Pre-fortis Clipping Patterns: A Cross Linguistic Examination of L2 English Production and Perception by Asian Learners of English

Nguyen Van Anh Le

***Abstract*—In English, it is widely observed that vowels are longer before voiced consonants than before voiceless ones such as English. However, in Mandarin Chinese, Vietnamese, Japanese, and Korean, the distribution of voiced-voiceless stop contrasts, and long-short vowel differences are vastly different from English. The purpose of this study is to determine whether these targeted learners' L2 English production and perception change in terms of vowel duration as a function of stop voicing. The production measurements in the database of Asian learners revealed a distinct effect than the one observed in native speakers. There were no evident vowel lengthening patterns. The results of the perceptual experiment with 24 participants indicated that individuals tended to prefer voiceless stops when preceding vowels were shortened, but there was no statistically significant difference between intermediate, upper-intermediate, and advanced-level learners. However, learners demonstrated distinct perceptual patterns for various vowels and stops. The findings have valuable implications for L2 English speech acquisition.**

***Index Terms*—Voiced/voiceless stops, preceding vowel duration, L2 English, Asian languages**

I. Introduction

It is widely accepted that adults' speech production and perception are influenced to some extent by their native language's phonological system (L1). While learning a second language (L2), learners frequently struggle with producing and perceiving distinct sounds that exist only in L2 but not in L1. Numerous studies have demonstrated that foreign language learners and native speakers differ in their ability to produce and perceive L2 phonetic contrasts [1]. For example, unlike in English, where stops have a voicing contrast and can occur in syllable final positions, Mandarin lacks a voicing contrast for stops and allows them only at syllable onset, or in Japanese, where codas are classified into two types: those that end in a nasal and those that end in a geminate consonant (i.e., the consonant closes one syllable and serves as onset of the next syllable). As a result, the majority of learners from these L1 backgrounds struggle with producing and recognizing word-final stop voicing contrast in English. According to previous study conducted on Chinese speakers of English, voiced and voiceless stops are erased, and vowels are added after the final stop of the syllable [1, 2]. Similarly, Saunders [3] argued that vowel reduction was the favored manner of pronunciation for Japanese English (JPE). According to previous research, Mandarin Chinese speakers are adept at perceiving released syllable-final voicing, but struggle with unreleased syllable-final stop voicing [4, 5]. Indeed, Chinese speakers produced the lowest average vowel duration ratio in this study when compared to native English speakers.

It was suggested that vowel duration fluctuates as a function of the voicing of the next consonant is a language-universal phenomena; this tendency has been validated for English in numerous investigations [6–9]. Vowel length differences between voiced and voiceless consonants have been found to play a significant role in the perception of these consonants in numerous minimum pairs of CVC type in English. Vowel obstruents with durations less than or equal to 200 milliseconds were found to be voiceless, whereas those with durations greater than or equal to 300 milliseconds were determined to be voiced [10]. The phonological structure of a language determines how much an adjacent voiced or voiceless consonant influences the duration of its preceding vowel [11]. The "voicing effect" [12] or "pre-fortis clipping" are the terms for this occurrence [12, 13]. Vowel duration has been studied as a perceptual signal for post-vocalic consonant voicing in connection to the relationship between English consonants and preceding vowels, and the differences in vowel length were found to be sufficient to cue the perception of voiced and voiceless consonants in English [10, 14]. A significant finding was that a prolonged vowel serves as a cue to the listener that a voiced consonant is about to follow. For English [15, 16], for German [17], for French [18], for Korean [11], for Japanese [19], and for Chinese [20] are among the earliest traceable references to this phenomenon. Keating [21] also made a claim indicating that L2 vowel lengthening patterns are language-specific, which determined by speaker's native language.

Following these, a large number of studies confirmed the effect's existence in these languages and in an ever-growing list of others. Surprisingly, no known language has been claimed to have the opposite effect, that is, longer vowel durations preceding voiceless consonants than preceding voiced consonants. We focused on this issue in this study to address the following questions about Asian learners:

1) What are the English vowel length patterns before voiced and voiceless consonants among four languages?

2) Do they differ in their L2 English production and perception in terms of the preceding vowel duration as a function of the following word-final stop voicing?

3) Is there a distinction between intermediate, upper intermediate, and advanced learners in terms of perception?

II. Literature Review

There are several studies that have examined vowel

Manuscript received December 9, 2022; revised January 20, 2023; accepted February 23, 2023.

Nguyen Van Anh Le is with the Sophia University, Japan.
E-mail: lenguyenvananh1411@gmail.com (N.V.A.L.)

doi: 10.18178/ijlll.2023.9.2.387

lengthening before voiced coda produced by Dutch, Arabic, and French speakers of English, and no significant vowel lengthening in their L2 has been observed [22–24]. This collection of studies has shown that the durational difference between vowels preceding contrast voicing codas is markedly greater in English than in other languages; and these results have first evidenced against vowel lengthening before voiced codas as universal. To that end, this paper continues at examining the vowel lengthening patterns of speakers from other L1 backgrounds (i.e, Chinese, Vietnamese, Korean, and Japanese) to explore whether vowel duration in L1 determines vowel lengthening patterns in L2. More specifically, if L1 Japanese speakers will be found to produce longer vowels before English voiced codas than voiceless ones, it could be supposed that the presence of a phonemic vowel length distinction in speaker's native language can determine their L2 vowel length distinction. In other words, it could be said that the acquisition of L2 vowel patterns depends on L1 vowel system in which the vowel length will be affected. It also may imply that the long and short vowel phonemes in Japanese allow their speakers to acquire long and short L2 allophones. On the other hand, if no vowel lengthening will be found on Japanese speakers, the claim that L2 vowel lengthening patterns depends on L1 vowel system could be weakened.

Moreover, it is expected that learners whose language without vowel quantity distinction as Korean hardly distinguish vowel lengthening before English voiced codas. If L1 Korean speakers will be found to effectively produce longer duration of vowels before voiced codas than voiceless ones, it may also weaken the claim that L2 vowel lengthening patterns are affected by L1 vowel system, and vice versa, L1 Korean speakers' inability to produce long and short English allophones could be explained by their inexistence of long and short vowels in their language system.

Since the languages mentioned above does not have word-final voicing contrast, it is our interest to investigate whether those native speakers are aware of this vowel length effect in English word-final voicing distinction. In other words, we want to find out whether the speakers of different language backgrounds use vowel length as a perceptual cue for English syllable-final voicing judgment.

The purpose of this paper is to examine whether L2 speakers of English can interpret English coda voicing distinction in the speech, and the target vowels used in this paper are not only monophthongs but also diphthongs to show a closer look on the vowel lengthening patterns of these four major languages in Asia. Moreover, this research is to quantify the extent to which pre-consonantal vowels produced by Asian bilingual speakers in each of four languages differing in duration to reflect language-specific settings triggered by the local segmental environment. This research endeavors to accomplish two broad objectives. To begin, I sought acoustic data that could aid in identifying the duration of English vowels preceding voiced/voiceless codas in four Asian languages. The second primary objective is to shed light on the debate over reported cross-linguistic differences through an analysis of four related but contrasting languages.

Comparing languages that differ in the presence or degree of vowel durational differences can reveal cross-linguistic variation differences among Japanese, Korean, Chinese, and Vietnamese.

English phonology has a phonemic contrast between voiced and voiceless codas, while Chinese and Vietnamese do not have aspiration in coda position. In addition, neither voiced-voiceless stop contrasts nor long-short vowel contrast exist in Chinese. While Korean has no long or short vowel contrast, Japanese and Vietnamese do have phonemic distinction between long and short vowels but Japanese still has no obstruent in its syllable codas [3].

The current research examines the acoustic correlates of vowel duration in two distinct contexts: monosyllabic environments and word-isolation disyllabic environments, both of which terminate with voice contrasting obstruents.

In China, Korea, Japan, and Vietnam, the teaching of English as a second language has primarily focused on an articulatory phonetics approach based on a contrastive examination of the first and target languages [25], [26]; while pronunciation instruction is more focused, little attention is paid to provide students with a suitable frame of reference for observing, mimicking, and comprehending the learning of pronunciation, as well as to testing, providing feedback, and correcting their errors [27].

The effect of voicing contrasts and lexical stress on vowel duration will be explored in this paper, as it was previously suggested that stressed syllables had longer duration than unstressed syllables [28, 29].

Previous studies have claimed that adult learners struggle to produce L2 contrasts [30–33] which may be related to perceptual assimilation of both L1 and L2 sound system, and due to their proximity to L1 qualities, the accuracy with which L2 sounds are produced may be limited. To substantiate this, Flege [34–37] employed the Speech Learning Model (SLM) to argue that when new sound categories are introduced, people's sound systems undergo reorganization. Best [38] also provides a Perceptual Assimilation Model (PAM), claiming that when non-native speakers deal with L2, they rely on their native phonemic systems, which means that if an L2 sound is comparable to their L1 system, the sound would be adopted successively. They will, however, struggle to acquire an L2 sound if assimilation into the L1 sound category is problematic.

To native English speakers, the phonemic distinction between "bed" and "bet" is the final consonant's voicing; this voicing can then be categorized as a distinguishing signal, as it is one of the methods to differentiate minimal pairs. As indicated above, the length of the vowel preceding the final consonant can also be used to identify the target in a minimal pair, though this is a redundant cue. When students are learning a second language, they attempt to gain the ability to recognize and generate distinctions between L2 phonemes and allophones by transferring patterns from their first language.

Previous research has revealed data on vowel duration among speakers of languages other than English. However, the vowel lengths of Chinese, Korean, Vietnamese, and Japanese speakers were not described. Are these Asian systems similar in terms of vowel lengthening patterns preceding voiced and voiceless codas? Are these patterns determined by the L1 vowel system's features or by universal factors? Are these patterns, in particular, language-specific or universal?

The goal of this study is to characterize the vowel duration produced by four groups of non-native English speakers to evaluate how speakers of these four languages acquire the English allophonic vowel duration contrast. In which, Chinese, Vietnamese-L1 English learners (Group 1), whose L1 is tone languages and have no audible released codas. Korean-L1 English learners (Group 2), whose L1 has no vowel quantity distinction (no VQD); and Japanese-L1 English learners (Group 3), whose L1 is pitch accented language, has no aspirated codas and has phonemic length contrast (VQD).

III. Method

Both the production materials and the perceptual stimuli were created specifically for this project.

The speakers were recruited from prestigious colleges in Japan, Korea, Vietnam, and China, and as a result, their English proficiency was equivalent to or more than the intermediate level.

IV. Experiments

A. Production

The speakers of the following languages were recruited for participants: Japanese (JPN), Chinese (CN), Korean (KR), Vietnamese (VN), and English (NS). They were divided into groups according to their L1 for the experiment. The population consisted of eight individuals from four different countries and two individuals from the United States (NS: native speakers). Each of them was a graduate student. Except for the NS, they were all non-native English speakers who began formal school-based English education in their own countries during their adolescence (China, Korea, Japan, and Vietnam). To that goal, they had all spent more than a decade studying English.

A speech production experiment was conducted to compare the performance of the above Asian bilinguals, who learned English as a second language in adulthood, with that of native English speakers, who learned the language simultaneously. All participants were recruited and compensated for their time. They were recruited by personal contact with the researcher after responding to an invitation letter issued via email to engage in a research study.

The JPN, CN, KR, and VN groups consisted of eight participants aged 18–33 from Tokyo, Beijing, Seoul, and Da Nang, respectively. Each speaker possessed normal speech and hearing abilities and had no prior history of communication difficulties or intellectual disabilities. Procedural instructions were given to the participants to pronounce each target word inserted within a carrier sentence. The participants were given sufficient time and direction to perform the recording work efficiently. Three times each target word was displayed. To facilitate a better stimuli creation, all recordings were saved as 44.1kHz, 16-bit WAV files.

The stimuli used in this experiment were manipulations of real words. The stimuli are composed of common and uncommon words, where frequent words can trigger the subject's lexical processing, allowing them to be produced more easily than rare ones. Twenty-six words in the form of minimal pairs and twelve disyllabic words with target vowels on unstressed syllables were randomly recorded for each subject in a soundproof room using a SONY PCM-D50 recorder and SONY ECM-959A microphone. The total of thirty-eight words was placed in a carrier phrase to ensure that the rising or falling intonation did not affect the participants' performance as it might with a single word list. Additionally, the phrases containing target words were presented in a random order via PowerPoint slides. English was employed as the instruction language to engage the target language, and each session lasted approximately 10 minutes.

The distribution of vowels covers a range of vowel heights: one low vowel /ɒ/, one mid-low vowel /æ/, one mid vowel /ɛ/, and one mid-high vowel /ɪ/. Participants were instructed to create utterances in which the target word contrasted in voicing value for the post-vocalic consonant and were placed within the frame sentence "Say please," as the consonant-initial word following the target word created an environment that could prevent the linking of sounds.

Results

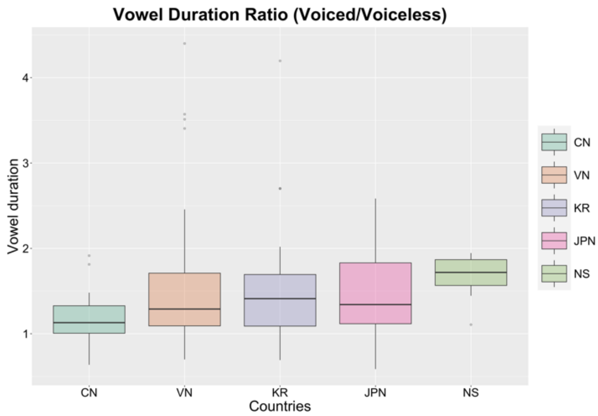

Fig. 1. Vowel duration ratio between voiced to voiceless pairs by participants' native language group.

As previously observed, vowels were substantially longer in voiced situations than in voiceless settings, with a mean ratio of 1:1.51 for voiced to voiceless. The ratio of vowel durations (VDR) by participant native language group is depicted in Fig. 1. A one-way ANOVA revealed a significant difference in average VDR between language groups, $F(8, 333) = 16.66$, $p < 0.001$. The test, however, revealed no statistically significant difference between the native and non-native English groups ($p = 0.54$). The VDR was distributed evenly across all groups.

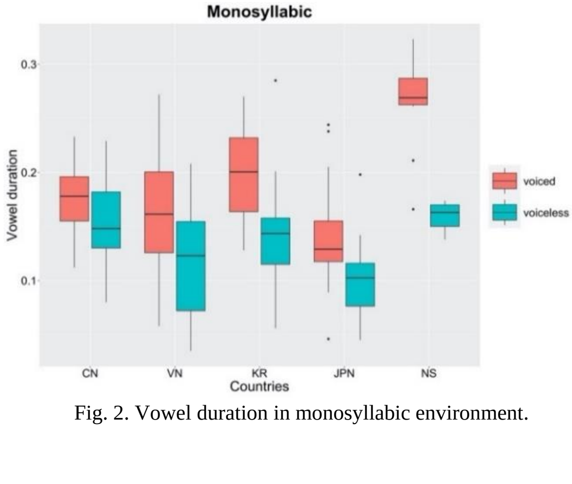

Fig. 2. Vowel duration in monosyllabic environment.

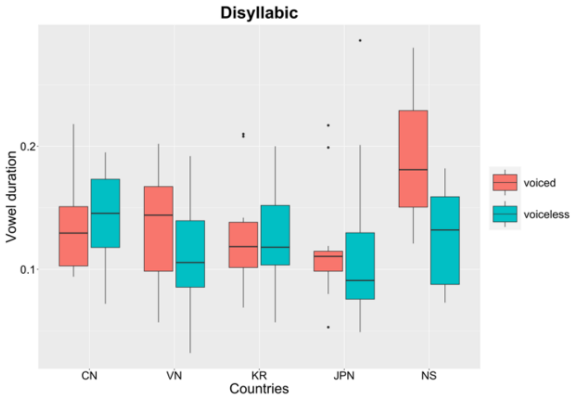

Fig. 3. Vowel duration under disyllabic environment.

In English monosyllabic environment as in Fig. 2, the vowel is longer when the postvocalic consonant is voiced. For every language group, it can be seen that the average vowel duration was higher for vowels that preceded voiced consonants than voiceless consonants. However, the difference between average vowel durations pre-voiced and pre-voiceless of four non-native English speakers is not significant ($p > 0.5$). More specifically, the first tonal group including Chinese and Vietnamese produced very similar vowel lengthening patterns in monosyllabic environment ($F= 0.95$, $p > 0.5$). All four groups showed similar pattern for longer vowel before voiced than voiceless codas. However, there was significant difference of vowel duration ratio of non-native compared to those of the native group ($F= 0.58$, $p < 0.01$). The native group showed the highest average pre-voiced duration and the lowest average pre-voiceless duration of vowels, the difference between pre-voiced and pre-voiceless duration of this group is significant, $F(11, 12) = 11.944$, $p < 0.001$.

The voicing status of the postvocalic consonant has no effect on the duration of the speech interval between the releases of two stops flanking a stressed vowel. However, in disyllabic environment as in Fig. 3, within tonal group, different vowel lengthening patterns of Chinese and Vietnamese speakers have been shown in which Chinese speakers lengthened vowel more before voiceless than voiced codas. This finding shows a lack of support for expectation indicating that Chinese and Vietnamese speakers may share similar patterns due to the tones those languages possess. More surprisingly, Korean group showed the same ratio for pre-voiced and pre-voiceless vowel duration, $F(13, 11) = 1.097$, $p = 0.8$. It is well-known that Korean speakers struggle to produce the long-short vowel contrast because Korean lacks the long-short vowel contrast [19, 39–41]. The contrast between long and short vowels entails differences in vowel quality as well as length differences [42].

B. *Perception*

We recruited individuals with three levels of English competence from prestigious colleges in Japan, Korea, China, and Vietnam. Subjects with TOEFL scores greater than 95, IELTS scores greater than 6.5, or TOEIC scores greater than 945 were classified as advanced level; those with TOEFL scores greater than 72, IELTS scores greater than 5.5, or TOEIC scores greater than 785 were classified as upper-intermediate level (B2); and the remainder were classified as intermediate level. Thus, we recruited 24 participants, eight of whom were advanced, eight upper-intermediate, and eight intermediate learners. They were all born and raised in their own countries and speak their mother language. They claimed to have normal hearing and volunteered to participate in the experiment in exchange for a compensation.

In a quiet environment, each participant did the 2AFC (two-alternative forced choice) identification task using Inquisit [43]. After the participants heard the sound stimulus CVC, they were asked to indicate which word they believed it was by pressing "1" or "2" on the keyboard or "3" to listen once more. Before the experiment began, the author instructed the participants for around five minutes to verify that they understood and could follow the task instructions precisely. The identification tasks consisted of 126 stimuli (3 coda articulation points × 2 voicing criteria × 7 vowel types × 3 duration steps), which were given in a random order. The entire experiment took roughly 20 minutes to complete. Each stimulus's choice and response time were recorded.

Inquisit was used to test the subjects and the stimuli were delivered via earphone. The exam was delivered to participants through email as a link; once they clicked the link, the items were randomly assigned and shown to the subjects; they then selected their response using an online platform. Each item was shown twice, followed by a 5-second quiet period as they deliberated on the responses.

This section's word list is derived primarily from the list used throughout the production experiment. It has a total of 42 words, each of which represents one of 42 possible rhyming types (3 articulation points in the coda × 2 voicing requirements × 7 vowel types). Seven vowels from American English [ɪ, e, æ, aɪ, ɒ, i: əʊ] were chosen as stimuli to represent high-mid-low, front-back, and monophthong and diphthong vowels. They were joined with three pairs of voiced/voiceless stops at word-final locations (labial [b, p], dental [d, t], and velar [g, k]); thus, 21 minimal pairs were obtained. Each word is separated into three distinct vowel length stages (0.25, 0.5, and 0.75 shorter respectively compared to the original vowel). An adult native English speaker was requested to serve as the speaker. The recording was done digitally in a soundproof room at 16 bit and 44.1 kHz. The speaker produced each of the 21 pairings three times in a typical and consistent manner. These stimuli were converted to digital format and stored in the computer.

Results

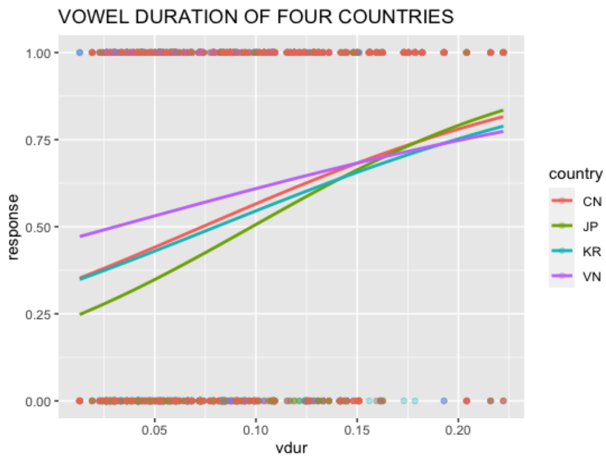

Fig. 4. Voiced (1.0)-voiceless (0.0) response as a function of vowel duration (msec) of four countries.

Although none of the four languages in this study include tense and lax contrasts in their L1 inventory, they generally were able to determine voiced-voiceless response as a function of vowel duration as shown in Fig. 4. However, only the Korean and Japanese groups achieved statistical significance ($p < 0.001$).

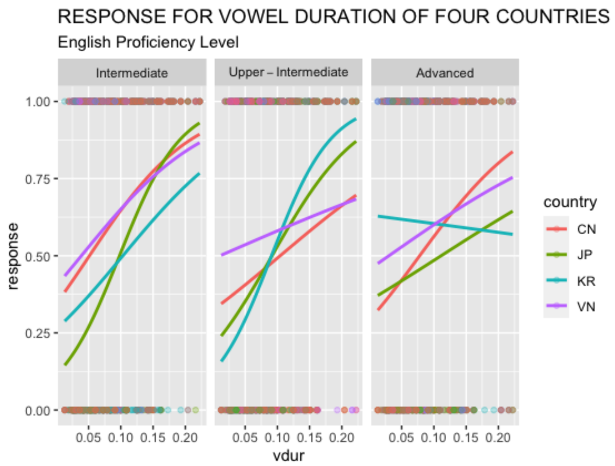

Fig. 5. Voiced (1.0)-voiceless (0.0) response as a function of vowel duration of different English proficiency levels of four countries.

On the basis of their English skill levels, the EFL learners were divided into Intermediate, Upper-Intermediate, and Advanced groups. Surprisingly, the graph in Fig. 5 shows that regardless of the group of non-native English speakers, experienced groups generated and interpreted voiced-voiceless response as a function of vowel duration less accurately than somewhat inexperienced groups.

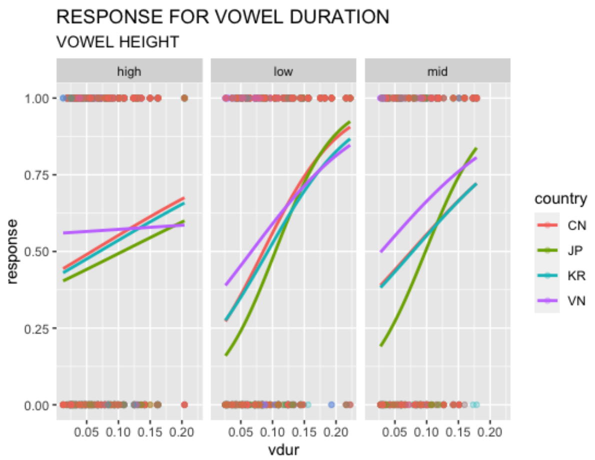

Fig. 6: Voiced (1.0)-voiceless (0.0) response as a function of vowel duration depending on different vowel height.

Figs. 6 and 7 summarize the effect of different vowels for the experiment's responses. The experiment also examines the relationship between perceived vowel duration and vowel quality and place of articulation (Fig. 8), and a logistic regression analysis was conducted to check the statistical significance. Vowel duration step; vowel quality [ɪ, e, æ, aɪ, ɒ, i:, əʊ]; English proficiency, country, and voicing of the original coda were all fixed effects; subject and word were random factors. Additionally, the experiment examines the effect of vowel categories on the perceived duration of vowels. If the relationship between vowel height and duration is a continuum, the same effect observed across categories should also be apparent within them: Vowels with a lower F1 are perceived as being longer than those with a higher F1. Even though no significant difference between low, middle, and high vowels was observed, we can see that the response rate of four countries for vowel duration of low vowels was quite consistent.

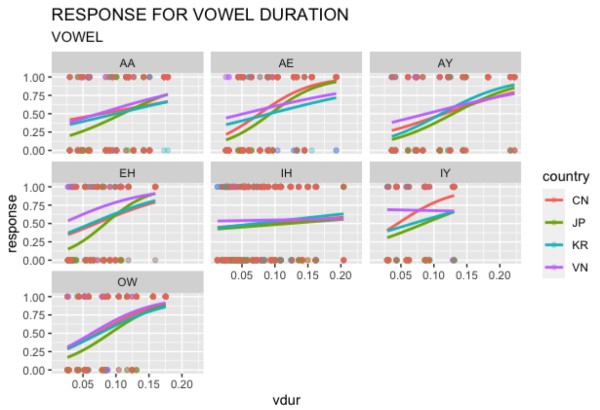

Fig. 7. Voiced (1.0)-voiceless (0.0) response as a function of vowel duration depending on different vowels (AA=[ɒ], AE=[æ], AY=[aɪ], EH=[e], IH=[ɪ], IY=[i:], OW=[əʊ]).

Overall, participants differentiated voiced-voiceless response as a function of vowel duration between other vowels except /ɪ/ most accurately and the difference between groups is significant ($p < 0.05$). Fig. 7 shows the response for vowel duration of separate vowels, vowel IH (i.e. /ɪ/) was discriminated differently and unsuccessfully compared to other vowels.

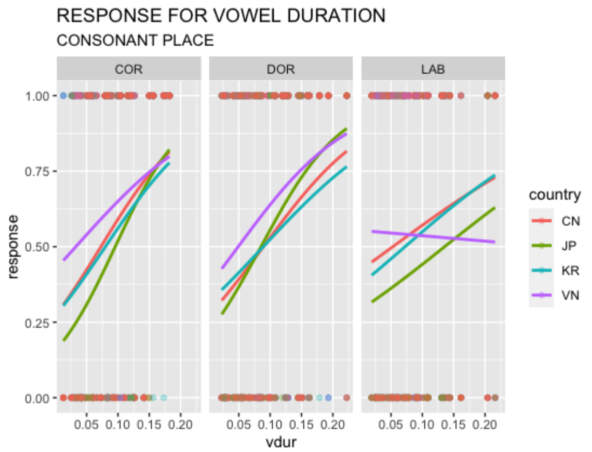

Fig. 8. Voiced (1.0)-voiceless (0.0) response as a function of vowel duration depending on consonant place.

The logistic regression analysis showed a significantly different pattern for the response for vowel duration at labial place of articulation ($p < 0.01$). Especially, as can be seen in the figure above, the Vietnamese has the lowest response rate for vowel duration at this place. In English, the latter phoneme /p/ is phonetically significant, whereas in Vietnamese, it is complimentary to its voiced companion /b/. Although it appears in the initial position in several loan terms, not all Vietnamese pronounce it as a voiceless consonant. The bilabial /p/ is rarely found at initial position in Vietnamese, but mostly in the final position and unreleased. The Vietnamese are unable to pronounce the initial and middle locations of English /p/, they pronounce /b/ instead.

The final and medial voiceless and voiced fricatives in English are difficult for them to differentiate since in Vietnamese there is no medial consonant and all of the final consonants are either voiceless stops or nasals. They're also having trouble with English's /p/ allophones. Their phonetic habits make it difficult to pronounce /p/ with a "puff of air" [26]. These disparities, which are unique to English, create significant tensions and challenges for the Vietnamese learners of English. They frequently bring their own phonemic habits to English and find it difficult to mimic proper pronunciation.

V. Discussion

The results of the production experiment for monosyllabic words fairly support the universality hypothesis that vowels are universally longer before voiced consonants than voiceless consonants as vowel lengthening patterns were observed in four cases where the L2 was English and the L1 was a variety of different languages. However, the data analysis process for the production experiment also revealed that disyllabic words exhibited inconsistent patterns, necessitating additional research. This contradicts the claims that vowels are universally longer before voiced consonants than voiceless consonants, and the claim that physical properties of speech signals such as vowel duration difference of voicing are "supplied by universal rules" [11].

In the perception experiment, participants demonstrated a reasonable ability to discriminate voiced-voiceless contrast in the context of variable vowel durations. While the majority of Asian learners are unaware of the English phonological rule to produce longer vowels preceding voiced stops, they automatically apply it to their perception. However, there is no differentiation between intermediate, upper-intermediate, and advanced level learners; the lack of a clear distinction between English proficiency levels may indicate that adult learners fail to adopt these subtle phonetic qualities. The current study aims at examining the association between accurate voiced-voiceless response with the variable vowel duration perception and L2 English proficiency levels achieved by Asian EFL learners; nevertheless, the findings indicate that there is no correlation between the voicing effect of English vowel duration perception and L2 proficiency. Generally, the significant effect was detected only when the preceding vowel is the high-front /ɪ/. Numerous earlier investigations have established that native Korean speakers lack a comprehension of the spectral distinction between the tense and lax contrast /i/-/ɪ/ because their L1 inventory lacks the phonological classification of tense and lax subgroups [26, 44]. However, the outcomes of this study may imply that Korean and Japanese English learners with normal proficient L2 fluency can discern voiced-voiceless response as a function of vowel duration regardless of tense-lax contrast. Furthermore, numerous studies have demonstrated that while Korean EFL learners disproportionately rely on the temporal characteristic to differentiate between tense and lax contrasts [5, 26, 44], the findings also indicate that Korean learners with relatively lower proficient L2 fluency ratings are surprisingly adept at discriminating voiced-voiceless response as a function of vowel duration in this study. In the case of other vowels, the results indicated that Asian English learners rarely discern distinctions in a spectral manner. The results, however, indicated an unexpected finding: the discrimination between voiced-voiceless response as a function of vowel duration was greater at the lower English level than at the higher or medium levels.

VI. Conclusion

This study added to the evidence indicating that Asian learners automatically adhere to the English phonological rule that vowels preceding voiced stops are longer than those preceding voiceless stops in their production and perception. When the vowel is a diphthong, their production performance suffers. In all the experiments, vowels that were produced with a voiced coda were more likely to be identified as longer than those produced with a voiceless coda. However, this concept for the evolution of voicing-conditioned vowel duration predicts that regardless of the listener's native language, the acoustic influences of the original coda will affect listeners' perception of vowel duration accordingly. The relationship between voiced-voiceless distinction with respect to vowel duration perception and L2 proficiency was evident only for the preceding high vowel vs front tense and lax vowels /i:/ and /ɪ/. The higher level of English proficiency ratings, the less they distinguished the voiced-voiceless distinction. Additionally, while there was no interaction effect of vowel and English proficiency level on the discrimination of voiced-voiceless distinction of adjacent vowel, and the higher level demonstrated poorer performance than the middle and lower levels.

Appendix

TABLE A1: Production Wordlist

Built-Build	Light-Lied	Goat-Gold	Pad-Pat
Lock-Log	Tote-Told	Bet-Bed	Pid-Pit
Sight-Side	Cap-Cab	Dock-Dog	Pod-Pot
Phallus-Pizzas	Racquet-Candid	Robot-Lingcod	Got-God
Fallout-Newfound	Rowboat-Household	Bailiff-Massive	

TABLE A2: Perception Wordlist

Pit-Pid	Mip-Mib	Pick-Pig
Bet-Bed	Vep-Veb	Peck-Peg
Beet-Deed	Beep-Beeb	Leek-League
Pat-Pad	Cap-Cab	Back-Bag
Light-Lied	Vipe-Vibe	Mike-Maig
Goat-Gold	Tope-Tobe	Coke-Cold
Got-God	Top-Tob	Lock-Log

Conflict of Interest

The author declares no conflict of interest.

References

[1] J. E. Flege, "The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification," *Journal of Phonetics*, vol. 15, no. 1, pp. 47–65, 1987, doi: https://doi.org/10.1016/s0095-4470(19)30537-6

[2] S. Weinberger, *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, Cambridge, MA: Newbury House Publishers, 1987.

[3] N. J. Saunders, "Morphophonemic variation in clusters in Japanese English," *Language Learning*, vol. 37, no. 2, pp. 247–272, 1987, doi: https://doi.org/10.1111/j.1467-1770.1987.tb00567.x

[4] J. E. Flege, "Using visual information to train foreign-language vowel production*," *Language Learning*, vol. 38, no. 3, pp. 365–407, 1989, doi: https://doi.org/10.1111/j.1467-1770.1988.tb00417.x

[5] Y. Zhang, *Distributional Learning of Extrinsic Vowel Duration Differences by Mandarin Native Speakers*, 2017.

[6] A. S. House and G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels," *The Journal of the Acoustical Society of America*, vol. 25, no. 1, pp. 105–113, 1953, doi: https://doi.org/10.1121/1.1906982.

[7] G. E. Peterson and I. Lehiste, "Duration of syllable nuclei in English," *The Journal of the Acoustical Society of America*, vol. 32, no. 6, pp. 693–703, 1960, doi: https://doi.org/10.1121/1.1908183

[8] A. S. House, "On vowel duration in English. Scitation," *Acoustical Society of America ASA*, September 1, 1961, doi: https://asa.scitation.org/doi/abs/10.1121/1.1908941

[9] K. R. Kluender, R. L. Diehl, and B. A. Wright, "Vowel-length differences before voiced and voiceless consonants: An auditory explanation," *Journal of Phonetics*, vol. 16, no. 2, pp. 153–169, 1988, https://doi.org/10.1016/s0095-4470(19)30480-2

[10] L. J. Raphael, "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," *The Journal of the Acoustical Society of America*, vol. 51, no. 4B, pp. 1296–1303, 1972, doi: https://doi.org/10.1121/1.1912974

[11] M. Chen, "Vowel length variation as a function of the voicing of the consonant environment," *Phonetica*, vol. 22, no. 3, pp. 129–159, 1970. doi: https://doi.org/10.1159/000259312

[12] F. Mitleb, "Voicing effect on vowel duration is not an absolute universal," *The Journal of the Acoustical Society of America*, vol. 71, no. S1, 1982. doi: https://doi.org/10.1121/1.2019285

[13] J. C. Wells, S. Ramsaran, and A. C. Gimson, "Syllabification and allophony," in *Studies in the Pronunciation of English: A Commemorative Volume in Honour*, A. C. Gimson, Ed. London: Routledge, 1990, pp. 76–86.

[14] P. Denes, "Effect of duration on the perception of voicing," *The Journal of the Acoustical Society of America*, vol. 27, no. 4, pp. 761–764, 1955, doi: https://doi.org/10.1121/1.1908020

[15] E. A. Meyer, *English Sound Duration: An Experimental Phonetic Study*, Uppsala, 1903. (in German)

[16] B. Lindblom, "Speech production. Vowel duration and a model of lip mandible coordination," *The Journal of the Acoustical Society of America*, vol. 72, no. 6, pp. 2039–2039, 1982, doi: https://doi.org/10.1121/1.388647

[17] E. A. Meyer and A. Noreen, *Nordic Studies: Dedicated to Adolf Noreen P°A His 50th Anniversary March 13, 1904*, pp. 347–356, 1904, Uppsala: Appelberg. (in Norwegian)

[18] A. Grégoire, "Influences of plosive consonants on the duration of preceding syllables," *Phonetics Review*, no. 1, pp. 260–292, 1911. (in French)

[19] K. Yoneyama and M. Kitahara, "Voicing effect on vowel duration: Corpus analyses of Japanese infants and adults, and production data of English learners," *Journal of the Phonetic Society of Japan*, vol. 18, no. 1, pp. 30–39, 2014.

[20] H. Ding, Y. Zhan, J. Yuan, and S. Liao, "Production of English stops by mandarin Chinese learners," *Speech Prosody 2018*, 2018, doi: https://doi.org/10.21437/speechprosody.2018-179

[21] P. A. Keating, "Phonetic and phonological representation of stop consonant voicing," *Language*, vol. 60, no. 2, p. 286, 1984, doi: https://doi.org/10.2307/413642

[22] J. E. Flege and R. Port, "Cross-language phonetic interference: Arabic to English," *Language and Speech*, vol. 24, no. 2, pp. 125–46, 1981, doi: https://doi.org/10.1177/002383098102400202

[23] M. Mack, "Voicing-dependent vowel duration in English and French: Monolingual and bilingual production," *The Journal of the Acoustical Society of America*, vol. 71, no. 1, pp. 173–178, 1982. doi: https://doi.org/10.1121/1.387344

[24] A. G. B. Elsendoorn, "Quality and quantity in English by dutchmen: Two parameters inducing double dutch," *De Gruyter Mouton*, July 11, 2011, doi: https://doi.org/10.1515/9783110846270.57

[25] G. Couper, "The value of an explicit pronunciation syllabus in ESOL teaching," *Prospect*, vol. 18, no. 3, pp. 53–70, 2003.

[26] H. V. Van, "Teaching English at Vietnamese tertiary level: Which register should we teach general, academic or a combination of the two?" in *Teaching English to Speakers of Other Languages in the Internationalization of Higher Education in Vietnam*, Hanoi: Australian Embassy, 2007.

[27] S. L. McKay, "Teaching English as an international language: The role of culture in Asian contexts," *The Journal of Asia TEFL*, vol. 1, no. 1, pp. 1–22, 2004.

[28] Fry, D. B. "Duration and intensity as physical correlates of linguistic stress," *The Journal of the Acoustical Society of America*, vol. 27, no. 4, pp. 765–768, 1955, doi: https://doi.org/10.1121/1.1908022

[29] M. Gordon and T. Roettger, "Acoustic correlates of word stress: A cross-linguistic survey," *Linguistics Vanguard*, vol. 3, no. 1, 2017, doi: https://doi.org/10.1515/lingvan-2017-0007

[30] S. Oyama, "A sensitive period for the acquisition of a nonnative phonological system," *Journal of Psycholinguistic Research*, vol. 5, no. 3, pp. 261–283, 1982, doi: https://doi.org/10.1007/bf01067377.

[31] G. Stevens, "Age at immigration and second language proficiency among foreign-born adults," *Language in Society*, vol. 28, no. 4, 1999. doi: https://doi.org/10.1017/s0047404599004030.

[32] J. E. Flege and S. Liu, "The effect of experience on adults' acquisition of a second language," *Studies in Second Language Acquisition*, vol. 23, no. 4, pp. 527–552, 2001, doi: https://doi.org/10.1017/s0272263101004041.

[33] J. E. Flege and I. R. MacKay, "Perceiving vowels in a second language," *Studies in Second Language Acquisition*, vol. 26, no. 1, 2004, doi: https://doi.org/10.1017/s0272263104026117.

[34] J. E. Flege, "Factors affecting degree of perceived foreign accent in English sentences," *The Journal of the Acoustical Society of America*, vol. 84, no. 1, pp. 70–79, 1988, doi: https://doi.org/10.1121/1.396876.

[35] J. E. Flege and H. Winitz, "The production and perception of foreign language speech sounds," in *Human Communication and Its Disorders: A Review*, NJ: Ablex Publishing, 1992, Norwood, pp. 224–401.

[36] J. E. Flege and W. Strange, "Second language speech learning theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Baltimore: York Press, 1995, pp. 233–277.

[37] J. E. Flege and D. Birdsong, "Age of learning and second-language speech," in *Second Language Acquisition and the Critical Period Hypothesis*, Hillsdale, NJ: Lawrence Erlbaum, 1999, pp. 101–132.

[38] C. T. Best and W. Strange, "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Timonium, Maryland: York Press, 1995, pp. 171–204.

[39] H. Lee, O. Susan, and V. Martin, "Developing the oral proficiency of Korean university students through communicative interaction," in *Teaching English to Koreans*, Hollym Corporation Publishers, 2003, pp. 29–48.

[40] E. Cho, "Washback on the CSAT English test on high school students' language learning," dissertation, 2010.

[41] Y. J. Kang, T.-J. Yoon, and S. W. Han, "Frequency effects on the vowel length contrast merger in Seoul Korean," De Gruyter Mouton, October 1, 2015, doi: https://doi.org/10.1515/lp-2015-0014

[42] A. C. Gimson, *An Introduction to the Pronunciation of English*, 4th ed. London: Arnold, 1989.

[43] "Millisecond Software." Inquisit 6. [Online]. Available: https://www.millisecond.com/

[44] J. E. Flege, O.-S. Bohn, and S. Y. Jang, "Effects of experience on non-native speakers' production and perception of English vowels," *Journal of Phonetics*, vol. 25, no. 4, pp. 437–470, 1997, doi: https://doi.org/10.1006/jpho.1997.0052.

Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).