UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Herrera-Bennett, A., Puah, S., Hasenbein, L., & Wildgruber, D. (2022). Emotion-modulated Recall: Congruency Effects of Nonverbal Facial and Vocal Cues on Semantic Recall. *Collabra: Psychology*, *8*(1). https://doi.org/10.1525/collabra.31601

Cognitive Psychology

Emotion-modulated Recall: Congruency Effects of Nonverbal Facial and Vocal Cues on Semantic Recall

Arianne Herrera-Bennett[1] [a], Shermain Puah[2], Lisa Hasenbein[3], Dirk Wildgruber[4]

[1] Psychology, University of California, Davis, USA; Psychology, Ludwig Maximilian University of Munich, Germany, [2] Psychology, Ludwig Maximilian University of Munich, Germany, [3] Psychology, Ludwig Maximilian University of Munich, Germany; Education Sciences and Psychology, University of Tübingen, Germany, [4] Psychiatry and Psychotherapy, University of Tübingen, Germany

Keywords: bimodal integration, semantic recall, emotional prosody, facial expression, congruency effects

https://doi.org/10.1525/collabra.31601

Collabra: Psychology

Vol. 8, Issue 1, 2022

The current study had two main goals: First, to replicate the 'bimodal integration' effect (i.e. the automatic integration of crossmodal stimuli, namely facial emotions and emotional prosody); and second, to investigate whether this phenomenon facilitates or impairs the intake and retention of unattended verbal content. The study borrowed from previous bimodal integration designs and included a two-alternative forced-choice (2AFC) task, where subjects were instructed to identify the emotion of a face (as either 'angry' or 'happy') while ignoring a concurrently presented sentence (spoken in an angry, happy, or neutral prosody), after which a surprise recall was administered to investigate effects on semantic content retention. While bimodal integration effects were replicated (i.e. faster and more accurate emotion identification under congruent conditions), congruency effects were not found for semantic recall. Overall, semantic recall was better for trials with emotional (vs. neutral) faces, and worse in trials with happy (vs. angry or neutral) prosody. Taken together, our findings suggest that when individuals focus their attention on evaluation of facial expressions, they implicitly integrate nonverbal emotional vocal cues (i.e. hedonic valence or emotional tone of accompanying sentences), and devote less attention to their semantic content. While the impairing effect of happy prosody on recall may indicate an emotional interference effect, more research is required to uncover potential prosody-specific effects. All supplemental online materials can be found on OSF (https://osf.io/am9p2/).

Introduction

To date, the *bimodal integration* literature has systematically demonstrated, across a range of discrete emotions (i.e. happy, fear, angry, sad), enhanced perception of emotional information when cues are conveyed simultaneously via different channels (e.g., face and voice). Specifically, emotions of faces were identified more quickly and accurately when paired with words or sentences spoken in the congruent emotional prosody (de Gelder & Vroomen, 2000; Dolan et al., 2001; Massaro & Egan, 1996). Notably, because such effects were found even when subjects were tasked to ignore the accompanying voice, findings have led authors to conclude that individuals naturally and automatically integrate unattended emotional information, such as emotional prosody (Ethofer et al., 2006). Interestingly, while a body of evidence exists on the facilitating effects of crossmodal congruency on *perception* of emotional stimuli (e.g., Paulmann & Pell, 2011; for review, see Klasen et al., 2012), research has seemingly neglected to investigate effects on subsequent *recall*.

Congruency effects, specifically in the context of multisensory integration, have been strongly conceptualized in connection with semantic meaning, whereby here the term 'semantic' is used to convey a "correspondence of incoming signals" (for review, see Doehrmann & Naumer, 2008, p. 137). For instance, matching emotional connotations conveyed at the level of verbal content and affective prosody (e.g., happy words spoken in a happy prosody) have been found to lead to faster lexical processing, even when subjects were not explicitly instructed to attend to the emotional content or tone, or when prosody and content varied from trial to trial (Nygaard & Queen, 2008). Authors proposed that emotional prosody may serve as a top-down constraint on linguistic processing, whereby "the integration of emotional tone of voice with linguistic content occurs relatively early in the processing of spoken language" (Nygaard & Queen, 2008, p. 1025). Such findings emphasize the natural interaction of verbal and nonverbal aspects of speech during access and recognition of spoken language (Nygaard & Queen, 2008).

The facilitatory effects of multisensory presentations

a aherrerabennett@ucdavis.edu

(e.g., auditory-visual pairs) for *recall* also appear to be critically conditional on whether both sources of information - taken together - are semantically meaningful, such as enhanced memory performance for a visual object (e.g., image of a bell) when presented alongside a meaningful sound (e.g., sound of bell) versus a sound that is only arbitrarily related (e.g., pure tone; Murray et al., 2004). In this way, the extent to which multisensory stimuli meaningfully correspond has been argued to play a critical role in modulating how information is processed, and how memory and behavior is subsequently affected (see e.g., Doehrmann & Naumer, 2008; Laurienti et al., 2004; Lehmann & Murray, 2005; Spence, 2007).

A logical question that follows, but to the best of our knowledge has yet to be empirically addressed, is whether there are differential recall effects for audiovisual (face/voice) information that is *emotionally* congruent versus incongruent. Specifically, given the evidence that individuals naturally integrate the qualitative nature of unattended emotional information (i.e. vocal prosody), when evaluating emotional faces (de Gelder & Vroomen, 2000; Dolan et al., 2001; Ethofer et al., 2006; Massaro & Egan, 1996), and the fact that emotional prosody and verbal content appear to be processed in parallel (e.g., Nygaard & Queen, 2008), there are grounds to expect that congruent emotional conditions may potentially facilitate the integration and recall of spoken verbal *content*. Put differently, is the automatic integration of unattended emotional information limited to nonverbal cues, i.e. emotional prosody, or does it additionally have effects on integration and retention of verbal content?

To this end, the current study had two main goals: First, to run a high-powered replication of the bimodal integration effect, drawing upon key design elements from past crossmodal studies (i.e. Davis et al., 2011; de Gelder & Vroomen, 2000; Dolan et al., 2001; Ethofer et al., 2006; Massaro & Egan, 1996). And second, to investigate the role of nonverbal emotional cues, i.e. emotional facial expressions (EF) and emotional prosody (EP), on semantic recall. Specifically, we asked whether a match in facial and vocal emotion facilitates recall for semantic content of spoken sentences. In this way, the current study consists of both a confirmatory replication effort (i.e. bimodal integration effects), as well as an exploratory extension (i.e. recall effects) of past work.

Our study borrowed from previous bimodal integration designs which made use of a two-alternative forced-choice (2AFC) task (de Gelder & Vroomen, 2000; Dolan et al., 2001; Massaro & Egan, 1996). Specifically, subjects were instructed to classify the emotion of a face (as 'angry' or 'happy') while ignoring a concurrently presented sentence (spoken in an angry, happy, or neutral prosody), after which they were administered a surprise (new/old) recognition task. In doing so, we explored whether the presence of matching, nonverbal, facial and vocal emotional cues led, not only, to *facilitated perception* of emotional expressions (confirmatory replication), but also to *facilitated recall*[1] of concurrently presented vocal content (exploratory extension). Our work hopes to shed light on an open question in the literature regarding the perception of face-voice audiovisual content, namely: To what extent does the automatic integration of peripheral information implicate qualitative aspects (i.e. emotional prosody) versus semantic elements (i.e. sentence meaning) of peripheral content?

Effects of Emotion on Recall of Accompanying Neutral Stimuli

The literature concerning the effects of emotion on recall are, unsurprisingly, mixed. Ample evidence exists to support both the *enhancing* as well as *impairing* effects of emotional stimuli on recall of accompanying neutral information. Some authors have emphasized the need to consider context-related influences (e.g., task priorities, or the relationship between neutral and emotional stimuli) in order to explain contradictory findings (for review see Riegel et al., 2016; West et al., 2017).

Several competing theories have been put forth to account for the inconsistencies in empirical findings. Based on the priority-binding theory, emotion has an *enhancing* effect because neutral information paired alongside emotional information is given processing priority, and hence, is recalled better (Mackay et al., 2004). Guillet and Arndt (2009) demonstrated that neutral words that preceded or followed emotionally arousing taboo words were recalled better than when presented in less arousing contexts, i.e. neutral or negatively-valenced contexts. The authors argued that heightened arousal from the emotional taboo words triggered binding mechanisms and enhanced the association in memory between the pair of words.

In contrast, the attention-narrowing hypothesis (Easterbrook, 1959) posits an *impairing* effect of emotional stimuli on recall of co-occurring neutral information. It suggests that emotional information heightens arousal and draws attention toward the arousing emotional stimulus, leaving reduced attentional capacity for the neutral information in the periphery. As a result, memory for the neutral information is impaired when presented alongside emotional information. Davis et al. (2011) found that when neutral words were interleaved with angry faces, recall for words was *impaired*. This finding supports the idea of attentional narrowing, specifically toward emotional information that communicates threat. It should be noted, however, that not all emotional stimuli were found to produce impairing effects; rather, differential effects of discrete facial emotions were observed. Specifically, when neutral words were interleaved with fearful faces, recall for words was *enhanced*, on account of fear faces triggering attention toward the environment

1 Our study paradigm uses a recognition task to investigate effects on memory. We understand that within the memory literature, recall and recognition are recognized as different processes. Due to the exploratory nature of the current study, we use the term "recognition" more specifically when referring to our new/old task, but use the term "recall" more broadly when synthesizing the past literature on memory effects, and when describing the results of the manuscript.

to extract salient threat information (Davis et al., 2011). Moreover, recall for angry faces exceeded that of fearful faces, signalling a trade-off effect between memory for central (i.e. faces) and peripheral (i.e. neutral words) information. Such findings speak to the complexity of emotions and emotional expressivity, as it concerns their effects on recall.

Finally, Mather and Sutherland (2011) offered the arousal-biased competition theory to explain divergent findings, which conceptualizes emotion effects (i.e. *enhancing* vs. *impairing* effects) as a function of the perceived priority of stimuli: "Arousal amplifies the effects of competition, improving perception of high priority information and weakening perception of low priority information" (Mather & Sutherland, 2011, p. 3). According to this theory, prioritization of information can arise from bottom-up influences (e.g., perceptual salience), top-down influences (e.g., expectations, goal relevance), or an interaction of both. For example, Sutherland and Mather (2012) found that under negatively-arousing conditions, memory for letters of high perceptual salience was *enhanced*, while memory for low-salience letters was *impaired*. Similarly, Sakaki et al. (2014) found that top-down goal relevance could determine the effects of emotionality on recall. For instance, using a classic oddball paradigm, authors found that when a neutral object was immediately followed by a negative emotional target, recall of the neutral object was *impaired*. The pattern of effects however was reversed when participants were tasked to prioritize the neutral information leading up to the emotional target; in these instances, the presence of negative emotional stimuli *enhanced* recall of preceding neutral stimuli.

The literature outlined above underscores a key idea regarding the impact of emotionality on recall: The extent to which the presence of emotional stimuli produces *beneficial* versus *hindering* effects on recall of accompanying neutral stimuli, is a function of contextual influences. Notably, the context of the task itself can play a salient role, insofar as determining the relationship between the emotional and neutral stimuli. For instance, in contexts where neutral and emotional stimuli compete for attentional resources (e.g., Flanker design), one might anticipate more evidence in support of attentional narrowing or prioritization toward emotional stimuli, at the expense of neutral stimuli (i.e. central-peripheral trade-off effect). By contrast, in paradigms where both neutral and emotional stimuli are task-relevant, one might expect more evidence of facilitating effects of emotional stimuli on recall of accompanying neutral information. Here, the terms 'central' and 'peripheral' do not refer to the spatial location of stimuli, but rather denote the relevance or attentional salience of stimuli (e.g., cues), made either explicit from task instructions or implied from the nature of the task (for further discussion, see Kensinger et al., 2007).

With respect to our specific task paradigm, the presented facial expressions are considered central to the 2AFC task, whereas the accompanying neutral sentences (i.e. spoken semantic information) are considered peripheral or task-irrelevant, on account of the fact that participants are explicitly tasked to ignore them. That said, we know from the crossmodal integration literature that participants naturally incorporate peripheral cues (i.e. vocal stimuli) into their decisions about central stimuli, with congruent conditions *facilitating* task performance, and incongruent conditions *impairing* task performance. As such, the delineation between central and peripheral information should be more ambiguous in crossmodal tasks, such as ours. Moreover, our paradigm is further complexified by the fact that two sources of emotional expressivity (i.e. facial expressions and vocal prosody) are present. For these reasons, more research is warranted to explore the expected effects of emotional stimuli on recall, specifically in the context of crossmodal designs.

Finally, it is worth noting that the existing literature on the effects of emotional expressivity on recall, specifically in the context of audiovisual (i.e. facial/vocal) presentations, is also mixed. In particular, in a review by Kauschke et al. (2019), authors assessed the valence effects (i.e. positivity or negativity biases) of studies that used facial and vocal stimuli, and reported no clear pattern of effects or systematic biases across modalities. They too emphasized the role of contextual variables, such as methodological differences, to potentially explain the heterogeneity in observed valence effects. Importantly, Kauschke and colleagues (2019) also noted that the majority of the reviewed studies adopted unimodal designs, and thus could speak only to modality-specific valence differences. As such, authors recommended that future research should also aim to explore the mutual influences or interplay between facial and vocal modalities in perception and processing of stimuli.

To this end, the current study seeks not only to replicate the effects of crossmodal integration on perception of emotional stimuli, but also to contribute to and broaden the understanding of emotional and contextual congruency effects on the automatic processing and subsequent recall of accompanying neutral stimuli (specifically, semantic information).

Research Questions

The current research extends our understanding of how individuals tend to integrate concurrently presented information from different modalities, i.e. facial and vocal channels. Specifically, we investigate whether automatic integration of emotional prosody (EP) is limited to effects on identification of emotional facial expressions (EF), or if it also facilitates the integration and retention of semantic content. Moreover, we explore whether there is an interaction between central emotional stimuli (i.e. EF) and emotional congruency of accompanying peripheral stimuli (i.e. EP) on semantic recall. In sum, our study includes a replication of the bimodal integration effect on emotion perception (i.e. accuracy and RT), and an exploratory investigation of potential effects on subsequent recall.

Key experimental design decisions in our study were borrowed directly from relevant past paradigms (incl. Davis et al., 2011; de Gelder & Vroomen, 2000; Dolan et al., 2001; Ethofer et al., 2006; Massaro & Egan, 1996), and are outlined in detail below. Although our paradigm includes methodological differences from previous studies, we believe that these deviations are minor with respect to the theoretical understanding of critical elements required to replicate the confirmatory effects in question (Zwaan et al.,

2018).

Hypotheses

Before collecting the final replication sample, a small pilot study (n = 37, after exclusions) was conducted, which allowed us to obtain initial approximations of effect size estimates for our exploratory research questions. Hypotheses were thus informed by relevant past research as well as our pilot study results. For confirmatory effects (i.e. accuracy and RT), our pilot data mostly corroborated past findings: All main effects were systematically found to be statistically significant. Interaction effects, however, displayed some inconsistencies with previous results: While the interaction between EF and EP on accuracy was statistically significant, the effect of happy prosody on judgment of neutral faces was counterintuitive. Additionally, the interaction effect on RT was non-significant. Pre-registered predictions for the replication data (i.e. all confirmatory and exploratory hypotheses) are now elaborated.

Confirmatory Hypotheses for Facial Emotion Identification. We expected to replicate the bimodal integration effects on accuracy and RT, namely faster and more accurate facial emotion identification for congruent emotional trials. Specifically, for accuracy, main effects of EF and EP, as well as interaction of EF x EP, are expected to be significant. For RT, main effects of EF and EP are expected to be significant; given, however, the essentially non-existent interaction effect (η_p^2 = .03) in our pilot study, we expected the interaction on RT to once again be non-significant.

Exploratory Hypotheses for Recall Effects. In order to explore the potential enhancing (vs. impairing) effects of facilitated perceptual integration on subsequent semantic recall, we used pilot data to compare sentence recognition rates for trials with congruent EF and EP (e.g., happy face paired with happy prosody) versus trials with incongruent EF and EP (e.g., happy face paired with angry or neutral prosody). 2-way repeated-measures ANOVA revealed non-significant main effects of EF and Congruency (of EP), $ps \geq .463$, but a nearly significant (i.e. nearly below the corrected alpha-level of .0167) interaction between EF and Congruency (p = .019): Notably, when happy faces were paired with incongruent prosodies, mean semantic recall rates were greater compared to congruent trials (approx. 78% vs. 58% accuracy, resp.). The opposite pattern, however, was observed for angry and neutral trials, though these pairwise differences were much less pronounced. Based on these preliminary pilot results, we expected non-significant main effects, but a significant interaction effect of EF x Congruency, on semantic recall. While we also measured face recall, it was not central to our research aim, and therefore we did not specify predictions.

Methods

Sample Inclusion Criteria

Subjects were between 18-40 years of age, with German native-level fluency, normal (or corrected-to-normal) vision, and no history of hearing impairment or mental illness. Participants were excluded if they displayed extreme pretest scores for negative or positive affect (i.e. outliers, +/- 3 standard deviations from sample means; PANAS-SF; Watson et al., 1988). Trials in which no response was given within the allotted time limit (i.e. missing data) were dropped from the analysis.

Sample Size Determination

Sample size was determined using a simulation-based approach (ANOVA Power shinyapp; Lakens & Caldwell, 2019) and was based on expected effect sizes obtained from the pilot data (n = 37, after exclusions). Given that our pilot sample was twice as large as previous study samples (i.e. N = 15, Massaro & Egan, 1996; N = 16, de Gelder & Vroomen, 2000), and pilot effect sizes were generally smaller than those reported in the past literature, we ran a power analysis based on our more conservative and precise estimates. Effects of interest included the confirmatory effects on accuracy and RT, and the exploratory interaction effect on sentence recall. Sample size was adjusted to ensure that even the smallest effect of interest (main effect of EP on RT, η_p^2 = .12) would be accounted for. Results of the simulations determined a planned N of 100 for the replication study: Specifically, running a 2-way repeated-measures ANOVA (specifying a correlation between within-subject factors of 0.5, and Bonferroni adjustment for multiple comparisons), power is 90% or above for all effects of interest listed above (for power analysis details, see https://osf.io/gb72c/).

Data Collection Restrictions

Due to unforeseen circumstances (COVID-19), data collection was stopped prematurely (replication study n = 49, after exclusions). Based on power simulations (see above), computed power levels for an N of 50 yields 90% power or above for all effects of interest, with the exception of the main effect of EP on RT (powered at approx. 76%). As such, even despite data collection restrictions, we argue that the sample size obtained should afford sufficient statistical power to draw meaningful inferences. For these reasons, we report the results of only the replication data in the current manuscript. For additional results, supplementary materials include a direct comparison between pilot data and replication data effects, an assessment of all effects on a set of replicability metrics, as well as the results of the pooled data (see online supplementary information https://osf.io/y3amf/).

Materials

Visual stimuli for the 'bimodal integration task' consisted of 18 faces from the KDEF database (Karolinska Directed Emotional Faces; Lundqvist et al., 1998), and selected specifically among the subset of faces identified as most representative of each of the emotions (see Goeleven et al., 2008, for KDEF validation study). All faces were pretested on a separate sample of 12 participants; only those stimuli that were correctly identified across individuals at least 70% of the time were used (for more details on the pretest study, see https://osf.io/n56v9/). This pretest protocol is borrowed directly from the approach applied in Ethofer et al. (2006). The final set of faces was split evenly

between emotion (i.e. 6 happy, 6 angry, and 6 neutral faces) and gender (50% male, 50% female).

Vocal stimuli consisted of spoken sentences rather than use of single words presented in isolation. This decision was in keeping with recommendations from the literature that emphasizes presenting verbal material within a broader context (e.g., sentences or stories) for ecological validity (for reviews, see Citron, 2012, and Riegel et al., 2016). Specifically, vocal stimuli were borrowed from the database used in Ethofer et al.'s (2006) study: 18 German spoken sentences, each 14 syllables long (duration of 2.2 – 2.7 sec) and emotionally neutral in content. All sentences were similarly pretested with minimum 70% identification accuracy rates, and split evenly between emotional prosody and gender. For the recognition phase of the experiment, 18 additional faces and 18 additional sentences (from the same databases described) were introduced as foil stimuli.

It should be noted that past bimodal integration designs have observed effects across a range of visual stimuli sets and presentations: This includes still images, i.e. the Ekman and Friesen (1976) series, presented in their original format (de Gelder & Vroomen, 2000) or in a morphed continuum format (Dolan et al., 2001; Ethofer et al., 2006), as well as animated faces, i.e. Face39 program (Massaro & Egan, 1996). Similarly, systematic effects have been observed across different vocal stimuli formats, such as repetition of the same word (Massaro & Egan, 1996), repetition of the same sentence (de Gelder & Vroomen, 2000), or use of unique sentences (Ethofer et al., 2006). Moreover, these effects have been replicated across different pairs of emotions (e.g., happy vs. sad; happy vs. angry; happy vs. fear), and across different languages and prosodies. For these reasons, we believe that the specific choice of stimuli, in the current study, should not introduce significant variability with respect to confirmatory effects (i.e. accuracy and RT).

Finally, our design also involves a pre-/post-test measure of state affect (PANAS-SF; Watson et al., 1988), explicit valence and arousal ratings for face stimuli (more details below), as well as a distractor task (Edinburgh Handedness Inventory; Oldfield, 1971), and demographics.

Experimental Procedure

The current study consisted of a lab-based within-subjects block design, using tasks programmed in E-Prime 2.0, version 2.0.10.242 (Psychology Software Tools, Pittsburgh, PA). All tasks were performed on a computer unless noted otherwise. Total study duration was estimated around 20-30 minutes. Experimental procedure began with administration of the consent form (paper-pencil), after which the first affect pretest measure (PANAS-SF) was administered (paper-pencil). Next, the test phase began with the 'bimodal integration task' (two short practice blocks and two test blocks), followed by a post-test affect measure, demographics, and distractor task (all paper-pencil). Participants then underwent a surprise 'recognition task' (2 test blocks), followed by explicit valence and arousal ratings (see details for all tasks below). The experimental procedure ended with administration of the debriefing form and awarding participation compensation.

Bimodal Integration Task. The bimodal integration task (see Figure 1) is a 2AFC task: In each trial, a visual stimulus (facial expression) was paired simultaneously with an auditory stimulus (spoken sentence); face and voice matched for gender. Onset of the visual stimulus started 1,650ms after onset of the audio; because visual presentation was set at 350ms (in keeping with past designs; de Gelder & Vroomen, 2000; Dolan et al., 2001), and audio files were ~2s in length, both modalities finished approximately at the same time. Participants were tasked to use the LEFT and RIGHT keys to identify the face emotion (as 'happy' or 'angry') as soon as the face appeared on screen, and to make their choice as quickly and as accurately as possible (max. response time = 3s; inter-trial interval (ITI) = 2s fixation mark). Arrow direction was counterbalanced across participants. Only key presses made after the onset of the visual stimulus were recorded in E-Prime (i.e. no selections based on purely *unimodal* audio information were collected). Participants were told to ignore the voice and focus on the face when making judgments. The entire task was made up of 2 blocks, with 18 bimodal trials each, presented in a randomized order: First and second blocks were identical in nature (i.e. same set of face/sentence pairings), but re-randomized in order to avoid primacy and recency recall effects; repetition of stimuli was borrowed from the Davis et al. (2011) paradigm[2]. Given the fully-crossed design (3 EF x 3 EP), each of the 9 bimodal conditions were presented twice within each block (i.e. total of four measurements).

Recognition Task. The surprise recognition task consisted of 2 blocks (*face recall* and *sentence recall* blocks, separated by a self-paced interval), counterbalanced for order. Both blocks were identical in format, wherein participants made recognition judgments, i.e. identified a stimulus as "old" or "new". Each block consisted of 36 trials, presented in randomized order, of which 18 contained stimuli previously presented in the bimodal integration task (i.e. test stimuli), intermixed with 18 that were new (i.e. foil stimuli). Test and foil stimuli sets were counterbalanced across individuals. Subjects used the LEFT vs. RIGHT arrow keys to make old/new choices; direction counterbalanced across participants. Task was self-paced (ITI = 2s fixation mark).

Explicit Valence and Arousal Ratings. Valence and arousal ratings were taken for each of the 18 face test stimuli (in keeping with the Davis et al., 2011, design). Trials were presented in a randomized order. In each trial, using a 9-point Likert scale, participants were first asked to pro-

2 To the best of our knowledge, Davis et al.'s (2011) study is one of the few designs that combined face and word information to investigate subsequent recognition effects. Building off this work, we utilized a similar number of trials, and implemented the same block design format and recognition paradigm. Specifically, Davis et al. (2011) used a 2-block design, whereby each block consisted of 20 trials, presented twice (re-randomized order). Their recall phase included a recognition task made up of the 20 test stimuli and 20 additional foil stimuli (for a total of 40 old/new trials).

vide a face valence rating ("How positive or negative is the expression on this face?" from (1) *Very Negative* to (9) *Very Positive*), followed by a face arousal rating ("How high or low is the emotional arousal of the expression on this face?" from (1) *Very Low* to (9) *Very High*). Task was self-paced (ITI = 2s fixation mark).

Planned Analyses

Confirmatory Effects for Facial Emotion Identification: Planned Analyses. Two 2-way repeated-measures ANOVAs, with two within-subject factors, were planned: EF (angry, neutral, happy) x EP (angry, neutral, happy). The same analysis was planned for each dependent variable (DV), namely: accuracy (i.e. proportion of correct 2AFC selections) and mean RT, averaged across trials. Inferences are based on a nominal alpha-level of $\alpha = .05$[3].

Exploratory Effects on Semantic Recall: Planned Analyses. One 2-way repeated-measures ANOVA, with two within-subject factors, was planned: EF (angry, neutral, happy) x Congruency (congruent, incongruent). Semantic recall was computed as mean sentence recognition hit rates, averaged across trials. Additionally, a 2-way repeated-measures ANOVA of EF (angry, neutral, happy) and EP (angry, neutral, happy) was planned to identify any main effect of EP on recall. Though we specify only one effect of central interest (namely, the interaction of EF x Congruency), nevertheless, we recognize that planned analyses are exploratory in nature, and were thus corrected for two analyses, each with 3 potential effects of interest (i.e. two main effects, one interaction effect); Bonferroni-corrected alpha ($\alpha = .008$)[4]. This is in accordance with recommendations by Cramer et al. (2016) for exploratory multiway ANOVA in order to maintain a family-wise type-1 error rate of 5%.

Additional Analyses. The same exploratory analyses specified for semantic recall effects were also planned for face recall with Bonferroni-corrected alpha ($\alpha = .008$). Additionally, to compare overall recognition rates of faces versus sentences, paired-samples t-test (non-directional) was run after applying a *correction for intrusions* formula (to correct for false alarms; Davis et al., 2011):

Correction for intrusions = (hits – false alarms) / (total number of old stimuli) (1)

Finally, in order to check for differences in valence and arousal of test face stimuli, a 1-way repeated-measures MANOVA, with EF (angry, neutral, happy) as within-subject factor, was conducted on valence and arousal ratings. *Note:* For all aforementioned (M)ANOVAs, simple effects tests (with Bonferroni-correction specified) were run following significant interactions.

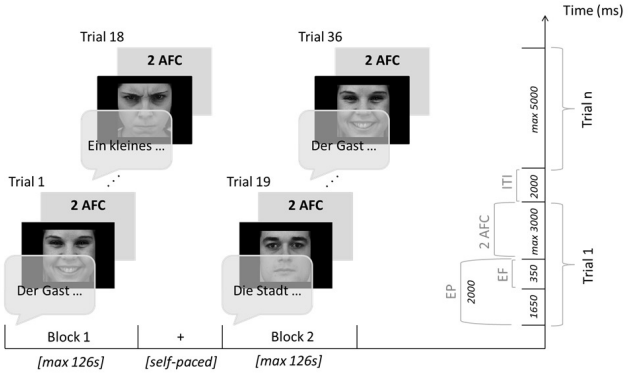

Figure 1. Bimodal Integration Task Design

Note. Within-subject design with two blocks (represented along the x-axis): Each block contains 18 trials, presented in randomized order, and separated by a self-paced break. Y-axis breaks down the timing of the trials: Emotional Face (EF) = 350s, is paired simultaneously with Emotional Prosody (EP ~ 2s); SOA (stimulus onset asynchrony) = 1,650ms. Audiovisual presentation is followed by a two-alternative forced-choice (2AFC) task (time limit = 3s); inter-trial interval (ITI) = 2s. Max. block duration = 126s; max. testing phase (discounting the break) = 252s ~ 4.2 mins. KDEF images used in figure as examples are: "AF25HAS", "AF20ANS", "AM06NES".

Results

Sample

Initial replication sample was $N = 52$; three participants were excluded based on the aforementioned exclusion criteria. Thus, all analyses were based on a final sample of $n = 49$, which comprised 73.5% females, 24.5% males (2.0% opted not to specify), ranging between the ages of 19-40 years ($M = 24.6$, $SD = 4.8$). All participants spoke German fluently (89.8% native speakers), and were primarily made up of students (91.8%).

Bimodal Integration Effects: Accuracy and RT

Accuracy. Two-way repeated-measures ANOVA revealed significant main effects of EF ($F(1.12, 53.65) = 291.75$, $p < .001$, $\eta_p^2 = .86$), and EP ($F(2, 96) = 8.39$, $p < .001$, $\eta_p^2 = .15$), as well as a significant interaction effect between EF and EP ($F(2.90, 139.39) = 7.96$, $p < .001$, $\eta_p^2 = .14$), on accuracy rates (see Figure 2, left plot). Results demonstrated ceiling effects for accuracy rates of angry and happy faces; in other words, accuracy for emotional faces were not significantly affected by accompanying prosodies. Post-hoc pairwise comparisons demonstrated that when neutral faces were paired with an angry prosody (as opposed to a neutral prosody), participants were significantly more likely to label these neutral faces as 'angry' ($M_{diff} = .15$, $SE = .05$, $p = .006$, 95% CI [.04, .27]). Surprisingly, this same pattern

3 In our original pre-registration, we applied a Bonferroni-correction ($\alpha = .0167$) for three effects of interest (i.e. two main effects, and one interaction effect). This, however, is the recommendation offered for exploratory (rather than *a priori* confirmatory) hypotheses (see Cramer et al., 2016). Alpha-level was therefore updated ($\alpha = .05$); change in alpha-level had no effect on the results.

4 In the original pre-registration, Bonferroni-correction account for multiple effects of interest ($\alpha = .0167$), but forgot to account for the two separate ANOVA analyses; alpha-level was therefore updated ($\alpha = .008$).

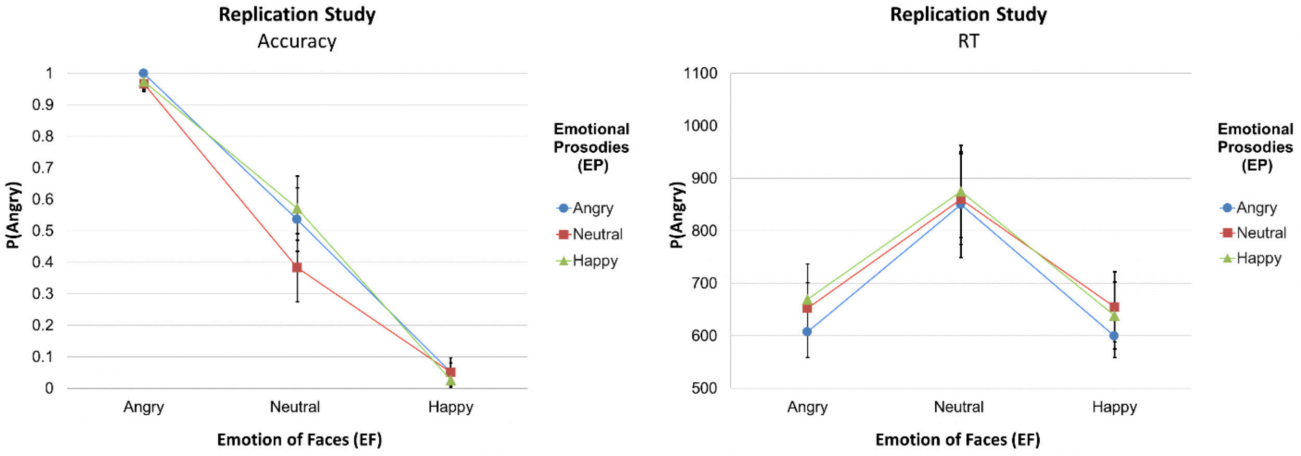

Figure 2. Bimodal Integration Effects

Note. Left plot: Accuracy (y-axis) is operationalized as proportion faces identified as 'angry'. Right plot: RT (y-axis) is measured in milliseconds (ms); higher values represent slower RTs. Error bars represent the 95% confidence intervals.

was observed when neutral faces were paired with a happy prosody; that is, participants again were more likely to classify these neutral faces as 'angry' (M_{diff} = .19, *SE* = .04, p < .001, 95% CI [.08, .30]). The latter finding contradicts prior work which has typically observed a greater tendency to label neutral faces as 'happy' when accompanied by a happy prosody.

RT. Two-way repeated-measures ANOVA revealed significant main effects of EF (F(1.20, 57.72) = 50.41, p < .001, η_p^2 = .51), and EP (F(1.78, 85.30) = 8.83, p = .001, η_p^2 = .16), and a non-significant interaction effect between EF and EP (F(2.80, 134.35) = 0.68, p = .554, η_p^2 = .01), on RT (see Figure 2, right plot).

Overall, findings for confirmatory effects were mostly consistent with past work. Emotional faces were identified significantly more quickly than neutral faces. Additionally, effects of prosody were more visible under ambiguous conditions, i.e. neutral face trials. The presence of emotional vocal information influenced the way in which neutral faces tended to be classified. One notable finding was the unexpected influence of happy prosody described above. Potential reasons for this inconsistency are discussed further in the discussion.

Sentence Recall Effects

Two-way repeated-measures ANOVA between EF (angry, neutral, happy) and Congruency (congruent, incongruent) revealed a significant main effect of EF (F(2, 96) = 7.65, p = .001, η_p^2 = .14), a non-significant main effect of Congruency (p = .719), and a non-significant interaction effect between EF and Congruency (p = .087). Main effect of EF showed that, overall, recall for sentence content was greater when paired alongside emotional faces rather than neutral faces (see Figure 3, left plot), with greater recall rates for trials with happy versus neutral faces (M_{diff} = .10, *SE* = .03, p = .009, 95% CI [.02, .18]), and for trials with angry versus neutral faces (M_{diff} = .07, *SE* = .03, p = .046, 95% CI [.00, .15]).

In order to check whether results might be explained by main effects of EP, a 2-way repeated-measures ANOVA, with EF and EP as within-subject factors, was run. Notably, the analysis revealed a significant main effect of EP (F(2, 96) = 5.88, p = .004, η_p^2 = .11), and a non-significant interaction effect between EF and EP (p = .983). On average, sentences spoken in a happy prosody were recalled more poorly than sentences spoken in an angry prosody (M_{diff} = .09, *SE* = .03, p = .045, 95% CI [.00, .17]) or in a neutral prosody (M_{diff} = .10, *SE* = .03, p = .006, 95% CI [.02, .18]). Given this finding, there may be some grounds to speculate an overall 'impairing' effect of happy EP on sentence content retention, within this specific paradigm (see Figure 3, right plot). Moreover, this is consistent with the above finding, whereby trials with happy faces led to poorer recall rates when paired with a happy (i.e. congruent) EP, as opposed to the incongruent prosodies. Interpretation of reported effects are discussed in more detail in the discussion.

Face Recall Effects

No clear pattern of effects for face recall rates was observed across conditions. For both analyses, all effects fell above the alpha-corrected significance level (α = .008). Specifically, two-way repeated-measures ANOVA, with EF and Congruency as within-subject factors, produced non-significant main and interaction effects ($ps \geq .089$; see Figure 4, left plot). Similarly, 2-way repeated-measures ANOVA, with EF and EP as within-subject factors, revealed no significant main or interaction effects ($ps \geq .033$; see Figure 4, right plot).

Overall Sentence versus Face Recall

Mean recognition rates (after correcting for false-alarms; see Eq. 1 above) was compared between sentences and faces. Overall, recall for sentences (M = .49, SD = .20) was significantly greater than recall for faces (M = .35, SD = .19), t(48) = 3.51, p = .001. These results align with previous findings which show greater recognition of words than faces (Davis et al., 2011).

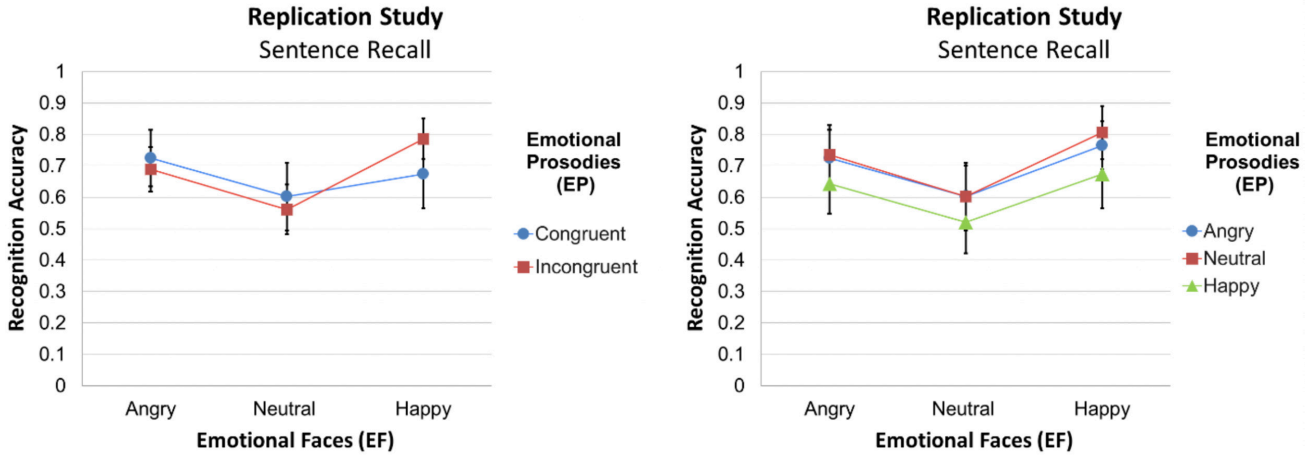

Figure 3. Sentence Recall Effects

Note. Left plot: EF & Congruency on sentence recall. Right plot: EF & EP on sentence recall. Y-axis represents mean recognition accuracy for sentence recall. Error bars represent the 95% confidence intervals.

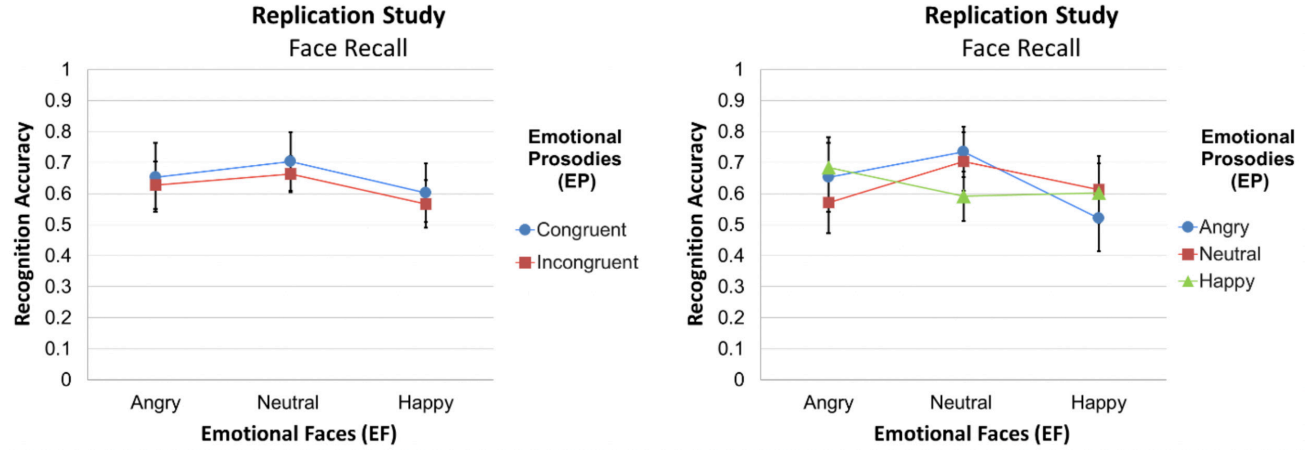

Figure 4. Face Recall Effects

Note. Left plot: EF & Congruency on face recall. Right plot: EF & EP on face recall. Y-axis represents mean recognition accuracy for face recall. Error bars represent the 95% confidence intervals.

Valence and Arousal Ratings

One-way repeated-measures MANOVA revealed expected effects of Emotion (angry, neutral, happy) on valence and arousal ratings for face stimuli. Angry faces were rated significantly more negatively than neutral faces (M_{diff} = 2.70, *SE* = .13, $p < .001$, 95% CI [2.38, 3.02]), and happy faces were rated significantly more positively than both neutral (M_{diff} = 2.95, *SE* = .11, $p < .001$, 95% CI [2.69, 3.21]) and angry faces (M_{diff} = 5.65, *SE* = .17, $p < .001$, 95% CI [5.22, 6.08]). With regard to arousal ratings, both angry faces (M_{diff} = 2.58, *SE* = .35, $p < .001$, 95% CI [1.72, 3.44]) and happy faces (M_{diff} = 3.60, *SE* = .23, $p < .001$, 95% CI [3.04, 4.15]) were rated as significantly more emotionally arousing than neutral faces; happy faces were also rated as significantly more arousing than angry faces (M_{diff} = 1.02, *SE* = .27, $p = .001$, 95% CI [0.36, 1.67]).

Discussion

At present, it is known that multimodal emotional cues, especially facial expressions (i.e. visual modality) and speech prosody (i.e. auditory modality), concurrently interact and modulate how presented information is processed (for review, see Klasen et al., 2012). Enhanced perception and recall of multisensory information appears to critically depend on the extent to which there is a match between both sources of information, such as: faster and more accurate perception of faces when accompanied with a congruent emotional tone (e.g., sad face paired with a sad voice; de Gelder & Vroomen, 2000), faster lexical processing for words whose linguistic content matched the emotional tone (e.g., happy words spoken in a happy prosody; Nygaard & Queen, 2008), and greater recall for audio-visual pairs that semantically matched (e.g., image of a bell paired with the sound of a bell; Murray et al., 2004). Additionally, whether the presence of emotional stimuli has an enhancing versus impairing effect on recall of accompanying neutral stimuli,

depends on contextual influences, such as how neutral and emotional stimuli relate, and the demands of the task (e.g., task-relevance or prioritization of stimuli; see Riegel et al., 2016).

To date, little research has explored the mutual influence or interplay between facial and vocal modalities in the processing and recall of stimuli (see Kauschke et al., 2019). While a body of evidence does exist on the facilitatory effects of crossmodal congruency on *perception* of emotional stimuli (e.g., Klasen et al., 2012; Paulmann & Pell, 2011), research has seemingly neglected to investigate effects on subsequent *recall*.

To this end, the present study asked whether the automatic integration of unattended emotional information is limited to nonverbal cues, i.e. emotional prosody, or if it additionally had effects on retention of verbal content. To answer these questions, we expanded upon previous paradigms in order to assess the extent to which the presence of an emotional prosody (presented alongside a face with a congruent vs. incongruent emotional expression) not only influences individuals' performance on the given task (i.e. speed and accuracy of classifying emotional faces), but additionally, subjects' performance on a subsequent surprise recall task measuring retention of sentence content. In this way, the current study consisted of both a confirmatory replication of bimodal integration effects (i.e. accuracy and RT), as well as an exploratory extension (i.e. recall effects).

Because one main goal of the study was to replicate bimodal integration effects, we incorporated key design elements from past relevant paradigms (de Gelder & Vroomen, 2000; Dolan et al., 2001; Ethofer et al., 2006; Massaro & Egan, 1996) into our methodology. This included task instructions, 2AFC task design, and facial/vocal stimuli properties (incl. presentation and duration). Additionally, to implement our exploratory extension, we borrowed design elements from the Davis et al. (2011) paradigm, including approximate number of trials, block format, and recognition task. One added strength in our design was the addition of the valence and arousal ratings for the face stimuli. Up until now, bimodal integration studies have not measured nor reported these ratings. Inclusion of such measures highlights the need to specify and measure auxiliary assumptions concerning the role of valence and arousal of stimuli in crossmodal research. Another added strength was the inclusion of neutral prosody stimuli. Certain past designs (e.g., de Gelder & Vroomen, 2000), compared the effects of emotional prosody (i.e. presence of happy vs. sad voice) against the absence of prosody altogether (rather than the presence of a neutral voice). We argue that a more systematic and valid investigation of crossmodal influences, on both perception and recall, should entail comparisons across bimodal conditions (rather than a combination of unimodal and bimodal conditions).

Given our specific paradigm, we speculated that we might observe an overall trade-off in terms of central versus peripheral task performance: Conditions which increased attentional capture toward the central task (i.e. face classification task) may bias attentional resources away from processing and retaining peripheral information (i.e. sentence content).

With regard to the confirmatory hypotheses, our results corroborated past work, with the exception of one unexpected finding. Consistent with past findings, participants displayed both faster and more accurate recognition of emotional faces when faces were paired with a congruent emotional prosody (de Gelder & Vroomen, 2000; Dolan et al., 2001; Massaro & Egan, 1996). This underscores the nature of crossmodal emotion perception: That is, the phenomenon whereby affect-relevant information or cues, from discrete sources, are automatically integrated when presented simultaneously. Surprisingly, however, our data showed that when neutral faces were accompanied by a happy prosody, participants had a greater tendency to classify them as 'angry' rather than 'happy'. This contradicts prior work which has typically shown a bias toward judging neutral faces in the direction of the accompanying emotional prosody (e.g., de Gelder & Vroomen, 2000).

While at face value, this finding appears conflicting, it is possible that it too was a result of the top-down influence of emotional prosody on processing. Given the nature of our paradigm, we explain this as follows: We know from the pretest results that, in the absence of an accompanying voice, and in the context of forced choice (i.e. 2AFC task), neutral expressions were more likely to be classified as 'angry' than 'happy'. We can thus speculate that when contrasted against a happy prosody, neutral faces may have appeared particularly negative, and in turn, more likely to be classified as 'angry'. In general, our findings contribute to a broader understanding of the bimodal integration effect. Specifically, when it comes to face judgments, the presence of accompanying prosodic information may produce *intuitive* results when the accompanying emotional prosody *reinforces* one's perceptions of a face, whereas it may produce *counterintuitive* results when the accompanying emotional prosody *contrasts* one's perceptions of a face.

One could alternatively assume that this finding might be due to problems with the stimuli (i.e. neutral expressions and/or 'happy' sentences not accurately conveying the intended emotion). We would argue that this explanation is implausible given the following: a) facial stimuli were selected from a validated database (Goeleven et al., 2008); b) facial stimuli were additionally rated by our sample on valence and arousal dimensions, and demonstrated expected ratings; c) 'happy' and 'angry' sentence stimuli were borrowed from Ethofer et al. (2006); and d) all stimuli were pretested on an independent sample. Thus, there is little reason to question the effectiveness of the stimuli themselves, and we posit that our findings should be interpreted to reflect the specific interaction between facial and vocal cues, under the current task demands.

With regard to the exploratory hypotheses, we did not observe a main effect of congruency on sentence recall rates. In other words, unlike accuracy and reaction time measures, recall performance was not facilitated by a match between emotion in the face and voice. Thus, while it is clear that individuals incorporated the peripheral vocal information into their decisions during the task, this may have been restricted to surface-level features (i.e. hedonic valence or emotional tone) as opposed to semantic content, given that recall rates across congruent and incongruent conditions were found to be comparable. These findings lend support for the arousal-biased competition theory

(Mather & Sutherland, 2011) insofar as reflecting the prioritization of incoming stimuli and the influence of top-down goals. In the case of our paradigm, increased priority is placed toward goal-relevant stimuli (i.e. emotions in the faces). While vocal information is not task-relevant, we know from the bimodal integration literature that congruency between face and voice emotions enhances perception, whereas the content of the spoken sentences lends no added goal-relevant value. Therefore, according to this theory, we should expect that individuals process information selectively based on the task at hand (i.e. integrating the prosody but not the content).

The one overall pattern we found was a main effect of the vocal modality on sentence recall, namely an overall 'impairing' effect of happy EP on sentence content retention. Again, we would argue that this finding is unlikely to be explained by a problem with the happy vocal stimuli for the reasons stated above. Rather, our finding aligns with the idea of an emotional interference effect (Anderson & Shimamura, 2005; Christianson, 1992; Touryan et al., 2007); that is, "emotion can interfere with the processing of unrelated stimuli via the modulation of attention" (West et al., 2017, p. 756). West and colleagues (2017) appeal to this reasoning to explain the negative effect of emotional prosody (i.e. happy or fearful) on word recall. Moreover, authors observed a greater interference effect for fearful prosody, which they took as support for a negativity bias. Within the context of our paradigm, because the accompanying sentences are not relevant to the task, we would expect that fewer attentional resources would be allocated toward processing their semantic content when spoken in an emotional, as opposed to neutral, prosody. While our findings partly corroborate those of West et al. (2017), in terms of happy prosody leading to worse recall than neutral prosody, we did not find that negative (i.e. angry) prosody produced worse recall.

Based on the current state of the literature, it is difficult to speculate why we only observed an interference effect for the happy prosody. First, while ample research has been conducted on the valence effects of faces and words, very few studies to date have explored how prosodic characteristics of speech influence word processing (for review on valence effects, see Kauschke et al., 2019). Therefore, it is unclear whether the effects of prosody, on sentence processing and recall, are dependent on emotional valence (i.e. positive vs. negative) or may hinge on the type of emotion being conveyed (e.g., happy, fearful, angry, etc.). Moreover, it is uncertain how these potential effects would operate under our specific task demands, where central and peripheral information are naturally integrated.

Another important consideration is methodological factors: In West et al. (2017), word learning was the central task, whereas in our paradigm, sentences constituted peripheral information and were task-irrelevant. Kauschke et al. (2019) stressed the importance of task features on valence effects for word-processing studies; depending on the type of task (e.g., lexical decision vs. memory task), as well as stimulus characteristics (e.g., word as target item vs. distractor), the effect of emotional valence on processing may differ. Additionally, the results of the current study are based on use of a recognition task; it is possible that the use of a free recall task may influence outcomes. Taken together, further research on prosody-specific effects is warranted to better understand the role of emotional prosody on word processing and recall.

Finally, we also explored the effects of emotional expressions and vocal prosody on face recall. Our data showed no main effects of modality or congruency on face recall; in other words, neither the type of emotion expressed in the face or voice, nor whether facial and vocal emotions matched, led to significantly greater retention rates of faces. Additionally, the interaction effect between modality and congruency on face recall was non-significant. We recommend that such preliminary results for face recall be considered inconclusive at this stage, and be substantiated by further research. It should be also noted that we did not observe trade-off effects for sentence and face recall. In other words, while happy prosody trials yielded relatively lower recall rates for sentence content (i.e. peripheral stimuli), this was not accompanied by relatively higher recall rates for faces (i.e. central stimuli).

Conclusion

The current experiment sought to explore the effects of emotional expressions and emotional prosody on the perception of facial expressions (i.e. speed and accuracy of identifying emotions of faces) and retention of semantic information (i.e. recall rates for content of concurrently spoken sentences). Specifically, we investigated whether congruent conditions (i.e. when emotion in the face and voice matched) led to faster and more accurate recognition of face emotions, as well as greater recall for sentence content. While congruent conditions were found to facilitate perception of faces, replicating past findings, we did not find an overall congruency effect when it came to sentence content recall. In other words, on average, recall rates for sentence content were not found to significantly differ between congruent and incongruent trials. What we did find, however, was an overall effect of prosody on sentence recall; namely, significantly poorer recall rates for trials accompanied with a happy prosody, as compared to trials paired with neutral or angry prosodies.

Taken together, our findings suggest that when individuals integrate facial and vocal cues, information in the periphery (i.e. accompanying sentences) is processed more superficially; integration is restricted to surface-level features such as hedonic valence or emotional tone, rather than semantic content. This would explain why the presence of accompanying vocal information demonstrated the expected congruency effects for central task performance, but did not yield differences in terms of peripheral sentence content recall. We also found poorer overall recall rates for sentences spoken in a happy tone, suggesting that happy prosody may pose an emotional interference effect on semantic processing: When faces are accompanied by a happy prosody, attention may be further shifted away from task-irrelevant stimuli (i.e. sentences) in favor of task-relevant stimuli (i.e. faces). The present findings alone do not suffice for making broader claims about prosody-specific valence effects; however, we feel that this is a line of research that merits to be addressed with further research. Finally, our study did not

find support for the idea of trade-off effects for sentence and face recall. We believe that a more complex framework is needed to explain the potential interactions of face and voice modalities on recall effects.

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no conflicts of interest or competing interests to declare that are relevant to the content of this article, and no relevant financial or non-financial interests to disclose.

Ethics Approval

Approval was obtained from the ethics committee of the Ludwig-Maximilians-Universität (Ethikkommission der Fakultät 11). The procedures used in this study adhere to the tenets of the Declaration of Helsinki. All participants provided written informed consent prior to participation. Participation in the study was completely voluntary, and subjects could withdraw at any time.

Author Contributions

AH generated the idea for the study and was responsible for the study design. AH, SP, and LH were jointly responsible for data collection, data analysis, and manuscript writing. DW contributed study materials and was involved in revisions of the manuscript at multiple stages. All authors approved the final submitted version of the manuscript.

Data Accessibility Statement

The data (i.e. anonymized data sets, syntax, output, tasks, supplemental online materials) that support the findings of this study are openly available on the OSF platform under the main project *"Emotion-Modulated Recall"* (https://osf.io/am9p2/). Direct links to project pre-registration (https://osf.io/v9ghm/) and supplementary information (https://osf.io/y3amf/).

All code (SPSS syntax) needed to reproduce reported results are openly available on OSF (see above).

Submitted: August 11, 2021 PST, Accepted: November 16, 2021 PST

REFERENCES

Anderson, L., & Shimamura, A. P. (2005). Influences of emotion on context memory while viewing film clips. *The American Journal of Psychology*, *118*(3), 323–337.

Christianson, S.-A. (1992). Emotional Stress and Eyewitness Memory: A Critical Review. *Psychological Bulletin*, *112*(2), 284–309.

Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*(3), 211–226. https://doi.org/10.1016/j.bandl.2011.12.007

Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., Waldorp, L. J., & Wagenmakers, E.-J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, *23*(2), 640–647. https://doi.org/10.3758/s13423-015-0913-5

Davis, F. C., Somerville, L. H., Ruberry, E. J., Berry, A. B. L., Shin, L. M., & Whalen, P. J. (2011). A tale of two negatives: Differential memory modulation by threat-related facial expressions. *Emotion*, *11*(3), 647–655. https://doi.org/10.1037/a0021625

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*(3), 289–311. https://doi.org/10.1080/026999300378824

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, *1242*, 136–150. https://doi.org/10.1016/j.brainres.2008.03.071

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*(17), 10006–10010. https://doi.org/10.1073/pnas.171288598

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*(3), 183–201. https://doi.org/10.1037/h0047707

Ekman, P., & Friesen, W. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., Reiterer, S., Grodd, W., & Wildgruber, D. (2006). Impact of voice on emotional judgment of faces: An event-related fMRI study. *Human Brain Mapping*, *27*(9), 707–714. https://doi.org/10.1002/hbm.20212

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*(6), 1094–1118. https://doi.org/10.1080/02699930701626582

Guillet, R., & Arndt, J. (2009). Taboo words: The effect of emotion on memory for peripheral information. *Memory & Cognition*, *37*(6), 866–879. https://doi.org/10.3758/MC.37.6.866

Kauschke, C., Bahn, D., Vesker, M., & Schwarzer, G. (2019). The Role of Emotional Valence for the Processing of Facial and Verbal Stimuli—Positivity or Negativity Bias? *Frontiers in Psychology*, *10*, 1654. https://doi.org/10.3389/fpsyg.2019.01654

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*(4), 575–591. https://doi.org/10.1016/j.jml.2006.05.004

Klasen, M., Chen, Y.-H., & Mathiak, K. (2012). Multisensory emotions: Perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*(4), 381–392. https://doi.org/10.1515/revneuro-2012-0040

Lakens, D., & Caldwell, A. R. (2019). *Simulation-Based Power-Analysis for Factorial ANOVA Designs*. https://doi.org/10.31234/osf.io/baxsf

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*(4), 405–414. https://doi.org/10.1007/s00221-004-1913-2

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, *24*(2), 326–334. https://doi.org/10.1016/j.cogbrainres.2005.02.005

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institut*, *91*(639), 2.

Mackay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition*, *32*(3), 474–488. https://doi.org/10.3758/BF03195840

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3*(2), 215–221.

Mather, M., & Sutherland, M. R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspectives on Psychological Science*, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage*, *21*(1), 125–135. https://doi.org/10.1016/j.neuroimage.2003.09.035

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017–1030. https://doi.org/10.1037/0096-1523.34.4.1017

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, *35*(2), 192–201. https://doi.org/10.1007/s11031-011-9206-0

Riegel, M., Wierzba, M., Grabowska, A., Jednoróg, K., & Marchewka, A. (2016). Effect of emotion on memory for words and their context: Emotion and Memory for Words and Their Context. *Journal of Comparative Neurology*, *524*(8), 1636–1645. https://doi.org/10.1002/cne.23928

Sakaki, M., Fryer, K., & Mather, M. (2014). Emotion Strengthens High-Priority Memory Traces but Weakens Low-Priority Memory Traces. *Psychological Science*, *25*(2), 387–395. https://doi.org/10.1177/0956797613504784

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61–70. https://doi.org/10.1250/ast.28.61

Sutherland, M. R., & Mather, M. (2012). Negative arousal amplifies the effects of saliency in short-term memory. *Emotion*, *12*(6), 1367–1372. https://doi.org/10.1037/a0027860

Touryan, S. R., Marian, D. E., & Shimamura, A. P. (2007). Effect of negative emotional pictures on associative memory for peripheral information. *Memory*, *15*(2), 154–166. https://doi.org/10.1080/09658210601151310

Watson, D., Anna, L., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

West, M. J., Copland, D. A., Arnott, W. L., Nelson, N. L., & Angwin, A. J. (2017). Effects of emotional prosody on novel word learning in relation to autism-like traits. *Motivation and Emotion*, *41*(6), 749–759. https://doi.org/10.1007/s11031-017-9642-6

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. https://doi.org/10.1017/S0140525X17001972

SUPPLEMENTARY MATERIALS

Peer Review History

Download: https://collabra.scholasticahq.com/article/31601-emotion-modulated-recall-congruency-effects-of-nonverbal-facial-and-vocal-cues-on-semantic-recall/attachment/79166.docx?auth_token=q10MBewmnvJYWxZDWYmV

Supplemental Materials

Download: https://collabra.scholasticahq.com/article/31601-emotion-modulated-recall-congruency-effects-of-nonverbal-facial-and-vocal-cues-on-semantic-recall/attachment/79167.pdf?auth_token=q10MBewmnvJYWxZDWYmV