Multimodal recognition of emotions in music and language

Alice Mado Proverbio¹,² and Francesca Russo¹,²

Abstract
We investigated through electrophysiological recordings how music-induced emotions are recognized and combined with the emotional content of written sentences. Twenty-four sad, joyful, and frightening musical tracks were presented to 16 participants reading 270 short sentences conveying a sad, joyful, or frightening emotional meaning. Audiovisual stimuli could be emotionally congruent or incongruent with each other; participants were asked to pay attention and respond to filler sentences containing cities’ names, while ignoring the rest. The amplitude values of event-related potentials (ERPs) were subjected to repeated measures ANOVAs. Distinct electrophysiological markers were identified for the processing of stimuli inducing fear (N450, either linguistic or musical), for language-induced sadness (P300) and for joyful music (positive P2 and LP potentials). The music/language emotional discordance elicited a large N400 mismatch response ($p = .032$). Its stronger intracranial source was the right superior temporal gyrus (STG) devoted to multisensory integration of emotions. The results suggest that music can communicate emotional meaning as distinctively as language.

Keywords
event-related potentials, musical emotions, STG, multimodal stimuli, mismatch response, incongruence, reading

The aim of this study was to investigate if, through which mechanism, and how clearly music can transmit distinct emotional states (Panksepp, 1994). While it is known that semantics and prosody can convey precise emotional hues, the neural mechanism with which music is able to transmit emotions is still under investigation. Coutinho and Dibben (2013) found that reported emotions to music and speech prosody could be predicted from a set of common psychoacoustic features including loudness, tempo/speech rate, melody/prosody contour, spectral structure, sharpness, and roughness. This hints at a common mechanism for extracting the emotional content of auditory affective information. Music is indeed able to arouse specific emotions in

¹Department of Psychology, University of Milano-Bicocca, Milan, Italy
²Milan Center for Neuroscience, Milan, Italy

Corresponding author:
Alice Mado Proverbio, Department of Psychology, University of Milano-Bicocca, Piazza dell’Ateneo Nuovo 1, U6, 20126 Milan, Italy.
Email: mado.proverbio@unimib.it
the listener. The most frequently reported musical emotions are calm/relaxation, happiness/joy, nostalgia, pleasure/enjoyment, sadness/melancholy, energy, and love/tenderness (see Juslin, 2013; Juslin, Liljeström, Västfjäll, Barradas, & Silva, 2008).

As for the brain mechanisms involved, we know, for example, that fearful music activates the amygdala (e.g., Baumgartner, Lutz, Schmidt, & Jäncke, 2006; Koelsch et al., 2013), that atonal music induces fear bradycardia (Proverbio et al., 2015), that listening to heroic music activates the ventral striatum (Vuilleumier & Trost, 2015). Again, listening to sad music activates the insula and the cingulate cortex (Janata, 2009; Trost, Ethofer, Zentner, & Vuilleumier, 2012), but the hidden mechanisms for which musical auditory frequencies are interpreted as meaningful by brain structures remain unclear. Paquette, Takerkart, Saget, Peretz, and Belin (2018) proposed that music could tap into neuronal circuits that have evolved primarily for the processing of emotional vocalizations, such as laughter and crying (see also Proverbio, De Benedetto, & Guazzzone, 2020; Proverbio, Santoni, & Adorni, 2020 for an integrated model of emotion comprehension in speech, vocalizations and music). However, the precision and relative universality with which music can transmit emotional meanings remains a rather unexplored issue.

The two-dimensional circumplex model of emotional states (Russell, 1980) predicts that emotions can be conceptualized as varying in intensity, along a mild/strong gradient, and in polarity (positive vs. negative). In this view, relaxation and sadness would be comparable in intensity (although varying in polarity). Here, we only considered fear, joy, and sadness feelings, while relaxation state was not considered as an emotion for several reasons. First, because of the specificity of language as aid for transmitting emotions: relaxing written words are experienced as neutral ones, as shown by preliminary tests. Moreover, as shown by stimulus validation, sad sentences (describing child abuse, rape, orphan stories, etc.) were more arousing than theoretically assumed by this model (i.e., more arousing than scary words). Furthermore, a two-dimensional model of emotional states (considering the pleasure–displeasure and deactivation–activation dimensions) does not hold for brain activation approaches. In fact, neural circuits supporting the various emotions are non-coincident, so that one area (e.g., insula, or cingulate cortex) might be more active in association with a supposedly “deactivating” emotion (e.g., sadness), and less active in association with an “activating” emotion (e.g., happiness). It seemed therefore more proper conceptualizing emotions as distinct-states supported by devoted neural circuits (e.g., Panksepp, 1994).

One possible way of investigating how music can convey clear emotional information is to compare music with other means of transmitting emotions (e.g., language, prosody, lyrics, facial expressions). Several studies explored the integration between affective information conveyed by facial expressions and music (e.g., Jeong et al., 2011; Jolij & Meurs, 2011; Proverbio & De Benedetto, 2018; Proverbio, Camporeale, & Brusa, 2020) or images and music (Spreckelmeyer, Kutas, Urbach, Altenmüller, & Münte, 2006), finding enhanced N400 responses to emotional discordance. Using magnetoencephalography, Parke, Perry, and Goodin (2016) also found enhanced N400 responses to word emotional discordance during sentence reading (e.g., “My mother was killed and I felt great”). In classical N400 paradigms, pairs of semantically congruent words (such as “tea” and “coffee”), as opposed to pairs of incongruent items (such as “milk” and “hydrocarbon”) are presented during electroencephalogram (EEG) recording. The second item of an incongruent pair typically elicits a measurable negative potential known as the N400 response, reflecting the conjoined processing of two incongruous information (Kutas & Federmeier, 2000; Lau et al., 2008).

To explore the ability of music to transmit emotions, in the present study, clear emotional meanings conveyed by written language were combined with piano musical fragments, which
might be emotionally discordant, or not with the verbal message, by means of N400 paradigm. Music has rarely been compared with linguistic stimuli, except for single words (e.g., Goerlich et al., 2012), while it has often been compared with lyrics and prosody (Morton & Trehub, 2001; Vidas et al., 2018).

The aim of the present study was multifold. By measuring the amplitude of event-related potential (ERP) responses as a function of music emotional content and congruence with linguistic information, we hoped to gather information about how precisely listeners comprehend the emotional content of music. (1) We wished to assess whether music was able to transmit emotions (joy, sadness, and fear) as distinctive as those conveyed by semantics of language. At this aim, emotional sentences were congruently or incongruently paired with piano music excerpts transmitting sadness, joy, or fear. If music clearly conveyed an emotional significance, a N400 response to incongruent music/sentence pairing would have been observed. (2) Furthermore, we wished to investigate whether the valence of emotional content of stimuli (positive vs negative) differentially modulated ERP responses. Specifically, it was hypothesized that fear (conveyed by music or written language) would be associated with enlarged N400-like anterior response, as previously found negative emotions conveyed by music (Proverbio et al., 2020), speech prosody (Grossman, Striano, & Friederici, 2005; Schirmer, Kotz, & Friederici, 2002), and vocalizations (Proverbio et al., 2020). Again, we expected that positive emotions such as joy (as opposed to sadness) would enhance the amplitude of P2 and late positivity (LP) as found for speech prosody (Paulmann, Bleichner, & Kotz, 2013; Paulmann & Kotz, 2008), music and vocalizations (Proverbio et al., 2020), and language (Kanske & Kotz, 2007). Finally, we expected that P300 component, reflecting the arousal response induced by a stimulus, would be greater to negative than positive emotions (Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000).

Material and methods

Participants

Twenty university students (10 males and 10 females) volunteered for the EEG study. They ranged in age from 21 to 28 years ($M_{age} = 23.65$ years, $SD = 1.75$) and had a high level of education (Diploma, $n = 6$, University degree, $n = 14$). Experiments were conducted with the understanding and written consent of each participant, with approval from the local Ethical Committee (Prot. N. RM-2019-176) and in compliance with APA ethical standards. Four participants were successively discarded for excessive EEG artifacts. Additional methodological details can be found in Supplemental Materials online.

Stimulus and procedure

Stimulus validation. Twenty university students (10 males and 10 females) took part in the validation procedure. They ranged in age from 22 to 28 years ($M_{age} = 24.15$ years, $SD = 2.08$) and had a high level of education (diploma, $n = 4$, university degree, $n = 16$).

The stimuli used in the validation phase and in the ERP study were auditory and visual. The auditory stimuli consisted of 24 piano short musical pieces, validated in the study by Vieillard and colleagues (2008) as transmitting specific feelings to the listener, namely, joy ($n = 8$), sadness ($n = 8$), and fear ($n = 8$). In their study, 59 students performed two categorization tasks, based on recognition and subjective experience during listening to musical fragments labeled as happy, sad, and scary. For each labeled emotion, listeners performed a rating on a 10-point
In all cases, the intended emotion was clearly recognized, with more than 80% of correct responses (joy = 99%, sadness = 84%, and fear = 82%).

The visual stimuli were 300 Italian sentences eliciting the same types of emotions: joy ($n = 100$), sadness ($n = 100$), and fear ($n = 100$) created on purpose for the present study. They elicited the same types of emotions of musical stimuli: joy (e.g., “The mother heard the first word of her infant”), sadness (e.g., “He fought hopelessly all his life against cancer”), and fear (e.g., “He opened his eyes wide and found himself submerged by the rubble”).

To validate the sentences emotional content, participants of this study were asked to read them and decide which emotional content they expressed among joy, sadness, and fear. Participants were also asked to indicate the intensity of the emotion induced, by means of a 3-point Likert-type scale (“I feel”: 0 = calm, 1 = activated, and 2 = very activated) taken from a modified version of the self-assessment Manikin tool (Bradley & Lang, 1994). This evaluation tool is based on nonverbal responses and directly measures the activation associated with the emotional reaction of the person linked by an object or an event.

The validation results showed that 270 out of 300 sentences were evaluated coherently by at least 80% of evaluators, while 90% of sentences were categorized as reflecting the same emotion by 18 out of 20 evaluators. These sentences were selected as stimuli for the EEG recording study and were equally paired to congruent and incongruent musical pieces as shown in Table 1.

**Table 1.** Example of stimulus pairs as a function of congruency condition and specific emotional content. The final set comprised 270 emotional stimuli (90 joyful, 90 sad, and 90 frightening sentences) plus 30 neutral sentences containing a city name (e.g., “Many composers lived in Wien in the 1800s” or “The carbonara is a typical dish from Rome”) which acted as target stimuli.

<table>
<thead>
<tr>
<th>Incongruent</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phrase</strong></td>
<td>Joy (e.g., Opened anxiously his birthday present)</td>
<td>Sad</td>
</tr>
<tr>
<td><strong>Music</strong></td>
<td>Joy (e.g., The little girl crawled for the first time)</td>
<td>Frightening</td>
</tr>
<tr>
<td><strong>Phrase</strong></td>
<td>Fear (e.g., The snarling dog pounced on her)</td>
<td>Joyful</td>
</tr>
<tr>
<td><strong>Music</strong></td>
<td>Sadness (e.g., Her one-year-old child died of meningitis)</td>
<td>Joyful</td>
</tr>
<tr>
<td><strong>Phrase</strong></td>
<td>Fear (e.g., He remembered the threatening words of his tormentor)</td>
<td>Sad</td>
</tr>
<tr>
<td><strong>Music</strong></td>
<td>Sadness (e.g., That gentleman spent his entire life alone)</td>
<td>Frightening</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Congruent</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phrase</strong></td>
<td>Joy (e.g., He breathed clean, fresh air as he pedaled)</td>
<td>Joyful</td>
</tr>
<tr>
<td><strong>Music</strong></td>
<td>Sadness (e.g., That young boy died in a work accident)</td>
<td>Sad</td>
</tr>
<tr>
<td><strong>Phrase</strong></td>
<td>Fear (e.g., The cellar door slammed shut behind him)</td>
<td>Frightening</td>
</tr>
</tbody>
</table>

Procedure. Participants comfortably sat in a dark, acoustically and electrically shielded test area facing a screen located 120 cm from their eyes. They were instructed to gaze at a central
translucent yellow dot that served as a fixation point, and to avoid any eye or body movements during the recording session.

The sentences, written in white Arial narrow size 10 font on a black background, were presented centrally and distributed over 2/3 lines, to facilitate the reading without saccadic movements. Sentences occupied a visual angle of approximately $3°30′$ in length and $1°30′$ in height.

The presentation of the auditory and visual stimuli was subdivided into eight experimental runs lasting ~ 2 min each (112 s). In total, 54 music tracks ($n=18$ joy, $n=18$, sadness, and $n=18$ fear), 270 affective sentences ($n=90$ joy, $n=90$ sadness, and $n=90$ fear), and 30 target sentences were presented across runs. In this way, each run contained emotionally congruent and incongruent music/language associations, depending on the emotional content of music and language, and their respective congruence or incongruence.

Every musical fragment lasted ~ 12 s and was paired to five sentences. Thirty filler sentences containing city names were randomly distributed in each of the eight runs. Sentences lasted 1,700 ms and were intermixed by an Inter-stimulus Interval randomly varying from 400 to 600 ms. To get enough stimulus repetitions for ERP averaging musical stimuli were repeated twice. Musical fragments were normalized with Audacity to $−1$ dB, and leveled at $−70$ dB. Musical stimuli were validated by Vieillard et al. (2008).

The experimental task consisted of responding, as quickly and accurately as possible, by pressing a key on a joystick with the right index to sentences containing a city name (e.g., “The historical buildings of Graz are very famous”). The task was fictitious, and served to draw the subjects’ attention toward the stimulation without revealing the study’s aim.

**EEG recordings**

EEG signals were continuously recorded from 128 scalp sites at a sampling rate of 512 Hz using tin electrodes mounted in an elastic cap. Further details are reported in Supplemental Materials online.

**Results**

Hit percentage was 98.55%, while the mean response time was 967 ms. Behavioral data were not analyzed because there was no factor of variability, since targets were just rare attention-capturing stimuli.

**Linguistic ERPs**

The ANOVA carried out on linguistic N170 response did not show any significant effect of Emotional content, indicating how stimuli were perfectly comparable across classes from the perceptual, orthographic, and lexical point of view (see Figure 1). Full inferential statistics can be found in Supplemental Table 1.

The ANOVA performed on the N450 linguistic component showed a significant effect of hemisphere, with a greater N450 amplitudes over left ($−1.66 \mu V; SD=0.38$) than right hemisphere ($−.35 \mu V; SD=0.55$). The interaction of emotion $\times$ hemisphere factors was also significant. Post hoc comparisons showed that N450 was larger to frightening ($−2.01 \mu V; SD=0.39$) than joyful ($−1.60 \mu V; SD=0.50$) or sad sentences ($−1.38 \mu V; SD=0.41$) over the left hemisphere. The interaction of electrode $\times$ hemisphere was also statistically significant. Post hoc comparisons showed that N450 was of greater amplitude at left anterior frontal ($−2.10 \mu V; SD=0.37$) than other sites ($p=.002$).
The ANOVA performed on P300 amplitudes showed a significant effect of emotional content factor. Post hoc comparisons ($p = .016$) indicated that sad sentences elicited larger positivities ($–.47\mu V; SD = 0.58$) than frightening ($–1.40\mu V; SD = 0.50$) or joyful sentences ($–1.45\mu V; SD = 0.62$), as can be also appreciated by looking at waveforms of Figure 1.

The ANOVA carried out on P2 mean area amplitudes revealed a significant effect of the emotional content factor. Post hoc comparisons showed that joyful music elicited a larger P2 component ($1.05\mu V; SD = 0.42$), than frightening ($0.69\mu V; SD = 0.40$) or sad music ($0.33\mu V; SD = 0.44$). This result can be appreciated by looking at waveforms of Figure 2 and was confirmed by post hoc comparisons ($p = .030$). The ANOVA also revealed a significant effect of hemisphere, indicating larger P2 responses over the left ($1.26\mu V; SD = 0.42$) than right hemisphere ($0.12\mu V; SD = 0.41$).

The ANOVA carried out on N400 mean area amplitudes revealed a significant effect of Emotional content. Post hoc comparisons ($p = .003$) showed that N400 responses elicited by
frightening music were much greater ($-1.80 \mu V; SD = 0.44$) than to joyful ($-0.79 \mu V; SD = 0.50$) or sad music ($-1.01 \mu V; SD = 0.48$), as can be clearly appreciated in Figure 2.

The ANOVA carried out on LP mean area amplitudes showed a significant effect of emotional content factor. Post hoc comparisons ($p = .017$) showed that joyful music elicited a wider LP component ($1.02 \mu V; SD = 0.42$) than sad music ($0.45 \mu V; SD = 0.45$) or frightening music ($0.40 \mu V; SD = 0.44$).

The ANOVA also highlighted the significance of electrode factor indicating greater LP amplitudes over inferior frontal than frontal electrode sites. The ANOVA also yielded the significance of emotional content $\times$ hemisphere factors. Post hoc comparisons indicated much larger LPs to joyful than sad and frightening music over right hemispheric sites (see Figure 2) and a lack of discriminative effects over left hemispheric sites (RH: joyful = $1.2 \mu V; SD = 0.43$; sad = $0.38 \mu V; SD = 0.46$; frightening: $0.48 \mu V; SD = 0.48$. LH: joyful = $0.87 \mu V; SD = 0.44$; sad = $0.75 \mu V; SD = 0.47$; frightening = $0.55 \mu V; SD = 0.54$).

**Figure 2.** Grand-Average ERPs Recorded at Left and Right Sites in Response to Different Categories of Auditory Stimuli.

Early P2 and later LP positive responses were greater to joyful than other music types over inferior frontal sites. N400 was much greater to frightening music than other music types over posterior brain areas.
Audiovisual ERPs

The ANOVA performed on the N400 mean area amplitudes showed a significant interaction of congruency $\times$ hemisphere. Post hoc comparisons showed greater N400 responses to incongruent than congruent audiovisual stimuli especially over the right hemisphere ($p < .001$; congruent: LH = $-2.30 \mu V$, $SD = 0.40$; RH = $-1.05 \mu V$, $SD = 0.57$. incongruent: LH = $-2.066 \mu V$, $SD = 0.42$; RH = $-1.61 \mu V$, $SD = 0.58$). This effect can very well observed in Figure 3.

An analysis of the swLORETA inverse solution of surface potentials was also performed to identify the intracranial generators responsible for the cortical activity recorded when auditory and visual information conflicted from the emotional point of view (i.e., in the incongruent condition). The most active area was the right superior temporal gyrus (STG, BA22), followed, in order of magnitude, by the left fusiform area (BA37), the left Wernicke area (BA39), prefrontal areas bilaterally (BA10), and inferior frontal areas (BA10/BA11). The source reconstruction can be observed in Figure 4, while a list of active brain areas is reported in Supplemental Materials online (Supplemental Table 2).

Discussion

The study’s aim was to investigate the neural mechanism subserving the comprehension of the emotional meaning of language and music, as well as their conjoined processing and integration within the semantic system. The effect of emotional incongruity between a phrase and a musical soundtrack was observed with an implicit paradigm. Sentences, eliciting three different emotional states (joy, fear, and sadness) and musical traces expressing the same emotional categories were congruently or incongruently paired, while subjects engaged in a fictions task.

Several ERP components were analyzed. Some of them were associated with language processing (N170, N450, and P300), others with the encoding of the musical stimuli (P2, N400, and LP) while N400 reflected the conjoined processing of auditory and visual stimuli and therefore the congruence/incongruence effect. The amplitude of N170, sensitive to orthographic stimulus properties (Maurer, Zevin, & McCandliss, 2008) did not differ as a function of emotional class, thus suggesting an optimal perceptual and linguistic matching of stimuli across classes.

Language: frontal N450 larger to fear

The first visual response showing a valence-dependent modulation was the anterior N450, which much larger to frightening than sad/happy sentences. Similar enhanced anterior N400 responses to negative than positive content was found by Schirmer et al. (2002) reporting greater N400 amplitudes to negative emotional prosody. Grossman et al. (2005) recorded ERP responses in infants listening to words characterized by a happy, angry, or neutral prosody and found larger negative potentials to negative that other prosodic intonations. An enhanced N400 response to negatively than positively valenced stimuli was also found by Proverbio, De Benedetto, and Guazzzone et al. (2020) reporting larger N400s to negative versus positive vocalizations (e.g., crying as opposed to laughter), and to negative versus positive music. Intriguingly, a similar N400 enhancement during processing of negative vs. positive speech was also reported (Proverbio, Santoni, & Adorni, 2020). This possibly suggests the existence of a common mechanism for extracting and comprehending the emotional content of music, speech, and vocalizations, based on the harmonic spectrum and melodic structure of prosody, voice, and music.
Figure 3. (a) Grand-Average ERP Waveforms Relative to the Perception of Audiovisual Stimuli. Incongruent Pairs Elicited a N400 Response Over Anterior Sites (400–500 ms). The Graphic Highlights the Right Hemispheric Asymmetry of the N400 Response to Incongruence. (b) Isocolor Topographic Maps of N400 Surface Voltage (Front View). The Maps Plot the Average Voltage Recorded in the Latency Range Corresponding to the N400 Peak.
P300 response was greater to sad than happy/frightening sentences. This fits with previous literature (Bernat, Bunce, & Shevrin, 2001) reporting larger P300s to negative than positive words/sentences. Here, the fact that P300 was larger to sad than frightening sentences suggests that the former were more arousing than the latter stimuli. Indeed, P300 response is larger to more arousing than less arousing stimuli (Cuthbert et al., 2000). Brain activation induced by reading sad events might reflect affective relevance of the stories depicted (e.g., descriptions of persons suffering from excruciating pain) inducing empathic distress (Cuthbert et al., 2000; Schupp et al., 2004).

Music: positive responses (P2 and LP) larger to joy

P2 component was larger to joyful than frightening/sad music. This response would play a crucial role in the analysis of physical characteristics, such as intensity, frequency, tonality, stimulus, as well as in selective attention (Kotz & Paulmann, 2011; Morris, Steinmetzger, & Tøndering, 2016) which might suggest that joyful musical pieces were more complex from the perceptual point of view. Indeed, they were more articulated, and fast, with a higher number of notes per second, than sad/fearful music. Interestingly, P2 was larger over the left than right hemisphere which recalls the hemispheric asymmetry model predicting a left hemisphere role for positive emotions (Rodway & Schepman, 2007; Schmidt & Trainor, 2001) and a right hemisphere role for negative emotions such as fear/sadness (Ross, Thompson, & YenkoSk, 1997).
Later, LP deflection was greater to happy than sad/frightening music. This is strikingly similar to what found for music, nonverbal vocalizations, and speech by previous investigations (Proverbio, De Benedetto, & Guazzzone, 2020; Proverbio, Santoni, & Adorni, 2020). The inferior frontal distribution of this component (larger to positive stimuli) also resembles previous functional magnetic resonance imaging (fMRI) findings by Belin et al. (2008) and Koelsch et al. (2006) relative to positive vocalizations and pleasant music.

**Music: N400 larger to fear**

The auditory N400 response (450–550 ms) was larger to frightening music, as opposed to sad and happy music. Quite consistently, larger auditory N400s to negative prosody (Schirmer et al., 2002), speech (Grossman et al., 2005), human vocalizations, and music (e.g., crying; Proverbio et al., 2020) were previously reported, hinting at a common mechanism for comprehending the emotional content of music.

**Audiovisual stimuli: N400 larger to incongruent emotions**

N400 response showed a strong increase in amplitude induced by the emotional incongruence of audiovisual stimuli (regardless of the specific content, e.g., joy, fear, or sadness), especially over the right anterior areas. For example, reading a joyful phrase while listening to a sad music induced an increase in N400 amplitude. To investigate the cortical areas responsible for comprehending the emotional significance of audiovisual stimuli, a swLORETA was performed on the N400 response to incongruent pairs. The most active electromagnetic dipole was the right STG (BA22). It would play a key role in audiovisual integration (Jeong et al., 2011), and in the extraction of the emotions conveyed by speech and concurrent visual stimuli (e.g., lip motion). Previous studies have shown how STG is involved in processing multisensory emotional information (Pehrs et al., 2014; Salmi et al., 2017), prosody (Beaucousin et al., 2007), and music (Angulo-Perkins et al., 2014; Khalfa et al., 2008; Proverbio & De Benedetto, 2018). In our study, also very active according to swLORETA was the left fusiform gyrus, which is known as the “visual word form area” (Polk et al., 2002; Proverbio, Zani, & Adorni, 2008). The third more active region explaining N400 response to affective incongruence was the left middle temporal gyrus (BA39), also known as “extended Wernicke’s area” involved in word associations and semantics, as well as in speech prosody integration.

Also active were found bilaterally the medial and superior frontal areas (BA10), described as key structures in music listening (Altenmüller, Siggel, Mohammadi, Samii, & Münte, 2014; Khalfa et al., 2005; Proverbio & De Benedetto, 2018) and supporting working memory and attention. In addition, swLORETA showed the activation of the right inferior frontal gyrus (BA45) and the left orbitofrontal cortex (BA11) during the coding of incongruent stimuli. Inferior frontal gyrus would play a key role in the syntactic processing of language (Zaccarella, Meyer, Makuuchi, & Friederici, 2017). Its activity during the processing of incongruent audiovisual pairs might reflect the detection of a syntactic violation in the incoming information. Indeed, neuroimaging studies (Janata et al., 2002; Koelsch, 2011) suggest that music-syntactic processing involves the pars opercularis of the inferior frontal gyrus, which perfectly fits with our data.

The current theoretical framework predicts that music is interpreted by brain areas (e.g., STG) normally coding the emotional hues of human vocalizations (Paquette et al., 2018; Proverbio, De Benedetto, & Guazzzone, 2020; Proverbio, Santoni, & Adorni, 2020) and the presents findings fully support this hypothesis. However, simple piano melodies were only used in
this study, so that the role of musical structure (harmonic content, melodic profile, tempo, rhythmic structure, and timber), in transmitting specific emotional information, need to be further investigated. At this purpose, future studies involving musical fragments more complex in nature and played by instruments other than piano, should be carried out.

Conclusion

The present data showed that the extraction and integration of the emotional content of audiovisual stimuli is an automatic and fast occurring process. Piano musical fragments were able to transmit clear-cut information about their emotional content as early as 250 ms post-stimulus, while semantic comprehension showed later effects (450 ms). Distinct ERP markers were identified for the processing of fear (N450, either linguistic or musical), for language-induced sadness (P300) and for joyful music (positive P2 and LP potentials). Overall, this evidence supports the undoubted ability of music to convey and communicate emotions quite distinctively and effectively as the verbal messages. A key role in this process would be played by the right STG, thanks to its multimodal and affective properties in the processing of music, vocalizations, and speech.

Acknowledgements

The authors are very grateful to all participants and to Francesco De Benedetto, Roberta Adorni, Andrea Orlandi, and Alessandra Brusa for their technical help.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was supported by the 2018-ATE-0003 28882 grant entitled “Neural encoding of the emotional content of speech and music” from the University of Milano-Bicocca to AMP.

ORCID iD

Alice Mado Proverbio https://orcid.org/0000-0002-5138-1523

Supplemental material

Supplemental material for this article is available online.

References


Psychology of Music 00(0)


