

“Do You Hear What I Hear?": On the Elusive Nature of Vietnamese Tones

Mai Chi Nguyen
Department of Modern Languages
Liberty University
1971 University Blvd.,
Lynchburg, Virginia 24515 USA

Faculty Advisor: Dr. Jaeshil Kim

Abstract

The acoustic cues for tone identification are known to come from auditory impressions and phonetic properties such as “pitch contours and height of tones” (Brunelle, 2009), but this account is insufficient to explain more complex tonal systems such as Vietnamese. Vietnamese is a tonal language, with six different tones: ngang (T1), huyền (T2), sắc (T3), nặng (T4), hỏi (T5), and ngã (T6). Brunelle (2009), in his research on dialectal variations of Vietnamese tones and their perceptual cues, claimed that lexical tones in Vietnamese are mainly characterized by pitch, glottalization, and duration, all of which directly and indirectly influence tone perception. Diverging from Brunelle (2009), this research is more concerned with the production of Vietnamese tones, especially of the nặng tone (T4), which is considered as the hardest tone to acquire. This study debunks the commonly held assumption that pitch and glottalization account for the difficulty of acquiring T4, and claims that the difficulty mainly comes from the varied realizations of this tone. Eighty pronunciation samples of the word ‘ma’ in six tones were collected from the Northern, Central, and Southern Vietnamese speakers. The variations of T4 are analyzed into three phonetic features: duration, pitch contour, and glottalization. The findings showed that T4 has three variations: a short falling tone in the Northern dialect (T4a), a contour tone with a falling-rising pattern in the Southern dialect (T4b), and a discontinuous contour tone found predominantly in young Southern female speakers aged 13-19 (T4c). Generally, T4 is characterized as a falling tone, with a word-final glottalization and short duration. However, these factors are susceptible to change. The duration of T4 is twice as long in the Southern dialect than the Northern, and pitch contours in the Southern dialect are more nuanced. Interestingly, almost half of the Southern subjects realize glottalization not word-finally but word-medially. This word-medial glottalization is more prevalent in females than males, and in the youth than adults, accounting for 60% and 87% of the samples, respectively. The findings of this study provide valuable insights into the linguistic understanding and pedagogical implications of Vietnamese tones.

Keywords: Vietnamese, tonal languages, phonetics

1. Introduction

1.1 The Mismatch between Speech Perception and Speech Production

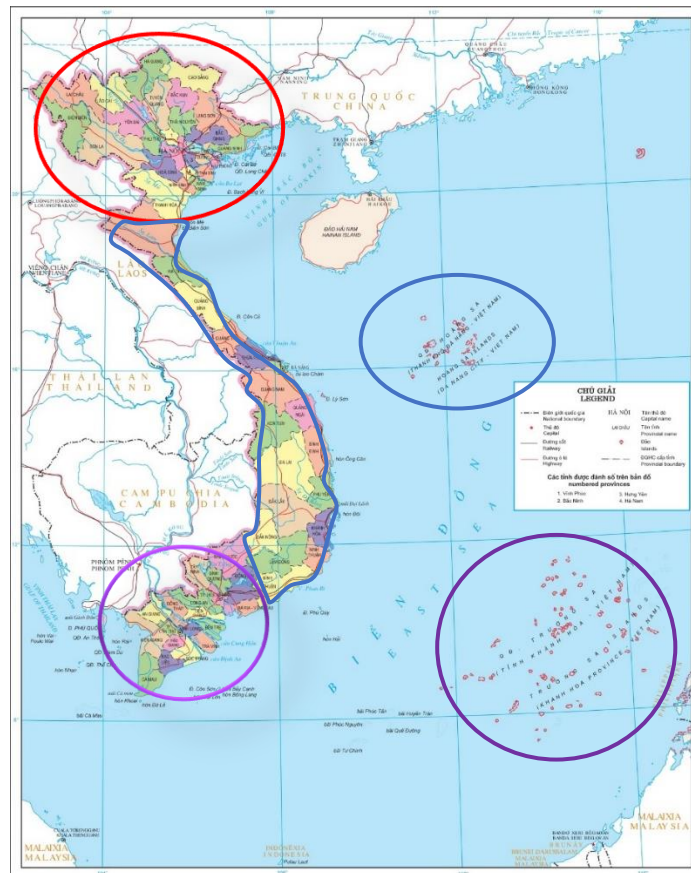
The mismatch between speech perception and speech production is prevalent but most of the time, is not within the consciousness of native speakers. Such discrepancy is found in Vietnamese speakers in their production of one of the six Vietnamese tones, tone nặng (T4). Speakers realize this tone in three distinct ways but underlyingly perceive it as one tone, which is extremely puzzling for Vietnamese learners, for learners perceive those realizations as three different tones. This research collects and presents some general pronunciation patterns and overall characteristics of

T4, shows the three possible realizations, and claims that these realizations are the main reason that makes the acquisition process of T4 particularly challenging for learners.

1.2 Overview of the Vietnamese Language

Vietnamese is a Viet-Muong language classified under the Mon-Khmer branch within the Austro-Asiatic language family (Pham, 2001). The language is highly isolating, determines grammatical relations through word order, and takes SVO as the default word order. The language does not allow consonant clusters, but it has a wide inventory of monophthongs, diphthongs, and triphthongs (Thompson, 2016). Generally, Vietnamese dialects are divided into three based on the three main regions: Northern, Central, and Southern (Vu, 1981). The map below marks the Northern, Central, and Southern region with red, blue, and purple, respectively. The Northern dialect is considered the standard dialect and is widely used in national media (Brunelle, 2009).

Image 1. A Map of Vietnam



1.3 Vietnamese Tones

Tone is a prosodic unit that could provide lexical meanings or grammatical functions of a word (Yip, 2002). This suprasegmental feature of speech can be analyzed physically, acoustically, and perceptually. Physically and acoustically, tone is viewed through fundamental frequency, which Yip (2002) described as the number of vibrations the vocal fold makes in speech. Perceptually, people normally associate tone with pitch variation. This means how low or how high the pitch is will determine the type of tones (*ibid*).

Vietnamese is a tonal language with six tones: sắc (T1), huyền (T2), nặng (T4), hỏi (T5), and ngã (T6). Tones in this language do not provide the grammatical function but the lexical meaning of words. The following table includes the six tones, with their Vietnamese names, contour pattern, English descriptions, and examples. In the examples, the

word is transcribed with IPA symbols but the tone diacritics are kept as in the Vietnamese orthography for the sake of simplicity.

Table 1. Vietnamese Tones

Number	T1	T2	T3	T4	T5	T6
Vietnamese name	ngang	huyền	sắc	nặng	hỏi	ngã
Contour pattern	high (H)	high-low (HL)	low-high (LH)	high-low-glottalization (HLG)	high-low-high (HLH)	High- glottalization-high (HGH)
English description	Level tone	Falling tone	Rising tone	Falling tone with ending glottalization	Contour tone with rising and falling pattern	Contour tone with rising and falling pattern with mid-glottalization
Example	[ma] 'ghost'	[mà] 'but'	[má] 'cheek'	[mạ] 'rice seeds'	[mả] 'tomb'	[mã] 'horse'

This table provides a definition that aligns to the standard Northern Vietnamese accent. The two figures below from Bauman et. al. (2009) give a visual representation of the tones in the Northern and Southern dialect.

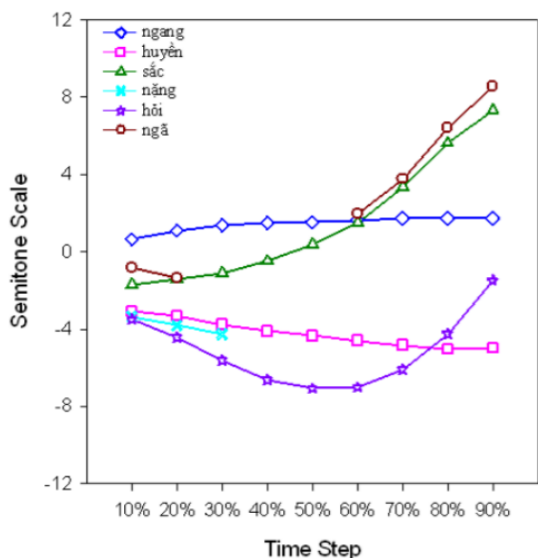


Figure 1. Northern Vietnamese tone system

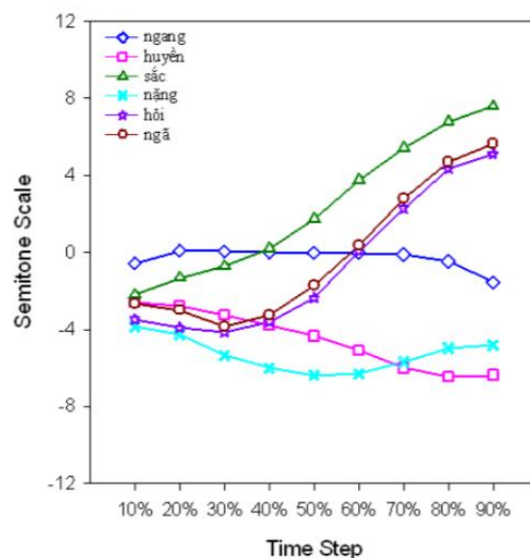


Figure 2. Southern Vietnamese tone system

T1, T2, T3 are level tones, meaning they have constant, falling, and rising pitch variation, respectively, while T5 and T6 are contour tones meaning there is a combination of falling and rising pitch variation patterns within these two tones. Even though T5 and T6 both follow a falling and rising pattern, T6 is perceived as a discontinuous tone in the Northern dialect due to the presence of a mid-glottal stop while most Southerners do not make a clear distinction between these two tones (Brunelle, 2009). Thus, most of the time, Southern speakers must rely on context to determine whether T5 or T6 is being used. However, T4 is by far the most complex tone out of the six. In the Northern dialect, it is realized as a level tone, which has an abrupt falling with a word-final glottalization and in the Southern dialect, as a contour tone with a falling and rising pattern, yet with a starting pitch lower than that of T5.

2. Literature Review

Vietnamese tones are studied and analyzed through fundamental frequency, pitch contour, voice quality, and duration (Brunelle, 2009; Hoang, 1986; Nguyen & Edmonson, 1997). Researchers conclude that, generally, F0, voice quality, pitch contour, and sometimes, duration function as the acoustic cues for tonal recognition (Brunelle, 2009; Gsell, 1980; Hoang, 1986; Kirby, 2010; Tran, 1967; Vu, 1981, 1982). Northern speakers rely more on voice quality while pitch height and pitch contour are the primary cues that Southern speakers employ to distinguish one tone from another (Brunelle, 2009; Kirby, 2010). However, not all researchers agree on how Vietnamese tones are perceived. For instance, Kirby (2010) believes duration is nuanced enough to influence perception while Brunelle (2009) disagrees. Such conflicting findings and claims exist because speakers' perception is subjective and tend to change given their linguistic background. Kirby (2010) stresses that a speaker's primary dialect highly influences how he/she perceives tones.

The limit of past research on Vietnamese tones perception is that it mainly focuses on analyzing Vietnamese tones from a perception perspective. The research usually involves native speakers that listen to a pre-recorded pronunciation sample of Vietnamese words with different tones done by a Northern Vietnamese speaker. Due to this, the research cannot capture an implicit analysis from within the speakers. Hence, diverting from the conventional research methodologies, this research attempts to study Vietnamese tones from a production perspective, which gives an overview of how first language speakers actually produce the tones and perceive them.

3. Research Methodology

3.1 Subjects

A total of 80 subjects participated in the research with 42 (52.5%) females and 38 (47.5%) males. Northern, Central, and Southern subjects consist of 24, 7, and 49 subjects, respectively. For Northern subjects, all speakers and their parents were born and raised in the North, most of whom currently live in Hanoi city, the capital of Vietnam, while some others reside in the nearby cities. Almost all of the Southern speakers were born and raised in Ho Chi Minh City, with 3 speakers from the nearby cities and provinces. The Southern speakers' linguistic backgrounds are relatively more complex. All of them were born and raised in the South, but some of their parents were Northerners that migrated to the South around the 1950s. Due to the limited access to Central speakers and the main focus of the research on the other two dialects, only seven recordings were collected from Central subjects.

3.2 Materials

Initially, 124 recordings were collected from subjects with their pronunciation of the word [ma] in six tones in the following order: [ma] (T1), [mà] (T2), [má] (T3), [mạ] (T4), [mã] (T5), and [mã̃] (T6). To ensure the sound quality, only 80 recordings were chosen and then put into Praat for further phonetic analysis.

3.3 Procedure

Subjects were asked to record their pronunciation of the word [ma] in 6 tones with either Audacity or a speech recorder in a quiet room with minimal noise. The recording started with information about the speaker for classification purposes such as name, age, and place/country of residence and ended with the speaker's pronunciation of the word [ma] in 6 tones. Natural speed and pronunciation were encouraged. After shortlisting for sound quality, 80 recordings were chosen to be analyzed in Praat in five main parameters: highest frequency, lowest frequency, pitch difference, the duration between the highest and lowest frequency, and the absence/presence of a mid-glottal stop. All measurements were put into an Excel document. The identity of the participants was confidential, so the measurements were divided into age groups, gender, and dialects with no names associated.

4. Results

4.1 Duration

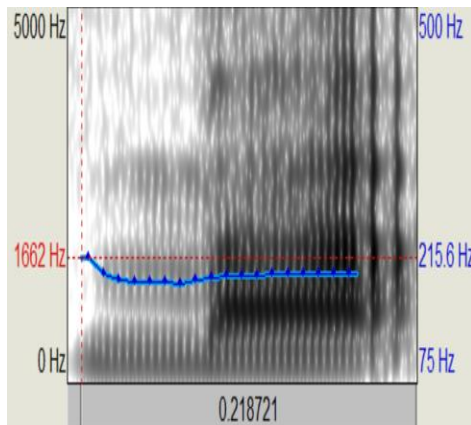
In comparison to the other tones of Vietnamese, T4 is a relatively short tone with an average duration of 228 milliseconds. T2 whose pitch contour is relatively similar to that of T4 has an average duration of 342 milliseconds, which is 1.5 times longer than that of T4. Out of the three dialects, Central speakers took the longest time to pronounce T4, approximately 293 milliseconds; Southern speakers took about 244 milliseconds, and Northern speakers had the shortest duration, about 146 milliseconds.

Table 1. General characteristics of T4

Duration from the highest-lowest pitch	158 milliseconds
Duration of the whole tone	228 milliseconds

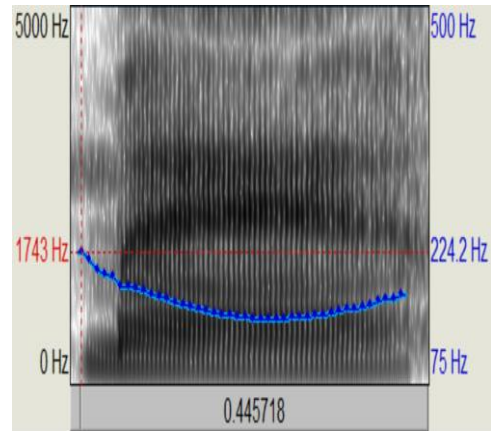
4.2 Pitch Contour

T4 is realized as a short falling tone in the Northern dialect and a contour tone with a falling-rising pattern in the Central and Southern dialect. Northern subjects showed almost the same results for the duration between the highest to lowest pitch point and the overall tone duration. This strongly indicates that speakers of this dialect mainly realize T4 as a sharp falling tone. As for the Central and Southern speakers, the contour falling-rising pattern was the norm and none of the subjects pronounced T4 as a falling tone. This contour nature also accounts for the longer duration of the tone in these two dialects compared to the Northern one.



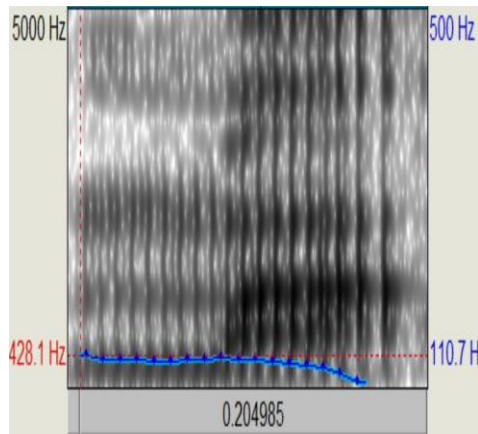
[mã]
‘rice seeds’

Figure 3. Northern female pronunciation of T4



[mã]
‘rice seeds’

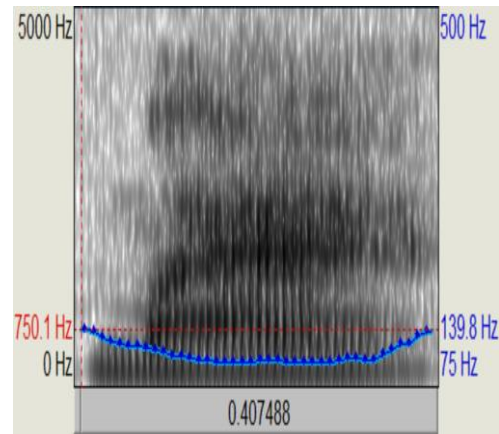
Figure 4. Southern female pronunciation of T4



[ma]

‘rice seeds’

Figure 5. Northern male pronunciation of T4



[ma]

‘rice seeds’

Figure 6. Southern male pronunciation of T4

4.3 Glottalization

Final glottalization is present in the Northern subjects’ pronunciation. Almost all Northern speakers pronounced T4 as a sharp falling tone with a final glottal stop. Glottalization is not present in almost every recording of Central and Southern speakers. The final glottalization may account for the short duration and the falling contour shape of T4 in the Northern dialect, for speakers must form a short falling tone, so the final glottal stop aids that process. However, since Central and Southern speakers need to produce a contour tone, they need longer time to create the falling-rising contour shape and therefore, the glottalization is not present.

4.4 Continuity and Discontinuity of T4

In addition to the two ways of realizing T4, which is either a short falling tone found in Northern subjects (T4a) or a falling-rising contour tone found in Southern speakers (T4b), a third realization, a discontinuous tone found predominantly in young Southern female speakers aged 13-19 (T4c), was discovered. Notably, it was neither recorded nor reported in any past research on Vietnamese tones. Out of the 80 recordings, 22 (28%) were discontinuous. Within the discontinuous tokens, 18, 3, 1 are from Southern, Northern, and Central subjects, respectively. Half of the discontinuous tokens are from 13-19-year-old female subjects, the other three discontinuous tokens are from Northern male subjects and one discontinuous case is found in the Central dialect and the speaker is a 21-year-old female born and raised in Da Nang. The third realization of T4, the discontinuous token (T4c) has a similar contour shape to that of the second realization (T4b), which is a contour falling-rising tone, except for the presence of a mid-glottalization.

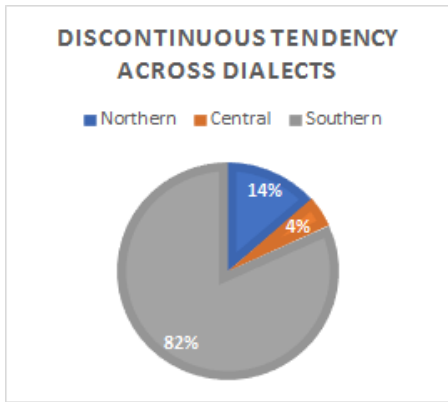


Figure 7. Discontinuous tendency across dialects

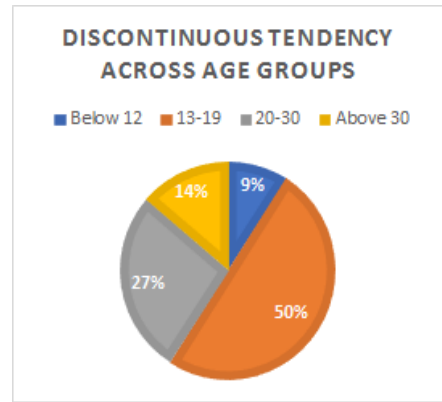
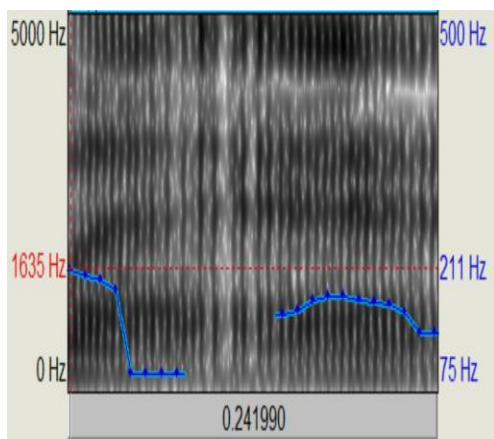
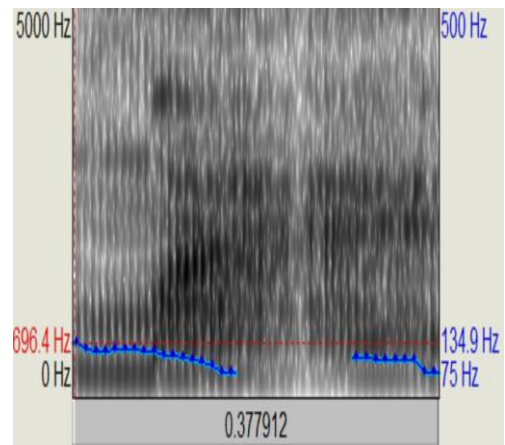


Figure 8. Discontinuous tendency across age groups



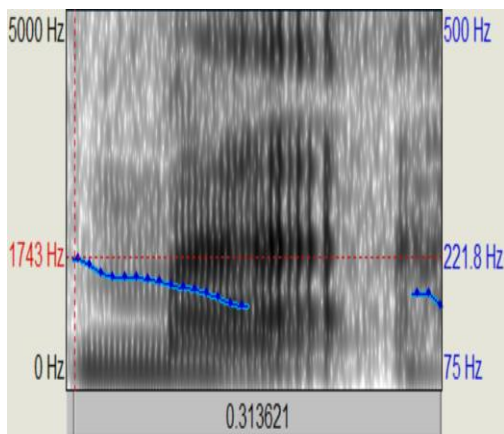
[mə]
'rice seeds'

Figure 9. Southern female discontinuous tone



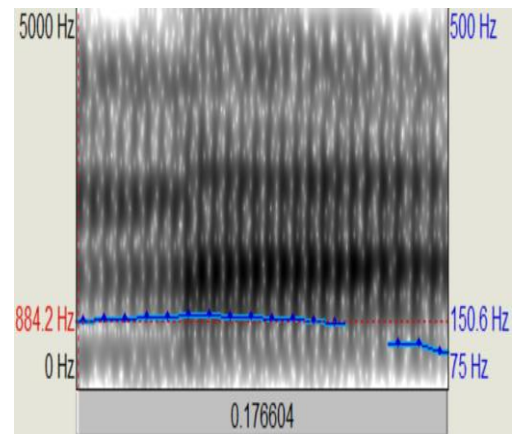
[mə]
'rice seeds'

Figure 10. Southern male discontinuous tone



[mə]
'rice seeds'

Figure 11. Central female discontinuous tone



[mə]
'rice seeds'

Figure 12. Northern male discontinuous tone

5. Discussion

The findings of this research suggest that the toneme “*nặng*,” T4, is realized as three allotones in Vietnamese: a short falling tone in the Northern dialect (T4a), a contour tone with a falling-rising pattern in the Southern dialect (T4b), and a discontinuous contour tone found predominantly in young Southern female speakers aged 13-19 (T4c). The third realization, T4c, was not recorded in any past research and raises an interesting question unto what triggers this realization among the young generation. However, this research is not able to give an exact cause of this phenomenon, for that lies beyond the research scope.

Several conjectures to account for this tendency that future research can consider are linguistic convergence and divergence, and that the glottal stop should be treated as a phenomenon of phonation rather than a segment. Southern speakers are making finer distinction within T4 itself, yet the third realization, T4c, seems to converge its contour shape with T6, tone “*ngã*.” As Kirby (2010) and Brunelle (2009) state, almost all Southern speakers do not make a distinction between T5 and T6 in running speech, whereas Northern speakers do. Thus, in a way, T4 is converging with T6 and diverting within T4 itself. Moreover, the glottal stop should be treated as a phenomenon of phonation rather than a segment of speech, for the glottal stop was present to indicate sudden pitch drop, but not as a consonant as in other languages such as Arabic, for instance. However, these conjectures require a more extensive research to determine their validation. The scope of this research was still relatively limited to discover the true reason that prompted the third realization.

6. Implications

The findings of this study provide valuable insights into the Vietnamese pedagogical methods. Generally, Vietnamese learners are more exposed to the standard Northern Vietnamese, which is considered an elite dialect. Such limited exposure hinders them from recognizing the diverse realizations of tones in practicality, so they tend to get puzzled. Thus, understanding the complexity of the tonal system helps shape better pedagogical strategies for future Vietnamese textbooks. For instance, learners can train their ears to adapt to these allotonic variations and can therefore, improve comprehension and enhance the learning experience. Moreover, the findings of this research can contribute to the field of linguistics by discovering the third realization of T4 that has not been reported in the literature yet. However, this research leaves much room for improvement because it was not able to present the underlying cause or condition of the innovative way of pronouncing T4 (T4c) and also why it is more common among females than males and why Southern speakers tend to manifest T4 in this way. Unfortunately, it is beyond the scope of this research to answer those questions. Thus, more extensive sociolinguistic research is needed to find out the cause for this innovative pronunciation of the *nặng* tone in Vietnamese.

7. Conclusion

In this research, Vietnamese tones were analyzed through the production perspective in which the pronunciations of native speakers of T4 were collected and analyzed. The finding of this research showed that T4 can be realized in three different ways: a short falling tone in the Northern dialect (T4a), a contour tone with a falling-rising pattern in the Southern dialect (T4b), and a discontinuous contour tone found predominantly in young Southern female speakers aged 13-19 (T4c). The third realization is a new discovery and is found predominantly in a certain group of speakers. Out of 80 pronunciation samples, 28% of the subjects have this pronunciation pattern and 50% of these discontinuous tokens are found in young people aged 13-19 years old. 82% of the subjects identified themselves as Southern speakers and 60% are females. Thus, the tendency for these discontinuous tokens is mainly found in young female speakers from the South of Vietnam. Such diverse realizations of this tone confirm that the challenge to acquire this tone may lie in the discrepancy of speakers’ production and perception of the tone itself.

8. Acknowledgements

This research was supported by the Liberty University Center for Research and Scholarship and the Department of Modern Languages at Liberty University. The author wishes to especially recognize Dr. Jaeshil Kim, Ph.D. who served

as the mentor and research advisor for this project. Dr. Kim has provided me the constant encouragement, thoughtful advice and insights, indefinite patience, and true inspiration for linguistics research. I would not have been able to pursue this research topic without her support. Dr. Kim played a big role in honing my research skills, especially in the field of linguistics. I also want to extend my thankfulness to all the participants and especially, to my mother. The participants were willing to join this research and provide me with valuable linguistic data and my mother has been a great assistance in helping me gather the data so I could choose the best quality recordings for further phonetic analysis.

9. References

1. Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37, 79–96.
2. Gsell, R. (1980). Remarques sur la structure de l'espace tonal en Vietnamien du Sud (parler de Saigon). *Cahiers d'Études Vietnamiennes* (4), 1–26.
3. Hoang, C. (1986). Suy nghĩ thêm về thanh điệu tiếng Việt. *Ngôn ngữ*, 69(3), 19-38.
4. Kirby, J. (2010). Dialect experience in Vietnamese tones perception. *Journal of Acoustical Society of America* 127(6), 3749-3757. doi: 10.1121/1.3327793
5. Nguyen, V. L., & Edmondson, J. (1997). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *The MonKhmer Studies Journal* 28, 1–18.
6. Pham, A. (2001). Vietnamese tone: Tone is not pitch. Ph.D. Dissertation, University of Toronto.
7. Tran, H. M. (1967). Tones and intonation in South Vietnamese. In L. Nguyen, H. M. Tran, & D. Dellinger (Eds.), *Series A—Occasional papers #9, papers in southeast Asian linguistics no. 1*. Canberra: Linguistics Circle of Canberra.
8. Vu, P. (1981). The acoustic and perceptual nature of tone in Vietnamese. Dissertation, Australian National University.
9. Vu, P. (1982). Phonetic properties of Vietnamese tones across dialects. In D. Bradley (Ed.), *Papers in Southeast Asian Linguistics*, 55-75. Sydney: Australian National University.
10. Yip, M. (2002). *Tone*. Cambridge, MA: Cambridge University Press.