COGNITION AND EMOTION, 2017
VOL. 31, NO. 2, 312–324
http://dx.doi.org/10.1080/02699931.2015.1110514

Routledge
Taylor & Francis Group

Fast and unintentional evaluation of emotional sounds: evidence from brief segment ratings and the affective Simon task

Tímea Folyi and Dirk Wentura

Department of Psychology, Saarland University, Saarbrücken, Germany

ABSTRACT

In the present study, we raised the question of whether valence information of natural emotional sounds can be extracted rapidly and unintentionally. In a first experiment, we collected explicit valence ratings of brief natural sound segments. Results showed that sound segments of 400 and 600 ms duration—and with some limitation even sound segments as short as 200 ms—are evaluated reliably. In a second experiment, we introduced an auditory version of the affective Simon task to assess automatic (i.e. unintentional and fast) evaluations of sound valence. The pattern of results indicates that affective information of natural emotional sounds can be extracted rapidly (i.e. after a few hundred ms long exposure) and in an unintentional fashion.

ARTICLE HISTORY
Received 28 October 2014
Revised 3 October 2015
Accepted 16 October 2015

KEYWORDS
Valence; emotional sounds; affective Simon task

Converging evidence suggests that people rapidly and unintentionally encode the affective content of an incoming stimulus (e.g. De Houwer & Eelen, 1998; Öhman, Flykt, & Esteves, 2001; Öhman & Mineka, 2001; Vuilleumier, 2005; Wentura, Müller, & Rothermund, 2014; Wentura & Rothermund, 2003). However, although evaluative (i.e. positive and/or negative) information is present in all stimulus modalities in our multisensory perception of the world, the majority of our knowledge on evaluative processing originates from visual studies. Accordingly, whereas it seems uncontroversial to claim that our auditory world has an immense potential in representing emotions, research on auditory affective processing is relatively sparse.[1]

The relative neglect of audition is astonishing given that our auditory system has unique operating and organising principles that allow it to take a vital role in monitoring our environment (e.g. King & Nelken, 2009). While visual perception has a limited spatial extent, audition is omnidirectional (i.e. it covers 360-degree in space). Compared with vision, auditory perception is less dependent on the spatial distance of the source, and sounds are perceived even if the sound source is occluded. Furthermore, auditory perception is characterised by an extensive pre-attentive system that allows for rapid detection of deviations in the auditory input from an internal model of the acoustic environment (e.g. Bendixen, SanMiguel, & Schröger, 2012; Näätänen, Paavilainen, Rinne, & Alho, 2007). Taking into consideration the importance of auditory perception in monitoring our environment, it seems straightforward that it should allow for efficient detection of emotional cues.

We argue that affective auditory research needs a clear understanding of the relevant specificities of the human auditory perception, which is in general not deducible by drawing analogies to the visual domain, and it needs to explore how the modality specific factors interact with affective processing. More specifically, the present study aimed to contribute to the growing field of auditory affective research by investigating the boundaries for evaluation of natural emotional sounds concerning the available time and the intentionality of the evaluation process. An important distinguishing characteristic of auditory stimuli is the fact that sounds carry temporally distributed information (e.g. Bregman, 1990; Griffiths & Warren, 2004; King & Nelken, 2009). While the evaluative content of emotional pictures—the most

CONTACT Tímea Folyi t.folyi@mx.uni-saarland.de

common type of stimulus in visual affective studies—is available instantly at stimulus onset, one can assume that the evaluative content of natural sounds becomes available only after a considerable exposure time due to temporal unfolding. Accordingly, brief (i.e. lasting approximately 1 s or less) natural emotional sounds have been employed only with caution in experimental paradigms investigating sound evaluation. For instance, Scherer and Larsen (2011) employed natural emotional sound segments as primes in their cross-modal evaluative priming study. Even though their primes had a considerably long duration (1 s), the authors introduced each sound in its full multi-second length before the priming task. They argued that the snippets would have only "reminded" participants of the previously introduced sounds. Thus, on the one hand, one can argue that the complex temporal structure of sounds would necessitate gradual extraction of evaluative information over time. On the other hand, a fast extraction of affective information from sounds seems to be crucial to detect and react to possibly beneficial or dangerous events rapidly. Hence, the claim of a slow affective processing in auditory modality seems ecologically invalid. For instance, we can easily imagine a situation of hearing a hostile human voice, a growling sound, or a siren, that causes us to react immediately (e.g. to orient toward the visual input of the sound source or to escape from the situation). Given these arguments, the first aim of our studies was to examine the boundary of sound exposure duration that is needed for extracting valence information explicitly from complex, naturally occurring emotional sounds. The second aim of the present studies was to investigate whether evaluative information can be—broadly speaking—automatically extracted from natural emotional sounds. "Automaticity" is an often used term to claim that certain processing outcomes are not the result of intentional, targeted, and (often) demanding behaviour. It is typically bound to the use of reaction time (RT)-based paradigms whose effects are—at least prima facie—not the result of intentional behaviour. We should hasten to add that a differentiated analysis shows that the term automaticity refers to a bundle of only loosely related characteristics (e.g. unintentional, uncontrollable, unconscious, efficient, fast; see Moors & De Houwer, 2006; Moors, 2015). We will focus here on fast and unintentional evaluation processes.

While there is converging evidence that valence of visual stimuli (e.g. emotional pictures, valent words) can be processed relatively automatically (for review, see e.g. Yiend, 2010; see also Wentura & Rothermund, 2003), we have relatively little knowledge about the automaticity of sound evaluation. Moreover, RT-based paradigms for assessing automatisms of evaluation are available almost exclusively in the visual modality. A possible reason for this discrepancy can be again the time-bound character of auditory processing. We can illustrate this issue with an exception to the dominantly visual RT-based paradigms for assessing automatic evaluations: Cross-modal evaluative priming effects have been demonstrated with evaluative auditory primes and evaluative visual targets (e.g. Carroll & Young, 2005; Goerlich et al., 2012; Marin, Gingras, & Bhattacharya, 2012; Scherer & Larsen, 2011; Schirmer, Kotz, & Friederici, 2002; Sollberger, Rebe, & Eckstein, 2003; Steinbeis & Koelsch, 2011; for a purely auditory version employing speech stimuli, see Degner, 2011). Evaluative priming effect refers to the phenomenon that the time needed to evaluate a target stimulus is considerably shorter when a preceding briefly presented prime stimulus has the same affective valence (i.e. the prime and target are affectively congruent) compared to when it has a different valence (i.e. the prime and target are affectively incongruent; for a review, see Klauer & Musch, 2003). Evaluative priming in the visual domain is almost exclusively focused on brief prime durations and brief stimulus onset asynchronies (SOA) of prime and target (i.e. $\leq$300 ms) because longer SOAs are thought to be influenceable by strategic behaviour and evaluation of the primes is thought to decay quickly. This is consistent with the finding that evaluative priming effects typically decrease with increasing SOA (Hermans, De Houwer, & Eelen, 2001; Klauer, Roßnagel, & Musch, 1997; see also Wentura & Degner, 2010). Therefore, presenting brief primes that still reliably convey evaluative information is crucial. Thus, to bypass the issue relating to the supposedly time-bound character of natural emotional sounds, musical primes were often employed in auditory-visual priming studies, as consonant and dissonant chords are assumed to transmit evaluative meaning with short exposure duration (Marin et al., 2012; Sollberger et al., 2003; Steinbeis & Koelsch, 2011; see above for an alternative solution by Scherer & Larsen, 2011). In contrast to this approach, in the present study we investigated whether ecologically valid, complex, natural emotional sounds (e.g. attack, bird singing, jackhammer, laughing) can be automatically (i.e. rapidly and unintentionally) evaluated in a RT-based paradigm already after very brief durations.

Thus, in the present studies, we employed two approaches: (1) We collected explicit valence ratings based on brief segments of natural emotional sounds; and (2) we used a RT-based paradigm to demonstrate automatic (in the sense of fast and unintentional) extraction of valence from natural sounds. Experiment 1 examined the boundary of the sound exposure needed for reliable explicit evaluative judgments. To this end, we collected explicit valence ratings of natural emotional sound segments with durations of 200, 400, and 600 ms. Additionally, to explore whether sound identification can mediate evaluative effects, we investigated whether semantic identification could occur based on the brief segments of emotional sounds. In Experiment 2, the evaluation of sound valence was measured indirectly in a paradigm requiring speeded responses to the presented stimuli. Specifically, Experiment 2 introduced an auditory variant of the affective Simon task (AST; De Houwer, Crombez, Baeyens, & Hermans, 2001; De Houwer & Eelen, 1998) to measure valence evaluation of natural sounds implicitly. In the AST, positive and negative stimuli (e.g. words in the visual version; here: emotional sounds) have to be categorised with regard to a *valence-neutral* task-relevant dimension which is varied orthogonally to valence (e.g. nouns vs. adjectives for the visual version; here: motion direction of sounds). Participants are instructed to respond, however, with evaluative responses (e.g. saying "good" for nouns and "bad" for adjectives for the visual version; here: e.g. saying "good" for a movement to the right and "bad" for a movement to the left). Accordingly, AST trials can be congruent (stimulus valence and response valence match) or incongruent (stimulus valence and response valence mismatch). The typical result of the AST is shorter RTs for congruent compared to incongruent trials. As the valence of the stimuli is irrelevant to the primary task (and participants are also often explicitly instructed to ignore stimulus valence), it is assumed to be processed automatically in the sense of unintentional and fast (see e.g. Bargh, 1992; De Houwer & Eelen, 1998).

Experiment 1

In Experiment 1, we investigated whether valence information of complex natural emotional sounds can be extracted rapidly and evaluated in an explicit fashion, and whether this evaluation can be mediated by early semantic identification of the sound source or content. To this end, we presented brief (200, 400, and 600 ms long) segments of natural emotional sounds sampled from the International Affective Digitized Sounds battery (IADS; Bradley & Lang, 2007). The IADS battery includes language-independent natural emotional sounds across a wide range of semantic categories, like environmental sounds (e.g. jackhammer) and human vocalizations (e.g. laughing). In a first sample of participants, we collected valence ratings on the brief segments of emotional sounds, and we tested whether the ratings of these segments mirror the valence ratings of the corresponding full-duration sounds (with approximately 6 s duration) based on the normative sample reported by Bradley and Lang (2007) and an own native German sample (see below). With a second sample of participants, we investigated whether sound identification (i.e. semantic identification of the source and/or the content of the sound) can occur based on these brief segments of emotional sounds. We used two measures of sound identification: (1) a rather coarse-grained identification of the sound source requiring the participants to differentiate whether a sound was produced by an animate or an inanimate agent; and (2) a more fine-grained identification of the sound regarding its content and source. Additionally, the second sample provided valence ratings on the full-length emotional sounds.

Methods

Participants

Sample 1 and 2 each had 30 participants (undergraduate students from Saarland University) who participated for monetary compensation (Sample 1: 22 females, aged 18–32 years, $Mdn = 23$ years; Sample 2: 19 females, aged 19–31 years, $Mdn = 24.5$ years).[2]

Materials

We selected 39 *positive* (e.g. applause, slot machine, bird singing), 39 *negative* (e.g. vomiting, attack, car wreck), and 39 *neutral* (e.g. office noise, walking, yawn) natural sounds from the IADS battery (Bradley & Lang, 2007). Our selection aimed to maximise the differences between the positive, negative, and neutral stimulus pools on normative valence ratings, thereby creating stimulus pools with non-overlapping rating ranges. A further selection criterion aimed to minimise silent periods in the 0–600 ms excerpts of the sounds. Mean normative valence ratings of the full-length stimuli on a 9-point scale ranging from very unpleasant (1) to very pleasant (9) were $M =$

7.02 (SD = 0.43; in the range from 6.31 to 7.90) for positive, M = 4.75 (SD = 0.40; in the range from 4.01 to 5.35) for neutral, and M = 2.41 scores (SD = 0.48; in the range from 1.57 to 3.08) for negative sounds, respectively. We coded the sounds as produced by animate or inanimate agents (21, 16, and 24 animate and 18, 23, and 15 inanimate sounds in positive, neutral and negative conditions, respectively). From each sound, we created three new sound files by extracting the 0–200, 0–400, and 0–600 ms segments. Sounds were organised into three stimulus sets. Each set contained the 39 positive, 39 negative, and 39 neutral sounds, with one-third of each valence category selected in the 200, 400, and 600 ms version, respectively. Thereby each set contained one version of all available sounds. Each participant received one of the three stimulus sets in a balanced design, and across participants we thus collected 10 rating scores for each version of each sound. Previous studies conducted at our lab suggest that an aggregate of 10 ratings secures high reliability of the aggregate measure.

Procedure

Each participant received 117 trials, featuring the sounds of one of the stimulus sets. Auditory stimuli were presented binaurally via headphones (HD-212 Pro, Sennheiser, Wedemark, Germany) in a comfortable loudness of approximately 70 dB(A). Trials were presented in an individually randomised order.

For Sample 1, each trial started with the presentation of a rating screen featuring a 9-point scale ranging from very unpleasant (−4) to very pleasant (+4), with zero as the neutral point. After 500 ms, a sound segment was presented. Participants were asked to rate the pleasantness of each sound by clicking on one of the nine scale points. The next trial started immediately after the response was registered.

For Sample 2, each trial started with the presentation of a fixation cross without auditory stimuli. After 500 ms, a sound segment was played. Thereafter, participants were asked to accomplish two tasks. First, participants had to categorise the presented sound according to whether it was produced by (an) animate or inanimate agent(s) by clicking to the corresponding category label. For instance, a person or an animal was considered as an animate agent, while musical instruments, tools, or natural phenomena (e.g. thunder) were considered as inanimate agents. Second, participants had to identify the sound by describing it in their own words. Participants were asked to type a one or two words long answer that ideally refers to both the sound source and the "nature" or content of the sound (e.g. "woman screams"). Participants were also encouraged to cover a complex situation by using only one word if it was apposite (e.g. "party"). The next trial started immediately after pressing the Enter key. Additionally, participants of Sample 2 were asked to perform a valence rating task on the *full-length* stimuli at the end of the experimental session. The procedure was identical with the procedure of Sample 1, but importantly, on each trial the emotional sounds were now played in their full-length version (6 s).

Design

We applied a 3 × 3 mixed factorial design on the valence ratings and on the two measures of sound identification with the a priori valence category (positive vs. neutral vs. negative) as the grouping factor and the duration of the sound segment (200 ms vs. 400 ms vs. 600 ms) as the repeated measures factor.

Results

All analyses are based on items as the units of analyses with values aggregated across participants.

Valence ratings

Valence ratings were transformed to a 1–9 scale to stay in line with the normative ratings provided by Bradley and Lang (2007). Results are presented in Table 1. First of all, it can be seen in the upper part of the table that the full-length rating provided by a German sample closely resembles the norm rating provided by Bradley and Lang (2007). Thus, there seem to be no important cultural differences between the two samples. Intraclass-correlations show that interrater-agreement was high for all ratings. However, it was considerably lower for the 200 ms rating than for the 400 and 600 ms ratings.

The pattern of means of the three valence conditions clearly reflects a differentiation into positive, neutral, and negative evaluations, and an increasing difference between the means of positive and negative ratings with longer sound durations. A 3 (valence: positive vs. neutral vs. negative) × 3 (duration: 200 ms vs. 400 ms vs. 600 ms) MANOVA for repeated measures with duration as a within-items factor and valence as a grouping factor on the valence ratings yielded a main effect of valence, $F(2,114) = 73.28$, $p < .001$, $\eta_p^2 = .562$, that was moderated by the sound duration, $F(4,228) = 16.30$, p

< .001, $\eta_p^2 = .222$. There was no significant main effect of duration, $F(2,113) = 2.12$, $p = .125$, $\eta_p^2 = .036$. To test valence differentiation in the different duration conditions, separate ANOVAs were performed for each duration condition with valence (positive vs. neutral vs. negative) as grouping factor. These analyses showed significant valence effects for all three durations, $Fs(2,114) > 26.60$, $p < .001$, $\eta_p^2 > .317$. To understand the interaction pattern, we tested the increase in valence differentiation for the two duration transitions (i.e. the transition from 200 to 400 ms and the transition from 400 to 600 ms): the first 3 (valence) × 2 (duration: 200 ms vs. 400 ms) planned interaction contrast was significant, $F(2,114) = 34.17$, $p < .001$, $\eta_p^2 = .375$, thereby signalling a gain in differentiation by using 400 ms excerpts compared with 200 ms snippets. The second 3 (valence) × 2 (duration: 400 ms vs. 600 ms) planned interaction contrast did not show significant differences, $F(2,114) = 1.70$, $p = .187$, $\eta_p^2 = .029$, thereby indicating that gain in differentiation by using 600 ms excerpts compared to those of 400 ms length is modest. The difference between the 200 ms condition on the one hand and the 400 and 600 ms conditions on the other hand can be seen additionally in the correlation coefficients of the ratings for the brief segments with the full-length ratings (see Table 1).[3]

Additionally, we carried out analyses focusing on duration effects within the a priori valence categories. We found a significant duration effect within the positive and negative valence category, $Fs(2,37) > 21.27$, $p < .001$, $\eta_p^2 > .534$ ($F < 1$ within the neutral category). Within the two valenced categories, we found significant gain in differentiation by using 400 ms excerpts compared with 200 ms excerpts, $Fs(1,38) > 22.40$, $p < .001$, $\eta_p^2 > .370$. However, gain in differentiation by using 600 ms excerpts compared to those of 400 ms length was modest, $F(1,38) = 4.06$, $p = .051$, $\eta_p^2 = .096$ within the positive category; and $F < 1$ within the negative category.

There are two more sources of evidence to evaluate the validity of the ratings. First, since our selection was category-focused (i.e. we a priori selected positive, neutral, and negative sounds such that the norm-rating distributions of the three samples were non-overlapping), we attempted to predict category membership on the basis of the 200, 400, or 600 ms ratings, respectively, using multinomial logistic regression. Corresponding to the results reported above, even the 200 ms rating significantly improved prediction in comparison to random assignment, $\chi^2(2) = 45.22$, $p < .001$ ($\chi^2[2] \geq 92.72$ for 400 and 600 ms). However, while classification accuracy (ACC) was 71.8% for the 600 ms rating, with only 1.7% severe misclassifications (i.e. classification of a positive sound as negative and vice versa), predictions based on the 200 ms rating were considerably weaker: Classification ACC was 59.0%, with 7.7% severe misclassifications (for 400 ms, classification ACC was 74.4%, with 6.0% severe misclassifications).

Second, standard deviations of mean norm ratings of the IADS are available. These can be considered as an index of relative ambiguity of evaluation. Thus, a new valid rating should be sensitive to this ambiguity; thereby it should be more predictive of the original norm ratings for less ambiguous sounds and less predictive for more ambiguous sounds. In statistical terms, we can assume the interaction term of a new valid rating and the ambiguity index to be significant when predicting the norm rating. This holds true for both the 400 ms rating, $\beta = -.12$, $t(116) = 2.07$, $p = .040$, for the product term, and the 600 ms rating, $\beta = -.15$, $t(116) = -2.73$, $p = .007$, but not for the 200 ms condition, $\beta = -.10$, $t(116) = -1.36$, $p = .178$.

Table 1. Mean valence ratings, intraclass-correlations (ICC), and correlations with the norm rating (r_n) and with the full-length sounds ratings based on a native German sample (r_G) for the three duration conditions of Experiment 1. Mean valence ratings are also provided for the full-length stimuli based on a normative (Full-Length$_n$) and a native German sample (Full-Length$_G$). Valence ratings range from very unpleasant (1) to very pleasant (9); *SD* in parentheses.

	Valence category					
	Negative	Neutral	Positive	ICC[a]	r_n	r_G
Full-Length$_n$	2.41 (0.48)	4.75 (0.40)	7.02 (0.43)			
Full-Length$_G$	2.36 (0.57)	4.49 (0.98)	6.39 (0.96)	.97	.93	
200 ms	3.73 (1.05)	4.59 (0.81)	5.32 (1.01)	.86	.58	.60
400 ms	3.05 (1.07)	4.56 (0.89)	5.99 (1.21)	.92	.77	.77
600 ms	3.03 (1.16)	4.71 (1.06)	6.32 (1.25)	.92	.78	.81

[a]Average intraclass-correlation for random raters (ICC [1, 10]) according to Shrout and Fleiss (1979).

Semantic identification

The results of the two measures on semantic identification of sounds are presented below and in Table 2.

Sound source categorisation. Participants were able to differentiate whether emotional sound segments were produced by animate or inanimate agents with remarkable precision (see Table 2). A 3 (valence: positive vs. neutral vs. negative) × 3 (duration: 200 ms vs. 400 ms vs. 600 ms) MANOVA[4] for repeated measures on sound source categorisation ACC yielded a main effect of duration, $F(2,113) = 6.67$, $p = .002$, $\eta_p^2 = .105$, with a

Table 2. Mean ACC (%) for sound source categorisation and mean ACC (ranging from fully incorrect [0] to fully correct [4]) for specific sound identification in the three duration conditions of Experiment 1, *SD* in parentheses.

	Valence category		
	Negative	Neutral	Positive
1. Sound source categorisation			
200 ms	84.1 (21.2)	74.9 (31.0)	83.6 (23.7)
400 ms	91.8 (15.5)	77.7 (29.7)	90.8 (20.6)
600 ms	92.8 (13.4)	78.5 (31.2)	90.5 (17.2)
2. Specific sound identification			
200 ms	1.49 (1.19)	0.80 (1.04)	1.99 (1.19)
400 ms	2.17 (1.06)	1.46 (1.24)	2.64 (1.24)
600 ms	2.36 (1.09)	1.70 (1.22)	2.86 (1.09)

significant difference between the 200 and 400 ms conditions, $F(1,114) = 11.46$, $p = .001$, $\eta_p^2 = .091$, but with no significant difference between the 400 and 600 ms conditions, $F < 1$; and a significant valence effect, $F(2,114) = 4.07$, $p = .020$, $\eta_p^2 = .067$ which is entirely due to lower ACC for neutral sounds compared to valent sounds, $F(1,115) = 8.14$, $p = .005$, $\eta_p^2 = .066$ ($F < 1$ for positive versus negative sounds). Duration and valence did not significantly interact, $F < 1$. Note that these MANOVA results have to be taken with some caution because source identification scores are extremely skewed (e.g. 64% of the 600 ms files have a score of 1).

Specific sound identification. Two native German raters assessed the correctness of the specific sound identifications by scoring the answers as "correct", "partly correct", or "incorrect". The label "partly correct" was applied to the situations when either the sound source or the content of the sound was not or was incorrectly described (e.g. "women" instead of "woman screams"). The interrater-agreement between the two raters was good in all duration conditions as shown by high intraclass-correlations, ICCs > .93. We aggregated the sums of the evaluations of the two raters; thus, the procedure resulted in a 5-point accuracy measurement (0–4; ranging from "judged as incorrect by both raters" to "judged as correct by both raters").

Table 2 presents the results of specific sound identification ACC. A 3 (valence: positive vs. neutral vs. negative) × 3 (duration: 200 ms vs. 400 ms vs. 600 ms) MANOVA for repeated measures on sound identification ACC yielded a main effect of duration, $F(2,113) = 47.78$, $p < .001$, $\eta_p^2 = .458$, with significant differences between the 200 and 400 ms conditions, $F(1,114) = 65.92$, $p < .001$, $\eta_p^2 = .366$, and the 400 and 600 ms conditions, $F(1,114) = 7.98$, $p = .006$, $\eta_p^2 = .065$. Additionally, a significant valence effect was found, $F(2,114) = 12.82$, $p < .001$, $\eta_p^2 = .184$, which is dominantly due to lower ACC for neutral sounds compared to valent sounds, $F(1,115) = 20.73$, $p < .001$, $\eta_p^2 = .153$, while $F(1,76) = 4.41$, $p = .039$, $\eta_p^2 = .055$ for positive versus negative sounds. Duration and valence did not interact, $F < 1$.

Evaluation and semantic identification

To obtain evidence for co-processing of semantic and evaluative features (which would encompass the possibility that semantic processing is a precondition of evaluation), we employed the following logic: If semantic processing would be a necessary precondition of evaluation, (non-neutral) sounds that are not identifiable for a given duration condition should be rated as neutral; sounds that are clearly identifiable should have received a marked valence rating—either a positive one for positive sounds or a negative one for negative sounds. Finally, (non-neutral) sounds with a medium accuracy score (i.e. sounds that were identified only by some raters) should have received moderate mean valence ratings as a result of some marked ratings (for those who identified) and some neutral ratings (for those who did not identify). Plotting specific identification scores (on the *Y*-axis) against the ratings (on the *X*-axis) should therefore yield a parabola-shaped scatterplot. Moreover, if the original ratings (scaled from −4 to +4) will be used, the parabola should have its vertex at $x = 0$. Therefore we regressed the specific identification scores of positive and negative sound files on the quadratic term of the original ratings only (i.e. we left out the first-order term) which forced the regression algorithm to fit a parabola with vertex $x = 0$ to the data. Note that this is a rather strong constraint. The quadratic relationship was significant for all durations, $\beta = .28$, $t(76) = 2.52$, $p = .014$ for 200 ms, $\beta = .25$, $t(76) = 2.28$, $p = .026$ for 400 ms, $\beta = .30$, $t(76) = 2.70$, $p = .009$ for 600 ms. The same kind of analysis using the source categorisation scores instead of the specific identification scores yielded non-significant results. However, this is probably due to the skewness of distributions.

Discussion

Results of Experiment 1 demonstrate clearly that valence can be extracted from very brief (i.e. a few hundred milliseconds long) segments of natural emotional sounds. Even valence ratings for durations as short as 200 ms are still reliable, although they are

slightly more ambiguous compared with ratings of the 400 and 600 ms segments. Evidence for these claims was derived from several sources. First, valence ratings of brief sound segments showed a clear differentiation between positive, neutral, and negative valence categories, which were defined a priori according to the norm ratings. Though ratings reflected significant valence differentiation in each duration condition, a significant interaction emerged between duration and valence: ratings showed clearer valence differentiation as exposure duration increased, with the largest increase in differentiation at the 200–400 ms transition. Second, evaluation of 200, 400, and 600 ms segments showed a close relationship with the normative valence ratings of the full-duration sounds. While the 600 ms rating (and with some slight limitation the 400 ms rating) seemed to behave almost like a re-rating of the full-length stimuli, the 200 ms rating was more equivocal. The inconsistency of the 200 ms rating was reflected in a lower correlation with the full-length ratings, a lower interrater-agreement, and less sensitivity to the ambiguity in the norm rating compared with the longer duration conditions.

Furthermore, we raised the question whether it is possible to extract complex semantic meaning during a few hundred milliseconds of presentation time, thus, whether very early evaluations—that is, evaluations based on the information content available after no longer than 200–600 ms sound exposure—can be driven by semantic meaning. We found that (1) participants could differentiate sounds produced by animate and inanimate agents with high precision; and (2) participants could identify the specific sounds still reliably, although with less precision. As expected, both the rather coarse-grained and the more specific index of sound identification showed higher precision as exposure duration increased, with the greatest increase in precision at the 200–400 ms transition. The more specific index of sound identification showed a close relationship with the evaluation of the sound fragments in all duration conditions, suggesting that sound identification could occur before or parallel with the early evaluations; thus, it is possible that early evaluative effects are based on semantic processing.

Results of Experiment 1 suggest that valence is evaluated in a similar way when a standard natural emotional sound of several seconds is available or when there is only a short snippet of sound to base the judgment on. While 400 and 600 ms segments were evaluated highly reliably (i.e. they appeared to be comparable to a re-rating of the full-length stimuli), 200 ms segments were evaluated still reliably but relatively more inconsistently compared with the longer durations. Our results thus suggest that 200 ms long exposure time is sufficient for a partial evaluation of natural sounds—at least under the conditions in which explicit instructions are given for evaluation. Taken together, our findings lend support to the notion that complex natural emotional sounds can be evaluated rapidly—at least in an intentional way—and fast evaluations can be mediated by early semantic identification of the sounds. Based on these results, in a second experiment we introduced an auditory version of the AST; with this task we investigated whether evaluation of natural emotional sounds can be automatic not only in the sense of fast, but also in the sense of unintentional.

Experiment 2

Experiment 2 introduced an auditory version of the affective Simon (AS) paradigm to assess automatic (i.e. unintentional and fast) evaluations of sound valence. AS effects have been demonstrated by employing a wide variety of stimuli (e.g. written words, schematic faces, simple stimuli associated with valence), task-relevant stimulus features (e.g. grammatical category, colour), and responses (e.g. by uttering valence category labels or affectively connoted words, or moving a manikin on the screen towards or away from the stimulus; see e.g. De Houwer & Eelen, 1998; Moors & De Houwer, 2001; Voß, Rothermund, & Wentura, 2003). In the auditory version of the AST, we used a purely perceptual stimulus feature as the relevant valence-neutral dimension that determines the correct response: Participants were required to classify the direction of an illusory movement of the sound source. This task can be performed successfully without intentional processing of the affective meaning of the sounds. We presented natural emotional sounds as stimuli and recorded verbal responses: Participants uttered "good" or "bad" depending on the direction of the illusory movement. RTs and error rates were analysed as a function of stimulus and response valence congruency.

Based on the results of Experiment 1, we assumed that valence information can be successfully extracted after a few hundred milliseconds of sound exposure. However, in contrast to Experiment 1 (i.e. presenting

brief snippets of sounds which were explicitly evaluated), in Experiment 2 we applied another approach to investigate rapid sound evaluations: We employed full-length auditory stimuli that required fast responses to a task-relevant feature. From the onset time of the task-relevant feature we can coarsely estimate the time of valence exposure. Taken into consideration the relative ambiguity of the 200 ms ratings in Experiment 1, we employed two parallel versions of Experiment 2: In *Experiment 2a*, we made the task-relevant feature (i.e. onset of virtual movement) available at 500 ms post sound onset (i.e. 500 ms feature start onset asynchrony; FSOA). This means that participants were exposed to the valence-relevant content slightly earlier than to the task-relevant information. In *Experiment 2b*, we used a synchronous version, that is, the task-relevant virtual movement started at the onset of the sound (0 ms FSOA).

Methods

Participants

In Experiment 2a, 57 students from Saarland University (39 females; aged 18–36 years, *Mdn* = 25 years; 4 left-handers) participated for monetary compensation. The data of four further participants were discarded because of extreme error rates (≥16.7%; i.e. far out values according to Tukey, 1977). In Experiment 2b, 52 students from Saarland University (30 females; aged 19–33 years, *Mdn* = 23 years; 6 left-handers) participated for monetary compensation.

Given a sample size of $N = 57$ in Experiment 2a (52 in Experiment 2b) and an α-value of .05 (two-tailed), effects of size $d = 0.49$ (0.51 in Experiment 2b; i.e. medium effects according to Cohen, 1988) can be detected with a probability of $1-\beta = .95$ (calculated with the aid of G*Power 3 software; Faul, Erdfelder, Lang, & Buchner, 2007).

Materials

20 *positive*, 20 *negative*, and 20 *neutral* sounds from the IADS battery were presented via headphones **(**HD-600, Sennheiser, Wedemark, Germany**)** with a maximal loudness of approximately 70 dB(A). Mean normative valence ratings—on a 9-point scale ranging from most unpleasant (1) to most pleasant (9)—were $M = 6.94$ ($SD = 0.51$) for positive, $M = 4.62$ ($SD = 0.52$) for neutral, and $M = 2.48$ ($SD = 0.54$) for negative sounds, respectively (Bradley & Lang, 2007). Additionally, four positive, four negative and four neutral IADS sounds were used in practice trials. To invoke the virtual sound movement, amplitude of the sounds was modulated in the following way: starting at 500 ms post onset (Experiment 2a) or starting at sound onset (Experiment 2b), intensity in one auditory signal channel of the stereo sound was reduced linearly over a 1000 ms interval by a total of 75%. We created two versions of each sound, one with an illusory movement from a central position toward the right side of the perceiver ("*moving to the right*" *sounds*) and one with an illusory movement to the left ("*moving to the left*" *sounds*; see e.g. Rosenblum, Carello, & Pastore, 1987).

Design

We employed a 2 (sound valence: positive vs. negative) × 2 (response valence: positive vs. negative) repeated measures design which reduces to a simple one-factorial congruency (congruent vs. incongruent) design. Neutral sounds were added to obtain a baseline measure against which to assess the effects of congruency (i.e. to obtain rough estimates of "costs" and "benefits").

Procedure

All participants were tested individually. Before the experiment, instructions emphasised that participants should attend only to the illusory movement of the sounds and ignore any other stimulus features. The experiment started with 12 practice trials. During the practice phase, participants received visual accuracy feedback after every trial. The experimental phase comprised 60 experimental trials, with 20 trials featuring positive, 20 trials featuring negative, and 20 trials featuring neutral sounds in an individually randomised order. Half of the sounds were presented in "moving to the left" and half of the sounds in "moving to the right" version, respectively, in a random order. Assignment of specific sounds to right and left moving versions was counterbalanced across participants.

An experimental trial started with the presentation of a fixation cross without auditory stimuli. The fixation cross remained on the screen until the end of the trial. After 1000 ms, a positive, negative, or neutral sound was played. Participants' task was to categorise the direction of the virtual movement by uttering "good" or "bad" ("gut" and "böse" in German, respectively) as quickly and accurately as possible. The assignment of response (saying "good" or "bad") to illusory movement direction (right or left) was counterbalanced

between participants. While a voice key apparatus recorded RT (i.e. the onset of the utterance), the response category was registered online by the experimenter, who was sitting in front of a second screen next to the participant but separated by a partition wall. That is, the experimenter pressed one key for response "good" and one key for response "bad"; if the voice key was triggered accidentally (e.g. by misutterances or by noises like coughing), a third key was used. After the vocal response was detected by the voice key, the auditory stimulus was terminated.

Results

RTs were calculated from the onset of the illusory movement (i.e. 500 ms post sound onset for Experiment 2a, at sound onset for Experiment 2b). RT analyses were restricted to trials with correct responses and error-free response recording (3.0% of the trials for both experiments were excluded because of incorrect or erroneous responses or non-reaction of the voice key). As an a priori criterion, RTs below 300 ms and above 2000 ms were discarded from further analyses (2.4% and 2.9% of the trials in Experiment 2a and 2b, respectively). Table 3 shows the mean RTs and error rates for the congruent and incongruent conditions, and for neutral sounds. (Table A1 in the Appendix shows the mean RTs and errors for the fully expanded design).

For Experiment 2a, the RT difference between congruent and incongruent trials was significant, $t(56) = 2.50$, $p = .015$, $d = 0.33$. The effect seems to be due mainly to the costs associated with the incongruent pairings: The mean RT for neutral sounds was almost identical to the RT for congruent trials, $|t| < 1$, but the difference between incongruent and neutral conditions was significant, $t(56) = 2.65$, $p = .011$, $d = 0.35$. Similar analyses on the error rates did not show any significant differences, all $|t|$s < 1.

Table 3. Mean RTs (in ms) and error rates (in %, in parentheses) as a function of stimulus and response valence congruency in Experiment 2.

	Experiment 2a (500 ms FSOA)[a]	Experiment 2b (0 ms FSOA)[a]
Neutral sounds	1020 (2.8)	1009 (2.6)
Congruent	1017 (3.0)	1021 (3.0)
Incongruent	1040 (3.3)	1027 (3.5)
AS effect[b]	24 [9]	6 [10]

[a]FSOA = Feature Start Onset Asynchrony.
[b]RT (incongruent) minus RT (congruent); standard errors in brackets.

For Experiment 2b, the incongruent-congruent RT difference was in the expected direction but fell short of significance, $t(51) = 0.66$, $p = .512$, $d = 0.09$. Neutral RTs were numerically faster than congruent RTs, but this difference was not significant, $t(51) = -1.12$, $p = .267$, $d = -0.16$. The difference between incongruent and neutral conditions was significant, $t(51) = 2.16$, $p = .035$, $d = 0.30$. Similar analyses on the error rates did not show any significant differences, all ts < 1.

Discussion

In Experiment 2, we used an auditory version of the AST in two variations: In one version, the task-relevant change started after half a second of exposure to the emotional sound (Experiment 2a); in the other version, it started at sound onset (Experiment 2b). We found a significant AS effect in Experiment 2a, that is, longer RTs when stimulus and response valence were incongruent rather than congruent. We have to emphasise that for successful task performance participants were not required to process the stimulus valence, as it was entirely task-irrelevant and not predictive of the task-relevant feature. Additionally, participants were explicitly instructed to ignore every characteristic of the sounds other than the task-relevant feature. Taken together, results of Experiment 2a support the interpretation that natural emotional sounds can be evaluated automatically, in the sense of fast and unintentional evaluation (see e.g. Bargh, 1992; De Houwer & Eelen, 1998).

In Experiment 2b, a similar RT pattern emerged as in Experiment 2a but fell short of statistical significance. A significant AS effect was thus found only in Experiment 2a, where exposure to the evaluative information started before the task-relevant manipulation, and not in Experiment 2b, where the onsets of the evaluative and the task-relevant information were synchronous. Thus, it seems—at least in the present paradigm—that a head start is needed for the valence information to facilitate or interfere with the behavioural response. However, the absence of a significant AS effect in Experiment 2b was largely due to the relatively long RTs in the congruent condition (i.e. numerically longer than the neutral condition RTs); the costs associated with the incongruent condition (relative to neutral) were significant and corresponded roughly to those found in Experiment 2a. We will return to this issue in the General Discussion.

General discussion

In the present study, we demonstrated that valence information can be extracted rapidly and even in an implicit fashion from natural emotional sounds. First, explicit valence ratings revealed that valence of natural emotional sounds can be evaluated validly even if only the first few hundred milliseconds of the sounds are presented. Valence ratings on the 400 and 600 ms long segments showed a clear-cut pattern: They firmly reflected the a priori sound valence and showed a strong relationship with the valence ratings of the full-length sounds. However ratings of natural sound segments with 200 ms duration also reflected valence reliably, they were slightly more ambiguous than the 400 and 600 ms ratings, thus suggesting that 200 ms long exposure may have allowed only partial evaluations. Despite of this relative ambiguity of the 200 ms ratings, results of Experiment 1 indicate that natural sounds can convey their affective meaning already after very brief exposure time. Second, we found evidence that this early evaluation, at least partly, can be driven by a rapid semantic identification of sounds. Third, we demonstrated that valence of natural sounds can be evaluated implicitly. We introduced an auditory version of the AST: In this task participants responded slower if the valence of the response and the valence of the sound mismatched. This effect emerged even though participants were instructed to ignore stimulus valence, and even though the task-relevant feature was varied orthogonally to valence and was purely perceptual (i.e. no semantic encoding of the sound was necessary). However, this effect became evident only if the task-relevant feature lagged behind the onset of the sound by half a second. There are at least three possible explanations for this pattern of results. First, one might speculate that the intentional processing of the task-relevant feature attenuated processing of other stimulus features, including valence, when presentation onset was synchronous. Second, taking into account the results of Experiment 1, the valence information provided in the first fraction of a second may be rather ambiguous. If so, some of the congruent trials may have in fact been processed as if they were incongruent. Given that the auditory AS effect (as found in Experiment 2a) seems to arise mainly from the costs associated with incongruency, this would result in a mean RT for the congruent condition that is (at least numerically) higher than the mean RT of the neutral condition. Third, if we assume that factors that can influence the relative automaticity of sound evaluation are cumulative, in Experiment 1, brief duration of exposure may have been compensated by increased intentional processing of sound valence (that in turn could govern increased attentional resources); while in Experiment 2, the lack of intentionality may have necessitated longer exposure time for sound evaluation to occur (see the argumentation of Moors, 2015). Hence, while evaluation of natural sounds emerged rapidly (i.e. at least partially already after 200 ms long exposure) when participants were explicitly instructed for evaluating the sounds, in an indirect RT-based paradigm sound evaluation occurred supposedly unintentionally but also might have emerged somewhat slower. However, note that a relatively short exposure time of 500 ms before the task-relevant manipulation was already sufficient for sound valence to influence behavioural responses to a task-relevant feature.

In summary, our results give support to the view that naturally occurring emotional sounds (e.g. environmental sounds, human vocalizations) can be evaluated rapidly and even without conscious intention. First, explicit valence ratings of brief sound segments showed that natural sounds can be evaluated reliably after very short (i.e. 600 and 400 ms, and with some limitation 200 ms) exposure. It means that although information content of sounds are typically distributed in time, valence information can be obtained after very short presentation time from natural emotional sounds. Second, evidence from a newly developed auditory version of the AST indicates that natural sounds can be evaluated not only rapidly but also in an implicit fashion. We can conclude that sound valence was processed automatically in the sense of involuntariness, as the valence information was completely irrelevant regarding the main task, and the task-relevant feature was purely perceptual (i.e. did not require "deep" processing for successful task performance) and it was not contingent on the stimulus valence. Moreover, participants were asked explicitly to ignore every other feature aside from the task-relevant modulation.

However, we cannot preclude the possibility that feature-specific attention allocation plays a role in the presented evaluation effects, that is, that auditory AS effects depend on attention allocation on evaluative stimulus features because verbal responses had to be uttered throughout the experiment which were strongly positively ("good!") or negatively ("bad!")

connoted. In recent years, the concept of feature-specific attention allocation has been suggested in different sub-domains of cognitive psychology (e.g. Folk, Remington, & Johnston, 1992; Kiefer & Martens, 2010; Spruyt, De Houwer, & Hermans, 2009; Spruyt, De Houwer, Hermans, & Eelen, 2007; see also Bermeitinger, Wentura, & Frings, 2011). Importantly, feature-specific attention allocation was suggested for the evaluative domain (see e.g. Everaert, Spruyt, & De Houwer, 2013; Spruyt et al., 2007, 2009; Spruyt, De Houwer, Everaert, & Hermans, 2012). Further research can elucidate the role of feature-specific attention allocation in the auditory AST. On a related note, it will be worthwhile to study the moderating effect of other affective variables just as motivation or mood (see e.g. Vermeulen, Corneille, & Luminet, 2007, for effect of mood on automatic evaluations in a variant of the visual AST).

Emotionally significant visual stimuli receive prioritised processing and are evaluated rapidly and unintentionally. The present results demonstrate that a similarly powerful affective processing takes place in audition that enables us to evaluate affectively significant sounds rapidly with an extremely high precision, automatically (in the sense of fast and unintentional evaluation), and even despite of the apparent drawback of temporally extended information conveyed by natural sounds. Besides of the theoretical implications, we believe that our results will be useful for further research in affective auditory processing in that they provide guidance in experimental design and supply researchers with a novel method to assess implicit evaluations of sounds.

Acknowledgements

The authors thank Ullrich Ecker for his helpful comments and Thorid Römer for assistance in data collection.

Funding

The present research was conducted within the International Research Training Group "Adaptive Minds" supported by the German Research Foundation [GRK 1457].

Notes

1. Despite of its relative neglect compared with visual affective research (which encompasses hundreds of published studies), there are remarkable attempts to investigate sound evaluation, which should be mentioned: There are studies on preferential processing of conditioned valence of sounds (Bröckelmann et al., 2011, 2013; Folyi, Liesefeld, & Wentura, 2015), on functional magnetic resonance imaging and electrophysiological correlates of complex emotional sounds such as environmental sounds, emotional vocalizations, and music (e.g. Czigler, Cox, Gyimesi, & Horváth, 2007; Grandjean et al., 2005; Koelsch, Fritz, von Cramon, Müller, & Friederici, 2006; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Sander, Frome, & Scheich, 2007; Sander & Scheich, 2001; Sauter & Eimer, 2010; Scott, Sauter, & McGettigan, 2009; Shinkareva et al., 2014), on identifying non-symbolic, low-level acoustic features that contribute to the evaluation of a wide range of sounds by using the approach of computational modelling (e.g. Weninger, Eyben, Schuller, Mortillaro, & Scherer, 2013), and on multisensory integration of emotional information (e.g. Dolan, Morris, & de Gelder, 2001; Pourtois, de Gelder, Bol, & Crommelinck, 2005).
2. Sample size was determined by considerations about the reliability of mean ratings (see Materials).
3. All correlations are associated with $p < .001$. However, due to the multimodal distribution of the norm ratings, inferential statistics might be biased. Thus, the correlations should be dominantly taken as a descriptive index of the correspondence between brief segments ratings and the full ratings.
4. Alternatively, we conducted a 3 (valence) × 2 (animacy category: animate vs. inanimate) × 3 (duration) MANOVA. All effects reported below are essentially the same in this analysis. Additionally, there were significant effects involving animacy. However, for the sake of succinctness and because these effects are rather uninteresting due to their ambiguity (i.e. they might be an effect of better discriminability of one category relative to the other or they might reflect a response bias), we report only the reduced analysis.

References

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology*, *105*, 181–199. doi:10.2307/1423027

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*, 120–131. doi:10.1016/j.ijpsycho.2011.08.003

Bermeitinger, C., Wentura, D., & Frings, C. (2011). How to switch on and switch off semantic priming effects for natural and artifactual categories: Activation processes in category memory depend on focusing specific feature dimensions. *Psychonomic Bulletin & Review*, *18*, 579–585. doi:10.3758/s13423-011-0067-z

Bradley, M. M., & Lang, P. J. (2007). *The International Affective Digitized Sounds: Affective ratings of sounds and instruction manual* (2nd ed., IADS-2, Tech. Rep. B-3). Gainesville, FL: University of Florida.

Bregman, A. S. (1990). *Auditory scene analysis. The perceptual organization of sound*. Cambridge, MA: MIT Press.

Bröckelmann, A.-K., Steinberg, C., Dobel, C., Elling, L., Zwanzger, P., Pantev, C., & Junghöfer, M. (2013). Affect-specific modulation of the N1m to shock-conditioned tones: Magnetoencephalographic correlates. *European Journal of Neuroscience*, *37*, 303–315. doi:10.1111/ejn.12043

Bröckelmann, A.-K., Steinberg, C., Elling, L., Zwanzger, P., Pantev, C., & Junghofer, M. (2011). Emotion-associated tones attract

enhanced attention at early auditory processing: Magnetoencephalographic correlates. *The Journal of Neuroscience*, *31*, 7801–7810. doi:10.1523/jneurosci.6236-10.2011

Carroll, K., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology A*, *58*, 1173–1197. doi:10.1080/02724980443000539

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Czigler, I., Cox, T. J., Gyimesi, K., & Horváth, J. (2007). Event-related potential study to aversive auditory stimuli. *Neuroscience Letters*, *420*, 251–256. doi:10.1016/j.neulet.2007.05.007

De Houwer, J., Crombez, G., Baeyens, F., & Hermans, D. (2001). On the generality of the affective Simon effect. *Cognition and Emotion*, *15*, 189–206. doi:10.1080/02699930125883

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion*, *12*, 45–62. doi:10.1080/026999398379772

Degner, J. (2011). Affective priming with auditory speech stimuli. *Language and Cognitive Processes*, *26*, 1710–1735. doi:10.1080/01690965.2010.532625

Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences*, *98*, 10006–10010. doi:10.1073/pnas.171288598

Everaert, T., Spruyt, A., & De Houwer, J. (2013). On the malleability of automatic attentional biases: Effects of feature-specific attention allocation. *Cognition and Emotion*, *27*, 385–400. doi:10.1080/02699931.2012.712949

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030–1044. doi:10.1037/0096-1523.18.4.1030

Folyi, T., Liesefeld, H. R., & Wentura, D. (2015). *Attentional enhancement for positive and negative tones at an early stage of auditory processing*. Manuscript submitted for publication.

Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience*, *24*, 1725–1741. doi:10.1162/jocn_a_00213

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*, 145–146. doi:10.1038/nn1392

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*, 887–892. doi:10.1038/nrn1538

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion*, *15*, 143–165. doi:10.1080/02699930125768

Kiefer, M., & Martens, U. (2010). Attentional sensitization of unconscious cognition: Task sets modulate subsequent masked semantic priming. *Journal of Experimental Psychology: General*, *139*, 464–489. doi:10.1037/a0019561

King, A. J., & Nelken, I. (2009). Unraveling the principles of auditory cortical processing: Can we learn from the visual system? *Nature Neuroscience*, *12*, 698–701. doi:10.1038/nn.2308

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–50). Mahwah, NJ: Erlbaum.

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. doi:10.1037/0278-7393.23.1.246

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. doi:10.1002/hbm.20180

Marin, M. M., Gingras, B., & Bhattacharya, J. (2012). Crossmodal transfer of arousal, but not pleasantness, from the musical to the visual domain. *Emotion*, *12*, 618–631. doi:10.1037/a0025020

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*, 1410–1421. doi:10.1016/S0028-3932(03)00017-4

Moors, A. (2015). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*. Advance online publication. doi:10.1146/annurev-psych-122414-033550

Moors, A., & De Houwer, J. (2001). Automatic appraisal of motivational valence: Motivational affective priming and Simon effects. *Cognition and Emotion*, *15*, 749–766. doi:10.1080/02699930143000293

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297–326. doi:10.1037/0033-2909.132.2.297

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544–2590. doi:10.1016/j.clinph.2007.04.026

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*, 466–478. doi:10.1037/0096-3445.130.3.466

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*, 483–522. doi:10.1037/0033-295X.108.3.483

Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, *41*, 49–59. doi:10.1016/S0010-9452(08)70177-1

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception*, *16*, 175–186.

Sander, K., Frome, Y., & Scheich, H. (2007). FMRI activations of amygdala, cingulate cortex, and auditory cortex by infant laughing and crying. *Human Brain Mapping*, *28*, 1007–1022. doi:10.1002/hbm.20333

Sander, K., & Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Cognitive Brain Research*, *12*, 181–198. doi:10.1016/S0926-6410(01)00045-3

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience*, *22*, 474–481. doi:10.1162/jocn.2009.21215

Scherer, L. D., & Larsen, R. J. (2011). Cross-modal evaluative priming: Emotional sounds influence the processing of emotion words. *Emotion*, *11*, 203–208. doi:10.1037/a0022588

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*, 228–233. doi:10.1016/S0926-6410(02)00108-8

Scott, S. K., Sauter, D., & McGettigan, C. (2009). Brain mechanisms for processing perceived emotional vocalizations in humans. In S. Brudzynski (Ed.), *Handbook of mammalian vocalizations: An integrative neuroscience approach* (pp. 187–198). Oxford: Academic Press.

Shinkareva, S. V., Wang, J., Kim, J., Facciani, M. J., Baucom, L. B., & Wedell, D. H. (2014). Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data. *Human Brain Mapping*, *35*, 3558–3568. doi:10.1002/hbm.22421

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. doi:10.1037/0033-2909.86.2.420

Sollberger, B., Rebe, R., & Eckstein, D. (2003). Musical chords as affective priming context in a word-evaluation task. *Music Perception*, *20*, 263–282. doi:10.1525/mp.2003.20.3.263

Spruyt, A., De Houwer, J., Everaert, T., & Hermans, D. (2012). Unconscious semantic activation depends on feature-specific attention allocation. *Cognition*, *122*, 91–95. doi:10.1016/j.cognition.2011.08.017

Spruyt, A., De Houwer, J., & Hermans, D. (2009). Modulation of automatic semantic priming by feature-specific attention allocation. *Journal of Memory and Language*, *61*, 37–54. doi:10.1016/j.jml.2009.03.004

Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007). Affective priming of nonaffective semantic categorization responses. *Experimental Psychology*, *54*, 44–53. doi:10.1027/1618-3169.54.1.44

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience*, *23*, 604–621. doi:10.1162/jocn.2009.21383

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Vermeulen, N., Corneille, O., & Luminet, O. (2007). A mood moderation of the extrinsic affective Simon task. *European Journal of Personality*, *21*, 359–369. doi:10.1002/per.621

Voß, A., Rothermund, K., & Wentura, D. (2003). Estimating the valence of single stimuli: A new variant of the affective Simon task. *Experimental Psychology*, *50*, 86–96. doi:10.1026//1618-3169.50.2.86

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585–594. doi:10.1016/j.tics.2005.10.011

Weninger, F., Eyben, F., Schuller, B. W., Mortillaro, M., & Scherer, K. R. (2013). On the acoustics of emotion in audio: What speech, music, and sound have in common. *Frontiers in Psychology, Emotion Science*, *4*, 1–12. doi:10.3389/fpsyg.2013.00292

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York: Guilford.

Wentura, D., Müller, P., & Rothermund, K. (2014). Attentional capture by evaluative stimuli: Gain- and loss-connoting colors boost the additional-singleton effect. *Psychonomic Bulletin & Review*, *21*, 701–707. doi:10.3758/s13423-013-0531-z

Wentura, D., & Rothermund, K. (2003). The "meddling-in" of affective information: A general model of automatic evaluation effects. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 51–86). Mahwah, NJ: Erlbaum.

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, *24*, 3–47. doi:10.1080/02699930903205698

Appendix

Table A1. Mean RTs (in ms) and error rates (in %, in parentheses) as a function of response and stimulus valence in Experiment 2.

	Experiment 2a (500 ms FSOA)[a]	Experiment 2b (0 ms FSOA)[a]
Positive response ("gut")		
Positive sound	1042 (3.9)	1051 (3.3)
Negative sound	1033 (2.6)	1031 (3.7)
Neutral sound	1023 (3.3)	1035 (3.5)
Negative response ("böse")		
Positive sound	1048 (3.9)	1023 (3.3)
Negative sound	992 (2.1)	990 (2.7)
Neutral sound	1016 (2.3)	983 (1.7)

[a] FSOA = Feature Start Onset Asynchrony.