The perception of parallel encoded emotional and linguistic prosody in Mandarin Chinese

Qianyutong Zhang[1,2], Shanpeng Li[1,3]

[1] MIIT Key Lab for Language Information Processing and Applications (LIPA), School of Foreign Language, Nanjing University of Science and Technology, China
[2]Institute of Linguistics, Shanghai International Studies University, Shanghai, China
[3]Lab of Anhui Language Resources Preservation and Research, Anhui University, China.

ABSTRACT

In speech prosody, emotional prosody (e.g., angry or neutral emotion) and linguistic prosody (e.g., declarative or interrogative intonation) are the two most common prosody functions. Successfully recognizing these different functions are crucial to daily communication. However, the two functions are usually parallel encoded in the same prosody contour, causing perceptual difficulties. In scarce previous studies focusing on the recognition of the two prosodic functions, most research focused on the non-tonal languages. It remains unknown how the two functions interacted when perceived by tonal language speakers. To answer this question, twenty-two native Mandarin speakers were participated to identify the linguistic and emotional information embedded in the one sentence. The results confirmed the interaction between linguistic and emotional function in Mandarin but with different orientations: the interrogative intonation impedes the identification of neutral emotion, but facilitates the angry emotion. On the other side, angry emotion impedes the identification of declarative intonation only.

Keywords: Parallel Perception, Linguistic Function, Emotional Function, Mandarin Chinese

1. INTRODUCTION

Speech communication is an essential part of our daily life, in which prosody is a reliable element for listeners to comprehend the speaker's intention [1]. Speech prosody is mainly used to express linguistic (e.g., declarative or interrogative intonation) and emotional information (e.g., angry or neutral emotion), manifested by the same acoustic parameters such as F0, duration, and intensity [2-4]. Emotional prosody is used to convey the speaker's emotional state. For example, if the speaker is angry, the acoustic features will have higher mean F0 (in most languages), larger intensity, and faster speech rate than emotionlessness (i.e., neutrality) [5]. At the same time, speech prosody also conveys linguistic information such as sentence type. In interrogative sentences, for example, the pitch contour is rising, and the mean F0 is higher than declarative counterparts [6].

According to the PENTA model [7-8], different communicative functions were encoded in parallel in the same prosody. However, previous studies have demonstrated that the decoding of these coinciding prosodic functions appears an interaction effect so that the listeners may have difficulty in identifying each prosody [3, 9-10]. To be specific, linguistic prosody like intonation will affect the recognition of emotion types. For example, Scherer et al. [11], found that the rising intonation of yes/no questions are perceived as more agreeable and polite. On the other hand, emotional prosody also interferes with the perception of linguistic prosody. Pihan et al. [3] revealed that the identification accuracy of interrogative/declarative contrasts was reduced in emotional stimuli because sentences with emotional prosody have greater pitch variability, so the linguistic signals marked by pitch direction become blurred. However, although these studies demonstrated the fact that the existence of one function of prosody does affect the perception of another prosodic function, they didn't reveal the detailed pattern in which how emotions and intonations affect the perception of each other. What's more, previous studies mostly focused on Indo-European languages such as English and German [3,11], thus we don't know whether people who speak Mandarin Chinese have similar perception results.

Despite the scarcity of Chinese studies in this field, the study of Mandarin Chinese is of great significance since it is a tonal language that has some differences from non-tonal languages. For example, in Mandarin Chinese, F0 works not only at the sentence level to convey intonation information similarly with non-tonal languages, but also at the syllable level to distinguish lexical meanings [12]. At the same time, the perception of prosody mainly depends on F0 [13], so the F0 differences may manifest themselves in the prosody perception process of Chinese and English utterances. Hence if the experiment on tonal Mandarin shares similar results with non-tonal English, it will be a better verification for a cross-language commonality that different prosodic functions have interactions during the perception of speech prosody.

Therefore, the present study aims to explore whether and how different functions of prosody influence Mandarin Chinese speakers' identification when emotional and linguistic prosody are simultaneously perceived. More specific, the current study explore whether and how the existence of speech emotions affects the identification of intonation information to distinguish interrogative and declarative sentences and whether and how intonations affect the perception of emotions by Mandarin Chinese speakers.

2. METHOD

2.1. Participants

Twenty-two native Mandarin Chinese speakers (12 female; 10 male) aged from 19 - 23 yrs. (mean = 21.5, SD = 1.64) participated in this perceptional experiment. All the participants speak Mandarin Chinese as their first language. All of them reported with normal hearing and normal or corrected-to-normal vision. They were provided written informed consent forms and were remunerated for their participation.

2.2. Stimuli

2.2.1. Corpus

The corpus consisted of 130 syntactically similar and semantically neutral sentences (e.g., "小马正在看电视Xiaoma zhengzai kan dianshi." / Mark is watching TV). Each sentence was produced with two different intonations (declarative or interrogative) and two different emotions (angry or neutral). These two factors are fully crossed, resulting in four conditions: (1) declarative intonation superimposed with a neutral emotion; (2) declarative intonation superimposed with an angry emotion; (3) interrogative intonation superimposed with a neutral emotion; (4) interrogative intonation superimposed with angry emotion. In addition, 130 fillers are also recorded, which are different from the target sentences in syntax and emotion.

2.2.2. Recording session

All target sentences were recorded by a female native Mandarin speaker. The speaker was instructed to produce the sentences presented randomly within four conditions and to use normal speech rate and loudness as they were having daily conversations. Before formal recording, the speaker looked through the textual materials to familiarize target sentences. Stimuli were recorded in a soundproof room in Phonetics Lab in Nanjing University of Science and Technology. The sounds were recorded using the cardioid condenser microphone Neumann U87Ai and the audio interface RME Fireface UCX and saved in WAV format with a 44.1 kHz sampling rate at a 16-bit resolution. All the recordings were normalized to amplitude of 70 dB HL (to correct for slight differences in recording levels across recording sessions). Finally, a total of 650 recordings (130 target sentences ╳ 4 speaking styles + 130 filers) were collected.

2.2.3. Validation test

To guarantee the ecological validity of stimuli recorded, a validation test was conducted after recording. Sixty-one native Chinese speakers who did not participate in the recording session were recruited. They were required to rate the degree of emotion using a 5-points Likert scale, from 1 (emotionless) to 5 (extremely angry), and to identify the intonations of the utterances using a two-alternative forced-choice (i.e., "declarative sentence" or "interrogative sentence"). The results of the validation test showed that the mean score of sentences with an angry emotion was 3.73, while that of the neutral emotion was 1.33. The mean accuracy of declarative and interrogative utterances was 99.04% and 88.43% respectively. In consideration of the quantity and quality of the corpus, those sentences scored less than 3 for angry or more than 2 for neutral, and the identification rate of declarative or interrogative utterances less than 75% were excluded for further experiments. Finally, 480 recordings were retained (120 target sentences ╳ 4 conditions).

2.3. Procedures

The perception experiment was conducted in a quiet computer classroom. Each participant was seated in front of a computer monitor and the recording was played through a headset. All the stimuli were randomly presented in the experiment, divided into 2 tasks. In each task, there were 250 trials (30 trials × 4 conditions + 130 fillers). Each trial involved one sentence, and each sentence presented in the task was different.

Before each stimulus, a beep was played to call the subjects' attention. When the audio was playing, a picture of a loudspeaker was displayed at the center of the screen. After each playback, a task was presented. The subjects should respond within a limited time (5 sec). There were two tasks, both of which were three-item forced choice identification tasks: In the emotion task, subjects were asked to ignore the sentence types and judge the emotion types of each utterance within **neutral** / **angry** / **other** emotion ("other" is set because fillers contain emotions other than angry and neutral); (2) In the intonation task, participants were asked to ignore speech emotion and judge the sentence type within **declarative** / **interrogative** / **other** intonation ("other" is set to balance with that in the emotion task). Subjects were required to press corresponding keys on the computer keyboard. And the sequence of tasks was balanced among the subjects. The subjects had about 10-minute rest between the two tasks.

2.4. Data collection and statistical analysis

Participants' responses were collected via PsychoPy [14]. The percentage accuracy was arcsine-transformed first to satisfy the normal distribution of linear models. All transformed mean accuracy values were analyzed using the linear mixed-effect model in R [15] with the "lmer" function in the "*lme4*" package [16]. The task (emotional vs. intonational task), and condition (neutral declaratives, neutral interrogatives, angry declaratives, and angry interrogatives) were set as fixed factors and the subjects was set as random factor. The *p* -values of the main and interaction effects were calculated using Satterthwaite approximation from the "ANOVA" function in the "*lmerTest*" package [17]. The observed significant interaction effect was further executed by the "emmeans" function in the "*emmeans*" package [18] to conduct post-hoc pairwise comparisons.

3. RESULTS

As Table 1 shows, the linear mixed-effect model revealed a significant two-way interaction effect between task and condition (F (3, 119) = 7.450, *p* = 0.0001), and also a significant main effect of task (F (1, 119) = 17.195, *p* ＜ 0.0001), as well as a significant main effect of condition (F (3, 119) = 26.194, *p* ＜0.0001). The post hoc results of the two-way interaction effect were reported in the following sections.

	df1	df2	*F*	*p*
Task	1	119	17.195	＜.0001
Condition	3	119	26.194	＜.0001
Task×Condition	3	119	7.450	.0001

Table 1: The omnibus effect of each main and interaction fixed effect from linear mixed-effect for identification accuracy rate.

3.1. The effect of intonation under emotion task

The Tuckey-HSD post hoc comparison for the two-way effect showed that, for intonation task, the accuracy of interrogative intonation (77.96%) was significantly higher than declarative intonation (49.80%; β = 0.367, SE = 0.085, t = 4.337, *p* < 0.0001) in angry utterances. On contrast, the accuracy of interrogative intonation (72.04%) was significantly lower than declarative intonation (97.71%; *β* = - 0.407, SE = 0.085, *t* = - 4.809, *p* <0.0001) in neutral utterances (see Fig. 1).

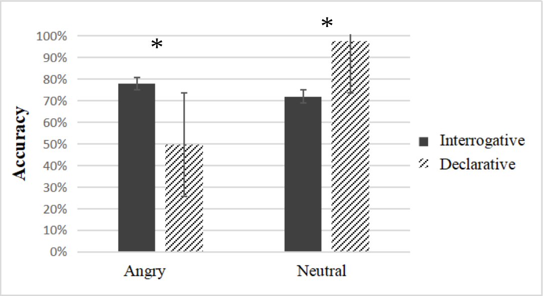

Figure 1: The mean identification accuracy of angry and neutral emotions under different intonations.

3.2. The effect of emotion under intonation task

For the emotion task, there was no significant difference between angry and neutrality in interrogative utterances (86.98% and 79.63% respectively; p = 0.4). However, the identification rate of angry emotion (79.68%) was significantly lower than neutral emotion (100%; *β* = - 0.346, SE = 0.085, *t* = -4.090, *p* < 0.001) in declarative utterances.

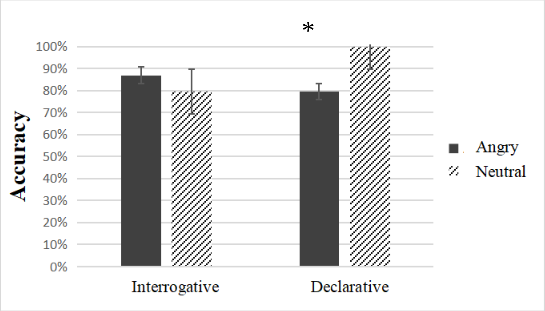

Figure 2: The mean identification accuracy of interrogative and declarative intonations under different emotions.

4. DISCUSSION

The current study explored the perception of different prosodic functions by native Mandarin speakers. The results revealed an interplay effect between linguistic and emotional prosody.

4.1. The influence of linguistic prosody on the perception of emotional prosody

The current study revealed that linguistic prosody (i.e., intonation) affects the perception of emotional prosody in Mandarin: interrogative intonation interferes with the identification of neutral emotion, but seems to improve angry emotion judgment. The role of declarative intonation was just in reverse: the identification of neutral emotion was improved, while that of angry emotion was declined when the carrier sentences were coupled with declarative intonation. This phenomenon might be attributed to the mean F0 and the final F0 movement [19]. Specifically, interrogative intonation has a higher mean F0 and a rising final F0 contour. However, these two prosodic cues are also used to convey emotionally exciting expressions like surprise, happy and anger [5]. Therefore, subjects might be more likely to mistake the emotionally neutral prosody for angry or other exciting emotions when they perceive the sentences with the interrogative intonation.

The accuracy of interrogative intonation was significantly higher than that of declarative intonation when sentences sounded angry, indicating that the interrogative intonation facilitates the perception of angry emotion. Similar to what has discussed above, interrogative intonation and angry emotion share both higher mean F0 and rising F0 contour, thus a superposition effect could be generated to make emotionally angry prosody seemed angrier under interrogative intonations.

4.2. The influence of emotional prosody on the perception of linguistic prosody

The results also revealed an effect of emotional prosody on the perception of declarative intonation but no effect on the identification of interrogative intonation. This is consistent with the results of Pihan et al. [3]. In their research, the discrimination rate of interrogative/declarative contrast was reduced when perceiving stimulus with emotional intonation. They explained that the pitch variability associated with emotion identification interfered with the pitch direction which is related to linguistic prosody perception. In angry sentences with high pitch variability, the stable signal of pitch direction in declarative intonation may have become covered, resulting in participants' lower accuracy of declarative intonation under angry emotion.

As for interrogatives, however, angry or neutral emotions do not affect the perception of interrogative intonation. It seems unconvincing since other results revealed a preliminary conclusion that different prosodic functions who share similar cues will have interactions. Nevertheless, this exceptional case has also been proved by Bryant and Fox Tree [20] that the linguistic function of interrogative prosody is not affected by the emotional function. From a prosody production perspective, McRoberts et al. [21] also proved that the final F0 rise of interrogative intonation does not decrease in emotional sentences. Combined with Pihan et al.'s speculation, it seems that the high pitch variability will not affect the prominently rising pitch direction in interrogative sentences.

Whilst the present study revealed the interaction pattern of emotional and linguistic prosody simultaneously decoded by Mandarin Chinese speakers, there are still some limitations in the current study. First, only declarative/interrogative contrasts were included as instances of linguistic prosody while prosody has many other linguistic functions, like tone, focus, prominence, etc. Similarly, only angry and neutral emotions were used as emotional stimuli in the current study, thus positive emotional valence and other emotional types should be explored in future studies. Second, since we only considered correct responses, it's not clear whether participants are failing to distinguish between emotional and neutral speech, or between anger and some other emotions. This can be further explored in future studies. Finally, the current study didn't control the final lexical tone of Mandarin sentences, though many studies have proved its influence on the identification of the interrogative sentences. Hence more related research is needed to confirm the conclusion of this study in future studies.

5. CONCLUSION

This study examined the interaction of linguistic and emotional prosody during speech perception in Mandarin. The results showed that linguistic intonation can affect the perception of emotional prosody. Specifically, interrogative intonation interferes with the identification of neutral emotion but facilitates the identification of angry emotion. Meanwhile, emotional prosody also affects intonation perception. Angry interferes with the perception of declarative intonation, while different emotions do not affect the perception of interrogative intonation. The present study enriches our knowledge of the synchronic perception of parallel encoded emotional and linguistic prosody by Mandarin Chinese speakers.

6. ACKNOWLEDGEMENTS

This research was supported by the National Social Sciences Foundation of China, No. 19BYY043 and the Fundamental Research Funds for the Central Universities, No. 30921012205.

7. REFERENCES

[1] Hirst, D. 2005. Form and function in the representation of speech prosody. *Speech Communication*, 46(3-4), 334-347.

[2] Eckstein, K., Friederici, A. 2006. It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of cognitive neuroscience*, 18(10), 1696-1711.

[3] Pihan, H., Tabert, M., Assuras, S., Borod, J. 2008. Unattended emotional intonations modulate linguistic prosody processing. *Brain and language*, 105(2), 141-147.

[4] Paulmann, S., Jessen, S., Kotz, S. 2012. It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, 50(7), 1609-1620.

[5] Wang, T., Lee, Y., Ma, Q. 2018. Within and Across-Language Comparison of Vocal Emotions in Mandarin and English. *Applied Sciences*, 8(12), 2629.

[6] Yuan, J. 2006. Mechanisms of Question Intonation in Mandarin. *Lecture Notes in Computer Science*, 19–30.

[7] Prom-On, S., Xu, Y., Thipakorn, B. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *The journal of the Acoustical Society of America*, 125(1), 405-424.

[8] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech communication*, 46(3-4), 220-251.

[9] Pell, M. 2001. Influence of emotion and focus location on prosody in matched statements and questions. *The Journal of the Acoustical Society of America*, 109(4), 1668-1680.

[10] Zora, H., Rudner, M., Montell, A. 2020. Concurrent affective and linguistic prosody with the same emotional valence elicits a late positive ERP response. *European Journal of Neuroscience*, 51(11), 2236-2249.

[11] Scherer, K., Oshinsky, J. 1977. Cue utilization in emotion attribution from auditory stimuli. *Motivation and emotion*, 1(4), 331-346.

[12] Ross, E., Thompson, R., Yenkosky, J. 1997. Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain and language*, 56(1), 27-54.

[13] Seddoh, S. 2002. How discrete or independent are "affective prosody" and "linguistic prosody"?. *Aphasiology*, 16(7), 683-692.

[14] Peirce, J., Gray, J., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Lindeløv, J. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, 51(1), 195-203.

[15] Team, R. 2013. R: A language and environment for statistical computing.

[16] Bates, D., Mächler, M., Bolker, B., Walker, S. 2014. Fitting linear mixed-effects models using lme4.

[17] Kuznetsova, A., Brockhoff, P., Christensen, R. 2017. lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82, 1-26.

[18] Lenth, R., Singmann, H., Love, J., Buerkner, P., Herve, M. 2018. Emmeans: Estimated marginal means, aka least-squares means. *R package version*, 1(1), 3.

[19] Grandjean, D., Bänziger, T., Scherer, K. 2006. Intonation as an interface between language and affect. *Progress in brain research*, 156, 235-247.

[20] Bryant, G., & Fox Tree, J. 2005. Is there an ironic tone of voice?. *Language and speech*, 48(3), 257-277.

[21] McRoberts, G., Studdert-Kennedy, M., Shankweiler, D. 1995. The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation. *Perception & Psychophysics*, 57(2), 159-174.