View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Hiroshima University Institutional Repository

*Glottometrics 8, 2004, 1-11*

# Are the effects of vowel repetition influenced by frequencies? A corpus study on CVCVCV-structured nouns with and without vowel repetition

*Katsuo Tamaoka, Hiroshima University, Japan*[1]
*Shogo Makioka, Osaka Women's University, Japan*
*Tadao Murata, Kyushu Institute of Technology, Japan*

**Abstract.** A psychological study by Tamaoka and Murata (2001) suggested that CVCVCV-structured nonwords (e.g., /kohomo/) with the same vowel repeated showed longer naming latencies than the same-structured nonwords without vowel repetition (e.g., /kohami/). One of the possible factors for prolonging vowel repetition could be the frequency of vowel repetition in Japanese. Thus, the present study calculated token frequencies for nouns with the same vowel repeated within a CVCVCV phonological structure, based on the Japanese lexical corpus (287,792,797 words) of Amano and Kondo (2000). The results showed that vowels were repeated among Japanese nouns with a CVCVCV string more frequently than the random possibility of 4 percent. In addition, nouns with the same vowels in the first and second positions (i.e., $V_1$ and $V_2$ in the $CV_1CV_2CV_3$) showed significantly higher occurrences than the random chance of 20 percent, whereas nouns with the same vowels in the second and third positions appeared at the random level (i.e., $V_2$ and $V_3$). Since it is expected that higher frequency enhances speed and accuracy in naming, phonological structures with the same vowel repeated can be expected to be more quickly and accurately named. Conflicting results between the present corpus study and the experimental study by Tamaoka and Murata (2001) excluded the possibility of the frequency of vowel repetition affecting the speed and accuracy of phonological processing.

*Keywords: vowel repetition, phonological structure, corpus study, Japanese nouns*

## 1. Introduction

A study by Tamaoka and Murata (2001) suggested that CVCVCV-structured nonwords with the same vowel repeated like /kohomo/ showed longer naming latencies than the same-structured nonwords without vowel repetition like /kohami/. The explanation proposed for this is the 'whack-a-mole' phenomenon. The vowel in the first CV mora (C referring to 'consonant' and V 'vowel') continues to have a high activation level even when the following CV morae are activated. When the same vowel is repeated throughout the CV morae, all the CV morae will be simultaneously excited to reach the activation level. To avoid confusing the continuous order of the CV mora string, sequential morae must be inhibited so as not to be activated to the same degree as the previous CV mora. This pattern of excitation and inhibition results in the decreased speed of phonological processing for nonwords. As for naming nonwords with no repeated vow-

[1] Address correspondence to: Katsuo Tamaoka, International Student Center, Hiroshima University, 1-1, 1-Chome, Higashihiroshima, Japan 739-8524. E-mail: ktamaoka@hiroshima-u.ac.jp

els, the processing of sequential order of the CV mora string is not affected by other morae. Thus, nonwords with varying vowels are named more quickly than nonwords which repeat vowels and, concomitantly, fewer errors are observed among nonwords with non-repeated vowels.

While the 'whack-a-mole' phenomenon was a psychological explanation, some linguists provide a different explanation from a phonological perspective. The Obligatory Contour Principle (OCP) refers to a linguistic constraint on similar or same phonological features from being repeated (e.g., Fukazawa, 2000; Ito & Mester, 1986; Kubozono, 1999; Kubozono & Ota, 1998; Leben, 1973; McCarthy 1986; Yip, 1988). Kubozono and Ota (1998) suggested the possibility that vowel dissimilation in Japanese may be a result of the OCP. For example, the two Japanese morphemes /nana/ ('seven') and /ka/ ('day') combine to form the compound word /nanoka/ ('the seventh day') instead of /nanaka/, which would seem to be the likely combination. This process of vowel dissimilation occurs so as to avoid vowel repetition of /a/ in sequence within the three mora CVCVCV word structure. Thus, it would be expected that naming visually-presented Japanese words and nonwords which violate the OCP (i.e., same vowel repetition in a series of CV strings) would result in slower processing speeds and higher error rates. Yet, the linguistic explanation of OCP does not conflict with the psychological explanation of the 'whack-a-mole' phenomenon.

Despite these psychological and linguistic explanations, 'frequency of vowel repetition' in Japanese could be a possible factor for prolonging vowel repetition. Therefore, the present study calculated type and token frequencies for nouns with the same vowel repeated within a CVCVCV phonological structure, based on the Japanese lexical corpus (287,792,797 words) of Amano and Kondo (2000). Study 1 examined the same vowels in all three consecutive V positions and Study 2 in two consecutive V positions. Both Studies 1 and 2 were used to examine whether the rate of vowel repetition in CVCVCV-structured Japanese nouns appears to be significantly greater or lesser than the random chance rate of occurrence.

## 2. Conditions for calculating word frequency

Three conditions were established for calculating word frequency. First, the random chance rate of occurrence was established using only CVCVCV-structured words. Under this condition, the possibility of a 3-mora CVCVCV-structured word, which contains the same vowel in three consecutive positions, is calculated as 4 percent (with the 5 different Japanese vowels of /a/, /e/, /i/, /o/ and /u/ in three positions calculated as $1/5^3 \times 5 = 1/5^2$). In the same way, the possibility of these words with the same vowel in chosen two positions in their CVCVCV structure is 20 percent (with the 5 different vowels in two chosen positions calculated as $1/(5\times5) \times 5 = 1/5$, the third vowel may be arbitrary).

Secondly, only nouns were selected from the word corpus of Amano and Kondo (2000), which still served as sufficient data for the purpose of the present investigation. As Japanese verbs and adjectives have grammatical inflections, they were not included in the present corpus study. For example, the Japanese verb /ugoku/, meaning 'to move', inflects as in /ugok<u>a</u>(nai)/, /ugok<u>i</u>(masu)/, /ugok<u>u</u>(toki)/, /ugok<u>e</u>(ba)/ and /ugok<u>o</u>R/. It uses all the five Japanese vowels of /a/, /i/, /u/, /e/ and /o/ in its grammatical inflections.

Thirdly, both frequencies of *type* and *token* were used for the purpose of this study. In type frequency, a single word is only counted once, regardless of how many times it is repeated in the written text. On the other hand, in token frequency (i.e., accumulative word frequency) a word is counted every time it appears in the text. Since a rare word (e.g., /guNzoR/ meaning 'ultra-marine') has the same type frequency of 1 as a frequently used word (e.g., /daigaku/ meaning

'university'), the present study considered token frequency as a better indicator of word frequency.

## 3. STUDY #1: Frequency of CVCVCV-structured nouns with the same vowels occurring in three consecutive vowel positions

Study 1 examined the occurrence of the same vowels in three consecutive V positions within CVCVCV-structured nouns. In this study, all nouns with a $CV_1CV_2CV_3$ structure had to share the same vowel in all $V_1$, $V_2$ and $V_3$ positions.

### 3.1. Lexical Corpus and Selection Procedure

As a result of the word frequency index created by Amano and Kondo (2000) from their study on accumulative word frequency (i.e., token frequency), a very large lexical corpus of 341,771 words was established from newspapers containing 287,792,797 words of accumulate frequency. All these words were taken from the *Asahi Newspaper* printed from 1985 to 1998. This is one of the largest and the most up-to-date word corpora created from calculating frequency of words in Japanese written texts. The present study utilized this corpus to investigate nouns with vowel repetition.

The programming language of MacJPerl 5.15r4J for Macintosh was used to run a calculation procedure. For Study 1, only nouns with a CVCVCV phonological structure were used. Thus, three mora nouns with VCVCV, CVVCV, CVCVV, VCVV, or VVV strings were not included. Therefore, the Japanese special long vowel /R/[1], where the same vowel appears twice without having a consonant between them, was not counted. In the same way, double vowels such as /ai/, /oi/, /ue/ were also excluded from the count. Other special sounds such as the nasal /N/ and the geminate /Q/ were also excluded as well as contrasted sounds such as /kya/, /myo/, /pyo/.

### 3.2. Results

Study 1 used two types of frequencies: *type* and *token* frequency. Type frequency only counts a word once and then is calculated by a simple addition of each word's frequency of '1' ($\Sigma N_i$), even though a word may appear repeatedly in a printed text. Token frequency, on the other hand, is calculated by taking the number of times each word appears in the text and adding all these frequencies together ($\Sigma W_f$). The present study took the .01 level of significance to reject the statistical null hypothesis, since the word corpus used was very large. These two frequency indexes of type and token frequency are reported separately as listed below.

#### 3.2.1. Type Frequency

As shown in Table 1, among the five Japanese vowels, the vowel /a/ was the most frequently repeated in three consecutive positions, found in 674 nouns or 61.50 percent of the total 1,096 nouns (both general and proper nouns) with vowel repetition. The second most frequently used vowel was /o/ found in 204 different nouns or 18.61 percent of the total nouns counted.

Ranking third was the vowel /i/, repeated in 158 different nouns or 14.42 percent of the total nouns counted. The vowel /u/ came in forth, repeated in 57 nouns or 5.20 percent of all nouns counted. The least repeated vowel was /e/, repeated in only 3 different nouns or 0.27 percent of all nouns counted. Kubozono (1999) explained that the three vowels of /a/, /i/ and /u/ are most frequently found within the various languages of the world. It is therefore reasonable to expect that these vowels will be repeated more often in a single noun in Japanese than the vowel /e/. Since the vowel /o/ is ranked at the top of the 'sound hierarchy' (Murata, 1984, 1990; Tamaoka & Murata, 1999), /o/ tended to be repeated more than the vowel /e/, although both these vowels have points of articulation within the middle of the vowel space. This tendency was observed in both general and proper nouns in the same way.

Table 1
Same Vowels in Three Consecutive V Positions

| | Word Frequency ($\Sigma N_i$) | | | Accumulative Word Frequency ($\Sigma W_f$) | | |
|---|---|---|---|---|---|---|
| Vowels | General Nouns | Proper Nouns | Total | General Nouns | Proper Nouns | Total |
| /a/ | 197 | 477 | 674 | 166,823 | 113,351 | 280,174 |
| /i/ | 118 | 40 | 158 | 36,620 | 1,664 | 38,284 |
| /u/ | 25 | 32 | 57 | 4,710 | 5,364 | 10,074 |
| /e/ | 2 | 1 | 3 | 136 | 11 | 147 |
| /o/ | 51 | 153 | 204 | 183,961 | 13,832 | 197,793 |
| Total | 393 | 703 * | 1,096 * | 392,250 * | 134,222 * | 526,472 * |
| Grand Total | 8,142 | 10,348 | 18,490 | 4,664,720 | 1,090,679 | 5,755,399 |
| Ratio | 4.83% | 6.79% **H** | 5.93% **H** | 8.41% **H** | 12.31% **H** | 9.15% **H** |

*Note 1* : * $p$ <.01.
*Note 2* : The random possibility of 3-mora CVCVCV nouns which have the same vowels in three consecutive V positions is 4 percent.
*Note 3* : The sign H refers to the frequency of nouns with vowel repetition which is significantly higher than random chance (4.00%).
*Note 4* : The grand total of 18,490 refers to the total number of nouns with a CVCVCV phonological structure out of the 341,771 nouns taken from the word corpus of Amano and Kondo (2000).
*Note 5* : The grand total of 5,755,399 refers to the total accumulative word frequency for the 18,490 nouns with a CVCVCV phonological structure.

The frequency of nouns with vowel repetition appearing in Japanese written texts was examined using Chebysheff's inequality theorem (see Maezono, 2002; Matsubara, Nawata & Nakai, 1994; Suzuki, 1999). The calculation of probability is provided by:

$$P(|X - m| \leq k\sigma) > 1 - 1/k^2$$

where the $m$ is a mean of a scattered variable $X$ and sigma ($\sigma$) is a standard deviation[2]. Using this measurement, 393 nouns (4.83%) were found to contain the same repeated vowels from among 8,142 general nouns with a CVCVCV structure, which fell within the range of the random chance of occurrence (326 times or 4.00%) at the .01 level of significance. In contrast, 703 nouns (6.79%) were found to have the same vowels repeated among 10,348 proper nouns, which was sigificantly higher than the 4 percent chance of random occurrence. A total of 1,096 nouns (5.93%) out of 18,490 repeated the same vowel in the CVCVCV phonological strings. This noun frequency was significant at the probability level of 1 percent. In short, type frequency indicated that the assimilation constraint causing vowel repetition in nouns with a CVCVCV phonological

structure which was observed in the total number of proper nouns and the total number of both general and proper nouns together, but not in the total number of general nouns alone.

### 3.2.2. Token frequency

As discussed in the introduction of this paper, accumulative word frequency or token frequency is considered to be more accurate in indicating occurrence of words than type frequency. Similar to type frequency, token frequency also showed a similar pattern in terms of vowels repeated in three consecutive positions. As shown in Table 1, these vowels are listed in the Japanese vowel kana order of /a/, /i/, /u/, /e/ and /o/ in all the categories of general nouns, proper nouns and the total of both.

An interesting tendency observed in token frequency is a high accumulative frequency of occurrence of Japanese nouns with vowel repetition. Although type frequency of general nouns did not show a significantly high occurrence of words with vowel repetition, token frequency in the category of general nouns was calculated as 392,250 (8.41%) out of the total of 4,664,720. According to Chebysheff's inequality theorem, this figure of token frequency was significantly higher ($p < .01$) than the random chance of occurrence of 4 percent. Therefore, these nouns with vowel repetition in three consecutive positions are often seen in written texts. For the category of proper nouns, type frequency was 134,222 (12.31%) out of 1,090,679, which was significantly higher than the random chance of occurrence ($p < .01$). The grand total of 526,472 (9.15%) out of 5,755,399 also showed significantly high occurrence of nouns with same vowel repetition ($p < .01$).

### 3.3. Discussion

Study 1 examined the existent to which the same vowels were repeated three times within a CVCVCV phonological structure in a corpus of Japanese nouns. Although type frequency of general nouns did not show significantly high occurrences of same vowel repetition, token frequency indicated significantly high repetition. Since type frequency only counts a word once, regardless of how often it is used, the index of token frequencies reflects actual appearance in written Japanese texts. Therefore, Study 1 concluded that vowels were repeated among Japanese general and proper nouns with a phonological CVCVCV string far more frequently (i.e., 9.15%) than the random chance level of four percent.

## 4. STUDY #2: Frequency of CVCVCV-structured nouns with the same vowel occurring in two consecutive vowel positions

Study 2 investigated the frequency of nouns with the same vowel occurring in two consecutive V positions within a $CV_1CV_2CV_3$ string. In this case, a vowel could be repeated in either of two ways: (1) $V_1$ and $V_2$ or (2) $V_2$ and $V_3$. The same vowels found in $V_1$ and $V_3$ were not considered to be in consecutive positions, so they were only considered for their frequency of occurrence which was used simply for comparing the other two conditions of (1) and (2). Results would then be expected to display either a lesser or a greater degree of word frequency than the random chance rate of occurrence of 20 percent among CVCVCV-structured Japanese nouns.

### 4.1. Lexical Corpus and Procedure

Study 2 made use of the same lexical corpus as in Study 1.

### 4.2. Results

The number of Japanese nouns having the same vowel in two consecutive positions is shown in Table 2. Words with the same vowel in three consecutive positions discussed in Study 1 are not included in the detail counts of each vowel in Table 2. For determining the significance level of one percent, counts of three-consecutive-vowel repetitions were included. According to the totals of both type and token frequencies, the type of vowel found in two consecutive positions among $CV_1CV_2CV_3$-structured nouns was similar to the vowel found in the order of three consecutive positions, with the most frequently-repeated vowel being /a/ and the least frequently-repeated vowel being /e/.

#### 4.2.1. Type frequency

General nouns with the same vowel in the $V_1$ and $V_2$ positions were counted 1,664 times (20.44%) out of 8,142 general nouns. Likewise, proper nouns showed a similar count of 2,138 times (20.66%) out of 10,348 proper nouns. The total of both types of nouns together indicated a percentage of 20.56 or a count of 3,802 times out of 18,490 nouns. Once frequency counts of the same vowel in three consecutive positions were included, the figures become 2,057 (25.26%) for general nouns, 2,841 (27.45%) for proper nouns and 4,898 (26.49%) for both together. As indicated by the upper arrow in Table 2, all these figures were significantly higher than the random chance rate of occurrence of 20 percent based upon the calculation from Chebysheff's inequality theorem ($p < .01$). Therefore, it is concluded that nouns with the same vowels in the first and second V positions of a CVCVCV phonological string occur more frequently than random chance.

Type frequency of general nouns which have the same vowel in the $V_2$ and $V_3$ positions was 13.68 percent or 1,114 times out of 8,142 general nouns. Including the same vowel in three consecutive positions, type frequency became 18.51 percent or 1,507 times. According to the calculation based on Chebysheff's inequality theorem, this frequency of occurrence did not significantly differ from the 20 percent random chance rate. Similarly, same vowel occurrence in $V_2$ and $V_3$ positions among proper nouns was 18.86 percent or 1.952 times (12.07 percent or 1,249 times excluding the same vowel in thee consecutive positions) out of 10,348 proper nouns, which did not significantly differ from the random chance level. The total of both general and proper nouns occurring 18.71 percent or 3,459 times (12.78 percent or 2,363 times without the same vowel in three consecutive positions) out of 18,490 nouns did not show a significantly lower or higher frequency than the random chance level. Therefore, the same vowel in the second and third positions within CVCVCV strings occurs at the random chance.

Although the $V_1$ and $V_3$ positions of a $CV_1CV_2CV_3$ string were not considered to be consecutive, type frequency was counted as a basis for comparison. General nouns with the same vowels in the first and third positions of a CVCVCV string were counted 1,740 times or 21.37 percent (1,347 times or 16.54 percent without the same vowel in the three consecutive positions) out of 8,142 general nouns. This frequency of occurrence was the same as the random chance. However, proper nouns with such vowel repetitions appeared 2,548 times or 24.62 percent (1,845

times or 17.83 percent without the same vowel in three consecutive positions) out of 10,348 proper nouns, which was significantly higher than the random chance rate of occurrence of 20 percent ($p < .01$). The total of both general and proper nouns with this type of vowel repetition showed a frequency of 23.19 percent or 4,288 times (17.26 percent or 3,192 times without the same vowel in three consecutive positions) out of 18,490 general and proper nouns. This figure was at the random chance rate of occurrence.

Table 2
Same Vowels in Two V Positions

| | Word Frequency ($\Sigma N_i$) | | | Accumulative Word Frequency ($\Sigma W_f$) | | |
|---|---|---|---|---|---|---|
| Vowels | General Nouns | Proper Nouns | Total | General Nouns | Proper Nouns | Total |
| (1) Same vowels in $V_1$ and $V_2$ | | | | | | |
| /a/ | 797 | 1,177 | 1,974 | 689,385 | 144,159 | 833,544 |
| /i/ | 188 | 288 | 476 | 129,141 | 19,437 | 148,578 |
| /u/ | 310 | 319 | 629 | 169,432 | 56,245 | 225,677 |
| /e/ | 32 | 24 | 56 | 80,639 | 3,638 | 84,277 |
| /o/ | 337 | 330 | 667 | 268,265 | 29,682 | 297,947 |
| Total | 1,664 | 2,138 | 3,802 | 1,336,862 | 253,161 | 1,590,023 |
| Ratio | 20.44% | 20.66% | 20.56% | 28.66% | 23.21% | 27.63% |
| Including $V_1=V_2=V_3$ | 2,057 * | 2,841 * | 4,898 * | 1,729,112 * | 387,383 * | 2,116,495 * |
| Ratio | 25.26% H | 27.45% H | 26.49% H | 37.07% H | 35.52% H | 36.77% H |
| Grand Total | 8,142 | 10,348 | 18,490 | 4,664,720 | 1,090,679 | 5,755,399 |
| (2) Same vowels in $V_2$ and $V_3$ | | | | | | |
| /a/ | 368 | 517 | 885 | 218,740 | 91,836 | 310,576 |
| /i/ | 340 | 320 | 660 | 136,673 | 14,337 | 151,010 |
| /u/ | 179 | 89 | 268 | 75,007 | 7,238 | 82,245 |
| /e/ | 66 | 17 | 83 | 55,385 | 707 | 56,092 |
| /o/ | 161 | 306 | 467 | 131,564 | 13,333 | 144,897 |
| Total | 1,114 | 1,249 | 2,363 | 617,369 | 127,451 | 744,820 |
| Ratio | 13.68% | 12.07% | 12.78% | 13.23% | 11.69% | 12.94% |
| Including $V_1=V_2=V_3$ | 1,507 | 1,952 | 3,459 | 1,009,619 | 261,673 | 1,271,292 |
| Ratio | 18.51% | 18.86% | 18.71% | 21.64% | 23.99% | 22.09% |
| Grand Total | 8,142 | 10,348 | 18,490 | 4,664,720 | 1,090,679 | 5,755,399 |
| (3) Same vowels in $V_1$ and $V_3$ | | | | | | |
| /a/ | 338 | 643 | 981 | 194,094 | 96,122 | 290,216 |
| /i/ | 562 | 741 | 1,303 | 266,958 | 68,852 | 335,810 |
| /u/ | 193 | 66 | 259 | 87,718 | 11,936 | 99,654 |
| /e/ | 129 | 22 | 151 | 22,312 | 1,752 | 24,064 |
| /o/ | 125 | 373 | 498 | 62,130 | 33,314 | 95,444 |
| Total | 1,347 | 1,845 | 3,192 | 633,212 | 211,976 | 845,188 |
| Ratio | 16.54% | 17.83% | 17.26% | 13.57% | 19.44% | 14.69% |
| Including $V_1=V_2=V_3$ | 1,740 | 2,548 * | 4,288 | 1,025,462 | 346,198 * | 1,371,660 |
| Ratio | 21.37% | 24.62% H | 23.19% | 21.98% | 31.74% H | 23.83% |
| Grand Total | 8,142 | 10,348 | 18,490 | 4,664,720 | 1,090,679 | 5,755,399 |

*Note 1* : * $p < .01$.

*Note 2* : The random possibility of 3-mora CVCVCV nouns which have the same vowels in the 1st and 2nd, 2nd and 3rd or 1st and 3rd V positions is 20 percent.

*Note 3* : The sign H refers to frequency of words with same vowel repetition that is significantly higher than random chance while the sign L refers to frequency of words with same vowel repetition which is significantly lower than random chance (no such cases).

*Note 4* : The grand total of 18,490 refers to the total number of nouns with a CVCVCV phonological structure out of the 341,771 nouns taken from the word corpus of Amano and Kondo (2000).

*Note 5* : The grand total of 5,755,399 refers to the total accumulative word frequency for the 18,490 nouns with a CVCVCV phonological structure.

These results suggest that the vowels occur frequently in the first and second V positions of CVCVCV string, but occur at the random chance level in the second and third V positions. The frequency of nouns with the same vowels in the first and third V positions seems to fall between the two previous conditions. Among general nouns the frequency was equal to the random chance, among proper nouns there were significantly high occurrences, but both together the frequency returns to the random chance level.

#### 4.2.2. Token Frequency

Token frequency of Japanese nouns with the same vowels in two consecutive positions was also examined and these figures were considered more representative of the natural occurrence of these nouns than those of type frequency.

General nouns with the same vowel in $V_1$ and $V_2$ positions of a $CV_1CV_2CV_3$ string were counted 1,729,112 times or 37.07 percent (1,336,862 times or 28.66 percent without the same vowel in three consecutive positions) out of 4,664,720 general nouns. Based upon the calculation of Chebysheff's inequality theorem, this count was significantly higher than the random chance rate of occurrence of 20 percent ($p < .01$). Proper nouns showed a relatively high frequency rate of 387,383 times or 35.52 percent (23.21 percent or 253,161 times without the same vowel in three consecutive positions) out of 1,090,679 proper nouns. This token frequency rate was also significantly higher than random chance ($p < .01$). The total of both general and proper nouns indicated a high frequency rate of 36.77 percent or 2,116,495 times (27.63 percent or 1,590,023 times without the same vowel in three consecutive positions) out of 5,755,399 general and proper nouns, which was significantly higher than random chance ($p < .01$). Overall, token frequency of the same vowels in $V_1$ and $V_2$ positions in nouns with a $CV_1CV_2CV_3$ structure indicated significantly higher occurrences than the random chance rate of 20 percent. Therefore, it is concluded that these nouns appear more frequently than just by random chance.

Token frequency of general nouns, which had the same vowel in the $V_2$ and $V_3$ positions of its $CV_1CV_2CV_3$ phonological string, was 21.64 percent or 4,664,720 times (13.23 percent or 617,369 times without the same vowel in three consecutive positions) out of 4,664,720 general nouns. This frequency of occurrence was at the random chance level of 20 percent. Similarly, the same vowel occurring in the second and third V positions among proper nouns was 23.99 percent or 261,673 times (11.69 percent or 127,451 times without the same vowel in three consecutive positions) out of 1,090,679 proper nouns, which was at the random chance level. The total frequency of both general and proper nouns was, naturally, at the random chance level, being counted 22.09 percent or 1,271,292 times (12.94 percent or 744,820 times without the same vowel in three consecutive positions) out of a total of 5,755,399 of both nouns together. Therefore, the same vowels occur at the random chance level in the second and third V positions of a CVCVCV string.

Token frequency of nouns with the same vowel in $V_1$ and $V_3$ positions of a $CV_1CV_2CV_3$ string was counted as a basis for comparison. General nouns with the same vowel in the first and third V positions were counted 1,025,462 times or 21.98 percent (633,212 times or 13.57 percent without the same vowel in three consecutive positions) out of 4,664,720 general nouns. This token frequency figure for general nouns was at the random chance level. In contrast, proper nouns with the same vowel in the first and third V positions appeared 346,198 times or 31.74 percent (211,976 times or 19.44 percent without the same vowel in three consecutive positions) out of 1,090,679 proper nouns, which was significantly higher than the 20 percent random chance

rate of occurrence. However, the total token frequency among both general and proper nouns was 1,371,660 times or 23.83 percent (845,188 times or 14.69 percent without the same vowel in three consecutive positions) out of 5,755,399 both types of nouns together. This figure was at the random chance level.

### 4.3. Discussion

Study 2 examined the rate of occurrence of nouns with the same vowels found in two consecutive V positions within their CVCVCV strings. As shown in Table 2, type and token frequencies showed a general trend that the same vowels were found more frequently than at the random chance level in the first and second V positions of CVCVCV-structured nouns, but basically occur at the random chance level in the second and third V positions and in the first and third V positions, excluding proper nouns of the first and third V positions.

## 5. General Discussion

The present study proved that the same vowels in the three consecutive V positions of CVCVCV strings of nouns are more frequently repeated than the random possibility of 4 percent in both cases of type and token frequencies. Furthermore, the same vowels occurring in two consecutive or determined V positions were also observed in type and token frequencies of CVCVCV-structured nouns. In the two consecutive or determined positions, the random chance level was 20 percent. The same vowels in the first and second V positions of CVCVCV string of nouns occur more frequently than the random chance level. On the other hand, both type and token frequency were found to be at the random chance rate of occurrence for nouns with the same vowels repeated in the second and third V positions in their CVCVCV strings. Accordingly, CVCVCV-structured nouns with the same vowels in the first and second V positions occur more frequently than random probability, whereas same vowels in the second and third V positions occur at the rate of random chance. The same vowel repetition of the first and third positions, though not consecutive but determined positions, showed mixed results; the random chance level among the general nouns, but significantly higher than the random chance level among the proper nouns. However, when both the general and proper nouns were examined together, it returned to the random chance level.

As depicted in Figure 1, an interesting trend was found when combining the results of token frequency concerning nouns with the same vowels in three consecutive and two consecutive or determined V positions (see also Tables 1 and 2). The percentage of nouns with the same vowels in three consecutive V positions overlaps with the percentage of nouns with the same vowels in two positions. In other words, the percentages of nouns with the same vowels in $V_1$ and $V_2$, $V_2$ and $V_3$, and $V_1$ and $V_3$ positions all include the same percentage of nouns with the same vowels in $V_1$, $V_2$ and $V_3$ positions (i.e., 9.15%). Therefore, when 9.15 percent was subtracted from all total percentages of each condition of the same vowels in two V positions, the results were 27.63 percent for $V_1$ and $V_2$ positions, 12.94 percent for $V_2$ and $V_3$ positions, and 14.69 percent for $V_1$ and $V_3$ positions.

Figure 1 uses 'H' to refer to a frequency of occurrence that is greater than random chance. Thus, according to the boxes with the arrow H, it can be seen that nouns with the same vowels occurring in $V_1$ and $V_2$ positions, as well as in $V_1$, $V_2$ and $V_3$ positions with a $CV_1CV_2CV_3$

structure, have a total token frequency of 36.77 percent. This suggests that once the same vowels occur in $V_1$ and $V_2$ positions, they are likely to be repeated in the $V_3$ position (i.e., 9.15% for $V_1 = V_2 = V_3$).

In summary, the findings of the corpus study indicated that vowels were repeated among Japanese nouns with a CVCVCV string more frequently than the random possibility of 4 percent. In addition, nouns with the same vowels in the first and second positions (i.e., $V_1$ and $V_2$ in the $CV_1CV_2CV_3$) showed significantly higher occurrences than the random chance of 20 percent, whereas nouns with the same vowels in the second and third positions appeared at the random chance level (i.e., $V_2$ and $V_3$). Since higher frequency enhances speed and accuracy in naming, phonological structures with the same vowel repeated can be expected to be more quickly and accurately named. However, Tamaoka and Murata (2001) found in their experimental study that native Japanese speakers named CVCVCV-structured phonemic strings with same vowel repetition less quickly and less accurately than unrepeated ones. This finding conflicts with the results of the present corpus study and excludes the possibility of the frequency of vowel repetition affecting the speed and accuracy of phonological processing.

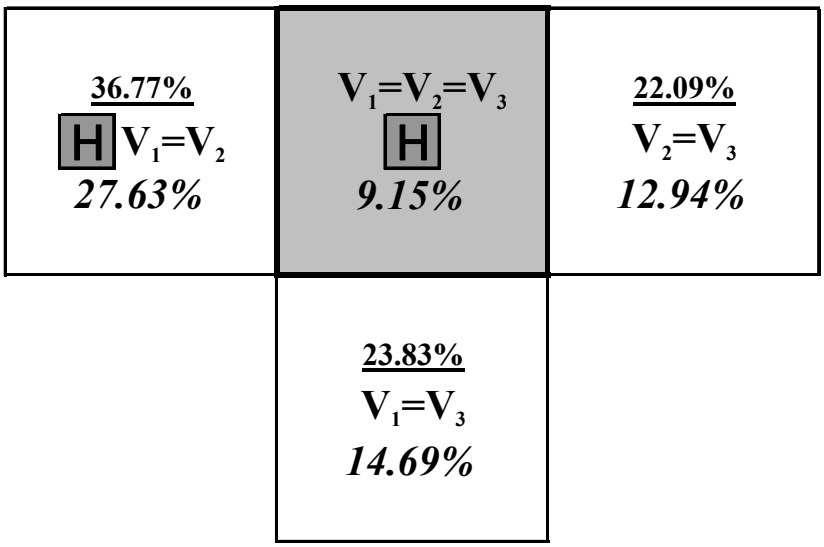


***Figure 1.*** Token frequency of Japanese nouns with vowel repetition

*Note 1* : A $V_1$, $V_2$ and $V_3$ refers to vowels in a $CV_1CV_2CV_3$ string.
*Note 2* : Percentages in Italics are the calculated frequencies of nouns out of the grand total of 5,755,399 CVCVCV-structured nouns.
*Note 3* : Percentages underlined indicate token frequency of that particular box plus the token frequency of nouns with the same vowels in $V_1=V_2=V_3$.

### *Notes*

[1] The pronunciation of words in this paper is transcribed using Japanese phonemic symbols which indicate three special sounds in Japanese: /N/ for nasal, /Q/ for geminate and /R/ for long vowel.

[2] The cutoff points at the 1 percent significant level for a type frequency of CVCVCV-structured general nouns is calculated as follows. The type frequency is 8,142 (see Table 1). A random probability is given by $1/5^3$ (three positions within five changeable vowels) ×5 (five different vowels) which equals 1/25. Thus, the mean $m$ of the variable X becomes 326 ($m = 8{,}142 \times 1/25 = 325.68$). The standard deviation $\sigma$ is given by $m$ multiplied by 24/25 and squared by one half, $\sigma = (8{,}142 \times 1/25 \times 24/25)^{1/2} = 17.68$. The probability at the 1 percent level of significance was established by $1 - 1/k^2 = 0.99$. Thus, the value $k$ was calculated as 10. Substituting all the values, the formula becomes $P(|X - 326| \leq 10 \times 17.68)$. After calculating this, the result was $P(149 \leq X \leq 503) \geq 0.99$. Since the actual type frequency is 393, this figure is within the range of 149 to 503. Consequently, the frequency 393 falls into a random chance range at the 1 percent level of significance. This calculation procedure is applied to all other frequencies in Studies 1 and 2.

**References**

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

**Fukazawa, H.** (2000). Typology of OCP on features. *Phonological Studies 3, 121-134.*

**Ito, J., & Mester, R.-A.** (1986). The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry 17, 49-73.*

**Kubozono, H.** (1999). *Nihongo no onsei [Japanese phonetics]*. Tokyo: Kenkyuusha Shuppan.

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

**Leben, W.** (1973). *Suprasegmental Phonology*. Doctoral dissertation submitted to the Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

**Maezono, Y.** (2002). *Gaisetsu kakuritsu tookei [An introduction to probability statistics]*. Tokyo: Saiensusha.

**Matsubara, N., Nawata, K., & Nakai, N.** (1994). *Tookeigaku nyuumon [An introduction to statistics]*. Tokyo: Tokyo University Press.

**McCarthy, J. J.** (1986). OCP effects: Gemination and Antigemination. *Linguistic Inquiry 17, 207-263*.

**Murata, T.** (1984). Jinkoo onomatope niyoru nihongo onsei haieraakii [Sound hierarchy of artificial onomatopes in Japanese]. *Linguistic Studies 85, 68-90.*

**Murata, T.** (1990). AB type onomatopes and reduplications in English and Japanese. In *Linguistic fiesta festschrift for professor Hisao Kakehi's sixtieth birthday: 257-272*. Tokyo: Kuroshio Shuppan.

**Suzuki, G.** (1999). *Joohooryoo kijun niyoru tookei kaiseki nyuumon [An introduction of information-based statistics]*. Tokyo: Kyoodansha.

**Tamaoka, K., & Murata, T.** (1999). Nihongo boin-no onsei haieraakii: Boin-o shotoo'on tosuru 3-paku koosei no muimi tsuzurigo-no meimei kadai kara [The hierarchical structure of Japanese vowels: A naming task investigation of 3-mora nonwords with initial vowel sounds]. *The Science of Reading 43, 79-89.*

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

**Yip, M.** (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry 19, 65-100.*