Advances in Social Science, Education and Humanities Research, volume 631
Proceedings of the 2021 International Conference on Social Development and Media Communication (SDMC 2021)

The Perception of Tone 2 and Tone 3 in Mandarin
An Experimental Study on Tones with Low Falling Contours in Monosyllabic Words

Ruiting Dong*

School of Humanities, University of Amsterdam, Amsterdam, 1017 CE, Netherlands
*Corresponding author. Email:20152917@stu.nun.edu.cn

ABSTRACT
This paper explores tone2 and tone3 in monosyllabic words in Mandarin, aiming to illustrate the contributing factors that influence the perception of tone2 and tone3 with low-falling contours. Previous studies mainly focus on inflection position and onset or final F0 values. As the pitch value of the inflection point is a vital acoustic parameter for tone3 identification, this study brings in inflection F0 values as one of the focuses. To reveal the crucial features of tone2 and tone3 identification, this study applies a series of synthetic sounds with identification tasks to investigate three factors that influence tone perception in Mandarin, inflection F0 values, final F0 values, and inflection position, both as independent variables and combined variables. Empirical results suggest that an individual variable cannot determine the perception of tone2 and tone3. Only combinations of variables can decisively alter tone identification. The inflection position plays a minor role in identifications of tone2 and tone3, contradicting with findings from previous researches. The status of inflection F0 values seems to be crucial in tone3 perception. The relationship between inflection F0 values and inflection positions is mutually promotion and compensation for tone3 identification. Final F0 values function as compensation for the emphasis on T2 perception caused by inflection F0 values. The perception of tone2 and tone3 does not show a categorical pattern at a stimuli continuum with varying inflection F0 values.

Keywords: *Tone2, Tone3, Perceptual pattern, Identification, Mandarin.*

1. INTRODUCTION

There are four basic tones in standard Mandarin Chinese, the high-level tone (tone1), the rising tone (tone2), the falling-rising tone (tone3), and the falling tone (tone4). In the following paper, 'T2' and 'T3' refer to tone2 and tone3 in this paper for conciseness. Both T2 and T3 depict rising tone contours. In nature speech, T2, the rising tone in standard Chinese, does not always depict an exact rising tone contour. Instead, the tone contour can sometimes show a rising contour followed by a slightly falling onset. This pattern is to some extent similar to T3, the falling-rising tone, in standard Chinese [1][2][3][4]. Tones contours of T2 and T3 in Figure 1 show the achievability of T2 with a falling onset. With similar tone contours, it can be difficult for speakers to perceive and differentiate between T2 and T3. Chuang et al. conducted a study on identifying the four tones in Mandarin Chinese and found out that participants frequently made mistakes in identifying between T2 and T3 [5]. Furthermore, the confusion among participants of the two tones is asymmetrical. Specifically, the error rate of incorrectly perceiving T3 is two times higher than that of incorrectly perceiving T2. Chuang's results indicate that the perception and identification of T2 and T3 are not as easy as the perception of other tone pairs [5].

Copyright © 2022 The Authors. Published by Atlantis Press SARL.
This is an open access article distributed under the CC BY-NC 4.0 license -http://creativecommons.org/licenses/by-nc/4.0/.

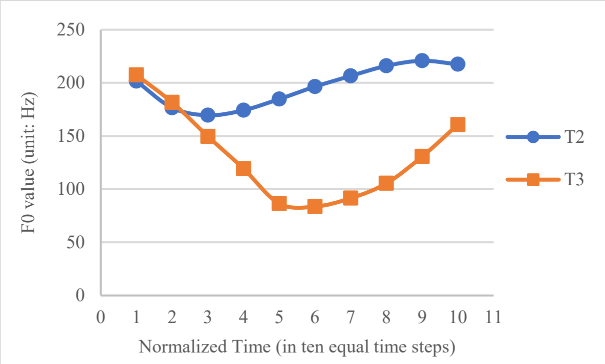

Figure 1 Examples of T2 and T3 contours with falling onsets [5].

1.1. Perception Pattern of T2 and T3

Categorical perception, as a vital pattern in phonetic perception, refers to a continuum of stimuli being perceived as scattered and limited categories, in other words, phonemes. Participants can hardly recognize the phonetic differences in one category but can be sensitive to the differences between categories [6]. Opposite to categorical perception, if a continuum of stimulus cannot be perceived as scattered and limited categories, named as phonemes, then the pattern belongs to continuous perception.

Previous studies show that some scholars suppose that the perception pattern of T2 and T3 is continuous, while other scholars hold the view that the perception pattern is categorical. Liu did an empirical study and illustrated that tone perception of Chinese native speakers did not show a categorical pattern on a T3 continuum, neither at the inflection position nor the onset F0 value [7].

However, Liu's results faced challenged by other empirical studies. Shen and Lin studied the categorical perception of T2 and T3 among Chinese native speakers [8]. They created a tone continuum of T2 and T3 by changing the position of inflection. The results suggest that the perception pattern of T2 and T3 is categorical. Wang and Li provided further support for Shen and Lin's arguments as they figured out the categorical boundary of inflection positions and final F0 values in the perception of T2 and T3 [9]. Wang and Li discovered another explanation of Liu's data by altering Liu's analytical methods. They concluded that T2 and T3 perception does show a categorical trend at onset F0 values and inflection positions [9]. In a later study, Wang and Qin stated that keeping the onset F0 value unchanged, when the final F0 is high, the perception of T2 and T3 demonstrates an obvious categorical pattern along with the inflection point moving back [10].

1.2. Influential Factors of T2 and T3 Differentiation

Empirical studies on the perception of T2 and T3 suggest that pitch values and tone contours are the two main acoustic parameters in Mandarin Chinese tone perception. Fang and Jin applied a multi-dimensional scaling analytical approach to explore the acoustic parameters of tone perception in Mandarin, which results illuminate that pitch values and rate of pitch changes are the two acoustic dimensions of tone perception [11]. In other words, listeners mainly rely on pitch values (pitch values of the fundamental frequency) and rate of pitch changes (tone contours) to perceive tones.

Yang et al. suggest that it is impossible to synthesize four tones in Mandarin through duration or intensity adaption after conducting a series of speech synthesis trials [12]. However, by altering only the F0 values, the correct identification rates of the four synthetic tones can reach 95%. This conclusion supports that the F0 value is likely to be a crucial acoustic parameter for tone perception in Mandarin.

Shen et al. suggest that the inflection position in falling-rising tones and the initial tone falling part are crucial to differentiate between T2 and T3 [8]. As mentioned above, Wang and Li discover that the perception of T2 and T3 among mother-tongue speakers show a categorical pattern on inflection positions and final F0 values [10]. Therefore, the inflection position and the final F0 value play essential roles in T2 and T3 perception.

In general, the inflection position and F0 values exert great impacts on the perception of T2 and T3 in Mandarin. Therefore, in this study, F0 values and inflection positions are treated as the main research emphases.

1.3. T2 and T3 with Low-falling Onsets

Most scholars agree that "low in pitch value" is a significant feature of T3 [13][14]. But whether the "low in pitch value" should be treated as "low-flat" or "low-falling" revealed in tone contours is still in controversy.

Yip suggests that it is unnecessary to distinguish "low-flat" and "low-falling" as they are both variations of T3 [15]. Unlike Yip's view, Cao conducted a study based on monosyllabic level tones and figured out that participants did not show a clear tendency towards low-level tones, Hence, it is not appropriate to describe T3 as low-flat. Wang et al. provide further support for the critical status of low-falling in T3 perception. In a continuum of stimuli with flat onsets but varying in inflection positions, Chinese native speakers perceive stimuli as T2 in almost all inflection positions, and the perception ratio of T2 identification is over 85 percent

[9]. Qin experimented with a series of synthetic low-level monosyllable continuums and found out that despite the final F0 values, native speakers tend to perceive the low-level rising tone as the rising tone in monosyllables instead of T3 [16]. Previous studies indicate that it is inappropriate to regard “low-flat” as the phonological basis of T3 perception. Therefore, this study treats tones with low-falling contours as the research targets.

1.4. Research Questions and Hypotheses

Previous empirical studies with regards to tone categorical perception provide many referential conclusions and experimental methods. However, the crucial and dominant factor(s) that influence the differentiation and perception of T2 and T3 needs further clarification. Besides, whether the perception of T2 and T3 is categorical or not is still controversial. According to the tone contours of T2 and T3 in Figure 1, compared to T2, the falling contour of T3 shows a more considerable range of pitch changes and the falling duration of T3 is longer, in other words, lower inflection F0 values, lower final F0 values, and latter inflection position. Therefore, focusing on T2 and T3 with low-falling contours in monosyllabic words, this study applies identification experiments to investigate whether inflection F0 values, final F0 values and inflection position solely or jointly influence the perception of T2 and T3. Previous studies on the perception of T2 and T3 mainly focus on the inflection position and onset/final F0 values with no empirical studies on inflection F0 values. Therefore, in experiment 2, this paper explores the perception pattern of T2 and T3 at synthetic stimuli continuum with different inflection F0 values.

Based on Yang’s emphasis on F0 values in tone perception, I suggest that the F0 values play a crucial role in T2 and T3 perception compared to other phonological features. Therefore, I hypothesize that in monosyllabic words, F0 values influence the perception of T2 and T3 as a dominant factor. A lowered inflection F0 leads to a decrease in the identification of T2. Combined with a lowered final F0 or/and a latter inflection position, a lowered F0 can result in a sharper decrease in T2 identification. In terms of the perception pattern of T2 and T3, I hypothesize that it presents a categorical perception pattern when native speakers perceive T2 and T3 stimuli with different inflection F0 values.

2. EXPERIMENT 1

2.1. Stimuli

The monosyllable used in this study is phoneme /ba/ in T2 and T3. The original four sounds were recorded by a female and a male Mandarin speaker from Beijing. Phonetic analytical software Praat was applied to synthesize all stimuli used in this study [17]. The duration for all stimuli is 450 milliseconds.

There were 26 stimuli (13 female stimuli and 13 male stimuli) presenting as the primary experimental materials. All the synthetic sounds were created based on T2 syllables. The synthetic process was to make the original T2 sounds get similar to T3 and show some acoustic features of T3. The stimuli materials can be divided into four groups.

The first group includes four original sounds of T2 and T3 with low-falling contours. A detailed list with acoustic values for the four sounds is given in Table 1, and tone contours for the four sounds are shown in Figure 2. Stimuli number of each sound consists of a letter “f” (representing the female speaker) or “m” (representing the male speaker), and a serial number.

Table 1 Acoustic parameter of Stimuli 01-02

Stimuli Number	Inflection F0	End F0	Inflection position
f01	219	298	0.13s
f02	104	221	0.23s
m01	102	150	0.13s
m02	51	107	0.25s

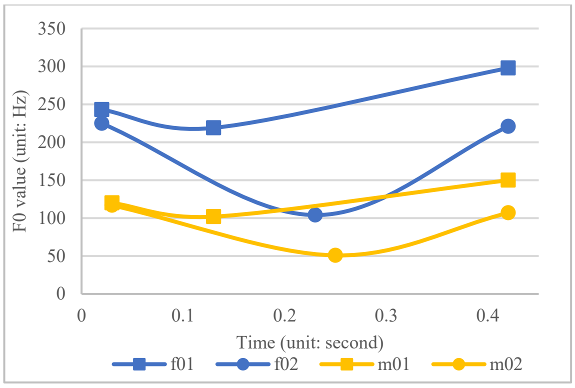

Figure 2 Tone contours of stimuli 01-02

The second group focuses on the F0 value variable, including six synthetic sounds with lowered inflection F0 or/and lowered final F0. The acoustic parameters and tone contours of the six sounds are given in Table 2 and Figure 3.

Table 2 Acoustic parameter of stimuli 03-06

Stimuli Number	Inflection F0	End F0	Inflection position
f03	104	298	0.13s
f04	219	242	0.13s
f05	104	242	0.13s
m03	51	150	0.13s
m04	102	116	0.13s
m05	51	116	0.13s

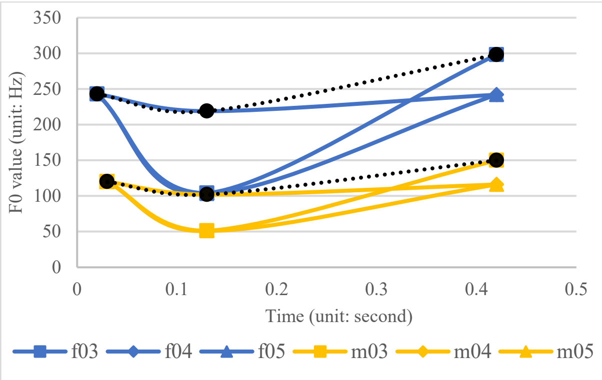

Figure 3 Tone contours of stimuli 03-05

The third group is the inflection-position group, including four synthetic sounds with inflection positions moving backward. The acoustic parameters and tone contours of the four sounds are given in Table 3 and Figure 4.

Table 3 Acoustic parameter of stimuli 06-07

Stimuli Num	Inflection F0	End F0	Inflection position
f06	219	298	0.18s
f07	219	298	0.23s
m06	102	150	0.19s
m07	102	150	0.25s

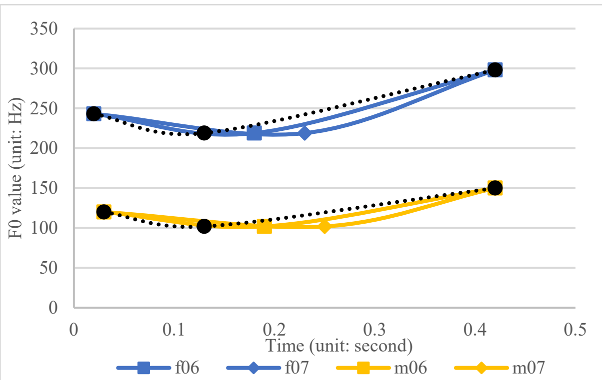

Figure 4 Tone contours of stimuli 06-07

The fourth group is the combination group, including 12 synthetic sounds with various combinations of F0 values and inflection position changes. The acoustic parameters and tone contours of the 12 sounds are given in Table 4 and Figure 5.

Table 4 Acoustic parameter of stimuli 08-13

Stimuli Num	Inflection F0	End F0	Inflection position
f08	104	298	0.18s
f09	219	242	0.18s
f10	104	298	0.23s
f11	219	242	0.23s
f12	104	242	0.18s
f13	104	242	0.23s
m08	51	150	0.19s
m09	102	116	0.19s
m10	51	150	0.25s
m11	102	116	0.25s
m12	51	116	0.19s
m13	51	116	0.25s

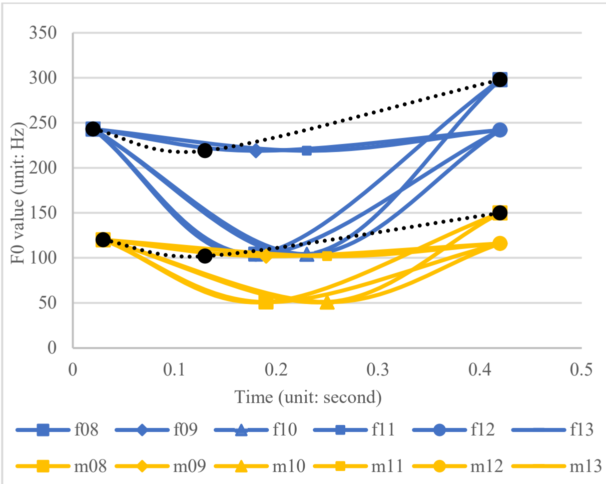

Figure 5 Tone contours of stimuli 08-13

2.2. Participants

40 female and 40 male Chinese native speakers participated in this experiment. All participants were born in China and speak Mandarin as their mother tongue with education backgrounds of junior college. The age of participants ranges from 20 to 40. All participants have normal hearing and have no knowledge about phonetics or phonology. Ten participants were excluded due to their excessive error rates.

2.3. Process

Experiment 1 was an identification task with 26 synthetic sounds organized randomly. All stimuli were repeated twice. There were in total 52 tasks in each questionnaire for each participant to finish. A link to the questionnaire was provided for participants. Participants were required to read instructions and fill in their personal information first. Participants needed to go through a trial to get familiar with the form of tasks. Then in the actual experiment, they were asked to click on the "play" button to listen to each sound file and then make a forced choice from T1, T2, T3, and T4. Each sound can be played twice.

2.4. Data analysis

The identification rate of a stimulus to be perceived as T2 or T3 can be represented as the Equation (1):

$$p = \frac{freq\ of\ Ss\ perceiving\ T2\ or\ T3}{num\ of\ Ss{*}rep\ num\ of\ stimuli} \quad (1)$$

As the answers for each task can be T1, T2, T3, or T4, the error rate needs to be counted, and errors need to be eliminated from the final dataset. Therefore, a further formula for precise identification rate of T2 or T3 is given as Equation (2):

$$p_{Final} = \frac{freq\ of\ Ss\ perceiving\ T2\ or\ T3}{num\ of\ Ss{*}rep\ num\ of\ stimuli - freq\ of\ Ss\ perceiving\ T1\&T4} \quad (2)$$

Equation (3) for error rate is also provided:

$$p_E = \frac{freq\ of\ Ss\ perceiving\ T1\&T4}{num\ of\ Ss{*}rep\ num\ of\ i} \quad (3)$$

'*p*' refers to the identification rate; 'P_{Final}' refers to the identification rate with errors eliminated; 'P_E' refers to the error rate; '*freq*' refers to the frequency; '*Ss*' refers to participants; '*num*' refers to number; '*rep num*' refers to repetition number.

2.5. Results and analyses

The identification rates, the error rates, and the final identification rates with errors eliminated of T2 and T3 for each stimulus are listed in Table 5 and Figure 6.

Table 5 Identification rate of T2 and T3 & error rate

Stimuli	P(T2)	P(T3)	P_E
m&f01	0.71	0.26	0.028
m&f02	0.29	0.67	0.038
m&f03	0.51	0.47	0.022
m&f04	0.50	0.30	0.200
m&f05	0.36	0.62	0.022
m&f06	0.69	0.28	0.034
m&f07	0.64	0.32	0.034
m&f08	0.39	0.56	0.050
m&f09	0.42	0.36	0.222
m&f10	0.38	0.59	0.022
m&f11	0.39	0.30	0.303
m&f12	0.31	0.65	0.038
m&f13	0.29	0.68	0.025

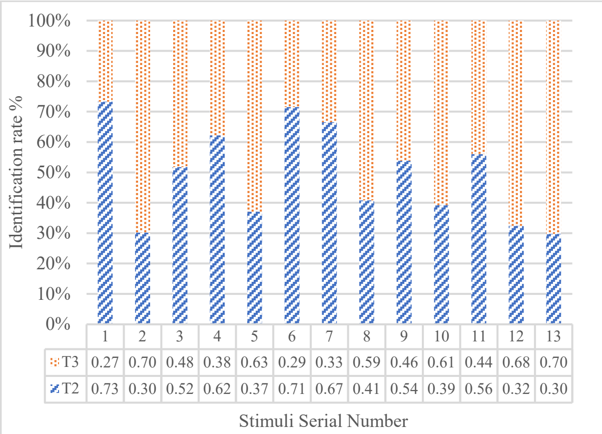

Figure 6 Identification rate of T2 and T3 (error eliminated)

From Figure 6, we can see that the identification rate of T2 and T3 did not reach 75%, even for the original T2 and T3. Low identification rates support Chuang's argument that the perception of T2 and T3 can be challenging compared to other tone perception pairs [5]. The identification rates of T2 and T3 in stimuli 01 and 02 are treated as the referential rates of control group to be compared with other experimental groups.

2.5.1. F0 values (Stimuli 03-05)

When only the inflection F0 value is lowered, compared to stimulus 01, the identification rate of T2 decreases to 52%, which is slightly higher than that of T3; when only the final F0 value is lowered, the identification rate of T2 is 62%, which does not decrease as low as that of stimulus 03. The situation is different when both inflection and final F0 values are lowered. The identification rate of T2 for stimulus 05 is 37%, which indicates that the combination of lowered inflection F0 values and lowered final F0 values enhance the T3 dentification and weaken the T2 identification. Therefore, it seems that the combination of inflection F0 and final F0 greatly influence the perception of T2 and T3.

2.5.2. Inflection position (Stimuli 06-07)

For both stimuli in this group, when only the inflection position moves backward, although the identification rate of T2 shows a tiny decrease compared to stimulus 01, namely, 71% and 67%, the rates of T2 for both stimuli are still over 65%. Therefore, it is likely that the inflection position does not play a dominant role in influencing the perception of T2 and T3. The comparison between stimulus 06 and 07 illustrates a tendency that the identification rate of T2 decreases along with the inflection position moving further back.

2.5.3. Combined variables (Stimuli 08-13)

With lowered inflection F0 values and latter inflection position (Stimuli 08 & 10), the identification rate of T2 decreases to around 40%, precisely 41% and 39%. By adding the inflection position variable, the identification of T3 is improved. Similar to the findings in stimuli 06 & 07, a weak tendency can be detected that with the inflection position moving to the mid-duration, the identification rate of T2 slightly decreases.

For stimuli 09 and 11, with lowered final F0 values and latter inflection position, the identification rates of T2 for the two stimuli are still around 55%. Although lowered F0 values and latter inflection influence the perception of T2 and T3, acoustic features for T3 perception are insufficient, resulting in participants' arbitrary choices between T2 and T3.

The identification rates of T2 for stimuli 12 and 13 are 32% and 30%. The identification rate of T2 for stimulus 13 is the same as the original T3 sound, as the acoustic features of stimuli 02 and 13 are almost identical. The comparison between the identification rates of T3 for stimuli 12 and 13 reveals a similar weak positive correlation between the falling proportion and the identification rate of T3.

2.5.4. Excessive error rates (Stimuli 04 & 09 & 11)

Another point that needs attention is relevant to the error rate. From Table 5, the error rate of stimuli 04, 09, and 11 all exceeds 20%, which means that more than 15 in 80 participants made a choice other than T2 and T3. Among the errors, more than 90% of the participants chose T1, the level tone, when hearing stimuli 04, 09, and 11. The identification rates of T1 increase because of the small pitch variation range in stimuli 04, 09, and 11. These three stimuli depict nearly-flat tone contours, similar to tone contours of level tones, as shown in Figure 7. Hence, participants show a more complicated perception variation for these three stimuli.

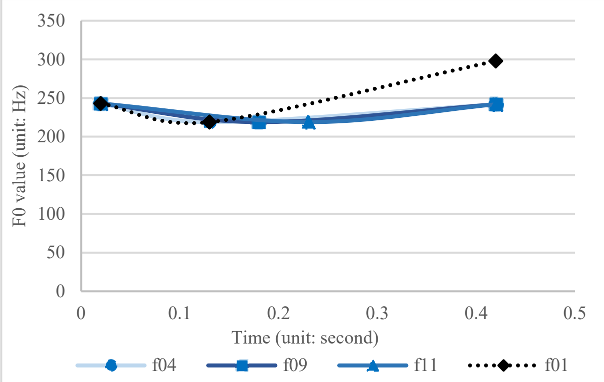

Figure 7 Tone contours of stimuli 04, 09, and 11

In conclusion, although the changes in inflection F0 values of a stimulus cannot decrease the identification rate of T2 sharply to under 40%, it can be an influential factor that alters the acoustic features of a stimulus from T2 to T3. Stimuli with any change combinations that include lowered inflection F0 values (stimuli 05, 08, 10, 12, 13) have an identification rate around or lower than 40%. The status of inflection F0 values is more dominant compared to other factors in aspects of T2 and T3 perception. Final F0 values and inflection position also play secondary roles in affecting the perception of T2 and T3. Furthermore, the delayed degree of inflection position is in proportion to the decrease of the T2 identification rate.

Based on previous studies and the results of experiment 1, the inflection F0 value does show its crucial status in the perception of T2 and T3. In experiment 2, inflection F0 is treated as the only variable. A continuum of synthetic sounds with different inflection F0 values is used as the materials of experiment 2 to see if there is a categorical boundary of inflection F0 values in T2 and T3 perception and whether the perception of T2 and T3 depicts a categorical pattern.

3. EXPERIMENT 2

3.1. Stimuli

A continuum of stimuli is synthesized by making gradual deductions from the inflection F0 value of the T2 syllable. Ten uniformly-spaced F0 values of the inflection point are set from 219 Hz to 104.7Hz (12.7 for one step) in accordance with F0 values of the original T2 and T3 stimuli. The F0 values of the onset and final point are set the same as the original T2 stimulus. The inflection point remains its position in the original T2 stimulus. The duration for all stimuli is adjusted to 450 milliseconds. The tone contours of the ten stimuli are given in Figure 8.

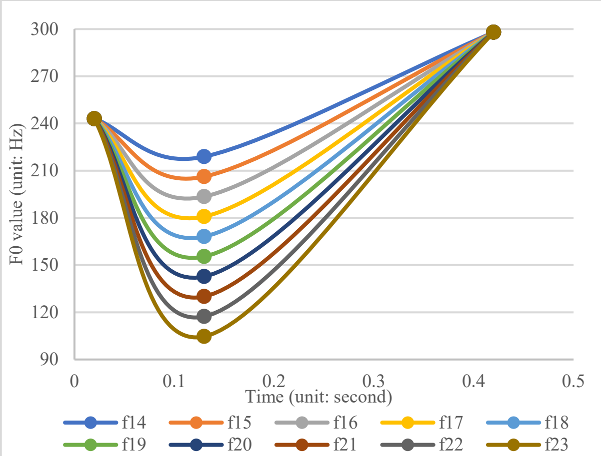

Figure 8 Tone contours of stimuli 14-23

3.2. Participants and Process

Ten synthetic sounds were organized randomly, with each stimulus repeated twice. There are in total 20 tasks. Similar to experiment 1, participants were asked to listen to each stimulus and then make a forced choice from T1, T2, T3, and T4. Participants and data analysis are the same as experiment 1.

3.3 Results and analyses

The identification rates, the error rates, and the final identification rates with errors eliminated for each stimulus are shown in Table 6 and Figure 9.

Table 6 Identification rate of continuum 14-23

Stimuli	P(T2)	P(T3)	P_E
f14	0.76	0.18	0.056
f15	0.77	0.19	0.044
f16	0.69	0.28	0.038
f17	0.70	0.29	0.013
f18	0.69	0.28	0.025
f19	0.60	0.38	0.019
f20	0.50	0.47	0.031
f21	0.54	0.41	0.050
f22	0.52	0.44	0.038
f23	0.54	0.44	0.025

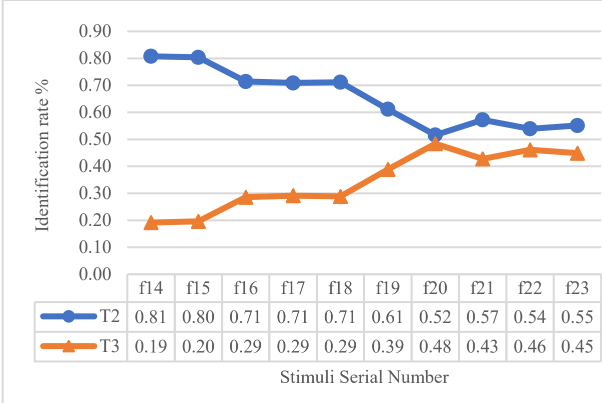

Figure 9 Identification rate of T2 and T3 for continuum f14-f23

In Figure 8, we can see the identification rate curves of T2 and T3 perception. The identification rate of T2 reaches 80% for the first two stimuli, and the identification rate of T3 is as low as 20%. However, throughout the continuum, the lowest identification rate of T2 is still higher than 50%, and the highest identification rate of T3 is only 48%.

The proper decrease of identification rate of T2 begins from stimulus f18, reaching its minimum at f20, and then fluctuates slightly above 50%. Correspondingly, the identification rate curve of T3 reaches its maximum at f20 and then fluctuates below 50%. The two curves do not have a proper intersection. Therefore, when the final F0 values and the inflection position are set as invariants in accordance with the T2 stimulus, the perception of T2 and T3 with a continuum of different inflection F0 values did not show a categorization trend, and no categorical boundary can be detected.

The results in experiments 1 and 2 suggest that with only one variable, it is nearly impossible to determine the perception of T2 and T3. According to the results in experiment 1, the lowered inflection F0 plays a vital important role in T3 identification, but the inflection F0 value is inconclusive in determining T2 and T3 perception. Therefore, the perception of T2 and T3 does not rely on a single factor. At least two factors need to combine to decisively influence the perception of T2 and T3, in which case a proper categorical pattern can probably be detected.

4. DISCUSSION

4.1. Inflection Position as a Secondary Factor

In contradiction to previous findings, the inflection position did not show that determining influence on the perception of T2 and T3. One possibility could result from different position settings. In Qin's experiment, the inflection position can go back until the falling contour occupies 70% of the whole duration [16]. The categorical boundary of inflection position when the final F0 is located in a high-level pitch range appears near 300ms in a 500ms-long stimulus. In this study, the inflection position can probably show its effect when the position is set to further back. A follow-up experiment can be conducted.

4.2. Relationships of inflection position, inflection F0, and final F0

A tone perception can be enhanced or weakened from two dimensions, slopes and duration. Chao suggests that tone is revealed by the functional relationship between pitch value and duration [18]. Previous studies support that duration plays a significant role in tone contour perception. Cao conducted a speech synthesis experiment by gradually shortening a 300ms zero-onset vowel with a falling contour to 60ms and found out that there is a duration threshold for participants to perceive the tone contour. Participants cannot identify the pitch changes if the duration is not long enough [19]. The longer the falling duration is, the better one can perceive the tone contour. The same pattern goes with the slopes, the rate of pitch changes. The steeper the tone contour goes, the more precise the perceptual clue becomes.

The inflection F0 value is a double-edged factor, as it makes both the falling contour and the rising contour easier to perceive. The contradictory information confuses participants when choosing between T2 and T3. Based on lowered inflection F0, the addition of diminished final F0 value or/and latter inflection position provide further perceptual clues for participants to identify T3. In the following paragraphs, I discuss the inner relationship of lowered inflection F0, lowered final F0, and latter inflection position and analyze the interactions within these factors.

The interactions can be divided into two types, promotion and compensation. Both promotion and compensation provide further perceptual support for tone identification. Promotion refers to the function of enhancing the target tone perception, whereas compensation refers to the function of neutralizing the irrelevant tone perception.

4.2.1. Promotion and compensation between lowered inflection F0 values and latter inflection positions

The effect of the latter inflection position on lowered inflection F0 in T3 identification includes both promotion and compensation. Latter inflection position leads to four changes in stimulus, flatter falling slope, steeper rising slope, longer falling duration, and shorter rising duration. The two changes in duration are beneficial for T3 identification. Lowered inflection F0 brings about a steeper rising slope and a steeper falling slope.

Steeper falling slopes, caused by lowered inflection F0, and longer falling duration, caused by latter inflection position, promote the strong perception of falling contours.

Steeper rising slopes resulting from lowered inflection F0 values are compensated by the shorter duration brought by latter inflection positions. In turn, flatter falling slopes caused by latter inflection positions can be compensated by steeper falling slopes caused by lowered inflection F0. Therefore, the compensation effect between lowered inflection F0 values and latter inflection positions is mutual.

4.2.2. Compensation effect of lowered final F0 values on lowered inflection F0 values

A virtuous compensation of lowered final F0 for lowered inflection F0 enhances the perceptual clue for T3 identification. Steeper rising slopes, resulting from lowered inflection F0, can be compensated by decreasing the rising slopes with lowered final F0 values. The flatter rising slope makes it harder for participants to perceive the acoustic feature of T2.

In this paragraph, I compare the two variable combinations, the combination of lowered inflection F0 & latter inflection (stimuli 08 and 10: average T3 identification rate 60%) and the combination of lowered inflection F0 & final F0 (stimulus 05: T3 identification rate 63%). Lowered final F0 values focus on adjustments in terms of slopes, while latter inflection positions focus on allocation in terms of duration. One explanation for the higher identification rate of T3 in stimulus 05 can be that the slope change is more efficient than duration change in T3 identification. However, this hypothesis needs further experimental support.

5. CONCLUSION

This study explores the perception of T2 and T3 with low-falling contours in monosyllabic words. Based on the primary results of two identification experiments, this paper illustrates the roles that inflection F0 values, final F0 values, and inflection positions in the perception of T2 and T3. The perception pattern of T2 and T3 is also analyzed at a continuum of stimuli with varying inflection F0 values. The main conclusions are as follows:

Firstly, as hypothesized, inflection F0 value acts as an influential factor as an independent variable in T2 and T3 perception. Although the inflection F0 value cannot determine the perception of T2 or T3, combined with lowered final F0 values, lowered inflection F0 values contribute to a sharp increase in identification of T3.

Secondly, in contradiction with arguments from previous studies, the inflection position as an independent variable does not alter the perception of T2 and T3 to a large extent. Not latter enough could be one possibility, as compared to previous experimental materials and results. In addition, T2 and T3 cannot be differentiated by solely one factor, neither the F0 values nor the inflection position. Combining all three acoustic parameters as joint factors influences the identification of T2 and T3 the most.

Thirdly, the relationships between inflection & final F0 values and between inflection F0 values & inflection positions are discussed. The lowered inflection F0 value and latter inflection position mutually promote the identification of falling contours and compensate each other to weaken the perception of rising contours. The lowered final F0 value offers compensation for steep rising slopes caused by the lowered inflection F0 value.

Fourthly, at a continuum of synthetic T2 to T3 stimuli with various inflection F0 values, the identification of T2 and T3 presents no clear categorical tendency.

For future research, some follow-up studies can be conducted in different experimental settings. For instance, asking participants to choose between T2 and T3 in each task can probably make participants sensitive to more specific perceptual features. In terms of perception pattern, from this study, it is clear whether the identification rate curves of T2 and T3 will continue rising and falling if the inflection F0 values continue to decrease according to the trends of the two curves. Therefore, future studies can investigate the categorical pattern of T2 and T3 perception with continuums of combined variables. In addition, as mentioned in the final discussion, whether a change in slopes plays a more critical role than duration in T2 and T3 perception is still hypothetical, which requires further primary evidence.

REFERENCES

[1] Shen, J. (1985). The pitch range of tones and the intonation in Beijing Mandarin. In T. Lin, & L. Wang, Experimental studies on speech sound in Beijing Mandarin (pp. 73-125). Beijing: Peking University Press.

[2] Wu, Z., & Lin, M. (1989). An introduction to experimental phonetics. Beijing: Higher Education Press.

[3] Shi, F., & Wang, P. (2006). A statistical analysis of the monosyllabic tones in Beijing Mandarin. Studies of the Chinese language, 33-41.

[4] Dong, R. (2020). An experimental analysis of tone sandhi in the Harbin dialect-focusing on some irregular disyllabic words. Amsterdam: University of Amsterdam.

[5] Chuang, C., Hiki, S., Sone, T., & Nimura, T. (1972). The acoustic features and perceptual cues of the four tones of standard colloquial Chinese. Proceedings of the 7th International Congress of Acoustics, 297-300.

[6] Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychology, 431-461.

[7] Liu, J. (2004). Perceiving the boundary between the lexical rising tone and the falling-rising tone. In F. Shi, & Z. Shen, A Festschrift in honor of professor William S-Y. Wang on his seventieth birthday (pp. 222-234). Tianjin: Nankai University Press.

[8] Shen, X., & Lin, M. (1991). A perceptual study of Mandarin Tone 2 and 3. Language and Speech, 34-39.

[9] Wang, Y., & Li, M. (2010). The effects of tone pattern and register in perceptions of Tone 2 and Tone 3 in Mandarin. Acta Psychologica, 899-908.

[10] Wang, Y., & Qin, X. (2015). The identification and discrimination of tone2 and tone3 in Mandarin. Linguistic Sciences, 337-352.

[11] Fang, Z., & Jin, L. (1992). The multi-dimensional scaling of tone perception. In P. Kuang, & J. Zhang, Studies of the Chinese Language-the 5th International Proceedings for Cognitive Science (pp. 145-155). Beijing: Science Press.

[12] Yang, S., & Xu, Y. (1988). An acoustic-phonetic orientated system for synthesizing Chinese. Speech Communication, 317-325.

[13] Lin, T., & Wang, L. (1992). An introduction of Phonetics. Beijing: Peking University Press.

[14] Cao, W. (2010). A study of level tone perception in Mandarin. The Studies of Chinese Language, 536-543.

[15] Yip, M. (2005). Tone. Beijing: Peking University Press.

[16] Qin, X. (2012). The influence of language experience on categorical perception of Mandarin tones. Beijing: Peking University Press.

[17] Boersma, P. & Weenink, D., 2005. Praat: doing phonetics by computer (Version 4.3.12) [Computer Program]. Available at: http://www.praat.org/ [Retrieved 11 May 2005].

[18] Chao, Y. (2002). A methodology of an experimental study on tones in Mandarin Chinese. In Z. Wu, & X. Zhao, Linguistic Essays by Yuenren Chao (pp. 27-36). Beijing: The Commercial Press.

[19] Cao, J. (1994). A comparative time-frequency analysis between /ao/ and /ou/ in standard Mandarin. Applied Acoustics, 356-360.