A perceptual study of the five level tones in Hmu (Xinzhai variety)

Wen Liu

School of Literature, Shandong University, China

liuwen3214@163.com

Abstract

Previous studies have shown that the perception is categorical when tones have different contours, whereas continuous when tones have the same contour. In this study, a perceptual experiment of the five level tones in Hmu (Xinzhai variety) was conducted to further examine this conclusion. Results show that in the identification test, continua between different level tones have different boundary width, which has a negative correlation with the pitch interval of two level tones. In the discrimination test, though there is no peak in discrimination curve, the discrimination accuracy reveals an important phenomenon that the accuracy is approximately 50% between two neighboring level tones, but higher when the level tones have a larger pitch interval. Besides, the boundary width is highly correlated with the discrimination accuracy (e.g., the narrower the boundary width, the higher the discrimination accuracy). These results reveal the basic characteristic of continuous perception, especially for level tones. Finally, the results also demonstrate that the category in categorical perception is not equal to phonological category.

Index Terms: speech perception, tone, five level tones, Hmu

1. Introduction

Categorization is a distinctive feature of speech perception. Categorical perception means that listeners can divide the continuous speech sounds into discrete phonological categories. The small change in acoustic continuum can be perceived among different phonological categories, while the same acoustic difference cannot be perceived in the same category [1-2]. The principle of categorical perception was developed, in part, to explain how listeners cope with the many-to-one mapping between acoustic patterns and phonological categories [3]. Generally, categorical perception has three characteristics [4]: (1) In the identification function, there is a sharp boundary between two categories; (2) In the discrimination function, the accuracy peak is at the category boundary, but is at or near chance level within category; (3) The discrimination function can be predicted from the identification function.

In terms of linguistics, speech perception contains two major aspects: segment (consonant and vowel) and suprasegment (e.g., tone). Previous studies have shown that stop consonants belong to categorical perception [5], while vowels belong to continuous perception [6-7]. At present, there is no agreement on whether tone perception is categorical or continuous. In general, the perception is categorical when tones have different contours, such as high rising vs. high level tones in Mandarin (e.g., [8-11]) and Cantonese [3], whereas continuous when tones have the same contour, such as three level tones in Thai (e.g., [12]) and two rising tones in Cantonese (e.g., [3]). Based on our research, the reason for categorical perception lies in that there is a significant slope difference between level and contour tone, which can be easily perceived by native speakers. However, there is only pitch height difference among level tones, and it is difficult for native speakers to perceive the continuous change.

Given that former studies mainly focused on the different contours in tone perception, this study chooses a language (Hmu (Xinzhai variety)) with different pitch heights that have the same contour. This language is spoken by people who reside in a small mountain village, Guizhou Province, China. It is regarded as the eastern dialect of the Hmongic language [9], which is classified as a member of the Hmongic subgroup of the Hmong-Mien language family [14-15]. It has eight contrastive phonemic tones in isolated syllables: five level tones, two rising tones and a falling tone [16]. Note that five level tones in one language are an important characteristic in view of tone typology. Those tones are marked using the five letter tone system, where 5 indicates the highest level and 1 the lowest [17], that is, T11, T22, T33, T44, and T55 respectively. Note that though the five level tones can be distinguished from pitch in tonal space, the lowest level tone (T11) has also been reported to use breathy voice [16, 18-19].

This study expects to find characteristics and mechanism of tone perception among the five level tones in Hmu (Xinzhai variety) by adopting behavioural experiment to test the native speakers' perception. In addition, whether the tone inventories with the same contour can bring about some effects on the sensitivity of tone perception is also examined. The answer to such questions not only strengthens our knowledge of tone perception, but also helps to provide direct/indirect evidence for discussing the relationship between category in categorical perception and phonological category.

2. Materials and methods

2.1. Materials

Ten tone continua (T11-T22, T11-T33, T11-T44, T11-T55, T22-T33, T22-T44, T22-T55, T33-T44, T33-T55, T44-T55) were constructed by manipulating f0 from each of the five syllables ([ta11], 'die'; [ta22], 'to throw'; [ta33], 'underground'; [ta44], 'frost'; [ta55], 'to come') to the others, with the PSOLA (pitch-synchronous overlap and add) function in Praat [20]. The five syllables were pronounced by a 25-year-old male native speaker. The recording was taken using *Audition* 2.0 at a sampling rate of 44.1 kHz, 16-bit resolution.

Table 1: *Original normalized 11 f0 values (Hz) of the five level tones.*

	1	2	3	4	5	6	7	8	9	10	11
T11	135	136	137	137	136	136	136	135	134	136	140
T22	167	165	166	165	165	165	164	161	162	163	161
T33	191	189	188	190	190	189	189	189	188	187	187
T44	209	208	210	214	216	216	213	211	213	215	215
T55	231	232	234	238	238	237	239	240	240	240	240

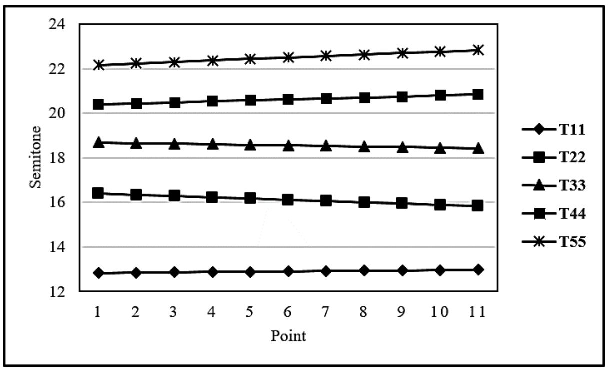


Figure 1: *Pitch trajectories of the five level tones.*

The major procedures of resynthesizing stimuli were: 1) Extract f0 values of original tones in Praat, each vowel was divided into 11 parts of equal duration (Table 1 and Figure 1). 2) Maintain the duration of original tones based on the direction of resynthesis, e.g., if resynthesising T11-T22 based on T11, then the duration of T11 is used. The duration of T22 is used when resynthesising T22-T33 based on T22. The duration of original tones was retained in the process of resynthesis in this study because there were no significant differences in duration among the five level tones at $p = 0.05$. 3) By manipulating the 11 f0 values accordingly, 10 steps of stimuli in each of the ten tone continua were resynthesized. Therefore, a total of 100 stimuli were obtained. Figure 2 showed the schematic diagram of f0 manipulation of the 10 stimuli for the ten tone continua.

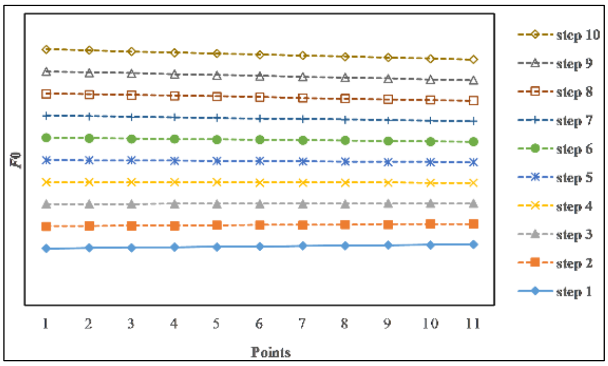


Figure 2: *Schematic diagram of 10-step f0 manipulation. The solid line indicates the normalized f0 values of original tone. The dashed lines indicate the f0 values of resynthesized samples.*

2.2. Participants

Thirty-six participants (20 M and 16 F, mean age=27.4, SD=10.3) were recruited for this study. All of them were native speakers. No participant reported any speech, language, or hearing difficulty.

2.3. Procedure

In this experiment, each participant was asked to perform identification and discrimination tasks. The two tasks were presented in random order across participants. Participants' responses and reaction time (RT) were recorded.

In the identification task, the stimuli in each continuum were presented to participants in random order, and they were asked to press the left or right arrow key on the keyboard to identify stimuli they heared (two-alternative forced choice, 2AFC). The 10 stimuli of each continuum were repeated twice, forming a total of 200 trials. The order of the stimuli presentation was randomized across participants.

In the discrimination task, the AX mode was used. For each tone continuum, a total of 26 pairs were presented in random order, 16 consisted of two different stimuli separated by two steps (different pairs), in either forward (1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 8-10) or reverse order (3-1, 4-2, 5-3, 6-4, 7-5, 8-6, 9-7, 10-8), the other 10 consisted of the 10 stimuli on the continuum each paired with itself (same pairs), namely, 1-1, 2-2, 3-3, 4-4, 5-5, 6-6, 7-7, 8-8, 9-9, 10-10. So, the ten continuum formed a total of 260 trials. Stimuli were presented in pairs with a 500 ms inter-stimulus interval (ISI). This ISI duration was selected to maximize differences in the performance of between-versus within-category discrimination [7]. After hearing each pair, participants were instructed to judge whether the two stimuli were the same or different.

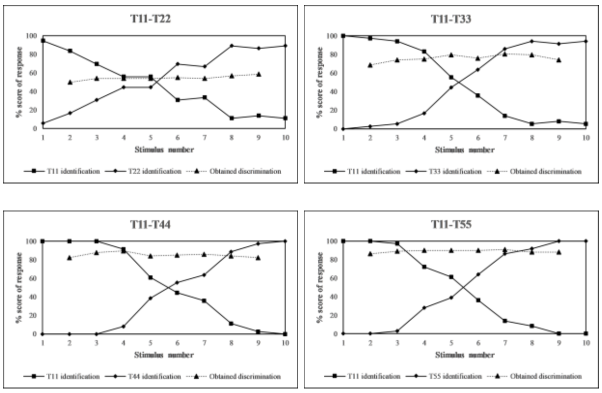
Prior to the formal experiments, participants were asked to perform a practice block using 20 stimuli selected randomly from the stimuli pool. These two practice tasks were designed primarily to familiarize them with task requirements in order to stabilize their performance in the following experiment.

2.4. Data analysis

To investigate the effects of pitch on identification and discrimination performance, boundary position, boundary width, and discrimination accuracy were obtained in this study. For a particular stimulus, the identification score was defined as the percentage of responses with which participants identified that stimulus as being either one or the other. Boundary position and width were assessed by probit analyses of identification curves: the boundary position was defined as the 50% crossover points, and the boundary width was defined as the linear distance between the 25% and 75% as determined by the mean and standard deviation obtained from probit analysis [8]. To obtain the discrimination scores of each pair, calculation was made using the formula described by Xu, Gandour, and Francis [11].

3. Results

Identification and discrimination curves are shown in Figure 3. The estimated boundary position and width, discrimination accuracy, are shown in Table 2.



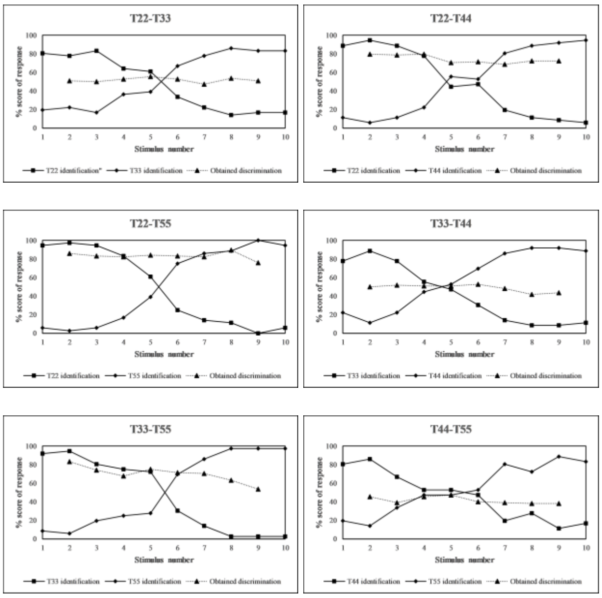


Figure 3: *Identification (solid lines) and discrimination (dashed line) curves of the five level tones.*

Table 2: *Derived boundary position and width, discrimination accuracy (%), and pitch interval (Hz) for each tone continuum.*

Continuum	Boundary position	Boundary width	Discrimination accuracy	Pitch interval
T11-T55	5.45	2.13	88.89	10.10
T11-T44	6.07	2.45	85.30	7.65
T22-T55	5.36	2.33	83.45	7.32
T11-T33	5.50	2.33	75.93	5.26
T22-T44	5.35	3.19	74.07	4.87
T33-T55	5.16	2.63	69.79	4.84
T11-T22	5.01	4.33	54.28	2.78
T22-T33	5.12	4.9	51.74	2.48
T33-T44	4.58	4.04	48.73	2.39
T44-T55	5.00	5.43	41.44	2.45

Table 2 demonstrates a phenomenon, which is that there is a link between tone continuum and boundary width. To be specific, if two level tones in a tone continuum have a small pitch interval, the boundary width is large, and vice versa. However, if two level tones have a large pitch interval, the value of boundary width will be smaller. Those tone continua can be divided into four groups based on pitch intervals: Group 1 (T11-T22, T22-T33, T33-T44, T44-T55), Group 2 (T11-T33, T22-T44, T33-T55), Group 3 (T11-T44, T22-T55), and Group 4 (T11-T55). There is a statistically significant difference between groups as determined by one-way ANOVA ($F(3,6) = 15.087$, $p = 0.003$).

Though there is no obvious peak in the discrimination curve, some results that are different from previous studies are observed. For example, the discrimination accuracy is not the same, which varies depending on tone continua. The discrimination accuracy approximates 50% (the chance level) between two neighboring level tones, such as T11-T22, T22-T33, T33-T44, and T44-T55. However, if the two level tones have a moderate pitch interval, such as T11-T33, T22-T44, and T33-T55, the discrimination accuracy improves significantly, ranging from 70% to 80%. If the pitch interval is large, such as T11-T44, T22-55, and T11-T55, the accuracy is on a fairly high level, between 80% and 90%. There is also a statistically significant difference between continua as determined by one-way ANOVA ($F(9,70) = 127.495$, $p < 0.001$).

When integrating data of identification and discrimination test, it is not difficult to find that the boundary widths are highly correlated with discrimination accuracies between different level tones, where the correlation coefficient is 0.8911 (see Table 2). To be specific, the smaller the boundary width, the higher the discrimination accuracy. On the contrary, there is a lower discrimination accuracy when the two level tones have a big boundary width.

Table 2 also illustrates the influence of pitch interval on the discrimination accuracy. In our experiment, 10 steps of stimuli in each of the ten tone continua were resynthesized, and the pitch interval of one step varied in each continuum. It is not difficult to see that the pitch intervals are highly correlated with the discrimination accuracies between different level tones, and the correlation coefficient is 0.8737. That is, the discrimination accuracy is high when the pitch interval of one step is large and vice versa. The discrimination accuracy is about 50%, 70%, 80%, and 90% when the mean pitch interval of one step is 3Hz, 5Hz, 7Hz, and 10Hz respectively.

4. Discussion

Previous studies have presented conflicting results with respect to the categorical nature of tone perception. The main discrepancy between Wang [10] and Abramson [12] is that they used different tone contours. Studies later showed that tone contour plays an important role in (the mode of) tone perception. In addition to categorical and continuous perception, Hallé, Chang, and Best [8] demonstrated another view, which was that there is a gradient in the degree of categorical perception. That is to say, there is no strict dichotomy between categorical and continuous.

If defining categorical perception based on the traditional definition, which is that there is a crossover point in the identification curve corresponding to the peak in the discrimination curve, there is no doubt that the perception of the five level tones in Hmu (Xinzhai variety) belongs to continuous perception because there is no sharp boundary in identification curve and no peak in discrimination curves. However, the discrimination curve of the five level tones shows a very important phenomenon, which has not been reported in past studies. Although Abramson [12] considered the three level tones as non-categorical perception in Thai, a phenomenon should be noticed: the discrimination tests yielded a high level of discrimination accuracy across the continuum with no effects of boundaries, ranging from 92% to 96%, which is far greater than the chance level (50%). Similarly, Francis, Ciocca, and Ng [3] also found that the discrimination curve of three level tones is flat, and the discrimination accuracy ranges from 70% to 80%. What those studies share in common is that their discrimination curves are leveled, which is different from our study which shows that the discrimination accuracy varies with the pitch intervals of different level tones. Besides, a rule was found from the data of the five level tone, which is that the boundary width in identification test is highly correlated with the discrimination accuracy in discrimination test, and the mean pitch interval is also highly correlated with the discrimination accuracy. This phenomenon illustrates that when the stimuli pairs have a large pitch interval, they are easier to be judged as different by listeners, which can be reflected by the narrow boundary width and the high discrimination accuracy, whereas the boundary width is wide and the discrimination accuracy low when the stimuli pairs have a small pitch interval. Based on such understandings, the fact we have to admit that the continuous perception obtained from tones with the same contour has its own characteristic, which is different from the perception of tones with different contours.

More importantly, every tone in a tonal language has its phonological category because it can be used to distinguish lexical meaning. However, the phonological category of level tones cannot be found based on the categorical perception. In fact, phonemic perception involves many factors, and the results of Hmu (Xinzhai variety) illustrate that it is inappropriate to simply use the performance of identification and discrimination test as the equivalent of phonological category.

5. Conclusion

In this study, the perception of five level tones in Hmu (Xinzhai variety) was investigated. From the identification and discrimination curves, changes of parameters from one level tone to another are less abrupt than the classical categorical perception, such as the high level and high rising tone in Mandarin Chinese. This is also reflected in the broader boundary width in identification curves. Meanwhile, a very important phenomenon was observed in this study, which is that the boundary width has a negative correlation with the pitch intervals of different level tones. If the continuum has a small pitch interval, the boundary width is large, and vice versa. In the discrimination task, though there is no peak in discrimination curve, the discrimination accuracy varies with pitch intervals (e.g., the larger the pitch interval, the higher the discrimination accuracy). Moreover, the boundary widths are highly correlated with the discrimination accuracies (e.g. the smaller the boundary width, the higher the discrimination accuracy). Compared to previous studies, the results of this study strengthen our understanding of the so-called ‘continuous perception’.

6. Acknowledgements

This study was supported by The Fundamental Research Funds of Shandong University (No. 2019GN086). I would like to thank Zhenghui Yang for the assistance of data collecting in the fieldwork.

7. References

[1] A. Liberman, "Some Results of Research on Speech Perception", *The Journal of the Acoustical Society of America*, vol. 29, no. 1, pp. 117-123, 1957. Available: 10.1121/1.1908635.

[2] A. Liberman, F. Cooper, D. Shankweiler and M. Studdert-Kennedy, "Perception of the speech code.", *Psychological Review*, vol. 74, no. 6, pp. 431-461, 1967. Available: 10.1037/h0020279.

[3] A. Francis, V. Ciocca and B. Chit Ng, "On the (non)categorical perception of lexical tones", *Perception & Psychophysics*, vol. 65, no. 7, pp. 1029-1044, 2003. Available: 10.3758/bf03194832.

[4] B. Repp, "Categorical perception: Issues, methods, findings", in Lass N. J. (eds.), *Speech and Language: Advances in Basic Research and Practice*, N. Lass, Eds. New York: Academic Press, 1984, pp. 243–335.

[5] A. Liberman, K. Harris, H. Hoffman and B. Griffith, "The discrimination of speech sounds within and across phoneme boundaries", *Journal of Experimental Psychology*, vol. 54, no. 5, pp. 358-368, 1957. Available: 10.1037/h0044417.

[6] D. Fry, A. Abramson, P. Eimas and A. Liberman, "The Identification and Discrimination of Synthetic Vowels", *Language and Speech*, vol. 5, no. 4, pp. 171-189, 1962. Available: 10.1177/002383096200500401.

[7] D. Pisoni, "Auditory and phonetic memory codes in the discrimination of consonants and vowels", *Perception & Psychophysics*, vol. 13, no. 2, pp. 253-260, 1973. Available: 10.3758/bf03214136.

[8] P. Hallé, Y. Chang and C. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners", *Journal of Phonetics*, vol. 32, no. 3, pp. 395-421, 2004. Available: 10.1016/s0095-4470(03)00016-0.

[9] G. Peng, H. Zheng, T. Gong, R. Yang, J. Kong and W. Wang, "The influence of language experience on categorical perception of pitch contours", *Journal of Phonetics*, vol. 38, no. 4, pp. 616-624, 2010. Available: 10.1016/j.wocn.2010.09.003.

[10] W. Wang, "Language change", *Annals of the New York Academy of Sciences*, 208, pp. 61–72, 1976.

[11] Y. Xu, J. Gandour and A. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction", *The Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 1063-1074, 2006. Available: 10.1121/1.2213572.

[12] A. Abramson, "The noncategorical perception of tone categories in Thai", in *Frontiers of Speech Communication Research*, B. Lindblom and S. Ohman, Eds. London: Academic Press, 1979, pp. 127–134.

[13] F. Wang, *A Sketch of the Hmongic Language*. Beijing: Nationalities Press, 1985.

[14] M. Ratliff, *Hmong-Mien language history*. Canberra: Pacific Linguistics, 2010.

[15] F. Wang and Z. Mao, *Reconstruction of the Sound System of Proto-Hmong-Mien*. Beijing: China Social Sciences Press, 1995.

[16] W. Liu, Y. Lin, Z. Yang and J. Kong, "Hmu (Xinzhai variety)", *Journal of the International Phonetic Association*, pp. 1-18, 2018. Available: 10.1017/s0025100318000336.

[17] Y. Chao, "A system of tone letters", *Le maître phonétique*, vol. 45, pp. 24–27, 1930.

[18] W. Liu, "An acoustic, perceptual, and neurophysiological study of the five level tones in Hmu (Xinzhai variety)", in *UCLA Phonetics Seminar*, 2018.

[19] W. Liu and J. Kong, "A study on phonation patterns of tones in Hmu (Xinzhai variety)", in *The 9th Conference of International Evolution Linguistics* (CIEL9), 2017.

[20] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer", 2009. Available: http://www.fon.hum.uva.nl/praat/.