Published in final edited form as:
Emotion. 2019 March ; 19(2): 219–233. doi:10.1037/emo0000429.

Automaticity in the Recognition of Nonverbal Emotional Vocalizations

César F. Lima[1,2,3], **Andrey Anikin**[4], **Ana Catarina Monteiro**[2], **Sophie K. Scott**[3], and **São Luís Castro**[2]

[1]Faculty of Psychology and Education Sciences, University of Porto, Portugal

[2]Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal

[3]Institute of Cognitive Neuroscience, University College London, UK

[4]Division of Cognitive Science, Department of Philosophy, Lund University, Sweden

Abstract

The ability to perceive the emotions of others is crucial for everyday social interactions. Important aspects of visual socio-emotional processing, such as the recognition of facial expressions, are known to depend on largely automatic mechanisms. However, whether and how properties of automaticity extend to the auditory domain remains poorly understood. Here we ask if nonverbal auditory emotion recognition is a controlled deliberate or an automatic efficient process, using vocalizations such as laughter, crying, and screams. In a between-subjects design (N = 112), and covering eight emotions (four positive), we determined whether emotion recognition accuracy (1) is improved when participants actively deliberate about their responses (compared to when they respond as fast as possible), and (2) whether it is impaired when they respond under low and high levels of cognitive load (concurrent task involving memorizing sequences of six or eight digits, respectively). Response latencies were also measured. Mixed-effects models revealed that: recognition accuracy was high across emotions, and only minimally affected by deliberation and cognitive load; the benefits of deliberation and costs of cognitive load were significant mostly for positive emotions, notably amusement/laughter, and smaller or absent for negative ones; response latencies did not suffer under low or high cognitive load; and high recognition accuracy (approximately 90%) could be reached within 500 ms after the stimulus onset, with performance exceeding chance-level already between 300-360 ms. These findings indicate that key features of automaticity, namely fast and efficient/effortless processing, might be a modality-independent component of emotion recognition.

Keywords

automaticity; cognitive load; deliberation; emotion recognition; nonverbal vocalizations

Correspondence to César F. Lima, Instituto Universitário de Lisboa (ISCTE-IUL), Avenida das Forças Armadas, 1649-026 Lisboa, Portugal. cesar.lima@iscte-iul.pt.

Author Note
This research was supported by grants from the Portuguese Foundation for Science and Technology, COMPETE and FEDER programs (UID/PSI/00050/2013; POCI-01-0145-FEDER-007294), and from the Bial Foundation (N29/08). During the preparation of the manuscript, CFL was supported by an FCT Investigator Grant from the Portuguese Foundation for Science and Technology (IF/00172/2015).

Introduction

The human voice is a primary tool for emotional communication. Similarly to facial expressions or body postures, nonverbal vocalizations such as laughter, crying or sighs, provide a window into the intentions and emotions of others. Nonverbal vocalizations are distinct from emotional speech regarding their underlying production and perceptual mechanisms (Pell et al., 2015; Scott, Sauter, & McGettigan, 2010), and they reflect a primitive and universal form of communication, which can be compared to the use of voice by other species (Gruber & Grandjean, 2017; Juslin & Laukka, 2003; Sauter, Eisner, Ekman, & Scott, 2010; Scherer, 1995). Forced-choice classification studies indicate that listeners can recognize a wide range of emotions in vocalizations, even when they are heard in isolation and without contextual information (e.g., Belin, Fillion-Bilodeau, & Gosselin, 2008; Lima, Castro, & Scott, 2013; Sauter, Eisner, Calder, & Scott, 2010; Schröder, 2003; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009). This includes the recognition of negative emotions, such as anger, fear, or disgust, as well as of positive ones, such as amusement, achievement, or pleasure (Sauter & Scott, 2007) and seldom-studied states like awe, compassion, or enthusiasm (Simon-Thomas et al., 2009). Vocalizations are typically recognized with high accuracy, often above 70-80% correct on average (e.g., 68% in Belin et al., 2008; 86% in Lima et al., 2013; 70% in Sauter Eisner, Calder, et al., 2010; 81% in Schröder, 2003), and listeners' responses can be predicted from the low-level acoustic attributes of the stimuli, including their temporal features, amplitude, pitch, and spectral profile (Lima et al., 2013; Sauter, Eisner, Calder, et al., 2010). Perceiving vocal cues and evaluating their emotional meaning involves several brain systems, including the superior temporal gyri, motor and premotor cortices, and prefrontal systems, namely the inferior frontal gyrus, along with subcortical regions such as the amygdala (e.g., Banissy et al., 2010; Bestelmeyer, Maurage, Rouger, Latinus, & Belin, 2014; Lima et al., 2015; Scott et al., 1997; Warren et al., 2006). Three- to seven-month-old infants already show specialized brain responses to crying vocalizations in regions involved in affective processing (Blasi et al., 2011), children as young as 5-7 years are proficient at recognizing a range of positive and negative vocal emotions (Sauter, Panattoni, & Happé, 2013), and emotion recognition accuracy remains high across the adult life span (Lima, Alves, Scott, & Castro, 2014).

Although this provides compelling evidence that humans are tuned to decode emotional information in vocalizations, far less is known about the cognitive processes underlying this socio-emotional skill. Specifically, it remains unclear whether vocal emotion recognition depends on controlled deliberate processes or on processes that are relatively automatic. One possibility is that, when evaluating the emotional meaning of a vocal expression, listeners engage in controlled processes to consider different alternatives, based on which they formulate an effortful judgment about the expression. In line with this hypothesis, for the processing of emotional speech and facial expressions, several studies have reported associations between emotion recognition performance and executive and attentional abilities, both in healthy (Borod et al., 2000) and clinical groups (Breitenstein, Van Lancker, Daum, & Waters, 2001; Hoaken, Allaby, & Earle, 2007; Lima, Garrett, & Castro, 2013). It has also been found that attention significantly modulates brain responses to vocal emotional information (Bach et al., 2008; Sander et al., 2005). Additionally, in everyday social

interactions, vocalizations are typically embedded in rich contexts, where their meaning depends, for instance, on emotional cues from other modalities (e.g., facial expressions), on verbal information, or on whether they are produced in a volitional or a spontaneous way (Anikin & Lima, 2017; Scott, Lavan, Chen, & McGettigan, 2014). A significant degree of flexible situated processing could therefore be routinely required.

Alternatively, vocal emotion recognition could proceed in a largely automatic manner. The idea that important aspects of socio-emotional processing are highly automatic has a long history (e.g., Bargh, 1994; Ekman, 1977; Öhman, 1986; Öhman, Flykt, & Esteves, 2001), and it is often tied to accounts of emotion and cognitive processes as evolutionary adaptations (e.g., Bargh et al., 2012; Öhman, 1986; Tracy & Robbins, 2008). Most research on this topic has been conducted in the visual domain and on preconscious automaticity, focusing on how some processes operate outside of awareness, in an unintentional and uncontrollable fashion. For instance, fear-relevant pictures (e.g., snakes) are detected faster than fear-irrelevant ones (e.g., mushrooms) in visual search tasks, even if they are presented away from the spotlight of attention or in the context of a large number of distractors (Öhman et al., 2001), suggesting a preattentive detection of emotional stimuli. Subliminally presented facial expressions generate automatic facial mimicry (Dimberg, Thunberg, & Elmehed, 2000), elicit early event-related potentials (ERP) similarly to consciously perceived expressions (starting 140 ms after face onset; Kiss & Eimer, 2008), and can influence evaluations of subsequently presented stimuli (e.g., Winkielman, Berridge, & Wilbarger, 2005).

Critically, there are several forms and features of automaticity, and relevant to the current study is the observation that the conscious, explicit recognition of emotions in facial expressions can also show features of automatic processes, namely fast and efficient processing. *Efficiency* refers to the extent to which a process can be completed effortlessly, with minimal involvement of controlled cognitive resources. In other words, efficient processes can operate even when controlled resources are occupied with other tasks (for reviews on automaticity, Bargh, 1994; Bargh, Schwader, Hailey, Dyer, & Boothby, 2012; Moors & De Houwer, 2006; Tzelgov, 1997). Consistent with this, Tracy and Robins (2008) found that facial expressions could be accurately recognized, both under time constrains (within 600 ms) and in suboptimal attentional conditions, under cognitive load. That is, emotion recognition remained accurate when the controlled processes available for emotion recognition were limited by a concurrent mnemonic task that competed for central resources. This was observed for ‘basic’ emotions and for more complex ones, such as embarrassment, pride and shame. Additionally, encouraging participants to carefully deliberate about their response, as compared to when they responded as quickly as possible, benefited performance only slightly, and only for some emotions (4 out of 8 in Experiment 1, and 3 out of 10 in Experiment 2), further suggesting that facial emotion recognition is supported by fast and efficient processes, that are relatively independent of deliberation. This also has implications for everyday social interactions, where facial expressions typically have to be recognized quickly and in conditions of considerable noise and distraction. However, whether these findings extend to modalities outside of vision remains unknown. ERP evidence indicates that emotional vocalizations are differentiated from neutral sounds within 150 ms of exposure (Sauter & Eimer, 2010; see also Liu et al., 2012), and such

differentiation is observed even if detecting emotional sounds is irrelevant to the task (Pinheiro, Barros, & Pedrosa, 2015). This suggests an early automatic processing of emotional salience, but no studies have examined automaticity in the conscious access to the emotional meaning of vocalizations.

In the current study, we ask if, and to what degree, vocal emotion recognition proceeds efficiently (i.e., with minimal effort) or under controlled deliberate processing. Participants judged whether or not vocalizations expressed a given emotion category (yes/no decision) under one of four conditions, in a between-subjects design. In the deliberated condition, they were instructed to carefully deliberate about their response in order to be as accurate as possible, thus maximizing the engagement of controlled processing. In the fast condition, participants were instructed to respond as quickly as possible, following their first impressions. In two cognitive load conditions, participants were also instructed to respond as quickly as possible, and had to simultaneously perform a memory task, thus minimizing the amount of controlled cognitive resources available. The memory task consisted of rehearsing sequences of six (low load condition) or eight (high load condition) digits, which participants had to hold in memory for later recall. Similar cognitive load manipulations have been effectively used in previous studies on automaticity of emotion recognition and social judgments (e.g., Aviezer, Dudarev, Bentin, & Hassin, 2011; Bargh & Tota, 1988; Gilbert & Osborne, 1989; Tracy & Robbins, 2008), and more broadly in cognitive research involving dual-task paradigms (e.g., Karatekin, 2004; Ransdell, Arecco, & Levy, 2001). We hypothesized that, if vocal emotion recognition depends on controlled deliberate processes to an important extent, (1) careful deliberation should be associated with significantly higher recognition accuracy, as compared to when participants respond fast or under load; and (2) there should be a relationship between the level of cognitive load and recognition accuracy. The higher the load, the lower the emotion recognition performance. On the other hand, if vocal emotion recognition is an efficient process, recognition accuracy should be relatively stable over different levels of cognitive load, i.e., it should remain high when controlled resources are limited. It should also be independent of careful deliberation.

Two other questions are addressed. First, we included a wide range of positive (achievement, amusement, pleasure, relief) and negative emotions (anger, disgust, fear, and sadness) to explore if the putative role of controlled processes varies across categories, and if it relates to broader affective dimensions, namely arousal and valence. High arousal is associated with a larger early differentiation between emotional and neutral vocalizations (Sauter & Eimer, 2009), and distinct brain systems are engaged depending on the arousal and valence properties of vocalizations (Warren et al., 2006). Regarding valence, positive vocalizations could involve relatively more controlled and flexible processing than negative ones, considering evidence that they might be more dependent on learning and context: as compared to negative vocalizations, positive ones seem to vary more across cultures (Sauter, Eisner, Ekman, et al., 2010), do not elicit selective brain responses as early in development (Blasi et al., 2011), evidence for their rapid detection in the adult brain is less consistent (Liu et al., 2012; Sauter & Eimer, 2010), and data from infants suggest that they are modulated by learning during development (Soderstrom, Reimchen, Sauter, & Morgan, 2017).

Second, we measured latencies in addition to response accuracy, to ask if cognitive load produces slower responses, as it could be predicted if controlled processing played a preponderant role during emotion recognition. Crucially, we examine the relationship between latencies and accuracy when participants were instructed to respond as quickly as possible, to estimate how quickly participants can reach accuracy levels above chance, i.e. how fast vocal emotion recognition can be. For emotional speech, evidence from gating experiments indicates that listeners can recognize emotions rapidly, with performance reaching high accuracy levels after hearing approximately 400-800 ms of an utterance (Jian, Paulmann, Robin, & Pell, 2015; Rigoulot, Wassiliwizky, & Pell, 2013), but for nonverbal vocalizations this question remains unanswered. Finally, as control measures, we examined participants' auditory perceptual abilities and collected information about their musical training, to ensure that any potential effects of controlled processing could not be attributed to these confounds. Both auditory perceptual abilities and musical training predict vocal emotion recognition in the context of emotional speech (Globerson, Amir, Golan, Kishon-Rabin, & Lavidor, 2013; Lima & Castro, 2011).

Method

Participants

One hundred and twelve undergraduate students from the University of Porto took part in the study for course credit or payment (Mage = 20.8 years; SD = 2.6; 95 female). They were randomly allocated to one of four conditions, in a between-subjects design: (1) deliberated; (2) fast; (3) low load; and (4) high load (n = 28 in each condition). Participants had normal or corrected-to-normal visual acuity, normal hearing, and were tested in individual sessions lasting around 45 minutes. Thirty-nine participants reported having had formal musical training, including instrumental practice (M = 5.7 years of training; SD = 4.7). The number of trained and untrained participants (χ^2 = 2.95, df = 3, p = .40) and the number of years of training (F[3,108] = 0.72, p = .54) were similar across conditions.

Informed consent was obtained from all participants, and the study was performed in accordance with the relevant guidelines and regulations.

Stimuli

The experimental stimulus set consisted of 80 brief purely nonverbal vocalizations (e.g., laughs, screams, sobs, sighs; emblems such as 'yuck' were not included). They were taken from validated corpora used in previous studies (Lima et al., 2014; Lima et al., 2013; Sauter, Eisner, Calder, et al., 2010; Sauter & Scott, 2007) and expressed eight emotions, four positive and four negative ones (10 tokens per emotion): achievement, amusement, pleasure, relief, anger, disgust, fear, and sadness. Eight speakers, four women and four men (aged 27 to 43 years), generated these vocalizations. The validation procedures showed that all vocalizations are recognized with high accuracy, and that their acoustic features provide sufficient information to permit automatic emotion classification and to predict listeners' emotion responses. The final set of expressions used here was selected based on a pilot study (N = 40, 20 provided categorization accuracy data and 20 provided intensity, arousal and valence data; none of these participants took part in the main study). We ensured that (1) all

emotion categories were matched for duration (F[7,72] = 0.87, p = .54), categorization accuracy (likelihood-ratio test, L = 11.9, df = 1, p = .10), and perceived intensity (L = 8.4, df = 7, p = .30); and that (2) positive and negative emotions were similar in duration (F[1,78] = 0.12, p = .73), intensity (L = 0.00, df = 1, p = 1), arousal (L = 0.0002, df = 1, p = .99), and categorization accuracy (L = 0.03, df = 1, p = .85). The characteristics of the stimuli are summarized in Table 1.

Design and Procedure

In all the four conditions, participants completed 16 blocks of 12 trials each (total 192 trials). Each block was assigned a target emotion (there were two blocks per emotion), and participants performed a yes/no decision, indicating whether each of the 12 vocalizations of that block expressed the target emotion or not (e.g., amusement in the amusement block). Five vocalizations in each block expressed the target emotion (*experimental* trials), and seven did not (*filler* trials); these non-target expressions included one example of each of the remaining seven emotions. The vocalizations used for the filler trials were selected from the same corpora as the experimental expressions, expressed the same emotion categories, were generated by the same speakers, and consisted of 112 stimuli (14 per emotion). There was no overlap between the experimental and filler stimulus sets (total number of unique vocalizations = 192), and none of the vocalizations was presented more than once throughout the experiment.

The order of the vocalizations was randomized within each block, and the order of the blocks was pseudo-randomized, ensuring that the two blocks of the same emotion were not presented consecutively. The vocalizations were played through headphones, and no feedback was given concerning response accuracy. A short familiarization phase preceded the task, and participants were informed about all the emotions that they would be asked to recognize (the emotion labels were introduced, alongside illustrative real-life scenarios for each emotion; for details, see Lima et al., 2013). Responses were collected via key presses (the order of the ‘yes’ and ‘no’ keys was counter-balanced across participants), and the stimuli were presented using SuperLab version 4.0 (Abboud, Schultz, & Zeitlin, 2006), running on an Apple MacBook Pro. Both response accuracy and latencies were collected. Latencies were measured from the onset of the vocalization until the key press.

In the *deliberated* condition, participants were instructed to respond as accurately as possible, and encouraged to take their time to think carefully before making a decision. In the *fast* condition, participants were instructed to make their decisions as quickly as possible, and encouraged to follow their first impressions in completing the task. In the two conditions with *load*, the instructions were the same as in the fast condition (i.e., participants made their decisions as quickly as possible), but participants were asked to perform a second task in addition to the emotion recognition one: before the beginning of each block, a sequence of digits was presented on the screen for 25 seconds, and participants were instructed to use that time to memorize it; they then completed the emotion recognition block, and were asked to recall the sequence of digits afterwards (for a similar procedure, Tracy & Robins, 2008). In the *low load* condition, the sequences to be memorized had 6 digits, and in the *high load*

Europe PMC Funders Author Manuscripts

condition they had 8 digits. Similarly to the emotion recognition task, no feedback was given concerning response accuracy.

Psychoacoustic Tasks

Participants' frequency discrimination and processing speed thresholds were determined using a 'two-down one-up' adaptive staircase procedure (Hairston & Maldjian, 2009). In the frequency discrimination task, participants listened to two 300 ms steady pure tones in each trial, and indicated which one was the highest. One of the tones was always presented at the same frequency (1000 Hz) and the other one at a higher frequency, varying adaptively from 1 to 200 Hz higher. The initial frequency difference was 100 Hz; correct responses led to progressively smaller differences until participants stopped responding correctly, and incorrect responses led to progressively larger differences, until participants responded correctly again. In the processing speed task, participants also indicated which of two tones in each trial was the highest, but what varied adaptively was the time difference between the onset of the first and of the second tones (stimulus onset asynchrony, SOA); correct responses led to progressively shorter SOAs, and incorrect responses led to longer SOAs (the higher tone was always presented at 660 Hz, and the lower one at 440 Hz; the initial SOA was 100 ms, and it varied between 1 ms and 150 ms). Both the frequency discrimination and the processing speed tasks ended after 14 reversals (i.e., changes in the direction of the stimulus difference), and thresholds were calculated using the arithmetic mean of the last 8 reversals. The initial step size was 10 Hz in the frequency discrimination task (10 ms in the processing speed task), it was divided by 2 after 4 reversals, and a final step size of 1 Hz (1 ms in the processing speed task) was reached after 8 reversals. This process converged on perceptual thresholds associated with a performance level of 70.7%.

Statistical Analysis

The effects of condition, emotion, and of other predictors on emotion recognition performance were examined in a series of logistic generalized linear mixed models (GLMM) for unaggregated data, with random intercepts per participant and per vocalization (separate analyses were conducted for hit rates and false alarms; significance was tested using likelihood ratio tests, L). These frequentist GLMMs were fit using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). They were complemented with Bayesian inference, which was used to contrast specific conditions and combinations of conditions, and to estimate the effects of valence and arousal on accuracy and on the contrasts between conditions. An advantage of employing Bayesian methodology is its flexible technique for controlling for multiple comparisons, namely shrinkage of regression coefficients (Kruschke, 2014). When simultaneously estimating a large number of coefficients (for example, 32 in a model with interaction between condition and emotion), we used shrinkage by imposing a horseshoe prior on all coefficients except the intercept (Carvalho, Polson, & Scott, 2009). All beta-coefficients in models with shrinkage are assumed to belong to the same distribution, the parameters of which are estimated from the data. This ensures that multiple comparisons between factor levels do not inflate the risk of false positives, helping to avoid type I errors.

For other Bayesian analyses without shrinkage, we specified mildly informative conservative priors centered at zero, as this improves convergence of complex mixed models and guards against over fitting (McElreath, 2015). Posterior distributions were summarized by taking the median and 95% credible interval (CI) over individual steps in the Markov Chain Monte Carlo (MCMC). Unless strong priors are specified, Bayesian CIs are often numerically comparable to confidence intervals, but more intuitive to interpret (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016): they contain a certain proportion of the posterior probability. That is, given the model and the observed data, the most credible value of an estimated parameter is 95% likely to lie within its 95% CI. When contrasting two conditions (e.g. deliberated vs. three remaining conditions), the entire CI indicates the most credible values for the difference; if it does not include zero, this can be taken as evidence in favor of an actual difference between those conditions. All generative models were fit using the Stan computational framework (http://mc-stan.org/) and the brms package (Buerkner, 2017).

The effects of condition and emotion on the time participants took to correctly recognize the target expressions were also examined. Errors and outliers (latencies below 250 ms or exceeding the mean of each participant by 3 *SD*) were not included in this analysis. Latencies were approximately normally distributed after a log transformation, and Gaussian models were applied.

The time needed to perform the task with accuracy above chance level (50%), and to reach peak accuracy level, was estimated by examining the relationship between latencies of all responses (correct and incorrect) and overall emotion recognition accuracy (including hits and correct rejections of filler expressions) in the timed conditions (fast, low load and high load). This relation was nonlinear over the full range of response latencies and not satisfactorily captured by a logistic model with a polynomial term. To model this latency-accuracy function, we therefore used smooth regression, namely generalized additive mixed models (GAMM; Wood, 2006). This model was fit using the brms package (Buerkner, 2017) with random intercepts per participant and per vocalization and a smoothing term for log transformed latencies. As above, mildly informative conservative priors were used.

All analyses were performed in R 3.2.2 (https://www.r-project.org). The code used for data analysis and the full data set are provided in Supplemental Materials.

Results

Vocalizations were recognized with high accuracy (M = 90.8% hits across conditions, i.e., correctly pressing ‘yes’ when the vocalization expressed the target emotion), well above the chance level (50%). Accuracy rates were high across conditions, even under the two levels of cognitive load: 94.6% in the deliberated condition, 89.8% in the fast condition, 91.4% in the low load condition, and 87.6% in the high load condition. These high rates cannot be explained by a bias to use the ‘yes’ key for any vocalization, as false alarms (i.e., incorrectly pressing ‘yes’ for filler vocalizations) were low (M = 6.7%), also across conditions: 5.4% in the deliberated condition, 8.2% in the fast condition, 5.4% in the low load condition, and 7.7% in the high load condition. Thus, participants’ ability to recognize that a particular vocalization did not express the target emotion was also high. The median of posterior

distribution and 95% CI for hits and false alarms are depicted in Figure 1, separately for each condition and emotion.

Recognition of Target Emotions Across Conditions

Although accuracy rates were generally high, differences between conditions were significant ($L = 16.7$, $df = 3$, $p < .001$). The effect of emotion ($L = 27.8$, $df = 7$, $p < .001$) and the interaction between condition and emotion ($L = 60.9$, $df = 21$, $p < .001$) were also significant, indicating that deliberation and cognitive load affected accuracy differently across emotions (accuracy differences between conditions, i.e., the magnitude of the effects, are depicted in Figure 2, separately for each emotion). To follow up on these effects, we first focused on whether deliberation improved the recognition of target emotions. Hit rates were 2.8%[1] higher in the deliberated as compared to the fast condition (95% CI [0.8, 5.5]), and 3.2% higher as compared to the average of the three other conditions (95% CI [1.6, 5.1]), indicating a significant, yet small, benefit of thinking carefully before responding (deliberated vs. low load conditions, +1.9%, 95% CI [0.1, 4.3]; deliberated vs. high load conditions, +4.7%, 95% CI [2.3, 8.1])[2]. Looking at specific emotions, the benefits of deliberation (vs. three other conditions) were 5.6% for amusement (95% CI [2.2, 11.0]) and 1.6% for pleasure (95% CI [0.1, 4.0]). For the remaining emotions, the trend was in the same direction, but the 95% CI included zero, providing no clear evidence for a benefit (see Figure 2). When comparisons were conducted between the deliberated and each of the other conditions separately, benefits were found for amusement across comparisons (deliberated vs. fast conditions, marginal effect +0.9%, 95% CI [-0.2, 3.8]; deliberated vs. low load conditions, +6.7%, 95% CI [2.3, 14.4]; deliberated vs. high load conditions, +8.0%, 95% CI [2.8, 16.5]), and additionally for pleasure and relief in the deliberated vs. high load comparison (pleasure, +3.8%, 95% CI [0.7, 9.5]; relief, (+3.4%, 95% CI [0.3, 9.1]).

We then focused on the potential negative effects of cognitive load. Hit rates were generally similar in the fast condition as compared to the average of the two load conditions, suggesting that there was no general cost of recognizing the target expressions under divided attention (+0.6% in the fast condition, 95% CI [-2.0, 3.1]). The only exception was amusement, for which accuracy was 5.6% higher in the fast condition than in the load conditions (95% CI [1.3, 11.8]). When the two load conditions were directly compared, accuracy rates were only marginally higher (+2.8%) in the low vs. high load conditions (95% CI [-0.1, 6.3]). The benefits of low vs. high load were apparent for relief (+3.9%, 95% CI [0.9, 9.3]), pleasure (+3.8%, 95% CI [0.8, 9.8], and disgust (+2.6%, 95% CI [0.1, 6.9]), but for the remaining emotions the evidence for a benefit was less clear.

These findings provide evidence for positive effects of deliberation and negative effects of cognitive load in the recognition of vocal emotions, but small and limited to a reduced set of mostly positive emotions, particularly amusement. To test if the pattern remained unaltered when emotion recognition was more difficult, and thus potentially more dependent on

[1]Here and elsewhere, reported difference scores are taken from the estimated models and not from the observed data, i.e., they reflect predicted (fit) values

[2]For completeness, we have also used a more standard frequentist approach to evaluate this and the remaining main comparisons of the current work (Wald tests or t-tests with Bonferroni correction for multiple comparisons). The general pattern of results was consistent with the Bayesian inferences, as detailed in Supplemental Materials.

effortful processing, we replicated the analysis focusing on the vocalizations that were least well recognized in the main experiment (5 vocalizations for each emotion, 40 in total). Accuracy for this subset of 40 vocalizations was 85.8% on average (vs. 95.9% for the remaining 40 vocalizations; see details in Supplemental Materials), and the interaction between condition and emotion was again significant ($L = 44.9$, $df = 21$, $p = .002$). The benefits of deliberation were only slightly larger than in the analysis on the full set of vocalizations (+5.2% for deliberated vs. fast condition, 95% CI [1.5, 9.5]; +6.1% for deliberated vs. three remaining conditions, 95% CI [3.1, 9.5]), and, importantly, the 95% CI excluded 0 for amusement only (+10% in the deliberated condition, 95% CI [2.6, 18.8]). Hit rates were similar in the fast condition as compared to the average of the two load conditions (+1.4% in the fast condition, 95% CI [-3.0, 5.7]), except for amusement (+9.5% in the fast condition, 95% CI [0.9, 19.4]). Hit rates were also similar between the high and the low load conditions (+3.7% in the low load condition, 95% CI [-1.4, 9.1]). The benefits of low vs. high load were most evident for the same emotions as in the full analysis, though the effects were only seen at the trend level: relief (+4.5%, 95% CI [-0.2, 12.2]), pleasure (+4.1%, 95% CI [-0.7, 11.5], and disgust (+4.7%, 95% CI [-0.8, 13.1]). Thus, we found no evidence for stronger effects of condition, or for a different pattern of results, even when emotion recognition was more challenging.

Additionally, we also wanted to ensure that the pattern of emotion-specific results (i.e., effects of deliberation and cognitive load mostly for positive emotions) was not an artifact of other attributes of the stimuli such as ambiguity, emotional intensity and duration. Based on data from the pilot study, we computed categorization accuracy and perceived intensity for each stimulus and included these measures, along with stimulus duration, as covariates in the model for predicting accuracy in different testing conditions in the main experiment. As expected, higher categorization accuracy ($L = 7.1$, $df = 1$, $p < .007$) and higher intensity ($L = 20.1$, $df = 1$, $p < .001$) at the pilot stage predicted higher recognition accuracy in the main experiment. In contrast, duration had no effect ($L = 0.06$, $df = 1$, $p = .81$). Crucially, adding pilot accuracy, intensity and duration as covariates did not change the pattern of emotion-specific effects across conditions. As before, the benefit of deliberation (vs. three other conditions) was particularly clear for amusement (+5.3%, 95% CI [2.4, 9.4]) and pleasure (+2.1%, 95% CI [0.4, 5.0]), while for the other emotions the 95% CI included zero. The emotion-specific negative effects of cognitive load were also replicated: accuracy was 5.2% higher in the fast condition than in the load conditions (95% CI [1.4, 10.3]) for amusement (for the remaining emotions the 95% CI included zero); and the benefits of low vs. high load were apparent for relief (4.2%; 95% CI [1.1, 9.3]), pleasure (4.9%; 95% CI [1.2, 10.9], and disgust (3.2%; 95% CI [0.1, 7.9]), but for the remaining emotions the evidence for a benefit was less clear.

False Alarms Across Conditions

The effect of condition on false alarm rates was marginally significant ($L = 7.9$, $df = 3$, $p = .05$), and the Condition x Emotion interaction was significant ($L = 39.9$, $df = 21$, $p = .01$; main effect of emotion, $L = 3.4$, $df = 7$, $p = .84$). Deliberation was associated with slightly fewer false alarms, both when compared with the fast condition (-1.5%; 95% CI [0.2, 3.1]) and when compared with the average of the three remaining conditions (-0.8%; 95% CI [0.0,

Europe PMC Funders Author Manuscripts

1.8]; see Figure 2b, 'Overall' metric; deliberated vs. low load conditions, 0.0%, 95% CI [-1.0, 0.9]; deliberated vs. high load conditions, -0.9%, 95% CI [-2.5, 0.1]). Looking at specific emotions, the benefits of deliberation (vs. three other conditions) were apparent for amusement (-1.3%; 95% CI [0.1, 3.3]), but not for the remaining emotions. When comparisons were conducted between the deliberated and each of the other conditions separately, no differences were apparent, apart from a benefit for amusement in the deliberated vs. high load comparison (-3.1%, 95% CI % [-7.3, -0.3]).

No evidence for general negative effects of cognitive load on false alarms was found; there was actually a tendency for lower false alarms in the two load conditions as compared to the fast condition (-0.9%; 95% CI [-2.5, 0.2]), with no difference between the low and high load conditions (-0.9% in the low load condition; 95% CI [-2.4, 0.2). Looking at specific emotions, only amusement was associated with fewer false alarms in the low load vs. high load conditions (-2.6%, 95% CI [-6.6, 0.0]).

Potential Roles of Valence and Arousal

We took arousal and valence ratings of the experimental vocalizations, i.e., perceived arousal and valence based on the pilot study, and examined how these dimensions modulated accuracy rates. No effect of arousal was found (main effect of arousal, $L = 0.1$, $df = 1$, $p = .74$; interaction Arousal x Condition, $L = 6.1$, $df = 3$, $p = .10$). However, valence significantly predicted how participants recognized the target expressions: higher valence (i.e., more positive vocalizations) was associated with higher hit rates, primarily in the deliberated condition. In other words, positive vocalizations, more than negative ones, significantly benefited when participants were encouraged to take their time to think about their responses (interaction Valence x Condition, $L = 21.4$, $df = 3$, $p < .001$; main effect of valence, $L = 3.0$, $df = 1$, $p = .09$; see Table 2). Additionally, the magnitude of the benefits of deliberation was numerically larger for more positive vocalizations, i.e., there was a positive relationship between valence and the magnitude of accuracy differences between the deliberated and fast condition, and between the deliberated and the three remaining conditions. These associations are illustrated in Figure 3a. Recognition accuracy for sounds of the most positive valence was predicted to be 2.8% higher in the deliberated as compared to the fast condition (95% CI [1.2, 5.7]), while for sounds of the most negative valence the difference was only 1.6% and non-significant (95% CI [-2.3, 5.7]). The difference between the deliberated and the three remaining conditions was 4.1% (95% CI [2.3, 7.0]) for the most positive vocalizations, and only 1.0% (95% CI [-2.4, 4.0]) for the most negative ones. A similar relationship was found between higher valence and differences in accuracy between the low and high cognitive load conditions (Figure 3a). The costs of higher load were 3.4% for the most positive vocalizations (95% CI [0.4, 8.1]), as compared to 2.1% and non-significant (95% CI [-1.7, 6.2]) for the most negative ones. Altogether, these findings suggest that the benefits of deliberation and the costs of cognitive load are relatively higher for more positive vocalizations.

Given that the effects of condition on hit rates were most apparent for amusement, we asked whether the modulatory effect of valence was solely driven by amusement vocalizations, or whether it was a more general effect. In an analysis excluding amusement vocalizations, the

interaction Valence x Condition remained significant (L = 12.2, df = 3, p = .01; main effect of valence L = 7.4, df = 1, p = .01): as can be seen in Table 2, even after excluding amusement vocalizations, valence was positively associated with hit rates primarily in the deliberated condition, and an additional positive association was also found in the low load condition. We also found that the numerical associations between valence and the magnitude of the benefits of deliberation (as well as the magnitude of the benefits of low vs. high load) remained similar after excluding amusement vocalizations (Figure 3b).

Latencies Across Conditions

On average, participants took 1332 ms to correctly recognize the target vocalizations in the deliberated condition, 969 ms in the fast condition, 975 ms in the low load condition, and 979 ms in the high load condition. Figure 4 depicts response latencies for each condition and emotion. The main effects of condition (L = 53.1, df = 3, p < .001) and emotion (L = 40.6, df = 7, p < .001) were significant, as was the interaction between condition and emotion (L = 60.5, df = 21, p < .001). As expected, latencies were higher in the deliberated as compared to the remaining conditions (+347 ms, 95% CI [252, 435]), confirming that participants did follow the instructions and took a longer time to think about their responses in this condition. As can be seen in Figure 4, the difference was significant for all emotions, and it varied between +212 ms for relief (95% CI [139, 303]) and +357 for sadness (95% CI [251, 466]). More importantly, when looking at the potential negative effects of cognitive load on the time taken to respond, we found no differences between the fast condition and the average of the two load conditions (-1 ms in the fast condition, 95% CI [-74, 87]), and no differences between the low load and the high load conditions (-3 ms in the low load condition, 95% CI [-80, 81]). For both contrasts, when looking at specific emotions, the 95% CI included 0 in all cases. We thus found no evidence for a cost of cognitive load in terms of the time participants needed to recognize the target expressions.

How quickly can nonverbal vocalizations be recognized?

To estimate how quickly participants could determine whether or not vocalizations expressed the target emotions, we examined the effect of latencies on overall accuracy (i.e., correct detection of target expressions and correct rejection of filler ones). These analyses were focused on the fast, low load and high load conditions together, as they all encouraged participants to be quick, and are therefore suitable to ask questions about the minimum amount of time needed for accurate responses (a separate analysis was conducted on the deliberated condition for completeness). Figure 5a shows observed accuracy in these three conditions as a function of latencies. Although only 2.5% of responses were provided under ~500 ms, thus increasing the margin of uncertainty within this range of latencies, accuracy rates significantly above 50% can already be seen between 300 and 360 ms (Figure 5a). Furthermore, performance reaches ~90% by ~500 ms and plateaus by ~600 ms. It is noteworthy that, when participants were encouraged to focus on being accurate and to take their time to respond (deliberated condition), no responses were faster than ~500 ms, as indicated by the separate analysis (Figure 5b). Additionally, the few responses provided between 500 and 600 ms were already highly accurate, further confirming that this time is sufficient to perform the task with high accuracy levels. Complementary analyses, separately for each emotion, showed that this time window is associated with high accuracy for all

emotions. Both in the conditions that emphasized fast responses and in the deliberated condition, we see that beyond a certain amount of time accuracy starts to decline: after ~1000 ms in the fast conditions, and after ~1500-2000 ms in the deliberated condition. This possibly reflects hesitation for vocalizations that might be more difficult to recognize, or a negative effect of taking more than a certain amount of time to ponder about responses.

Memory Task

Performance on the memory task (cognitive load conditions) was 9.6% (95% CI [7.8, 11.3]) higher in the low load (6 digits) condition (M = 91.0%, 95% CI [89.9, 92.0]) than in the high load (8 digits) condition (M = 81.4%, 95% CI [80.0, 82.7]). These percentages correspond to correctly recalling 5.5 digits on average (out of 6) in the low load conditions, and 6.5 (out of 8) in the high load condition. This finding confirms that the high load condition was indeed significantly more demanding than the low load one.

There was no overall relationship between performance levels in the memory task and performance levels in the emotion recognition task (L = 2.1, df = 1, p = .15), suggesting that vocal emotion recognition was not directly compromised by the amount of cognitive resources devoted to the second task. This was found across the two cognitive load conditions (Memory Task x Condition interaction, L = 0.18, df = 1, p = .67).

Psychoacoustic Processing, Musical Training, and Emotion Recognition

Participants' frequency discrimination thresholds were 65.2 Hz in the deliberated group, 60.4 Hz in the fast group, 55.3 Hz in the low load group and 33.5 Hz in the high load group. Processing speed thresholds were 101 ms in the deliberated group, 90 ms in the fast group, 98 ms in the low load group, and 72 ms in the high load group. For the two measures, there were no differences across groups, apart from an unexpected advantage of the high load group vs. deliberated group (frequency discrimination: main effect, F[3,107] = 3.5, p = .02; high load vs. deliberated groups, t = -2.9, p = .005; processing speed: F[3,107] = 2.7, p = .02; high load vs. deliberated groups, t = -2.4, p = .02). No associations were found between psychoacoustic thresholds and the recognition of target emotional expressions, though (frequency discrimination, L = 2.6, df = 1, p = .11; interaction Frequency Discrimination x Condition, L = 1.3, df = 3, p = .72; processing speed, L = 1.8, df = 1, p = .18; interaction Processing Speed x Condition, L = 2.4, df = 3, p = .49). The recognition of target emotional expressions was also not influenced by musical training (L = 1.9, df = 1, p = .17). In contrast, psychoacoustic thresholds were strongly predicted by musical training: participants with more years of musical training had lower thresholds, indicating better psychoacoustic processing abilities (frequency discrimination, R^2 = .27, F[1,109] = 41.1, p < .001; processing speed, R^2 = .17, F[1,109] = 22.1, p < .001).

Discussion

The current study examined whether emotion recognition in nonverbal vocalizations is a controlled deliberate or an automatic effortless process. To that end, we determined the effects of deliberation and cognitive load on response accuracy and latencies, covering a wide range of positive and negative emotions. We present four novel findings. First, emotion

recognition accuracy was generally high, and relatively stable across conditions: both the benefits of deliberation and the costs of cognitive load were small, and only observed for a reduced subset of emotions. Second, the deliberation and cognitive load effects were mostly seen for positive emotions, notably amusement/laughter, and they relate to the valence properties of the vocalizations more generally. Third, higher levels of cognitive load were not associated with costs in the time taken to correctly recognize vocalizations. Fourth, analyses of latency-accuracy functions indicated that high recognition accuracy (approximately 90% correct) can be reached within 500 ms of exposure to the vocalizations, with performance exceeding chance level accuracy already between 300-360 ms of exposure. These findings are discussed in the next paragraphs.

Although many studies have addressed automaticity in socio-emotional processing, the emphasis has often been on visual stimuli and on preconscious mechanisms, such as how subliminally presented facial expressions elicit emotional responses and modulate cognitive processes in an unintentional and uncontrollable way (e.g., Dimberg et al., 2000; Kiss & Eimer, 2008; Winkielman et al., 2005). Less is known about the automatic components of auditory emotional processing, and particularly concerning conscious, goal-directed mechanisms. These are mechanisms that involve higher-order conscious processes, such as explicit evaluations of emotional expressions, but that can show important features of automaticity, namely efficiency, i.e., an ability to operate with minimal dependence on controlled resources (Bargh, 2012). Our findings that vocal emotion recognition accuracy remained high in dual task conditions, under cognitive load, and improved only minimally when participants carefully deliberated about their responses, extend to the auditory domain previous results on the recognition of facial expressions (Tracy & Robins, 2008). They indicate that, like facial expressions, vocalizations can be recognized and discriminated with a high degree of efficiency, even under different levels of attentional distraction. This has implications for understanding the cognitive mechanisms underlying vocal emotional processing, but also for everyday social interactions, which are rapidly changing and require the simultaneous processing of multiple sources of information, often under suboptimal conditions of distraction and noise. It is thus highly functional to be able to quickly and effortlessly evaluate the meaning of vocal expressions, while simultaneously performing other tasks (e.g., keep a conversation; process emotional cues from other modalities).

Interestingly, we also found that the effortless nature of vocal emotion recognition might extend to stimuli that are relatively more ambiguous, as indicated by the analysis of the subset of least well recognized vocalizations. This effortlessness is further reflected in the time participants took to respond. Emotions were categorized as quickly when controlled resources were taxed by a competing task, as when the task was performed under full attention, i.e., latencies did not suffer under cognitive load, both when the load was low and when it was high. Taken together, these findings add to previous ERP research in important ways (Liu et al., 2012; Pinheiro et al., 2015; Sauter & Eimer, 2010), demonstrating that the automaticity of emotion decoding in nonverbal vocalizations can be seen at different stages of processing: in the early (unintentional) neural differentiation of emotional sounds vs. neutral ones, and in the later high-order processes involving the conscious access to the specific emotional meaning of vocalizations. An important consideration is whether the high accuracy rates that we obtained, and the small effects of cognitive load and deliberation

observed, truly reflect the efficiency of the mechanism, or rather a task-related bias, i.e., a tendency to use the 'yes' key regardless of whether the vocalization expressed or not the target emotion. However, the analysis of false alarms speaks against this interpretation: they were generally low across conditions (under 9% on average), indicating that (1) participants used the 'yes' key mostly when the vocalizations indeed expressed the target emotion, and that (2) the ability to decide that a vocalization does not express a given emotion also involves efficient mechanisms. Consistent with this, there were no overall costs of cognitive load in terms of the percentage of false alarms, and the benefits of deliberation were negligible.

Not only were the effects of deliberation and cognitive load small, but we also found that they varied across vocal emotions. Converging evidence from emotion-specific analyses and from an analysis of the valence properties of the stimuli (perceived valence) suggests that positive vocalizations benefited relatively more than negative ones from controlled deliberate processing. This was particularly evident in the case of amusement/laughter, for which both benefits of deliberation and costs of cognitive load were consistently found. However, this effect appears to be more general, since it extended to other positive emotions, namely pleasure and relief. Furthermore, perceived valence of the experimental stimuli significantly modulated the benefits of deliberation and costs of cognitive load[3]. Thus, while both negative and positive vocalizations can be efficiently recognized, it could be that the recognition of positive vocalizations is more susceptible to contextual/task effects, i.e., their processing might be relatively less automatized. One possibility is that this relates to the social function of positive emotions, as it was previously argued to account for the fact positive vocalizations of achievement, pleasure and relief are not universally recognized, while negative vocalizations are (Sauter, Eisner, Ekman, et al., 2010). The communication of positive emotions facilitates social cohesion and affiliative behaviour, mostly with in-group members – with whom it is highly advantageous to build and maintain social connections – and their meaning could therefore be culturally variable, more dependent on learning, and contextually situated. Indeed, three- to seven-month-old infants already show selective brain responses to crying vocalizations, but the same was not found for laughter, possibly reflecting an earlier specialization for negative vs. positive vocalizations (Blasi et al., 2011). ERP evidence from adults indicates that the early automatic differentiation between emotional and neutral vocalizations might be more robust for negative as compared to positive vocalizations (Liu et al., 2012; Sauter & Eimer, 2010). Recent behavioral evidence from infants further suggests that the discrimination of positive vocalizations is modulated by learning throughout development (Soderstrom et al., 2017). Additionally, in everyday social interactions, positive vocal expressions often convey different meanings depending on context, and it might therefore be advantageous that their interpretation incorporates deliberate processes to some extent. Laughter is a clear illustration of this: while it is typically taken as an expression of positive affect, laughter can reflect a variety of distinct emotional states (e.g., polite agreement; affection; amusement; anxiety; embarrassment), it can be associated with a spontaneous genuine reaction or with a more voluntary

[3]One of the negative expressions (disgust) was also affected by cognitive load, but we refrained from emphasizing this finding because it was observed for one contrast only (low load vs. high load conditions), and it was a marginally significant effect.

communicative act (e.g., social laughter; McGettigan et al., 2015; Scott et al., 2014), and it can even be perceived as a negative expression, for instance if associated with insults and bullying (Otten, Mann, van Berkum, & Jonas, 2017). The interpretation of laughter (and of other positive vocalizations) could thus routinely involve deliberate processes to allow for the flexible consideration of contextual cues to optimize performance. An alternative to such social function account would be that our emotion- and valence-specific effects are related to acoustic ambiguity, i.e., it could be that acoustic cues are more ambiguous (and more similar) across positive vocalizations, making them more susceptible to task condition effects. However, ambiguity in acoustic cues would arguably be reflected in recognition accuracy differences (ambiguous vocalizations would be more difficult to recognize), and we have shown in a follow-up analysis that the pattern of results remains unchanged when stimuli differences in pre-test accuracy and emotional intensity are accounted for. It thus seems unlikely that the reported findings are reducible to differences in low-level acoustic cues.

Interestingly, our findings suggest a moderating role of valence (but not arousal) in the degree of automaticity of vocal emotion recognition, whereas previous ERP evidence suggested a moderating role of arousal (but not valence) in the magnitude of the rapid neural detection of vocal emotions (Sauter & Eimer, 2010). This emphasizes the importance of considering the affective dimensions of the stimuli, in addition to specific emotion categories, if we are to gain a mechanistic understanding of vocal emotional processing (see also Lima et al., 2014; Warren et al., 2006). While the early neural detection of emotional salience might be more determined by the arousal properties of vocalizations, the higher-order explicit interpretation of emotional meaning might be more determined by their valence and associated complexity of social functions.

Additional evidence for the notion that vocal emotion recognition is an efficient and fast process was provided by the analysis of the relationship between latencies and accuracy. We examined emotion categorization accuracy as a function of the time taken to respond, and were able to estimate, both the minimum amount of time needed to reach accuracy levels above chance (~300-360 ms) and the amount of time needed to reach peak performance (~500-600 ms). It is important to note that the average duration of the vocalizations was ~1000 ms, and so participants were able to accurately recognize emotions well before they were exposed to the full expressions. These findings extend to nonverbal vocalizations the results previously obtained in the context of emotional speech, and using a gating paradigm. In this paradigm, stimuli are gated to different durations, thereby limiting the amount of temporal and acoustic information that participants can use to recognize emotions (Jian et al., 2015; Rigoulot et al., 2013). Above-chance accuracy rates can be observed after only 200 ms, and performance reaches high levels after approximately 400-800 ms of exposure to the utterance, a time window roughly similar to the one obtained in the current study. For the recognition of facial expressions, Tracy and Robins (2008) showed that accurate emotion discrimination could occur within 600 ms. More recently, using dynamic stimuli, Martinez, Falvello, Aviezer, and Todorov (2016) found that 250 ms of exposure to facial expressions might be enough for accuracy recognition, with performance rapidly increasing with longer exposures (500 ms and 1000 ms) and then reaching a plateau. In future studies it will be of interest to directly compare the time-course of emotion recognition across different types of

Europe PMC Funders Author Manuscripts

stimuli, both within the auditory modality (nonverbal vocalizations and emotional speech) and across modalities (auditory and visual modality). This is particularly relevant in light of ERP evidence showing distinct neural responses to nonverbal vocalizations and emotional speech (Pell et al., 2015), and behavioral evidence showing differences in emotion recognition accuracy across modalities (Hawk, van Kleef, Fischer, & van der Schalk, 2009).

The findings of the current study raise other interesting questions for future research. First, although we showed a similar pattern of deliberation and cognitive load effects across different levels of stimulus difficulty, emotion recognition accuracy was generally high. Thus, it remains to be determined if the degree of automaticity in vocal emotion recognition uncovered here, using stimuli previously validated to communicate the intended emotions in a clear way (Lima et al., 2013; Sauter, Eisner, Calder, et al., 2010), is also seen for highly ambiguous stimuli. Studies covering a wider range of stimulus ambiguity, and systematically manipulating this variable, will shed light on this question. Second, we focused on the recognition and discrimination of emotion categories, in line with the dominant approach in emotion research. However, recent work has shown that listeners can also reliably make more nuanced socio-emotional inferences from vocalizations, namely regarding emotional authenticity, i.e., to judge whether a vocalization reflects a genuine emotional state or a more volitional communicative act (Anikin & Lima, 2017; Lavan, Scott, & McGettigan, 2016; McGettigan et al., 2015; Scott et al., 2014). It will be interesting to ask whether the degree of automaticity is similar or different for the processing of different aspects of vocalizations. Finally, more developmental studies will be important in order to shed light, both on the relative role of learning/skill acquisition vs. predispositions in the automaticity of vocal emotional processing (e.g., Bargh et al., 2012) and on the potentially different trajectories of positive and negative vocal expressions.

To conclude, the present study forms the first demonstration that the recognition of nonverbal emotional vocalizations is a fast and efficient process. These are both key features of automatic processes, and they are relevant for the demands of everyday social interactions. Building on previous evidence from facial expressions, we showed that human vocalizations can be recognized fast and accurately, even when controlled cognitive resources are taxed by a concurrent task. Consistent with this, intentionally engaging in controlled deliberate processes improved emotion performance only minimally. These findings extend the automatic properties of emotion recognition to the auditory modality, and have implications for current debates on the neurobiology of vocal communication and on the automaticity of socio-emotional processes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Abboud, H, Schultz, WH, Zeitlin, V. SuperLab, Stimulus presentation software (Version 4.0). San Pedro, California: Cedrus Corporation; 2006.

Aviezer H, Dudarev V, Bentin S, Hassin RR. The automaticity of emotional face-context integration. Emotion. 2011; 11:1406–1414. DOI: 10.1037/a0023578 [PubMed: 21707150]

Anikin A, Lima CF. Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. The Quarterly Journal of Experimental Psychology. 2017; doi: 10.1080/17470218.2016.1270976

Bach DR, Grandjean D, Sander D, Herdener M, Strik WK, Seifritz E. The effect of appraisal level on processing of emotional prosody in meaningless speech. NeuroImage. 2008; 42:919–927. DOI: 10.1016/j.neuroimage.2008.05.034 [PubMed: 18586524]

Banissy MJ, Sauter D, Ward J, Warren JE, Walsh V, Scott SK. Suppressing sensorimotor activity modulates the discrimination of auditory emotions but not speaker identity. The Journal of Neuroscience. 2010; 30:13552–13557. DOI: 10.1523/JNEUROSCI.0786-10.2010 [PubMed: 20943896]

Bargh, JA. The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognitionHandbook of Social Cognition. Wyer, RJ, Srull, TK, editorsHillsdale, NJ: Erlbaum, Inc; 1994. 1–40.

Bargh JA, Tota ME. Context-dependent automatic processing in depression. Accessibility of negative constructs with regard to self but not others. Journal of Personality and Social Psychology. 1988; 54:925–939. DOI: 10.1037/0022-3514.54.6.925 [PubMed: 3397867]

Bargh JA, Schwader KL, Hailey SE, Dyer RL, Boothby EJ. Automaticity in social-cognitive processes. Trends in Cognitive Sciences. 2012; 16:593–605. DOI: 10.1016/j.tics.2012.10.002 [PubMed: 23127330]

Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015; 67:1–48. DOI: 10.18637/jss.v067.i01

Belin P, Fillion-Bilodeau S, Gosselin F. The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. Behavior Research Methods. 2008; 40:531–539. DOI: 10.3758/BRM.40.2.531 [PubMed: 18522064]

Bestelmeyer PE, Maurage P, Rouger JM, Belin P. Adaptation to vocal expressions reveals multistep perception of auditory emotion. The Journal of Neuroscience. 2014; 34:8098–8105. DOI: 10.1523/JNEUROSCI.4820-13.2014 [PubMed: 24920615]

Blasi A, Mercure E, Lloyd-Fox S, Thomson A, Brammer M, Sauter D, et al. Murphy DGM. Early specialization for voice and emotion processing in the infant brain. Current Biology. 2011; 21:1220–1224. DOI: 10.1016/j.cub.2011.06.009 [PubMed: 21723130]

Borod JC, Pick LH, Hall S, Sliwinski M, Madigan N, Obler LK, et al. Tabert M. Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. Cognition and Emotion. 2000; 14:193–211. DOI: 10.1080/026999300378932

Breitenstein C, Lancker DV, Daum I, Waters CH. Impaired perception of vocal emotions in Parkinson's disease: Influence of speech time processing and executive functioning. Brain and Cognition. 2001; 45:277–314. DOI: 10.1080/026999300420011 [PubMed: 11237372]

Buerkner PC. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software. 2016; 80:1–28. DOI: 10.18637/jss.v080.i01

Carvalho, CM; Polson, NG; Scott, JG. Handling sparsity via the horseshoe. Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS); 2009. 73–80.

Dimberg U, Thunberg M, Elmehed K. Unconscious facial reactions to emotional facial expressions. Psychological Science. 2000; 11:86–89. DOI: 10.1111/1467-9280.00221 [PubMed: 11228851]

Ekman, P. Biological and cultural contributions to body and facial movementThe Anthropology of the Body. Blacking, J, editorNew York: Academic Press; 1977. 39–84.

Gilbert DT, Osborne RE. Thinking backward: Some curable and incurable consequences of cognitive busyness. Journal of Personality and Social Psychology. 1989; 57:940–949. DOI: 10.1037/0022-3514.57.6.940

Globerson E, Amir N, Golan O, Kishon-Rabin L, Lavidor M. Psychoacoustic abilities as predictors of vocal emotion recognition. Attention, Perception and Psychophysics. 2013; 75:1799–1810. DOI: 10.3758/s13414-013-0518-x

Gruber T, Grandjean D. A comparative neurological approach to emotional expressions in primate vocalizations. Neuroscience and Biobehavioral Reviews. 2017; 73:182–190. DOI: 10.1016/j.neubiorev.2016.12.004 [PubMed: 27993605]

Hairston WD, Maldjian JA. An adaptive staircase procedure for the E-Prime programming environment. Computer Methods and Programs in Biomedicine. 2009; 93:104–108. DOI: 10.1016/j.cmpb.2008.08.003 [PubMed: 18838189]

Hawk ST, van Kleef GA, Fischer A, van der Schalk J. "Worth a thousand words": Absolute and relative decoding of nonlinguistic affect vocalizations. Emotion. 2009; 9:293–305. DOI: 10.1037/a0015178 [PubMed: 19485607]

Hoaken PNS, Allaby DB, Earle J. Executive cognitive functioning and the recognition of facial expressions of emotion in incarcerated violent offenders, non-violent offenders, and controls. Aggressive Behavior. 2007; 33:412–421. DOI: 10.1002/ab.20194 [PubMed: 17683105]

Jiang X, Paulmann S, Robin J, Pell MD. More than accuracy: Nonverbal dialects modulate the time course of vocal emotion recognition across cultures. Journal of Experimental Psychology: Human Perception and Performance. 2015; 41:597–612. DOI: 10.1037/xhp0000043 [PubMed: 25775176]

Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin. 2003; 129:770–814. DOI: 10.1037/0033-2909.129.5.770 [PubMed: 12956543]

Karatekin C. Development of attentional allocation in the dual task paradigm. International Journal of Psychophysiology. 2004; 52:7–1. DOI: 10.1016/j.ijpsycho.2003.12.002 [PubMed: 15003369]

Kiss M, Eimer M. ERPs reveal subliminal processing of fearful faces. Psychophysiology. 2008; 45:318–326. DOI: 10.1111/j.1469-8986.2007.00634.x [PubMed: 17995905]

Kruschke, J. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. 2nd ed. London: Academic Press; 2014.

Lavan N, Scott SK, McGettigan. Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. Journal of Nonverbal Behavior. 2016; 40:133–149. DOI: 10.1007/s10919-015-0222-8

Lima CF, Alves T, Scott SK, Castro SL. In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. Emotion. 2013; 14:145–160. DOI: 10.1037/a0034287 [PubMed: 24219391]

Lima CF, Castro SL. Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. Emotion. 2011; 11:1021–1031. DOI: 10.1037/a002452 [PubMed: 21942696]

Lima CF, Castro SL, Scott SF. When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. Behavior Research Methods. 2013; 45:1234–1245. DOI: 10.3758/s13428-013-0324-3 [PubMed: 23444120]

Lima CF, Garrett C, Castro SL. Not all sounds sound the same: Parkinson's disease affects differently emotion processing in music and in speech prosody. Journal of Clinical and Experimental Neuropsychology. 2013; 35:373–392. DOI: 10.1080/13803395.2013.776518 [PubMed: 23477505]

Lima CF, Lavan N, Evans S, Agnew Z, Halpern AR, Shanmugalingam P, et al. Scott SK. Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems. Cerebral Cortex. 2015; 25:4638–4650. DOI: 10.1093/cercor/bhv134 [PubMed: 26092220]

Liu T, Pinheiro AP, Deng G, Nestor PG, McCarley RW, Niznikiewicz MA. Electrophysiological insights into processing nonverbal emotional vocalizations. NeuroReport. 2012; 23:108–112. DOI: 10.1097/WNR.0b013e32834ea757 [PubMed: 22134115]

Martinez L, Falvello VB, Aviezer H, Todorov A. Contributions of facial expressions and body language to the rapid perception of dynamic emotions. Cognition and Emotion. 2016; 30:939–952. DOI: 10.1080/02699931.2015.1035229 [PubMed: 25964985]

McElreath, R. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC Press; 2015.

McGettigan C, Walsh E, Jessop R, Agnew ZK, Sauter DA, Warren JE, Scott SK. Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. Cerebral Cortex. 2015; 25:246–257. DOI: 10.1093/cercor/bht227 [PubMed: 23968840]

Moors A, De Houwer J. Automaticity: A theoretical and conceptual analysis. Psychological Bulletin. 2006; 132:297–326. DOI: 10.1037/0033-2909.132.2.297 [PubMed: 16536645]

Morey RD, Hoekstra R, Rouder JN, Lee MD, Wagenmakers EJ. The fallacy of placing confidence in confidence intervals. Psychonomic Bulletin & Review. 2016; 23:103–123. [PubMed: 26450628]

Öhman A. Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analyses of emotion. Psychophysiology. 1986; 23:123–145. DOI: 10.1111/j.1469-8986.1986.tb00608.x [PubMed: 3704069]

Öhman A, Flykt A, Esteves F. Emotion drives attention: Detecting the snake in the grass. Journal of Experimental Psychology: General. 2001; 130:466–478. DOI: 10.1037/0096-3445.130.3.466 [PubMed: 11561921]

Otten M, Mann L, van Berkum JJ, Jonas KJ. No laughing matter: How the presence of laughing witnesses changes the perception of insults. Social Neuroscience. 2017; 12:182–193. DOI: 10.1080/17470919.2016.1162194 [PubMed: 26985787]

Pell MD, Rothermich K, Liu P, Paulmann S, Sethi S, Rigoulot S. Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. Biological Psychology. 2015; 111:14–25. DOI: 10.1016/j.biopsycho.2015.08.008 [PubMed: 26307467]

Pinheiro AP, Barros C, Pedrosa J. Salience in a social landscape: Electrophysiological effects of task-irrelevant and infrequent vocal change. Social Cognitive and Affective Neuroscience. 2015; 11:127–139. DOI: 10.1093/scan/nsv103 [PubMed: 26468268]

Ransdell S, Arecco MR, Levy CM. Bilingual long-term working memory: The effects of working memory loads on writing quality and fluency. Applied Psycholinguistics. 2001; 22:113–128.

Rigoulot S, Wassiliwizky E, Pell MD. Feeling backwards? How temporal order in speech affects the time course of vocal emotion recognition. Frontiers in Psychology. 2013; 4:367.doi: 10.3389/fpsyg.2013.00367 [PubMed: 23805115]

Tracy JL, Robins RW. The automaticity of emotion recognition. Emotion. 2008; 8:81–95. DOI: 10.1037/1528-3542.8.1.81 [PubMed: 18266518]

Sander D, Grandjean D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. NeuroImage. 2005; 28:848–858. DOI: 10.1016/j.neuroimage.2005.06.023 [PubMed: 16055351]

Sauter D, Eimer M. Rapid detection of emotion from Human vocalizations. Journal of Cognitive Neuroscience. 2010; 22:474–481. DOI: 10.1162/jocn.2009.21215 [PubMed: 19302002]

Sauter DA, Eisner F, Calder AJ, Scott SK. Perceptual cues in nonverbal vocal expressions of emotion. The Quarterly Journal of Experimental Psychology. 2010; 63:2251–2272. DOI: 10.1080/17470211003721642 [PubMed: 20437296]

Sauter DA, Eisner F, Ekman P, Scott SK. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proceedings of the National Academy of Sciences. 2010; 107:2408–2412. DOI: 10.1073/pnas.0908239106

Sauter D, Panattoni C, Happé F. Children's recognition of emotions from vocal cues. British Journal of Developmental Psychology. 2013; 31:97–113. DOI: 10.1111/j.2044-835X.2012.02081.x [PubMed: 23331109]

Sauter DA, Scott SK. More than one kind of happiness: Can we recognize vocal expressions of different positive states? Motivation and Emotion. 2007; 31:192–199. DOI: 10.1007/s11031-007-9065-x

Scherer KR. Expression of emotion in voice and music. Journal of Voice. 1995; 9:235–248. DOI: 10.1016/S0892-1997(05)80231-0 [PubMed: 8541967]

Schröder M. Experimental study of affect bursts. Speech Communication. 2003; 40:99–116. DOI: 10.1016/S0167-6393(02)00078-X

Scott SK, Lavan N, Chen S, McGettigan C. The social life of laughter. Trends in Cognitive Sciences. 2014; 18:618–620. DOI: 10.1016/j.tics.2014.09.002 [PubMed: 25439499]

Scott, SK, Sauter, D, McGettigan, C. Brain mechanisms for processing perceived emotional vocalizations in humansHandbook of Behavioral Neuroscience. Stefan, MB, editorLondon: Academic Press; 2010. 187–197.

Scott SK, Young AW, Calder AJ, Hellawell DJ, Aggleton JP, Johnsons M. Impaired auditory recognition of fear and anger following bilateral amygdala lesions. Nature. 1997; 385:254–257. DOI: 10.1038/385254a0 [PubMed: 9000073]

Simon-Thomas ER, Keltner DJ, Sauter D, Sinicropi-Yao L, Abramson A. The voice conveys specific emotions: Evidence from vocal burst displays. Emotion. 2009; 9:838–846. DOI: 10.1037/a0017810 [PubMed: 20001126]

Soderstrom M, Reimchen M, Sauter D, Morgan JL. Do infants discriminate non-linguistic vocal expressions of positive emotions? Cognition and Emotion. 2017; 31:298–311. DOI: 10.1080/02699931.2015.1108904 [PubMed: 27900919]

Tzelgov, J. Automatic but conscious: That is how we act most of the timeAdvances in Social Cognition. Wyer, JRS, editorVol. X. New York: Psychology Press; 1997. 217–230.

Warren JE, Sauter DA, Eisner F, Wiland J, Dresner MA, Wiser RJS, Rosen S, Scott SK. Positive emotions preferentially engage and auditory-motor "mirror" system. The Journal of Neuroscience. 2006; 26:13067–13075. DOI: 10.1523/JNEUROSCI.3907-06.2006 [PubMed: 17167096]

Winkielman P, Berridge KC, Wilbarger JL. Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. Personality and Social Psychology Bulletin. 2005; 31:121–135. DOI: 10.1177/0146167204271309 [PubMed: 15574667]

Wood, S. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC Press; 2006.

Europe PMC Funders Author Manuscripts

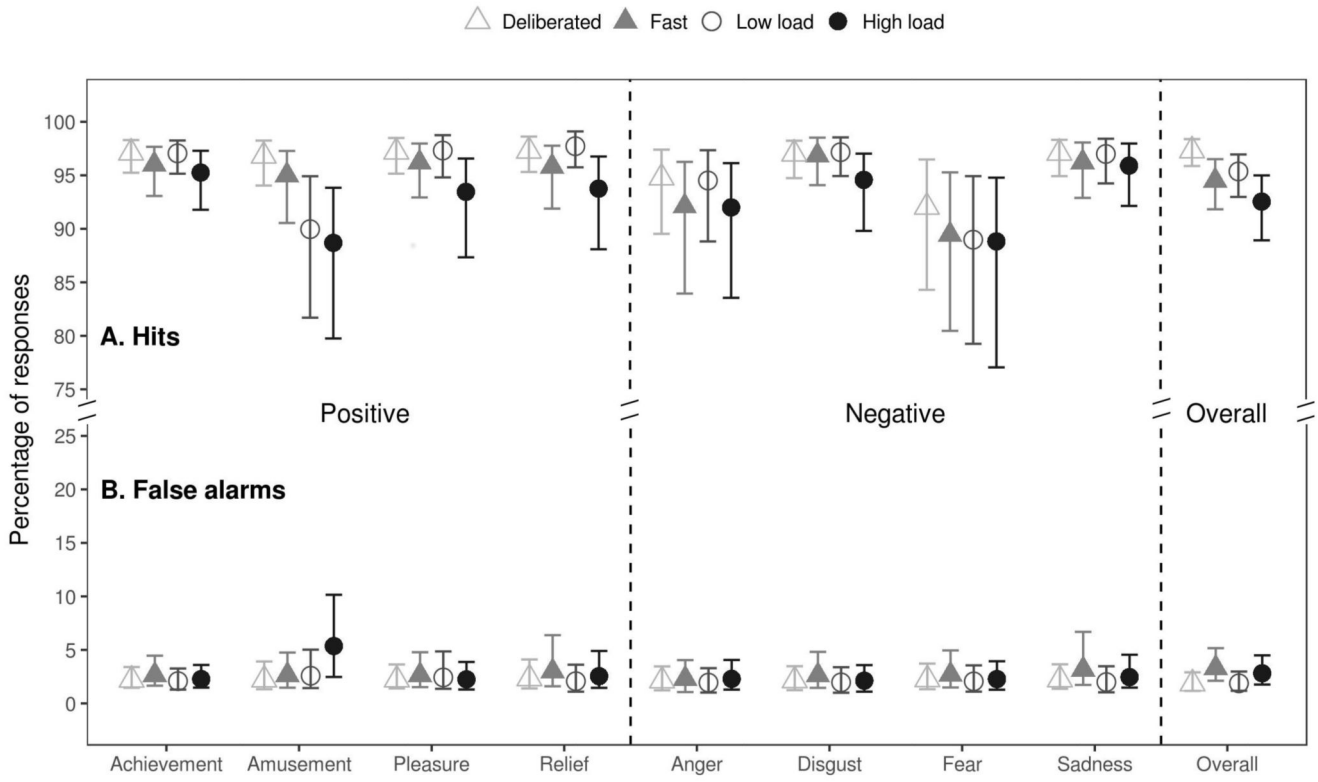

Figure 1.

Hit rates (a) and false alarms (b) for each condition and emotion (Overall corresponds to all emotions combined). The median of posterior distribution and 95% CI are presented. The analyses included 8960 trials for hits and 12544 trials for false alarms.

Europe PMC Funders Author Manuscripts

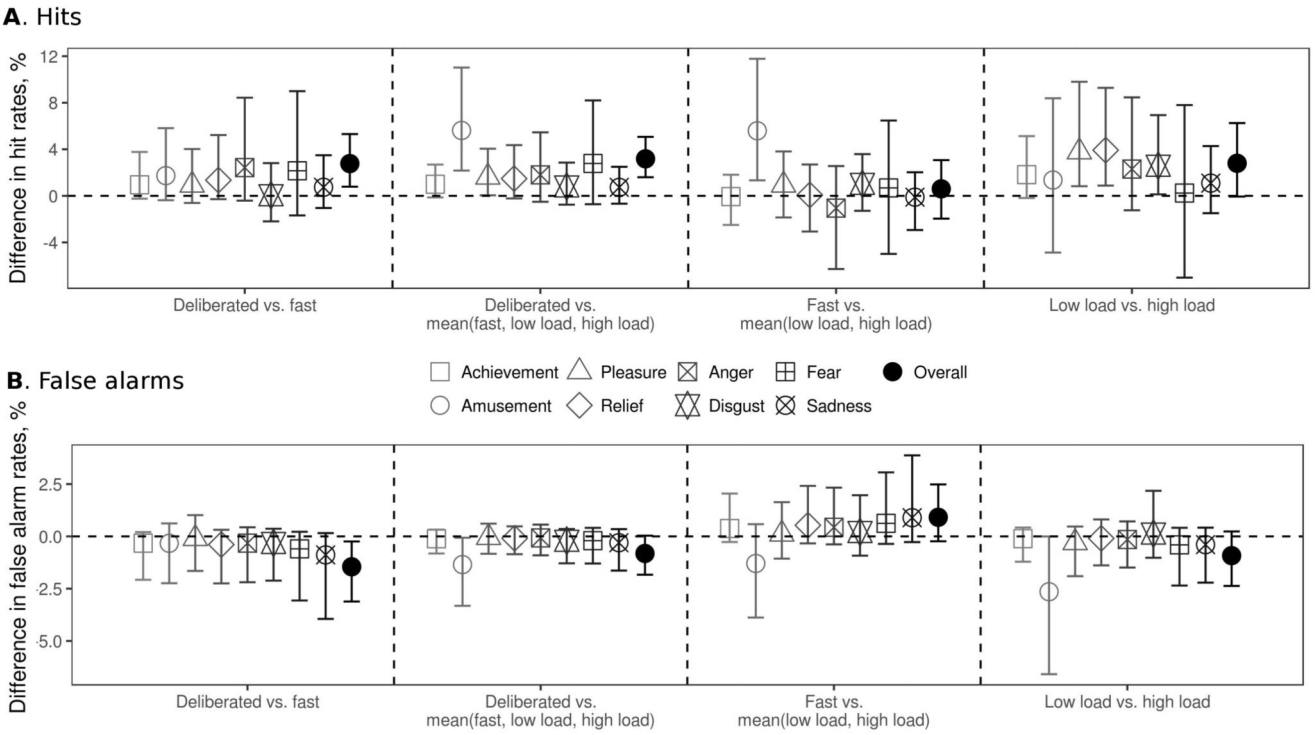

Figure 2.

Magnitude of the difference between conditions in hit rates (a) and false alarms (b), separately for each emotion and for all emotions combined (Overall). The most plausible estimate of the difference between conditions and 95% CI are presented. Evidence for a difference between conditions can be directly inferred from the figure, corresponding to when the 95% CI excludes 0.

Europe PMC Funders Author Manuscripts

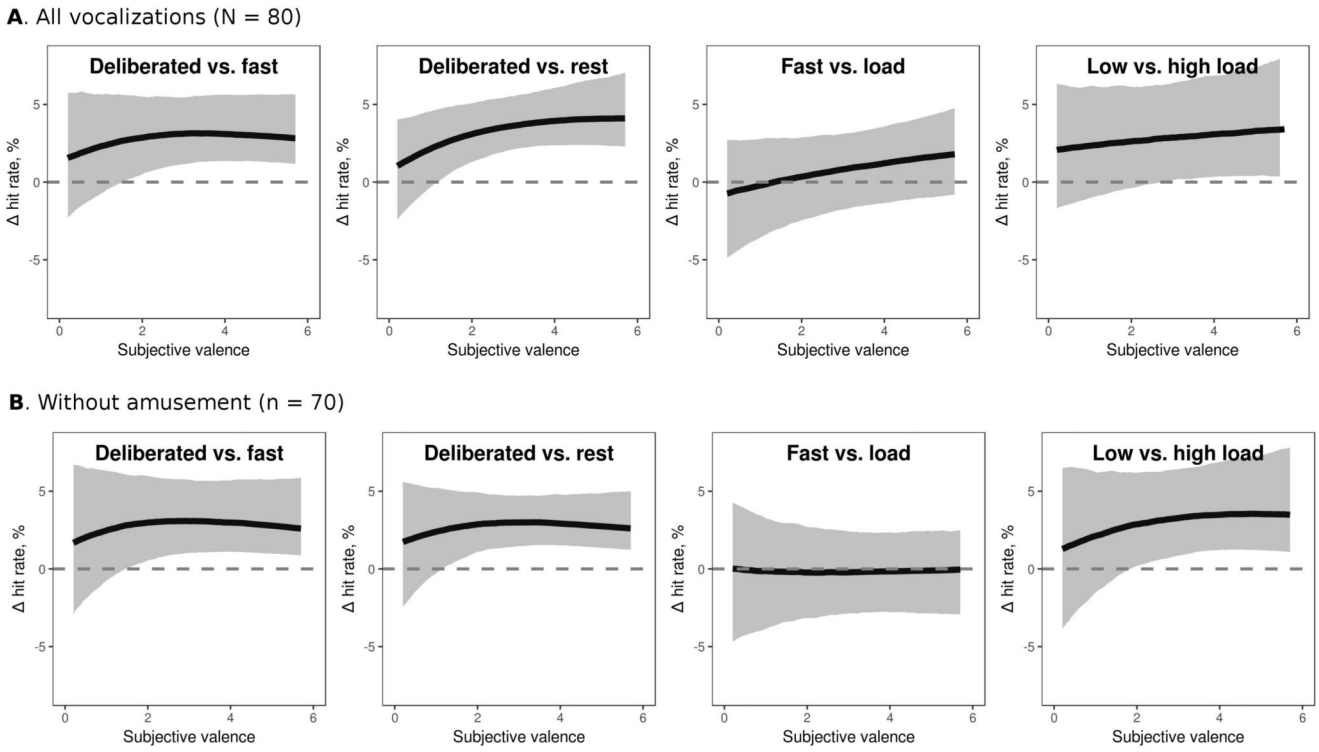

Figure 3.

Relationship between the valence of vocalizations and the magnitude of the difference in hit rates between conditions. The gray shaded area shows the 95% CI.

Europe PMC Funders Author Manuscripts

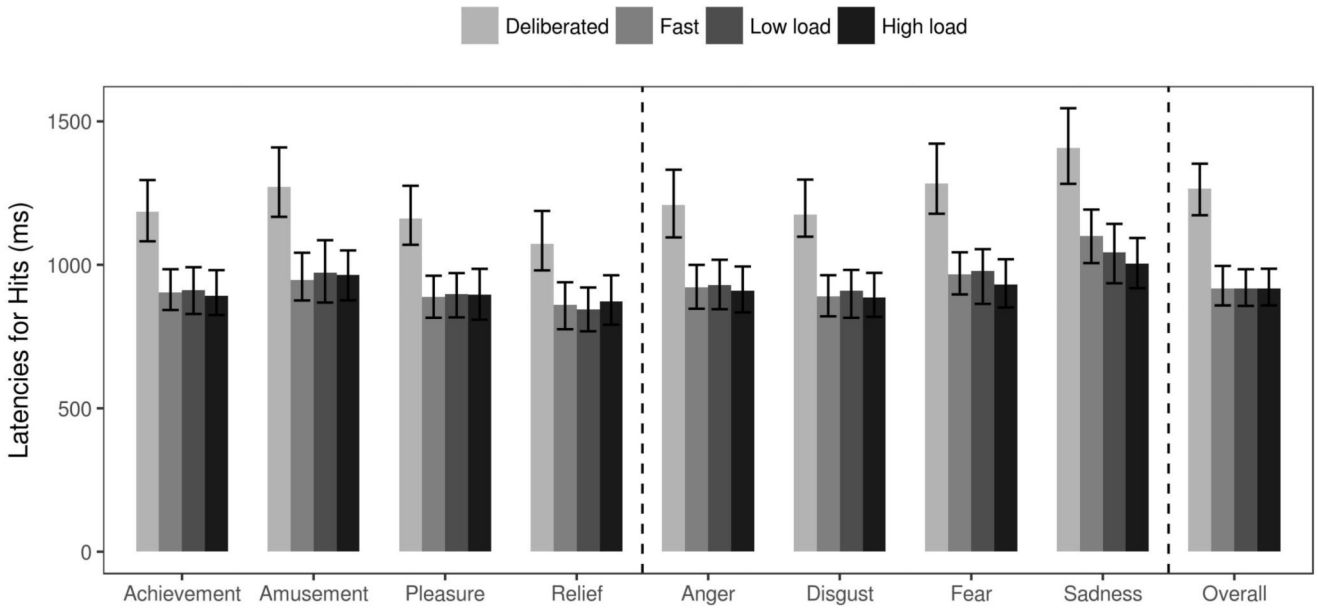

Figure 4.

Response latencies for correctly recognized target emotional expressions, separately for each condition and emotion. The median of posterior distribution and 95% CI are presented. The analysis included 8022 trials.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

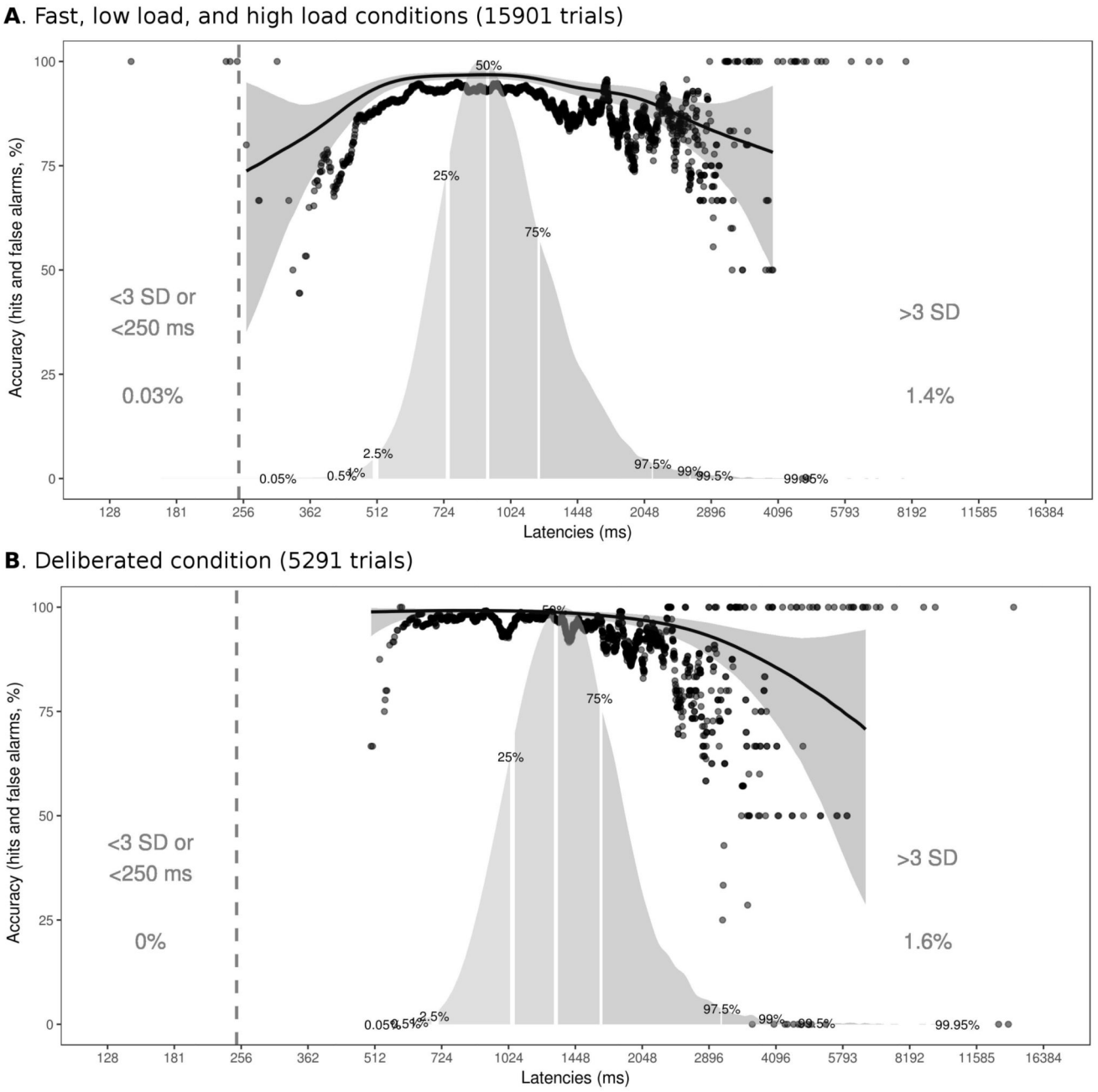

Figure 5.

Accuracy of emotion recognition as a function of response latencies (for both experimental and filler expressions), with overlaid density plots showing the distribution of latencies. The black dots correspond to observed accuracy, including hits and correct rejections, averaged over bins of ±25 ms. The vertical dashed lines and text labels indicate the cut-offs for outlier exclusion and the percentage of excluded trials. The solid smooth regression line shows the predicted accuracy (median of posterior distribution and 95% CI within gray shaded area), for all trials excluding outliers.

Europe PMC Funders Author Manuscripts

Table 1

Characteristics of the experimental nonverbal emotional vocalizations (*n* = 10 per emotion, total 80). Standard deviations are given in parentheses.

Stimulus Type	Duration (ms)	Accuracy (%)	Intensity (0-6)	Valence (0-6)	Arousal (0-6)
Positive					
Achievement	1018 (237)	80.0 (11.8)	4.5 (0.3)	5.2 (0.4)	5.3 (0.3)
Amusement	1000 (244)	91.0 (6.1)	4.5 (0.8)	4.9 (0.4)	4.6 (0.7)
Pleasure	1114 (177)	86.5 (14)	4.9 (0.4)	4.4 (0.4)	2.5 (0.5)
Relief	916 (226)	89 (6.6)	4.7 (0.5)	3.4 (0.3)	2.0 (0.4)
Average	1012 (225)	86.6 (10.6)	4.6 (0.6)	4.5 (0.8)	3.6 (1.5)
Negative					
Anger	1048 (170)	85 (12.9)	4.8 (0.5)	1.0 (0.3)	4.4 (0.5)
Disgust	920 (427)	91 (9.4)	4.8 (0.5)	1.1 (0.2)	3.2 (0.5)
Fear	914 (295)	79 (10.7)	4.5 (0.6)	1.6 (0.4)	4.1 (1.1)
Sadness	1083 (278)	87.5 (16.5)	4.4 (0.6)	1.0 (0.5)	2.5 (0.6)
Average	991 (304)	85.6 (13)	4.6 (0.6)	1.2 (0.4)	3.6 (1.1)

Note: Perceptual data are based on a pilot study with $N = 40$; accuracy data were obtained using a forced-choice emotion recognition task ($n = 20$); intensity, valence and arousal data were obtained using 7-point rating scales (0-6), with higher values indicating higher perceived emotion intensity, positive valence, and higher arousal ($n = 20$).

Europe PMC Funders Author Manuscripts

Table 2

Beta coefficients for valence of vocalizations as a predictor of hit rates, separately for each condition. Values represent the median of posterior distribution and 95% CI on the logit scale (when it includes 0 = no effect).

Condition	Full set of vocalizations (n = 80)	Vocalizations with lowest accuracy (n = 40)	All vocalizations except amusement (n = 70)
Deliberated	0.38 [0.20, 0.58]	0.38 [0.15, 0.61]	0.42 [0.19, 0.65]
Fast	0.14 [-0.02, 0.3]	0.16 [-0.04, 0.35]	0.17 [-0.01, 0.35]
Low Load	0.08 [-0.08, 0.25]	0.16 [-0.04, 0.36]	0.31 [0.10, 0.51]
High Load	0.02 [-0.13, 0.19]	0.05 [-0.14, 0.24]	0.11 [-0.07, 0.29]