



Semantics–Prosody Stroop Effect on English Emotion Word Processing in Chinese College Students With Trait Depression

Fei Chen*, Jing Lian, Gaode Zhang and Chengyu Guo

School of Foreign Languages, Hunan University, Changsha, China

This study explored the performance of Chinese college students with different severity of trait depression to process English emotional speech under a complete semantics–prosody Stroop effect paradigm in quiet and noisy conditions. A total of 24 college students with high-trait depression and 24 students with low-trait depression participated in this study. They were required to selectively attend to either the prosodic emotion (happy, sad) or semantic valence (positive and negative) of the English words they heard and then respond quickly. Both prosody task and semantic task were performed in quiet and noisy listening conditions. Results showed that the high-trait group reacted slower than the low-trait group in the prosody task due to their bluntness and insensitivity toward emotional processing. Besides, both groups reacted faster under the consistent situation, showing a clear congruency-induced facilitation effect and the wide existence of the Stroop effect in both tasks. Only the Stroop effect played a bigger role during emotional prosody identification in quiet condition, and the noise eliminated such an effect. For the sake of experimental design, both groups spent less time on the prosody task than the semantic task regardless of consistency in all listening conditions, indicating the friendliness of basic emotion identification and the difficulty for second language learners in face of semantic judgment. These findings suggest the unneglectable effects of college students' mood conditions and noise outside on emotion word processing.

Keywords: semantics–prosody Stroop, English, emotion word processing, trait depression, college students

OPEN ACCESS

Edited by:

Maha Atout,
Philadelphia University, Jordan

Reviewed by:

Süleyman Kasap,
Van Yüzüncü Yil University, Turkey
Zilong Xie,
University of Kansas Medical Center,
United States

*Correspondence:

Fei Chen
chenfeianthony@gmail.com

Specialty section:

This article was submitted to
Public Mental Health,
a section of the journal
Frontiers in Psychiatry

Received: 04 March 2022

Accepted: 06 May 2022

Published: 06 June 2022

Citation:

Chen F, Lian J, Zhang G and
Guo C (2022) Semantics–Prosody
Stroop Effect on English Emotion
Word Processing in Chinese College
Students With Trait Depression.
Front. Psychiatry 13:889476.
doi: 10.3389/fpsy.2022.889476

INTRODUCTION

Among all sources and respects of emotional communication cues, the comprehensive process of multisensory integration is typically employed to reach a locutionary conveyance. This ability to perceive and combine both linguistic messages (i.e., verbal content meaning) and paralinguistic messages (i.e., non-verbal cues by pragmatic context, body language, and tone of voice) facilitates sophisticated social interaction (1, 2). Yet, the co-occurring semantic meaning and emotional cues in utterances simultaneously are not always presented in a consistent state, and the very discrepant messages combined may lead to delays or even challenges to a correct interpretation of true emotional expression (3–6).

As two prime channels for emotional speech interaction, verbal content and prosodic information mainly bridge the daily communication linguistically and emotionally. The general semantic meaning of speech enjoys the main content of emotional expression, but often the

paralinguistic messages serve as completion and exterior presentation in physical forms (7). Therefore, verbal content acts as the most common means, and prosodic information is one of the most fundamental aspects of social interaction (8). Prosodic cues even present a clearer emotional tendency, particularly when verbal form representation is obstructed due to implicitness or other language barriers (9). By means of changes in pitch, loudness, speech rate, and pauses (10), emotional prosody reveals various non-verbal respects of language that make speakers convey emotional cues in conversation (11). But in real communication practice, emotional prosody can be isolated from semantic information, and in return interacts with verbal content, as a consequence of irregular verbal expression [e.g., sarcasm; (12, 13)].

Empirical research under the Stroop effect paradigm (14) examined the emotional interactions with these informative dimensions through cross-channel and cross-modal experiments (9, 15). With participants facing congruent and incongruent stimuli under different modalities, the inter-competence between linguistic information and paralinguistic information in emotional speech processing would be presented, suggesting relative dominance of either of them (16, 17). Participants performed better with congruent stimuli, while the prolonged reaction time and poorer accuracy rate were caused by specific but conflicting stimuli, which was in line with the congruence-induced facilitation effect and on the other hand, the incongruence-induced interference effect (18–20).

However, interpersonal and social interactions pose challenges for major depressive people, since major depression is closely connected with cognitive impairments in memory and executive functions (21, 22). Depression, a mood disorder marked especially by sadness, inactivity, difficulty in thinking and concentration, has increasingly become a major threat to human life and has arisen significant interest both in the pathological characteristics and the social performance of the patients (23–25). And people with the depression-related illness often display a quite fixed pattern of negative thinking about experience, values, and the whole world generally, and the correct social interaction and interpretation can be compromised (26).

As a subclinical state of depression, trait depression is the exact and frequently occurred tendency of an individual to experience depressive emotions (27). Being regarded as being below the diagnostic criteria for depression clinically, trait depression shares some similarities with depression on cognitive and physiological deficits (28), including pessimism, inferiority, loneliness, and unworthiness (27). As mentioned above, people with major depressive disorder (MDD) presented no self-positivity bias, and they even presented self-negativity bias, which connected more closely with negative information, leading to more automated processing of negative information in the environment (29–31). The lack of self-positive bias makes trait-depressive individuals make people succumb to depression disorders more easily, meaning the group of people who have not yet developed depression disorder but mostly are susceptible. The trait mirrors the long-term emotional stability of their state of mind (32). Although the introduction of various experimental designs and assessment scales availed research for MDD patients,

the emotional speech processing for people with trait depression lacks solid evidence. Studies on emotional speech processing in trait-depressive people are quite scarce (33), partly due to the lack of attention on this specific group of people with mood disorders tendency, and partly due to the lack of a scales for the professional assessment of depressive state and trait (26, 34).

In view of previous studies employing the Stroop-like paradigm to investigate emotional processing, only a few of them focused on college students with trait depression. What is still worth mentioning is mainly the variants of the experimental design. First, studies on emotional speech processing exploring the interaction between word information and emotional prosody are rich. The congruency of affective prosody and word valence facilitated the emotional processing, which was corroborated by later studies (35, 36). The relatively salient role of paralinguistic prosodic information over semantics in emotion word processing was presented with both cross-channel and cross-modal behavioral evidence (9). Second, many previous studies on emotional processing performed on participants with MDD showed quite consistent results. The key role of correct interpretation of emotional signals across verbal and non-verbal channels can be worse (37, 38). The cognitive patterns displayed by MDD patients presented the impaired perception of positive cues and the enhanced attention to negative cues as well in emotional communication (39, 40). Such biased emotional perception has been attested by plenty of studies *via* face recognition (41–43) and a few studies *via* voice recognition (44). These are in accordance with findings at the neurophysiologic level presenting a reduction of activation in frontal and limbic areas in MDD patients (45, 46). Third, Gao et al. (47) presented the mechanisms of the “bilingual advantage effect” under the condition of different languages in the Stroop task. It turned out that skilled bilinguals performed better and presented stronger inhibitory control ability under the condition of the first language than monolinguals. And these bilinguals possessed better information monitoring ability and conflict resolution, which shed some light on the variants of the Stroop paradigm in terms of language capability (48).

To date, very few studies on emotional processing employed vocal speech as auditory materials to explore the performance of college students with trait depression. In the research field of psychology and sociology, the study concerning emotional conflict of college students with trait depression under the Stroop paradigm variants in the visual modality merely examined the different responses of participants under emotionally consistent and inconsistent conditions between words and facial expressions (33), showing emotional consistency effects (i.e., the fact that participants had higher accuracy and shorter response time under the word–face consistent condition). Gao et al. (33) found that the accuracy of the high trait depression group was significantly higher than that of the low trait depression group in all conditions. But the response time of the high trait depression group was significantly lower under the condition of emotional inconsistency, partly because participants tended to use a kind of processing strategy to complete the cognitive task more conveniently, according to the Emotion Infusion Theory proposed by Bower (49). Therefore, high trait-depressive

participants may have a state of readiness and be able to judge the valence of emotion and face more quickly.

Within the existing literature, the studies concerning emotional prosody were examined either in MDD patients or under the simplified semantics–prosody paradigm (in lack of word valence in some experiments). Of all, the marked inclination of emotional conflicts and emotional prosody in participants with MDD seems quite certain and a truism in a general way. And studies ranging from facial signals to human voice and even cross-modal are increasingly mature and complete. Yet, research on emotional prosody in trait-depressive college students under a complete semantics–prosody Stroop effect paradigm is still quite poor. Furthermore, relevant studies were all conducted under the ideal experimental condition, rather than under background noise with ecological value. Moreover, individual differences were rarely considered as a significant factor affecting participants' performance in certain experimental circumstances. Different levels of second language proficiency and personal state of mind are not negligible. So, this study will discuss the interaction of semantic content and emotional prosody during emotion word processing by human voice under a complete Stroop effect paradigm, with different severity of Chinese trait-depressive college students as participants, trying to explore the differences between and within groups, and then to shed light on the undiscovered land.

The current study applied the experimental protocols of Schirmer and Kotz (36) and investigated the English emotion word processing in Chinese college students with trait depression through cross-channel experiments. In terms of participants, the second language proficiency and their severity of trait depression varied more or less because of the well-known individual differences, embodying some of the individuals' proficiency in speech perception (50). For these second language learners of English, the aural English words, to some extent, turn into a language barrier as the second language proficiency, but it is less likely to appear the ceiling effect since all English words we selected in this experiment as language materials are "everyday words" with fair verbal valence. These words were produced with happy and sad emotions, which were the two most distinguished and uncontroversial emotions shared across cultures (51, 52). Besides, participants' long-term state of mind with depressive emotions exerts influence on the perception of emotional prosody (53). Experiment 1 employed both semantic valence judgment with and without prosody–congruency stimuli (i.e., the semantic task), and emotional prosody judgment with and without semantic–congruency stimuli (i.e., the prosody task) to explore the altered perception of speech emotions. Based on the poor performance of MDD on semantic and emotional prosody work, people with trait depression might also present prolonged response time and insensitive emotion recognition on the emotion word processing through verbal and non-verbal channels, thus less Stroop effect in semantic valence judgment. Following the same protocols, Experiment 2 stimulated a more authentic locutionary situation by means of the speech-shaped noise (i.e., an energetic environmental degradation), which added difficulties and interference in emotion word perception both linguistically and prosodically, to break the limit of the singular

laboratory environment and reach conclusions with broader sense (54, 55). In this case, we hypothesized that the noisy condition might aggravate the emotional perception difficulty for the high trait-depressive group.

In a nutshell, with the second language–based and psychology-related behavioral study, we aimed to explore further the mechanisms of emotional perception in college students with trait depression specifically. By contrasts between different severity of trait depression and different levels of listening conditions, practical patterns of the Stroop-like paradigm and theoretical frameworks of emotional processing would be enriched in greater detail, which could facilitate the effective probe of nature about multiple channels of the cognitive process for clinical populations.

MATERIALS AND METHODS

Participants

In total, 48 Chinese college student volunteers (24 men and 24 women) were recruited for this experiment, all born in China and native Mandarin speakers. Age varied from 18 to 26 years. They were graduate or undergraduate students who had English as their second language, and they have passed CET-4 (College English Test Band 4). All participants' trait depression scores were evaluated based on the Chinese version of the State-Trait Depression Scale (ST-DEP), which was proposed by Spielberger (56) and then translated into Chinese by Lei et al. (26). With evidence of being highly valid and more focused on the assessment of cognitive and affective factors, it serves as an effective measure to distinguish between depressive state and trait (57). With a full score of 16–64, students with higher scores would be regarded as participants with high-severity trait depression and likewise, college students with lower scores would be regarded as participants with low-severity trait depression in the current study (26). Specifically, the high-trait group ($n = 24$) comprised 11 men and 13 women who scored above 40 but no more than 64 in the T-DEP, while the low-trait group ($n = 24$) contained 13 men and 11 women who scored above 16 but no more than 30.

Furthermore, all participants were tested for their English skills with the LexTALE test, an efficient vocabulary test to measure L2 language proficiency (58), and phonological short-term working memory (WM) with a digit-span test, the information held temporarily for use in immediate activities such as reasoning and decision making, which serves as a significant indicator of language learning ability (59). In addition, they fulfilled the self-rating of Emotional Intelligence Test [SREIT; (60)], a 33-item scale concerning mental representation and utilization of emotions. The demographic characteristics of participants are presented in **Table 1**. As displayed, there did exist significant group differences between the high trait depression group and low trait depression group in terms of T-DEP and SREIT, but not in the age, LexTALE, and WM.

All participants were right-handed with normal or corrected vision, and only those who reported no history of speech, hearing impairment, no musical training, or had no experience of major

TABLE 1 | Basic information of participants.

	High-trait group (n = 24)		Low-trait group (n = 24)		p
	M	SD	M	SD	
Chronological age	21.96	2.07	22.25	1.87	0.611
T-DEP	44.92	4.98	25.54	4.01	<0.001***
SREIT	113.26	12.37	129.63	13.76	<0.001***
LexTALE	54.74	5.82	55.48	7.46	0.703
WM	27.42	1.72	27.01	3.39	0.595

Means (and standard deviations) of chronological age, T-DEP, SREIT, LexTALE (L2 vocabulary size), and WM for the high-trait group and low-trait group. T-DEP, Trait Depression Scale; SREIT, Self-Rating of Emotional Intelligence Test; WM, working memory. *** $p < 0.001$.

depressive therapy were recruited in the current study (61). This study was approved by the (institution redacted for peer review) to ensure proper compliance with the informed consent procedure. Participants completed the informed consent at the study and got reimbursed for their participation.

Stimuli

The stimuli we employed in the study contained 120 English words (60 verbs and 60 adjectives) carefully selected from “The Oxford 3000,” a list of the 3,000 most important words to learn in English from the Oxford English Corpus, and “The Longman Communication 3000,” a list of the 3,000 most frequent words in spoken and written English that account for 86% of the language, to avoid rare words and therefore guarantee the understandability for these second language learners of English. The whole stimulus set (American spelling) included 60 positive words and 60 negative words based on a pilot study of valence ratings obtained from four advanced English speakers and one native speaker who did not participate in either of the experiments. They were instructed to judge the semantic valence of the words in a randomized order on a 5-point scale from -2 (highly negative) to 2 (highly positive). Negative words had a mean valence of -1.43 ($SD = 0.24$), and positive words of 1.44 ($SD = 0.35$), showing no significant difference in valence strength (positive words were rated just as strong as negative words). Additionally, word frequency was counted by means of the Corpus of Contemporary American English [COCA; (62)]. As shown in **Table 2**, the positive words presented a similar word frequency as the negative words; positive and negative words showed comparable syllable numbers.

A Canadian male speaker (35 years old) produced all English words clearly in a quiet setting with happy and sad prosody (240 stimuli = 120 words \times 2 prosodic categories), which were subsequently normalized to the same duration (1,000 ms). The pitch, however, was different between happy and sad prosodies ($p < 0.001$). Specifically, words read in happy prosody had an average pitch of 154.07 Hz ($SD = 43.04$) and words read in sad prosody of only 98.72 Hz ($SD = 6.67$), which is in line with the acoustic attributes of happy and sad utterances (63). Thus, pitch variations in accordance with the speaker’s emotion

TABLE 2 | Word frequency and syllable numbers of selected English words.

	Type	M (SD)		p
		Positive	Negative	
Word frequency	Positive	44813.18 (21884.29)	39013.72 (20774.27)	0.14
	List 1	41913.77 (21433.09)	List 2 41913.13 (21638.33)	0.99
	Adjectives	40596.93 (21929.90)	Verbs 43229.97 (21051.11)	0.50
Syllable numbers	Positive	2.03 (0.78)	Negative 1.97 (0.74)	0.63
	List 1	1.98 (0.77)	List 2 2.02 (0.75)	0.81
	Adjectives	2.07 (0.84)	Verbs 1.94 (0.66)	0.34

serve as assistant effects for listeners to complete the prosody-identification task (64).

Moreover, though the same words were employed as stimuli in two experiments, we added noise (SNR = 10 dB) to the audio files for Experiment 2 to create a noisy condition. The whole stimuli were divided into two lists (List 1 and List 2), with each list containing 30 positive and 30 negative words spoken by happiness and sadness prosody, conveying both semantic meaning and prosodic emotion to participants simultaneously. Thus, under different instructions of tasks, participants accordingly pay selective attention to either semantic valence information or emotional prosody information of the auditory stimuli. Notably, each list was presented under arrangement on different tasks. Therefore, half of the participants already having heard one list of words under semantic instruction would hear the other list of words under prosodic instruction and *vice versa*.

Task and Procedures

The whole experiment was conducted in a quiet room and each participant was seated in a comfortable chair facing a computer monitor, a noise-canceling headphone, and a Chronos box [an E-Prime-based device with high accuracy of response time; (65)]. Two tasks were performed for participants to selectively attend to word valence information (semantic task: positive or negative) or emotional prosody information (prosody task: happy or sad) of auditory stimuli under corresponding instructions in quiet (Experiment 1) or noisy (Experiment 2) environment. In both experiments, instructions and auditory stimuli were presented by E-Prime [Version 3.0; (66)] on the computer, with the stimulus presentation program customized in advance. Having received the standard auditory information of English words through the noise-canceling headphone (Sennheiser HD280 Pro) binaurally at a comfortable sound intensity level (65 dB SPL), participants offered their responses by pressing the corresponding button of Chronos as quickly and as accurately as possible to indicate their judgments. The accuracy and response time recorded by Chronos would then serve as the measurement.

Participants were told to complete the practice session first as a familiarization process with four spoken words in the semantic task and prosody task, respectively, and these eight words were not used in the real experiment. After participants responded, the instant accuracy would be presented on the screen. Those who reached at least 80% accuracy would enter the formal task phase.

TABLE 3 | Results of linear mixed effects model on reaction time (full presentation with results of both Experiment 1 and Experiment 2).

Effect	Chi-square	p
Severity	4.26	0.039*
Congruency	123.90	<0.001***
Task	69.93	<0.001***
Condition	7.96	0.005**
scale_Trial	28.57	<0.001***
scale_Digit span	0.14	0.706
LexTALE	5.21	0.022*
Severity:Congruency	0.01	0.910
Severity:Task	6.45	0.011*
Congruency:Task	9.38	0.002**
Severity:Condition	17.01	<0.001***
Congruency:Condition	0.06	0.801
Task:Condition	13.56	<0.001***
Severity:Congruency:Task	0.06	0.808
Severity:Congruency:Condition	1.14	0.285
Severity:Task:Condition	0.07	0.785
Congruency:Task:Condition	2.26	0.133
Severity:Congruency:Task:Condition	4.45	0.035*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In formal experiments, for word valence and emotional prosody in each experimental stimulus, participants were instructed to pay main attention to only one respect, though the twofold pairings with two different channels presented either congruency (positive-happy, negative-sad) or incongruency (positive-sad, negative-happy). Specifically, under the instruction of semantic information, participants would identify word valence as positive or negative while ignoring the prosody in this “semantic task.” On the other way around, in the “prosody task,” participants would judge emotional prosody as happy or sad under the instruction of prosody information while ignoring word valence. Instructions were visually presented on the computer screen to make sure participants’ full understanding of each task. Stimuli were presented in a randomized fashion within different tasks to each participant. Unlike the familiarization session, no instant feedback of accuracy would be presented on the screen to avoid the unnecessary distraction of participants, and the session would proceed to the next trial if no response was recorded within 5,000 ms.

Experiment 2 followed the same procedure of the protocols and employed the same word in the quiet environment of Experiment 1, only the speech-shaped noise at a fixed signal-to-noise ratio (SNR = 10 dB) was affiliated to create a noisy environment in Experiment 2, with the effect of energetic masking (67, 68). The SNR of 10 dB was determined based on a pilot study, which reached the lowest SNR level with a minimum accuracy of 80% in both tasks. To avoid the fatigue effect, there was a short break between two experiments, which were presented to participants in random order. The whole experiment took approximately 1 h.

Data Analyses

For statistical analyses, a range of calculations were performed in R [Version 4.1.2; (69)]. For the collected data, 48 subjects

participated in two experiments, with each experiment containing 2 tasks and each task containing 120 items, 23,040 data were obtained in total ($48 \times 2 \times 2 \times 120 = 23,040$ observations). As for the preliminary data filtering process, only data with reaction time between 100 ms to 2,500 ms were counted as acceptable in the experiment to enhance data validity, since neither the excessive speed nor the noticeable delay in response time was admitted in the study. Then, we eliminated incorrect responses, and 18,809 observations were kept eventually. Besides, we also performed a log transformation to reaction time data since in many perceptual experiments response time exhibits positive skewness (70). Furthermore, to compare the inter-group difference of T-DEP scores, SREIT scores, and WM between high-trait group and low-trait group, we employed two-sample *t*-tests with the R package of *ez* (71).

In general, a linear mixed-effect model (LMM) was constructed using the R package of *lme4* (72). Considering the huge difference between types of data, the trial number of words and digit-span scores of participants were centered and therefore reached normalization. When fitting all the LMMs in the analyses of the two identification data, “Reaction time” was counted as the dependent variable. Fixed factors: “Severity (high vs. low),” “Congruency (congruent vs. incongruent),” “Task (semantic vs. prosodic),” “Condition (quiet vs. noisy),” and all their interactions; two random factors: “Participant” and “Item,” were included in the model. And controlled co-variants were LexTALE scores, working memory, and normalized trial number. By-participant random intercepts and slopes for all possible fixed factors were included in the initial model (73), which was compared with a simplified model that excluded a specific fixed factor using the analysis of variance (ANOVA) function in *lmerTest* package (74). The model was fitted to optimize it. Besides, Tukey’s *post hoc* tests were employed using *lsmeans* packages (75) to elaborate the significant interaction effects when necessary.

RESULTS

Statistics suggested that the mean age of participants is 22.10 ($SD = 1.96$, range = 18–26) years, and they have received an average of 15.83 ($SD = 1.84$) years of formal education. **Table 3** presents the fullest results of the LMM on these participants’ reaction time across two experiments, showing a significant four-way interaction of “Severity” \times “Congruency” \times “Task” \times “Condition” [$\chi^2(1) = 4.45$, $p < 0.05$], which was further separately analyzed under two different conditions (quiet and noisy conditions), namely, Experiment 1 and Experiment 2.

Experiment 1 (Quiet Condition)

Figure 1A shows high-severity and low-severity trait-depressive participants’ reaction time in semantic-emotion interference tasks in quiet condition. As displayed in **Table 4**, in the quiet condition, a significant two-way interaction of “Severity” \times “Task” was found [$\chi^2(1) = 4.14$, $p < 0.05$]. The following *post hoc* tests showed that when participants conducted

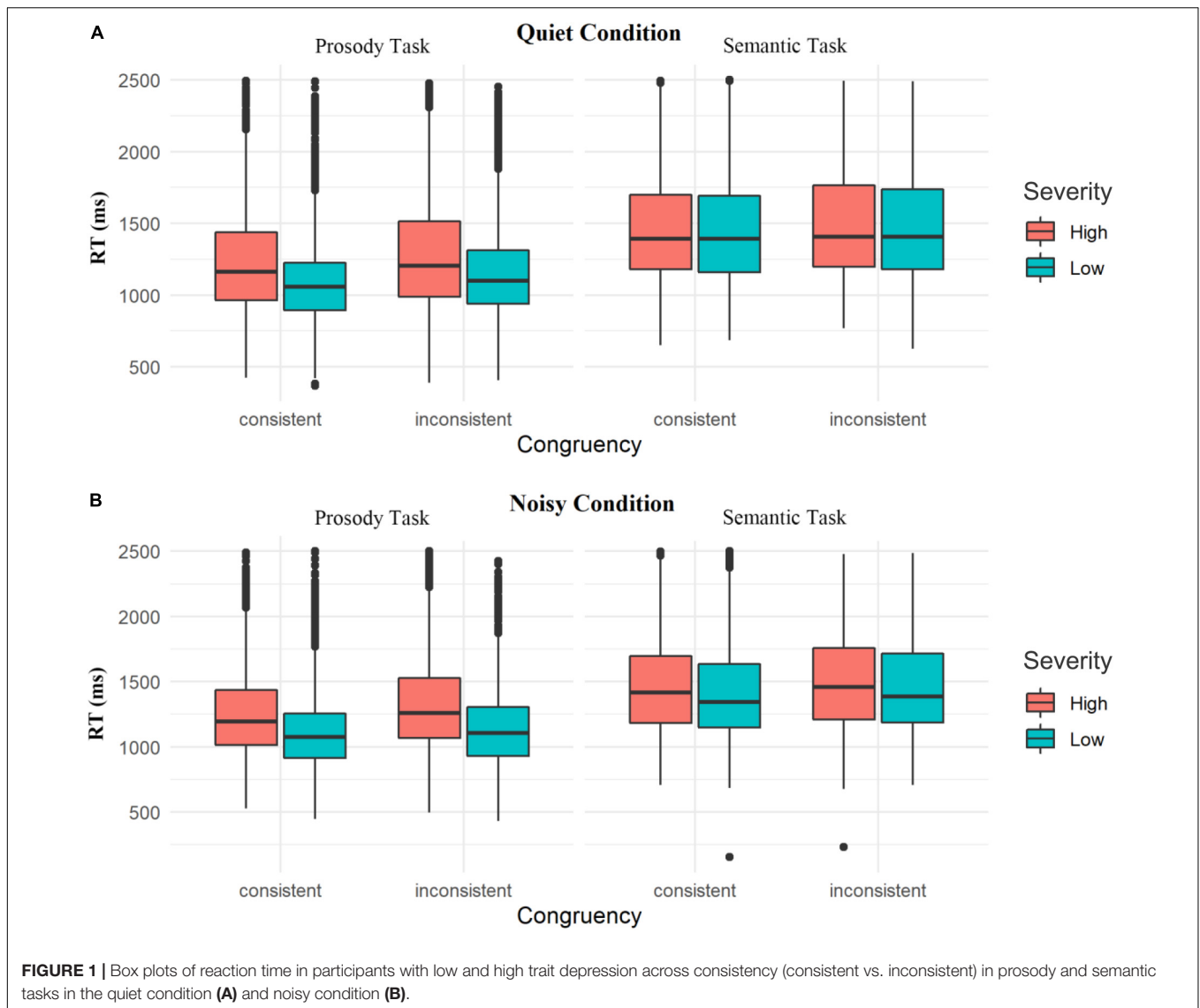


FIGURE 1 | Box plots of reaction time in participants with low and high trait depression across consistency (consistent vs. inconsistent) in prosody and semantic tasks in the quiet condition (A) and noisy condition (B).

the prosody task, high-trait group reacted slower than low-trait group ($\beta = 0.094$, $SE = 0.047$, $z = 1.978$, $p < 0.05$), but there exists no such significant difference when they conducted semantic task ($\beta = 0.008$, $SE = 0.027$, $z = 0.278$, $p = 0.781$); in general, participants reacted faster in the prosody task regardless of their trait depression scores ($ps < 0.001$).

As **Table 4** displays, a significant two-way interaction of “Congruency” \times “Task” was found as well in the quiet condition [$\chi^2(1) = 9.82$, $p < 0.01$]. More specifically, when performing both prosody ($\beta = -0.054$, $SE = 0.007$, $z = -7.700$, $p < 0.001$) and semantic ($\beta = -0.022$, $SE = 0.007$, $z = -3.035$, $p < 0.01$) tasks, both high-trait group and low-trait group spent less time under consistent situation compared with inconsistent situation. The results displayed that participants tended to be more affected by semantic congruency (or not) in the prosody task than be affected by prosody congruency (or not) in the semantic task.

There was also a clear fact that regardless of consistency or not (the Stroop effect set on the experiment), they spent less

time when they conducted prosody task rather than semantic task ($ps < 0.001$).

Experiment 2 (Noisy Condition)

Figure 1B displays high-trait and low-trait groups’ reaction time in semantic and prosody tasks under a noisy condition. As displayed in **Table 5**, in the noisy condition, LMM revealed a clear main effect of “Congruency” [$\chi^2(1) = 34.61$, $p < 0.001$] in the noisy condition, while the two-way and three-way interactions between “Congruency” and any other factors failed to reach significance (all $ps > 0.05$). Notably, while “Congruency” and “Task” produced a two-way interaction and the Stroop effect therein made differences in the two tasks in a quiet condition, no such interaction was found in the noisy condition.

Besides, a significant two-way interaction of “Severity” \times “Task” was detected in the noisy condition [$\chi^2(1) = 83.11$, $p < 0.001$]. When participants conducted the prosody task ($\beta = 0.126$, $SE = 0.035$, $z = 3.611$, $p < 0.001$), the

TABLE 4 | Linear mixed-effects model with severity, congruency, task as the fixed effects and the logarithm of reaction time as dependent variables in Experiment 1.

Effect	Chi-square	<i>p</i>
Severity	2.34	0.126
Congruency	56.36	<0.001***
Task	60.25	<0.001***
scale_Digit span	0.35	0.553
scale_Trial	14.06	<0.001***
LexTALE	5.88	0.015*
Severity:Congruency	0.63	0.427
Severity:Task	4.14	0.042*
Congruency:Task	9.82	0.002**
Severity:Congruency:Task	2.43	0.119

p* < 0.05, *p* < 0.01, ****p* < 0.001.

group difference was quite clear: high-trait group took a longer time to identify emotions expressed by noisy-interferential spoken English words than low-trait group, which closely resembled the results in the quiet condition. However, there was no such significant difference between two groups when they conducted the semantic task ($\beta = 0.036$, $SE = 0.035$, $z = 1.035$, $p = 0.301$).

Besides, in the noisy condition, both high-trait and low-trait groups took a shorter time to complete the prosody task than the semantic task ($ps < 0.001$). Primarily, these data presented that Chinese college students tended to identify emotions faster than L2 verbal content under all listening conditions.

DISCUSSION

So far, the question remains unresolved as to how people with a propensity to depression process emotional cues of the second language during daily communication. To fill the research gap of emotion word processing for second language learners with different severity of trait depression, the current study investigated the interaction of semantic content and emotional prosody under a complete Stroop effect paradigm by Chinese college students with trait depression in quiet and noisy environments. It was designed to address the following three research questions. First, we tried to investigate the differences between high trait-depressive group and low trait-depressive group in emotion word processing. Second, we were interested in the general mechanisms of the Stroop effect on emotion word processing in two severity of trait-depressive groups. And finally, we aimed to figure out whether any change in English emotion word processing would be posed by the influence of noise. The following discussions tried to answer these research questions based on relevant findings.

Specific for High Trait-Depressive Group: Bluntness Toward Emotions

For the issue of Chinese college students in terms of the emotional prosody identification, results of the current study showed that in two experiments, the response time of the high-trait depression group was significantly longer than that of the low-trait depression group regardless of the congruency condition.

TABLE 5 | Linear mixed-effects model with severity, congruency, task as the fixed effects and the logarithm of reaction time as dependent variables in Experiment 2.

Effect	Chi-square	<i>p</i>
Severity	4.65	0.031*
Congruency	34.61	<0.001***
Task	1539.82	<0.001***
scale_Digit span	0.00	>0.999
scale_Trial	19.97	<0.001***
LexTALE	2.07	0.150
Severity:Congruency	0.00	>0.999
Severity:Task	83.11	<0.001***
Congruency:Task	0.72	0.395
Severity:Congruency:Task	1.50	0.220

p* < 0.05, *p* < 0.01, ****p* < 0.001.

This finding of the semantics-prosody Stroop experiment, however, is not congruent with previous findings of the word-face Stroop experiment with trait-depressive college students (33). They found that the response time of the high-trait depression group was significantly shorter than that of the low-trait depression group under the condition of emotional inconsistency. In order to explain this, the authors adopted the Affect Infusion Model (49), meaning participants took strategies in advance and processed information more conveniently, potentially accounting for this phenomenon. So, it is the earlier readiness for cognitive processing more conveniently that assisted the high-trait depression group to react faster.

The poorer performance of emotional processing in high-trait depression people is generally in line with previous studies of emotion-related judgment in patients with MDD. Previous studies presented their impaired recognition of emotions in the visual modality (i.e., facial expressions) or auditory modality (i.e., emotions are expressed vocally). They seemed to show deficits in the correct perception of affective prosody (76). And in most rating experiments, MDD tended to skew the recognition of emotional stimuli into two directions: the tendency toward negative emotional stimuli, and the bluntness of positive stimuli (38). Since trait-depressive undergraduates are associated with low heart rate variability and more specifically its parasympathetic component, which is considered a physiological index of emotion regulation capacity (77), they are less competent to regulate their emotions and reach controlled sensory processing. Participants got poor concentration toward outside information with increasing severity of depression.

Both in quiet and noisy conditions, the results of the current study showed clear contrasts of reaction time between different trait-depression groups in the prosody task, while no such significant contrast was found in the semantic task. This is plausible due to the closer connection between the prosody task and the effect on the long-term psychology of participants.

General for Participants: Extensive Existence of Stroop Effect

Variants of Stroop effect protocols, as behavioral experiments, were considered as an exploration of the primitive operations of cognition, offering clues to the fundamental process of attention

and an ideal tool for the research of automatic processing (78). Results of the current study showed the consistency facilitation effect, meaning that participants took a shorter response time under emotional consistency conditions, which is congruent with previous findings (9, 18, 19). Interestingly, the high-trait depression group lacked activated sensitivity toward emotion perception, then they might have been less affected by the change of emotional prosody when they conducted semantic tasks. However, the lack of two-way interaction of “Congruency” and “Severity” in Experiment 1, indicated that the effect of congruency of stimuli from two channels did not vary between the high-trait group and the low-trait group. Besides, the main effect of “Congruency” in Experiment 2 symbolized the “independence” of the Stroop effect from the mental state of individuals inside (i.e., participants’ trait depression severity) and environmental influences outside (i.e., quiet and noisy listening conditions). The results were roughly analogous, jointly indicating an extensive existence of the Stroop effect.

Notably, the anterior cingulate cortex and dorsolateral prefrontal cortex were reported to remain active when resolving conflict (79), indicating the brain activity in Stroop interference. And the widespread view about the Stroop effect in cognition told that mental skills (e.g., reading) are automatic once they were acquired through repetitive and extensive practice (80). Cattell (81) suggested an automatic process in the cognitive science of the Stroop effect. The automatic process was regarded as unintentional, uncontrolled, unconscious, and fast (82). Just as in the word-color experiment initially, participants in the current study could not “resist” processing word meaning in the prosodic task or pay attention to emotional prosody in the semantic task, and the interference made the difference.

Noise in Emotion Word Processing: Masking Effect

While the Stroop effect presented high automation, it did make a varying difference between semantic task and prosody task in the quiet condition. More specifically, in Experiment 1, under the interaction of “Congruency” and “Task,” “Congruency” exerted a higher influence in the prosody task ($***p < 0.001$) than in the semantic task ($**p < 0.01$). In other words, all participants were more affected by the Stroop effect in emotion identification with the interference of English word content, which might relate to experimental design. One potential factor to account for this could be that, compared with the word valence judgment in the semantic task, identifying emotions as happy or sad in the prosody task was easier for participants, which could be proved by the reduced response time. First, we selected the two most uncontroversial emotions that share across multi-cultures as the basic emotions. Unlike other finer emotions (e.g., suspicious, surprising, sarcasm, regret), familiarity and understandability increased the response efficiency. Therefore, it is very much unlikely for participants to misinterpret them. Second, given the significant pitch difference of audio stimuli between happy tone and sad tone, there existed an obvious contrast, with happy tone much higher than sad tone ($p < 0.001$), while no such significant pitch contrasts between positive and negative words were observed ($p = 0.808$). This was not surprising since the

happiness expression was always presented with explicit and unneglectable acoustic cues, such as higher pitch and quicker speed (64). Thus, participants reached faster responses within a short time. Third, although college students in this experiment have received an average of 15.83 years of education and learned English from a young age, they did not achieve perfect scores in LexTALE ($M = 55.11$). In semantic task, listening to English audio files only once and reacting within 5 s for second language learners could be a demanding task of pretty pressure, accompanied by a significant decrease of attention toward the emotional prosody of English words and a less Stroop effect. On the other hand, in the prosody task, going much easy on the emotion identification could always leave much other room for attention to verbal content, and the semantics-prosody channel congruency (positive-happy, negative-sad) or incongruency (positive-sad, negative-happy) mattered more. This aligned with the perceptual load theory: to what extent the task-irrelevant stimuli are processed is determined by whether there are spare attentional resources left when they are used to process the task-relevant stimuli (83).

However, this varying degree of Stroop effect between two tasks was not consistently observed in Experiment 2, where the listening background changed from quiet to speech-shaped noise. The interaction of “Congruency” and “Task” did not reach statistical significance in the noisy condition. Primarily, it is likely that the audio files accompanied by the noise created a relatively harsh environment for listeners to make their responses quickly. Unlike in the quiet condition where listeners could rely on the integrated prosody of words to identify emotions, they were probably forced to hear every syllable with much more effort to do the same job. And for these second language learners, mishearing only one syllable under a noisy condition could lead to loss and confusion about the lexical meaning of the whole word in the semantic task. Thus, the prosody tasks did not appear to be as easy and convenient due to the impact of noise. The increased difficulty of both tasks posed challenges for listeners to allocate their limited attention and seek the optimal solutions.

Moreover, the perception of English speech under noisy conditions occupied many more cognitive resources, including WM and inhibitory ability (84). After all, having controlled for language skills, WM, and emotional intelligence, the better inhibitory ability still predicted higher problem-solving accuracy (85). Factors such as noisy environment and other languages can adversely affect the speech perception process, leading to an increased difficulty for full understanding and a longer time to decode what was heard accurately (86). Since adverse listening conditions impair the encoding of speech signals, which means listeners have to allocate processing resources to separated aspects (87). (88, 89) also pointed out that the perceptual load of the current task participants conducted determined the allocation of cognitive resources during the process of selective attention. If the perceptual load of the current task was relatively low, and only a part of the attention resources was consumed in the process, then the extra attention resources would spare automatically to process the distractive stimuli, thus producing the distractive effect. On the contrary, if the perceptual load of the current task was high and the limited attention resources

were exhausted at once, the distractible stimulation unrelated to the task could not be perceptually processed, so the distractor effect will not be produced. In all in, in the current experiment, the additional cognitive, linguistic, and perceptual resources to understand English speech in noise, heavily consumed an individual's cognitive resources, leading to less Stroop effect.

Limitations and Future Directions

There exist several limitations in the current study. First, the conclusions were limited to trait-depressive Chinese college students of age 18–26 years. Given the extensive distribution of this mentally impaired population of all age groups, data, and information of only a fraction of college students, with even indistinctive second language competence in hearing, might limit how we interpret the model. Whether the results mentioned above reflect the psychological features of more common people remains unclear. Thus, a larger size of participants with marked characteristics is highly needed to draw more compelling views, with the assessing scales being of high validity. Second, compared with some previous studies adopting the Stroop-like paradigm, this research only focused on the binary cross-channel contrasts of audio emotional stimuli (semantic vs. prosodic), without applying more access to communication channels and modalities (e.g., facial expressions, videos). For the experimental design, practical settings are highly feasible. For instance, more types of noises with effects of closer authentic communication simulation or even real-life environment (i.e., babble noise), diverse emotions with finer classification sharing across cultures (i.e., surprise, sarcasm), multiple approaches to keeping abreast of language and psychological research to comprehend the neurological basis (i.e., event-related potential measures). Finally, it would be beneficial for further investigations on the clinical group of MDD to apply the current findings to the clinical setting.

CONCLUSION

In summary, this current study investigated psychological mechanisms of English emotion word processing under the semantics–prosody Stroop effect paradigm in quiet and noisy listening backgrounds, with Chinese college students with trait depression as participants. It was proved that the high trait depression group showed evident bluntness toward emotions compared with the low trait depression group in emotion word processing. And the widely existed Stroop effect affects the emotion word processing automatically, regardless of

participants' trait severity (i.e., high trait or low trait) and listening conditions (i.e., quiet or noisy). The results also showed that participants tended to be more affected by the Stroop effect when they conducted prosody tasks and recognized emotions than being affected in the semantic task and identified English word valence. However, such contrast was not observed with a background of speech-shaped noise, indicating the masking effect of noise on cognitive processing. Taken together, these findings provide evidence supporting the emotional processing deficit of high trait-depressive people and congruence-induced facilitation effect in widespread Stroop effect, which provide a reference on the cross-linguistic special group with multi-listening conditions for future studies and offer fairly basic evidence for clinical application of the trait depression.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the School of Foreign Languages, Hunan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FC conceived and designed the study, performed the statistical analysis, and offered the financial support. JL collected and analyzed the data, and wrote the first draft of the manuscript. GZ designed the study and interpreted the data. CG participated in the statistical analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partly supported by the Social Science Foundation of Hunan Province (No. 20ZDB003) and Fundamental Research Funds for the Central Universities, Hunan University (No. 531118010660).

REFERENCES

1. Kasap S, Tanhan F. The effect of body posture on foreign language anxiety. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*. (2019) 37:46–65.
2. Schroeder J, Kardas M, Epley N. The humanizing voice: speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychol Sci*. (2017) 28:1745–62. doi: 10.1177/0956797617713798
3. Grandjean D, Bänziger T, Scherer KR. Intonation as an interface between language and affect. *Prog Brain Res*. (2006) 156:235–47. doi: 10.1016/S0079-6123(06)56012-1
4. Paulmann S, Pell MD. Is there an advantage for recognizing multi-modal emotional stimuli? *Motiv Emot*. (2011) 35:192–201. doi: 10.1007/s11031-011-9206-0
5. Wilson D, Wharton T. Relevance and prosody. *J Pragmat*. (2006) 38:1559–79. doi: 10.1016/j.pragma.2005.04.012
6. Yang L, Campbell N. Linking form to meaning: the expression and recognition of emotions Through prosody. *Proceedings of the 4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis*. (Kolkata: ISCA) (2001).
7. Song Y, Zhong J, Jia Z, Liang D. Emotional prosody recognition in children with high-functioning autism under the influence of emotional intensity:

- based on the perspective of emotional dimension theory. *J Commun Disord.* (2020) 88:106032. doi: 10.1016/j.jcomdis.2020.106032
8. Morton JB, Trehub SE. Children's understanding of emotion in speech. *Child Dev.* (2001) 72:834–43. doi: 10.1111/1467-8624.00318
 9. Lin Y, Ding H, Zhang Y. Prosody dominates over semantics in emotion word processing: evidence from cross-channel and cross-modal stroop effects. *J Speech Lang Hear Res.* (2020) 63:896–912. doi: 10.1044/2020_JSLHR-19-00258
 10. Hellbernd N, Sammler D. Prosody conveys speaker's intentions: acoustic cues for speech act perception. *J Memory Lang.* (2016) 88:70–86. doi: 10.1016/j.jml.2016.01.001
 11. Bach DR, Buxtorf K, Grandjean D, Strik WK. The influence of emotion clarity on emotional prosody identification in paranoid schizophrenia. *Psychol Med.* (2009) 39:927–38. doi: 10.1017/S0033291708004704
 12. Mitchell RL. Age-related decline in the ability to decode emotional prosody: primary or secondary phenomenon? *Cogn Emot.* (2007) 21:1435–54. doi: 10.1080/02699930601133994
 13. Wittfoth M, Schröder C, Schardt DM, Dengler R, Heinze HJ, Kotz SA. On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. *Cereb Cortex.* (2010) 20:383–92. doi: 10.1093/cercor/bhp106
 14. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol.* (1935) 18:643–62. doi: 10.1037/h0054651
 15. Filippi P, Ocklenburg S, Bowling DL, Heege L, Güntürkün O, Newen A, et al. More than words (and faces): evidence for a stroop effect of prosody in emotion word processing. *Cogn Emot.* (2017) 31:879–91. doi: 10.1080/02699931.2016.1177489
 16. Barnhart WR, Rivera S, Robinson CW. Different patterns of modality dominance across development. *Acta Psychol.* (2018) 182:154–65. doi: 10.1016/j.actpsy.2017.11.017
 17. Spence C, Parise C, Chen Y-C. The colavita visual dominance effect. In: Murray MM, Wallace MT editors. *The Neural Bases of Multisensory Processes.* (Abingdon: CRC Press/Taylor & Francis) (2012).
 18. Liu P, Rigoulot S, Pell MD. Culture modulates the brain response to human expressions of emotion: electrophysiological evidence. *Neuropsychologia.* (2015) 67:1–13. doi: 10.1016/j.neuropsychologia.2014.11.034
 19. Pell MD, Jaywant A, Monetta L, Kotz SA. Emotional speech processing: disentangling the effects of prosody and semantic cues. *Cogn Emot.* (2011) 25:834–53. doi: 10.1080/02699931.2010.516915
 20. Schwartz R, Pell MD. Emotional speech processing at the intersection of prosody and semantics. *PLoS One.* (2012) 7:e47279. doi: 10.1371/journal.pone.0047279
 21. Kindermann SS, Kalayam B, Brown GG, Burdick KE, Alexopoulos GS. Executive functions and P300 latency in elderly depressed patients and control subjects. *Am J Geriatr Psychiatry.* (2000) 8:57–65. doi: 10.1097/00019442-200002000-00008
 22. Zakzanis KK, Leach L, Kaplan E. On the nature and pattern of neurocognitive function in major depressive disorder. *Neuropsychiatry Neuropsychol Behav Neurol.* (1998) 11:111–9.
 23. Ebmeier KP, Donaghey C, Steele JD. Recent developments and current controversies in depression. *Lancet.* (2006) 367:153–67. doi: 10.1016/S0140-6736(06)67964-6
 24. Gotlib IH, Lee CM. The social functioning of depressed patients: a longitudinal assessment. *J Soc Clin Psychol.* (1989) 8:223–37. doi: 10.1521/jscp.1989.8.3.223
 25. Levendosky AA, Okun A, Parker JG. Depression and maltreatment as predictors of social competence and social problem-solving skills in school-age children. *Child Abuse Neglect.* (1995) 19:1183–95. doi: 10.1016/0145-2134(95)00086-N
 26. Lei ZH, Xv R, Deng SB, Luo YJ. Reliability and validity of the Chinese version of state-trait depression scale in college students. *Chin Ment Health J.* (2011) 25:136–40.
 27. Costa PT, McCrae RR. *Neo Personality Inventory-Revised (NEO PI-R)*. Odessa, FL: Psychological Assessment Resources (1992).
 28. Terracciano A, Tanaka T, Sutin AR, Sanna S, Deiana B, Lai S, et al. Genome-wide association scan of trait depression. *Biol Psychiatry.* (2010) 68:811–7. doi: 10.1016/j.biopsych.2010.06.030
 29. Hallion LS, Ruscio AM. A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychol Bull.* (2011) 137:940–58. doi: 10.1037/a0024355
 30. Iijima Y, Takano K, Boddez Y, Raes F, Tanno Y. Stuttering thoughts: negative self-referent thinking is less sensitive to aversive outcomes in people with higher levels of depressive symptoms. *Front Psychol.* (2017) 8:1333. doi: 10.3389/fpsyg.2017.01333
 31. Takano K, Iijima Y, Sakamoto S, Raes F, Tanno Y. Is self-positive information more appealing than money? Individual differences in positivity bias according to depressive symptoms. *Cogn Emot.* (2016) 30:1402–14. doi: 10.1080/02699931.2015.1068162
 32. Rittlerband LM, Spielberger CD. Construct validity of the Beck depression inventory as a measure of state and trait depression in nonclinical populations. *Depress Stress.* (1996) 2:123–45. doi: 10.1016/S0165-0327(97)00094-3
 33. Gao C, He CS, Yan XZ. Emotional conflicts of college students with trait depression under the word-face Stroop paradigm. *J Heihe University.* (2019) 8:56–8.
 34. Endler NS, Macrodimitris SD, Kocovski NL. Depression: the complexity of self-report measures. *J Appl Biobehav Res.* (2000) 5:26–46. doi: 10.1111/j.1751-9861.2000.tb00062.x
 35. Fazio RH, Sanbonmatsu DM, Powell MC, Kardes FR. On the automatic activation of attitudes. *J Pers Soc Psychol.* (1986) 50:229–38. doi: 10.1037/0022-3514.50.2.229
 36. Schirmer A, Kotz SA. ERP evidence for a sex-specific stroop effect in emotional speech. *J Cogn Neurosci.* (2003) 15:1135–48. doi: 10.1162/089982903322598102
 37. Emerson CS, Harrison DW, Everhart DE. Investigation of receptive affective prosodic ability in school-aged boys with and without depression. *Neuropsychiatry Neuropsychol Behav Neurol.* (1999) 12:102–9.
 38. Schlipf S, Batra A, Walter G, Zeep C, Wildgruber D, Fallgatter A, et al. Judgment of emotional information expressed by prosody and semantics in patients with unipolar depression. *Front Psychol.* (2013) 4:461. doi: 10.3389/fpsyg.2013.00461
 39. Kan Y, Mimura M, Kamijima K, Kawamura M. Recognition of emotion from moving facial and prosodic stimuli in depressed patients. *J Neurol Neurosurg Psychiatry.* (2004) 75:1667–71. doi: 10.1136/jnnp.2004.036079
 40. Sloan DM, Bradley MM, Dimoulas E, Lang PJ. Looking at facial expressions: dysphoria and facial EMG. *Biol Psychol.* (2002) 60:79–90. doi: 10.1016/S0301-0511(02)00044-3
 41. Dalili MN, Penton-Voak IS, Harmer CJ, Munafo MR. Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychol Med.* (2015) 45:1135–44. doi: 10.1017/S0033291714002591
 42. Garrido-Vásquez P, Jessen S, Kotz SA. Perception of emotion in psychiatric disorders: on the possible role of task, dynamics, and multimodality. *Soc Neurosci.* (2011) 6:515–36. doi: 10.1080/17470919.2011.620771
 43. Krause FC, Linardatos E, Fresco DM, Moore MT. Facial emotion recognition in major depressive disorder: a meta-analytic review. *J Affect Disord.* (2021) 293:320–8. doi: 10.1016/j.jad.2021.06.053
 44. Péron J, El Tamer S, Grandjean D, Leray E, Travers D, Drapier D, et al. Major depressive disorder skews the recognition of emotional prosody. *Prog Neuro Psychopharmacol Biol Psychiatry.* (2011) 35:987–96. doi: 10.1016/j.pnpb.2011.01.019
 45. Canli T, Sivers H, Thomason ME, Whitfield-Gabrieli S, Gabrieli JDE, Gotlib IH. Brain activation to emotional words in depressed vs healthy subjects. *Neuroreport.* (2004) 15:2585–8. doi: 10.1097/00001756-200412030-00005
 46. Epstein J, Pan H, Kocsis JH, Yang Y, Butler T, Chusid J, et al. Lack of ventral striatal response to positive stimuli in depressed versus normal subjects. *Am J Psychiatry.* (2006) 163:1784–90. doi: 10.1176/ajp.2006.163.10.1784
 47. Gao S, Jiao JL, Liu Y, Wen SX, Qiu XS. Bilingual advantage in Stroop under the condition of different languages. *J Psychol Sci.* (2017) 40:315–20.
 48. Tse CS, Altarriba J. The effects of first-and second-language proficiency on conflict resolution and goal maintenance in bilinguals: evidence from reaction time distributional analyses in a stroop task. *Bilingualism.* (2012) 15:663–76. doi: 10.1017/S1366728912000077
 49. Bower GH. Mood congruity of social judgments. 1st ed. In: Forgas JP editor. *Emotion and Social Judgments.* (New York, NY: Garland Science) (1991). p. 31–53. doi: 10.4324/9781003058731-3
 50. Ehrman ME, Leaver BL, Oxford RL. A brief overview of individual differences in second language learning. *System.* (2003) 31:313–30. doi: 10.1016/S0346-251X(03)00045-9

51. Ekman P. Are there basic emotions? *Psychol Rev.* (1992) 99:550–3. doi: 10.1037/0033-295X.99.3.550
52. Ekman P, Cordaro D. What is meant by calling emotions basic. *Emot Rev.* (2011) 3:364–70. doi: 10.1177/1754073911410740
53. Cusi AM, Nazarov A, Holshausen K, MacQueen GM, McKinnon MC. Systematic review of the neural basis of social cognition in patients with mood disorders. *J Psychiatry Neurosci.* (2012) 37:154–69. doi: 10.1503/jpn.100179
54. McLaughlin DJ, Baese-Berk MM, Bent T, Borrie SA, Van Engen KJ. Coping with adversity: individual differences in the perception of noisy and accented speech. *Atten Percept Psychophys.* (2018) 80:1559–70. doi: 10.3758/s13414-018-1537-4
55. Speck I, Rottmayer V, Wiebe K, Aschendorff A, Thurow J, Frings L, et al. PET/CT background noise and its effect on speech recognition. *Sci Rep.* (2021) 11:22065. doi: 10.1038/s41598-021-01686-5
56. Spielberger CD. *State-Trait Depression Scales (Form X-1)*. Palo Alto, CA: Mind Garden (1995).
57. Krohne HW, Schmukle SC, Spaderna H, Spielberger CD. The state-trait depression scales: an international comparison. *Anxiety Stress Coping.* (2002) 15:105–22. doi: 10.1080/10615800290028422
58. Lemhöfer K, Broersma M. Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behav Res Methods.* (2012) 44:325–43. doi: 10.3758/s13428-011-0146-0
59. Baddeley A. Working memory: looking back and looking forward. *Nat Rev Neurosci.* (2003) 4:829–39. doi: 10.1038/nrn1201
60. Schutte NS, Malouff JM, Hall LE, Haggerty DJ, Cooper JT, Golden CJ, et al. Development and validation of a measure of emotional intelligence. *Pers Individ Differ.* (1998) 25:167–77. doi: 10.1016/S0191-8869(98)00001-4
61. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia.* (1971) 9:97–113. doi: 10.1016/0028-3932(71)90067-4
62. Davies M. *The Corpus of Contemporary American English (COCA)*. (2008). Available online at <https://www.english-corpora.org/coca/> (accessed October 01, 2021).
63. Schirmer A, Kotz SA, Friederici AD. Sex differentiates the role of emotional prosody during word processing. *Brain Res. Cogn. Brain Res.* (2002) 14:228–33. doi: 10.1016/S0926-6410(02)00108-8
64. Scherer KR, Banse R, Wallbog HG. Emotion inferences from vocal expression correlate across languages and cultures. *J Cross Cult Psychol.* (2001) 32:76–92. doi: 10.1177/0022022101032001009
65. Babjack DL, Cernicky B, Sobotka AJ, Basler L, Struthers D, Kisc R, et al. Reducing audio stimulus presentation latencies across studies, laboratories, and hardware and operating system configurations. *Behav Res Methods.* (2015) 47:649–65. doi: 10.3758/s13428-015-0608-x
66. Psychology Software Tools. *World-Leading Stimulus Presentation Software E-Prime 3.0*. Pittsburgh, PA: Psychology Software Tools (2021).
67. Helfer KS, Freyman RL. Aging and speech-on-speech masking. *Ear Hear.* (2008) 29:87–98. doi: 10.1097/AUD.0b013e31815d638b
68. Tun PA, Wingfield A. One voice too many: adult age differences in language processing with different types of distracting sounds. *J Gerontol B Psychol Sci Soc Sci.* (1999) 54B:317–27. doi: 10.1093/geronb/54B.5.P317
69. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2018).
70. Baayen RH, Milin P. Analyzing reaction times. *Int J Psychol Res.* (2010) 3:12–28. doi: 10.21500/20112084.807
71. Lawrence MA. *ez: Easy Analysis and Visualization of Factorial Experiments. R Package Version 4.4-0*. (2016). Available online at: <https://CRAN.R-project.org/package=ez> (accessed November 01, 2021).
72. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* (2015) 67:1–48. doi: 10.18637/jss.v067.i01
73. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Memory Lang.* (2013) 68:255–78. doi: 10.1016/j.jml.2012.11.001
74. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *J Stat Softw.* (2017) 82:1–26.
75. Lenth RV. Least-squares means: the R package lsmeans. *J Stat Softw.* (2016) 69:1–33. doi: 10.18637/jss.v069.i01
76. Uekermann J, Abdel-Hamid M, Lehmkämpfer C, Vollmoeller W, Daum I. Perception of affective prosody in major depression: a link to executive functions? *J Int Neuropsychol Soc.* (2008) 14:552–61. doi: 10.1017/S1355617708080740
77. Yan WR, Wang ZH. The characteristic of heart rate variability in state-trait depressive undergraduates. *Proceedings of the 14th National Conference of Psychology*. Xi'An: (2011).
78. MacLeod CM. Half a century of research on the stroop effect: an integrative review. *Psychol Bull.* (1991) 109:163–203. doi: 10.1037/0033-2909.109.2.163
79. Banich MT, Milham MP, Atchley R, Cohen NJ, Webb A, Wszalek T, et al. fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *J Cogn Neurosci.* (2000) 12:988–1000. doi: 10.1162/08989290051137521
80. Besner D, Stolz JA, Boutilier C. The stroop effect and the myth of automaticity. *Psychon Bull Rev.* (1997) 4:221–5. doi: 10.3758/BF03209396
81. Cattell JM. The time it takes to see and name objects. *Mind.* (1886) 11:63–5. doi: 10.1093/mind/os-XI.41.63
82. Moors A, De Houwer J. Automaticity: a theoretical and conceptual analysis. *Psychol Bull.* (2006) 132:297–326. doi: 10.1037/0033-2909.132.2.297
83. Lavie N. Perceptual load as a necessary condition for selective attention. *J Exp Psychol.* (1995) 21:451–68. doi: 10.1037/0096-1523.21.3.451
84. Gordon-Salant S, Shader MJ, Wingfield A. Age-related changes in speech understanding: peripheral versus cognitive influences. In: Helfer KS, Bartlett EL, Popper AN, Fay RR editors. *Aging and Hearing: Causes and Consequences*. (Basel: Springer International Publishing) (2020). p. 199–230. doi: 10.1007/978-3-030-49367-7_9
85. Coderre EL, Filippi CG, Newhouse PA, Dumas JA. The stroop effect in kana and kanji scripts in native Japanese speakers: an fMRI study. *Brain Lang.* (2008) 107:124–32. doi: 10.1016/j.bandl.2008.01.011
86. Van Engen KJ, Peelle JE. Listening effort and accented speech. *Front Hum Neurosci.* (2014) 8:577. doi: 10.3389/fnhum.2014.00577
87. Murphy DR, Craik FIM, Li KZH, Schneider BA. Comparing the effects of aging and background noise on short-term memory performance. *Psychol Aging.* (2000) 15:323–34. doi: 10.1037/0882-7974.15.2.323
88. Lavie N, Cox S. On the efficiency of visual selective attention: efficient visual search leads to inefficient distractor rejection. *Psychol Sci.* (1997) 8:395–6. doi: 10.1111/j.1467-9280.1997.tb00432.x
89. Lavie N, Fox E. The role of perceptual load in negative priming. *J Exp Psychol.* (2000) 26:1038–52. doi: 10.1037/0096-1523.26.3.1038

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Lian, Zhang and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.