

Affective Priming Effects of Musical Sounds on the Processing of Word Meaning

Nikolaus Steinbeis¹ and Stefan Koelsch²

Abstract

Recent studies have shown that music is capable of conveying semantically meaningful concepts. Several questions have subsequently arisen particularly with regard to the precise mechanisms underlying the communication of musical meaning as well as the role of specific musical features. The present article reports three studies investigating the role of affect expressed by various musical features in priming subsequent word processing at the semantic level. By means of an affective priming paradigm, it was shown that both musically trained and untrained participants evaluated emotional words congruous to the affect expressed by a preceding chord faster than words incongruous to the preceding chord. This behavioral effect was accompanied by an N400, an ERP typically linked with semantic processing, which was spe-

cifically modulated by the (mis)match between the prime and the target. This finding was shown for the musical parameter of consonance/dissonance (Experiment 1) and then extended to mode (major/minor) (Experiment 2) and timbre (Experiment 3). Seeing that the N400 is taken to reflect the processing of meaning, the present findings suggest that the emotional expression of single musical features is understood by listeners as such and is probably processed on a level akin to other affective communications (i.e., prosody or vocalizations) because it interferes with subsequent semantic processing. There were no group differences, suggesting that musical expertise does not have an influence on the processing of emotional expression in music and its semantic connotations. ■

INTRODUCTION

The question if, how, and what music is capable of communicating has roused scholarly interest for some time (Swain, 1997; Sloboda, 1986; Meyer, 1956). Recent empirical demonstrations have shown that, under certain circumstances, music appears to be capable of conveying semantically meaningful concepts (Koelsch et al., 2004). However, to date, more rigorous empirical demonstrations of the mechanisms underlying the communication of meaning have been lacking. The present study investigates the previously proposed role of emotional expression in music in communicating meaning.

The concept of emotion, affect, or emotional expression in music has received increased attention recently (see Juslin & Västfjäll, 2008, for a review). A distinction must be drawn between the kind of processes that lead to the recognition of emotions expressed in music and emotions elicited in the listener in response to the music. Whereas the former entails the recognition and categorization of sounds into discrete categories by virtue of their affective quality, the latter refers to the emotional state of the listener as a result of the emotional expression of the music. In the present context, emotion, affect, and emotional ex-

pression are used exclusively to refer to the recognition of emotions and not the feeling aspect.

Intuitively, the expression of an emotion would appear to be the most obvious way in which music can communicate. Of all signals music contains, emotional ones are the most prevalent, regardless of whether one feels emotions in response to music or simply recognizes their expression (Juslin & Västfjäll, 2008; Juslin, 2003). Thus, by communicating an emotion, however basic, music can refer to a variety of different affective states, which are, more or less, unanimously understood by listeners familiar with the musical idiom (Juslin, 2003). Recent evidence even suggests that certain emotions portrayed in music may be universally recognized, as Westerners and people totally unfamiliar with Western music show a statistically significant degree of agreement when classifying Western pieces as happy, sad, or scary (Fritz et al., 2009).

Discussions on how music can give rise to meaning have outlined several pathways for this to occur, such as by means of extra-musical associations, the mimicry of real-life features or occurrences, as well as tension-resolution patterns and emotional expression (Koelsch et al., 2004; Swain, 1997; Meyer, 1956). Whereas there is evidence for the first three (Steinbeis & Koelsch, 2008a; Koelsch et al., 2004), a direct link between emotional features and meaning has not been established. The expression of an emotion in music can be recognized very fast (under 1 sec; Daltrozzo & Schön, 2009; Bigand, Filipic, & Lalitte, 2005). It is likely

¹Max-Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, ²University of Sussex, Falmer, Brighton, UK

that this recognition also entails the activation of other concepts associated with that emotion (i.e., the expression of sadness in music will automatically lead to the activation of concepts such as *funeral* or *separation* which are associated with the recognized emotion). The coactivation of related concepts suggests that recognizing an emotion in music could have an effect on the processing of emotional information in other domains, such as language, which is coded in semantic form (i.e., through the *meaning* of the word). Such a mechanistic account of how emotional expression in music can be meaningful is in line with general theoretical accounts of priming, such as spreading activation (Collins & Loftus, 1975).

Recent models on music processing and its links to emotion perception and meaning have advanced the notion that each and every musical feature is capable of expressing an emotion, which are recognized as such, and which in turn can activate associated meaningful concepts (Koelsch & Siebel, 2005). The present study explores three such musical features to test this hypothesis: consonance/dissonance, mode (major/minor), and timbre (Experiments 1–3, respectively). It is important to note that such individual musical features do not resemble music as such, but represent fundamental constituents of (major–minor tonal) music. By means of a cross-modal affective priming paradigm, it was tested whether single musical features varying in affective valence can prime the semantic processing of subsequently presented words.

Cross-modal paradigms have been successfully employed both for studies on semantic priming (Holcomb & Anderson, 1993) and on affective priming (Schirmer, Kotz, & Friederici, 2002, 2005; Schirmer & Kotz, 2003). Affective priming typically entails the presentation of an affectively valenced (i.e., pleasant or unpleasant) prime stimulus followed by an affectively valenced target stimulus. Either the stimulus valence of the prime matches with that of the target (i.e., pleasant–pleasant or unpleasant–unpleasant) or it does not (i.e., pleasant–unpleasant, unpleasant–pleasant). Theory states that the processing of an affective target should be influenced by the valence of the preceding prime stimulus (Musch & Klauer, 2003), either by facilitating matched target processing or delaying mismatched target processing. Whereas these paradigms have been primarily employed to assess the psychology of evaluative processes, they have also been used to assess the general influence of affect in stimulus processing (Musch & Klauer, 2003). In addition, the literature on priming semantic processing with environmental sounds provides some useful insights into the ability of nonverbal material to prime the processing of word meaning (Orgs, Lange, Dombrowski, & Heil, 2006, 2007; Van Petten & Rheinfelder, 1995).

There are several issues relevant for conducting an affective priming experiment, particularly the SOA and the experimental task. Findings have, so far, suggested that affective priming only works with SOAs at 200 msec or less (Klauer, Rosnagel, & Musch, 1997; Fazio, Sanbonmatsu,

Powell, & Kardes, 1986). With longer SOAs, the priming effect disappeared, from which it was inferred that the affective activations are short-lived and the resulting priming effect is due to automatic processes, rather than strategic ones (McNamara, 2005). Because the present research questions were very similar to the ones addressed by Schirmer and Kotz (2003), an SOA of 200 msec, as was used in their study, was also presently employed. Tasks used in affective priming paradigms typically involve either the identification of a target attribute, pronouncing the target, or most frequently evaluating the target. The latter task was employed for the present set of experiments.

The present article aimed at testing whether a specific musical feature is capable of expressing affect, which is perceived as such by the listener and which has an influence on subsequent word processing at the semantic level. Primes always consisted of chords manipulated either in their consonance/dissonance, their mode (major/minor), or their timbre. The manipulation of each of these features has been shown to affect emotional responses (see Introduction section of each experiment for details), which ought to transfer onto the subsequent processing of word content. Word targets were presented visually 200 msec after the onset of the prime (see also Figure 1). Participants had to decide whether the word target had a pleasant or an unpleasant meaning. Each word was presented twice, either matching or not matching the valence of the preceding musical prime. Dependent variables were the speed and accuracy of target word evaluation. In addition, an EEG was recorded and ERPs were analyzed. The primary component of interest to these analyses was the N400, which has been shown to reflect semantic processing (Koelsch et al., 2004; Kutas & Federmeier, 2000; Kutas & Hillyard, 1980). Thus, if these single musical properties can convey meaning, the N400 ought to be sensitive to the match between musical prime and word target.

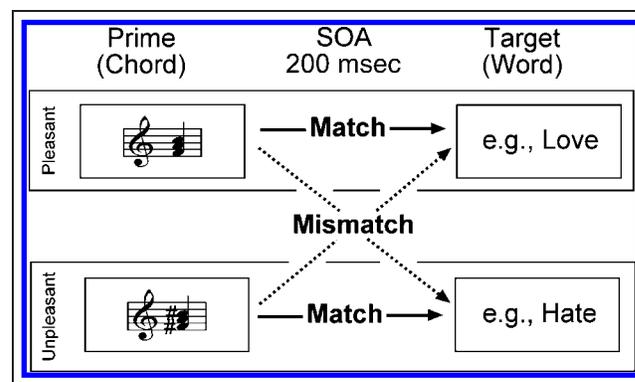


Figure 1. Design of the affective priming paradigm. Chords are used as primes and words as targets. To test whether certain musical features are capable of conveying meaning information, chords are varied along affective dimensions of the musical feature under investigation (Experiment 1: consonance/dissonance; Experiment 2: major/minor; Experiment 3: timbre).

With each variation of a musical feature carried out in separate experiments, it was hypothesized that the affective information contained in the acoustic parameter of a musical stimulus communicates meaning, and thus, that congruent prime–target pairs elicit a smaller N400 amplitude compared to incongruent pairs. In addition, congruency between target and prime should also affect the response times and accuracy of target evaluation, where congruent target words should elicit faster and more correct responses than incongruent target words. To investigate effects of musical training on semantic processing, two groups of subjects were measured: highly trained musicians and nonmusicians. There are several ERP studies reporting differences between the two groups with regard to basic perceptual processes (Wong, Skoe, Russo, Dees, & Kraus, 2007; Schön, Regnault, Ystad, & Besson, 2005; Tervaniemi, 2001) and musical expectancies in both adults (Schön, Magne, & Besson, 2004; Koelsch, Schmidt, & Kansok, 2002; Besson & Faita, 1995) as well as children (Jentschke, Koelsch, Sallat, & Friederici, 2008; Magne, Schön, & Besson, 2006). However, because there are no previous studies investigating training effects on processing the affective expression of musical features and its influence on semantic word processing, no directed hypotheses were made regarding ERP and behavioral differences between groups.

EXPERIMENT 1: CONSONANCE/DISSONANCE

Introduction

The aim of this experiment was to examine if acoustic roughness is capable of communicating meaning. Psychoacoustically, it has been suggested that the perception of harmonic roughness, specifically consonance and dissonance, is a function of the regularity of frequency ratios with which the simultaneously presented tones resonate (Plomp & Levelt, 1965). Typically, consonant music is perceived as pleasant sounding and dissonant music as unpleasant sounding. For instance, both infants (Zentner & Kagan, 1996) and adults (Sammler, Grigutsch, Fritz, & Koelsch, 2007; Koelsch, Fritz, von Cramon, Müller, & Friederici, 2006; Blood, Zatorre, Bermudez, & Evans, 1999) show a preference for consonance over dissonance, and functional neuroimaging experiments have shown that consonant/dissonant stimuli elicit activity changes in limbic and paralimbic brain structures known to be involved in emotional processing (Koelsch et al., 2006; Blood et al., 1999). This can be considered as strong evidence that harmonic roughness can modulate affective responses in music listeners. Additional brain structures typically involved in the coding of acoustic roughness include the auditory brainstem (superior olivary complex and inferior colliculus) and thalamus as well as the primary auditory cortex (for details, see Koelsch & Siebel, 2005). Hence, acoustic roughness appears to contain information capable of signaling affective categories, such as pleasantness and unpleasantness, thereby communicating basic emotional information. It was therefore hypothesized

that target words congruous with this information of a preceding musical prime stimulus (consonance = pleasant; dissonance = unpleasant) would elicit a smaller N400 than incongruous target words. In addition, it was hypothesized that this priming effect would also be reflected in faster and more accurate responses for congruous than for incongruous target words.

Methods

Participants

Twenty musically untrained (i.e., no formal musical training received—10 women) volunteers participated in the experiment. The same experiment was also carried out with highly musically trained participants, which, however, have already been published elsewhere in combination with data from an fMRI experiment (Steinbeis & Koelsch, 2008b). On average, participants were 24.75 years old ($SD = 2.51$). All subjects were right-handed, native German speakers, with normal or corrected-to-normal vision, and no hearing impairments.

Materials

The prime stimulus material consisted of 48 chords of piano timbre, of which 24 were consonant and, therefore, pleasant sounding and of which 24 were dissonant¹ and, therefore, unpleasant sounding. The consonant stimuli were major chords, presented in root position (e.g., C–E–G–C), or as six–four chords (e.g., G–C–E–G). Dissonant stimuli involved two types, one using the following superposition of intervals: augmented fourth, fourth, minor second (e.g., C–F#–B–C) and another one, namely, a superposition of minor second, fourth, and augmented fourth (e.g., C–C#–F#–C). Both consonant and dissonant chords were played in each of the 12 keys of the chromatic scale, leading to 24 chords in each affective category (see www.stefan-koelsch.de/meaning_of_musical_sounds for examples of the stimuli). Chords were 800 msec long, created using Cubase (Steinberg Media Technologies GmbH, Hamburg, Germany), exported with the Grand option (piano timbre) and modified with Cool-Edit (sampling rate = 44.1 kHz; 16-bit resolution). To verify that dissonant chords possess greater roughness than consonant chords, additional analyses were carried out using an established algorithm for calculating acoustic roughness (Parncutt, 1989). The mean roughness of consonant chords was 0.139 ($SD = 0.0304$) and for dissonant chords 0.375 ($SD = 0.032$). Using a paired-sample *t* test, it was shown that the difference in roughness between consonant and dissonant chords was highly significantly different [$t(23) = -48.353, p < .0001$]. Experimental target words comprised 24 pleasant (e.g., love, joy, pleasure, courage) and 24 unpleasant (e.g., hate, disgust, fear, rage) words.

To evaluate the emotional perception of the stimulus material, a behavioral experiment was conducted with an independent group of subjects, some of which were highly

musically trained (12 years of formal musical training; $n = 20$) and untrained (no formal musical training; $n = 20$). The data showed that on a scale of 1 to 5, where 1 meant pleasant and 5 unpleasant, consonant and dissonant chords were significantly different from one another in their perceived pleasantness, which was verified using a paired-samples t test [consonant = 1.7 and dissonant = 3.9; $t(23) = 25.778, p < .0001$]. There were no group differences in the valence ratings. On average, pleasant words were 5.7 and unpleasant words 5.5 letters long (see Appendix 1). In the same rating experiment, it was established that on a scale of 1 to 5, where 1 meant pleasant and 5 unpleasant, the affective meaning of pleasant and unpleasant words was perceived to differ significantly, as indicated by a paired-samples t test [pleasant = 1.7 and unpleasant = 4.4; $t(23) = 32.135, p < .0001$].

Additionally, pleasant and unpleasant words were not found to differ in terms of the abstractness or concreteness of their content, with approximately equal number of both abstract and concrete words within and between each affective category.

Procedure

For each chord, one pleasant and one unpleasant target word were chosen, which was done randomly and altered for each participant. Each chord was played twice, followed once by a congruous word and once by an incongruous word (see also Figure 1). There were, therefore, four experimental conditions: match and mismatch conditions for pleasant chords as well as match and mismatch conditions for unpleasant chords. There were 96 trials in total, with 24 pleasant match trials, 24 pleasant mismatch trials, 24 unpleasant match trials, and 24 unpleasant mismatch trials. Trials were pseudorandomized and presented over two blocks of 48 trials.

The experiment was conducted in a sound-proof and electrically shielded cabin. Participants were seated in a comfortable self-adjustable chair facing a computer screen approximately 1.2 m away. Chords were presented from two loudspeakers positioned to the left and right of the participant. Visual targets appeared 200 msec following the onset of the chord on the screen in front. Participants were instructed to decide as fast and accurately as possible whether the meaning of the word was pleasant or unpleasant. Responses were made with a button-box, pressing left for pleasant and right for unpleasant, which was switched after the first half of the experiment. As soon as a response was made, this terminated the presentation of the chord as well as the word. A practice run preceded the experiment and was repeated if necessary.

EEG Recording and Analysis

The EEG was recorded using Ag–AgCl electrodes from 60 locations of the 10–20 system and referenced to the left

mastoid. The ground electrode was applied to the sternum. In addition, a horizontal electrooculogram was recorded, placing electrodes between the outer right and outer left canthus, for subsequent removal of eye movement-related artifacts. A vertical electrooculogram was recorded, placing an electrode above and below the right eye. Electrode resistance was kept below 5 k Ω and the EEG was recorded at a sampling rate of 500 Hz. The data were filtered off-line using a band-pass filter with a frequency range of 0.25–25 Hz (3001 points, finite impulse response) to eliminate slow drifts and reduce muscular artifacts. To remove eye movement-related artifacts, data were excluded if the standard deviation of the horizontal eye channels exceeded 25 μ V within a gliding window of 200 msec. To eliminate movement-related artifacts and drifting electrodes, data were excluded if the standard deviation exceeded 30 μ V within a gliding window of 800 msec. ERP averages were computed with a 200-msec prestimulus baseline and a 1000-msec ERP time window.

For statistical analysis, ERPs were analyzed by repeated measures ANOVA as univariate tests of hypotheses for within-subject effects. Electrodes were grouped into four separate ROIs: left anterior (AF7, AF3, F9, F7, F5, F3, FT9, FT7, FT5, FT3), right anterior (AF8, AF4, F10, F8, F6, F4, FT10, FT8, FT6, FT4), left posterior (TP9, TP7, CP5, CP3, P9, P7, P5, P3, PO7, PO3), and right posterior (TP10, TP8, CP6, CP4, P10, P8, P6, P4, PO8, PO4). To test for specific patterns of scalp distribution, anterior and posterior ROIs established the factor AntPost and left and right ROIs established the factor hemisphere. The time window for statistical analysis of the ERPs was 300–500 msec, based on visual inspection and time windows used in previous studies (Koelsch et al., 2004). Only trials in which participants had evaluated the targets correctly entered the statistical analysis. To test for an effect of prime valence on target processing, the factors prime (pleasant/unpleasant) and target (pleasant/unpleasant) were entered into the analysis separately. A significant interaction between prime and target was taken as an affective priming effect, indicating ERP differences between congruous targets and incongruous targets. For display purposes of ERPs, congruous and incongruous trials are depicted without differentiating further along valence. However, a graph is included showing mean ERP size for each of the four conditions averaged over all ROIs. After the statistical evaluation, ERPs were filtered for better legibility with a low-pass filter of 10 Hz (301 points, finite impulse response).

Results

The same paradigm has already been carried out with a group of musicians published elsewhere (Steinbeis & Koelsch, 2008a). To assess if there were any group differences between the trained musicians and a group of non-musicians, an additional factor, training, was included in the present analysis of both the behavioral and the ERP data.

Behavioral Results

The data showed that participants evaluated the affectively congruous target words faster than affectively incongruous target words (see Figure 2). This effect was effectively only present when the prime was dissonant and not when the prime was consonant. The factors prime and target were entered into a repeated measures ANOVA as within-subject and training as between-subject factor. Analysis of the reaction times revealed a significant two-way interaction between factors prime and target [$F(1, 38) = 15.46, p < .001$]. There were no interactions with the factor training or any other factors or any main effects (for all tests, $p > .6$). To check whether this interaction still holds when only analyzing the group of musically nontrained subjects, the ANOVA was run only for that group. Analysis of the reaction times revealed a significant two-way interaction between factors prime and target [$F(1, 19) = 17.88, p < .001$]. There were no interactions with any other factors or any main effects (for all tests, $p > .7$).

As an additional analysis, both the congruent and the incongruent trials were analyzed as a single factor congruence in a repeated measures ANOVA. The analysis of the reaction times revealed a significant effect of congruence [$F(1, 19) = 26.813, p < .0001$].

The analysis of performance accuracy revealed a high performance of 98.9%. There were neither significant interactions nor any main effects, showing that error rates were not sensitive to the relationship between valence of prime and target.

ERP Results

The ERP data reveal a larger N400 for incongruous targets words than for congruous target words. This effect was glob-

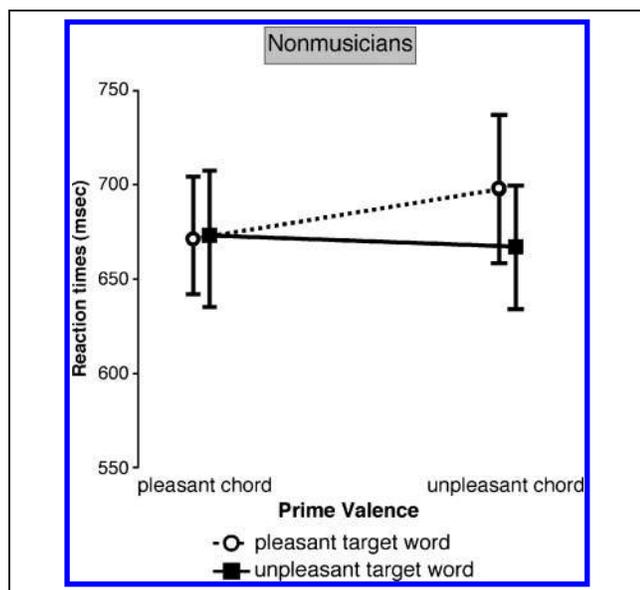


Figure 2. Experiment 1: Mean reaction times (± 1 SEM) for evaluative decisions on pleasant and unpleasant word targets.

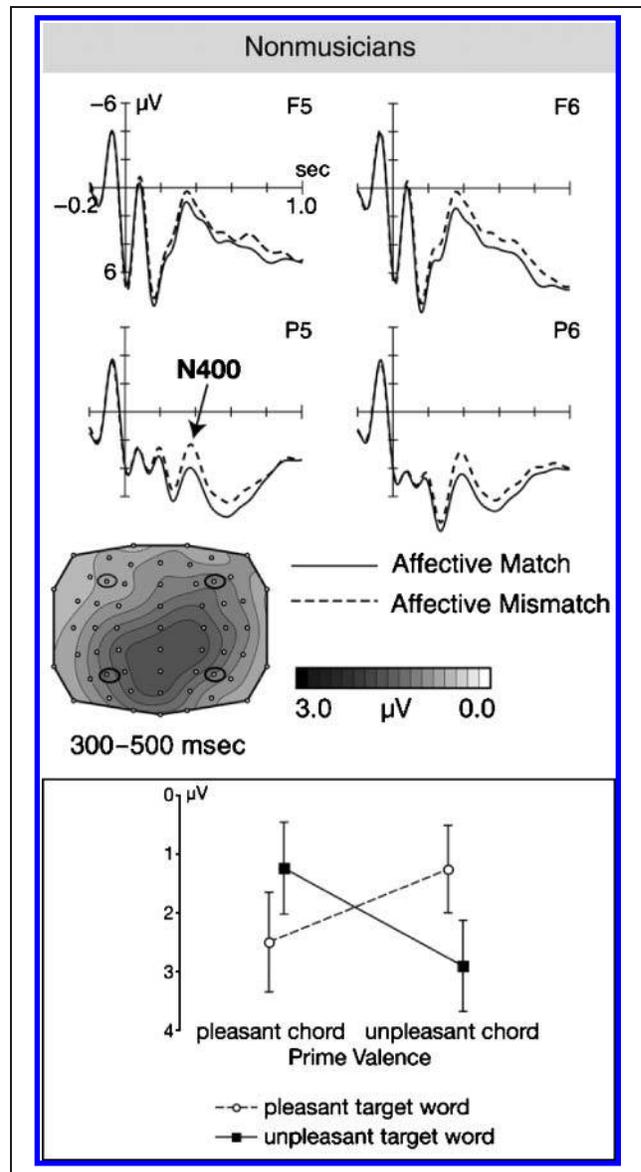


Figure 3. Experiment 1: ERPs locked to onset of the target word. ERPs in response to an affective mismatch between prime and target valence (dashed line) resulted in a larger N400 between 300 and 500 msec compared to ERPs in response to a prime–target match (solid line). The effect is distributed broadly over the scalp. The inlaid box displays mean ERPs over all ROIs between 300 and 500 msec for each condition.

ally distributed and maximal between 300 and 500 msec. Analysis of the ERPs in the time window of 300–500 msec for both groups revealed a significant two-way interaction between factors prime and target [$F(1, 38) = 20.65, p < .001$], indicating a larger N400 for incongruous target words compared to congruous target words. There were no interactions with the factor training or any other factors or any main effects (for all tests, $p > .8$). To check whether this interaction still holds when only analyzing the group of nonmusicians, the ANOVA was run only for that group and showed a significant two-way interaction between factors prime and target [$F(1, 19) = 10.82, p < .01$]. Despite a visually suggestive larger effect over posterior regions, this

was not borne out statistically (see Figure 3). There was no interaction with the factors AntPost or hemisphere or any other significant main effects (for all tests, $p > .5$). ANOVAs for earlier (100–300 msec) as well as later (500–700 msec and 700–900 msec) time windows revealed no main effects or interactions (for all tests, $p > .6$).

For reasons outlined above, the present group represents a more representative segment of the population than trained musicians. The data are therefore discussed below.

Discussion

Participants showed a larger N400 for the incongruous target chords compared to the congruous target chords, which in turn was accompanied by a behavioral effect, whereby congruous word targets were evaluated significantly faster than incongruous word targets. This behavioral effect, however, was only present when the prime was dissonant and not when consonant.

Seeing that the N400 has been taken to indicate semantic processing, the present findings suggest that harmonic roughness is capable of communicating affectively meaningful signals. This indicates that harmonic roughness already communicates meaningful information (Koelsch & Siebel, 2005), which in turn can transfer onto the processing of other meaningful concepts. Whereas this has already been demonstrated for a set of highly musically trained subjects (Steinbeis & Koelsch, 2008a), the present findings extend this to a group without any formal musical training, a more representative sample of the population.² Thus, irrespective of musical training, basic musical features appear to be able to communicate meaning.

This discrepancy between the ERP and the reaction time data with regard to processing the lexical items after consonant chords suggests that even though the brain appears to process the different affective meaning, this is not reflected in the behavior. This would imply that ERP measures are more sensitive to the difference in affective meaning between words and their relationship to a previously built up affective context. The reasons for this discrepancy are so far unclear and require further investigation. However, the present data demonstrate that the affective context established by chords varying in harmonic roughness is capable of influencing the subsequent processing of lexical affect, leading to integration costs in the case of an incongruent affective pairing (as indicated by the presence of the N400) and an effect in the reaction times between pleasant and unpleasant words after dissonant chords.

The absence of an affective priming effect for accuracy of responses can be accounted for by the very high performance producing a ceiling effect. The task was relatively easy and accuracy may not have been sensitive to the congruency of prime–target pairs (for similar reasoning, see Schirmer & Kotz, 2003). The fact that this effect could be observed for both musically trained (see Steinbeis & Koelsch, 2008b) and untrained participants (as indicated by a nonsignificant difference between the two groups re-

ported above) suggests that expertise does not modify the processing of affectively semantic properties contained in basic features of the auditory input. This appears to be in line with some previous findings, where both musicians and nonmusicians were equally able to correctly classify the emotion of a musical piece, based on no more than 1 sec of the music (Bigand et al., 2005).

Several mechanisms have been proposed to account for the various behavioral priming effects found in the literature (for reviews, see McNamara, 2005; Musch & Klauer, 2003; Neely, 1991), such as spreading activation, expectancy-based priming, and semantic matching. The first of these is a mechanism argued to operate automatically, whereby the representation of each entry in the mental lexicon is connected to words closely related in meaning. The activation of one entry will automatically spread to activate closely related words. Compared to both expectancy-based priming (whereby subjects generate a list of words possibly connected to the prime) and semantic matching (whereby the subject scans preceding information when coming across a new item), which are both highly controlled processes, spreading activation is fast-acting, of short duration, and does not require attention or awareness (Shiffrin & Schneider, 1977; Collins & Loftus, 1975). Affective priming typically functions only at SOAs at or below 200 msec, which has also been argued to reflect the automatic nature of affective processing and evaluative decisions (Musch & Klauer, 2003). Thus, spreading activation would appear to be a likely mechanism, which can explain the observed priming effects. Therefore, single chords may activate affective representations, which spread onto affectively related representations, in this case, affective target words. This hypothesis of purported underlying mechanisms could be tested by varying the SOA and observing the persistence or absence of the effects. In addition, any final claims on the automaticity of the present effects can only be established by means of an implicit task, which was not the case in the present experiment. Thus, although we believe that given the short SOA and previous literature on affective priming (Musch & Klauer, 2003) that the present effects constitute those of automatic priming, this has yet to be established in further empirical study.

This discussion of the possible mechanisms underlying the presently observed effect allow for a proper contextualization of the findings. As has been argued above and shown in previous studies, certain psychoacoustic properties, such as consonance and dissonance, give rise to the expression of certain emotional qualities (i.e., pleasantness or unpleasantness). These qualities are presumably quickly recognized by the auditory system by means of a general mechanism dedicated to decode the affective information contained in acoustic information. This information is then classified into its affective category (the specificity of which is still unclear), and by virtue of its classification, which coactivates related affective concepts, represents something meaningful to the listener (i.e., an emotional concept), which in turn can influence subsequent semantic processing.

At present, no clear answer can be given for whether the observed behavioral priming effect is the result of facilitated responses to the matched target words, an inhibitory effect to the mismatched target words, or a mixture of both. It is well documented that inhibition is small or nonexistent for SOAs shorter than 300 msec (McNamara, 2005; de Groot, 1984; Neely, 1977), which indicates that the present priming effect is the result of facilitated responses on matched target words. Even though an investigation using neutral target words would be required to adequately respond to this issue, it has been shown that the use of these in affective priming studies does not always constitute a reliable baseline (Schirmer & Kotz, 2003).

The use of an evaluative decision task implies conflict at the response level in addition to the one at the level of affective meaning, as the evaluative decision requires giving incongruous responses for incongruous trials (Wentura, 2000). Thus, incongruous targets represent a mismatch on a higher level (affective meaning) as well as a lower (response) level. Whereas it cannot be ruled out that the observed behavioral effect can be accounted for by a basic response conflict/tendency explanation, the neural data suggest that this alternative account cannot be supported. ERP studies of stimulus–response conflict typically report a fronto-central N200 component (340–380 msec), presumably generated in caudal anterior cingulate cortex (cACC; van Veen & Carter, 2002). The present data, however, suggest a distinctly different ERP component, both by virtue of its latency as well as its distribution, which is highly reminiscent of the N400 typically found for semantic violations. In addition, the present effect was significant only in the time window of 300 to 500 msec and did not extend into later time windows despite its visual appearance, which suggests that this component is unlikely to constitute a general mismatch mechanism that is likely to extend into later time windows. Also, recent data suggest that the conflict represented by the presently used paradigm does not recruit cACC, which would imply the engagement of areas processing conflict at the response level, but rather the right inferior frontal gyrus as well as the middle temporal gyrus in musicians (Steinbeis & Koelsch, 2008a), the latter of which has been consistently linked with the processing of meaning in a wide variety of domains (Koelsch et al., 2004; but see Patterson, Nestor, & Rogers, 2007 for a thorough review). Thus, although the behavioral effect may, in part, be explained by stimulus–response conflict, the underlying neural processes still suggest that acoustic roughness is a musical feature which is capable of both communicating affective meaning and priming the processing of subsequently presented affective words. Whereas the evaluative decision task is the most frequently used task for affective priming measures, one may also wish to try other more implicit tasks, such as lexical-decision tasks, which have been successfully used for semantic priming studies (McNamara, 2005).

To see if the presently observed effect can be replicated for other musical features, this was tested in two further experiments.

EXPERIMENT 2: MODE (MAJOR–MINOR)

Introduction

The aim of this experiment was to test whether major and minor chords can also communicate affect and thereby influence subsequent processing of word meaning. This was tested to investigate effects of more fine-grained pitch-interval analysis on semantic processing: Detailed processing of the pitch relations between the tones of a chord is required to determine whether the chord is a major or minor chord, and such processing appears to involve both posterior and anterior areas of superior temporal cortex bilaterally (for details, see Koelsch & Siebel, 2005). The employment of this condition enabled us to compare effects of auditory feature extraction (decoding of acoustic roughness) and of the decoding of intervals on semantic processing. The term pitch interval is used here to refer to the subtle difference between a major and a minor chord. The difference between the interval superposition of a major chord consists of a major and a minor third, whereas that of a minor chord consists of a minor and a major third. Decoding whether a chord is major or minor thus requires a relatively fine-grained analysis of pitch intervals.

The experimental literature on a link between major/minor mode and emotion has a long history. An early study showed that major pieces of music are classified more often as happy than music pieces in a minor key, and minor pieces are classified more often as sad than major pieces (Hevner, 1935). Recently, this has found further empirical support in studies designed to investigate whether emotion conveyed by music is determined most by musical mode (major/minor) and tempo (slow/fast; e.g., Gagnon & Peretz, 2003). Using the same set of equitone melodies, participants had to judge whether they sounded happy or sad. It was found that musical mode was a highly significant predictor for listeners' judgments, with major melodies being rated significantly more often as happy than minor melodies, which in turn were rated significantly more often as sad than major melodies. Thus, there is considerable evidence in favor of major/minor mode to communicate affective meaning such as happiness and sadness. It was therefore hypothesized that target words congruous with the emotional connotation of a musical prime stimulus (major chord = happy, minor chord = sad) would elicit a smaller N400 than incongruous target words. In addition, the evaluative decision on congruous target words was hypothesized to be faster and more accurate than on incongruous target words.

Methods

Participants

Twenty musically trained (10 women) and 20 musically untrained (10 women) volunteers participated in the experiment. On average, musically trained participants were 22.7 years of age ($SD = 3.82$) and musically untrained participants were 22.88 years of age ($SD = 2.44$). Musicians

had received approximately 12 years of musical training (mean = 12.32; $SD = 4.33$; all of them played the piano and most of them string instruments). All subjects were right-handed, native German speakers, with normal or corrected-to-normal vision, and no hearing impairments.

Materials

The prime stimulus material consisted of 48 chords, of which 24 were in a major chords and, therefore, happy sounding, and of which 24 were minor and, therefore, sad sounding. Major chords were presented either in root position (e.g., C–E–G–C) or as six–four chords (e.g., G–C–E–G). Analogously, minor chords were presented in root position (e.g., C–E flat–G–C), or as six–four chords (e.g., G–C–E flat–G). Both major and minor chords were played in each of the 12 keys of the chromatic scale, leading to 24 chords in each affective category (see www.stefan-koelsch.de/meaning_of_musical_sounds for examples of the stimuli). Chords were created spanning an octave, which ranged from C4 to C5. Sound files were created in Cubase (Steinberg Media Technologies GmbH, Hamburg, Germany), exported with the Grand option and modified using Cool-Edit (sampling rate = 44.1 kHz, 16-bit resolution). Chords were 800 msec long. The difference in roughness between major and minor chords was also calculated using the same procedure as in Experiment 1. The mean roughness of major chords was 0.139 ($SD = 0.0304$) and of minor chords 0.164 ($SD = 0.04$). The difference in roughness between consonant and dissonant chords was highly significantly different as indicated by a paired-samples t test [$t(23) = -6.536, p < .0001$].

Experimental target words comprised 24 words with a happy (e.g., success, gift, jest, fun) and 24 words with a sad (e.g., loss, tear, misery, woe; see Appendix 2) meaning. On average, happy words were 5.4 letters and sad words 5.3 letters long.

A previous rating experiment conducted with both musically trained and untrained participants established that, on a scale of 1 to 5, where 1 meant happy and 5 sad, major and minor chords were significantly different from one another in their perceived happiness, which was verified using a paired-samples t test [major = 2.5 and minor = 3.6; $t(23) = 12.489, p < .0001$]. There were no group differences in the happiness ratings. Additionally, major and minor chords were rated by both groups on the pleasantness/unpleasantness dimension and were found not to differ significantly as shown by paired-samples t test (major = 2.56 and minor = 2.61; $p > .3$), again showing no difference between the groups. It was also established that on a scale of 1 to 5, where 1 meant happy and 5 sad, the affective meaning of happy and sad words was perceived to differ significantly as indicated by a paired-samples t test [happy = 1.6 and sad = 4.4; $t(23) = 40.921, p < .0001$]. There were no group differences in the happiness ratings. Additionally, happy and sad words were found to not differ in terms of the abstractness or concreteness of their con-

tent, with approximately equal numbers of abstract and concrete words in each affective category.

For each chord, one happy and one sad target word were chosen, which was done randomly and altered for each participant. Thus, each chord was played twice, followed once by a congruous word and once by an incongruous word. There were, therefore, four experimental conditions: match and mismatch conditions for happy chords as well as match and mismatch conditions for sad chords, with 96 trials in total (24 happy match trials, 24 happy mismatch trials, 24 sad match trials, and 24 sad mismatch trials). Trials were pseudorandomized and presented over two blocks of 48 trials each. Participants were instructed to decide as fast and accurately as possible whether the meaning of the word was happy or sad.

Procedure, ERP recording, and data analysis were the same as in Experiment 1. Seeing that both musicians and nonmusicians participated in the experiment, the additional between-subject factor, training (musically trained/musically untrained), entered all statistical analyses to test for any differences resulting from musical training.

Results

Behavioral Results

The data showed that participants evaluated the affectively congruous target words faster than affectively incongruous target words (see Figure 4). The factors prime and target were entered into a repeated measures ANOVA in addition to the between-subject factor training. Analysis of the reaction times revealed no significant three-way interaction between the factors prime, target, and training, ($p > .6$), but a significant two-way interaction between factors prime and target [$F(1, 38) = 12.11, p < .001$]. There were no further interactions or main effects (for all tests $p > .6$). It has to be pointed out that although there was no statistical difference between the groups in their behavioral effect, musicians do appear to show a somewhat stronger effect than the nonmusicians when happy chords were used as primes. This difference was only nominal, however.

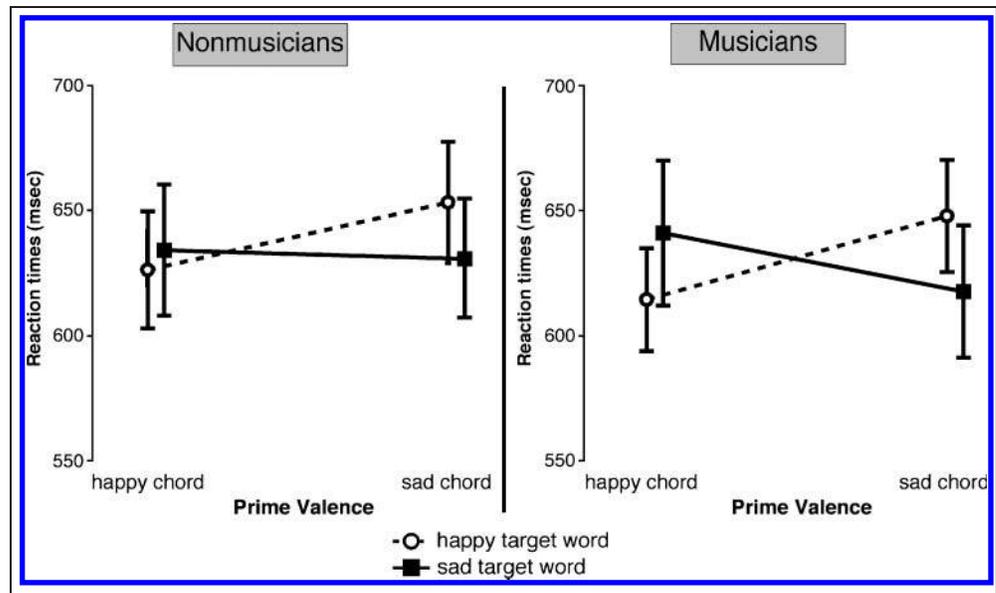
As an additional analysis, both the congruent and the incongruent trials were analyzed as a single factor congruence in a repeated measures ANOVA. The analysis of the reaction times revealed a significant effect of congruence [$F(1, 38) = 15.321, p < .001$].

The analysis of performance accuracy revealed high performance of both groups (musicians: 97.7%; nonmusicians: 96.2%). There were neither significant interactions nor any main effects, showing that error rates did not differ between groups nor were they sensitive to the relationship between valence of prime and target (for all tests, $p > .5$).

ERP Results

The ERP data reveal a larger N400 for incongruous target words than for congruous target words. This effect

Figure 4. Experiment 2: Mean reaction times (± 1 SEM) for evaluative decisions on happy and sad word targets.



was globally distributed and maximal between 300 and 500 msec. Analysis of the ERPs in the time window of 300–500 msec revealed no significant three-way interaction between the factors prime, target, and training ($p > .8$), but a significant two-way interaction between factors prime and target [$F(1, 38) = 33.69, p < .0001$], indicating a larger N400 for incongruous target words compared to congruous target words for both musically trained and musically untrained participants with a broad scalp distribution (see Figure 5). There were no further interactions with any of the other factors, nor were there any significant main effects. ANOVAs for earlier (100–300 msec) as well as later (500–700 msec and 700–900 msec) time windows revealed no main effects or interactions (for all tests, $p > .7$).

Discussion

Both groups of participants showed a larger N400 for target words that mismatched in valence with the preceding chord prime, which was accompanied by a significant behavioral priming effect. These findings strongly suggest that even a brief presentation of musical mode is capable of communicating the expression of an emotion, which influences the subsequent processing of verbally presented affective information. Because the affective information is encoded in the meaning of the word, these findings can be taken to imply that musical mode can affect the processing of language meaning on an affective level. It is striking that as little as the manipulation of one semitone (the difference between major and minor chords) is sufficient to communicate affective meaning. This shows that the analysis of pitch intervals (required to differentiate major from minor chords) is linked to establishing meaning in music and lends empirical support for such an idea ex-

pressed in previous theoretical outlines (Koelsch & Siebel, 2005). This effect was observed for both musically trained and untrained participants, which provides further evidence to that obtained in Experiment 1, suggesting that expertise does not modify the processing of affectively semantic properties contained in basic musical features. There was no priming effect found in the accuracy of responses, which fits with the data reported in Experiment 1. Similarly, the task may have been too easy, and thus, produced ceiling effects, whereupon the accuracy scores would be insensitive to the prime–target relationship.

It may be argued that this experiment is merely a replication of Experiment 1 because the manipulation of a harmonic interval automatically entails a manipulation of harmonic roughness (as indicated by the calculation described in the Methods). A previous rating experiment, however, indicated that, whereas major and minor chords were perceived as differing on the happy/sad dimension, they were not rated as significantly different on the pleasant/unpleasant dimension (see Methods). This can be taken as evidence that even though harmonic roughness was manipulated, the manipulation of a semitone suggested a different or perhaps additional affective meaning to that conveyed by harmonic roughness (i.e., happy/sad as opposed to pleasant/unpleasant). In addition, the roughness scores procured by the analyses showed that the difference in roughness was far greater between consonant and dissonant chords (mean = 0.2354; $SD = 0.0238$) than between major (consonant) and minor chords (mean = 0.0228; $SD = 0.017$) [$t(23) = 66.55, p < .0001$]. There were no differences in the size of the behavioral effect or the ERPs between Experiments 1 and 2, which therefore suggest that fine pitch discrimination is a more likely candidate to have led to the effects observed in the present experiment, rather than the decoding of acoustic roughness. This, in turn, suggests that the present experiment provides

evidence for subtle differences in harmonic intervals to be capable of communicating affective meaning not significantly mediated via harmonic roughness.

Similar to Experiment 1, one may level a rival account of the present effects in terms of a conflict at the response level, given the nature of the explicit task. As has been argued, however, the ERP typically associated with such a conflict is the N200, which, although also a negativity, is distinct from the N400 both in terms of distribution and latency. The present ERP bears the classical hallmarks of the N400 in being broadly distributed over the scalp and only significant between 300 and 500 msec. Although the neural generators have not been directly assessed in the present experiment, it is assumed that the middle temporal gyrus rather than rACC is involved in the operation of this task. This, however, is merely based on fMRI data of a related but not an identical experiment (Steinbeis & Koelsch, 2008a). Ideally, this experiment would be carried out using an implicit task, whereupon alternative accounts in favor of a response-conflict could be ruled out on methodological grounds.

EXPERIMENT 3: TIMBRE

Introduction

The aim of this study was to see if instrumental timbre is also capable of communicating affect and thereby influencing the processing of subsequent word meaning. So far, there is little literature on a relationship between instrumental timbre and emotion (but for a review, see Juslin & Laukka, 2003) and virtually none systematically exploring which aspects of timbre may link with the expression and perception of emotion, by explicitly manipulating this. This may have partly to do with the fact that timbre has been difficult to define empirically. Definitions of musical timbre have commonly been made more in terms of what it is not, rather than what it is, whereby it was argued that timbre refers to those aspects of sound quality other than pitch, loudness, perceived duration, spatial location, and reverberant environment (Menon et al., 2002; McAdams, 1993). Generally, it seems to have been agreed on that timbre is the tonal color or texture that allows one to distinguish the same note played by two different instruments.

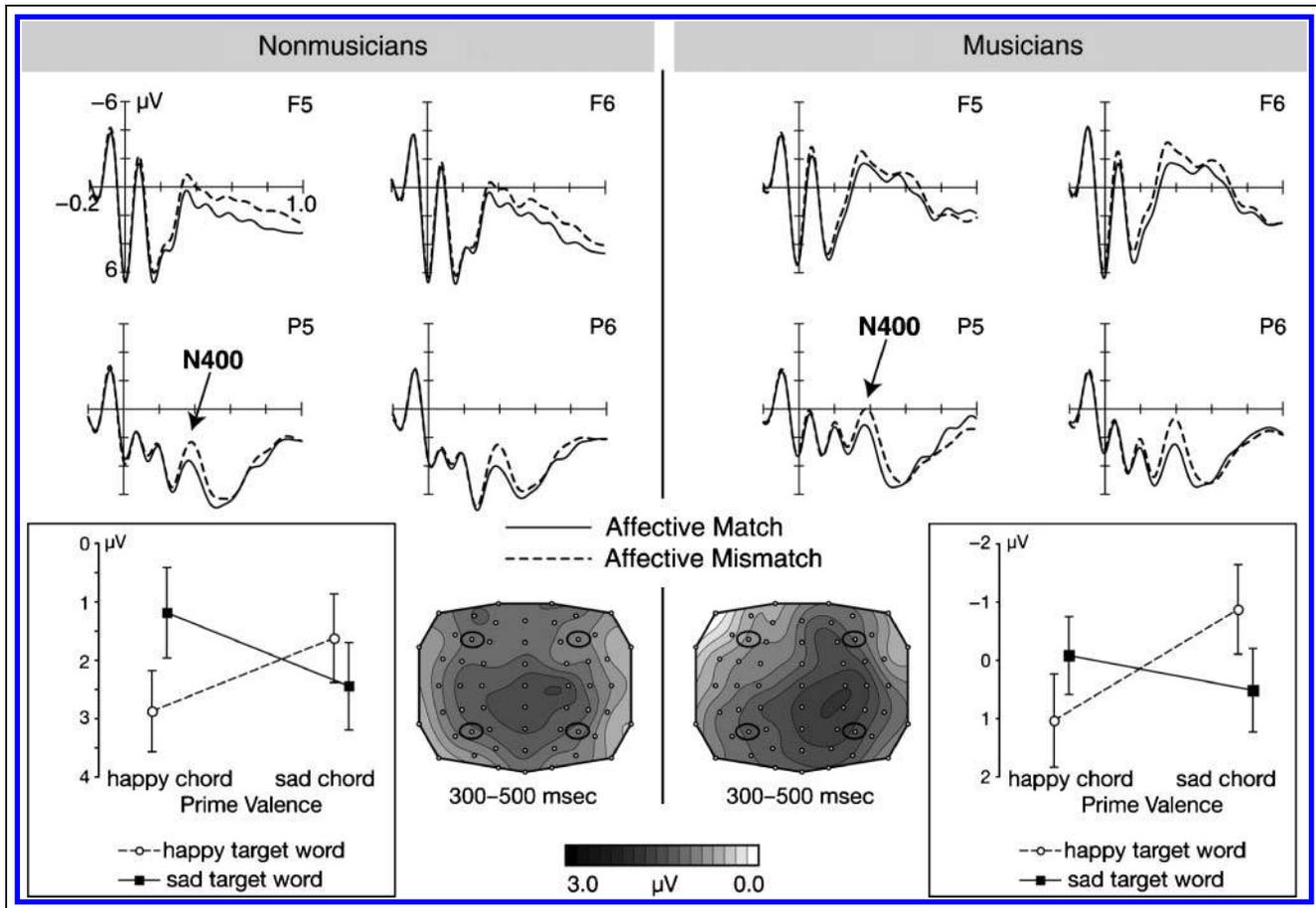


Figure 5. Experiment 2: ERPs locked to onset of the target word. ERPs in response to an affective mismatch between prime and target valence (dashed line) resulted in a larger N400 between 300 and 500 msec compared to ERPs in response to a prime–target match (solid line). Both musicians and nonmusicians show the effect, which is broadly distributed over the scalp. The inlaid boxes display mean ERPs over all ROIs between 300 and 500 msec for each condition.

By use of multidimensional scaling (MDS), a variety of psychoacoustic parameters have been identified to correlate with the perception of different timbres. These have included attack time of the sound, as well as various spectral parameters, such as the spectral centroid (which measures the average frequency of a spectrum, weighted by amplitude), the spectral flux (a measure of the change within a frequency spectrum over the duration of the signal), as well as the spectrum fine structure (which measures the attenuation of even harmonics of the sound signal; McAdams, Winsberg, Donnadieu, Soete, & Krimphoff, 1995). A recent study has shown that in dissimilarity ratings, attack time, spectral flux, and spectrum fine structure are used most saliently to differentiate between different timbres (Caclin, McAdams, Smith, & Winsberg, 2005), however, more work is required to fully understand which parameters are used for the perceptual analysis of timbre.

There is some evidence in the literature on speech prosody that the timbral quality of vocal utterances (e.g., the distribution of energy in a frequency spectrum) is strongly related to the communication of an emotion: In a study which set out to test to what extent listeners can infer emotions from vocal cues, Banse and Scherer (1996) recorded the vocalization of 14 different emotions as portrayed by professional actors, which were then classified blindly by judges and underwent a psychoacoustic analysis. Specific psychoacoustic patterns could be identified with each vocalized emotion and it was shown that the expression and perception of anger, fear, sadness, and joy partly depend on the relative energy in the high- versus low-frequency spectra, which is known to characterize vocal timbre.

To date, the only explicit systematic exploration of the perception of emotion in musical timbre is a study using the mismatch negativity (Goydke, Altenmüller, Möller, & Münte, 2004). Violin timbres differing in emotional expression (happy or sad) were used to create “emotional” standards and deviants. It was found that emotional deviants elicited an MMN, which was interpreted by the authors as the fast and accurate perception of emotional timbre. However, the study confounded perceptual differences and emotional expression, as happy and sad timbres differed in terms of basic perceptual features. Thus, apart from the ratings taken prior to the experiment, this does not constitute a clear piece of evidence that musical timbres can communicate emotions.

Although there is no direct evidence on the perception of emotions expressed in musical timbre, the work on timbre in vocal productions suggests that this feature of the acoustic input may be more generally capable of expressing emotions. The present study employed chords of two types of timbres, one subjectively pleasant and another subjectively unpleasant to investigate whether the expressed pleasantness of musical timbre has an influence on subsequent semantic processing. Given the results of our previous experiments, it was hypothesized that target words congruous with the timbre-dependent valence of a musical prime should elicit a smaller N400 than incon-

gruous target words. In addition, the evaluative decision on congruous target words was hypothesized to be faster and more accurate than on incongruous target words.

Methods

Participants

Fifteen musically trained (8 women) and 18 musically untrained (10 women) volunteers participated in the experiment. On average, musically trained participants were 25.4 years of age ($SD = 3.77$) and musically untrained participants were 24.6 years of age ($SD = 4.01$). Musicians had received approximately 14.6 years of musical training (mean = 14.61; $SD = 5.21$; all of them played the piano and most of them string instruments). All subjects were right-handed, native German speakers, with normal or corrected-to-normal vision and no hearing impairments.

Materials

The prime stimulus material consisted of 48 major chords, of which 24 had a pleasant and 24 an unpleasant musical timbre (see www.stefan-koelsch.de/meaning_of_musical_sounds for examples of the stimuli). The major chords were presented either in root position (in C: C–E–G–C) or as six–four chords (in C: G–C–E–G) and played in each of the 12 keys. Chords were created in Cubase (Steinberg Media Technologies GmbH, Hamburg, Germany). Pleasant sounding chords were exported with the Grand option of Cubase and sounded like normal chords played on a piano. Unpleasant sounding chords were exported with the tin-drum option of Cubase and sounded considerably harsher and unpleasant. They were then modified using Cool-Edit (sampling rate = 44.1 kHz, 16-bit resolution). Chords were 800 msec long.

To allow for a more fine-grained (but by no means exhaustive) analysis, which parameters may be relevant for the perception of pleasantness in instrumental timbre, the presently used stimuli were analyzed with regards to two parameters relevant for perceiving timbre: attack time and spectral centroid. Attack time was calculated by extracting the root-mean-square (RMS) of the signal over a time window of 10 msec, with a gliding window of 1 msec of the entire signal. RMS is a statistical measure of the magnitude of a varying quantity (in this case, amplitude of a sound) and derived by means of the following formula: $\sqrt{[x]^2}$, where x denotes the arithmetic mean. The time from the beginning of the sound to the maximum RMS was calculated, which constituted the attack time. The spectral centroid measures the average frequency, weighted by amplitude of a spectrum. The standard formula for the average spectral centroid of a sound is: $c = \sum c_i/i$, where c_i is the centroid of one frame and i is the number of frames for the sound. A spectral frame is the number of samples, which given the present stimuli was 44.1 kHz. The frequency spectrum of each sound was calculated, by means of a Fast-Fourier-Transformation (FFT), with

a size of 2048 sampling points to obtain an optimal estimate of spectral resolution given the present sampling rate of the signal. Perceptually, the spectral centroid has been associated with the brightness of a sound (Schubert, Wolfe, & Tarnopolsky, 2004).

Attack time and spectral centroid were calculated for each chord, averaged for each timbre and compared using paired-samples *t* tests. The attack time was found to differ significantly between the pleasant (177 msec) and the unpleasant timbre (115 msec) [$t(23) = 4.85, p < .001$]. The spectral centroid was also found to differ between the two timbres and was considerably lower for the pleasant (402 Hz) than for the unpleasant timbre (768 Hz) [$t(23) = -10.79, p < .001$]. Chords with the unpleasant timbre appear to have a significantly earlier attack time as well as a brighter sound compared to chords with the pleasant timbre.

A previous rating experiment conducted with both musically trained and untrained participants established that on a scale of 1 to 5, where 1 meant pleasant and 5 unpleasant, piano-timbre and tin-drum-timbre chords were significantly different from one another in their perceived pleasantness as shown by a paired-samples *t* test [piano = 1.7 and tin-drum = 4.3; $t(23) = 38.224, p < .0001$]. There were no group differences in the pleasantness ratings.

Experimental target words, matching, randomization, and presentation procedures, as well as ERP recording and data analysis, were the same as in Experiment 1 (see Appendix 1).

Results

Behavioral Results

The data showed that participants evaluated the affectively congruous target words faster than affectively incongruous

target words (see Figure 6). The factors prime and target were entered into a repeated measures ANOVA in addition to the between-subject factor training. Analysis of the reaction times revealed no significant three-way interaction between the factors prime, target, and training ($p > .7$), but a significant two-way interaction between factors prime and target [$F(1, 33) = 32.17, p < .0001$], suggesting that both musically trained and untrained participants evaluated affectively congruous target words faster than affectively incongruous target words (see Figure 6). There were no further interactions or main effects (for all tests $p > .6$). Despite this, it is worth pointing out that whereas the effect for the nonmusicians is small when preceded by the pleasant chord and large when preceded by the unpleasant chord, this pattern is reversed for the musically trained participants.

As an additional analysis both the congruent and the incongruent trials were analyzed as a single factor congruence in a repeated measures ANOVA. The analysis of the reaction times revealed a significant effect of congruence [$F(1, 33) = 32.488, p < .0001$].

The analysis of performance accuracy revealed high performance of both groups (musicians: 96.9%; nonmusicians: 94.9%). There were neither significant interactions nor any main effects, showing that error rates did not differ between groups nor were they sensitive to the relationship between valence of prime and target.

ERP Results

The ERP data reveal a larger N400 for incongruous target words than for congruous target words (see Figure 7). This effect was globally distributed and maximal between 300 and 500 msec. Analysis of the ERPs in the time window of 300–500 msec revealed no significant three-way

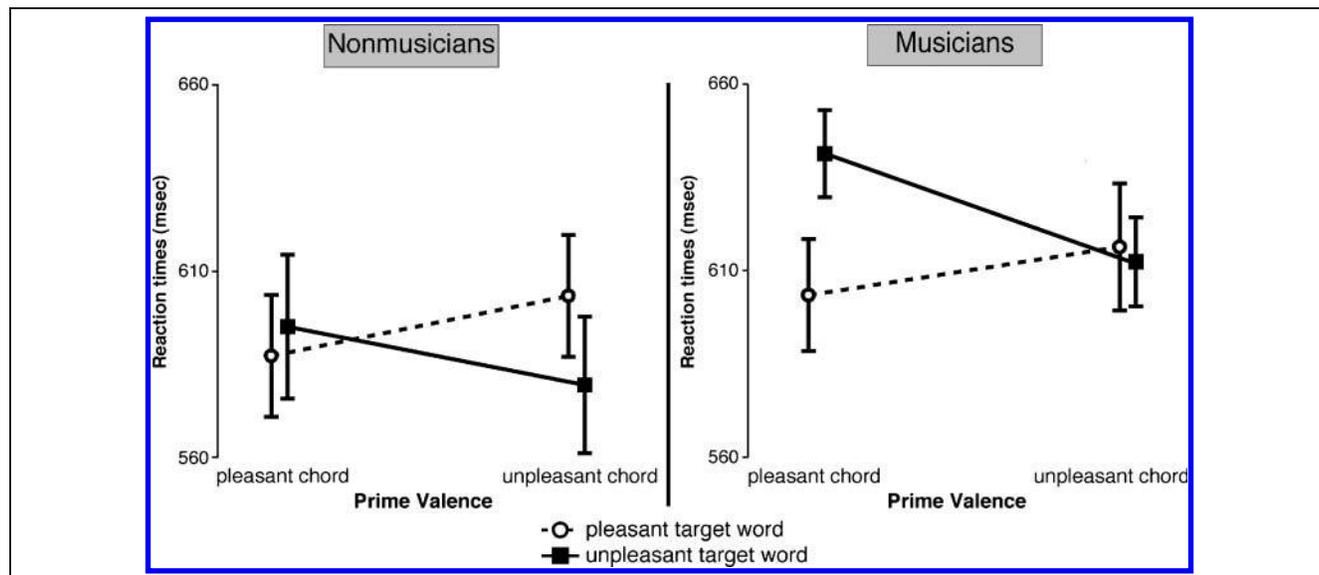


Figure 6. Experiment 3: Mean reaction times (± 1 SEM) for evaluative decisions on pleasant and unpleasant word targets.

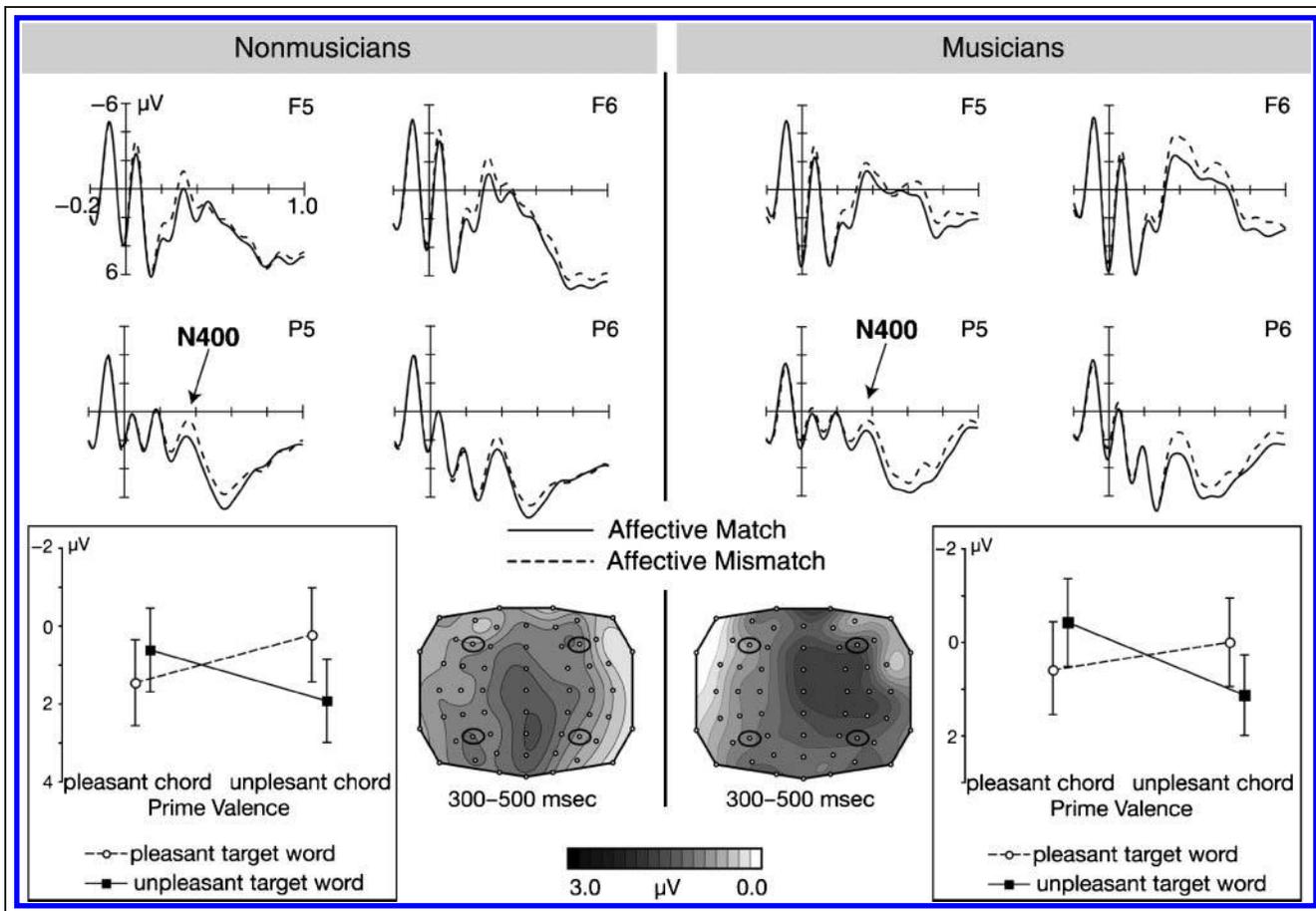


Figure 7. Experiment 3: ERPs locked to onset of the target word. ERPs in response to an affective mismatch between prime and target valence (dashed line) resulted in a larger N400 between 300 and 500 msec compared to ERPs in response to a prime–target match (solid line). Both musicians and nonmusicians show the effect, which is broadly distributed over the scalp. The inlaid boxes display mean ERPs over all ROIs between 300 and 500 msec for each condition.

interaction between the factors prime, target, and training ($p > .4$), but a significant two-way interaction between factors prime and target [$F(1, 33) = 17.88, p < .001$], indicating a larger N400 for incongruous target words compared to congruous target words for both musically trained and musically untrained participants with a broad scalp distribution (see Figure 7). There were no further interactions with any of the other factors or any significant main effects. Despite of suggestive visual evidence of further ERP effects in later time windows, these could not be statistically confirmed (for all tests, $p > .8$).

Comparison of ERPs for All Three Experiments

To test for differences in the N400 effect elicited by each of the three parameters, we conducted an additional ANOVA in the time window of 300–500 msec with the within-subject factors prime, target, and the between-subject factors training and experiment over all regions of interest. There was no significant interaction between the factors prime, target, and experiment, indicating that the N400 did

not differ between the three experiments, neither in amplitude nor distribution.

Discussion

This study demonstrates that timbre appears to be capable of communicating affect and the perception of which can transfer onto the subsequent processing of affective meaning at the word level. Both musically trained and untrained participants showed a larger N400 for target words mismatched in valence to the preceding chord prime compared to the matched target words, which was accompanied by a behavioral priming effect. This provides some support for a link between the emotional expression of instrumental timbre and the establishment of meaning in music, as has been hypothesized (Koelsch & Siebel, 2005).

As in the first two experiments, no differences resulting from musical training were found, neither in the behavioral data nor in the ERP data, suggesting once more that musical expertise does not modify the processing of affectively meaningful properties contained in basic features of

the auditory input. Similarly to the preceding two experiments, no priming effect was found in the accuracy of responses. As the accuracy data show, the task was very easy for participants, and accuracy may therefore not have been sensitive to the prime–target relationship due to ceiling effects.

The present study shows that different musical timbres can lead to a differential perception in terms of their emotional expression and these findings constitute one of the first pieces of evidence that this is the case. It must be acknowledged that timbre is a psychoacoustically challenging phenomenon and its perception consists of a wide variety of different factors. It has been shown that the perception of instrumental timbre depends on more than one psychoacoustic property (Caclin et al., 2005) and therefore it cannot be said with any degree of certainty that any single one of those argued to be involved (e.g., attack time and spectral centroid) is uniquely responsible. It may even be, that neither attack time nor the spectral centroid have any bearing on the perceived affect, but rather another acoustic dimensions not explicitly assessed in the present experiment, such as spectral flux or the spectral fine structure. Differences in timbre are perceived on the basis of the relative interplay of these psychoacoustic properties and it is presumably the unique composition arising out of these, which make a timbre more or less pleasant sounding. Future work may want to focus more on the systematic exploration of each of the relevant (and further) psychoacoustic factors, and to investigate to what extent these may be responsible for the perceived valence of instrumental timbre. Ideally, sounds would undergo parametric modulation for each of the parameters known to influence the perception of timbre and the effects on how pleasant or unpleasant they sound could be measured by explicitly asking participants or by means of a more implicit paradigm, such as the present one, where the attention is directed toward the lexical targets. Not knowing exactly which factors play a role in the perception of affect expressed by different timbres represents a limitation of the present study. However, the present data provide evidence that manipulating the timbre of musical sounds has an effect on the perceived affect of the sound, which in turn, is capable of influencing the subsequent processing of emotional word meaning.

Like in the first two experiments, given the nature of the task, a response-conflict could be argued to explain the presently observed effects. This would mean, however, finding an N200 typically peaking between 340 and 380 msec with a fronto-central scalp distribution. The present component is, like in the other two experiments, widely distributed, maximal at 400 msec and significant between 300 and 500 msec only. In addition to the purported neural generators, which as argued above, are likely to be located in the medial temporal gyrus, we would argue that the presently observed effect constitutes an N400, which typically reflects the integration costs at the semantic level, brought about by a mismatch in affective expression

of a musical prime and the emotional meaning of a subsequently presented word.

GENERAL DISCUSSION

To test the hypothesis whether the emotion that can be expressed by various musical constituents can convey information that is capable of influencing the processing of meaning, several cross-modal priming experiments were conducted. Similar to semantic priming studies, single chords were followed by target words which either matched or did not match the affect communicated by the chord. It was found that target words matching in valence to the preceding chord were evaluated faster than mismatched target words. In addition, ERP results showed an increase in the N400 amplitude in response to mismatched target words. The N400 has been seen as a classical index of semantic processing (Kutas & Federmeier, 2000), and therefore, the present findings indicate that the emotion signaled by musical features can interfere with processing word meaning, which in turn suggests that musical features can communicate basic affective signals, which are understood as meaningful.

The set of studies demonstrate a link between the emotional expression of musical features and meaning using several musical parameters previously known to elicit emotions in listeners. Consonance/dissonance, mode (major/minor), and instrumental timbre are all integral aspects of musical information, each of which is presumably processed and analyzed separately (Koelsch & Siebel, 2005). Here it is shown that these individual aspects of the acoustic input are all capable of signaling affective meaning. Whereas this idea has already been put forward (Koelsch & Siebel, 2005), these studies provide the empirical evidence for this assumption. It is likely that the mechanism underlying this process is the basic ability to perceive emotional signals in one's auditory environment via the processing of several acoustic signals. These signals are rapidly interpreted in terms of their emotional expression, which are then linked to associated affective concepts. Once this stage has been reached, the information is capable of interfering with other types of affective information, which in the case of verbal input, is coded in the meaning of the word.

The fast and accurate perception and recognition of an emotion would appear to be an important evolutionary advantage increasing the chances for survival. In normal human exchange, musical signals are not the primary means of communication, instead of which gestures, facial expressions, and, most importantly, speech are used efficiently. It has been shown that the speech signal, apart from the use of words, contains important information signaling the emotional state of the speaker (Schirmer & Kotz, 2006; Banse & Scherer, 1996). It may be possible to conceive of similar mechanisms of perceptual analysis applied to both speech and music to scan for information signaling changes in the affective state. Even though the purpose of such

mechanisms is lost on chords (as there is no defined signaler), it is not unlikely that mechanisms used for more relevant information (e.g., the emotional state of con-specifics) are also applied to music perception. Such ideas have also been formulated more recently in a theory on “superexpressive voices,” whereby some of the emotional appeal of music is derived from its similarities to human vocal expression of emotion (Juslin & Västfjäll, 2008). There is also some recent evidence that the perception of affective information contained by single chords recruits the same neural structures as the ones responsible for processing prosodic affect, namely, the posterior superior temporal sulcus (Steinbeis & Koelsch, 2008a; Grandjean et al., 2005).

Whereas music is capable of expressing both “basic” (Krumhansl, 1997) and “aesthetic” (Zentner, Grandjean, & Scherer, 2008) emotions, the present set of studies investigated only affective categories such as pleasant/unpleasant and happy/sad and also did so using an impoverished set of stimuli. Naturally, there is a lot more to music than individual parameters, which are pieced together to provide a coherent whole. Seeing that an effect of emotional expression and subsequent influence on meaning processing has been shown for individual musical features, it is likely that this assertion would also hold for longer musical sequences or pieces. However, this should be the subject of further empirical investigation. One therefore ought to view the present findings as the beginning, whereby affective categories are shown to communicate affective meaning.

The data do not speak for a link between the feeling of an emotion in response to and the perception of an emotion expressed by musical features. No data were obtained which could inform on the emotional state of the participants. It seems plausible that the perception or recognition of an emotion expressed by music is sufficient to prime-associated concepts and that this does not need to occur via the feeling of an emotion. However, this entire issue is underresearched and requires more empirical attention to be able to give a definite answer.

The present data give no indication of differences between musically trained and untrained participants with regards to the perception of affective meaning communicated by various musical features. Looking at both the behavioral responses and the ERP data, both groups seem to be influenced to the same degree by the affective congruity of chord–word pairs. This suggests that musical expertise has little effect on emotion and meaning-related processes, which is in line with previous studies. It was previously shown (Steinbeis, Koelsch, & Sloboda, 2006) that there were no differences between musicians and nonmusicians in emotional responses to manipulations in harmonic expectancy violations, and Bigand et al. (2005) demonstrated that there were no group differences in the perception of emotions in music.

In sum, the present three experiments show that the emotion expressed by various musical features (e.g., acous-

tic roughness, pitch intervals, and timbre) is capable of interfering with subsequent processing of affective word meaning and, therefore, suggests that individual musical features communicate signals which are processed as affectively meaningful. This appears to be understood regardless of musical training and the recognition of which may be the result of the acoustic analysis of affect, which can be applied to speech and acoustic signals generally. These data provide the first evidence that several individual features of the musical input are capable of communicating meaning, albeit on a basic affective level. It is very likely that this ability can be extended to other musical features, such as melody and rhythm. This evidence constitutes an extension of previous work on the meaning of music (Koelsch et al., 2004) by showing that emotion is a specific route to meaning in music, albeit the current conclusion rests

APPENDIX 1: WORDS USED IN EXPERIMENTS 1 AND 3

<i>Pleasant</i>	<i>Unpleasant</i>
Agreeableness	Aversion
Charme	Fear
Harmony	Evil
Friend	Disgust
Peace	Frustration
Cheerfulness	Dread
Pleasure	Danger
Grace	Violence
Good	Poison
Salvation	Atrocity
Love	Horror
Reward	Anguish
Glamour	Anger
Attraction	Hatred
Calm	Anxiety
Treasure	Suffering
Beauty	Murder
Protection	Damage
Blessing	Pain
Consolation	Fright
Goodwill	Nausea
Miracle	Misfortune
Wish	Fury
Goal	Rage

APPENDIX 2: WORDS USED IN EXPERIMENT 2

<i>Happy</i>	<i>Sad</i>
Success	Poverty
Party	Destitution
Banquet	Grave
Feast	Grief
Joy	Lament
Joyfulness	Grievance
Present	Illness
Luck	Sorrow
Humour	Burden
Ideal	Suffering
Idol	Plight
Jubilation	Detriment
Joke	Misery
Comedy	Misfortune
Laughter	Anguish
Delight	Plague
Festivity	Problem
Gag	Torture
Fun	Worry
Game	Tear
Triumph	Sadness
Advantage	Affliction
Wit	Loss
Bliss	Balefulness

on the basis of very low-level musical features and not whole pieces of music. The experiments therefore represent one of the first systematic analyses of which single aspects constituting music can communicate meaning. Although this is presently restricted to basic emotional categories, future work may focus on extending this to a wider and more subtle range of emotional shades and semantic connotations, taking the psychoacoustic information contained in the musical signal into account.

Reprint requests should be sent to Nikolaus Steinbeis, Max-Planck Institute for Human Cognitive and Brain Sciences, Stephanstr 1a, 04103 Leipzig, Germany, or via e-mail: steinbei@iew.uzh.ch.

Notes

1. Strictly speaking, dissonant stimuli were not chords, but for the sake of simplicity, we will use the term “chord” here also with reference to the dissonant stimuli.

2. In Germany, only 14% of teenagers and adults continue to play a musical instrument.

REFERENCES

- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.
- Besson, M., & Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparisons of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1278–1296.
- Bigand, E., Filipic, S., & Lalitte, P. (2005). The time course of emotional responses to music. *Annals of the New York Academy of Sciences, 1060*, 429–437.
- Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience, 2*, 382–387.
- Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America, 118*, 471–482.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.
- Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21*, 1882–1892.
- de Groot, A. B. M. (1984). Primed lexical decision: Combined effects of the proportion of the related prime–target pairs and the stimulus-onset asynchrony of prime and target. *Quarterly Journal of Experimental Psychology A, 36*, 253–280.
- Fazio, R., Sanbonmatsu, D., Powell, M., & Kardes, F. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., et al. (2009). Universal recognition of three basic emotions in music. *Current Biology, 19*, 573–576.
- Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to “happy–sad” judgements in equitone melodies. *Cognition and Emotion, 17*, 25–40.
- Goydke, K. N., Altenmüller, E., Möller, J., & Münte, T. F. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Brain Research, Cognitive Brain Research, 21*, 351–359.
- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. K., Scherer, K. R., et al. (2005). The voice of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8*, 145–146.
- Hevner, K. (1935). The affective character of the major and minor modes in music. *American Journal of Psychology, 47*, 103–118.
- Holcomb, P., & Anderson, J. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes, 8*, 379–411.
- Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience, 20*, 1940–1951.
- Juslin, P. N. (2003). Communicating emotion in music performance: Review and theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion*:

- Theory and research* (pp. 309–337). Oxford: Oxford University Press.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814.
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioural and Brain Sciences*, *31*, 559–575.
- Klauer, K. C., Rossnagel, C., & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 246–255.
- Koelsch, S., Fritz, T., von Cramon, D. Y., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: An fMRI study. *Human Brain Mapping*, *27*, 239–250.
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*, 302–307.
- Koelsch, S., Schmidt, B. J., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related potential study. *Psychophysiology*, *39*, 657–663.
- Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*, 578–584.
- Krumhansl, C. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, *51*, 336–353.
- Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.
- Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
- Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioural and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*, 199–211.
- McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford: Oxford University Press.
- McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*, 177–192.
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.
- Menon, V., Levitin, D. J., Smith, B. K., Lemke, A., Krasnow, B. D., Glazer, D., et al. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage*, *17*, 1742–1754.
- Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
- Musch, J., & Klauer, K. C. (Eds.) (2003). *The psychology of evaluation*. Mahwah, NJ: Erlbaum.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.
- Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition*, *62*, 267–272.
- Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2007). Is conceptual priming for environmental sounds obligatory? *International Journal of Psychophysiology*, *65*, 162–166.
- Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987.
- Plomp, R., & Levelt, W. J. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, *38*, 548–560.
- Sammler, D., Grigutsch, M., Fritz, T., & Koelsch, S. (2007). Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, *44*, 293–304.
- Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*, 1135–1148.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30.
- Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Brain Research, Cognitive Brain Research*, *14*, 228–233.
- Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Brain Research, Cognitive Brain Research*, *24*, 442–452.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341–349.
- Schön, D., Regnault, P., Ystad, S., & Besson, M. (2005). Sensory consonance: An ERP study. *Music Perception*, *23*, 105–118.
- Schubert, E., Wolfe, J., & Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the International Conference on Music Perception and Cognition, North Western University, Illinois* (pp. 112–116).
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.
- Sloboda, J. A. (1986). *The musical mind*. Oxford: Oxford University Press.
- Steinbeis, N., & Koelsch, S. (2008a). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*, 1169–1178.
- Steinbeis, N., & Koelsch, S. (2008b). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS One*, *3*, e2226.
- Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*, 1380–1393.

- Swain, J. (1997). *Musical languages*. New York: Norton.
- Tervaniemi, M. (2001). Musical sound processing in the human brain. Evidence from electric and magnetic recordings. *Annals of the New York Academy of Sciences*, *930*, 259–272.
- Van Petten, C., & Rieffers, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related potential measures. *Neuropsychologia*, *33*, 485–508.
- van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, *14*, 593–602.
- Wentura, D. (2000). Dissociative affective and associative priming effects in the lexical decision task: Yes versus no responses to word targets reveal evaluative judgment tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 456–469.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422.
- Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*, 494–521.
- Zentner, M. R., & Kagan, J. (1996). Perception of music by infants. *Nature*, *383*, 29.

This article has been cited by:

1. Stefan Koelsch. 2011. Towards a neural basis of processing musical semantics. *Physics of Life Reviews* . [[CrossRef](#)]
2. Miriam D. Lense, Elisabeth M. Dykens Musical Interests and Abilities in Individuals with Developmental Disabilities **41**, 265-312. [[CrossRef](#)]