When voices get emotional: A study of emotion-enhanced memory and impairment during emotional prosody exposure

Cyrrielle Chappuis 1, Didier Grandjean 1,2

1 Swiss Center for Affective Sciences, University of Geneva
2 Neuroscience of Emotion and Affective Dynamics, University of Geneva
cyrielle.chappuis@unige.ch, didier.grandjean@unige.ch

Abstract

Emotional expressions influence memory by both impairing and enhancing attention and perception at encoding, consolidation and recall. In three studies we compare angry, happy and fearful emotional prosodies using a single-word presentation paradigm, and investigate the effects of emotional vocal expressions on immediate and delayed recall for central emotional items and their periphery. Overall the results support a prosody-induced emotional enhancement of memory, as well as a memory facilitation for the periphery of specific emotions. The fact that emotional enhancement of memory by prosody was replicated in three studies suggests a robust effect, which to our best knowledge has not been observed before with vocal expressions.

Index Terms: emotion, prosody, memory, trade-off, facilitation

1. Introduction

Emotional stimuli are known to benefit from prioritized processing when appearing among neutral ones [1]. The cocktail party effect is a good example of such an advantage: in a crowded party, hearing a sudden exclamation has been found to capture attention [2]. Moreover, dichotic exposure to prosody showed that emotional prosody was automatically processed among neutral prosodies [3]. In this case and in this report, the term “emotional prosody” is used to refer to perceived emotion embedded in linguistic vocalizations and recognized as such [4]. Can emotional prosody also influence memory processes? A wealth of literature and daily life experience suggest that emotional stimuli are better memorized than neutral stimuli, an effect known as emotion-enhancement of memory (EEM) [5-9]. On the other hand, several studies point out the existence of retroactive and proactive memory interference due to emotionality [10]. It suggests that emotions convey and imply a tendency to broaden or narrow the attention span: negative arousing stimuli lead to memory enhancement for emotional items and impairment for the periphery of emotional items [5,11-14] while positive arousing stimuli lead to memory facilitation for the periphery of emotional items [15-18]. A specific emotion's communicative value has also been linked to distinctive effects [19,20]. Indeed, fear and anger are both labeled as negative-arousing emotions, but they do not convey the same information. For instance, perceiving a fearful expression indicates an indirect threat in the environment, and an angry expression indicates a direct threat (the emotional source itself). Moreover, fearful events have been shown to lead to better recall of the periphery (facilitation) whereas angry ones led to a better memorization of the emotional event itself, to the cost of its periphery [21]. In the present study, we presented participants with fear, anger and happy vocal expressions, and expected different patterns of peripheral memorization depending on their communicative value. Indeed, since happy and fearful prosodies should both enhance memory for peripheral items (facilitation hypothesis), we expected angry prosody to be associated with poorer peripheral memorization (impairment hypothesis).

Currently, most experiments in the EEM as well as memory facilitation and impairment literature imply visual tasks. However, what has better communicative value than language itself? Prosody has been shown to convey specific traits depending on the speaker’s emotional state [22]. Moreover, emotional prosody has been found to be automatically processed [23], which supports the conception that emotional auditory stimuli could benefit, as visual stimuli do, from prioritized processing when competing for resources with neutral stimuli [1]. Recently, Schirmer, Chen, Chin, Tan and Hong (2013, [24]) demonstrated that sad vocal expression had an impact on memorization and long-term representation of presented words. Sadly pronounced words were recognized more often and evaluated more negatively than neutral words.

To our best knowledge, the advantage of emotional prosody in memory as well as impairment and facilitation effects for auditory stimuli have not been reported before. We tested whether happy, angry and fearful prosodies were better remembered than neutral prosody, as well as their impacts on peripheral information.

2. Experiment 1

2.1. Method

Participants. Twenty four (12M/12F) native French speakers were recruited. Participants' age ranged from 18 to 35.

Materials. Ninety-two neutral common nouns [32] were recorded in the sound deadening booth at the Geneva Brain and Behavior Laboratory with the help of four stage actors (2M/2F). Sounds were normalized and a pilot study was then conducted to ensure that the emotional tones were valid and reliable. Selected words were split into four lists of words (one voice per list) and counterbalanced within the lists. The lists were made up of twenty neutrally-expressed and three emotionally-expressed words (a happy one, an angry one and a fearful one) which remained in the same position. Each word immediately preceding an emotional item constituted the "impairment condition" (E-1), whereas words that immediately preceded control items represented their controls (C-1). The same procedure was applied to create the "facilitation condition" with items directly following emotional (E+1) and control (C+1) items. Words were presented with a 1000ms inter-stimuli interval. The experiment was run on Eprime 2 Professional (www.pstnet.com/products/e-prime).

Design and procedure. Participants were successively presented with the four lists, and were instructed to memorize the words they heard through headphones, in order to recall
them immediately after each list. We measured the number of words recalled for each condition in all four lists.

In line with the literature on the EEM effect, we expected emotional stimuli (E) to be recalled more often than control stimuli (C). Furthermore, we expected a memory impairment effect to occur for peripheral stimuli, in that target stimuli immediately preceding emotional ones (E−1) would be recalled less than the ones preceding controls (C−1). We also expected a facilitation effect for stimuli directly following emotional stimuli (E+1), in that they would be recalled more often than stimuli following controls (C+1).

2.2. Results

Given the binary (hit/miss) nature of our dependent variable we used a generalized linear mixed model using binomial distribution (GLMM) to perform our analyses. Indeed, this particular model allows us to integrate random effects including primary and recency effects, type of words, as well as participants. Analyses showed that emotional stimuli were better remembered than neutral ones ($Z = 4.28, p < .0001$) (see Figure 1). However, stimuli immediately following emotional ones (E+1) were not found to be remembered more often than stimuli following neutral ones, referred to as C+1 ($Z = 1.64, p > .1$). Likewise, E−1 stimuli were not recalled differently than C−1 stimuli, failing to provide support for an impairment ($Z = -0.58, p > .5$). Further analyses showed that happy ($Z = 2.03, p < 0.05$), angry ($Z = 3.72, p < 0.001$) and fearful ($Z = 3.88, p < 0.001$) stimuli all participated in the facilitation effect on emotional stimuli, suggesting the main effect was not emotion-specific.

Figure 1. Recall rates for emotionally pronounced words (E) and neutrally pronounced words (C) in Experiment 1. Brackets represent standard errors of the means.

3. Experiment 2

In a second study we plan to replicate the results of the first study, and to test the inter-stimuli interval (ISI) effect on memory for emotional vocal expression. Several studies showed that the memory span of a list of words depends on its length and ISI [33-35]: the longer the list lasts, the less is recalled. On the other hand, retroactive impairment seems to follow a gradient: the closer the emotional stimulus is to the others, the higher the impairment is [36]. We hypothesized that the less time participants had, the greater the emotion effects would be. So we expected a higher impairment effect with shorter ISI.

3.1. Method

Participants. Twenty four (12M/12F) French native speakers were recruited. Participants’ age ranged from 18 to 35 years. One subject (male) was excluded for misunderstanding the instructions.

Materials. The materials and procedure were identical to Experiment 1. However, inter-stimulus intervals (ISIs) varied between the lists with 250ms, 500ms, 750ms, and 1000ms ISI.

3.2. Results

The GLMM analysis revealed that emotional words were recalled more often than neutral ones ($Z = 3.61, p < .001$), all ISI conditions taken together. However, contrast analyses did not show any significant difference in the recall of emotional words between the ISI conditions ($Z = 1.07, p < .1$). Hence, we managed to replicate the main result of the Experiment 1, namely the EEM effect, across all ISI conditions, which affords for a quite reliable effect. However, no significant difference in the recall of emotional versus neutral items was found between the ISI conditions, which leads to conclude that the EEM effect was sufficiently strong and thus not to be influenced by, for example, the shortest ISI.

4. Experiment 3

Having replicated our first and main effect of EEM, we conducted a third study aiming at investigating whether this phenomenon could be enhanced or altered by concurrent visual cues, and a longer consolidation time.

Most of our perceptions are a combination and an integration of several percepts from different sensory modalities, and result in a multi-modal ensemble. Indeed, multi-modal integration has been shown to influence encoding, by modifying, enhancing or deteriorating the percepts [25-27,38].

Thus, in this third experiment, we divided participants into three different integration conditions: (1) auditory condition (same as study 1 and 2), (2) a congruent audiovisual condition, and (3) an incongruent audiovisual condition. A greater EEM effect was expected for congruent audiovisual cues than for incongruent or for auditory stimuli only.

Furthermore, in the first two studies we presented here, there were no different delays between the encoding and the recall stages, as the recall was only immediate. We acknowledge that for this reason the results can hardly be compared to those of studies using longer delays, as different processes might be involved. Indeed, whereas the modulation hypothesis states that memory traces for emotional stimuli are strengthened over time with the help of amygdala-medial temporal lobe mechanisms [13], it has been suggested that immediate EEM effects might be mainly explained by cognitive processes [28]. Thus we added a 24-hour delayed recognition test to assess whether emotional prosodic advantages could benefit from long-term advantages in memory.
4.1. Method

Participants. Seventy-two French native speakers (mean age = 23.34, SD = 4.36) were recruited at the University of Geneva (36M/36F). Three were excluded from the analyses for misunderstanding the instructions.

Material. The same set of auditory stimuli was used as in experiments 1 and 2. However, static facial expressions for the four speakers (one identity for each speaker, and four facial expressions) were created using FACEgen (Singular Inversions, www.facegen.com, [37]) and FAC5gen softwares [29]. Static images were mainly chosen for practical reasons, since each auditory stimulus has a different length making it difficult to associate to an animation. When visual stimuli (faces) were presented along with auditory stimuli (words), we made sure they were aligned at the onset of the auditory stimulus, and that their duration was the same. A 500ms inter-stimulus interval was applied between words.

Design and procedure. Participants were randomly assigned to one of the three integration conditions. The first condition, referred to as “auditory condition” consisted of a replication of Experiment 1 and 2, because stimuli were only auditory. The second condition, referred as “audio-visual congruent” consisted of presentations of emotionally matched faces and words (e.g., an angry face paired with an angry voice). The third condition consisted of audio-visual stimuli presented incongruently, such as voices and faces’ emotional category never matched (e.g., an angry face paired with a happy voice, or a neutral face paired with an angry voice). In all audio-visual conditions, participants were not informed about faces and only instructed to listen carefully to the voices. Participants then followed the exact same procedure described in Experiment 1 including an immediate recall test after each list (T1). A recognition test was conducted 24 hours later (T2), in which words used at T1 as well as new words were presented one at a time on a screen with the software Eprime. Participants were asked to press a key to indicate whether the word was “old” (seen at T1) or “new”. Fifty-nine out of sixty-nine participants took part in the experiment at T2.

4.2. Results

To test our hypothesized main effect of emotion (across all integration conditions), we performed a GLMM analysis on the recall at T1. The contrast analysis showed that emotionally-pronounced words (E) were recalled more often than neutrally pronounced words (C) (Z = 3.67, p < 0.001). This result is a replication of the EEM effect found in the two previous experiments, independently of integration conditions. A simple effect analysis showed that angry vocal expressions (Z = 4.58, p < 0.0001) and happy vocal expressions (Z = 2.37, p < 0.05) were both better remembered than neutral ones, whereas fearful vocal expressions did not lead to any different recall than neutral expressions (Z = 1.02, p > .3). A contrast analysis for the E-1 and C-1 (impairment effect) was not significant (Z = -.53, p > .5). Furthermore, we expected integration conditions to have an impact on the EEM effect, however neither incongruent audiovisual cues (Z = -.19, p > .5) nor auditory cues alone (Z = -.99, p > .3) lead to a significantly different recall than congruent audiovisual cues. To test the effect of delay on EEM, we performed the same analysis using T1 and T2 as factors and observed non-significant differences between emotional and neutrally-pronounced words across T1 and T2 (Z = -1.01, p > .3). However, contrast analyses of emotion category (E+1, C+1) across test modalities (T1, T2) showed that words immediately following emotionally-pronounced words were remembered more often at T2 than T1 (Z = 2.94, p < 0.01). Further analyses showed that this broadening was merely driven by happy vocal expressions (Z = 2.02, p < 0.05) and fearful vocal expressions (Z = 2.72, p < 0.01).

5. Discussion

This study sheds light on the specific status of emotional vocal expression in memory processing. Experiment 1 yielded support for a better remembrance of emotionally-pronounced items than for neutrally-pronounced items, followed by experiments 2 and 3 which replicated the EEM effect. These results support the conception of a prioritized processing of emotion, and its effect on memory processes. To our best knowledge, such results have not been observed before with emotional prosody. In Experiment 3, participants were presented with either auditory cues alone, congruent audiovisual cues or incongruent audiovisual cues, but this manipulation did not yield any significant difference in EEM as expected. One way to explain it is by suggesting that emotional auditory cues and visual cues were processed independently. In retrospect, we suppose visual cues were too easy to ignore while performing the task since they were not useful to achieve it, leading participants to rather focus on the auditory cues themselves. In addition, in experiment 3 we used a 24-hour delay and observed a better remembrance of emotionally-pronounced items than neutrally-pronounced items, which reveals that prosodic emotion traces are better maintained over time than neutral traces. Moreover, the facilitation effect observed in the third study after a 24-hour delay accounts for a specific consolidation process over time for items presented after the emotional ones (E+1), in agreement with the modulation hypothesis [13]. Indeed the difference between immediate and delayed recall is crucial, since there is a growing claim in the literature stating that the modulation hypothesis does not account for immediate memory enhancement and facilitation. According to the modulation model, emotional arousal causes adrenergic and cortisol hormones to be released, and leads the amygdala to modulate hippocampal activity, thus enhancing consolidation for emotional items [13]. Besides, we would like to point out that this enhanced consolidation of an emotional item’s periphery over time was driven solely by happy and fearful items. Yet, as mentioned previously, happiness and fear both have “exploratory” communicative values, and were expected to broaden the attentional focus, and hence memorization, for the event at the periphery of the emotional event, whereas anger was expected to lead to a focus on the emotional event itself [21]. These original results strengthen the necessity to replicate such an experiment in order to better observe the specificity of the different emotional impacts on memory and attention processes by adding attentional measures to this design. However, although these results are in line with the modulation hypothesis, Talmi (2013, [30]) has suggested that early emotional-enhancement of memory such as we can observe at T1 and in study 1 and 2 can be better explained by cognitive factors such as attention, distinctiveness, and/or organization which might trigger the allocation of cognitive resources during encoding. A consequence of this assumption is that experimental methodology plays an important part in the emotional enhancement of memory and might explain variations in the previously reported results in the literature. In the present research, we performed three experiments and
replicated the main EEM effect despite the modification of several conditions including ISIs and integration modalities. However, our procedure only allowed for testing hits and misses of presented words, and not the quality of the recall. As mentioned before, emotional prosody might have a long-term effect on the representations of words [24], and several studies have previously shown that emotional stimuli could enhance vividness. We suggest that further experiments should assess both the quantity as well as the vividness of memories. Finally, we would like to point out that even if prosodic stimuli were evaluated prior to the experiments reported in this paper, it is possible that some emotional expressions were more arousing than others. Indeed, arousal is crucial for emotional-enhancement of memory, facilitation and impairment effects [11, 31] and we did not control for the potential differences of arousal levels for fearful, happiness and angry prosodies. Although fear and anger stimuli are usually more arousing than happy stimuli, the observed difference in memory is unlikely to be solely related to arousal. Arousal levels could be derived from physiological measures and they are possibly associated with the variance of observed effects. A further analysis of prosodic parameters using data-driven measures would also allow us to directly link EEM effects to acoustical parameter modulations.

6. Conclusion

The present study provides strong and original evidence for the enhancement of memory by vocal emotional expressions in three different experiments. It demonstrates that vocal emotional expressions are relevant cues in the environment which lead to distinct orientation patterns, and benefit from early emotional-memory enhancement, as well as long-term advantages in memory. The lack of memory impairment results can be explained by the large number of factors influencing emotional memory: arousal level, attention, cognitive factors, consolidation time, stimuli and methodology, valence and relevance. Further research is needed to investigate how emotion in voice impacts on encoded and recalled information. Emotional prosody conveys many relevant cues from which a listener can make inferences, and as such it is a crucial part of our social communication. Its impact on cognitive processes such as memory is still understudied. The way speech impacts on our representations and memories is important to better understand how emotional information is processed, encoded, consolidated and retrieved in everyday life or during traumatic situations.

7. References


