Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

# Perceptual assimilation and discrimination of falling, level, and rising lexical tones by native English speakers

*Irina Shport*

Louisiana State University, U.S.A.
ishport@lsu.edu

## Abstract

Perceptual assimilation model posits that the perceived relation of non-native sounds to native sound categories predicts the discriminability of non-native sounds [1]. This study examines whether the degree of perceived similarity between Vietnamese tones and English intonation patterns predicts the accuracy of tone-contrast discrimination. In a categorization task, ten native English speakers were asked to match falling, level, and rising tone words with five English intonation patterns. Then, in a discrimination task, they were asked to judge the similarity of Vietnamese tones to each other. In the first task, speakers perceived both of the falling and level tones as most similar to the statement intonation (*Right*), and the rising tone as most similar to the question intonation (*Right?*). These cross-language mappings averaged at 74%, 65% and 73% of responses, respectively, suggesting relatively reliable categorization of tones and predicting relatively low discrimination accuracy for the level-falling tone contrast. In the second task, however, level-falling tones were discriminated significantly better than level-rising tones. This suggests that assimilation of tones to different intonation patterns is not the sole predictor of tone discrimination accuracy.

**Index Terms**: tones, intonation patterns, cross-language mapping, discrimination, perceptual assimilation model

## 1. Introduction

Theories of cross-linguistic speech perception posit that listeners perceive non-native sounds in relation to native sound categories [1, 2]. According to Perceptual Assimilation Model (PAM), this perceived relation predicts the discriminability of non-native sound contrasts, and, ultimately, sound learning [1]. Few studies, however, provided data on both the perceived phonetic similarity between non-native and native sounds and discrimination of sound contrasts [3, 4]. The PAM prediction was directly tested in two previous studies: for perception of American English /ɹ/ and /l/ in native Japanese speakers and for perception of Thai lexical tones in native speakers of Mandarin, Cantonese, and English. Similar to [4], the goal of this study was to test the PAM prediction by providing both of the perceived similarity and discrimination data for lexical tones. The focus, however, was on perception of selected Vietnamese lexical tones in native English listeners. The perceptual mapping work has not been done before with these two particular languages.

The cross-linguistic mapping between lexical tones of a tone language, such as Vietnamese, and intonation patterns of a non-tone language, such as English, is of a particular theoretical and practical interest. Theoretically, it is not clear whether PAM is applicable to sound categories across different domains in intonational phonology: namely, word-level lexical tones and phrase-level intonation patterns. Practically, lexical tones are notoriously difficult for speakers of non-tone languages to acquire. Thus, it would be important to know whether speakers of non-tone languages can consistently associate lexical tones with particular native intonation patterns, and whether this cross-linguistic mapping predicts discrimination of tone contrasts.

Previous research on tone perception is not conclusive with regard to the above-formulated question. Some studies suggest that lexical tones may not be consistently identified in terms of intonation patterns, and thus they are uncategorizable, while other studies report that such cross-language mapping is possible [5, 6]. Lexical tones and intonation patterns are coded by variation in fundamental frequency ($F_0$), and their perceived similarity may be based on the overall similarity of their $F_0$ contours. (It may not be driven by function because lexical tones differentiate word meaning and intonation patterns differentiate discourse-pragmatic contexts).

In this study, Vietnamese tones and English intonation patterns that have comparable $F_0$ contours and modal voice quality were selected for the categorization and discrimination experiments. They are illustrated in Figures 1 and 2, respectively [7]. The selection was made using previous literature, consultations with two informants, and a pilot recording. In Vietnamese, level tone A1, falling tone A2, and rising tone B1 have modal voice quality, at least in some dialects and in the pronunciation of this study's informant [8, 9]. This study was focused on perceptual assimilation of these three tones to English intonation patterns and pairwise perceptual discrimination of the three tone contrasts.

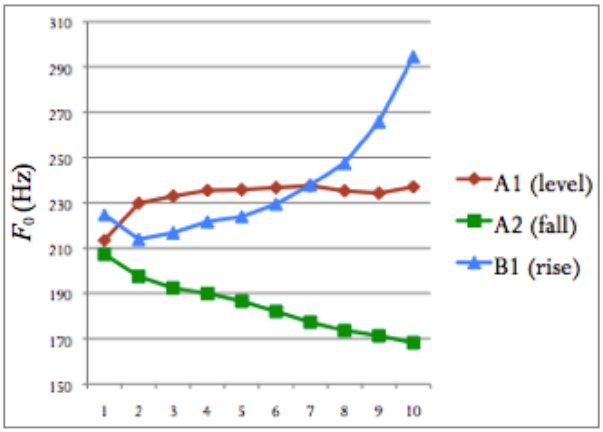


Figure 1: *Time-normalized fundamental frequency ($F_0$) contours of three lexical tones in Vietnamese 'ma' and 'nong' words produced by a female native speaker and averaged over six exemplars for each tone.*


doi: 10.21437/SpeechProsody.2016-109

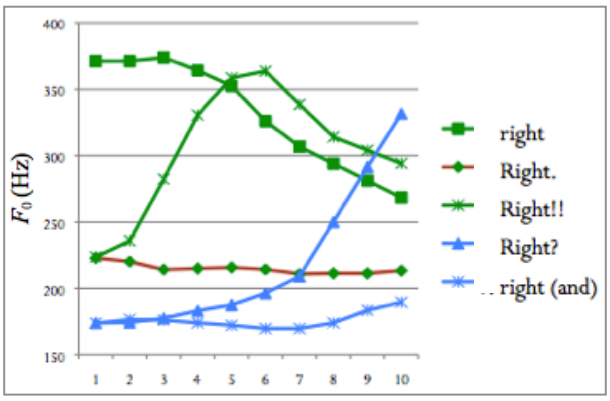


Figure 2: *Time-normalized fundamental frequency ($F_0$) contours of five intonation patterns in English 'right' produced by a female native speaker and averaged over three exemplars for each pattern.*

The selection of English intonation patterns was more complicated. Statement and question intonation patterns are often compared to lexical tones because they have rising and falling $F_0$ contours associated with them (*Right?* and *Right*); however, these associations are not unique [10, 11]. Exclamations and commands may have definitive falls (*Dreadful!! Stop that noise*); incomplete utterances (open lists) and short commands may have rises (*Chicken or beef? Go on*), let alone widespread high-rising terminals in declarative sentences [11]. Given a variety of discourse-pragmatic functions associated with rising and falling $F_0$ contours in English, the choice of English intonation patterns that the rising and falling tones can possibly map onto is not straightforwardly restricted. In [4], for example, five intonation patterns were chosen: question, statement, command, uncertainty, and "flat pitch". In the current study, the command and uncertainty patterns were avoided because of the one-to-many correspondence problem: Each of these patterns may have either a rising or falling $F_0$ contour [11, 12]. The uncertainty pattern was also accompanied by glottalization in a pilot recording. The incomplete pattern was selected because it is characterized by an $F_0$ rise albeit a smaller one than in a question pattern (*right (and)*). The exclamation pattern with a definitive fall (*Right!!*) was selected because in a pilot recording the statement pattern (*Right*) consistently had a flat $F_0$ contour, which resembled the description of "single affirmative words" [13]. Finally, the backchannel pattern (*right*) was selected because it may also have a flat $F_0$ contour [13]. All English intonation patterns used in this study as possible assimilation categories are shown in Figure 2.

In sum, this study investigated the perceptual assimilation and discrimination of falling, level, and rising Vietnamese lexical tones in native speakers of North American English, naïve to tone languages. In the first categorization experiment, the overall shape of the $F_0$ contour was expected to influence cross-language mapping among non-native and native categories. The cross-language mapping patterns were expected to predict the outcomes of the second discrimination experiment.

# 2. Methods

## 2.1. Participants

Ten native speakers of the Southern American English participated in this study as volunteers: 6 females and 4 males between the ages of 19-31. All of the participants were naïve to tone languages. They self-reported normal hearing; this was confirmed in a hearing test (at or under 25 dB at each of 250, 500, 1000, 2000, 4000, and 8000 Hz).

## 2.2. Materials

Three Vietnamese lexical tones (level A1, falling A2, and rising B1 in Figure 1) and five English intonation patterns (backchannel, statement, exclamation, question, and incomplete in Figure 2) were chosen to examine the perceived similarity between Vietnamese and English prosodic categories [8-13]. Nine Vietnamese tone words were recorded for each of the target tones by a female native speaker of Central Vietnamese, e.g., *ma* 'ghost', *mà* 'but', and *má* 'mother.' All of these words were produced as single-syllable utterances. A visual examination of the word $F_0$ contours in Praat confirmed that each of them had a level, falling, or rising contour, similar to the ones shown in Figure 1 [7].

English 'right' utterances were recorded in scripted dialogues by a female native speaker of North American English. These utterances were chosen because they were also monosyllabic, similar to the Vietnamese words. They were produced with different intonation patterns depending on the dialogue context. The $F_0$ contours of the 'right' utterances are shown in Figure 2. These excised utterances were used in the categorization experiment for the illustrative purpose only, when instructions were provided to the participants.

## 2.3. Procedure

Both experiments were conducted in a sound-proof room, one participant at a time. Praat was used to present the stimuli and record participants' responses [7]. First, the degree of perceived similarity was assessed in a categorization experiment. Vietnamese stimulus words were presented to the participants for two kinds of auditory evaluation. First, the participants were asked to identify each Vietnamese token as an instance of some English intonation pattern category. They responded by clicking one of the answer buttons on a computer screen: 'right,' 'Right.' 'Right!!' 'Right?' or 'right (and).' Then, immediately after, they were asked to rate the token for goodness-of-fit of just identified English category on the scale from 1 to 7 by clicking the respective button.

Next, tone discrimination was tested in an ABX experiment with 720 'different' trials (e.g., A1-A2-A2) and 480 'same' trials (e.g., A1-A1-A1). After hearing a triplet of tone words, listeners decided which word was an odd-man-out, if any. They responded by clicking one of the answer buttons on a computer screen: 'first,' 'second,' third,' or 'none.'

## 2.4. Analysis

For the data from the categorization experiment, chi-square and *t*-test analyses were conducted on identification responses and goodness-of-fit responses, respectively (3 tones x 9 word forms x 3 repetitions x 10 listeners). For each of the three stimulus tones, if any intonation pattern was selected over

50% of the time, it was classified as "categorized" (C); otherwise, it was classified as "uncategorized" (U)" [4]. If two tones were both categorized as the same intonation pattern using the 50% criterion, then this tone pair was classified as either single-category (SC) or category goodness (CG). In these cases, if the goodness ratings for the two tones did not differ significantly according to a *t*-test, it was classified as SC; otherwise, it was classified as CG [1, 4]. If two tones were both classified as different intonation patterns using the 50% criterion, then this tone pair was classified as two category (TC). Based on these cross-linguistic mapping patterns, the following discrimination accuracy of tones was predicted: TC > CG > SC [1].

For the data from the discrimination experiment, d-prime scores were calculated to assess each listener's sensitivity for each of the three contrasts in the discrimination test - A1-A2, A1-B1, and A2-B1 [14]. Then, the scores were submitted to the analysis of variance with Contrast as a fixed factor (3), and further examined in pairwise comparisons with Bonferroni corrections for multiple comparisons.

# 3. Results

## 3.1. Categorization of Tones

Chi-square tests showed that the *Right* and *right* responses were used with similar frequency for each of the three target tones (n.s.). Therefore, these responses were collapsed in further analyses. Figure 3 illustrates the effect of tone on categorization responses, [$X^2(6, N = 810) = 562.84$, $p < .001$]. The results showed that the listeners perceived both of the level A1 and falling A2 tones as most similar to the statement/backchannel intonation (*Right/right*), and the B1 tone as most similar to the question intonation (*Right?*). The difference between the A1 and A2 tones in terms of the categorization as *Right/right* was not significant.

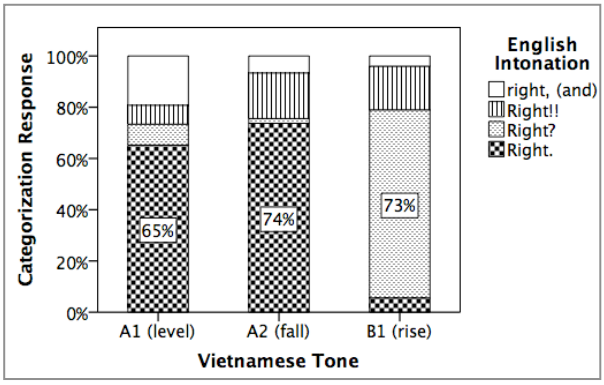


Figure 3: Categorization of the three Vietnamese tones in terms of English intonation patterns by native English speakers.

Mean goodness-of-fit index for each of the tone-intonation mappings identified above using the 50% response criterion is shown in Table 1. The difference between the goodness-of-fit of the level A1 and falling A2 tones to the same intonation pattern category was significant, [$t(373) = -3.71$, $p < .001$]. As the A1 and A2 tones were perceptually assimilated to the same native category to a different degree, this mapping type was classified as category-goodness assimilation (CG). A1-A2 and A2-B1 mappings were classified as two-category assimilation (TC). The discrimination accuracy was predicted to be better in TC than in CG [1, 4]. However, this prediction was not borne out as the results of the discrimination experiment showed, as reported below.

Table 1. *A summary of mapping between Vietnamese lexical tones and English intonation patterns, according to Perceptual Assimilation Model.*

| Tone Pairs | Categorization (> 50%) | Goodness-of-Fit (*M*) | Mapping Type |
|---|---|---|---|
| A1 - A2 | *Right/right* | 4.48 - 4.99 | CG |
| A1 - B1 | *Right/right - Right?* | 4.48 - 5.77 | TC |
| A2 - B1 | *Right/right - Right?* | 4.99 - 5.77 | TC |

## 3.2. Discrimination of Tone Contrasts

The analyses of d-prime score variance revealed that the effect of contrast was significant in tone discrimination responses, [$F(2,18) = 12.88$ $p < .001$]. Specifically, the A1-A2 contrast was discriminated better than the A1-B1 contrast, [$F(1,9) = 7.98$, $p = .020$], but not the A2-B1 contrast (n.s.). The difference between the discrimination of the A1-B1 and A2-B1 contrasts was also significant $F(1,9) = 40.71$, $p < .001$]. Figure 4 illustrates these results.

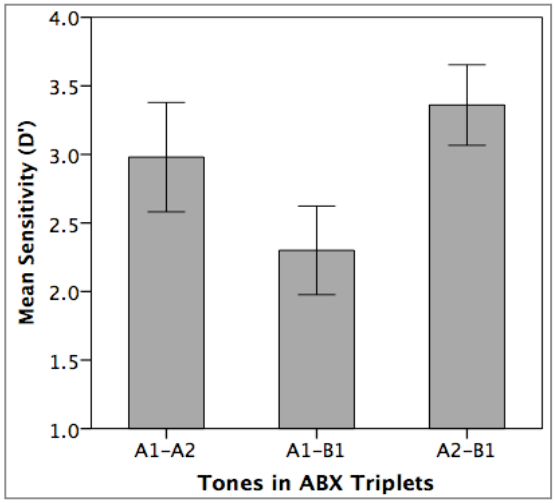


Figure 4: Discrimination of the three Vietnamese tone contrasts by native English speakers.

# 4. Discussion

This study examined the possibility of cross-language mapping between lexical tones and intonation patterns, and tested the consequences of these mappings for sound discriminination, as predicted by PAM [1, 4]. English listeners were able to reliably identify falling, level, and rising lexical tones as similar to native intonation patterns of statement/backchannel and question. This suggests that the mapping between word-level and phrase-level $F_0$ contours is possible, at least when these contours are manifested over single-syllable utterances.

These results are different from the report in [4] that no strong cross-language assimilation was found between Thai

lexical tones and English intonation patterns, and thus, lexical tones are 'uncategorized" speech sounds for native English speakers. In both studies, the 50% criterion was used to assess the strength of assimilation patterns. By this criterion, the assimilation patterns were definitely strong (consistent) in this study as they averaged at 65%, 74% and 73% of responses (Figure 3). Perhaps, other methodological choices in the two studies can explain the discrepancy in results. First, Reid at al. used intonation patterns that may have either rising or falling $F_0$ contour, such as the command and uncertainty patterns [11, 12]. Inconsistent categorization in [4] could have stemmed from this "one label - two contours" type of the response category. Second, the process of categorization seemed to rely on metalinguistic awareness of intonation patterns more heavily in [4] than in this study. However, such awareness in participants may be low in general, which could have affected their responses [15]. Lastly, not only target Thai tones but also the mode of tone perception was varied in [4], namely, auditory-visual, auditory-only, and visual-only. In this complicated study design, native English speakers might have been less consistent in their categorization responses than in our relatively simple experiment design.

Falling, level, and rising lexical tones occur in many tone languages, including Thai and Vietnamese, which were investigated in [4] and in this study, respectively. Table 2 provides a summary of cross-language mappings between these three tones and English intonation patterns. A comparison of the findings suggests that the falling and level tones tend to assimilate to the same intonation pattern in English, whereas the rising tone tends to assimilate to a different intonation pattern, whatever particular choices of categorization labels are. (Similar results were reported for cross-language tone assimilation, e.g., Thai, on one side, and Mandarin and Cantonese, on the other side, [4]). Such consistency across languages suggests that something may be special about rising as opposed to non-rising tones [4, p. 586]. Perhaps, rising $F_0$ contours are less subjected to the perceptual normalization due to the $F_0$ declination [16], and this influences the categorization of rising tones.

Table 2. *Categorization of three lexical tones by native English speakers: A comparison between two studies.*

| Tones | Reid et al. [4] | This study |
|---|---|---|
| Falling | Flat Pitch | Statement / Backchannel |
| Level | Statement / Flat Pitch | Statement / Backchannel |
| Rising | Command | Question |

Cross-language mapping patterns and the degree of perceived similarity between non-native and native $F_0$ contours did not fully predict the discrimination accuracy of tones in this study. Based on the categorization results, tone discrimination was expected to be worse in the level versus falling contrast than in the other two contrasts. This expectation was not borne out, and the tone contrast with the lowest discrimination accuracy turned out to be the level versus rising contrast. This result may be explained by a perceptual normalization effect as suggested by an anonymous reviewer [16]. Namely, it may be the case that listeners normalize for $F_0$ declination in level tones, which would result in the perception of plateau $F_0$ contours as having a rise. Alternatively, this relatively low discrimination accuracy may be explained by a native language bias of English listeners naïve to tone languages: In tone discrimination tasks, they pay attention to the acoustic dimension of $F_0$ height rather than to the dimension of $F_0$ direction [17, 18]. Figure 1 shows that the level and rising tones in this study were much higher in $F_0$ than the falling tone. Listeners who paid most attention to this dimension could have found the level and rising $F_0$ contours confusable with each other, but distinct from the falling contour.

The two explanations of the relative difficulty in discerning the difference between the level and rising tones outlined above are not mutually exclusive. Both perceptual normalization for $F_0$ declination and perceptual bias towards the $F_0$ height dimension could have determined the observed discrimination accuracy of level-rising < level-falling, rising-falling in native English listeners. The finding that this discrimination accuracy order holds regardless of speakers' native language background (a tone or a non-tone language) further supports the hypothesis that more than one factor is at play here [6]. So and Best found that the discrimination accuracy was lower for the Mandarin level-rising tone contrast than for the rising-falling contrast in all of their Cantonese, Japanese, and Canadian English groups. Overall, such findings suggest that for accurate predictions of sound discrimination accuracy we should consider not only perceived acoustic-phonetic similarity as in PAM, but also universal and native language biases in processing of $F_0$ information.

This study has several limitations. First, the experiments were deliberately confined to simple designs in which the cross-language mapping of selected prosodic categories could be easily tested. Second, the selection of five English intonation patterns as candidates for mapping was motivated by previous literature and pilot recordings; however, due to a large individual variability among speakers, the patterns used here could be idiosyncratic to our English informant [15, 18]. Different English intonation patterns could have been selected, as in [4]. These methodological choices impose some limitations on generalization of the findings to the cross-language mapping of lexical tones and English intonation patterns. Last but not least, the perceived similarity between tones and intonation patterns was predicted only on the basis of their overall $F_0$ contours. While such approach is warranted by the PAM framework, it does not answer the question of whether in categorization tasks English listeners use their native language phonological categories that differentiate discourse-pragmatic contexts or simply compare $F_0$ contours of native and non-native sounds. In order to investigate this question, a mapping experiment between two nonse words with $F_0$ contours acoustically similar to English intonation patterns can be conducted in future research.

## 5. Acknowledgements

The author is thankful to Tâm Nguyễn at the Vietnam National University-HCMC and Nữ Nguyễn at Louisiana State University (LSU), for their help with Vietnamese stimuli recordings; Chris Barrett (LSU) for her recordings of English intonation patterns; and LSU community for data collection and support. The feedback of two anonymous reviewers had terrific suggestions that benefited the result discussion in this paper. This work was supported by the Manship Grant, Humanities & Social Sciences Summer Research Funding, LSU.

# 6. References


[1] C. T. Best, "A Direct Realist view of cross-language speech perception," *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 171- 204, 1995.

[2] P. Kuhl and P. Iverson, "Linguistic experience and the "Perceptual Magnet Effect," *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 121- 154, 1995.

[3] S.G. Guion, J.E. Flege, R. Akahane-Yamada, and J.C. Pruitt, "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," *Journal of the Acoustical Society of America*, vol. 107, no. 5, pp. 2711-2724, 2000.

[4] A. Reid, D. Burnham, B. Kasisopa, R. Reilly, V. Attina, N.X. Rattanasone, and C.T. Best, "Perceptual assimilation of lexical tone: The roles of language experience and visual information," *Attention, Perception, & Psychophysics,* vol. 77, pp. 571-591, 2015.

[5] P.A. Hallé, Y.-C. Chang, and C.T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," *Journal of Phonetics*, vol. 32, pp. 395-421, 2004.

[6] C.K. So and C.T. Best, "Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences," *Language and Speech*, vol. 53, no. 2, pp. 273-293, 2010.

[7] P. Boersma and D. Weenink, *Praat: doing phonetics by computer* [software], 2014.

[8] M. Brunelle, "Tone perception in Northern and Southern Vietnamese," *Journal of Phonetics*, vol. 37, pp. 79-96, 2009.

[9] J.P. Kirby, "Illustration of the IPA: Vietnamese (Hanoi Vietnamese)," *Journal of the International Phonetic Association*, vol. 41, no. 3, pp. 381-392, 2011.

[10] J.B. Pierrehumbert, *The phonology and phonetics of English intonation*. Bloomington, Indiana: Indiana University Linguistics Club, 1980.

[11] J.C. Wells, *English intonation*: Cambridge University Press, 2006.

[12] J. Pierrehumbert and J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in P. Cohen, J. Morgan, and M.E. Pollack (eds), *Intentions in communication*. Cambridge, Mass.: MIT Press, pp. 271-311, 1990.

[13] S. Benus, A. Gravano, and J. Hirschberg, "The prosody of backchannels in American English," in *ICPhS XVI – International Congress for Phonetic Sciences, August 6-10, Saarbrücken*, *Germany, Proceedings, 2007,* pp. 1065-1068.

[14] N.A. Macmillan and C.D. Creelman, *Detection theory: A user's guide*. New Jersey: Lawrence Erlbaum Associates, Publishers, 2005.

[15] T. Bent, *Perception and production of non-native prosodic categories*. Northwestern University, Evanston, 2005.

[16] J. Pierrehumbert, "The perception of fundamental frequency declination," *Journal of Acoustical Society of America*, vol. 66(2), pp. 363-369, 1979.

[17] J.T. Gandour and R.A. Harshman, "Crosslanguage differences in tone perception: A multidimensional scaling investigation," *Language and Speech,* 21, pp. 1-33, 1978.

[18] B. Chandrasekaran, P.D. Sampath, and P.C.M. Wong, "Individual variability in cue-weighting and lexical tone learning," *Journal of the Acoustical Society of America*, 128(1), pp. 456-465, 2010.