Running head: ERP STUDY ON EMOTIONAL PROSODY

An ERP Study on Emotional Prosody among Cantonese speakers

Ming LUI

The Chinese University of Hong Kong

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Philosophy in Psychology

May 2003

Abstract

The present study aimed at studying the interaction between emotional prosody and semantic information in Cantonese speech communication, as well as the gender differences in emotional prosody processing among Chinese. Participants listened to semantically positive and negative words spoken with either happy or sad prosody. They were asked to make forced-choice semantic judgment of whether the words they heard were positive or negative while ignoring the emotional prosody. Behaviorally, participants responded more slowly and less accurately to incongruent than congruent happy prosody stimuli. While men did not differ in reaction time for positive and negative words, women responded significantly slower in response to negative words than to positive words. The event-related potential (ERP) data showed that significantly larger negativity was elicited in right hemisphere for women in response to stimuli, while there was no significant difference between the left and right hemisphere ERP amplitude for men. Moreover, women but not men, showed a larger positivity towards prosodic-semantic incongruent stimuli compared to congruent stimuli between 750 and 1400 msec post-stimulus onset. Results of the present study suggest that although an emotional Stroop effect exists among Cantonese speakers, there were differences between the present data and past findings in terms of the processing time courses and ERP components. Finally, the gender difference in emotional prosody processing was specific to sad prosody and negative meaning words for Cantonese speakers.

摘要

本實驗旨在研究廣東話語氣及語意之間的互動影響，以及其性別差異。實驗參加者聆聽帶有快樂或哀傷語氣的正面或負面詞語。他們被要求在作出語意決定時忽略其刺激物所帶有的語氣。實驗結果顯示當刺激物帶有快樂語氣時，若語氣與語意不一致，參加者反應顯著較慢。男性對正面及負面語意的反應時間沒有顯著分別，但女性對負面詞語反應顯著較慢。腦電波數據顯示，當女性回應刺激物時右腦有顯著較負極的電位，但男性在這方面則沒有顯著分別。女性回應語氣及語意不一致的刺激物時有較大正極電位。研究結果顯示在行為上情感 Stroop 效應存在於廣東話使用者。在腦電波活動方面，此次研究結果則與過去所獲數據有差別。廣東話使用者語氣和語意處理的性別差異集中於哀傷語氣及負面詞語。

Acknowledgment

I would like to show my deep thankfulness to my thesis supervisor, Prof. Trevor B. Penney, and Dr. Annett Schirmer for their inspirations and guidance during the proceeding of the project. I would also like to thank Prof. H. C. Chen for his comments on the stimulus construction, and Prof. Catherine McBride for being my thesis examiner. Also, I cordially thank Elaine Ng-Hoi Ning and Pelen Yip Pak Yam for their support during the data collection. And I thank Joe Lau Tak Fu sincerely for emotional support during the writing of the thesis. Finally I would like to give all my thanks to the Lord who is the stream of strength in my life.

Table of Contents

An ERP Study on Emotional Prosody among Cantonese speakers

One unique characteristic of humans is the highly advanced system of communication that enables the passage of knowledge and wisdom across time and space. Apart from the sharing of knowledge, another important function of human communication is the sharing of subjective experience- feelings and emotions during social interaction. Though verbal communication is probably the most direct and common among various ways of human interaction, the most accurate emotional meaning is often expressed through non-verbal cues like emotional prosody. An example which can fully demonstrate the importance of nonverbal cues in conveying meaning is the use of sarcasm, in which the surface meaning is often opposite to the underlying message intended to be delivered. Due to the specific nature of emotional prosody in human communication, it is intriguing to study the interaction of prosody and semantics in speech processing, which may further reveal the underlying mechanisms of semantic recognition in humans. Moreover, the cross-cultural study of emotional prosody processing is particularly important for revealing whether human language processing mechanism is innate. Past studies generally supported the existence of a genetic origin for the ability to recognize meaning from emotional prosody, as it appears to emerge consistently by the first 6 years of life (Matsumoto & Kishimoto, 1983) and it is observed across different cultures (Scherer, Banse, & Wallbott, 2001). It is important to study the behavioral pattern and neural activities during emotional prosody processing in both tonal and non-tonal language speakers, as well as language users in different cultures, so as to reveal the picture of the interaction between nature and nurture factors on language development.

In the following sections, before discussing the effect of emotional prosody on speech processing, models of word recognition will first be evaluated. We will look at when and how emotional information takes a role in semantic processing as proposed in some word recognition models. Then we will introduce the nature of emotional prosody, followed by a discussion of the potential factors, including gender, lateralization, characteristics of tonal languages, and cultural factors, which may modulate the interaction between emotional prosody and semantic processing. Finally, the investigation methods used for the study of emotional-prosodic interaction in the present study will be described.

Models of word recognition

When evaluating the past studies on word recognition, we see that a three-stage model which describes stages ranging from the primary sensory input to the semantic unit (Marslen-Wilson, 1987) is generally supported. The first stage is lexical access, in which the phonological or visual word form activates its corresponding entry in the mental lexicon. The mental lexicon is a hypothesized human cognitive storage system which registers "the central abstract representation of a word's semantic, syntactic, and morphological properties" (Marslen-Wilson, 1987) and each word has a corresponding entry in it. Some researchers believe that the stage of lexical access is purely stimulus driven (Soto-Faraco, Sebastian-Galles, & Cutler, 2001).

The second stage is lexical selection and it is an intermediate stage between lexical form and contextual meaning. It was proposed that if the linguistic input is sufficient for word identification, the word together with its semantic (and emotional), syntactic and morphologic properties gets selected from the lexicon (Hagoort & Brown, 2000). This process is particularly important in the case of lexical ambiguity in which appropriate meaning has to be decided on the basis of contextual information.

The final stage, namely lexical integration, is post-lexical and relates the properties of a word selected from the lexicon to the context in which it has been presented. While the second stage lexical access is regarded as a "reflex-like and effortless behavior, which cannot be controlled by the subject" (Hagoort & Brown, 1994), the lexical integration is assumed to be a controlled process "that can be guided by the subject's awareness of the informational content of the discourse".

There is another well-supported model of word recognition, known as the "TRACE" model (McClelland & Rumelhart, 1981; McClelland & Elman, 1986). We may try to integrate the "TRACE" interactive activation model with the three-stage model proposed by Zwitserlood (1989). It is supposed that the processes described in the TRACE model lie within the stages of lexical access and lexical selection proposed by Zwitserlood (1989). In the TRACE model, the authors propose that features, phonemes and word units are organized hierarchically. Feature units are first excited, which then excite phonemes, and finally activate words. All levels are highly interactive and activated units in the same level inhibit each other. The activation of word units by the phoneme unit proposed by the TRACE model is in some sense analogous to the lexical access stage described above, which is a stage when phonological form activates its corresponding entry in the mental lexicon. Moreover, when certain word units are activated to the largest extent, the lexical selection is performed (Connine, 1987).

Emotion and semantics

The component of emotion in word recognition lies in its inextricably intimate relationship with the semantic representation in the mental lexicon. It is proposed that the emotional aspects of a word constitute a major component that is first accessed from the mental lexicon among all semantic features (Osgood, Suci & Tannenbaum, 1957). In speech perception, one may regard the emotions conveyed in a speaker's voice (i.e., emotional prosody) as providing contextual information similar to the semantic context established by words (Schirmer, Kotz, & Friederici, 2002).

Emotional Prosody

Prosody is a feature of human speech determined by the physical components of fundamental frequency (F0), duration, intensity, harmonics and spectral energy distribution (Bolinger, 1964). Its significance lies in both linguistic and emotional expression values. Prosody influences linguistic categories such as tone, stress, vowel length, phrasing and (linguistic) intonation, as well as expressive categories, such as (affective) intonation, timbre, tempo and melody (Bolinger, 1964).

The acoustic parameters that Scherer (1989) used to measure the emotional content of vocalization partially overlap with the prosodic parameters introduced by Bolinger (1964). He listed four major classes of vocal parameters: vocal intensity, vocal frequency, vocal quality, and vocal resonance. In a production and perception study of emotional prosody, Banse and Scherer (1996) collected the acoustic profiles of 14 different emotions as vocalized by 6 actresses and 6 actors. The results showed that each emotion has a characteristic profile for the parameters mentioned above. For example, a sadness utterance consists of a decrease in F0, a restricted F0-variability, lower intensity and a slower speech rate. On the other hand, a happy utterance is usually characterized by a faster, louder, and higher fundamental frequency voice as compared to an utterance spoken in a sad state.

Potential modulating factors of the interaction between emotional prosody and semantics

Gender

It is well-established in sociocultural and psychological research that women have an advantage over men towards emotional expression and perception, both verbally and non-verbally (Brennan, Goddard, Wilson & Kinnear, 1991; Hall, 1978). With the exception of anger, which is found to be expressed more frequently by men than women (Hubbard, 2001), women are generally more expressive of various types of emotion (Hess, Senecal, Kirouac, Herrera, Phillippot, & Kleck, 2000). The advantage of women in emotional perception is also supported by a meta-analysis of 75 studies which investigated accuracy in nonverbal emotional cue recognition. A small but consistent female superiority in the recognition of emotions from voice, faces and gestures was found (Hall, 1978), and the differences were even strongest in studies that required the recognition of combined visual and auditory processing. A study by Rotter and Rotter (1988) on sex differences in recognizing emotionally negative facial expression revealed that men not only show exceptionally good performance in anger expression over women, but also have an advantage in anger perception. A sex difference in emotional prosody perception was further supported recently by Schirmer, Kotz and Friederici's (2002) study. In this study, the influence of emotional prosody on the processing of visually presented positive and negative target words was investigated. The stimulus material consisted of semantically neutral sentences spoken with happy or sad tones. The presentation of each sentence was followed by either a German word which was semantically related to the sentence final word or a pseudo word. The target word had a positive or a negative meaning, which matched or did not match the emotional prosody of the preceding sentence. Participants were asked to judge whether the target word was a word or non-word. The results supported the modulation of emotional prosody on word processing. However, the time course of this modulation differed for males and females. Behaviorally women were more accurate in recognizing emotional prosody at a significantly shorter interstimulus interval (women: 200 msec; men: 750msec) than men in a cross modal priming experiment. Women also showed electrophysiological priming effects with a smaller time interval between the prosodic prime and the visual target word compared to men. The results may suggest a more rapid integration of emotional-prosodic context and semantic information by women than men (Schirmer, Kotz & Friederici, 2002).

In another study by Schirmer and Kotz (in press), a sex difference in the emotional prosody-semantic interaction was also found. The stimuli used in the study were positive, neutral and negative words spoken with a happy, neutral and angry tone. Participants were asked to rate word valence while ignoring the emotional prosody in the prosodic judgment task or vice versa in the semantic judgment task. Behaviorally, men and women were not significantly different. They both responded faster and more accurately to congruent compared to incongruent prosodic-semantics stimuli, and the effects were more salient for the semantic judgment task than for the prosodic judgment task for both men and women. While behaviorally a sex difference was not indicated, electrophysiologically men and women did show a difference in the underlying mechanism. The effect on an event related potential (ERP) component which is sensitive to the interaction between prosody and semantics was significant only for women (Schirmer & Kotz, in press). The specific details of the electrophysiological differences obtained in the studies by Schirmer and colleagues are presented below.

Lateralization

A factor that is closely related to gender differences is hemispheric lateralization. It is claimed by some researchers that there is a right hemisphere lateralization for cognitive processing involving emotion (Blonder, Burns, Bowers, Moore, 1993; Heilman, Blonder, Bowers, & Valenstein, 2003). Prosody, which serves both linguistic and emotional functions, is found to be lateralized specifically by its functions. Right hemisphere lateralization is evident for processing of emotional prosody as opposed to left hemisphere lateralization for the linguistic function of semantic processing (Heilman, Blonder, Bowers, & Valenstein, 2003). Focusing on emotional prosody perception, it was found that the right lateralization begins with the processing of voice in the auditory cortex (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), as well as for the acoustic analysis of melodies (Zatorre, Evans, & Meyer, 1994). Studies using evoked potentials (Pihan, 1997), functional imaging (Buchanan et al., 2000), and lesion paradigms (Barrett, Crucian, Raymer, & Heliman, 1999) all support the role of the right hemisphere in emotional prosody processing. For example, Schmitt, Hartje, and Willmes (1997) demonstrated that when judging a large inventory of multimodal stimuli, right hemisphere damaged patients showed a much higher occurrence of impairment in emotional perception towards prosodic and facial cues than left hemisphere damaged patients. More recent patient studies have supported the involvement of the right frontoparietal regions (Starkstein, Federoff, Price, Leiguarda, & Robinson, 1994), the right frontal regions (Breitenstein, Daum, & Ackermann, 1998) and the right orbital frontal regions (Hornak, Rolls, & Wade, 1996) in emotional prosody processing. Functional imaging studies have even more precisely located activities in the right dorsal and ventral prefrontal cortex (George, Parekh, Rosinsky, Ketter, Kimbrell, & Heilman, 1996, Morris, Scott, & Dolan, 1999). On the other hand, in a fMRI study (Buchanan, Lutz, Mirzazade, Specht, Shah, Zilles, Jaencke, 2000) examining the neural substrates involved in emotional prosody and phonemic recognition, bilateral activities were observed. Significant activity was found in the right inferior frontal lobe during the recognition of emotional component compared with that during the recognition of phonemic component. In contrast, significant activity was found in the left inferior frontal lobe regions during the recognition of phonemic component compared to that during the recognition of emotional component. To conclude, the processing of emotional prosody may involve distributed regions in both hemispheres. For recognition involving more emotional processing, regions in the right hemisphere may take a larger role in the task, while for recognition involving more linguistic processing, the left regions may dominate.

Regarding the phenomenon of gender differences in emotional prosody perception, it may be logical to refer to lateralization differences in the explanation. Women are generally believed to be more bilateral in language processing while men are more left hemisphere lateralized, as supported by numerous fMRI (Shaywitz et al.,1995), dichotic listening (Hiscock et al., 1994), and tachistoscopic studies (Hiscock et al., 1994). Since the right hemisphere is thought to be crucial in the processing of emotion (e.g., Pell & Baum, 1997), it was proposed that women might profit from shorter neuronal connections between areas that process language and areas that process emotional prosody. Nevertheless, there also exist a number of studies showing no significant gender differences in lateralization of language functions (e.g., Schirmer, Kotz & Friederici, 2002).

Characteristics of tonal language

A tonal language (e.g., Cantonese) is a language having the property that the meaning can be changed by varying the tone of a syllable (So & Dodd, 1995). Tone in tonal languages is also characterized by the lack of an absolutely fixed fundamental frequency for each distinctive tone. From occasion to occasion, each tone is subject to the influence of factors like intonation, which can distort the tone register and contour (Connell et al., 1983). Still, native speakers are able to perceive tone-dependent lexical items in context despite the deviation in F0 from the norm or citation form.

In Cantonese there are six contrastive tones that distinguish different homophone words: high level, high rising, mid level, low falling, low rising, low level. Additionally, there are three entering tones that are allotones of the three level tones. Stress and tone are highly correlated in Cantonese, with high tones receiving greater stress.

Tonal languages like Cantonese are special in that prosody can serve at two levels of function. On the linguistic level, prosody is often decisive in word identification as it involves variation in tone. At the same time, an emotional message is often conveyed through speech melody. The dual function of prosody in tonal language, however, creates some constraints on the manifestation of the emotional expressive function. Research reveals that the pitch level (register), pitch range, and final syllable of tonal languages are all influenced by the emotional attitude of the speaker (Chao, 1968). Be that as it may, due to the linguistic function of tones in tonal language, the expression of affective prosody through these parameters is also restricted, especially on the modulation of F0. This kind of constraint does not usually occur in non-tonal languages (Ross, Edmondson & Seibert, 1986). Specifically, in tonal languages, the relative height among tones may serve the linguistic function of signaling different words. Owing to this, the range of variation for the expression of emotional prosody is limited so that the lexical information would not be disrupted. It is evident that tonal languages generally allow fewer possibilities than non-tonal languages for variations in both intonation and the free use of pitch information for emotional expression purposes (Lee, Vakoch & Wurm, 1996). Instead, segmental markers like intensity and timing parameters are significant for showing emotional meaning (Chao, 1968) due to the limitation in the variation of prosodic signal.

Be that as it may, one should note that the limitation of emotional function by prosody cues is a matter of speech signaling rather than speech perception. The limitation in tone variation for emotional expression in tonal language does not necessarily imply anything about emotional speech perception. Further research is needed to reveal the underlying mechanism of emotional perception from speech under the limited tone variation during emotional expression in tonal languages. It is possible that due to the relatively smaller range of emotional tone variation in tonal languages, native tonal language speakers may develop higher sensitivity in the perception of subtle variation in emotional tone, and may also be more attentive to such variation during speech perception. Alternatively, the restriction of emotional prosodic expression may also lead to reliance on non-verbal cues like facial expression or other verbal cues like the semantic context during communication.

The Cultural Factor in Chinese

Culture and socialization pose a significant influence on the development of language. Indeed, some authors claim that the relationship between linguistic and cultural aspects is so intimate that it is hard to isolate the "language per se" from the entire cultural pattern of practices (Kitayama & Ishii, 2002).

It is generally believed that there is a difference in communicative styles between individualistic and collectivistic cultures (Hall, 1976). Hall (1976) raised two types of communicative practices- "low-context" style and "high-context" style. "Low-context" type refers to the communication style in which a large proportion of information is conveyed by verbal context, while contextual cues like vocal tone serve a relatively minor role. This communication practice exists in many individualistic cultures (e.g., German, English), in which discourses are characterized as direct, succinct, personal and instrumental (Gudykunst, Ting-Toomey & Chua, 1988). In contrast, the "high-context" type refers to a communication style that has a low proportion of information conveyed by verbal content, and as a result contextual and non-verbal cues play a larger role in expressing meaning. This type of communication is common in collectivistic societies (e.g., Japan, China), in which people's conversation is often indirect, elaborate, contextual and more implicit (Gudykunst, Ting-Toomey & Chua, 1988). This is why in Japanese communicative practice, people often find that daily utterances become ambiguous when taken out of context (Barnlund, 1989).

Recent studies on cultural variations in cognition have demonstrated that people in eastern cultures are more sensitive to contextual information in a variety of inference tasks than are those in western cultures (Kitayama, 2000; Nisbett, Peng, Choi, & Norenzayan, 2001). People socialized in low-context cultures develop a well-practiced attentional bias towards verbal content, whereas speakers and listeners in high-context cultures tend to bias towards vocal tones during communication.

The origin of the difference in communication style between individualistic and collectivistic societies can be explained by the deep-rooted cultural beliefs in social interaction. For collectivistic societies, one major taboo in social interaction is the expression of emotions that may disrupt social harmony (Smith & Bond, 1993). While people in individualist cultures are guided primarily by the rule of clarity (Kim, 1994) during social interaction, speakers within collectivist cultures were guided by the politeness norm of avoidance of imposition on the hearer and hurting the hearer's feelings. Among Chinese, people who can hide their emotion across different emotionally arousing situation are often highly regarded. As an old Chinese proverb says, "泰山崩於前而色不變". This phrase is describing people who can keep their facial expression indifferent even when a hill falls down in front of them. Moreover, the expression of extreme happiness or anger during social interaction is regarded as a lack of manners and politeness. Argyle et al. (1986) also found that rules restraining social expression of anger and distress were more strongly endorsed in Japan and Hong Kong, which scored higher on collectivism than Italy or the United Kingdom, which scored high on individualism.

Recently a cross-cultural study (Kitayama & Ishii, 2002) was carried out to investigate the perception of emotional prosody in both individualistic (American) and collective (Japanese) societies. The prosody-semantic interference task revealed that Japanese respondents showed a moderate interference effect in both the semantic judgment and the prosodic judgment. In contrast, American respondents showed a strong interference effect of competing semantic information in the prosodic judgment, but no such interference was found in the semantic judgment. Overall, in the prosodic judgment task, the interference effect of competing semantic information was stronger for American than for Japanese, while for the semantic judgment task, the interference effect of competing prosody information was stronger for Japanese than for Americans. Kitayama et al. (2002) therefore argued that the effect of individualism-collectivism contributed to the different findings among Japanese and American participants. Yet, a result opposite to that of the American sample in Kitayama et al's (2002) study was found in a prosodic-semantic interference study (Schirmer & Kotz, in press) done in Germany, which is also regarded as an individualistic country. For the German participants, prosodic judgment was an easier and more automatic task and prosodic information was more likely to influence the semantic judgment than vice versa. The authors (Schirmer & Kotz, in press) suggested that differences in the emotional Stroop paradigm could be the main sources of discrepancy in the two studies' results. First, the speakers used in the two studies may have different salience of emotional expression. This can lead to differences in the difficulty level of the prosodic judgment task. The stronger the salience and clarity of emotional expression by the speaker, the easier the prosodic judgment task is due to increased level of automaticity of the emotional prosodic processing. This will in turn modulate the interaction between semantic and prosodic processing in the Stroop task. Another potential modulating factor is the stimulus category design. As revealed by Schirmer and Kotz's (in press) study, the presence of a neutral response category in addition to the positive and negative category may increase the task difficulty of the semantic judgment. Participants probably need more mental resources to check each word for its emotional strength and compare this to neutral for the decision to be made. As the semantic judgment becomes more difficult, it is more easily influenced by emotional prosodic information. In short, the authors suggested that the underlying principles governing the automaticity of emotional prosody and semantic processing and their interaction are flexible and can be modulated by various factors of the experimental design and participants' characteristics (Schirmer & Kotz, in press).

Study methods

Stroop task

The Stroop task has long been applied to the study of interference between cognitive processing in different dimensions. In the original Stroop task, words were presented as stimuli and participants were required to report the color of the words presented. When the color and word were incongruent, for example, the word GREEN appearing in red color, the responses of color naming by the participants were slower and less accurate than when the color of the stimulus was congruent with the word itself (e.g., the word "Red" in red color), or when the color and word were unrelated (e.g., #### in green). The slower and less accurate response due to conflicting sensory and semantic stimuli is known as the Stroop effect.

Studies on the interaction between emotional prosody and semantic processing have also applied Stroop-like tasks (e.g., Grimshaw, 1998; Kitayama & Ishii, 2002; Schirmer & Kotz, in press). In the study of Grimshaw (1998), the words "mad", "sad" and "glad" were presented in three different emotional tones that either matched or mismatched the semantic emotional valence of those words. Participants were required to respond according to judgment of semantic content or emotional prosody in the stimulus identification task. For the prosody judgment task, participants responded to prosodic-semantic incongruous trials less accurately and more slowly than congruous trials, while there was no difference between congruous and incongruous trials in the semantic judgment task. From these results we can deduce that participants performed more automatically in the semantic task than in the prosody task.

As mentioned in the previous section about Chinese cultural factors, similar results were found by Kitayama and Ishii (2002) for the North American participants they recruited. Yet the results from a German study (Schirmer & Kotz, in press) were contrary to the findings of Grimshaw (1998) and Kitayama and Ishii (2002), in which participants responded faster and more accurately in the semantic task.

<u>ERP Recordings</u>

Event-related potential (ERP) recording has been used to reveal temporal characteristics of human brain activities in numerous neuroscience studies over the past three decades. The ERP refers to the unit of voltage change over time originating from synchronously active neuron populations located mainly in the neocortex. The systematically aligned pyramidal cells over this area become polarized when excited, and the polarity difference causes the flow of extracellular current, and in turn produces electric potentials that can be recorded on human scalp (for detailed explanation please refer to Birbaumer & Schmidt, 1981). Since there is always spontaneous electric activity in the brain which may not be specific to stimulus processing, repeated recording of ERPs to a single event category is needed to obtain an averaged electric signal in the corresponding epochs so as to partial out the systematic event-related activity from the unsystematic spontaneous neuronal activity.

In the following, several ERP components which are commonly found during human cognitive processing are introduced. The P200 is a positive potential that has a latency of 200ms after stimulus onset. It is found to be sensitive to sensory stimulus aspects (Ritter, Simson, & Vaughan, 1983). The P300 is a later positivity that peaks around 500 msec post-stimulus onset. Of particular relevance here is its sensitivity to emotional stimuli. In research on the cognitive processing of emotional pictures (Johnson, Miller, & Burleson, 1986) and emotionally valenced words (Windmann & Kutas, 2001), P300 amplitudes were shown to be larger in response to emotional as compared to neutral stimuli. The N400 is a negativity with a peak latency of about 400 msec. It is commonly regarded as a component reflecting semantic processing. The N400 is found in response to visual as well as auditory word stimuli. The earliest reported N400 was in an experiment investigating the effect of semantic congruency of final words to the whole sentence (Kutas & Hilyard, 1980). Results showed that incongruent final words elicited a larger N400 amplitude as compared to congruent final words. Evaluating the N400 within the framework of the three-stage processing model of word recognition, some researchers suggested that the N400 is generated in early processes such as lexical access (Kutas & Hillyard, 1989; Besson, Fischler, Boaz, & Raney, 1992). However, a study by Chwilla, Brown, & Hagoort (1995) reported an absence of N400 semantic priming effects when the task was letter case discrimination, leading to the conclusion that N400 reflects lexical integration rather than lexical access.

Emotional prosody can be regarded as a type of context in word processing that guides the contextual integration of a word in a way similar to semantics. As emotional prosody is thought to carry primary semantic information, it was proposed to relate to the N400 component. In past studies (e.g., Schirmer & Kotz, in press; Schirmer, Kotz & Friederici, 2002) emotional prosodic modulation of the N400 component was observed. Words that were congruent in valence to the emotional-prosodic context were responded to faster and elicited a smaller N400 amplitude as compared to incongruent words (Schirmer & Kotz, in press; Schirmer, Kotz & Friederici, 2002). In view of this, the N400 is a focus of investigation in the present study of emotional speech perception.

<u>Objective and Research Questions</u>

The present study aims at investigating how emotional prosody interacts with semantic information in Cantonese speech communication. Due to the inconsistent findings from a German sample (Schirmer & Kotz, in press) and an American sample (Kitayama & Ishii, 2002), we would like to explore further the emotional prosodic and semantic processing during speech perception among low-context culture people and tonal language speakers (i.e., Cantonese) in Hong Kong. In the present study, we mainly followed the Stroop paradigm of a German study (Schirmer & Kotz, in press), but included the semantic judgment task only. Neutral stimuli were not included in the present study since it would increase task difficulty particularly in the semantic judgment task

(Schirmer & Kotz, in press) which may lead to a difficulty of comparison between the semantic and prosodic judgment task, as the latter is proposed to be carried out in a follow up study for further comparison. Another difference between Schirmer and Kotz's study and the present study is the use of sad instead of angry stimuli since it was found that sad stimuli were conveyed and perceived more accurately than angry stimuli from the results of a preliminary study (for details, please refer to the method section).

Due to the high-context communication style of Chinese culture and the reliance on tone in word identification by tonal language speakers, there is probably a more automatic prosodic processing system among our tonal language participants compared to previous western samples. It was therefore expected that a stronger interference between emotional prosody and semantic information would be found in the semantic judgment task. The Stroop effect is characterized by longer reaction times and correct rates for incongruous as compared to congruous trials. In the present study, a larger effect size was expected, as compared to the findings from previous western samples.

Another research question was whether the same gender difference of emotional prosody could be found in Chinese as in the previous German study (Schirmer & Kotz, in press). Back to the question about nature versus nurture in language development, if we suppose a larger influence of genetics on the development of emotional prosody processing, we may expect a similar pattern of gender difference in emotional prosody processing among tonal and non-tonal language people. Alternatively, if we expect a larger influence of "nurture" like post-natal language experience, we may look at the difference in the characteristics of tonal and non-tonal language which may lead to a difference in the emotional prosody processing among tonal and non-tonal language speakers. As tonal language speakers rely heavily on prosody for semantic processing of speech, we may expect men speaking a tonal language to have relatively higher level of automaticity in emotional prosody processing as compared to men speaking a non-tonal language. As a result, we may expect a different pattern of gender difference to be observed in the present study compared to that obtained from Schirmer and Kotz's (in press) study, in which men exhibited a smaller interaction between prosody and semantics during a semantic processing stage. In the present study, it was hypothesized that both men and women would not differ significantly in that both groups would respond slower and less accurately in the incongruent trials of the semantic judgment task. This is based on the assumption that both male and female tonal language speakers have higher sensitivity towards emotional prosody and therefore have a more automatic prosody processing system which leads to a stronger prosodic-semantics interference in the semantic judgment.

In addition to the behavioral data of reaction time and accuracy, the ERP data provides temporal information about the interaction between emotional prosody and semantic information during cognitive processing. The N400, which is a negative ERP component sensitive to the incongruency in semantic processing, should act as an indicator of interference between emotional prosodic and semantic information in our study. It was expected to have a larger difference in amplitude between incongruous and congruous trials as compared to previous findings in western samples. Regarding the proposed gender difference, we would expect to see an insignificant difference in N400 amplitude among women and men. In short, we would expect the results from the ERP recording to parallel the expected behavioral results in our study.

Method

Materials

Selection of word stimuli

All the word stimuli used in this experiment were disyllabic verbs that exist both in spoken and written Chinese in Hong Kong. Disyllabic words were chosen because one characteristic of Chinese is that most morphemes or monosyllabic words cannot be uniquely identified from its phonological form (Zhou & Wilson, 1997). Such phonological ambiguity of morphemes and the simplicity of syllabic structure have pushed the language to employ compounds in word formation, because disyllabic compound words are less phonologically ambiguous. This may also be a reason that disyllabic compounds make up a majority of word usage in Chinese (73.6% by type and 34% by token of words, Institute of Language Teaching and Research, 1986). In the present study, one precaution made when selecting word stimuli was the exclusion of any words that have homophones.

In a preliminary rating study, word valence, word familiarity and word type were assessed by 40 judges (Table 1). Judges were asked to rate all words on a 5-point scale ranging from "very negative" (–2) to very positive (+2) for emotional valence. The positive words had a mean valence of 1.277 (S.D.= .186) while negative words had a mean valence of -1.284 (S.D.= .220). Analysis by One-way ANOVA showed that the negative and positive word stimuli selected for use in the present experiment were not significantly different in their absolute valence values [$F(1, 138) = .039$, $p > .05$] (Table 1).

For word familiarity, a six-point-scale ranging from "extremely seldom heard or spoken" (1) to "extremely often heard or spoken" (6) was given to the judges to rate the frequency of encounter in daily conversation and listening. The positive words had a mean frequency of 3.572 (S.D.= .532) while negative words had an absolute mean frequency of

3.459 (S.D.= .299). Again, there was no significant difference found between the words in the positive and negative category [$F(1, 138)= 2.392$, $p> .05$] (Table 1).

A word type scale was constructed to assess whether those words chosen as stimuli were initially perceived as verbs because in Chinese most words can be used as different parts of speech even when they have exactly the same morphological form. Therefore we ensured that all selected stimuli most probably exist as verbs in daily usage. Included words were regarded by 48% or more of the judges to be a verb. No significant difference were found between the mean percentage of positive and negative words [$F(1, 138)= .035$, $p> .05$] (Table 1).

Abstractness of word seems to influence hemispheric specialization in processing (Prior, Cumming, & Hendy, 1984). For example, a number of studies indicate that the left hemisphere is superior in the recognition of abstract words, but not in the recognition of concrete words (e.g. Day, 1977). The right hemisphere is found to be significantly less successful in a competitive situation in dealing with abstract words (e.g. Ellis, & Shepherd, 1974; McFarland, McFarland, Bain, & Ashton, 1978). Moreover, an interaction of sex by ear by concreteness was found, suggesting a larger sex difference for abstract than concrete words, especially for the left ear. There was a tendency for males to be particularly poor with abstract words to the left ear. Furthermore, past ERP studies showed that the N400 is sensitive to the abstract-concrete distinction, with abstract words eliciting smaller amplitudes than concrete words (Holcomb, Kounios, Anderson, & West, 1999; Kounios & Holcomb, 1994). With regard to these findings, the proportion of abstract and concrete verbs was controlled in the present study to be similar in the positive and negative word conditions. We followed the classification criterion of Perani (1999): concrete verbs refer to those related to object manipulation (e.g., to write, to cut) while abstract verbs refer to those related to psychological states (e.g., to think, to hope). There were 19 (27.1%) concrete verbs and 51 (72.9%) abstract verbs out of the 70 positive words, while there were 22 (31.4%) concrete verbs and 48 (68.6%) abstract verbs among the 70 negative words (Table 1). No significant difference was found between the mean abstractness of positive and negative words [$F(1, 138) = .307$, $p > .05$].

Apart from abstractness, semantic transparency of the stimuli is also a potential concern in the present study. Semantic transparency of a compound is defined by the semantic relationship between a compound and its component morphemes (Zwitserlood, 1994). For a fully transparent compound, the meaning of the whole compound is related to the meaning of its composite words. Semantic opacity refers to the situation in which the relationship between the meaning of the whole compound and at least one of its constituents is not apparent (Zwitserlood, 1994). Compounds can be truly opaque and have no semantic relationship with any morphemes, or partially opaque and have a semantic relationship with only one morpheme. As our study involves ERP recording designed to investigate the temporal characteristics of the brain activities during word processing, the factor of time is also important. A behavioral study on the access of meaning from Chinese compound words showed that for opaque compounds, the whole word meanings were accessed earlier than the morpheme meanings (Liu & Peng, 1997). The morpheme meanings were inhibited by the whole word meanings in earlier stage and were not processed until 100ms. Supporting evidence included the lack of a priming effect for opaque words before 86ms, while there was a significant morpheme priming effect at 143ms SOA. This supports Peng et al.'s (1993) claim that whole word and morpheme representations exist, and that a top-down effect determines the whole word and morpheme units, which are two kinds of access units. On the other hand, for transparent compounds, the meaning of the morphemes facilitated whole-word meaning access. Since for transparent words, the morpheme meaning is related to the whole word meaning, this suggests that the whole word and morpheme overlap to some extent at the meaning level. When the activation level of the whole word meaning starts to rise, the morpheme receives facilitative feedback from the meaning level because of the overlap. The morpheme level will in return send more activation to the meaning level. It can be inferred that transparent and opaque words may have different timing in activating the lexical representation, particularly when the first character of the opaque words does not have similar meaning with the whole word. To prevent this timing difference from confounding the results of our present study, we ensured the proportion of transparent plus opaque words with the first character having the same meaning as the whole word was similar in the positive (57.1%) and negative word (55.7%) conditions (Table 1). No significant difference was found between the mean transparency of positive and negative words [$F(1, 138) = .029$, $p > .05$].

Stimuli recording

After the above selection procedures, the stimuli finally consisted of 70 positive and 70 negative verbs (see Table 2a & 2b for the whole list of words in Chinese and translated English). Nine native Cantonese speaking performance art school students with drama experience were given a list of positive and negative words and were asked to produce all words with happy, angry and sad tones and their outputs were recorded. Ten judges were then recruited to determine from five choices – happy, neutral, sad, angry and others- the prosody of the auditory words recorded from the individual performance art students. The student with the highest accuracy rate as shown by the judges' responses, a female, was finally recruited to do the recording of the real stimuli that were used in the study. The selected speaker was asked to produce the 70 positive and 70 negative words in happy and sad tones. Sadly spoken words were selected to be our study's stimuli because sadly spoken words (correct rate= 91.4%) were perceived much more accurately than angrily spoken words (56.4%) by the rating study judges. The positive and negative words were randomized and the speaker was asked to produce the whole set of words first with happy prosody and then with sad prosody. The speaker was advised not to vary the pattern of emotional expression of the stimuli within the same prosody category too much and the loudness of all stimuli was adjusted to be similar later with sound editing software. To make sure each word recorded was in the same acoustic context and to extinguish the influence of the acoustic context by preceding words, the recording of each word began with a carrier phrase "He said....". As the carried phrase was chosen to have a hard consonant ending in Cantonese, it was easily cut using a sound editing software during the preparation of the word stimuli. The words were recorded with a DAT recorder in a sound proof laboratory and were digitized at a 16-bit/ 44.1 KHz sampling rate.

Finally, 40 judges which had not participated in any of the preliminary rating studies described earlier were employed. The list of auditory word stimuli recorded by the female speaker were presented to the judges and they were asked to recite the words and give ratings ranging from 1 (very unhappy) to 5 (very happy) to the voice. The clarity of the stimuli was reflected by the accuracy of the judges' recitation and the valence of emotional prosody was evaluated by the judges' rating. Stimuli that were found to be unclear or too neutral in prosody were re-recorded by the female speaker.

Design

There were 4 stimulus conditions (2 prosody x 2 semantics)- happily spoken positive words, happily spoken negative words, sadly spoken positive words and sadly spoken negative words with 70 experimental trials in each condition. Three breaks separated the experimental trials into four blocks. The 280 experimental trials were arranged by pseudo-randomization so that the same word would not appear in the same half (either the first two blocks or the last two blocks) of the total trials (there was one repetition of each word since each word appeared in both happy and sad prosody). Moreover, not more than four repetitions of the same prosody occurred across consecutive stimuli. There were four sets of such randomization and participants were randomly assigned to one of the four arrangements. In order to familiarize participants with the task, the experimental session started with 12 practice trials, with three trials in each of the four stimulus conditions.

Participants

Forty-eight native Cantonese speakers studying as undergraduate students at the Chinese University of Hong Kong were recruited. They were given a payment of HKD $150 for their participation in the experiment. Twenty four were male and twenty four were female. Male participants had a mean age of 20.7 (S.D.= 1.1) and female participants had a mean age of 20.6 (S.D.= 0.8). All subjects were right-handed, had normal or corrected to normal vision and no hearing impairment.

Procedure

Participants were seated in a chair facing a computer monitor at a distance of 1 meter. They were told that they would hear a number of Cantonese two-character words which they had to judge as positive or negative in semantic meaning while ignoring the happy or sad emotional tone of the stimuli.

During each trial, participants first saw a cross in the middle of the screen and they were asked to fixate their eyes on the cross whenever it appeared. There was a delay of

300 msec between the fixation cross and the onset of the word. The fixation aimed at minimizing the eye movements in the EEG recordings. Participants were asked to place their left index finger on the left button and right index finger on the right button of a response box. They were required to press one of the two response keys as quickly and as accurately as possible during the judgment. To reduce baseline artifacts, the inter-trial intervals were 3000ms, 3100ms and 3200ms for a third of the trials each. If a participant's reaction time was longer than the inter-trial interval, a warning sentence "You were too slow" appeared and the ERP data of this trial was discarded. There was no feedback for incorrect responses, though the ERP data for incorrect responses were also discarded. Half of the participants were instructed to press the left button when they thought the word stimulus was positive and the right button when they thought the word stimulus was negative. The remaining participants had a reversed key-assignment.

The electroencephalogram (EEG) was recorded at a 256 Hz sampling rate from 50 electrodes mounted in an elastic cap (Quickcap, Neuroscan). A digital FIR low pass filter (70 Hz) was applied during EEG acquisition. The reference electrode was placed on the nose tip. In order to control for horizontal and vertical eye movements, a bipolar elecctrooculogram was recorded using 4 electrodes. Electrode resistance was kept below 5 kOhm. Event-related potential (ERP) averages were computed with a 150 msec baseline and a 2000 msec ERP time window. Trials containing eye blinks or movement artifacts were omitted from the data analysis following visual inspection. Grand averages were smoothed with a 10 Hz low pass filter for illustration only.

Statistical Analysis of ERP data

A series of initial ANOVAs were performed to analyze the prosody-semantics interaction for the mean amplitudes in 50 msec time windows from 0 msec to 2000 msec after auditory stimulus onset. To reduce the possibility of committing Type I Error that is associated with the large numbers of comparisons, only effects that reached significance in more than 2 consecutive time windows were regarded as significant. For significant effects, a second ANOVA was performed summarizing significant consecutive time windows into a larger time window. The within subjects variables included PROS (happy/sad), SEM (positive/negative), HEM (left/right), and AP (anterior/posterior), while SEX (male/female) was treated as a between subjects factor. The factor AP has two levels- ERPs measured at anterior and posterior regions of interest (ROIs). Also, the factor HEM has two levels- ERPs measured at right and left hemisphere ROIs. Electrodes were grouped into 4 ROIs: left anterior (F7, F5, F3, FT7, FC5, FC3, T7, C5, C3), right anterior (F8, F6, F4, FT8, FC6, FC4, T8, C6, C4), left posterior (TP7, CP5, CP3, P7, P5, P3, PO7, PO3, O1), and right posterior (TP8, CP6, CP4, P8, P6, P4, PO8, PO4, O2). Five follow-up ANOVAs in the time windows of 150-1050 msec, 750-1400 msec, 1000-1150 msec, 1150-1550 msec and 1300-1700 msec were conducted.

Results

Behavioral data

The group mean of reaction time and accuracy data are presented in Table 3. Reaction time and accuracy were analyzed separately as dependent variables using split plot analysis of variance. PROS (happy/sad) and SEM (positive/negative) were the within subjects factors while SEX (male/female) was a between subjects factor. All significant and marginally significant main effects and interactions are presented in Table 4.

Insert Table 3

Insert Table 4

Reaction time

Analysis of the reaction time data revealed a significant main effect of PROS [$F(1, 45)= 3321.98$, $p< .05$]. The effect of SEM was also significant [$F(1, 45)= 7.01$, $p< .05$]. The significant interaction between PROS and SEM [$F(1, 45)= 23.90$, $p< .05$] showed that for happily spoken words, participants took significantly longer to respond to words with negative meaning than to those with positive meaning ($p< .05$), whereas for sadly spoken words, participants did not differ significantly in reaction time for positive and negative words ($p > .05$) (see Figure 1). A significant interaction of SEM by SEX [$F(1, 45)= 4.73$, $p< .05$] was found. Analysis of SEM separately by SEX showed that men did not differ in reaction time for positive and negative words [$F(1, 22)= .135$, $p> .05$], while women responded significantly slower in response to negative word than to positive words [$F(1, 23)= 10.09$, $p< .05$] (see Figure 2).

Accuracy rate

The analysis of the accuracy rate revealed a significant PROS by SEM interaction [$F(1, 46)= 19.20$, $p< .05$]. When the stimuli were spoken with happy prosody, participants were significantly more accurate in judging positive words than negative words [$F(1, 46)= 24.20$, $p< .05$]; whereas in the sad prosody conditions, participants were more accurate in judging negative words than positive words [$F(1, 46)= 5.29$, $p< .05$] (see Figure 3). In short, the accuracy in response to prosody-semantics congruent conditions was higher than in incongruent conditions. The interaction between SEM and SEX was only marginally significant [$F(1, 46)= 3.50$, $p< .1$] (see Figure 4).

ERP data

The ERP waveforms of women and men are illustrated in Figure 5a and Figure 5b respectively. Three important ERP waveform characteristics were observed only among women. First, the negativity elicited in right hemisphere was larger in a time window from about 150 msec to 1050 msec. Second, within the time window from 750 msec to 1400 msec, the positivity was larger in prosodic-semantic incongruent trials (i.e. happily spoken negative words and sadly spoken positive words) compared to prosodic-semantic congruent trials (i.e. happily spoken positive words and sadly spoken negative words). Moreover, the larger positivity in prosodic-semantic incongruent trials compared to prosodic-semantic congruent trials was limited to happily spoken stimuli (i.e. happily spoken negative words versus happily spoken positive words) in posterior regions in the time window from 1000 msec to 1150 msec.

The summary of all significant and marginally significant effects (p-value not larger than .1) for the 50 msec time window analysis is presented in Table 5, while the significant interactions between prosodic and semantic information of the wider time window secondary post hoc comparisons are shown in Table 6. In the following, only the results of the wider time window analysis will be discussed.

Insert Table 5

Insert Table 6

The earliest time window which contains a significant interaction of HEM by SEX [$F(1, 46)= 7.89$, $p< .01$] is between 150 msec and 1050 msec. Separate analyses for each level of SEX showed that women show a significant effect of HEM [$F(1, 23)= 24.88$, $p< .0001$] while men do not [$F< 1$]. This means significantly larger negativity was elicited in right hemisphere for women in response to stimuli, while there was no significant difference between the left and right hemisphere ERP amplitude for men (Figure 5a & 5b).

There is a significant PROS by SEM by SEX interaction between 750 msec and 1400 msec [$F(1, 46)= 9.47$, $p< .01$]. Analysis of the PROS by SEM interaction separately for each level of SEX shows that the effect is significant for women only [F(1, 23)= 7.28, p< .05]. For women, the SEM effects were just marginally significant in both happy [F(1, 23)= 3.90, p= .0603] and sad [F(1, 23)= 3.69, p= .0671] prosody conditions. When listening to happily spoken words, women showed a larger positivity towards negative words, which are prosody-semantics incongruent, than positive words. When listening to sadly spoken words women showed a larger positivity towards positive words, which are also prosody-semantics incongruent, than negative words. In short, women showed larger positivity towards prosodic-semantic incongruent stimuli compared to congruent stimuli between 750 and 1400 msec (Figure 5a & 5b).

A significant PROS by SEM by AP by SEX interaction occurred between 1000 msec and 1150 msec [F(1, 46)= 6.36, p< .05]. Separate analyses for each level of SEX revealed that women, but not men showed a significant PROS by SEM by AP interaction [F(1, 23)= 21.04, p< .001]. Further analysis on women showed that the significant PROS by SEM interaction only occurred in the posterior region [F(1, 23)= 10.07, p< .01]. When women listened to happily spoken words, there was a larger ERP positivity [F(1, 23)= 9.78, p< .01] in the posterior regions when the words presented were semantically incongruent to the prosody (i.e., negative words with happy prosody) compared to semantically congruent words (i.e., positive words with happy prosody) (Figure 5a & 5b). When they listened to sad prosody words, there was no significant difference of posterior activity between positive and negative words [F(1, 23)= 2.61, p= .12] (Figure 5a & 5b).

A significant PROS by SEM by AP interaction occurred between 1150 msec and 1550 msec [F(1, 46)= 11.20, p< .01]. Further post-hoc comparison showed that the PROS by SEM interactions were not significant in either anterior or posterior regions, but the F-value was farther away from significance in anterior [F< 1] than in posterior region [F(1, 46)= 1.57, p= .22].

A significant PROS by SEM by HEM interaction was obtained in the late time window of 1300 msec to 1700 msec [F(1, 46)= 8.92, p< .01]. The PRO by SEM interaction was not significant for either the left or right hemisphere, though the effect in the right hemisphere was farther away from significance [F < 1] than in the left hemisphere [F(1, 46)= 1.53, p= .22]. All significant post hoc effects are summarized in Table 6.

Discussion

The present study aimed at examining emotional speech processing in Cantonese using a Stroop task. Behaviorally, the current findings generally showed a similar emotional Stroop effect as in past studies (e.g., Schirmer & Kotz, in press), as revealed by the significant interaction between emotional prosody and semantics for both reaction time and accuracy. Participants responded more slowly and less accurately to words with negative meaning than with positive meaning for happily spoken words. However, for sadly spoken words, participants did not differ much in reaction time for positive and negative words. It is possible that the effect of prosody on reaction time may have been confounded by stimulus word length. Sadly spoken words (969 msec) used in the present study on average last 357 msec longer than the happily spoken words (612 msec), and a past study (Schirmer & Kotz, in press) showed that the difference in word length modulated response latencies especially for the semantic task.

Another interesting behavioral finding is the sex difference in response times to negative and positive words. Follow-up analysis of the interaction between semantics and sex showed that the main source of difference in reaction time to positive and negative words came from women. It is intriguing to ask why women took much longer time to respond to negative words than to positive words, and why this lengthened latency of response was not found in men. Generally, non-depressed people have a slower response to negative stimuli (Pratto & John, 1991). Moreover, a study on negative affect processing in females showed that in response to lateralized negative visual stimuli, there was a selective lengthening of reaction time mediated by the right hemisphere during production of negative affect (Ladavas, Nicoletti, Umilta, & Rizzolatti, 1984). This suggests that the right hemisphere plays an important role in mediating the neural state subjectively referred to as sadness, and that the feeling of sadness is strongly linked with the motor organization that expresses this emotion. In the present study, a significantly larger negativity was elicited in right hemisphere among women in response to stimuli, whereas there was no significant difference between the left and right hemisphere ERP amplitude among men. As a result, the response selection and the production of the key press response may be mediated by the right hemisphere activity that leads to longer reaction time for women in response to negative words. It is notable that a semantics by sex interaction for response time was not reported in the German study (Schirmer & Kotz, in press) we tried to replicate. Apart from the difference between tonal and non-tonal language, this discrepancy could also be due to sad prosody stimuli used in the present study rather than angry prosody stimuli in the German study.

The finding of a gender difference specifically for negative stimuli was different from the prediction we made about gender differences in prosody processing in speakers of tonal languages. Instead of a narrowed gender difference in prosodic-semantic interaction during emotional prosody processing, male and female speakers of tonal languages seem to follow different processing mechanisms in processing emotional speech. Behaviorally, we found that women showed differential response to positive and negative words. Electrophysiologically, we can see that during the 150 to1050 msec epoch, women had larger ERP negativity in the right hemisphere as compared to the left hemisphere, while men did not show significant difference between the activity of left and right hemisphere. This suggests that in processing emotional speech, female speakers of a tone language rely more heavily on the right hemisphere while males are more bilateral. This is in contrast to findings of previous fMRI (Shaywitz et al.,1995), dichotic listening (Hiscock et al., 1994), and tachistoscopic studies (Hiscock et al., 1994), in which female non-tonal language speakers were shown to be more bilateral while male non-tonal language speakers are more left lateralized in language processing. The discrepancy between the current findings and those past studies could be due to the difference between tonal and non-tonal language, but can also be attributed to the focus of past studies on linguistic processing whereas the present study was focused on the emotional component of language processing.

The N400 component

In the present study there was a large and broad negativity in the 300 to 1500 msec epoch instead of the N400 normally elicited in semantic judgment tasks and as observed in the Schirmer and Kotz (in press) study. As a proposition, we may interpret the broad negativity in the 300 to 1500 msec time window as a broadening of the conventional N400. In most word and non-word visual stimuli studies, the N400 usually lasts only for 200 to 400 msec. However, it has been proposed that in certain situations, the time frame of the N400 may be modified. For example, a broadening of the N400 component was observed in a study of semantic integration using video-taped real-world event stimuli (Sitnikova, Kuperberg, & Holcomb, 2003). In that study, the duration of the N400 was greater than 600 msec at some electrodes, and the author attributed the prolonged time-course to variability in the timing of identification of different stimuli. If the timing of identification of stimuli varies from trial to trial, the N400 could be spread by the averaging process across trials and result in a dispersion over a broader range of time. This explanation for broad N400 effects was also proposed in some studies of spoken words (e.g., Holcomb & Neville, 1991) and American Sign Language videos (Neville et al., 1997) which also had variable stimulus identification time points. In the present study, one factor that was difficult to control for was the time point of the identification of the word's semantic meaning. Since the stimuli we used were disyllabic two-character words, the time point of identification could vary according to the semantic transparency of the word. For semantically transparent words, the meaning of the whole word is intimately related to its composite characters and therefore participants may have obtained enough semantic information for word identification at the end of the first syllable. However, if the word was semantically opaque, the participants may not have been able to identify the word until the end of the second syllable. Moreover, the duration of stimuli varied a lot. As a result, it is suspected that the broadening of N400 in the present study may be caused by the effect of multiple points of identification as suggested by previous studies. However, an N400 effect of congruency was not found, and the results will be discussed in the following paragraphs.

Late positivity component (LPC)

Another extraordinary finding in the present study is a late positivity component (LPC) present in the 750 msec to 1400 msec time window. A late positivity component was observed in a previous study (Sitnikova, Kuperberg, & Holcomb, 2003) which investigated the ERP differences between contextually appropriate and inappropriate objects appearing in video film clips of common activities. In that study, between 600-900-msec there were more positive potentials at the posterior electrodes in incongruent than in congruent conditions. The authors pointed out that in language comprehension studies the P600, which is a late positivity with a latency of around 600 msec, reflects a variety of syntactic processing difficulties (e.g., Osterhout & Holcomb, 1992), semantic violations, and processes of reanalysis triggered by error detection (e.g., Münte, Heinze, Matzke, Wieringa, & Johannes, 1998). The effect of incongruency on the P600 was also reported in a study of structural violations in music (Patel, Gibson, Ratner, Besson, &

Holcomb, 1998), which was found to be closely related to emotional prosody processing in many aspects (Besson, 1998; Palmer, Jungers, & Jusczyk, 2001; Patel & Daniele, 2003).

There was a gender difference observed in the LPC in the present study. Only women had a significantly larger amplitude in response to incongruent stimuli as compared to congruent stimuli, which reveals that women may have higher sensitivity to an incongruency between semantic and prosodic emotional information. It may also explain why women have a tendentially higher correct rate than men in all prosodic and semantic conditions (see Table 3). Besides, women also showed more differentiation of ERP responses in different stimulus conditions. In the epoch of 1000 to 1150 msec, women had a larger positivity in posterior regions in response to incongruent stimuli as compared to congruent stimuli. But that differentiation of neural activity to congruent and incongruent stimuli only occurred for happily spoken stimuli. For sadly spoken stimuli, there was no observed ERP difference between congruent and incongruent stimuli. It seems that in the time frame of 1000 to 1150 msec, women show word valence differentiation in ERP only for happily spoken words. Given we found a similar correct rate between happily and sadly spoken stimuli for women, women probably have similar neural sensitivity to happy and sad stimuli for recognition while the time course for processing happy and sad stimuli may differ.

<u>Location of interference in STROOP</u>

Whenever there are two or more sensory signals that are contradictory, interference between these signals will develop at certain point during the processing, leading to the degradation of processing outcome like lengthened reaction time and lower accuracy. Some researchers suggested the perceptual stage (Hock & Egeth, 1970), while others proposed the stimulus processing stage to be the starting point of the interference

(Seymour, 1977). There were also a number of researchers who claimed that the parallel activation of correct and incorrect response in the response preparation output stage was when the interference occurred (Duncan-Johnson & Kopell, 1981; DeSoto, Fabiani, Geary, & Gratton, 2001). In the following paragraph we try to infer the interference point between contradicting semantic and prosodic information in a Stroop-like paradigm using Cantonese speech stimuli.

The absence of N400 effect and presence of LPC

Before looking into the investigation of the interference point, we may first evaluate certain particularities in our present findings. One significant contradiction between our expected results and the present findings is the absence of a N400 amplitude difference between congruent and incongruent conditions. In the earlier German study, a significantly larger N400 negativity was found for emotionally incongruent stimuli compared to congruent stimuli for semantic judgment task among women (Schirmer & Kotz, in press). In the present study, there is no significant difference of N400 amplitude in response to emotionally congruent and incongruent stimuli in either men or women, while there was a gender difference in the late positivity found in the time window of 1500 to 2000 msec, as only women show significantly larger late positivity amplitude towards incongruent stimuli compared to congruent stimuli. Comparing and contrasting the electrophysiological results from Schirmer and Kotz's study and the present study, we may suggest that the interference between emotional prosody and semantics happens at a later processing stage among Cantonese speakers as compared to German speakers. For the semantic judgment task, participants should place much of their attention to the lexical judgment while trying to ignore the emotional prosodic context. Since tone is such a crucial source of information lexical decision among tonal language speakers, much of the mental resources may be placed on the processing of lexical tone information at an early stage. This could lead to a delay of emotional prosody information processing and therefore, the interference effect between semantic and emotional prosody information can occur at a later stage, which is revealed by the late positivity difference observed in a later time window. Moreover, as the response latency coincides with the late positivity, we could infer that the interference of the semantic and emotional prosodic information may lie within the response preparation output stage.

Conclusion

To conclude, behaviorally we observed a significant interaction between prosodic and semantic information processing in Cantonese similar to that found in past studies of non-tonal language, though electrophysiologically there exist differences between the present data and past findings in terms of the processing time courses and ERP components. In fact, the uniqueness and commonness of the interaction between emotional prosody and semantic processing systems in tonal language speakers as compared to non-tonal language speakers may lead us to query how the factor of nature and nurture interact to produce the pattern of human behavior across different cultures. We may argue a rather genetically prescribed and innate language mechanism for processing emotional prosody that contributes to the similar pattern of gender difference in processing emotional prosody among both tonal and non-tonal language speakers. Also, we may see variation in the timing of prosodic and semantic interference among tonal and non-tonal language speakers and that can be explained by the effect of post-natal learning experience under tonal and non-tonal language environment. Further studies are needed to reveal the underlying picture of language processing among tonal and non-tonal language speakers.

References

Argyle, M., Henderson, M., Bond, M., & Iizuka, Y. (1986). Cross-cultural variations in relationship rules. *International Journal of Psychology, 21(3)*, 287-315

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70(3),* 614-636.

Barrett, A. M., Crucian, G. P., Raymer, A. M., & Heliman, K. (1999). MTI: Spared comprehension of emotional prosody in a patient global aphasia. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 12(2),* 117-120.

Barnlund, D. C. (1989). *Communicative styles of Japanese and Americans: Images and realities.* Belmont, CA: Wadsworth.

Besson, M. (1998). Meaning, structure, and time in language and music. *Current Psychology of Cognition, 17(4-5)*, 921-950.

Besson, M., Fischler, I., Boaz, T., & Raney, G. (1992). Effects of automatic associative activation on explicit and implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(1),* 89.

Blonder, L. X., Burns, A. F., Bowers, D., & Moore, R. W. (1993). Right hemisphere facial expressivity during natural conversation. *Brain and Cognition, 21(1)*, 44-56.

Birbaumer, N. (1981). EEG and slow cortical potentials in anticipation of mental tasks with different hemispheric involvement. *Biological Psychology, 13*, 251-260.

Breitenstein, C., Daum, I., & Ackermann, H. (1998). Emotional processing following cortical and subcortical brain damage: Construction of the fronto-striatal circuitry. *Behavioral Neurology, 11*, 29-42.

Brennan, P. K., Goddard, M., Wilson, A. E., & Kinnear, L. (1991). Sex differences in smiling as measured in a picture taking task. *Sex Role, 24(5-6),* 375-382.

Bolinger, D. (1972). Accent is predictable (If you're a mind-reader). *Language, 48(93)*, 633-644.

Smith, P. B., & Bond, M. H. (1993). *Social psychology across cultures: Analysis and perspectives.* England: Harvester Wheat sheaf.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., Jaencke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research, 9(3),* 227-238.

Chang, T. N. C. (1958). Tones and intonation in the Chengtu dialect, *Phonetica, 2,* 60-84.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology, 32,* 274-285.

Connell, B. A., Hogan, J. T., & Rozsypal, A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics, 11*, 337-351.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26(5),* 527-538.

Crystal, D. (1976). *The English tone of voice*. New York: St Martins Press.

Day, J. (1977). Right hemisphere language processing in normal right handers. Journal of Experimental Psychology. *Human Perception and Performance, 3(3),* 518-528.

DeSoto, M. C., Fabiani, M., Geary, D. C., & Gratton, G. (2001). When in doubt, do it both ways: Brain evidence of the simultaneous activation of conflicting motor responses in a spatial Stroop task. *Journal of Cognitive Neuroscience, 13(4),* 523-536.

Duncan, J. C. C., & Kopell, B. S. (1981). The Stroop effect: Brain potentials localize the source of interference. *Science, 214(4523),* 938-940.

Ellis, H., & Shephard, J. (1974). Recognition of abstract and concrete words presented in left and right visual fields. *Journal of Experimental Psychology, 103,* 1035-1036.

George, M. S., Parekh, p. I., Rosinsky, N., Ketter, T., Kimbrell, T. A., & Heilman, K. H. (1996). Understanding emotional prosody activates right hemisphere regions. *Archives of Neurology, 53*, 665-670.

Grimshaw, G. M. (1998). Integration and interference in the cerebral hemispheres: Relations with hemispheric specialization. *Brain and Cognition, 36(2),* 108-127.

Gudykunst, W. B., Matsumoto, Y., Ting-Toomey, S., Nishida, T., Kim, K., & Heyman, S. (1996). The influence of cultural individualism-collectivism, self-construals, and individual values on communication styles across cultures. *Human Communication Research, 22,* 510-543.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38(11),* 1531-1549.

Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. *Perspectives on sentence processing. (pp. 45-80).* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85,* 845-857.

Hashimoto, A. (1972). *Studies in the Yue dialect.* Cambridge: C.U.P.

Heilman, K. M., Blonder, L. X., Bowers, D., & Valenstein, E. (2003). Emotional disorders associated with neurological diseases. In Heilman, Kenneth M. & Valenstein, Edward (Ed). *Clinical neuropsychology (4th ed.).* (pp. 447-478). London, Oxford University Press.

Hess, U, Senecal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R. E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cognition and Emotion, 14(5)*, 609-642.

Hiscock, M., Inch, R., Jacek, Hiscock-Kalil, C. & Kalil, K. M. (1994). Is there a sex difference in human laterality? I. An exhaustive survey of auditory laterality studies from six neuropsychological journal. *Journal of Clinical Experimental Neuropsychology, 16,* 423-435.

Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology, 83(2, Pt. 1),* 299-303.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual coding context-availability and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, memory and cognition, 25(3),* 721-742.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology, 19(4),* 286-300.

Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioral changes following ventral frontal lobe damage. *Neuropsychologia, 34*, 247-261.

Hubbard, P. (2001). Sex zones: Intimacy, citizenship and public space. *Sexualities, 4(1),* 51-71.

Institute of Language Teaching and Research (1986). *A frequency dictionary of Modern*

Chinese. Beijing: Beijing Language Institute Press.

Kim, H. S. (1994). Korean collective self-esteem scale. *Korean Journal of Social Psychology, 8(1)*, 103-116.

Kitayama, S. (2000). Cultural variations in cognition: implications for aging research. In P. C. Stern & L. L. Cartensen (Eds), *The aging Mind: The Opportunities in Cognitive Research* (pp. 218-237). Washington, DC: National Academy Press.

Kitayama, S., & Ishii, K. (2002). Word and voice: spontaneous attention to emotional utterances in two languages. *Cognition and Emotion, 16(1),* 29-59.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20(4),* 804-823.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Ioup, G., & Tansomboon, A. (1987). The acquisition of tone: A maturational perspective. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 333-349). New York: Newbury House/ Harper & Row.

Ladavas, E., Nicoletti, R., Umilta, C., & Rizzolatti, G. (1984). Right hemisphere interference during negative affect: A reaction time study. *Neuropsychologia, 22(4)*, 479-485.

Lee, Y. S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. *Journal of Psycholinguistic Research, 125,* 527-542.

Liu, Y. & Peng, D. L. (1997). Meaning access of Chinese compounds and its time course.

In *Cognitive Processing of Chinese and Related Asian Languages.* Hong Kong: Chinese University Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25(1-2),* 71-102.

Matsumoto, D., & Kishimoto, H. (1983). Developmental characteristics in judgments of emotion from nonverbal vocal cues. *International Journal of Intercultural Relations, 7(4),* 415-424.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18(1),* 1-86.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88(5),* 375-407.

McFarland, K., McFarland, M., Bain, J., & Ashton, R. (1978). Ear differences of abstract and concrete word recognition. *Neuropsychologia, 16*, 555-561.

Morris, J. S., Scott, S. K., & Dolan, R. J. (1999). Saying it with feeling: Neural responses to emotional vocalizations. *Neuropsychologia, 37*, 1155-1163.

Muente, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36(3),* 217-226.

Neville, H. J., Coffey, S. A., Lawson, D. S., & Fischer, A. (1997). Neural systems mediating American Sign Language: Effects of sensory experience and age of acquisition. *Brain and Language, 57(3),* 285-308

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review, 108,* 291-310.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Oxford, England: University of Illinois Press.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31(6),* 785-806.

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45(4),* 526-545.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87(1),* B35-B45.

Patel, A. D., Gibson, E., Ratner, J., Besson, M, & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10(6),* 717-733.

Pell, M. D., & Baum, S. R. (1997). The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. *Brain and Language, 57,* 80-99.

Perani, D., Cappa, S. F., Schnur, T, Tettamanti, M., Collina, S., Rosa, M. M., & Fazio, F. (1999). The neural correlates of verb and noun processing A PET study. *Brain, 122,* 2337-2344.

Pihan, H., Altenmueller, E., & Ackermann, H. (1997). The cortical processing of perceived emotion: A DC-potential study on affective speech prosody. *Neuroreport: An International Journal for the Rapid Communication of Research In Neuroscience, 18(3),* 623-627.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61(3),* 380-391.

Prior, M. R., Cumming, G., & Hendy, J. (1984). Recognition of abstract and concrete words in a dichotic listening paradigm. *Cortex, 20(1),* 149-157.

Ritter, W., Simson, R., & Vaughan, H. G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination trials. *Psychophysiology, 20(2)*, 168-179.

Ross, E. D., Edmondson, J. A., & Seibert, G. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics, 14,* 283-302.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior, 12(2),* 139-148.

Seymour, P. H. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology, 29(2),* 245-265.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. A., Fletcher, J. M., Shankwiler, D. P., Katz, L., & Gore, J. C. (1995). Sex differences in the functional organization of the brain for language. *Nature, 373,* 607-609.

Scherer, K. R. (1989). The essence of emotions: Semantic constituents of colloquial speech or multicomponent process patterns? Reply to T. Boll and R. Neppl's essay, "Understanding and goals of cognitive approaches to the analysis of specific emotions: Comments on K. R. Scherer's component-process model.", *Sprache and Kognition, 8(1),* 19-22.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross Cultural Psychology, 32(1),* 76-92.

Schirmer, A., & Kotz, A. (in press). ERP evidence for a gender-specific Stroop effect in emotional speech.

Schirmer, A., Kotz, A. S., & Friederici. A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14,* 228-233.

Schmitt, J. J., Hartje, W., & Willmes, K. (1997). Hemispheric asymmetry in the recognition of emotional attitude conveyed by facial expression, prosody and propositional speech. *Cortex, 33(1)*, 65-81.

Sitnikova, T., Salisbury, D. F., Kuperberg, G., & Holcomb, P. J. (2002). Electrophysiological insights into language processing in schizophrenia. *Psychophysiology, 39(6),* 851-860.

So, L. K. H., Dodd, B. J. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language, 22(3)*, 473-495.

Soto-Faraco, S., Sebastian-Galles, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45(3),* 412-432.

Starkstein, S. E., Federoff, J. P., Price, T. R., & Leiguarda, R. C. (1994). Neuropsychological and neuroradiologic correlates of emotional prosody comprehension. *Neurology, 44(3)*, 515-522.

Windmann, S., & Kutas, M. (2001). Electrophysiological correlates of emotion-induced recognition bias. *Journal of Cognitive Neuroscience, 13(5),* 577-592.

Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: Review, replication, and reanalysis. *Cerebral Cortex, 6(1),* 21-30.

Zhou, X., & Wilson, W. M. (1997). The abstractness of phonological representation in the Chinese mental lexicon. In *Cognitive Processing of Chinese and Related Asian Languages*. Hong Kong: Chinese University Press.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes, 9(3),* 341-368.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32(1),* 25-64.

Table 1. Descriptive data and significance values of the word valence, word familiarity, word type, abstractness and semantic transparency

	Word valence	Word familiarity	Word type
Positive stimuli (mean)	1.277	3.572	.758
Negative stimuli (mean)	-1.284	3.459	.762
F-value of the comparison between positive and negative stimuli	.039 (compared between absolute means) ($p > .05$)	2.392 ($p > .05$)	.035 ($p > .05$)

	Abstractness	Semantic transparency
Positive stimuli (frequency)	Concrete= 19(27.1%) Abstract= 51(72.9%)	Opaque= 30(42.9%) Transparent= 40(57.1%)
Negative stimuli (frequency)	Concrete= 22(31.4%) Abstract= 48(68.6%)	Opaque= 31(44.3%) Transparent= 39(55.7%)
Positive stimuli (mean)	1.729	1.571
Negative stimuli (mean)	1.686	1.557
F-value of the comparison between positive and negative stimuli	.307 ($p > .05$)	.029 ($p > .05$)

Table 2a. List of selected positive stimuli in Chinese and translated English

1	分享	Share
2	崇敬	Adore
3	欽佩	admire
4	守護	Guard
5	安慰	comfort
6	吸引	attract
7	忍耐	put up with
8	改善	improve
9	改進	improve
10	享有	benefit/ own enjoyably
11	享受	Enjoy
12	愛惜	cherish
13	欣賞	admire
14	表揚	commend
15	促進	promote
16	勉勵	encourage
17	珍惜	treasure
18	盼望	to long for
19	美化	embellish
20	恭候	Await
21	恭喜	congratulate
22	振作	Cheer
23	晉升	promote
24	接納	accept
25	推崇	praise
26	啓發	enlighten
27	喜愛	Like
28	喜歡	please
29	報恩	requite
30	報答	recompense
31	尊重	respect
32	尊敬	esteem
33	復活	resurrect
34	復興	revive
35	提升	elevate
36	提拔	promote
37	款待	entertain
38	發展	develop
39	發揮	elaborate
40	發揚	enhance
41	發奮	exert one's energy
42	答謝	appreciate for kindness
43	傳頌	spread
44	微笑	smile
45	感謝	move
46	愛戴	respect & support
47	愛護	care
48	道謝	to say thanks
49	達成	reach
50	愛慕	admire
51	鼓勵	encourage
52	暢遊	tour pleasantly
53	歌頌	extol
54	稱頌	extol
55	稱讚	eulogize
56	增強	reinforce
57	增進	enhance
58	慶祝	celebrate
59	熱愛	love fervently
60	賞識	recognize one's worth
61	奮鬥	exert
62	戰勝	defeat
63	激勵	inspire
64	優待	to give priority
65	幫助	help
66	讚揚	praise
67	讚賞	compliment
68	歡迎	welcome
69	敬重	venerate
70	變通	become flexible

Table 2b. List of selected negative stimuli in Chinese and translated English

71	勾引	entice
72	出賣	betray
73	包庇	shield
74	犯案	commit a crime
75	犯規	break rule
76	犯錯	violate
77	危害	endanger
78	妄想	irrational attempt
79	作弊	cheat (in exam)
80	妒忌	envy
81	折磨	torture
82	攻擊	attack
83	挨餓	endure hunger
84	抱怨	Blame
85	爭吵	squabble
86	非禮	molest sexually
87	侵犯	infringe
88	侵略	invade
89	侮辱	insult
90	威脅	threaten
91	挑剔	pick fault with
92	歪曲	falsify
93	虐待	ill-treat
94	要脅	threaten
95	迫害	persecute
96	倒閉	go bankrupt
97	剝削	chisel off
98	剝奪	deprive
99	埋怨	blame
100	恐嚇	threaten
101	浪費	waste
102	破壞	destroy
103	逃學	truancy
104	逃避	escape
105	強姦	rape
106	得罪	offend
107	勒索	extort
108	患病	fall ill
109	捱苦	suffer
110	處分	punish
111	處罰	punish
112	陷害	design to kill
113	喪失	lose
114	欺騙	cheat
115	滋擾	disturb
116	開除	dismiss
117	傳染	inflect
118	傷害	hurt
119	嫌棄	abandon
120	損害	damage
121	綁架	kidnap
122	解僱	dismiss
123	賄賂	bribe
124	過身	pass away
125	煽動	instigate
126	誘惑	tempt
127	敷衍	do thing perfunctorily
128	誹謗	libel
129	操縱	control
130	擔憂	be worried
131	諷刺	satirize
132	濫用	abuse
133	糟蹋	spoil
134	擺佈	manipulate
135	鎮壓	suppress
136	偽造	counterfeit
137	觸犯	violate
138	襲擊	raid
139	剝削	chisel off
140	剝奪	deprive

Table 3. Descriptive data of reaction time and correct rate

Reaction time				Correct rate			
Condition	Sex	Mean	S.D.	Condition	Sex	Mean	S.D.
Happy-	Male	1087.4085	88.1162	Happy-	Male	97.3833	2.5077
positive	Female	1089.7088	76.2401	positive	Female	97.6208	2.4456
	Total	1088.5832	81.3699		Total	97.5021	2.4533
Happy-	Male	1115.4517	90.8935	Happy-	Male	93.3833	4.1412
negative	Female	1135.9470	66.3749	negative	Female	95.9500	2.9067
	Total	1125.9174	79.1282		Total	94.6667	3.7694
Sad-	Male	1365.9721	87.4848	Sad-	Male	95.3500	5.7793
positive	Female	1359.6464	75.1955	positive	Female	95.7708	3.8304
	Total	1362.7420	80.6090		Total	95.5604	4.8549
Sad-	Male	1343.8541	88.6975	Sad-	Male	96.6625	2.7517
negative	Female	1373.6296	74.0511	negative	Female	97.7958	2.4574
	Total	1359.0586	82.0411		Total	97.2292	2.6436

Table 4. Repeated measured ANOVA of reaction time and correct rate: Significant and marginally significant main effects and interactions

Reaction time			Correct rate		
Source	df	F-value	Source	df	F-value
PROS	1, 45	3321.98***	SEM*SEX	1, 46	3.50#
SEM	1, 45	7.01*	PROS*SEM	1, 46	19.20***
			Happy: Pos vs Neg	1, 47	22.71***
			Sad: Pos vs Neg	1, 47	5.37*
SEM*SEX	1, 45	4.73*			
PROS*SEM	1, 45	23.90***			
Happy: Pos vs Neg	1, 46	40.07**			

*** $p < .001$; ** $p < .01$; * $p < .05$; # $p < .1$

Table 5. ERP 50 msec time windows with significant effects#

<table>
<tr><th></th><th>0</th><th>50</th><th>100</th><th>150</th><th>200</th><th>250</th><th>300</th><th>350</th><th>400</th><th>450</th><th>500</th><th>550</th><th>600</th><th>650</th><th>700</th><th>750</th><th>800</th><th>850</th><th>900</th><th>950</th><th>1000</th><th>1050</th><th>1100</th><th>1150</th><th>1200</th><th>1250</th><th>1300</th><th>1350</th><th>1400</th><th>1450</th><th>1500</th><th>1550</th><th>1600</th><th>1650</th><th>1700</th><th>1750</th><th>1800</th><th>1850</th><th>1900</th><th>1950</th></tr>
<tr><td>LR*SEX</td><td colspan="3"></td><td colspan="18">7.89 **</td><td colspan="19"></td></tr>
<tr><td>PROS*SEM*SEX</td><td colspan="15"></td><td colspan="13">9.47**</td><td colspan="12"></td></tr>
<tr><td>PROS*SEM*AP*SEX</td><td colspan="20"></td><td colspan="3">6.36*</td><td colspan="17"></td></tr>
<tr><td>PROS*SEM*AP</td><td colspan="23"></td><td colspan="8">11.20**</td><td colspan="9"></td></tr>
<tr><td>PROS*SEM*LR</td><td colspan="26"></td><td colspan="8">8.92**</td><td colspan="6"></td></tr>
</table>

*** p< .001; ** p< .01; * p< .05

For all the effects shown, df = 1, 46

Table 6. Post hoc comparisons for the interactions in the ERP data analysis

1) LR*SEX (Time window: 150..1050 msec)

Effect	df	F-value
Men: LR	1, 23	<1
Women: LR	1, 23	24.88****

2) PROS*SEM*SEX (Time window: 750..1400 msec)

Effect	Df	F-value
Men: PROS*SEM	1, 23	2.86
Women: PROS*SEM	1, 23	7.28*
Women: Happy PROS: SEM	1, 23	3.90#
Women: Sad PROS: SEM	1, 23	3.69#

3) PROS*SEM*AP*SEX (Time window: 1000..1150 msec)

Effect	Df	F-value
Men: PROS*SEM*AP	1, 23	1.13
Women: PROS*SEM*AP	1, 23	21.04***
Women: Anterior: PROS*SEM	1, 23	1.42
Women: Posterior: PROS*SEM	1, 23	10.07**
Women: Posterior: Happy PROS*SEM	1, 23	9.78**
Women: Posterior: Sad PROS*SEM	1, 23	2.61

**** p< .0001; *** p< .001; ** p< .01; * p< .05; # p< .1

Figure Captions

Figure 1. Reaction time in response to positive and negative words spoken with happy and sad prosody.

Figure 2. Reaction time of women and men in response to positive and negative words.

Figure 3. Correct rate in response to positive and negative words spoken in happy and sad prosody.

Figure 4. Correct rate of women and men in response to positive and negative words.

Figure 5a. The ERP waveforms of women in response to positive and negative words spoken with happy and sad prosody at electrodes F5, C5, P5, F6, C6, and P6.

Figure 5b. The ERP waveforms of men in response to positive and negative words spoken with happy and sad prosody at electrodes F5, C5, P5, F6, C6, and P6.

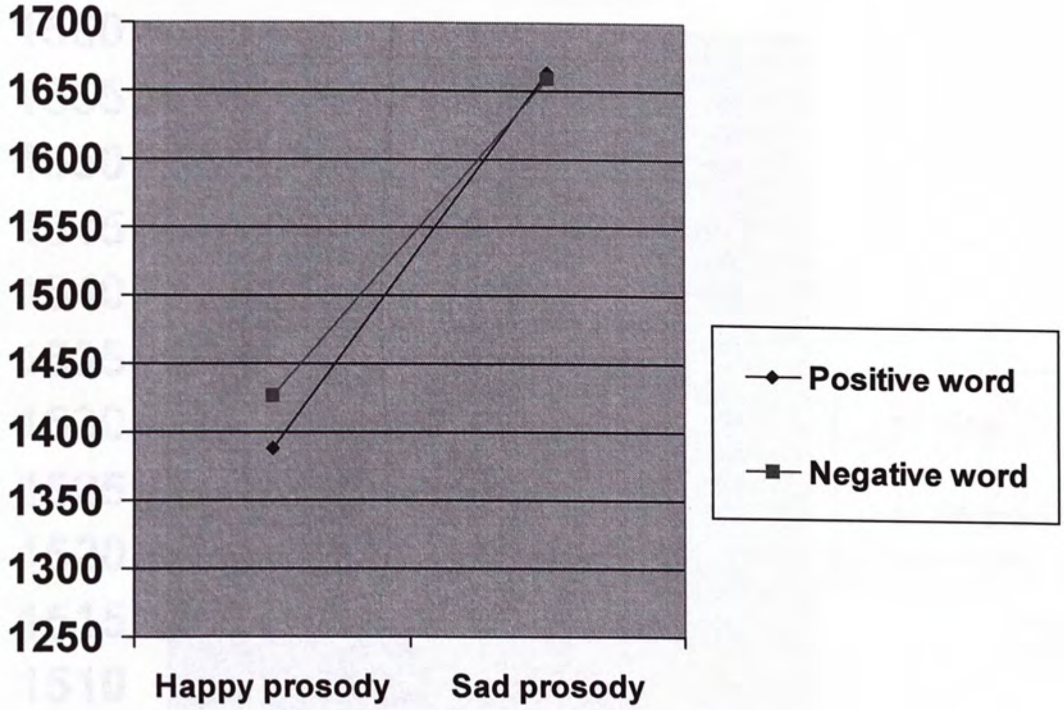
1700
1650
1600
1550
1500
1450
1400
1350
1300
1250
Happy prosody
Sad prosody
Positive word
Negative word

Figure 1.

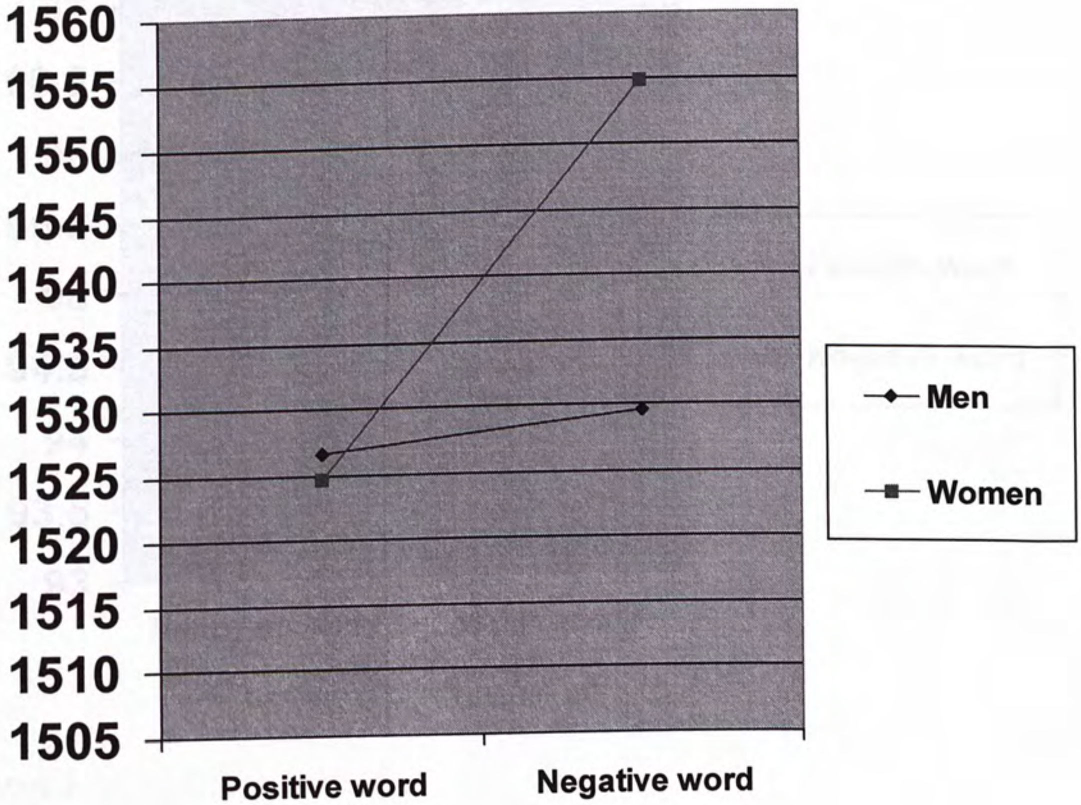
1560
1555
1550
1545
1540
1535
1530
1525
1520
1515
1510
1505
Positive word
Negative word
Men
Women

Figure 2.

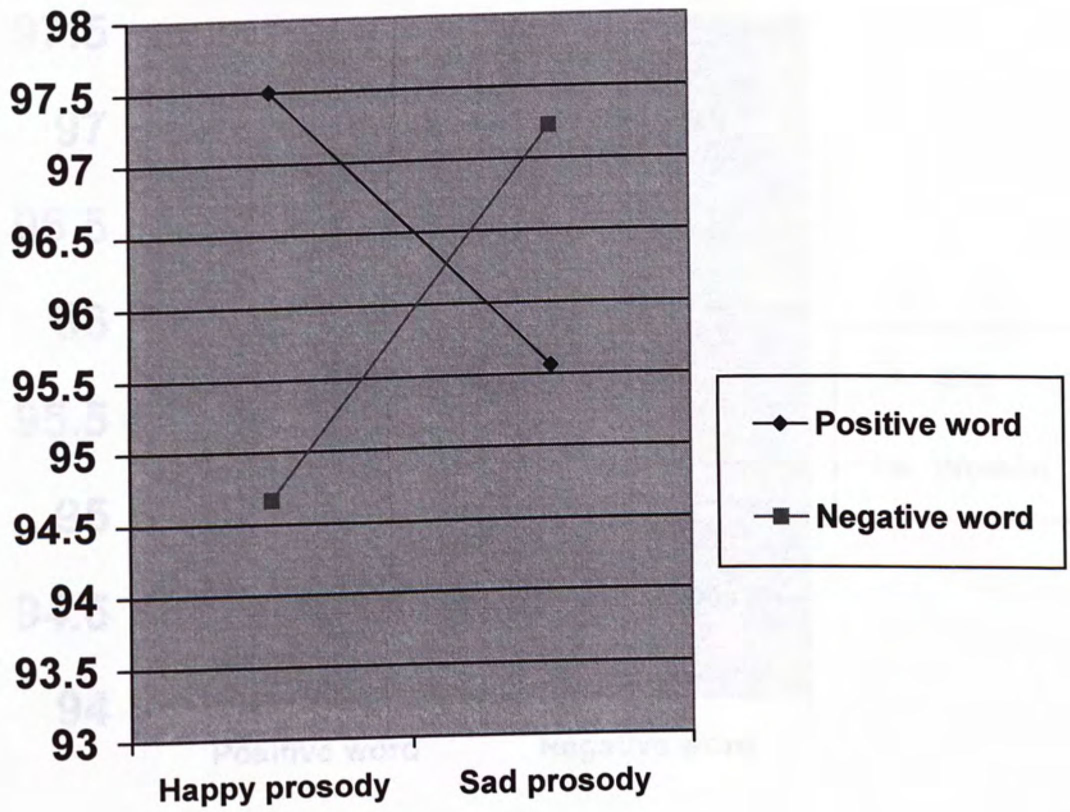
98
97.5
97
96.5
96
95.5
95
94.5
94
93.5
93
Happy prosody
Sad prosody
Positive word
Negative word

Figure 3.

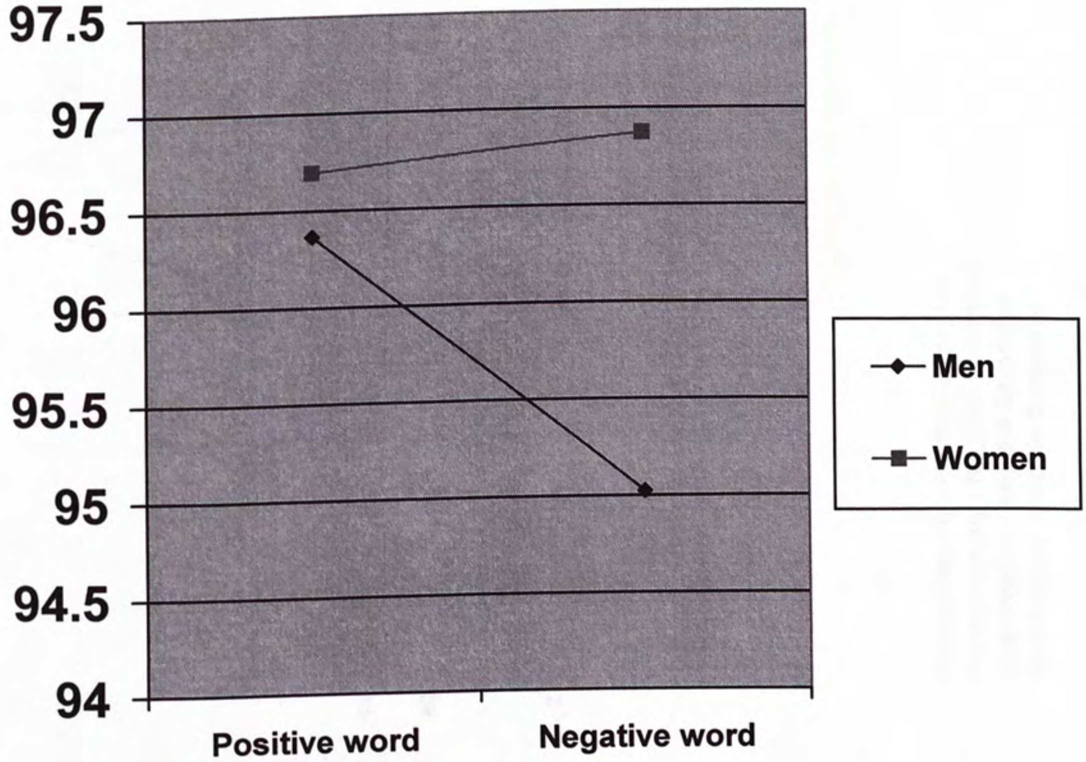
97.5
97
96.5
96
95.5
95
94.5
94
Positive word
Negative word
Men
Women

Figure 4.

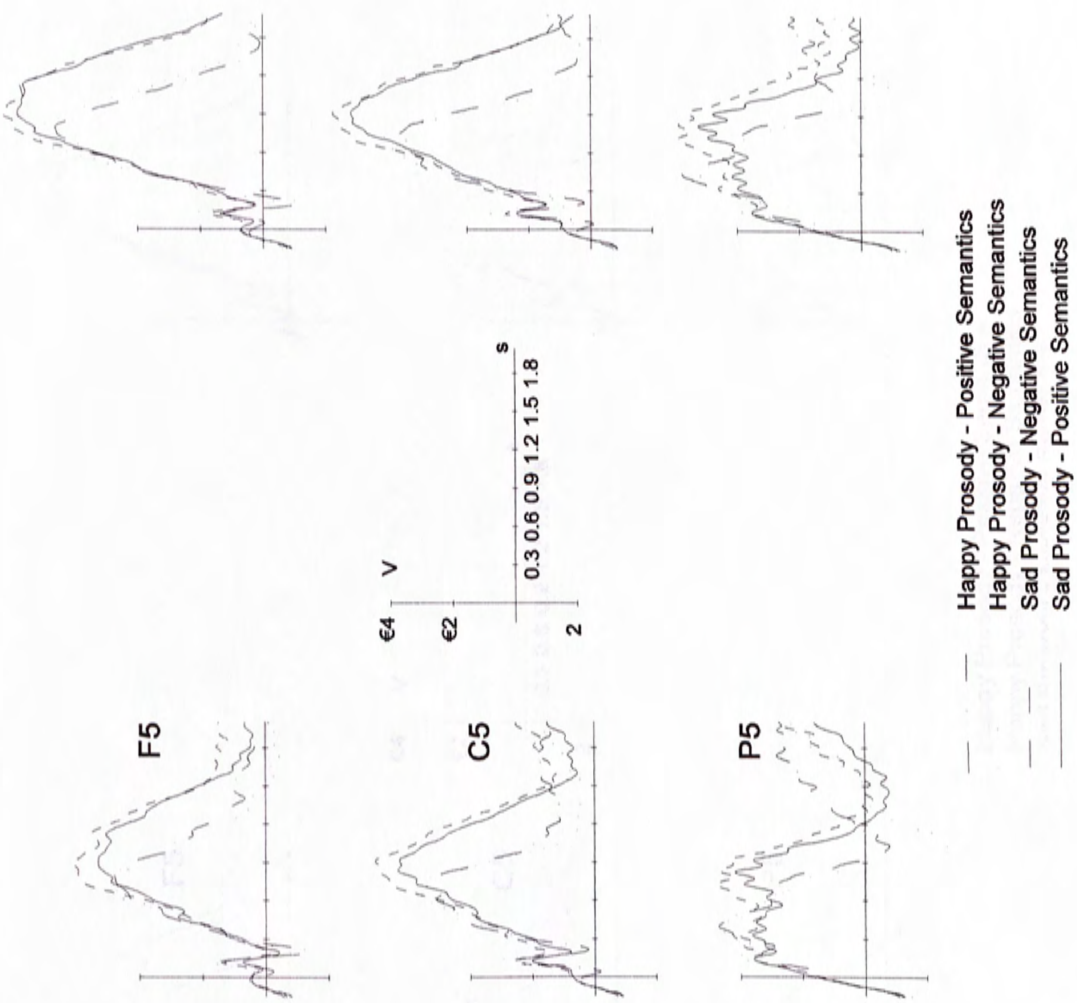
F5
C5
P5
€4
€2
2
V
0.3 0.6 0.9 1.2 1.5 1.8
s
Happy Prosody - Positive Semantics
Happy Prosody - Negative Semantics
Sad Prosody - Negative Semantics
Sad Prosody - Positive Semantics

Figure 5a.

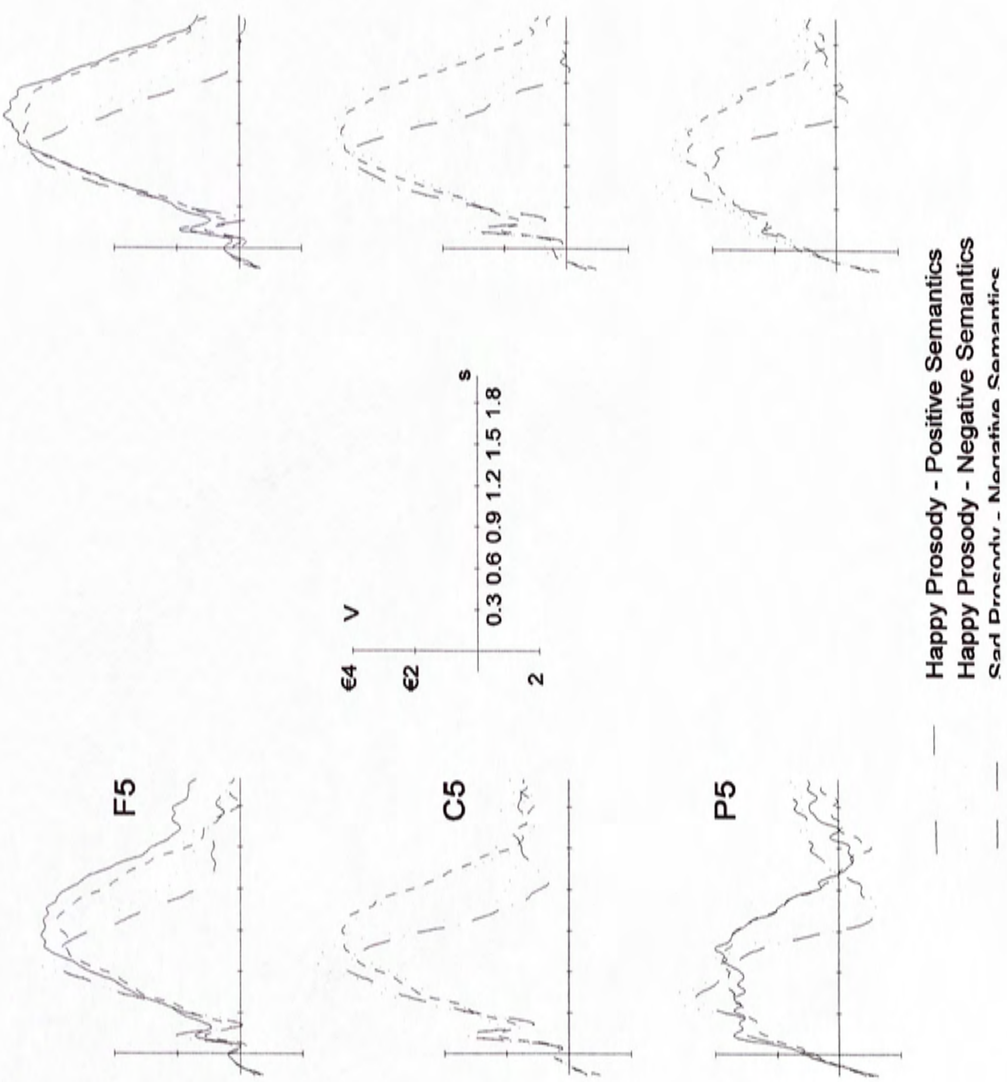
F5
C5
P5
€4
€2
2
V
0.3 0.6 0.9 1.2 1.5 1.8
s
Happy Prosody - Positive Semantics
Happy Prosody - Negative Semantics

Figure 5b.