 2005; Shimizu 2008). Therefore, Japanese speakers should pay more attention to the trajectory movement of diphthongs. But it would not be a matter of concern to produce English diphthongs in a way that diphthongs should be produced. The present study showed that VV sequences produced by Japanese speakers had diphthong-like features in certain environment, both from the time domain analysis and the spectral domain analysis. Some researchers even argued that utilising the pronunciations of five Japanese vowels is enough to teach the pronunciation of English diphthongs as a Lingua Franca (Uchida 2008, Shimizu 2011), that is, trying to produce VV sequences is enough to learn English diphthongs at the level of Lingua Franca.

On the contrary, English speakers were said to have a tendency that they tend to produce V1 longer and V2 shorter when they produce Japanese VV sequences (Nakagawa 1996). This tendency reflects the way that diphthongs are produced by native speakers. Roach (1983) pointed out that the first part of a diphthong is much longer and stronger than the second part and Ladefoged (2006) also noted that the last part of a diphthong is often so brief and transitory that it is difficult to determine its exact quality. Therefore, it is essential for speakers of English, which has phonemic diphthongs, to recognise that both vowels in VV sequences in Japanese are in different morae as the mora is a rhythmic unit in Japanese and each mora is produced largely at the same timing interval. There would be less effort to learn to produce VV sequences of Japanese with the qualities that Japanese speakers produce because Japanese speakers produce the vowels with the reduced qualities from the corresponding singletons when they are in VV sequences.

6.5 Limitations of the present study

Although the present study aimed to investigate a full range of VV sequences of a full range of accent contours, to fill gaps in the existing research literature, VV sequences of certain morphological conditions and speech styles were unable to cover the full range of the accent contour. Also, the recordings for VV sequences across the morphological conditions and the speech styles were held at different times, therefore, the numbers of participants were different between the two sessions and the speakers for the read speech (WI) and other spontaneous styles were different. Moreover, the speech rate was not controlled for the productions and it may influence the acoustic properties.

While analysing recordings for the present study, it was often the case that background noise rendered the analysis difficulty to interpret the acoustic properties. It could be dealt with by adjusting the dynamic range setting of Praat for some tokens but a number of tokens had to be omitted from the analysis for this reason. Although it was ideal to conduct the recordings in the laboratory condition for better sound quality, there is a risk that participants may produce speech unnaturally. The quiet but usual circumstance would make participants more relaxed and their speech is expected to be normal, casual and spontaneous. The balance between obtaining better quality and obtaining more natural speeches should be taken into consideration.

6.6 Future research

Taking the findings and the limitations for the present study into consideration, the intrapersonal differences between the controlled speech and the spontaneous speech could be investigated. Also, it is still uncertain whether the ‘semantic boundary’ that Kawakami (1977) and Saito (1997) referred to is the morphological boundary that Gore (2006) adopted and the word boundary that the present study adopted. It is assumed to be the unity of VV sequences; therefore, there may be some other conditions in which the realisations of VV sequences are clearly distinguished. Ultimately, it is desirable to find a crucial difference for VV sequences in Japanese that are realised as more diphthong-like or as hiatuses.

Chapter 7. Conclusion

The present research set out to investigate the vowel sequences of Japanese considering the full range of possible sequences and with a full range of accent contours. The first research question was to investigate the acoustic properties of the vowel sequences in Japanese across a range of different phonological conditions. VV sequences in Japanese were found consistently to have steady states for both vowels and a transition between the vowels. Also, the vowel qualities of the vowels in VV sequences are significantly reduced from the corresponding singletons.

Secondly, the extent to which the acoustic realisations of VV sequences vary as a function of the degree of cohesion between the two vowels in the sequence and as a function of different speech styles is considered. The morphological conditions and the speech style conditions are found to be significant factors in respect of variability in the realisations of VV sequences. The ratio for each interval greatly varied across the morphological conditions of VV sequences and the speech styles within which VV sequences are produced and the degrees of reduction are variable depending on the morphological condition, the speech style condition and the vowel identity.

Thirdly, the debate around whether Japanese VV sequences can be classified as diphthongs was considered. From a phonological point of view, it seems certain that VV sequences in Japanese are not diphthongs. This is because of the syllabicity of diphthongs and the fact that VV sequences are made up by two distinct vowels and each vowel occupies one mora and syllable. Therefore, VV sequences occupy two morae and two syllables. However, from the phonetic point of view, it is not straightforward to distinguish the realisations of VV sequences from diphthongs as these have been described in other languages. Whether VV sequences in Japanese are diphthongs or not depends on the point of reference. However, there are certain circumstances in which VV sequences do appear to be realised as more diphthong-like, having a longer transition ratio and reduced vowel quality.

The present research provides new findings relating to the acoustic properties of VV sequences of Japanese from a comprehensive data-set. However, there is still a need for much further work to be done across-languages on the realisation of diphthongs and vowel sequences in order to explore these issues in a better-informed manner.

References

Abercrombie, D. (1967) *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Aguilar, L. (1999) 'Hiatus and diphthong: Acoustic cues and speech situation differences', *Speech Communication,* 28(1), pp. 57-74.

Alam, F., Murtoza Habib, S.M. and Khan, M. (2008) 'Acoustic analysis of Bangla vowel inventory', *BRAC University CRBLP Technical report*.

Ashby, M. and Maidment, J. (2005) *Introducing phonetic sciences* Cambridge Cambridge University Press.

Asher, R.E. and Keane, E.L. (2005) 'Diphthongs in colloquial Tamil', in Hardcastle, W.J. and Beck, J.M. (eds.) *A Figure of Speech: a Festschrift for John Laver*. Mahwah, New Jersey: Laurence Erlbaum Associates, pp. 141–171.

Asu, E.L., Lippus, P., Niit, E. and Türk, H. (2012) 'The Acoustic Characteristics of Monophthongs and Diphthongs in the Kihnu Variety of Estonian', *Linguistica Uralica*, 48(3), pp. 161-170.

Cabre´, T. and Prieto, P. (2007) 'Exceptional hiatuses in Spanish', in Colina, F.M.n.-G.S. (ed.) *Optimality-theoretic studies in Spanish phonology*. Amsterdam & Philadelphia: Benjamins, pp. 205–238.

Carlyle, K. (1987) 'Vowel sequence and sonority in Breton', *Toronto Working Papers in Linguistics* 8, pp. 20-41.

Catford, J.C. (1977) *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Catford, J.C. (2001) *A practical introduction to phonetics*. 2nd edn. Oxford: Oxford University Press.

Cheung, Y.M. (2007) 'An Acoustical Analysis of the Vowels, Diphthongs and Triphthongs in Hakka Chinese ', *Proceedings of International Congress of Phonetic Sciences 2007*, pp. 841-844.

Chitoran, I. and Hualde, J.I. (2007) 'From hiatus to diphthong: the evolution of vowel sequences in Romance', *Phonology*, 24(1), pp. 37-75.

Clark, J. and Yallop, C. (1995) *An introduction to phonetics and phonology*. Oxford: Blackwell.

Clermont, F. (1992) 'Characterisation Of The Diphthongal Sound Beyond The F1-F2 Plane', *Proceedings for 4th Australasian International Conference on Speech Science and Technology*. Brisbane. pp. 298-304.

Cox, F. (2006) 'The Acoustic Characteristics of /hVd/ Vowels in the Speech of some Australian Teenagers', *Australian Journal of Linguistics*, 26(2), pp. 147-179.

Cruttenden, A. (2001) *Gimson's pronunciation of English*. 6th edn. London: Arnold.

Denes, P.B. and Pinson, E.N. (1993) *The speech chain; the physics and biology of spoken language* New York: W. H. Freeman.

Deterding, D. (1996) ‘Diphthong Measurements In Singapore English’ *Proceedings for 6th Australian International Conference on Speech Science and Technology*. Adelaide. pp. 61-66.

“diphthong, n.” *OED Online*. June 2004. Oxford University Press. 13 April 2013 <http://dictionary.oed.com/>.

Face, T.L. and Alvord, S.M. (2004) 'Lexical and acoustic factors in the perception of the Spanish diphthongs vs. hiatus contrast', *Hispania*, 87, pp. 553–564.

Fujisaki, H. and Higuchi, N. (1979) 'Temporal Organization of Segmental Features in Japanese Disyllables', *Annual Bulletin Research Institute of Logopedics and Phonitrics, Faculty of Medicine, University of Tokyo*, 13.

Garrido, M. (2007) 'Diphthongization of Mid/Low Vowel Sequences in Colombian Spanish', in al, J.H.e. (ed.) *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics*. Somerville, MA: Cascadilla Proceedings Project, pp. 30-37.

Gay, T. (1968) 'Effect of speaking rate on diphthong formant movements', *The Journal of the Acoustical Society of America*, 44(6), pp. 1570-1573.

Gore, M. (2006) *English and Japanese Diphthongs and Vowel Sequences*. Thesis (PhD) Reading: University of Reading.

Grissom, R.J. and Kim, J.J. (2005) *Effect sizes for research : a broad practical approach*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Harrington, J. and Cassidy, S. (1994) 'Dynamic and Target Theories of Vowel Classification: Evidence from Monophthongs and Diphthongs in Australian English', *Language and Speech*, 37(4), pp. 357-373.

Hansen, J.H.L. and Arslan, L. M. (1995) 'Foreign accent classification using source generator based prosodic features', *Proceedings of The International Conference on Acoustics, Speech, and Signal Processing '95*, pp. 836-839.

Hattori, S. (1967) *Phonetics [Onseigaku]*. Tokyo: Iwanami Shoten.

Hiraiwa, T. (1981) 'The metholology of the speech education of English', *Bulletin of Aichi University of Education (Educational Sciences)*, 30, pp. 1-8.

Hirasaka, F. and Kamata, S. (1981) 'English and Japanese diphthongs: An acoustic approach', *Sophia Linguistica*, 8/9, pp. 183–195.

Holbrook, A. and Fairbanks, G. (1962) 'Diphthong Formants and their Movements', *Journal of Speech and Hearing Research* 5, pp. 38-58.

Hualde, J.I. and Prieto, M. (2002) 'On the diphthong/hiatus contrast in Spanish: some experimental results', *Linguistics*, 40(2), pp. 217–234.

Inozuka, E. and Inozuka, H. (2003) *Introduction to Japanese Speech sounds - Discussion and Exercises- [Nihongo no Onsei Nyumon]*. Tokyo: Babel Press.

Jha, S.K. (1985) 'Acoustic analysis of the Maithili diphthongs', *Journal of Phonetics*, 13, pp. 107-115.

Jones, D. (1972) *An outline of English phonetics* 9th edn. Cambridge, New York: Cambridge University Press.

Kashima, T. (2002) *Basic Phonetics for teaching Japanese [Nihongo Kyouiku wo mezasu hito no tame no kisokara manabu Onseigaku]*. Tokyo: Three A network.

Kasuya, H. and Sato, S. (1990) 'Cho-on, Nirenboin to Nijyuboin - Nihongowasha to Eigowasha no Baai- [Prolonged sound, Two sequential Vowel and Diphthong -In the cases of Japanese speakers and English speakers-]', in Sugito, M. (ed.) *Kouza Nihongo to Nihongo Kyoiku [Lectures on Japanese and Japanese Teaching]*. Tokyo: Meiji Shoin.

Kawakami, S. (1977) *An Outline of Japanese Speech Sounds [Nihongo Onsei Gaisetsu]*. Tokyo: Ohfu.

Kent, R.D. and Read, C. (2002) *The acoustic analysis of speech*. 2nd edn. London: Singular.

Kitagawa, A. (2012) 'Acoustic characteristics of English diphthongs produced by Acoustic characteristics of English diphthongs produced by Japanese learners', *Proceedings of The 17th Conference of Pan-Pacific Association of Applied Linguistics*, pp. 99-100.

Koizumi, T. (2003) *Kaitei Onseigaku Nyumon [An Introduction to phonetics revised]*. Tokyo: Daigaku Shorin.

Kotrlik, J.W. and Williams, H.A. (2003) 'The Incorporation of Effect Size in Information Technology, Learning, and Performance Research', *Information Technology, Learning & Performance Journal;Spring2003*, 21(1), pp. 1-7.

Kubozono, H. (1999) *Nihongo no Onsei [Speech sounds of Japanese]*. Tokyo: Iwanami Shoten.

Kubozono, H. (2001) 'On the Markedness of Diphthongs', *Kobe Papers In Linguistics* 3, pp. 60-73.

Labrune, L. (2012) *The phonology of Japanese*. Oxford: Oxford University Press.

Ladefoged, P. (2006) *A course in phonetics*. Boston: Thomson/Wadsworth.

Laver, J. (1994) *Principles of phonetics*. Cambridge, New York: Cambridge University Press.

Lee, W.S. (2010) ‘Spectral analysis of the diphthongs in Peking dialect’, *Proceedings of the 9th Phonetics Conference of China*. Tianjin, China.

Lehiste, I. and Peterson, G.E. (1961) 'Transitions, Glides, and Diphthongs', *Journal of the Acoustical Society of America*, 33(3), pp. 268-277.

Lindau, M. (1985) 'Hausa vowels and dipthongs', *Studies in African Linguistics*, 16(2), pp. 161-182.

Lindau, M., Norlin, K. and Svantesson, J.-O. (1990) 'Some cross-linguistic differences in diphthongs', *Journal of the International Phonetic Association*, 20(1), pp. 10-14.

Llisterri, J. (1992) 'Speaking styles in speech research', *ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language*. Dublin, Ireland, 15-17 July 1992.

Manrique, A.M.B.d. (1979) 'Acoustic Analysis of the Spanish Diphthongs', *Phonetica*, 36, pp. 194-206.

Mauder, E. and van Heuven, V.J.J.P. (1996) 'On the rise and fall of Spanish diphthongs ', *Linguistics in the Netherlands 1996*, pp. 171 - 182.

Maxwell, O. and Fletcher, J. (2010) 'The acoustic characteristics of diphthongs in Indian English', *World Englishes*, 29(1), pp. 27-44.

Misono, K. and Hirasaka, F. (2008) 'Diphthongs and vowel sequences', *Bulletin of the College of Humanities, Kanto Gakuin University*, 113, pp. 87-103.

Nakagawa, K. (1996) 'Towards the acquisition of Japanese prosody for English native speakers', *Hokkai Gakuen University studies in culture* 7, pp. 95-118.

Nakamura, A., Suzuki, M., Minematsu, N., Hirose, K. and Makino, T. (2010) 'An experimental study on assessment of diphthongs of learners using pronunciation structure', *Proceedings of the 2010 Spring Meeting of the Acoustical Society of Japan*, 1-P-15, pp. 439-442.

Niwa, Y. (1956) 'A report on the English pronunciation of young students -mainly on the influence of Japanese-', *Bulletin of Nagoya University School of Education Affiliated Upper and Lower Secondary School*, 2, pp. 84-88.

O’Connor, J.D. (1973) *Phonetics*. Harmondsworth: Penguin.

Peng, G. (2007) ‘An Experimental Investigation of the Inter-relationship between the Diphthong and the tone in Fuzhou Chinese’ *Proceedings of 16th International Congress of Phonetic Sciences*. Saarbrücken. pp. 1301-1304.

Penny, L. (1992) ‘Acoustic Measurements of the Diphthongs of Women Speakers of General Australian English’, *Proceedings of 4th Australasian International Conference on Speech Science and Technology*. Brisbane, pp. 489-494.

Peterson, G.E. and Barney, H.L. (1952) 'Control methods used in a study of the vowels', *Journal of the Acoustical Society of America* 24, pp. 175-184.

Roach, P. (1983) *English phonetics and phonology: a practical course*. 3rd edn. Cambridge: Cambridge University Press.

Roengpitya, R. (2002) ’Different Durations of Diphthongs in Thai: a New Finding’ *Proceedings of the Twenty-Eighth Annual Meeting of the Berkeley Linguistics Society: Special Session on Tibeto-Burman and Southeast Asian Linguistics (2002),* pp. 43-54.

Roengpitya, R. (2007) 'A new look on diphthongs in Thai', *SEALS XIII Papers from the 13th Annual Meeting of the Southeast Asian Linguistics Society 2003*. Canberra, Australia. I. Shoichi et al, pp. 231-237.

Saito, Y. (1997) *An Introduction to Japanese Phonetics [Nihongo Onseigaku Nyumon]*. Tokyo: Sanseido.

Sarwar, A., Ahmed, S. and Tarar, A.A. (2004) 'Diphthongs in Urdu Language and Analysis of their Acoustic Properties', *CRULP Annual Student Report, 2003-2004*, pp. 9-14.

Shibatani, M. (1989) *The Languages of Japan* Cambridge ; New York: Cambridge University Press.

Shimizu, A. (2011) 'English as a Lingua Franca and the Teaching of Pronunciation', *Journal of the Phonetic Society of Japan*, 15(1), pp. 44-62.

Shimizu, K. (2008) 'L2 speech learning and theoretical implications', *Journal of Nagoya Gakuin University Language and culture*, 19(2), pp. 81-87.

Shimoda, H., Narahara, Y. and Tetsuya, O. (1973) 'A Contrastive Study of Some English and German Diphthongs', *Reports of the University of Electro-Communications* 24(2), pp. 319-324.

Simpson, A.P. (1998) 'Characterizing the formant movements of German diphthongs in spontaneous speech', in Lenders, W., Hess, W. and Portele, T. (eds.) *Computer, Linguistik und Phonetik zwischen Sprache und Sprechen, Tagungsband der 4. Konferenz zur Verarbeitung natürlicher Sprache - KONVENS-98*. Frankfurt. : Lang, pp. 192-200.

Svantesson, J.-O. (1984) 'Vowels and diphthongs in Standard Chinese', *Lund University, Department of Linguistics Working Papers*, 27, pp. 209-227.

Takayama, T. (2003) 'Gendai Nihongo no Onin to sono Kinou [Phonology of modern Japanese and its function] ', in Mitahara, Y. (ed.) *Onsei, Onin [Phonetics and Phonology]*. Tokyo: Asakura Shoten.

Takayama, T. (2010) 'Where did the "Mu-akusento hoogen" (the dialects without accentual distinction) come from?', *Studies in literature , Faculty of Humanities, Kyushu University*, 107. pp. 1-15.

Takubo, Y., Maekawa, K., Haruo, K., Honda, K., Shirai, K. and Nakagawa, S. (1998) *Onsei [Speech Sounds]*. Tokyo: Iwanami Shoten.

Tamaoka, K. and Makioka, S. (2004) 'Frequency of occurrence for units of phonemes,morae, and syllables appearing in a lexical corpus of a Japanese newspaper', *Behavior Research Methods, Instruments, & Computers 2004*, 36(3), pp. 531-547.

Tasko, S.M. and Greilick, K. (2010) 'Acoustic and articulatory features of diphthong production: a speech clarity study', *Journal of Speech, Language and Hearing Research*, 53(1), pp. 84-99.

Teshima, M. (2011) 'On Teaching English Pronunciation in Japanese Secondary Education : How It Is and How It Should Be', *Journal of the Phonetic Society of Japan*, 15(1), pp. 31-43.

Tsujimura, N. (2007) *An introduction to Japanese linguistics*. 2nd edn. Malden, MA: Blackwell Pub.

Tsukada, K. (2008) 'An Acoustic comparison of English monophthongs and diphthongs produced by Australian and Thai speakers', *English world-wide*, 29(2), pp. 194-211.

Tsukuma, Y. (2005) 'English pronunciation training for Japanese students in the beginner's class ', *Language and its influence*, 3, pp. 163-200.

Uchida, H. (2008) 'A benefit of pronunciation training: how orally practicing unfamiliar phonemes can positively affect aural identification of these phonemes ', *Bulletin of Tottori University of Environmental Studies* 6, pp. 39-48.

van Heuven, V.J., Edelman, L. and van Bezooijen, R. (2002) 'The pronunciation of /ɛi/ by male and female speakers of avant-garde Dutch', *Linguistics in the Netherlands 2002* pp. 61–72.

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

Vance, T.J. (2008) *The Sounds of Japanese*. New York: Cambridge University Press.

Weil, K.S., Fitch, J.L. and Wolfe, V.I. (2000) 'Diphthong changes in style shifting from Southern English to Standard American English', *Journal of Communication Disorders*, 33(2), pp. 151-163.

Yamaguchi, R. (2008) 'A method of teaching English pronunciation thorough Japanese phonetics', *Language and culture: Bulletin of Institute for Language Education, Aichi University* 19, pp. 119-131.

Yu, J., Li, A. and Wang, X. (2004) 'A Contrastive Investigation of Diphthongs between Standard Mandarin and Shanghai Accented Mandarin', *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*. Beijing, China. pp. 113-117.

Appendices

Appendix 1 List of words for VV sequences within real words

1. VVs with High-High accent contour

VVs	pronunciation	meaning
ai	[wadaɪ]	topic, theme
au	[ʦɯkaɯ]	to use
ae	[sakaeɾɯ]	be prosperous
ao	[hataoɾɪ]	weaver
ia	[taʧɪaɡaɾɯ]	to stand up
iu	[ɯmɪɯʃɪ]	a nudibranch
ie	[maɾɯmɪe]	apparent
io	[ɯʧɪotosɯ]	to shoot
ua	[ʥokɯakɯ]	vulgar
ui	[kakɯɪʦɯ]	standardized
ue	[ʦɯkɯe]	desk
uo	[kaʦɯo]	bonito
ea	[oteaɾaɪ]	toilet
ei	[ɯkeɪɾe]	acceptance
eu	[ateɯma]	teaser
eo	[sɯteɾeo]	stereo
oa	[ɪɾoaseɾɯ]	to bleach
oi	[omoi]	heavy
ou	[ʤɪdoɯ]	automatic
oe	[ʤɪɡoe]	natural voice

2. VVs with High-Low accent contour

VVs	pronunciation	meaning
ai	[takaɪ]	high
au	[kaɯ]	to keep (a pet)
ae	[kaeɾɯ]	to go back
ao	[egao]	smile
ia	[ɸɯʃɪawase]	unhappy
iu	[çɪɯɴ]	misfortune
ie	[kɪe]	embrace
io	[juːkɪoɴ]	aspirated sound
ua	[sɯaʃɪ]	barefoot
ui	[kɯɪkɪ]	area
ue	[kɯekɪ]	drudgery
uo	[gakɯoɴ]	musical tone
ea	[kea]	care
ei	[meɪɴ]	main
eu	[ʥeɯsɯ]	Zeus
eo	[neoɴ]	neon
oa	[koa]	core
oi	[omoi]	thoughts
ou	[noɯɾɪ]	in mind
oe	[koe]	voice

3. VVs with Low-High accent contour

VVs	pronunciation	meaning
ai	[kaɪdaɴ]	stairs
au	[kaɯ]	to buy
ae	[kaeɾɯ]	to change
ao	[taosɯ]	to defeat
ia	[kɪaʦɯ]	air pressure
iu	[ʃɪɯʧɪ]	treatment
ie	[mɪeɾɯ]	to see
io	[kɪokɯ]	memory
ua	[ɸɯaɴ]	anxiety
ui	[ʦɯɪka]	add
ue	[sɯe]	the end
uo	[mɯoɴ]	silence
ea	[keana]	pore
ei	[neɪkɪ]	sleep-breath
eu	[teɯʧɪ]	hand-made
eo	[meoto]	married couple
oa	[soakɯ]	inferiority
oi	[koiʃɪ]	pebble
ou	[koɯʃɪ]	calf
oe	[koeda]	twig

4. VVs with Low-Low accent contour

VVs	pronunciation	meaning
ai	[çɪgaɪ]	damage
au	[kokiʦɯkaɯ]	to overwork
ae	[soːɪɾekae]	reshuffleing
ao	[sɯnao]	obedient
ia	[sɯɾobakɪa]	Slovakia
iu	[kaɾɪɯdo]	hunter
ie	[mɯnɪeɾɯ]	meuniere
io	[toɾɪo]	trio
ua	[teɪkeʦɯaʦɯ]	low blood pressure
ui	[neʦɯɪ]	enthusiasm
ue	[ʤɪmbabɯe]	Zimbabwe
uo	[managaʦɯo]	butter fish
ea	[aɪdea]	idea
ei	[teɪneɪ]	careful
eu	[moʔteɯmaɾeta]	innate
eo	[paɾeo]	pareo
oa	[kokoa]	cocoa
oi	[itagakoi]	boarding
ou	[sahoɯ]	manner
oe	[monooboe]	memory

Appendix 2 List of family names and given names for VV sequences across word boundaries

1. List of family names and given names

Family names	Given names
Nakamura	Akiko
Musanagi	Ikuko
Nakamaru	Utako
Watanabe	Emiko
Yamamoto	Okiko

2. List of names presented to participants

VVs across WB	Names
a#i	Nakamura Ikuko
a#u	Nakamura Utako
a#e	Nakamura Emiko
a#o	Nakamura Okiko
i#a	Musanagi Akiko
i#u	Musanagi Utako
i#e	Musanagi Emiko
i#o	Musanagi Okiko
u#a	Nakamaru Akiko
u#i	Nakamaru Ikuko
u#e	Nakamaru Emiko
u#o	Nakamaru Okiko
e#a	Watanabe Akiko
e#i	Watanabe Ikuko
e#u	Watanabe Utako
e#o	Watanabe Okiko
o#a	Yamamoto Akiko
o#i	Yamamoto Ikuko
o#u	Yamamoto Utako
o#e	Yamamoto Emiko

Appendix 3 List of target words and corresponding questions for the VV sequences in the response style

VV	pronunciation	meaning	Questions to ask
ai	[ʃɪkaɪ]	a chairperson	Who does take charge of the meeting?
au	[kaɯ]	to buy	What is the opposite word of “sell”?
ae	[kaesɯ]	to return	What should you do after you borrow a book from a library?
ao	[taoɾeɾɯ]	to be fallen	What happens to a tree when it is cut down?
ia	[aʃɪato]	a footprint	What appears on the path when a person walk there?
iu	[mɪɯʧɪ]	a relative	What is a collective term of uncles, aunts and cousins?
ie	[kɪeɾɯ]	to disappear	What is the opposite word of “appear”?
io	[kakɪokɪ]	a message left behind	What do you leave when you run away from home?
ua	[gokɯakɯ]	outrageous	How do you describe a murder?
ui	[kɯɪʃɪmboɯ]	a glutton	What do you call a person who really likes to eat?
ue	[ʦɯe]	a walking stick	What do old people use when they walk outside?
uo	[ɾokɯoɴ]	a recording	What do you do with a cassette tape?
ea	[kakeaʃɪ]	to run	What do you do when you in a hurry?
ei	[keɪto]	woollen yarn	What do you knit for a sweater?
eu	[meɯe]	one’s elders	What is it called for an elderly or superior person?
eo	[bɪdeo]	a video recorder	What do you use when you record the TV programmes?
oa	[kokoa]	hot chocolate	What is the name of a drink made of chocolate?
oi	[ototoɪ]	the day before yesterday	What is the day before yesterday?
ou	[koɯʃɪ]	a calf	What are young cows called?
oe	[koe]	a voice	What is the sound for communication?

Appendix 4 List of pictures presented to participants for VVs in the description style

/ai/
Kokkaigijidou
(parliament house)

/au/
chigau
(different)

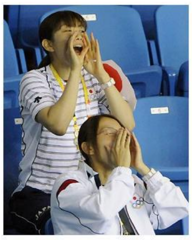
/ae/
kaesu
(to return)

/iu/
Oshiuri
(pusher)

/ui/
muite-iru
(peeling)

/eu/
teuchi-soba
(handmade buckwheat noodles)

/oi/
Koinobori
(carp pennants)



/oe/
Koe
(voice)

/ei/
Keisatsu
(policeperson)

/ou/
Koukou Yakyu (the senior high school baseball tournament)
Koushien (the stadium it takes place)

These pictures were cited from various websites. The following picture files are still online but others might be deleted.

/ae/: http://illpop.com/img_illust/season/apr03_a16.png

/oi/: http://ryubun.net/photo/2009/04/090428-0233.jpg

/oe/: http://beijing2008.nikkansports.com/badminton/photo/20080814-65447.html

/ei/: http://www.police.pref.kanagawa.jp/ps/74ps/74pic/74002_27.gif

/ou/: http://www.asahi.com/koshien2004/news/images/TKY200408220180.jpg

Appendix 5 Consent forms (Japanese and English)

実験参加同意書

題名:日本語話者による母音の発音

実験者:原　功

あなたは英国ニューカッスル大学のG.J ドカティ教授の指導の下で行われている原功の学位申請論文の一部となる実験に参加していただくことになりました。下記の情報をお読みいただき、参加の可否をお決めください。

1. 実験の目的

この実験の目的は、標準的な日本語を話す人による日本語と英語の発音を調べることにあります。録音された発音を分析することにより、日本語と英語の音声学的な特徴を比較するための基本的な情報が得られます。今回お願いしたのは、あなたが日本語の標準語が話される地域で生育し、標準語話者であると考えられるからです。

2. 実験の手続き

あなたはこれより、カードに印刷された単語や文を読み上げていただくことになります。全部で５つの部分から成り、終了までおよそ１時間かかります。あなたの声はデジタル録音機によって録音され、コンピュータ分析されることになります。また、あなたの住居地や英語力に関する質問にもお答えいただきます。

3. 謝礼について

この実験に参加していただいた方には粗品をご用意しております。なお、実験への参加を中断した方にも差し上げています。

4. 機密保持について

この調査において得られる、あなたを特定できる情報はすべて厳正に取り扱います。学位論文、雑誌論文、学会発表において結果を公表する際には、個人を特定できる情報は全て排除されま

す。

5. 拒否権について

この実験への参加は完全に任意であり、実験に参加しないことも、また途中で参加を取りやめることも自由です。また、その場合に取りやめたことを記録に残すことはありません。

6. 問い合わせ

この実験に関するお問い合わせは、以下のメールアドレスにて実験者または指導教官にお願いします。

原　功: isao.hara@newcastle.ac.uk
G.J ドカティ教授: G.J.Docherty@newcastle.ac.uk

7. 参加への同意

私は、上に書かれたことに関して十分に理解し、この実験に参加することに同意します。また、この用紙の控えを受け取りました。

________________________　____________
(参加者署名)　(日付)

________________________　____________
(実験者署名)　(日付)

Informed Consent Form for Project Participants

Project title: Pronunciations of Vowels by Japanese Speakers
Investigator: Isao Hara

You are invited to take part in a research project conducted by Isao Hara as a part of his doctoral thesis under the supervision of Professor G.J Docherty(Newcastle University, UK). The following information is provided to help you make an informed decision about whether or not to participate in this project.

1. Purpose of the study

The aim of this project is to investigate the pronunciation of Japanese and English by Native Japanese speakers of standard dialect. Analysis of speech recordings will provide the basis for a comparison of the phonetic characteristics of Japanese sounds and English sounds. You are invited to participate in this research because you are born and brought up the area where Tokyo dialect (Standard Japanese) is spoken and recognised as a speaker of Standard Japanese.

2. Procedure

You will be asked to read out words or sentences printed on a card. The reading task consists of 5 parts and it will take approximately one hour. Your reading will be recorded on the hard disk of a digital recorder and subsequently analysed by computer. You will also be asked to fill out a questionnaire relating to your place of residence and your proficiency in English.

3. Benefits and Compensation

There is a small gift to thank you for your time. You will receive it even if you withdraw from participation.

4. Confidentiality

Information obtained during this research which could identify you will be kept strictly confidential. The summarized findings with no identifying information may be published in the thesis or an academic journal and presented at a scholarly conference.

5. Right to refuse or withdraw

Your participation is completely voluntary. You are free to withdraw from participation at any time during the recording or thereafter or to choose not to take part in the research at all. There will no consequence for you of any description if you decide to do so.

6. Questions

If you have any questions about this research or desire information in the future regarding your participation, you can contact Isao Hara or Professor G.J Docherty at the following addresses.

Isao Hara: isao.hara@newcastle.ac.uk
Professor GJ Docherty: G.J.Docherty@newcastle.ac.uk

7. Agreement

I am fully aware of the nature and extent of my participation in this project as stated above. I hereby agree to participate in this project. I acknowledge that I have received a copy of this consent statement.

________________________ ____________

(Signature of Participant) (Date)

________________________ ____________

(Signiture of Investigator) (Date)