*SUVANNABHUMI* Vol. 14 No. 1 (January 2022) https://doi.org/10.22801/svn.2022.14.1.193

# The Perception and Production of Vietnamese Tones by Japanese, Lao and Taiwanese Second Language Speakers

Đào Mục Đích* · Anh Thư T. Nguyên**

**[ *Abstract* ]**

This study investigates the production and perception of Vietnamese tones by Japanese, Lao, and Taiwanese second language (L2) learners [n=30], comparing their performance in an Imitation task to that of Identification and Read-Aloud tasks. The results show that the Imitation task is generally easier for L2 speakers than the Identification and Read-Aloud tasks, suggesting that imitation is performed without some of the skills required by the other two tasks. It is also found that Lao and Taiwanese speakers outperform Japanese speakers, suggesting that prior experience with one tone language facilitates the acquisition of tone in another language. The result on speakers' tonal range show that L2 leaners have significantly narrower tonal F0 range than control Vietnamese speakers [n=11]. The results of error pattern analysis and tonal transcription also suggest that non-modal voice (glottal stop and creakiness) and contour tones (bidirectional fall-rise) are more difficult for L2 learners than modal voice tones (e.g., unidirectional contours: rising, falling, and level).

**Keywords:** Second language imitation, perception, production, Vietnamese tones, Japanese, Lao, Taiwanese Mandarin

* Lecturer, University of Social Sciences and Humanities, Vietnam National University - Ho Chi Minh City, dichvns@hcmussh.edu.vn

** Researcher, Mountain Creek, QLD 4557, Australia, thunguyen1972@gmail.com

## Ⅰ. Introduction

Research indicates that the acquisition of nonnative lexical tones is difficult for adult second language learners (So 2005; Wang, Spence, Jongman and Sereno 1999; Wayland and Guion 2004). Studies on the effect of speakers' L1 prosodic experience on the acquisition of nonnative tonal systems yield conflicting results. On the one hand, it has been shown that prior experience with one tone language may facilitate the acquisition of tone in another language. For example, in a perceptual training study on Thai tones, Wayland and Guion (2004) discover that native Chinese speakers (Taiwanese and Mandarin) significantly outperformed native English speakers in the discrimination and identification of Thai mid-level and low-level tone contrasts after a brief period of perceptual training. This is because the ability to track the change of F0 values, movement, and the direction of the movement at word level with one tone language, may be transferable to the discrimination of tones in an unfamiliar tonal system. On the other hand, conflicting results have also been reported, showing that direct experience with a lexical tone system failed to facilitate learning. For example, in a study of native Cantonese and Japanese speakers' perception of Mandarin tones, So (2005) reveals that native Cantonese speakers consistently demonstrated greater difficulties in distinguishing Mandarin tone 1 (high level) - tone 4 (falling) and tone 2 (rising) - Tone 3 (falling-rising) contrasts than native Japanese speakers before and after a brief period of training. Given the fact that Cantonese speakers have prior experience with lexical tone contrasts in their first language, while Japanese speakers only have prosodic experience with pitch and accent at the phrasal level in their L1, native Cantonese speakers' direct experience with a lexical tone system failed to facilitate learning, at least at the initial stage. In a recent study, Tsukada and Kondo (2018) examine the perception of Mandarin lexical tones by native speakers of Burmese who use lexical tones in their first language but are naïve to Mandarin. Unlike Mandarin tones, which are primarily cued by pitch, Burmese tones are cued by phonation type as well as pitch. Burmese listeners' discrimination accuracy was compared with that of Mandarin listeners and Australian English listeners to investigate

whether Burmese listeners can utilize their L1 experience in processing unfamiliar Mandarin tones. Their main findings were: 1) Mandarin listeners were more accurate in discriminating all tone pairs than non-native listeners; 2) Australian English listeners naïve to Mandarin were more accurate than similarly naïve Burmese listeners in discriminating all tone pairs except for T2-T4; and 3) Burmese listeners had the greatest trouble discriminating T2-T3 and T1-T2. Taken together, their results suggest that merely possessing lexical tones in L1 may not necessarily facilitate the perception of non-native tones, and that the active use of phonation type in encoding L1 tones may have played a role in Burmese listeners' less than optimal perception of Mandarin tones.

These inconsistent findings suggest that the relationship between learners' sensitivity to lexical tones or pitch accent, because of first language experience and its effect on learning a new tonal system, is not straightforward, at least at initial stages of learning. This study extends these findings by aiming to understand how different L1 experiences with prosodic features (pitch accent: Japanese and lexical tones: Lao and Taiwanese) affects L2 prosodic speech production and perception by examining the Vietnamese tone patterns of Japanese, Lao and, Taiwanese learners. The findings of this study will have an original and significant contribution. Firstly, it presents a novel comparison: the acquisition of Vietnamese as an L2, particularly by Japanese, Lao, and Taiwanese speakers, is still understudied. Secondly, it contributes to the understanding of the process and nature of second language acquisition and has implications to the teaching of Vietnamese as a second or foreign language.

### 1.1. L2 production models

Theories that account for the production of L2 speech, particularly L2 intonation, have been proposed. According to the *Markedness Differential Hypothesis* (MDH) (Eckman 1977), a language learner's areas of difficulty can be predicted, such that (1) those areas of the target language (TL) which differ from the native language (NL) and are more marked than the native language will be difficult; (2) the relative degree of difficulty of the areas of difference of the target

language which are more marked than the native language will correspond to the relative degree of markedness; and (3) those areas of the target language which are different from the native language, but are not more marked than the native language, will not be difficult. The MDH claims that NL-TL differences were necessary for explaining L2 learning difficulty only on the basis of differences between the native language and target language, but they were not sufficient; rather, one needed to incorporate typological markedness into the explanation. The hypothesis asserts that within the areas of difference between the NL and TL, marked structures are more difficult than corresponding unmarked structures. What follows immediately from this hypothesis is that not all NL-TL differences will cause equal difficulty. TL structures that are different from the NL but are not related by markedness principles to any other structures are predicted to cause no difficulty, while TL constructions which are related to other representations by markedness principles are predicted to cause learning problems. The degree of difficulty involved is predicted to correspond directly to the relative degree of markedness. It is predicted that in Vietnamese, non-modal voice (glottal stop and creakiness) and contour tones (bidirectional fall-rise) are more marked and thus are more difficult for L2 learners than modal voice tones (e.g., unidirectional contours: rising, falling, and level).

### 1.2. L2 perception of tones

Cross-linguistic studies were conducted to examine how speakers of tonal and nontonal languages differ in the perceptual processing of tones. With respect to F0 height and contour, research shows that the perceptual weights of these two dimensions are related to the linguistic experience of the listeners. Gandour (1983), using multidimensional scaling, examined the perception of tones by listeners of four tonal languages, including Mandarin, Cantonese, Taiwanese, and Thai, as well as by those of a non -tonal language, English. He observed that English listeners attached more importance to the height, and less to the contour dimension of F0 than did listeners of most tonal languages. Gandour reiterates that since English has no contrastive tones, English listeners directed their attention almost exclusively to F0 height. Lee, Vakock and

Wurm (1996) extend this research using a tone discrimination task. Both Cantonese and Mandarin tones were presented to Cantonese, Mandarin, and English listeners. They found that tone language speakers were better at discrimination of tones, in terms of speed and accuracy of their responses, than were non-tone language speakers. Tone language speakers seemed to acquire general tone discrimination abilities. Thus, it appears that listeners' strategy for tone perception depends to some extent on the linguistic function of pitch in their native language.

### 1.3. L2 imitation

In addition to studies on production and perception of L2 tones, there were also explorations on the link between these two domains. Many studies employ a speech imitation task, where performers listened to an audio stimulus and reproduced it (Flege and Eefting 1988; Fowler, Brown, Sabadini and Weihing 2003; Mitterer and Ernestus 2008; Mitterer and Müsseler 2013; Shockley, Sabadini and Fowler 2004). By comparing the target stimulus and the performers' reproductions, these studies conceive of a processing route from speech perception to production and establish the relations between them. L2 imitation merits thorough investigation not only because it permits an examination of both L2 perception and production, but also illustrates how these two modalities are coordinated in a single act. Additionally, since L2 perception and production have often been found to diverge from each other (e.g. Baker and Trofimovich 2006; Bohn and Flege 1997; Bradlow, Pisoni, Akahane-Yamada and Tohkura 1997), a comparison of second language learners' imitation with their perception and production patterns would more clearly reveal the relative contribution of perception, production, and other processing mechanisms in speech imitation. In a recent study, Hao and de Jong (2016) observed English speakers' learning of Mandarin tones, comparing their performance in an imitation task to that of identification and read-aloud tasks. The results show that the imitation task was generally easier for English speakers than the identification and read-aloud tasks, suggesting that imitation was performed without some of the skills required by the other two tasks. In an extension of these findings, in this study, we examine

Japanese, Lao and Taiwanese speakers' performance in identification, read-aloud, and imitation of the six Vietnamese tones.

### 1.4. Prosodic aspects in Vietnamese, Japanese, Lao and Taiwanese

Vietnamese is a tonal language which uses pitch to distinguish lexical meaning. Its standard Northern dialect has six lexical tones (Table 1 below describes the F0 contour and phonation types of the six Northern tones). Vietnamese tone is superimposed on monosyllables. Central and Southern Vietnamese each have one tone less because the fifth and the sixth tones (Curve and Broken tones respectively) have merged (Brunelle 2009a, 2009b; Emeneau 1951; Kirby 2010). The six Northern Vietnamese tones combine complex pitch contours with voice quality distinctions (Brunelle 2009b; Kirby 2010; Michaud 2004; Michaud, Vu, Amelot and Roubleau 2006; Nguyen and Edmondson 1997). Voice quality, particularly the laryngeal features of glottal stop, creakiness, and breathiness are distinctive tonal features characterizing Vietnamese tones at the phonetic level across dialects. Glottal stop/glottaliazation and creakiness, in addition to occurring as a regular feature on the Broken and Drop tones of the Northern dialect and the Curve tone of the Central dialect, also occur on some local variants of the Southern Drop tone (Brunelle 2009a; Kirby 2010; Vu 1981). Checked syllables, syllables closed by voiceless stops, bear one of two additional tones, which are sometimes considered allotones of tones Rising and Dropping (they will not be addressed here). In perception, glottalization and direction of contour are the dominant cues in Northern Vietnamese (Brunelle and Jannedy 2013).

<Table 1> Tones of Northern Vietnamese

| Tones | Diacritics | F0 Contours and phonation |
|---|---|---|
| Ngang Level | a (unmarked) | Level F0 contour, slight declination toward the end. |
| Sắc Rising | á (acute accent) | Rising F0 contour, starts lower than the level tone, fairly level in the first third and rising to the upper end of the pitch range. |
| Huyền Falling | à (grave accent) | Falling F0 contour, begins on a low pitch and falls gradually, sometimes with breathy voice. |
| Nặng Drop | ạ (subscript dot) | A low-falling tone with strong final laryngealization. |

| Tones | Diacritics | F0 Contours and phonation |
|---|---|---|
| Hỏi Curve | å (question mark) | Concave F0 contour, sharp fall and short rise and/or creakiness |
| Ngã Broken | ã (the tilde) | Concave F0 contour, a falling-rising tone with laryngealization/glottal stop in the middle. |

Japanese (Tokyo dialect) exhibits lexical contrasts based on pitch accent; that is, there are minimal pairs of words that are identical segment-wise, but can be distinguished in terms of their pitch contours. While what kind of pitch contour a particular word shows is often unpredictable for many lexical words, there are many phonological and morphological environments where the distribution of lexical accent is predictable, at least to some extent. In other words, there are some regularities regarding the phonological distributions of Japanese pitch accent (Kawahara 2015). Pitch accent in Japanese is fundamentally a word-level property, not a phrasal or sentence-level property, although they interact non-trivially with sentence-level intonational patterns. Japanese makes lexical contrasts in terms of pitch accent in two ways: (1) presence versus absence of pitch accent; and (2) if present, accent location. Unlike in many other tonal languages (Yip 2002), Japanese lexically uses only two levels of tonal heights (High and Low). Phonetically speaking, an accented vowel is assigned a High tone followed by a Low tone on the following vowel, resulting in an abrupt H(igh)-L(ow) fall in f0, whereas unaccented words do not show such a fall. Japanese also distinguishes words in terms of where pitch falls; i.e., in terms of accent location.

The tonal descriptions of Vientiane Lao range from five to seven tones. Most often the tones are described separately for live syllables, short dead syllables, and long dead syllables (for different analyses see e.g., Gedney 1972; Brown 1976; Strecker 1979). In the tone-count, the tones of the live syllables are listed first (e.g. 1-6), and the dead-syllable tones are simply added (e.g. 7- 10). It is generally agreed that there are two falling contour tones, one low-rising tone, and at least two level tones (the mid and low tones). Most authors also agree that Vientiane Lao has six tones: (1) rise, (2) mid rise, (3) high-fall, (4) mid-lower level contour, (5) low-level w/ glottalization, and (6) low-fall (Morev et al. 1979). By

comparing stressed tones in different prosodic contexts, Gårding and Svantesson (1994) emphasize the phonological analysis which indicates that live syllables have six contrastive tones and regard the tones of dead syllables as contextual variants of the live ones. Morev et al. (1979) reveal that tone 5 (low-level) involved glottalization and by Osatananda (1997) that syllables ending with -j and -w, end up with either a creaky voice or a glottal stop, regardless of tone height and shape.

There are four lexical tones in Taiwanese Mandarin: Tone 1, 2, 3, and 4 — high-level, rising, dipping, and high-falling respectively. There is also a neutral tone, and its pitch contour is decided by the preceding lexical tone. In Standard Mandarin, a syllable with any of the four lexical tones can have a neutral tone when it is unstressed. These unstressed syllables can be found in the second syllables of some disyllabic compounds. However, some syllables, mostly suffixes and particles, always have a neutral tone. Unlike Standard Mandarin, the neutral-tone syllables of Taiwanese Mandarin do not undergo tone loss. The neutral tone in Taiwanese Mandarin has a mid-low pitch target in disyllabic words, neutral-tone sequences, and novel formations (Huang 2012). Data show that Dipping Tone 3 (T3) is often accompanied by creaky phonation, which may at least in part be due to the already low F0 in the middle of the tone (Belotel-Grenié and Grenié 1994, 2004; Davison 1991; Kuang 2013). In fact, it has also been discovered that the Falling Tone 4 (T4) may exhibit some creaky phonation (Belotel-Grenié and Grenié 1994; Kuang 2013), again likely related to the low F0, and in this case, at the end of the tone. Moreover, with regard to T3, findings suggest that creaky phonation may, in fact, serve to enhance the perception of this tone (Yang 2011, 2015; Kuang 2013).

In a recent study, Đào and Nguyễn (2019) investigated the production and perception of Vietnamese tones by Korean second language learners [n=11], comparing their performance in an Imitation task to that of Identification and Read-Aloud tasks. The results show that the Imitation task was generally easier for Korean speakers than the Identification and Read-Aloud tasks, suggesting that imitation was performed without some of the skills required by

the other two tasks. The results on tonal F0 range and speakers' tonal range show that Korean leaners have significantly narrower tonal F0 range than control Vietnamese speakers. The results of error pattern analysis and tonal transcription in this study also suggest the effects of phonetic realizations of lexical tones in Vietnamese that are in interaction with language transfer from Korean phonology.

In brief, studies show that in addition to pitch height and pitch contours, voice quality, particularly the laryngeal features of glottal stop and creakiness, is a distinctive tonal feature characterizing tones at the phonetic level across three lexical tone languages, Vietnamese, Lao, and Mandarin. By contrast, phonation as a distinctive feature has not been reported for Japanese as a pitch accent language. Therefore, it is predicted that Japanese would have more difficulties acquiring the "marked" non-modal voice (glottal stop and creakiness) and contour tones (bidirectional fall-rise) such as Broken and Curve than Lao and Taiwanese speakers.

### 1.5. Study aims

The goal of this study is to understand how different L1 experiences with prosodic features (pitch accent: Japanese; and lexical tones: Lao and Taiwanese) affects L2 prosodic speech production and perception by examining the Vietnamese tone patterns of Japanese, Lao, and Taiwanese learners. The study addresses four research questions:

1. Does perception or production exert a stronger influence on imitation of tones?

2. What are the general error patterns of Vietnamese tones by Japanese, Lao, and Taiwanese learners?

3. How are tonal features (i.e., tonal range, tonal contours, and voice quality) produced by Japanese, Lao, and Taiwanese learners different from those produced by Vietnamese control speakers?

4. How do speakers' different L1 experiences with prosodic features (pitch accent versus lexical tones) affect L2 prosodic speech production and perception of Vietnamese tones? More specifically,

do Lao and Taiwanese speakers outperform Japanese speakers in perception and production of Vietnamese tones?

We hypothesize that if imitation employs phonological encoding [i.e., the process of retrieval of segmental and prosodic information, the generation of a syllabified phonological word, and the computation of the phonetic form of the intended utterance (Levelt, Roelofs and Meyer 1999)], in addition to perception and production, it would require the acquisition of more skills than those used in either Identification or Read-Aloud. Table 2 below shows how the three tasks match up with phonological encoding, perception, and production. Consequently, the participants' performance in the Imitation task should be constrained by their accuracy in the other two tasks. On the other hand, if imitation bypasses some aspects of phonological encoding, the learners' accuracy in imitation would not necessarily fall behind the other two tasks, but might actually be better. In addition to examining the learners' relative accuracy, their error distributions in the three tasks were compared to identify the sources of difficulty in their imitation of L2 sounds. To the extent that Identification and Read-Aloud tasks create different error patterns, the similarity between the learners' error patterns in the Imitation task and those in Identification and Read-Aloud tasks would reveal whether their performance in imitation has a predominantly perceptual or articulatory basis, or whether it is a joint effect of perceptual errors compounded by production inaccuracy (Hao and de Jong 2016).

<Table 2> How the three tasks match up with phonological encoding, perception, and production. * means “can be bypassed”

| **Identification** | **Imitation** | **Read-Aloud** |
|---|---|---|
| Auditory perception | Auditory perception (Phonological encoding) * Motor production | Phonological encoding Motor production |

## Ⅱ. Method

### 2.1. Participants

A control group of 11 Northern Vietnamese (Hanoi) speakers (7 female, 4 male) was included. They were international students at Macquarie University and had lived in Australia from 6 months to 1 year. Their average age is 35.3 (SD=7.2). The reasons why Northern Vietnamese (Hanoi) speakers were chosen is because the L2 learners learned Vietnamese with instructors of Northern (Hanoi) dialect. They all spoke Vietnamese with a Northern accent. The Northern Vietnamese tone system has a voice quality, a "marked" feature predicted to be difficult for L2 learners to acquire.

The L2 learners of Vietnamese consist of three groups: Japanese, Lao, and Taiwanese. Ten native Japanese speakers (6 females, 4 males), ten native Lao speakers (5 females, 5 males), and ten native Taiwanese Mandarin speakers (5 males, 5 females) were recruited from students/learners of the Department of Vietnamese Studies, University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City. They have lived in Vietnam for 6 months to 1 year. The Japanese speakers all came from Tokyo. Their average age is 23 years old (SD=1.7) and their average length of learning Vietnamese is more than one year (mean=14.3 months). The Lao speakers all came from the Vientiane Capital of Lao. Their average age is 22 years old (SD=1.5) and their average length of learning Vietnamese is more than one year (mean=15.6 months). The Taiwanese speakers' average age is 25 years old (SD=1.8) and their average length of learning Vietnamese is more than one year (mean=14.6 months). In the intermediate Vietnamese language courses, Japanese, Lao, and Taiwanese or other foreign learners basically learn subjects such as Vietnamese vocabulary, Vietnamese grammar, Vietnamese culture, communication skills [listening comprehension (in daily situation, radio and television)], conversations, reading (newspapers and formal documents), and writing (informal and formal). Because the L2 leaners started learning Vietnamese at the average age of 18.5 years, they can be considered late learners of L2. Also, since they were studying intermediate Vietnamese language courses, their level of Vietnamese can be considered as intermediate level.

### 2.2. Stimuli

The experiment used open syllables with the initial stop consonant /t-/ and the nine Vietnamese vowels /i/, /e/, /ɛ/, /ɯ/, /ɤ/, /a/, /u/, /o/, /ɔ/. These vowels were then embedded in /t_/ carrier words. Each word independently carried one of the six Northern Vietnamese tones (see Table 3). The total number of items included 9 simple vowels x 6 tones, totalling 54 items. The syllables used in the study are all legal syllables, most of which are high frequency words and thus were familiar to the participants because they appear in lesson materials and are used in everyday classroom speech of the learners.

Since the L2 learners learned Vietnamese with instructors of Northern (Hanoi) dialect, one male native speaker of Hanoi Vietnamese produced all the stimuli for the Identification and Imitation tasks, which were recorded at 44.1kHz using the built-in microphone of a laptop and Praat (Boersma and Weenink 2009). The stimuli were randomized in one block with an inter-stimulus interval of 6 seconds. The total duration of the block was 13 minutes. The same stimuli were presented in written form plus tone marks via Power point slides for the Read-Aloud task.

<Table 3> Vietnamese stimuli - Tones of Vietnamese

| ***Tones*** / ***Words*** | ***Level tone*** | ***Falling tone*** | ***Curve tone*** | ***Broken tone*** | ***Rising tone*** | ***Dropping tone*** |
|---|---|---|---|---|---|---|
| ti /ti/ | ti | tì | tỉ | tĩ | tí | tị |
| tê /te/ | tê | tề | tể | tễ | tế | tệ |
| te /tɛ/ | te | tè | tẻ | tẽ | té | tẹ |
| tư /tɯ/ | tư | từ | tử | tữ | tứ | tự |
| tơ /tɤ/ | tơ | tờ | tở | tỡ | tớ | tợ |
| ta /ta/ | ta | tà | tả | tã | tá | tạ |
| tu /tu/ | tu | tù | tủ | tũ | tú | tụ |
| tô /to/ | tô | tồ | tổ | tỗ | tố | tộ |
| to /tɔ/ | to | tò | tỏ | tõ | tó | tọ |
| tia/ tie/ | tia | tìa | tỉa | tĩa | tía | tịa |
| tưa /tɯɤ/ | tưa | từa | tửa | tữa | tứa | tựa |
| tua /tuo/ | tua | tùa | tủa | tũa | túa | tụa |

### 2.3. Procedures

#### 2.3.1. Identification task

The participants sat individually in a quiet room and listened to the

stimuli through a laptop. They were provided an answer sheet for all the stimuli without tone marks and were instructed to mark the tone of every syllable of the stimuli.

### 2.3.2. Imitation

The same audio stimuli used in the Identification task were randomized in a different order. The participants listened to each stimulus once through head phones and were asked to repeat it without any visual aid. Their responses were recorded using Praat on a laptop.

### 2.3.3. Read-Aloud task

The participants were asked to read aloud 54 stimuli presented on Powerpoint slides (one word for each slide) at their own pace. The order of the stimuli was randomized in a different order from the other tasks. Their responses were recorded using Praat on a laptop.

All the participants completed the Read-Aloud task last, while the order of the Identification and Imitation tasks was counterbalanced to avoid order and learning effects. There was a delay (a short break of 10 minutes) before the Imitation task started. The conditions of the experiments for control group and the L2 learners' groups were identical/ compatible.

### 2.3.4. Assessing accuracy in the Read-Aloud and Imitation tasks

The recordings were judged blindly by two phonetically trained native speakers of Vietnamese. They identified the tonal errors made by the participants. The two native speakers labelled the tone of each syllable/word with a choice among six lexical tones. When there was any disagreement between them, acoustic analysis was carried out using Praat with visual pitch contour to decide the label of the tone. The two native judges agreed on most of the tokens (the mean inter-rater agreement across three L1 languages was 95% for the Imitation task and 88% for the Read-Aloud task), and their divergence appeared to reflect ambiguity in the productions.

### 2.3.5. Analysis

To answer the first research question, L2 learners' performance in

Identification, Read-Aloud, and Imitation of the six Vietnamese tones were compared to examine whether perception or production exerts a stronger influence on the imitation of tones. The accuracy rates for each participant were calculated for each task (Identification, Read-Aloud, and Imitation) and tone (Level, Rising, Falling, Dropping, Curve, and Broken). The individual learners' accuracy rates were first transformed into Rationalized Arcsine Units (RAU) to make them more suitable for statistical analysis (Hao and de Jong 2016; Studebaker 1985). The rationalized arcsine transform was used to transform data obtained from speech related studies in order to make them suitable for parametric statistical analyses. The arcsine transformation expresses scores in radians and the rationalized arcsine transform adjusts these scores into units resembling percentages, making them easier to interpret. Formula (1) was used in most speech test cases, then the number of trials (N) was taken into account. In the following, S is the number of correct responses and N is the number of trials performed. Equation (2) converts radians into RAU.

(1) $\mathrm{AU} = {}^{asin}\sqrt{\frac{S}{N+1}} + {}^{asin}\sqrt{\frac{S+1}{N+1}}$

(2) $\mathrm{RAU} = (\frac{146}{\pi})*\mathrm{AU}\text{-}23$

The resulting RAU values were compared in a three-way mixed-effects model to determine the relative accuracy of the three tasks. The dependent variable was RAU. The fixed effects were L1 Languages (Japanese, Lao and Taiwanese), Tasks (Identification, Read-Aloud, and Imitation) and Tones (Level, Rising, Falling, Dropping, Curve, and Broken). The random factors were Items (54 words x 3 tasks= 162 monosyllabic words) and Speakers (10 Japanese, 10 Lao, and 10 Taiwanese speakers). A Tukey post-hoc test was then conducted to determine the significant differences among the levels of the main fixed effects. The results were reported in section 3.1.

Then, accuracy rates in the Imitation task were compared with those of Identification and Read-aloud tasks via Pearson correlation analyses on the RAU scores to examine whether the difficulty of the Imitation task was a result of perceptual imprecision, production

inaccuracy, or a combination. If perception was a more dominant factor in determining the imitation performance, the learners' accuracy patterns in the Imitation task should be most similar to those in Identification. Alternatively, if production exerted a stronger influence on imitation, the Read-Aloud and Imitation tasks should share a more correlated pattern. Results are reported in table 6.

In order to answer the second research question, the learners' mean percentage accuracy and error rates for the six Vietnamese tones in the three different tasks were calculated and summarized in confusion matrices. Results were reported in tables 7, 8 and 9.

In order to answer the third research question, acoustic analysis was performed on the Imitation and Read-Aloud data of four groups of speakers (three groups of L2 learners and control Vietnamese). The key acoustic parameters included F0 (Fundamental frequency) maximum and F0 minimum for each tone, the Fundamental frequency (F0 in Hz) at five equidistant points on the tone contour of each syllable rime, and a qualitative transcription of the F0 contours and voice quality. The F0 was measured because it is one of the key acoustic correlates that represent the tonal contour. The tones of the target words were segmented and a script was used to extract the F0 min, F0 max and F0 values at five equidistant points of the tonal contours of the target words (Boersma and Weenink 2017). Results were visually inspected for f0 doubling and halving: suspicious values were simply excluded.

Then the speakers' tonal range at five equidistant points on the tone contour was calculated by subtracting the F0 value of the highest tone from the lowest tone in the tonal contours for each speaker. In addition, the tone contours and voice quality (i.e., glottal stop and creakiness) of the target words were transcribed qualitatively with help from spectrographic and F0 display using the symbols shown in Table 4. The glottal stop is caused by a complete closure of the vocal folds, acoustically represented by irregular widely spaced pulses and gap in the spectrogram (such as at the end of Dropping tone and in the middle of Broken tone), while an incomplete closure of the vocal folds results in creakiness or mild laryngealization with fewer irregular pulses, such as at the end of Curve tone (Michaud 2004; Pham 2003). Results are reported in

section 3.4 and figures 3.4 and 5. Due to limited space, only results of the three most problematic tones (Broken, Curve, and Dropping) are reported since L2 speakers across three L1 languages generally performed well (with more native-like contours) on the three other tones (Level, Falling, and Rising: a mean of 85% and above).

<Table 4> Symbols and descriptions of tone contours and voice quality of the target words as shown on the spectrographic and F0 display

| Symbols | Contour descriptions | Symbols | Contour descriptions |
|---|---|---|---|
| R | sharp rise | FRc | sharp fall sharp rise and creaky |
| r | slight rise | Fg | sharp fall end with a glottal stop |
| F | sharp fall | Fc | sharp fall with creakiness at the end |
| f | slight fall | L | level |
| FR | sharp fall sharp rise | LR | level sharp rise |
| fR | slight fall sharp rise | Lr | level slight rise |
| fr | slight fall slight rise | LcR | level creaky sharp rise |
| Fr | sharp fall slight rise | cF | creaky sharp fall |
| FgR | sharp fall glottal stop and sharp rise | cRc | creaky sharp rise creaky |
| FcF | sharp fall creaky sharp fall | c | creaky |

A mixed-effects model was performed on the speakers' tonal F0 range as a dependent variable. The fixed effects included 7 groups (Vietnamese, Japanese Imitation, Japanese Read-Aloud, Lao Imitation, Lao Read-Aloud, Taiwanese Imitation, and Taiwanese Read-aloud) and genders (male vs. female). Since female pitch (F0) is normally higher than that of male speakers, we included gender as a factor and plotted F0 values of female and male speakers separately (Figure 6). The random factors included speakers (11 control Vietnamese and 30 L2 learners in the Imitation task and the Read-Aloud task) and items (54 words x 41 speakers =2214 words).

## Ⅲ. Results

### 3.1. L2 learners' accuracy rates in Identification, Read-Aloud, and Imitation tasks

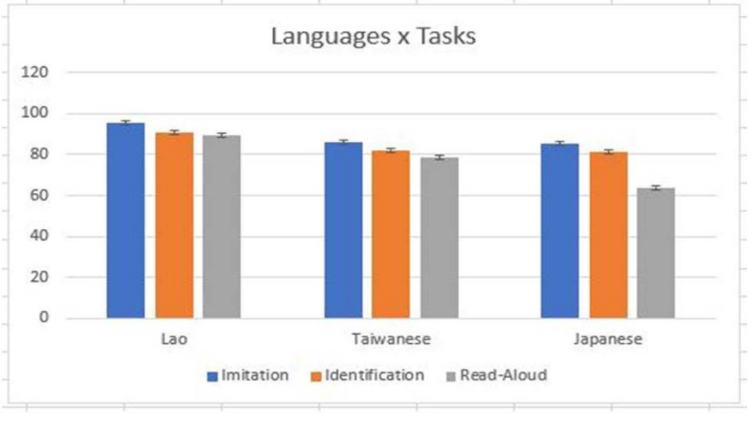
As shown in Table 5, the results of the mixed-effect model show that

all three main factors (Languages, Tasks, and Tones) are significant ($p<0.0001$). The interaction effects (Languages x Tasks, Languages x Tones and Tasks x Tones) are also significant ($p<0.05$). The three-way interaction effect (Languages x Tasks x Tones) is not significant.

<Table 5> F values and significant levels of the mixed effect model

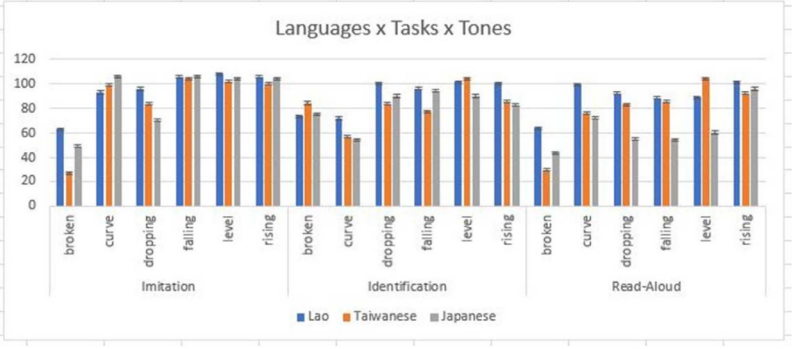
| Factors | F values and sig. levels |
|---|---|
| Languages | $F(2, 446)=12$, $p<0.0001$ |
| Tasks | $F(4, 438)=7$, $p<0.0001$ |
| Tones | $F(5, 438)=25.6$, $P<0.0001$ |
| Languages x Tasks | $F(2,438)=5.4$, $p<0.01$ |
| Languages x Tones | $F(10, 438)=2.1$, $p<0.05$ |
| Tasks x Tones | $F(20, 438)=4.6$, $p<0.0001$ |
| Languages x Tasks x Tones | $F(10, 438)=1.0$, $p=0.36$ ns. |

As shown in figure 1, the participants were more accurate in the Imitation task, followed by Identification and least accurate in Read-Aloud tasks across the three L1 speaker groups ($p<0.001$). Figure 1 also shows that compared to Japanese speakers, Lao and Taiwanese learners performed much better in all three tasks. A post-hoc pairwise analysis showed that Lao learners' accuracy RAU scores were significantly higher than those of Taiwanese ($p<0.001$) and Japanese ($p<0.0001$). The Taiwanese also significantly outperformed the Japanese across three tasks ($p<0.05$).




<Figure 1> Mean RAU values of Languages x Tasks. Y-axis: Mean RAU values

As shown in Figure 2, Broken and Curve tones are the least accurately performed for Identification and Read-Aloud tasks across three L1 speaker groups ($p<0.0001$). In the Imitation task, Broken tone was the worst performed across three L1 speaker groups ($p<0.0001$). In contrast, Level, Rising, and Falling tones are ones most accurately perceived and produced across three L1 speaker groups ($p<0.01$). The Dropping tone was more successfully imitated and identified than read aloud ($p<0.01$). Also, Lao speakers were more successful in imitating Broken tone and in reading aloud Broken and Curve tones than the other two groups.




<Figure 2> Mean RAU values of Languages x Tasks x Tones. Y-axis: Mean RAU values

<Table 6> Correlation results

| Tasks | Japanese | Lao | Taiwanese |
|---|---|---|---|
| Imitation vs. Read-Aloud | r=0.37, p<0.005 | r=0.47, p<0.0001 | r=0.65, p<0.0001 |
| Imitation vs. Identification | r=0.037, p=0.78 ns. | r=0.35, p<0.01 | r=0.045, p=0.76 ns. |
| Identification vs. Read-Aloud | r=0.034, p=0.79 ns. | r=0.45, p<0.0001 | r=0.20, p=0.15 ns. |

As shown in table 6, Pearson correlation analyses on the accurate RAU scores reveal that the correlation between Imitation and Read-Aloud task was significant at $p<0.005$ across three L1 language speakers. On the contrary, the correlation between Imitation and Identification did not reach significance for Japanese and Taiwanese speakers. The correlation between Read-Aloud and Identification was significant ($p<0.0001$) for Lao speakers only.

### 3.2. Tonal error patterns

#### 3.2.1. Japanese speakers

As shown in Table 7, in the Imitation task, there were two main patterns of error. The first was between Broken and Curve tones (47% of Broken was mispronounced in Curve tone). The second was between Dropping and Falling (30% of Dropping was mispronounced as Falling).

In the Read-Aloud task, the main confusion involved Broken and Curve (37% of Broken was confused as Curve), and between Curve and Rising (21% of Curve was read as Rising). Dropping was confused as Level (16%), Rising (20%) and Falling (7%). Falling was either mispronounced as Level (13%) or Rising (24%). Level was read as Rising (33%).

In the Identification task, the main pattern of errors was also between Broken and Curve (10% Broken was confused as Curve and 20% of Curve was misidentified as Broken). Curve was also identified as Rising (21%). In addition, Dropping was confused as Falling (6%) and Level (4%).

<Table 7> Proportional confusion matrix for 6 Vietnamese tones in the three tasks by Japanese speakers. The leftmost column indicates the target sounds, whereas the top row classifies the responses. Numbers in the cells represent the percentage of each response. The accurate responses are in bold. The prominent error pattern is italicized.

| Imitation | | | | | | |
|---|---|---|---|---|---|---|
| | broken | curve | drop | falling | level | rising |
| broken | **51** | *47* | 0 | 0 | 0 | 2 |
| curve | 2 | **98** | 0 | 0 | 0 | 0 |
| dropping | 0 | 0 | **70** | *30* | 0 | 0 |
| falling | 0 | 0 | 0 | **100** | 0 | 0 |
| level | 0 | 0 | 0 | 0 | **99** | 1 |
| rising | 0 | 0 | 0 | 0 | 0 | **100** |
| Read-Aloud | | | | | | |
| | broken | curve | drop | falling | level | rising |
| broken | **46** | *37* | 0 | 0 | 0 | 18 |
| curve | 1 | **73** | 0 | 2 | 2 | *21* |
| dropping | 0 | 3 | **54** | 7 | *16* | *20* |
| falling | 0 | 4 | 1 | **57** | *13* | *24* |

| | | | | | | |
|---|---|---|---|---|---|---|
| level | 0 | 4 | 0 | 3 | **59** | *33* |
| rising | 0 | 4 | 0 | 0 | 2 | **93** |
| | | | Identification | | | |
| | broken | curve | drop | falling | level | rising |
| broken | **83** | *10* | 0 | 0 | 0 | 7 |
| curve | *20* | **57** | 0 | 2 | 0 | *21* |
| dropping | 0 | 0 | **89** | 6 | 4 | 1 |
| falling | 0 | 0 | 4 | **92** | 2 | 1 |
| level | 0 | 0 | 0 | 2 | **89** | 9 |
| rising | 3 | 4 | 3 | 1 | 4 | **83** |

### 3.2.2. Lao speakers

As shown in Table 8, in the Imitation task, there were two main patterns of error. The first involved Broken and Curve tones (36% of Broken was mispronounced as Curve tone and vice versa; 12% of Curve was mispronounced as Broken). The second involved Dropping and Falling (8% of Dropping was mispronounced as Falling).

In the Read-Aloud task, the main confusion involved Broken and Curve (26% of Broken was confused as Curve; and vice versa, 6% of Curve was read as Broken). Dropping was confused as Falling (10%). Falling was either mispronounced as Level (7%) or Dropping (6%). Level was read as Falling (9%).

In the Identification task, the main pattern of errors was also between Broken and Curve (26% of Broken confused as Curve and 23% of Curve was misidentified as Broken). Curve was also identified as Falling (4%); and vice versa, Falling was confused as Curve (6%) and dropping (3%).

<Table 8> Proportional confusion matrix for 6 Vietnamese tones in the three tasks by Lao speakers. The leftmost column indicates the target sounds, whereas the top row classifies the responses. Numbers in the cells represent the percentage of each response. The accurate responses are in bold. The prominent error pattern is italicized.

| | | | Imitation | | | |
|---|---|---|---|---|---|---|
| | broken | curve | drop | falling | level | rising |
| broken | **63** | *36* | 0 | 0 | 0 | 1 |
| curve | *12* | **88** | 0 | 0 | 0 | 0 |
| dropping | 0 | 0 | **92** | *8* | 0 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| falling | 0 | 0 | 1 | **99** | 0 | 0 |
| level | 0 | 0 | 0 | 0 | **100** | 0 |
| rising | 1 | 0 | 0 | 0 | 0 | **99** |
| Read-Aloud | | | | | | |
| | broken | curve | drop | falling | level | rising |
| broken | **64** | *26* | 1 | 3 | 0 | 6 |
| curve | 6 | **92** | 0 | 2 | 0 | 0 |
| dropping | 0 | 2 | **88** | *10* | 0 | 0 |
| falling | 0 | 2 | 6 | **86** | 7 | 0 |
| level | 0 | 2 | 1 | 9 | **84** | 3 |
| rising | 0 | 1 | 0 | 2 | 1 | **96** |
| Identification | | | | | | |
| | broken | curve | drop | falling | level | rising |
| broken | **74** | *26* | 0 | 0 | 0 | 0 |
| curve | *23* | **70** | 0 | 4 | 1 | 1 |
| dropping | 0 | 1 | **97** | 2 | 0 | 0 |
| falling | 0 | 6 | 3 | **91** | 0 | 0 |
| level | 1 | 0 | 0 | 2 | **97** | 0 |
| rising | 3 | 1 | 0 | 0 | 0 | **97** |

### 3.2.3. Taiwanese speakers

As shown in Table 9, in the Imitation task, there were two main patterns of error. The first involved Broken and Curve tones (67% of Broken was mispronounced as Curve tone and vice versa; 6% of Curve was mispronounced as Broken). The second involved Dropping and Falling (17% of Dropping was mispronounced as Falling).

In the Read-Aloud task, the main confusion involved Broken and Curve (58% of Broken was confused as Curve and 8 % of Broken was read as Rising). Curve was confused as Falling (24%). Dropping was also read as Falling (18%); and vice versa, Falling was mispronounced as Dropping (14%).

In the Identification task, the main pattern of errors also involved Broken and Curve (11% of Broken was confused as Curve and 29% of Curve was misidentified as Broken). Curve was also confused as Rising (13%). Dropping was misidentified as Falling (15%). Falling was misidentified as Curve (17%). Rising was confused as Broken (11%).

In general, the results showed that L2 learners across the three L1 languages were most successful in producing and perceiving Level, Rising, and Falling tones. In contrast, they had problems perceiving Broken, Curve, and Dropping tones. In addition, they had

similar tonal error patterns, namely the confusion between Broken and Curve, and between Dropping and Falling.

<Table 9> Proportional confusion matrix for 6 Vietnamese tones in the three tasks by Taiwanese speakers. The leftmost column indicates the target sounds, whereas the top row classifies the responses. Numbers in the cells represent the percentage of each response. The accurate responses are in bold. The prominent error pattern is italicized.

| | Imitation | | | | | |
|---|---|---|---|---|---|---|
| | broken | curve | dropping | falling | level | rising |
| broken | **31** | *67* | 0 | 1 | 0 | 1 |
| curve | *6* | **94** | 0 | 0 | 0 | 0 |
| dropping | 0 | 0 | **83** | *17* | 0 | 0 |
| falling | 0 | 0 | 0 | **100** | 0 | 0 |
| level | 0 | 1 | 0 | 1 | **97** | 0 |
| rising | 0 | 4 | 0 | 0 | 0 | **96** |
| | Read-Aloud | | | | | |
| | broken | curve | dropping | falling | level | rising |
| broken | **33** | *58* | 0 | 0 | 0 | 8 |
| curve | 0 | **71** | 4 | *24* | 0 | 1 |
| dropping | 0 | 0 | **82** | *18* | 0 | 0 |
| falling | 0 | 1 | *14* | **85** | 0 | 0 |
| level | 0 | 0 | 0 | 0 | **100** | 0 |
| rising | 0 | 6 | 0 | 0 | *6* | **89** |
| | Identification | | | | | |
| | broken | curve | dropping | falling | level | rising |
| broken | **82** | *11* | 0 | 0 | 0 | 7 |
| curve | *29* | **57** | 0 | 1 | 0 | 13 |
| dropping | 0 | 1 | **83** | *15* | 0 | 0 |
| falling | 0 | *17* | 4 | **79** | 0 | 0 |
| level | 0 | 0 | 0 | 0 | **100** | 0 |
| rising | *11* | 4 | 0 | 0 | 0 | **85** |

### 3.3. Transcription of tone contours and voice quality

#### 3.3.1. The broken tone

As shown in Figure 3, the Vietnamese produced this tone with fall rise contour with a glottal stop in the middle (FgR: 50%; fgR: 20%) or a creakiness at the beginning (cR: 9%). In contrast, in the imitation task, the Japanese and Taiwanese learners failed to copy this feature and produced only fall rise contours (Japanese: FR: 19%; Taiwanese FR: 31%). This explains why native Vietnamese listeners heard L2 productions as a curve tone. Nevertheless, it is interesting that in the Read-Aloud task, many speakers of the three L1 groups

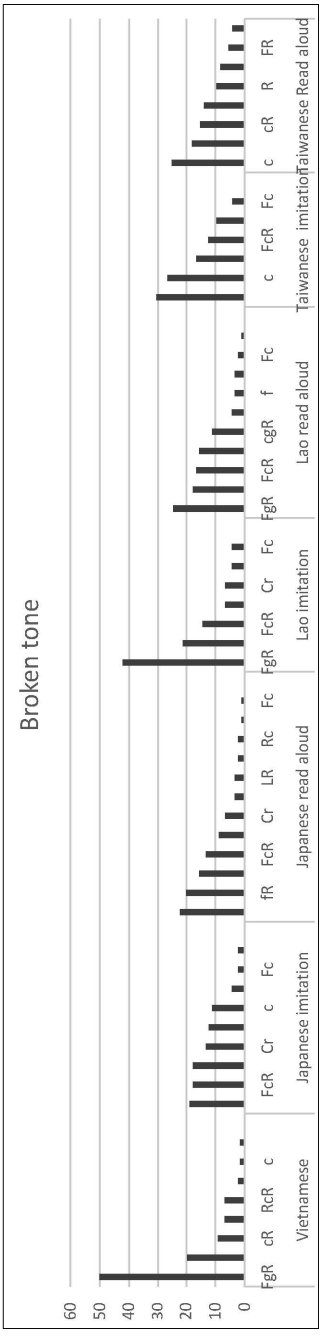
were able to produce the glottal stop in the middle of this tone contour (Japanese FgR: 22%; Lao FgR: 24%; and Taiwanese FgR: 18%) or the creakiness at the beginning of the tone contour (cR). This indicates that this unique voice quality feature may be learnable and that L2 learners can recognise this particular tone via visual written diacritic presented in the read aloud task, strongly suggesting the positive effect of formal teaching of tones, on Read-Aloud tasks.

### 3.3.2. The curve tone

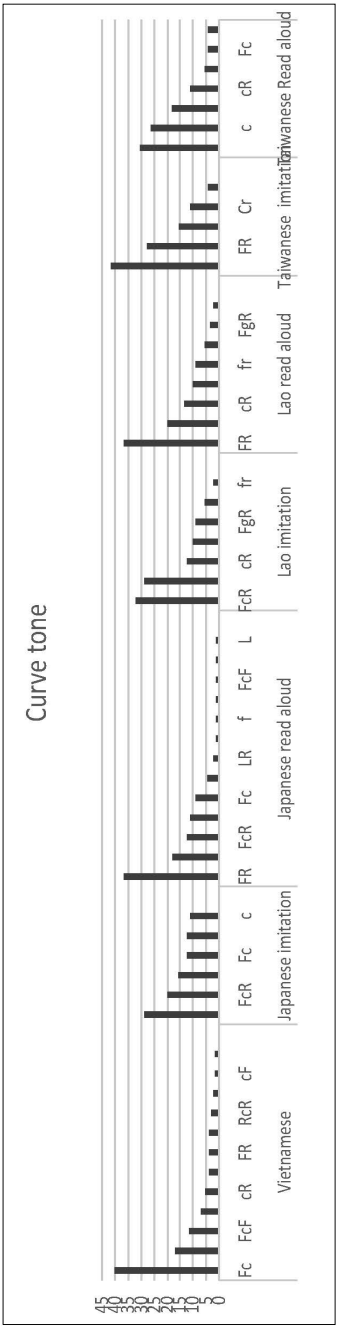
Figure 4 showed that native Vietnamese speakers produced this tone with either sharp fall contours that ended with creakiness or mild laryngealization (Fc: 40%); or a sharp fall sharp rise contours with creakiness in the middle (e.g., FcR: 17%; FcF:11%). In contrast, the L2 learners produced the curve tone only as a sharp fall sharp rise contour with no creakiness in the middle (Japanese Imitation: FR: 29%; Japanese Read-Aloud: FR: 37%; Lao Imitation FR: 29% Lao Read-Aloud 37%; Taiwanese Imitation: 28%; Taiwanese Read-Aloud: 31%). This shows that many L2 learners failed to employ voice quality on this tone, a distinctive feature in Northern Vietnamese tones. Nevertheless, some L2 learners could produce this creaky feature (Japanese Imitation: FcR: 20%; Japanese Read-Aloud: FcR: 12%; Lao Imitation FcR: 32%; Lao Read- Aloud: 20%; Taiwanese Imitation: c: 42%; FcR: 15%; Taiwanese Read-Aloud: c: 26%, FcR: 11%)

### 3.3.3. The dropping tone

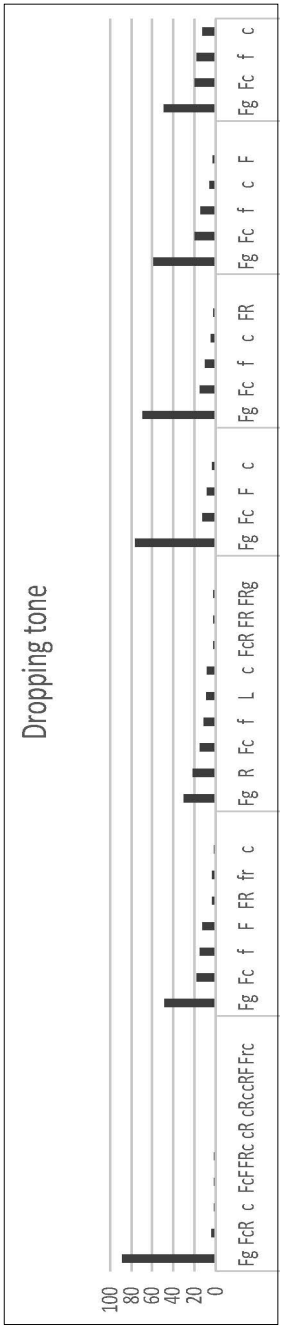
Figure 5 shows that while native Vietnamese speakers produced more sharp fall contours that ended with a glottal stop (Fg: 89%) for this tone, most of the L2 learners' contours also had a final glottal stop (Japanese Imitation: Fg: 48%; Japanese Read-Aloud: Fg: 30%; Lao Imitation: Fg: 77%; Lao Read-Aloud: 69%; Taiwanese Imitation: 58%; Taiwanese Read-Aloud: 49%). In contrast, some L2 learners produced the dropping tone as a sharp fall contour with no final glottal stop (Japanese Imitation: f: 14%; Japanese Read-Aloud: f: 11%; Lao Imitation: F: 8%; Lao Read-Aloud: f:10%; Taiwanese Imitation: f:14%; Taiwanese Read- Aloud: f: 18%). This is why the two native Vietnamese listeners perceived them as a falling tone (as shown in tonal error patterns in section 3.3 above).




<Figure 3> Percentage of frequency of tone produced – Broken tone



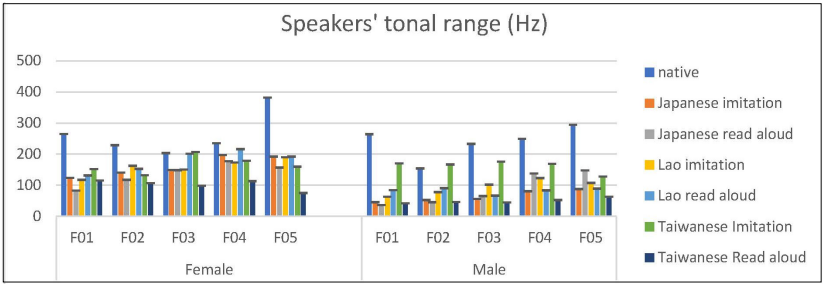

<Figure 4> Percentage of frequency of tone produced – Curve tone




<Figure 5> Percentage of frequency of tone produced – Dropping tone

### 3.4. Speakers' F0 range

The results of the mixed-effect model on the speakers' tonal range (F0: Hz) across five points of the tone contours show that the main factors of groups and the interactions of groups x genders are significant (groups [Vietnamese, Japanese Imitation, Japanese Read-Aloud, Lao Imitation, Lao Read-Aloud, Taiwanese Imitation, and Taiwanese Read-Aloud]: F (6, 51) = 8-32, $p<0.0001$, groups x genders: (F (6,51) = 2.7-11.6, $p<0.0001$) while gender is not significant (F (1, 24) = 0.4-2.5, p= 0.19 ns.) As shown in Figure 6, the L2 learners had a significantly narrower tonal F0 range than the control Vietnamese, across both male and female speakers and across two tasks, Imitation and Read-Aloud ($p<0.0001$). Taiwanese male speakers in Imitation task and Lao speakers in both Imitation and Read-Aloud tasks also have larger tonal range than other groups ($p<0.01$).




<Figure 6> Mean tonal F0 range (Hz) of speakers across five points of the tonal contours

## Ⅳ. Discussion and conclusion

In this section, we summarize and discuss the results by addressing the four research questions raised in section 1.5.

First, *does perception or production exert a stronger influence on imitation of tones?*

The results show that learners were significantly more accurate in imitating than in reading Vietnamese tones aloud or identifying

them. This indicates that L2 imitation may reduce the effect of phonological categorization. The learners' more accurate performance in the Imitation task also indicates that they do not lack the skills to correctly perceive or articulate Vietnamese tones. Their difficulty in the Identification and Read-Aloud tasks appears to come from the phonological encoding of tones, that is, the retrieval of segmental and tonal information and the attendant application of tone labels. This result is consistent with findings in previous studies (Hao and de Jong 2016; Hallé, Chang and Best 2004).

The positive correlation between Imitation and Read-Aloud tasks indicates that these two tasks share a common production error pattern: learners had problems producing Broken, Curve, and Dropping tones, suggesting that production exerts a stronger influence on imitation. It means that L2 learners seem to have difficulties with the motor production or articulation of these tonal features in both the imitation and the Read-Aloud tasks.

Second, *what are the general error patterns of Vietnamese tones by L2 learners?*

The analysis of the error patterns of the confusion matrices (Tables 7, 8 and 9) reveals that all the three tasks across three L1 languages shared a prominent error pattern, the mutual confusion between Broken and Curve tones, and between Dropping and Falling tones. One major type of errors was that Broken was produced as a Curve tone. That is, the Broken tone was produced with only falling and rising contours while it should have been a sharp falling- sharp rising contour with a glottal stop or a creaky feature in the middle of the tonal contour. Another common error involved Dropping and Falling tones. In Northern Vietnamese, the Dropping tone is sharp falling, short and marked by strong final laryngealization. L2 learners did not realize the final laryngealization of the tone and produced it as either a Falling tone or a Level tone. This result is strongly supported by the tone contour and voice quality transcription data. Particularly, many Japanese L2 learners failed to produce the glottal stop and/or creakiness in the Broken and Drop tones, thus making the two native Vietnamese listeners confuse many Broken as Curve tones and Dropping as Falling tones.

This can be said to be a negative transfer from their L1 since voice quality is not a distinctive feature in Japanese while it is in Vietnamese. Nevertheless, the Lao and Taiwanese speakers could produce this creaky feature which is also available in their tone systems, indicating a positive transfer from their L1s.

Furthermore, the confusion between Broken and Curve tones not only occurred in L2 learners but also in cross-dialectal data, e.g. Southern listeners identifying Northern tones (Kirby 2010). Brunelle and Jannedy (2013) showed that large numbers of Northern Vietnamese Broken tones were misidentified as Curve tone, which can be attributed to the fact that Broken and Curve are merged into a single tone in Southern Vietnamese and therefore, are not clearly distinct for the Southern Vietnamese subjects. In general, the results of this study suggested the effects of phonetic realizations of lexical tones in Vietnamese on L2 learners' perception and production.

Third, *how are tonal features (i.e., tonal range, tonal contours and voice quality) produced by L2 learners different from those produced by Vietnamese control speakers?*

The result on speakers' tonal range showed that L2 leaners had a narrower tonal F0 range than the Vietnamese. This mirrors our findings on heritage Vietnamese speakers, who either were born or who grew up in Australia, having more compact (narrower) tonal ranges than that of the contemporary Southern Vietnamese speakers in Vietnam (Đào and Nguyễn 2017) and Korean learners whose tonal ranges are narrower than control Vietnamese (Đào and Nguyễn, 2019). This is also in line with results by Tu et al. (2016) that Korean leaners generally have a narrower pitch range than Mandarin speakers, and similar to the results from Japanese speakers of L2 Mandarin (Tu et al. 2014) who also had narrower pitch range than control Mandarin speakers.

The results on tonal contour transcription showed that many L2 leaners failed to produce native-like tonal contours (i.e., their tonal contours were different from those of the control Vietnamese), indicating a deviation or underdevelopment of tonal distinctive features in L2 learners of lexical tones.

In addition, native Vietnamese listeners not only rely on tonal contours but also on voice quality feature in identifying tones as found in the two native Vietnamese listeners' perception patterns and consistent with results by Brunelle and Jannedy (2013). However, many L2 learners seem to ignore or fail to realize this feature, since they did not produce/copy a glottal stop in the middle of the Broken tone and creakiness or mild laryngealization (tense voice) at the end or in the middle of the Curve tone. This suggests that the voice quality feature is more "marked" (Eckman 1977; Hyman and VanBik 2004) and may take longer to acquire, in line with previous studies on heritage Vietnamese speakers (Đào and Nguyễn 2017) and first language acquisition data; Đoàn (1999) noted that the articulation of the creaky feature in the middle, and/or falling-rising of the Broken and Curve tones respectively, is difficult for Vietnamese children under three years of age. Nevertheless, some L2 learners adapted their production to these cues, indicated by the fact that a few of them successfully produced the glottal stop feature in the Broken and Drop tones in the read aloud task. This confirms that learners can develop their phonetic targets in active interaction with native Vietnamese speakers. This result suggests that this voice quality feature is learnable but requires explicit teaching in formal instruction, implying that neural plasticity is not hindered by neural maturation and is indeed possible after different forms and lengths of auditory training (Chandrasekaran et al. 2012; Skoe et al. 2014; Song et al. 2008).

Furthermore, the results also indicated that there was one consistent asymmetrical perceptual pattern among Vietnamese tones by L2 learners. Specifically, the tone pairs Broken-Curve, Dropping-Falling and Curve-Falling share phonetic similarities. Broken-Curve share fall-rise pitch contours, Dropping-Falling and Curve-Falling share a sharp falling contour particularly if the distinctive voice quality feature was not pronounced. This is consistent with suggestions by Polkka (1991, 1992) that a high degree of phonetic similarity between two non-native segments could increase perceptual difficulty for the listener. In fact, previous studies (Kiriloff 1969; So 2005; Wang, Spence, Jongman and Sereno 1999; Wayland and Guion 2004) have reported that non-native

language learners have great difficulties in producing and perceiving lexical tones that are similar.

Fourth, *how do speakers' different L1 experiences with prosodic features (pitch accent vs. lexical tones) affect L2 prosodic speech production and perception of Vietnamese tones?*

The results showed that Lao speakers significantly outperformed Taiwanese and Japanese speakers, and the Taiwanese significantly performed better than the Japanese in perception and the production of Vietnamese tones across three tasks. This suggests that prior experience with one tone language may facilitate the acquisition of tone in another language, consistent with Wayland and Guion (2004).

In addition, the misproduction and misperception of non-modal voice (glottal stop and creakiness) and contour tones (bidirectional fall-rise) in Vietnamese by L2 learners across three different L1 speaker groups supports the *Markedness Differential Hypothesis* (Eckman 1977), suggesting that voice quality is more "marked" and thus are more difficult for L2 learners than modal voice tones (e.g., unidirectional contours: rising, falling, and level). Furthermore, as predicted, it is found that the Japanese had more difficulties acquiring the "marked" non-modal voice (glottal stop and creakiness) such as Broken, Curve, and Dropping tones than Lao and Taiwanese speakers. This on the one hand suggests a negative transfer from their L1 Japanese and on the other, indicates that the active use of phonation type in encoding L1 tones may have played a role in the better perception of Lao and Taiwanese in the production of Vietnamese tones.

In summary, this study investigated the production and perception of Vietnamese tones by Japanese, Lao, and Taiwanese leaners, comparing their performance in an Imitation task to that in Identification and Read-Aloud tasks. The results showed that the Imitation task was generally easier for L2 speakers than the Identification and Read-Aloud tasks, suggesting that Imitation was performed without requiring the phonological encoding skills required by the other two tasks. It is also found that Lao and Taiwanese speakers outperformed the Japanese speakers, suggesting

that prior experience with one tone language may facilitate the acquisition of tone in another language. The result on speakers' tonal range showed that the L2 leaners have significantly narrower tonal F0 range than control Vietnamese speakers. The results of error pattern analysis and tonal transcription also suggest that non-modal voice and contour tones are more difficult for L2 learners than modal voice tones.

## Acknowledgements

We would like to thank the subjects for their voluntary participation in the experiment and the anonymous reviewers for their constructive comments. We also thank the editor for English corrections.

## References

Baker, Wendy and Trofimovich Pavel. 2006. Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL–International Review of Applied Linguistics in Language Teaching,* 44: 231-250.

Belotel-Grenié, Agnès and Grenié Michel. 1994. Phonation types analysis in Standard Chinese. *Proceedings of the 3rd International Conference on Spoken Language Processing* (ICSLP 1994), 343-346. https://www.iscaspeech.org/archive/icslp_1994/i94_0343.html (Accessed May 10, 2019).

Belotel-Grenié, Agnès and Grenié Michel. 2004. The creaky voice phonation and the organisation of Chinese discourse. *Proceeding of the International Symposium on Tonal Aspects of Languages With Emphasis on Tone Languages*. 5-8. https://www.isca-speech.org/archive/tal2004/tal4_005.html (Accessed August 18, 2019).

Boersma, Paul and Weenink David. 2017. Praat: doing phonetics by computer (version 6.0.26). http://www.praat.org (Accessed July 1, 2017).

Bohn, Ocke Schwen and Flege James Emil. 1997. Perception and production of a new vowel category by adult second language

learners. *Second language speech: Structure and process.* Jonathan Leather and Allan James, eds. 53-73. Berlin: Mouton de Gruyter.

Bradlow, Ann R., Pisoni, David B., Akahane-Yamada, R. and Tohkura Yoh Ichi. 1997. Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America,* 101: 2299-2310.

Brown, J. Marvin. 1976. Dead consonants or dead tones. *Tai linguistics in honor of Fangkuei Li.* Thomas W. Gething, Jimmy G. Harris and Pranee Kullavanijaya, eds. 28-38. Bangkok: Chulalongkorn University Press.

Brunelle, Marc. 2009a. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics,* 27(1): 79-96.

Brunelle, Marc. 2009b. Northern and Southern Vietnamese tone coarticulation: A comparative case study. *Journal of the Southeast Asian Linguistics Society*, 1: 49-62.

Brunelle, Marc and Jannedy Stefanie. 2013. The Cross-dialectal Perception of Vietnamese Tones: Indexicality and Convergence. *Linguistics of Vietnamese*, Daniel Hole and Elisabeth Löbel, eds. Amsterdam: John Benjamins.

Chandrasekaran, Bharath, Kraus Nina and Wong Patrick. C. 2012. Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology,* 107: 1325-1336.

Đào, Mục Đích and Nguyễn Thị Anh Thư. 2017. Vietnamese tones produced by Australian Vietnamese speakers. *Heritage Language Journal,* 14(3): 224-247.

Đào, Mục Đích and Nguyễn Thị Anh Thư. 2019. Korean L2 learners' perception and production of Vietnamese tones. *Journal of Second Language Pronunciation,* 5(2): 195-222. https://doi.org/https://doi.org/10.1075/jslp.17011.dao (Accessed August 18, 2019).

Davison, Deborah S. 1991. An acoustic study of so-called creaky voice in Tianjin Mandarin. In *UCLA Working Papers in Phonetics,* 78: 50-57.

Đoàn, Thiện Thuật. 1999. *Ngữ âm tiếng Việt* [Vietnamese phonetics]. Hà Nội: Đại học & Trung học chuyên nghiệp Press.

Eckman, Fred R. 1977. Markedness and the Contrastive Analysis Hypothesis. *Language Learning,* 27(2): 315-330.

Emeneau, Murray Barnson. 1951. *Studies in Vietnamese (Annamese) grammar*. Los Angeles: University of California Press.

Flege, James Emil and Eefting Wieke. 1988. Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America,* 83: 729-740.

Fowler, Carol A., Brown, Julie M., Sabadini Laura and Weihing Jeffrey. 2003. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language,* 49: 396-413.

Gandour, Jack. 1983. Tone perception in far eastern languages. *Journal of Phonetics,* 11: 149-175.

Gårding, Eva and Svantesson,Jan-Olof. 1994. An introductory study of tone and intonation in a Lao dialect, *Acta Linguistica Hafniensia*, 27(1): 219-233.

Gedney, William. J. 1972. A checklist for determining tones in Tai dialects. *Studies in linguistics in honor of George L Trager*. Estellie Smith, ed. 423-437. Mouton: The Hague.

Hallé, Pierre A., Chang Yuehchin and Best Catherine T. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics,* 32(3): 395-421.

Hao, Yen-Chen and De Jong Kenneth. 2016. Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics,* 54: 151-168.

Huang, Karen. 2012. A study of neutral - tone syllables in Taiwan Mandarin. PhD Dissertation. University of Hawaii at Manoa.

Hyman, Larry M. and VanBik Kenneth. 2004. Directional rule application and output problems in Hakha Lai tone. *Phonetics and Phonology*, Special Issue, Language and Linguistics, 5: 821-861.

Kawahara, Shigeto. 2015. The phonology of Japanese accent. *The handbook of Japanese phonetics and phonology.* Haruo Kubozono, ed. 445-492. Berlin: Mouton de Gruyter.

Kirby, James. 2010. Dialect experience in Vietnamese tone perception. *Journal of the Acoustical Society of America,*

127(6): 3749-3757.

Kiriloff, C. 1969. On the auditory discrimination of tones in Mandarin. *Phonetica,* 20: 63-67.

Kuang, Jianjing. 2013. Phonation in Tonal Contrasts. PhD Dissertation. University of California.

Lee, Yuh-Shiow, Vakoch Douglas and Wurm Lee H. 1996. Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5): 527-544.

Levelt, Willem J. M., Roelofs Ardi and Meyer Antje S. 1999. A theory of lexical access in speech production, *Behavioral and Brain Sciences,* 22: 1-75.

Michaud, Alexis. 2004. Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. *Phonetica,* 61: 119-146.

Michaud, Alexis, Vũ, Ngọc T., Amelot Angelique and Roubleau Bernard. 2006. Nasal realease, nasal finals and tonal contrasts in Hanoi Vietnamese: an aerodynamic experiment. *Mon-Khmer Studies,* 36: 121-137.

Mitterer, Holger and Ernestus Miriam. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition,* 109: 168-173.

Mitterer, Holger and Müsseler Jochen. 2013. Regional accent variation in the shadowing task: Evidence for a loose perception–action coupling in speech. *Attention, Perception, &* Psychophysics, 75: 557-575.

Morev, Lev Nikolaevich, Moskalyov Aleksei Alekseevich and Plam IUrii IAkovlevich. 1979. *The Lao language*. Moscow: USSR Academy of Sciences, Institute of Oriental Studies.

Nguyễn, Văn Lợi and Edmondson Jerold A. 1997. Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies,* 28: 1-18.

Osatananda, Varisa. 1997. Tone in Vientiane Lao. PhD Dissertation. University of Hawaii.

Pham, Hoa Andrea. 2003. *Vietnamese tone: A new analysis*. Routledge.

Polka, Linda. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of*

*the Acoustical Society of America,* 89(6): 2961-2977.

Polka, Linda. 1992. Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics,* 52(1): 37-52.

Shockley, Kevin, Sabadini Laura and Fowler Carol A. 2004. Imitation in shadowing words. *Perception & Psychophysics,* 66, 422-429.

Skoe, Erika, Chandrasekaran Bharath, Spitzer, Emily R., Wong, Patrick. C. and Kraus Nina. 2013. Human brainstem plasticity: the interaction of stimulus probability and auditory learning. *Neurobiology of Learning and Memory,* 109: 82-93.

So, Connie K. 2005. The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. *The Journal of the Acoustical Society of America,* 117(4): 2427. https://doi.org/10.1121/1.4786607 (Accessed August 18, 2019).

Song, Judy H., Skoe Erika, Wong, Patrick C. and Kraus Nina. 2008. Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience,* 20, 1892-1902.

Strecker, David. 1979. A preliminary typology of tone shapes and tonal sound changes in Tai the Lan Na A-tones. *Studies in Tai and Mon Khmer phonetics and phonology in honour of Eugenie J A Henderson*. Thiraphan Lo Thongkum et al., eds. 171-240. Bangkok: Chulalongkorn University Press.

Studebaker, Gerald A. 1985. A rationalized arcsine transform. *Journal of Speech, Language, and Hearing Research,* 28: 455-462.

Tsukada, Kimiko and Kondo Mariko. 2018. The Perception of Mandarin Lexical Tones by Native Speakers of Burmese. *Language and Speech*, 1-16.

Tu, Jung-Yueh, Hsiung Yuwen, Cha Jih-Ho, Wu Min-Da and Sung Yao-Ting. 2016. Tone production of Mandarin disyllabic words by Korean learners. *Proc. Speech Prosody*, 375-379.

Tu, Jung-Yueh, Hsiung Yuwen, Wu Min-Da and Sung Yao-Ting. 2014. Error patterns of Mandarin disyllabic tones by Japanese learners. *Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech-2014),* 2558-2562.

Vũ, Thanh Phương. 1981. The acoustic and perceptual nature of

tone in Vietnamese. PhD Dissertation. Australian National University.

Wang, Yue, Spence Michelle M., Jongman Allard and Sereno Joan. A. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America,* 106: 3649-3658.

Wayland, Ratree P. and Guion Susan G. 2004. Training English and Chinese listeners to perceive Thai Tones: A preliminary Report. *Language Learning,* 54(4): 681-712.

Yang, Ruo-Xiao. 2011. The phonation factor in the categorical perception of mandarin tones. *Proceedings of ICPhS XVII*, 2204-2207.

Yang, Ruo-Xiao. 2015. The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics,* 43: 453-472.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Received: June 7, 2021; Reviewed: Nov. 11, 2021; Accepted: Jan. 15, 2022