

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Tracking the Sound of Human Affection: EEG Signals Reveal Online Decoding of Socio-Emotional Expression in Human Speech and Voice

Xiaoming Jiang

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/66418>

Abstract

This chapter provides a perspective from the latest EEG evidence in how brain signals enlighten the neurophysiological and neurocognitive mechanisms underlying the recognition of socioemotional expression conveyed in human speech and voice, drawing upon event-related potentials' studies (ERPs). Human sound can encode emotional meanings by different vocal parameters in words, real- vs. pseudo-speeches, and vocalizations. Based on the ERP findings, recent development of the three-stage model in vocal processing has highlighted initial- and late-stage processing of vocal emotional stimuli. These processes, depending on which ERP components they were mapped onto, can be divided into the acoustic analysis, relevance and motivational processing, fine-grained meaning analysis/integration/access, and higher-level social inference, as the unfolding of the time scale. ERP studies on vocal socioemotions, such as happiness, anger, fear, sadness, neutral, sincerity, confidence, and sarcasm in the human voice and speech have employed different experimental paradigms such as crosssplicing, crossmodality priming, oddball, stroop, etc. Moreover, task demand and listener characteristics affect the neural responses underlying the decoding processes, revealing the role of attention deployment and interpersonal sensitivity in the neural decoding of vocal emotional stimuli. Cultural orientation affects our ability to decode emotional meaning in the voice. Neurophysiological patterns were compared between normal and abnormal emotional processing in the vocal expressions, especially in schizophrenia and in congenital amusia. Future directions highlight the study on human vocal expression aligning with other nonverbal cues, such as facial and body language, and the need to synchronize listener's brain potentials with other peripheral measures.

Keywords: affective voice, social communication, nonverbal cues, pragmatics, EEG/ERPs, empathy, anxiety, person perception

1. Introduction

Theoretical models based on electrophysiological studies have indicated early and late neurophysiological markers that index online perception of vocal emotion expressions in speech as well as other higher-order socioemotive expressions (e.g., confidence, sarcasm, sincerity, etc.), which roughly correspond to each hypothesized processing stage [1, 2]. Studies with event-related potentials (ERPs), which focused on the analysis of averaged electrophysiological response to a certain vocal or speech event, have enlightened neurocognitive processes at a fine-grained temporal scale. The early fronto-central auditory N1 is known to be associated with a wide range of auditory stimulus types as a measure of sensory-perceptual processing. In vocal emotion processing, N1 has been linked to the extraction of acoustic cues that differentiate different types of vocal signals, frequency, and intensity parameters [3, 4], and is unaffected by differences in emotional meaning. The fronto-central P200 has been associated with the early attentional allocation or relevance evaluation of vocal signals [2, 5], ensuring preferential processing of emotional stimuli. Differentiation of P200 amplitude can be found between basic emotions [6] or between emotional vs. neutral speech [3, 7], suggesting that this component may reflect an early function of “tagging” emotional or motivational relevant stimuli. The P200 tended to be associated with higher mean and range of f_0 , larger mean and range of amplitude of speech, and slower speech rate [6], implicating that the early P200 modulation is partially explained by early meaning encoding as well as continued sensory processing [8]. A late centro-parietal positivity (also named LPC) evoked by vocal emotion expressions has been defined as a positive-going wave starting about 500 ms post-onset of the vocal stimuli and perhaps sustaining until 1200 ms depending on stimulus features. The LPC is considered as reflecting continued or second-pass evaluative process of the meaning of vocal emotional signals [2, 5]. The LPC was larger in emotional vocal stimuli, leading to larger differences in the LPC amplitude among basic emotion types [6], suggesting a more elaborative processing vocal information at this stage. In addition to these ERP effects, a more delayed sustained positivity may reflect a listener's attempt to infer the goal of a speaker, especially when an expected way of speaking is mismatched in an utterance context [9]. These event-related potential components have provided a useful tool to examine the temporal neural dynamics of emotional decoding in voice and speech.

2. Neurophysiological studies on basic vocal emotion in speech and voice

Vocal emotion has been investigated mainly in vocalization and speech. A study compared the ERP responses toward the perception of three basic emotions (happiness, sadness, and anger) in vocalization vs. pseudo-speech (same as real-speech except the lexical-semantic contents were replaced by meaningless syllables [10, 11]) in a task when listeners were presented with emotional vocal expressions followed by emotional and neutral faces and were asked to judge the emotionality of the face. Pell et al. [11] showed that the vocalization and speech can be differentiated very early at about 100 ms. Vocalization elicited a larger, earlier, and more differentiated P200 between emotions, and a stronger and earlier late-positivity effect. These findings support a preferential decoding in the neurophysiological system of vocalization over speech-embedded emotions in the human voice. They also demonstrated

angry voice elicited the strongest P200 than the other expressions. In another study in which anger, happiness, and neutral vocalizations were compared, anger elicited a stronger positivity in the 50 ms while both anger and happiness elicited a reduced N100 and an increased P200 as compared with neutral vocalization [7]. These findings, taken together, suggest an early sensory registration of emotional information which is assigned increased relevance or motivational significance in decoding human vocalization.

Earlier ERP works have focused on how the brain responded to emotional transitions in the voice and to the transition in both voice and lexico-semantics simultaneously [13]. Using a crosssplicing technique, a leading phrase of a sentence was crossspliced with the main stem of a sentence either congruent or incongruent in prosody with the leading phrase. The onset of the crosssplicing point of the vocal expression in the main sentence elicited a larger negativity (350–550 ms) for a mismatch in both voice and lexico-semantics and a larger more right-hemispheric distributed positivity (600–950 ms) for a mismatch in voice only (pseudo-utterances: [3]; utterances with no emotional lexical items: [1]). The negativity suggested an effort of integrating the emotional information in both vocal and semantic channel with the context. The late positivity suggests a detection of acoustic variation in the vocal expression.

Some evidence further delineated the role of a specific acoustic feature in the ERP responses toward the vocal emotion decoding. For example, one EEG study compared the ERPs for the mismatching emotional prosody (a statement with neutral voice which was disrupted by an anger voice) and that for the matching prosody revealed an increased N2/P3 as compared with the matching prosody ([12]). The amplitude of the N2/P3 complex was reduced and the latency of such complex was more delayed when the intensity for that prosody was weakened. This finding suggests that emotional significance in the voice can be promoted by increased sound intensity. The role of a specific acoustic profile such as loudness of sound needs to be specified in vocal-emotion studies.

3. Neurophysiological studies on vocal sarcasm, sincerity and confidence

In order to evaluate whether and how basic emotional and higher-level social information (e.g., attitudinal) are manifested in the brain in a different manner, Wickens and Perry [13] compared the ERP responses to neutral, angry, and sarcastic expressions. These expressions began with a leading phrase (e.g., He has) in a neutral voice and were followed by an expression (e.g., a serious face) intoned with different voices. As compared with the neutral expression, both angry and sarcastic expression elicited an increased P200 and a late positivity effect (450–700 ms) with no amplitude difference between the two emotions. The angry voice also elicited an early N100 as compared with the other two expressions when listeners performed a probe-verification task. These findings revealed similar neurocognitive processes between basic emotion and interpersonal attitudes conveyed in the voice while the basic emotion seems to be registered earlier under certain conditions. Other studies revealed that the decoding of sarcasm involved similar neurocognitive processes to social intention perception. Rigoulot et al. [14] compared compliments with sincere vs. insincere tone of voice (What do you think of my presentation? I think it is very interesting) and found that the sincere compliment to the question elicited a larger P600 effect as compared with the

insincere one. This ERP effect was localized in the left insula which is associated with the action of lying and concealment.

Recent growing evidence has been accumulated in the field of decoding of speaker's feeling of (un)knowing using event-related potentials. In Jiang and Pell [15], vocal expression of confidence was manipulated such that statements which sounded very confident, somewhat confident, and unconfident and those which sounded neutral were presented to native English speakers. At the onset of the vocal expression, the confident expression elicited an increased positive response than the other two types of expressions. The unconfident expression elicited an increased P300 as compared with the confident and the neutral expression. The neutral voice produced a more-delayed positivity as compared with all confidence-intending expressions.

Two follow-up experiments further evaluated how the decoding of vocal confidence expression is impacted by the presence of additional linguistic cues which either congruent [16] or incongruent [17] with the tone of voice in statements which followed the linguistic cues. Different from the statements with no lexical cues, statements with congruent cues (e.g., I'm sure; Maybe) elicited an increased N1 and P2 for confident than for unconfident and close-to-confident expressions, and an enhanced delayed positivity in unconfident and close-to-confident expression than confident one. Moreover, the direct comparison between statements with and without a preceding lexical phrase elicited a reduced N1, P2, and N400 in those without a phrase [16]. The incongruent cues elicited different ERP effects at the onset of the main statement of confident and unconfident tones. The unconfident statement elicited an increased N400 or late positivity (depending on the listener's gender). The confident statement elicited a more delayed, sustained positivity effect. Source localization of these ERP effects revealed pre-SMA for N400, suggesting a difficulty in accessing the speaker meaning, and SFG, STG and insula underlying the late positivity effect, suggesting an increased demand of executive control to implement the attentional resources and socioevaluative processes [17]. These studies extended the neurocognitive model for basic vocal emotion and argued for a perspective of studying the neurophysiological mechanisms underlying decoding interpersonal and sociointeractive affective voice.

4. Modulation of brain responses toward vocal expression by other nonverbal expressions

One of the key questions in emotional communication is how decoding vocal information is aided by other nonverbal cues. The neurophysiological studies have focused on emotional processing when voice is paired with other nonverbal social cues (such as face). In a task when participants were asked to evaluate the actor's identity (e.g., monkey or not) rather than the emotion, the simultaneous presentation of vocal and facial expressions revealed some similar ERP correlates of emotional information as the vocal expression did [18]. The bimodal emotional cues elicited a larger P200 and P300 for happy and angry expressions and a larger N250 for neutral expression, suggesting that an implicit affective processing of audiovisual

information emerges as early as 200 ms. Using a priming paradigm in which a face was followed by a vocal expression of words either congruent or incongruent with the emotion of the facial expression (happiness vs. anger), Diamond and Zhang [19] revealed that the mismatch elicited an increased N400 followed by a late positivity. Further, source localization of these two effects revealed activations in the superior temporal gyrus and inferior parietal gyrus dominated in the right hemisphere.

The interaction between vocal and other nonverbal emotional information was also examined in detection of emotional change. In a study in which participants were presented with simultaneously presented vocal and facial expressions while being asked to detect the change of emotion from neutral to anger or happiness conveyed in voice or in face [20]. The P3 associated with the detection of the emotional categorical change in both voice and face was larger than the sum of the change in single channel (see also [21]). The N1 associated with the detection of early acoustic change was dependent on whether their attention was guided to the voice or the face, with the attention to the voice yielding to a N1 in bimodal change larger than the sum of the two single modal change conditions. These findings suggest the modulation of selection attention on voice-face integration during emotional change perception in early sensory processing.

5. Effects of task demand, listener characteristics, and speaker characteristics on brain responses toward vocal expression decoding

Decoding emotion from voice has suffered from many variations, one noticeable factor is the communication context. The task relevance modulates the level of explicitness of emotional processing of vocal expression. One study presented mismatching and matching emotional prosody to listeners and asked them to judge the emotional congruency (where the emotional information is task relevant), or to verify the consistency between a visually presented lexical item and the statement [22]. Three ERP effects were elicited: an early negativity effect from 150 to 250 ms regardless of task relevance and the pattern of mismatch, an early positivity from 250 to 450 ms only on angry voice which was preceded by a neutral voice but regardless task relevance, and a late positivity effect after 450 ms for the task that directed listener's attention to the emotional aspects of the vocal expression. Explicit task relevant processing emotionality enhanced vigilance in perceiving emotional change in the voice.

Vocal emotion decoding is also characterized according to the listener's characteristics. Developmental studies revealed neurophysiological correlates of emotional voice processing (especially negative emotion) were similar in children and adults [23]. Using emotional interjections ("ah"), Chronaki et al. [23] compared angry, happy, and neutral voices in 6- to 11-year-old typically developing children. The N400 was attenuated by angry than by other expression types over parietal and occipital regions. Comparing neurocognitive processes along stages of early human development merits further examinations [24].

Another topic is how listener's linguistic and cultural background affect their perception of vocal expressions. In a recent EEG study, native North-American English and Chinese speakers were asked to detect the emotion of the vocal or facial expression in a voice-face pair [25]. The emotional information between the voice and face was either congruent or incongruent. Both groups were sensitive to the emotional differences between voice and face, revealing lower accuracy and higher N400 amplitude for the incongruent voice-face pairs. However, English speakers showed more pronounced N400 enlargement and more reduced accuracy when vocal information was attended, suggesting that those from a Western culture suffered from a larger interference effect from irrelevant face information. Another study using a passive odd-ball paradigm in which the two groups of listeners were presented with deviant or standard facial expressions which were paired with a vocal expression or not [26]. Chinese speakers showed a larger mismatch negativity when vocal expression was presented together with a facial expression, suggesting that individuals from an eastern culture were more sensitive to an interference from task-irrelevant vocal cues. These findings implicate a role of cultural learning and different cultural practices in communication shape neurocognitive processes associated with the early perception of voice-face emotional cues.

Listener's biological sex has been central in modulating the integration of emotional information in vocal and verbal channels [27, 28]. Recent evidence extended this idea beyond the basic emotion. Jiang and Pell [15] examined the sex difference in evaluating confidence in both confidence- and neutral-intending vocal expressions and the associated neural responses. They revealed that the delayed positivity effect elicited by neutral-intending expression was only observed in female listeners, suggesting an inferential process aimed at deriving speaker meaning from nonexpression-intending vocal expressions. Their further analysis revealed that, when vocal statements were led by lexical phrases of some level of certainty (LEX + VOC), females elicited more pronounced N1 in confident expression and larger late positivity (550–1200 ms) in unconfident and close-to-confident expressions. When these statements were compared with those with only vocal cues signifying confidence (VOC only), reduced N1, P2 as well as N400 were observed in females [16]. These findings suggest the enhanced sensitivity to socioemotional information for females in vocal communication. Females and males also engage different strategies in resolving conflicting information in vocal expressions. Jiang and Pell [17] demonstrated that the conflicting message of vocal confidence expressions elicited different ERP effects in female vs. male listeners. The confident statement following an unconfident phrase elicited a larger delayed positivity only in a female participant; while the unconfident statement following a confident phrase elicited an N400 in a male participant and a P600 effect in male participants. These findings provided a picture of how mixed messages are dealt with in female vs. male brain: in face of a mismatch in vocal expressions, the female attempted to unify separate information to establish an integrated representation while the male updated the initially built representation by switching an alternative interpretation (for example, by saying “She has access to the building” in the unconfident voice following “I’m certain,” the speaker reveals some level of hesitation).

Given its sociointeractive nature, inferring a speaker meaning from interactive emotive expression is susceptible to listener's traits and personality characteristics. One factor which

has been ignored but should be evaluated is the individual's interpersonal sensitivity. Jiang and Pell [16, 17] measured individual's interpersonal sensitivity using interpersonal reactivity index (IRI) [29] and regressed the early and late ERP responses toward perceiving a certain level of confidence to the interpersonal sensitivity. They found that those who displayed higher IRI score revealed more pronounced delayed positivity effects in close-to-confident and unconfident congruent expressions [16] and in incongruent confident expressions preceded by an unconfident phrase [17]. A further examination of such individual difference revealed that a larger positivity for a female listener fully mediated their perceptual adjustment toward that incongruent expression (e.g., judging the incongruent confident expression to be less confident than the congruent one).

Listener's level of anxiety also places an important role in modulating their neural responses toward decoding vocal emotions. In Jiang and Pell [15], both early (N100) and late ERP responses (P200, late positivity) were associated with the one's trait anxiety with those exhibiting higher trait anxiety revealed a reduced N100 and late positive effect in both vocalization and speech but an enhanced P200 effect in vocalization. Jiang and Pell [17] further found that the P200 in response to the confident vs. unconfident vocal expression was larger in those who displayed a lower level of trait anxiety and such modulation mediated the reduced P200 in male listeners who showed reduced anxiety as compared with female listeners.

6. Brain responses toward vocal expression in clinical populations

The study on vocal emotion decoding in normal populations has provided a wide range of neurophysiological markers and experimental paradigms to examine how such process is impaired in a clinical context. Studies have been focusing on psychiatric-risk populations and neurodevelopment disorders.

A study used an oddball paradigm in which a group of healthy listeners with anxious and depressive tendencies and a group of controls detected the target of emotional stimuli from a sequence of neutral expressions [30]. The emotional expressions were presented in voice, in face, or in voice-face pair with congruent expressions. The amplitude of P3b in response to the deviant expression was reduced in the clinical group than the control group, only in voice-face presentation. This finding suggests the crossmodal design as an effective approach to increase the sensitivity of the P300 amplitude difference between healthy populations and those with clinical symptoms.

Another study used an auditory oddball paradigm in which anger or happy deviant vocal or nonvocal synthesized syllables (data) were presented in a sequence of neutral syllables to listeners with symptoms in schizophrenia and normal listeners [31]. A larger mismatch of negativity was elicited following the deviant angry voice and anger-bearing nonvocal sounds and such enlargement was decreased in those with schizophrenia. The weaker the MMN amplitudes, the more positive symptoms of schizophrenia. Using MMN responses to anger voice, anger-derived nonvocal sound could predict whether someone received a

clinical diagnosis of schizophrenia. These findings implicate that the emotional salience detection of voices differentiate the negative and positive symptoms in neuropsychiatric disorders at the preattentive level.

The emotional prosody was also examined in those with congenital amusia (a specific neurodevelopmental disorder featured as tone-deafness, [32]). Lu et al. [32] presented emotional words spoken with declarative or the question voice to the amusics and their healthy control. The N1 was reduced and the N2 was increased in incongruent voice. The modulation of N1 was intact whereas the change in N2 was reduced in amusics, suggesting an impaired conflict processing in amusia. The authors argued that the impaired discrimination of speech intonation among amusic individuals may arise from an inability to access information extracted at early processing stages.

7. Applications and future directions

One application of these studies is to build an artificial intelligence to decode brain signals which contribute to socioemotion understanding. Most of the studies use the acted (posted) vocal expression as testing materials, which were produced by professional actors, public speakers, or amateurs to portray an intended emotion. In real-life communication, the communicators may use such emotional pose to achieve certain communicative goals. Some research purpose, for example, the cultural display in vocal expression communication, may be specifically favored by using posed stimuli. However, a call for research on naturalistic, ecological, and observation-based stimuli is highly recommended. Therefore, a future study is to examine how the brain differentiates “real” vs. “fake” vocal expression by looking at the neurophysiological responses.

Another implication of using EEG signals to study vocal emotion decoding is to test the effectiveness of speech-coding strategies used in hearing aids for deaf listeners when they distinguish the emotions via prosody-specific features of language [33, 34]. In Agrawal et al. [33], statements simulated with different speech-encoding strategies differentiated the P200 in the happy expression and an early (0–400 ms) and late (600–1200 ms) gamma band power increase in vocal expressions of happiness, anger, and neutral. In Agrawal et al. [34], the P200 was differentiated by different simulation strategies in all types of emotions, and was larger in happiness than in other emotion types across speech-encoding strategies. These studies emphasized the importance of vocoded simulation to better understand the prosodic cues which cochlear impairment users may be utilizing to decode emotion in the voice. Further studies will also draw upon the merits of multimodal recording and synchronization of neurophysiological and peripheral physiological responses to decoding vocal expressions, including eye movement, pupil dilation, heart rate tracking, etc., to understand how different systems support the understanding of social and emotional information in speech and vocalizations.

Acknowledgements

Special thanks to Professor Dr. Marc D. Pell who leads the Neuropragmatics and Emotion Lab in the School of Communication Sciences and Disorders, the McLaughlin Scholarship and McGill MedStar by Faculty of Medicine, McGill University that were awarded to the author.

Author details

Xiaoming Jiang

Address all correspondence to: xiaoming.jiang@mail.mcgill.ca

School of Communication Sciences and Disorders, Neuropragmatics and Emotion Lab (Pell Lab), Faculty of Medicine, McGill University, Montréal, Canada

References

- [1] Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, 1151(1), 107–118. <http://doi.org/10.1016/j.brainres.2007.03.015>
- [2] Schirmer, A., & Kotz, S. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10, 24–30. <http://dx.doi.org/10.1016/j.tics.2005.11.009>
- [3] Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language*, 105(1), 59–69. <http://doi.org/10.1016/j.bandl.2007.11.005>
- [4] Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience*, 290, 175–184.
- [5] Kotz, S., & Paulmann, S. (2011). Emotion, language and the brain. *Language and Linguistic Compass*, 5, 108–125. doi: 10.1111/j.1749-818X.2010.00267.x
- [6] Paulmann, S., Bleichner, M., Kotz, S. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, 4, 345. doi: 10.3389/fpsyg.2013.00345.
- [7] Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport*, 23(2), 108–112. <http://doi.org/10.1097/WNR.0b013e32834ea757>

- [8] Schirmer, A., Chen, C., Ching, A., Tan, L., Hong, R. (2013). Vocal emotions influence verbal memory: Neural correlates and interindividual differences. *Cognitive, Affective and Behavioral Neuroscience*, 13, 80–93.
- [9] Van Overwalle, F., Van den Eede, S., Baetens, K., Vandekerckhove, M. (2009). Trait inferences in goal-directed behavior: ERP timing and localization under spontaneous and intentional processing. *Social, Cognitive & Affective Neuroscience*, 4, 177–190. doi: 10.1093/scan/nsp003.
- [10] Pell, M. D., Paulmann, S., Dara, C., Alasser, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37, 417–435. doi:10.1016/j.wocn.2009.07.005
- [11] Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, 111, 14–25. <http://doi.org/10.1016/j.biopsycho.2015.08.008>
- [12] Chen, X., Yang, J., Gan, S., & Yang, Y. (2012). The contribution of sound intensity in vocal emotion perception: Behavioral and electrophysiological evidence. *PLoS ONE*, 7(1), e30278. <http://doi.org/10.1371/journal.pone.0030278>
- [13] Wickens, S., & Perry, C. (2015). What do you mean by that?! An electrophysiological study of emotional and attitudinal prosody. *PLoS ONE*, 10(7), 1–24. <http://doi.org/10.1371/journal.pone.0132947>
- [14] Rigoulot, S., Fish, K., & Pell, M. D. (2014). Neural correlates of inferring speaker sincerity from white lies: An event-related potential source localization study. *Brain Research*, 1565, 48–62. <http://dx.doi.org/10.1016/j.brainres.2014.04.022>
- [15] Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, 66, 9–34. <http://dx.doi.org/10.1016/j.cortex.2015.02.002>
- [16] Jiang, X., & Pell, M. D. (2016). Neural responses towards a speaker's feeling of (un)knowing. *Neuropsychologia*, 81, 79–93. <http://dx.doi.org/10.1016/j.neuropsychologia.2015.12.008>
- [17] Jiang, X., & Pell, M. D. (2016). The feeling of another's knowing: How “mixed messages” in speech are reconciled. *Journal of Experimental Psychology: Human Perception and Performance*. 42(9), 1412–1428. <http://dx.doi.org/10.1037/xhp0000240>
- [18] Liu, T., Pinheiro, A., Zhao, Z., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Emotional cues during simultaneous face and voice processing: Electrophysiological insights. *PLoS ONE*, 7(2), e31001. <http://doi.org/10.1371/journal.pone.0031001>
- [19] Diamond, E., & Zhang, Y. (2016). Cortical processing of phonetic and emotional information in speech: A cross-modal priming study. *Neuropsychologia*, 82, 110–122. <http://doi.org/10.1016/j.neuropsychologia.2016.01.019>
- [20] Chen, X., Han, L., Pan, Z., Luo, Y., & Wang, P. (2016). Influence of attention on bimodal integration during emotional change decoding: ERP evidence. *International Journal of Psychophysiology*, 106, 14–20. <http://doi.org/10.1016/j.ijpsycho.2016.05.009>

- [21] Chen, X., Pan, Z., Wang, P., Yang, X., Liu, P., You, X., Yuan, J. (2015). The integration of facial and vocal cues during emotional change perception: EEG markers. *Social Cognitive and Affective Neuroscience*, 11(7), 1152–1161. <http://doi.org/10.1093/scan/nsv083>
- [22] Chen, X., Zhao, L., Jiang, A., & Yang, Y. (2011). Event-related potential correlates of the expectancy violation effect during emotional prosody processing. *Biological Psychology*, 86(3), 158–167. <http://doi.org/10.1016/j.biopsycho.2010.11.004>
- [23] Chronaki, G., Broyd, S., Garner, M., Hadwin, J. A., Thompson, M. J. J., & Sonuga-Barke, E. J. S. (2012). Isolating N400 as neural marker of vocal anger processing in 6–11-year old children. *Developmental Cognitive Neuroscience*, 2(2), 268–276. <http://doi.org/10.1016/j.dcn.2011.11.007>
- [24] Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, 141(6), 1266–1287. <http://doi.org/10.1037/bul0000002>
- [25] Liu, P., Rigoulot, S., & Pell, M. D. (2015). Culture modulates the brain response to human expressions of emotion: Electrophysiological evidence. *Neuropsychologia*, 67, 1–13. <http://doi.org/10.1016/j.neuropsychologia.2014.11.034>
- [26] Liu, P., Rigoulot, S., & Pell, M. D. (2015). Cultural differences in on-line sensitivity to emotional voices: Comparing East and West. *Frontiers in Human Neuroscience*, 9(May), 311. <http://doi.org/10.3389/fnhum.2015.00311>
- [27] Schirmer, A., Kotz, S. (2003). ERP evidence for a sex-specific stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, 15, 1135–1148. doi:10.1162/089892903322598102
- [28] Schirmer, A., Kotz, S., Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech. *Cognitive Brain Research*, 24, 442–452. doi:10.1016/j.cogbrainres.2005.02.022
- [29] Davis, M. (1983). Measuring individual differences in empathy: Evidence for a multi-dimensional approach. *Journal of Personality and Social Psychology*, 44, 113–126. <http://dx.doi.org/10.1037/0022-3514.44.1.113>
- [30] Campanella, S., Bruyer, R., Froidbise, S., Rossignol, M., Joassin, F., Kornreich, C., & Verbanck, P. (2010). Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations. *Clinical Neurophysiology*, 121(11), 1855–1862. <http://doi.org/10.1016/j.clinph.2010.04.004>
- [31] Chen, C., Liu, C.-C., Weng, P.-Y., & Cheng, Y. (2016). Mismatch negativity to threatening voices associated with positive symptoms in schizophrenia. *Frontiers in Human Neuroscience*, 10(July), 1–11. <http://doi.org/10.3389/fnhum.2016.00362>
- [32] Lu, X., Ho, H. T., Liu, F., Wu, D., & Thompson, W. F. (2015). Intonation processing deficits of emotional words among Mandarin Chinese speakers with congenital amusia: An ERP study. *Frontiers in Psychology*, 6(March), 1–12. <http://doi.org/10.3389/fpsyg.2015.00385>

- [33] Agrawal, D., Thorne, J., Viola, F., Timm, L., Debener, S., Büchner, A., Dengler, R., Wittfoth, M. (2013). Electrophysiological responses to emotional prosody perception in cochlear implant users. *NeuroImage: Clinical*, 2, 229–238.
- [34] Agrawal, D., Timm, L., Viola, F., Debener, S., Büchner, A., Dengler, R., Wittfoth, M. (2012). ERP evidence for the recognition of emotional prosody through simulated cochlear implant strategies. *BMC Neuroscience*, 13, 113.