RESEARCH ARTICLE

A study of the effect of auditory prime type on emotional facial expression recognition

Sameer Sethi[1]*, Dr. Marc D. Pell[2]

[1] Faculty of Science, McGill University, Montreal, QC
[2] School of Communication Sciences and Disorders, McGill University, Montreal, QC

*Email Correspondence:
sameer.sethi@mail.mcgill.ca

Keywords:
Vocalization: *auditory stimulus consisting of a non-verbal sound to convey an emotional meaning.*
Pseudoutterance: *auditory stimulus consisting of a verbal, speech-like (but nonsensical), sentence to convey an emotional meaning.*
Auditory Prime: *auditory stimulus that influences the interpretation of the subsequent stimuli (visual in this case).*
N400: *Negative amplitude observed in EEG data 400 msec after the onset of a stimulus; modulated by the congruency or incongruency of the prime-target pair.*

Abstract

Background: In this study, we investigated the influence of two types of emotional auditory primes - vocalizations and pseudoutterances - on the ability to judge a subsequently presented emotional facial expression in an event-related potential (ERP) study using the facial-affect decision task. We hypothesized that accuracy would be greater for congruent trials than for incongruent trials. This is due to the possibility that a congruent prime would allow the listener to implicitly identify the particular emotion of the face more effectively. We also hypothesized that the normal priming effect would be observed in the N400 for both prime types, i.e. a greater negativity for incongruent trials than for congruent trials.

Methods: Emotional primes (vocalization or pseudoutterance) were presented to participants who were then asked to make a judgment regarding whether or not a facial expression conveyed an emotion. Behavioural data on participant accuracy and experimental electroencephalogram (EEG) data were collected and subsequently analyzed for six participants.

Results: Behavioural results showed that participants were more accurate in judging faces when primed with vocalizations than pseudoutterances. ERP results revealed that a normal priming effect was observed for vocalizations in the 150 msec - 250 msec temporal window – where greater negativities were produced during incongruent trials than during congruent trials – whereas the reverse effect was observed for pseudoutterances. Few participants were tested (n = 7). Hence, this study is a pilot study preceding a further study conducted with a greater sample size (n = 25) and slight modifications in the methodology (such as the duration of auditory primes.)

Conclusions: Vocalizations showed the expected priming effect of greater negativities for incongruent trials than for congruent trials, while pseudoutterances unexpectedly showed the opposite effect. These results suggest that vocalizations may provide more prosodic information in a shorter time and thereby generate the expected congruency effect.

Introduction

On a daily basis, we are faced with the task of interpreting emotional meaning in voices by their prosody, the way they are said. A simple question such as "are you kidding?" can be asked with anger, sadness, confusion, or frustration. It can be interpreted wholly differently depending on its prosodic nature. Judging the prosody (and subsequently the meaning) of the question is up to the listener. Therefore, a popular topic of research in recent years has been to develop an understanding of the neural processes and timing involved emotional prosody integration. The investigation of how context influences semantic integration has been studied frequently through the presentation of an auditory emotional prime prior to presenting a visual target.

Priming Effects of Auditory Stimuli

Many studies have shown that the effects of prosodic cues of a voice are not limited to the interpretation of message's meaning, but also include priming effects in interpreting emotion conveyed by facial expressions (1). Information regarding the emotion of a voice is processed and integrated soon after hearing it (2), which helps explain the display of subsequent priming effects when judging a facial expression's emotions. The N400 – a characteristic negative amplitude at 400 msec post-stimulus – is particularly associated with semantic integration of emotion. Studies involving the presentation of a word after an auditory prime has established semantic context have shown that a greater negativity is generated in the N400 when the prime is emotionally unrelated to the visual stimulus (3). Paulmann & Pell

(2010) investigated the effects of the length of an emotional auditory prime on the ability of the listener to judge emotional facial expressions when they were congruent and incongruent with respect to the auditory primes. They tested both short (200 msec) and medium (400 msec) length primes in order to observe the relative amount of prosodic information required to observe such a priming effect. They found that the medium length prime generated the expected ERP pattern: a larger negativity in the N400 for incongruent trials compared to congruent trials. However, they found the reverse relationship when looking at the short length prime. This suggests that a short auditory prime may not provide the same "processing advantage" in establishing an emotional context for a facial expression. As such, long primes (longer than 400 msec) were used in the present study.

Prime Types

A factor that has not yet been studied in depth is the nature of the auditory prime used. Auditory stimuli can take the form of regular sentences with a particular prosody. However, this poses the risk of semantic context playing a role in the priming effect (in addition to the intended conveying of emotion). For this reason, this study made use of primes that did not pose the risk of establishing specific context to the listener.

The aim of the present study was to investigate the priming effect of vocalizations versus pseudoutterances in judging facial expressions by comparing the ERP pattern in the N400. *Vocalizations* refer to non-speech- like sounds that convey emotions, such as the sound "Ahhh!" to convey anger. *Pseudoutterances* resemble speech, but do not have a comprehensible meaning (an example of a pseudoutterance is "they nestered the flugs"). They are used in order to convey emotional meaning without conveying any semantic meaning. We presented both types of primes for their full duration rather than for a universal duration because the latency of the peak of emotional meaning varied from one prime to another. We created an implicit task requiring participants to answer a yes/no question regarding whether or not the presented face conveyed any of three pre-determined emotions – happiness, sadness, or anger - rather than requiring participants to determine the identity of the emotion. With respect to the behavioural data, it was hypothesized that accuracy would be greater for congruent trials than for incongruent trials due to the possibility that a congruent prime would allow the listener to implicitly identify the particular emotion of the face more effectively. With respect to the ERP data, we hypothesized that the normal priming effect would be observed in the N400 for both prime types - that is to say, a greater negativity for incongruent trials than for congruent trials. The justification for this is that both prime types convey emotional meaning without setting up any semantic context, despite the fact that they are intrinsically different in the way they convey emotion.

Methods

Participants

One left-handed and six right-handed English-speakers (6 female; mean age = 21.5 years, SD = 2.2 years) participated in the study. None of the participants reported any hearing impairments. All participants had normal or corrected-to-normal vision. Participants gave informed consent before participating in the study. Participants were compensated $10 per hour for their participation. The right-handed male participant's data was not used for the analysis, as his accuracy was less than 50%.

Stimuli

We used two types of emotional auditory cues (pseudoutterances and vocalizations) as primes and pictures of emotional or non-emotional facial expressions as targets. The primes and targets were paired for cross-modal presentation according to the 'facial affect decision task' (FADT), which involves the pairing of facial expressions with emotionally related or un-related vocal utterances. Both the auditory and visual stimuli were taken from established inventories (4-6).

Prime Stimuli. We used pseudoutterances and vocalizations of three emotion categories: anger, sadness, and happiness. The use of pseudoutterances allows for an emotion to be conveyed without being coupled with semantic information (7, 8). The pseudoutterances were presented by 10 speakers (6 female, 4 male), and were selected based on a validation study conducted by Pell *et al.* (5). Average recognition accuracy for the pseudoutterances were as follows: anger = 93%, sadness= 91%, happiness = 89%. The mean duration of the pseudoutterances was 1966.01 msec (s = 404.9 msec). The vocalizations were presented by 10 speakers (5 female, 5 male), and were constructed by Belin *et al.* (6). Average recognition accuracy for the vocalizations were as follows: anger = 75%, sadness = 77%, happiness = 81%. The mean duration of the vocalizations was 1533.13 msec (s = 799.70 msec). For both pseudoutterances and vocalizations, we presented ten items (one per speaker) for each of the three emotion categories (a total of 30 per participant). Both the pseudoutterances and vocalizations were presented in their entirety to preserve prosodic meaning.

Target Stimuli. The target stimuli consisted of 13.5 cm x 19 cm images with static facial expressions. The images were restricted to the actors' faces and were presented in colour. As with the prime stimuli, there were three distinct emotion types: anger, sadness, and happiness. Facial grimaces representing a non-emotion (a grimace) were presented in equal proportion to each emotion. The stimuli presented consisted of 10 different actors (6 female, 4 male) of various ethnicities (Caucasian, Asian, and African). We required participants to judge whether or not the target facial stimulus presented during each trial represented an emotion. Because the task was implicit, the

participant was not required to identify the specific emotion, but to simply answer "yes" or "no". Yes trials were those presented with faces in either of the three emotion categories (n=30 for each anger, sadness, happiness) and no trials were those that included the aforementioned facial grimaces (n=90). The target stimuli used in the study were retrieved from a previously established inventory (4). The accuracy of recognition for faces conveying a discrete emotion was seen to be the following: anger = 92%, sadness = 88%, happiness = 99% (4). The recognition rate for grimaces (non-emotion) was seen to be 60% in the same validation study. Fig. 1 shows an instance of each emotion as well as a facial grimace posed by one of the actors used in the study.

Fig. 1
Sample Facial Expression for Emotions and Grimace

Design

We required participants to make a yes/no judgment regarding whether or not the target stimulus in each trial represented a discrete emotion. The actual interpretation of the emotion in each trial was implicit since the participant had only to confirm or deny the presence of a discrete emotion in the face. The trial distribution was as follows: for each prime type (vocalizations and pseudoutterances), there were 30 trials for each combination of prime emotion (anger, sadness, happiness) and facial emotion (anger, sadness, happiness), and 90 additional trials for each combination of prime emotion and facial grimaces – a total of 1080 trials.

Procedure

We set up a 64-electrode EEG cap on each participant and seated him or her in a dimly lit room approximately 65 cm from the computer monitor. The visual stimuli on the monitor delivered the auditory stimuli through earphones at a volume set by the participant according to his or her comfort level. Participants were instructed to listen to the auditory stimulus, and carefully observe the face presented in order to subsequently make a yes/no judgment regarding whether or not it conveyed an emotion. The structure of each trial was as follows. First, we presented an auditory prime stimulus (vocalization or pseudoutterance) with a fixation cross present in the middle of the monitor screen. At the offset of the prime, we presented the visual target stimulus (emotional or grimace face) until the participant pressed *yes* or *no* on the response. After the participant's response, the instruction "Blink" was presented for 1000 msec to encourage the participant to blink at that time as opposed to during the following trial. Finally, we displayed a 1500 msec interstimulus interval with a fixation cross in the middle of the screen. The experiment was divided into groups of trials (180 trials each) – referred to as "blocks" – in order to give the participants consistent breaks during the experiment. Prior to the first experimental block, each participant completed two practice blocks consisting of 10 trials each. During the first practice block, we presented a face with an emotion or grimace and the participant was asked to make a yes/no judgment regarding whether or not the face conveyed an emotion. The purpose of this block was to help the participants distinguish faces that conveyed emotional meaning and those that did not prior to the experiment. During the second practice block, an auditory prime of either type preceded the presentation of the face, after which the participant was required to make the same yes/no judgment. There were two optional breaks within each of the six blocks (after every 60 trials), and an optional break between blocks. Each block lasted approximately 20 minutes and the entire experiment lasted approximately 3-3.5 hours (including set-up and removal of the EEG cap.)

ERP Recording and Data Analysis

We used sixty-four active Ag/AgCl electrodes (actiCAP) mounted in an elastic cap to record the EEG. The EEG was digitized at 1024 Hz during recording. We then digitized the data sets offline to a sampling rate of 250 Hz, re-referenced them to the average of the electrodes, and applied a band pass filter between 0.1 Hz and 40 Hz. Bipolar horizontal and vertical EOGs were recorded in order to help with trial rejection. We then ran an independent component analysis (ICA) and rejected trials with EOG activity greater than 50 μV. Overall, approximately 30-40% of the trials for each participant were rejected using this method. We discarded all *no* trials (i.e. trials with grimaces as the target stimulus) from this analysis, as well as incorrect *yes* trials (i.e. trials with an angry, sad, or happy target emotion where the participant incorrectly responded *no*). We averaged the remaining trials from the onset of the facial stimulus to 1000 msec after its onset.

Using ANOVA, we separately analyzed accuracy (i.e. correct yes trials) and ERP data for each condition (angry congruent, angry incongruent, etc.). We also used seven Regions of Interest (ROIs) as group factors for the analysis. The regions were divided as follows: left frontal electrodes (F3, F5, F7, FC3, FC5, FT7); right frontal electrodes (F4, F6, F8, FC4, FC6, FT8); left central electrodes (C3, C5, T7, CP3, CP5, TP7); right central electrodes (C4, C6, T8, CP4, CP6, TP8); left parietal electrodes (P3, P5, P7, PO3, PO7, O1); right parietal electrodes (P4, P6, P8, PO4, PO8, O2); and midline electrodes (FZ, Cz, CPz, Pz, POz, Oz). We used a temporal window of 440 msec – 540 msec for the mean amplitude analysis of the N400. After visual inspection, we hypothesized that the temporal window of 150 msec – 250 msec was also of interest.

Results

A two by two by three repeated measures ANOVA was carried out separately for the behavioural data (accuracy) and the ERP data. In both cases, the factors were prime type (vocalizations, pseudoutterances), prime-target congruency (congruent, incongruent), and emotion (anger, sadness, happiness), respectively.

Behavioural Results

Analysis of accuracy rates revealed a significant main effect of prime type [$F(1,5) = 19.22$, $p = 0.007$]. There were overall differences in the ability to judge whether a facial expression represented an emotion based on the prime type, with significantly greater mean accuracy for vocalizations ($\bar{x} = 88\%$, $s = 0.088\%$) than for pseudoutterances ($\bar{x} = 84\%$, $s = 0.10\%$). No significant main effects for prime-target congruency were observed. Moreover, no other significant effects with regards to emotion category or congruency status of trials were observed.

ERP Results

Temporal window of 440 msec – 540 msec. The analysis for the temporal window 440 msec – 540 msec revealed a significant main effect of prime type [$F(1,35) = 4.31$, $p = 0.045$]. Overall, we observed more negative ERP amplitudes for pseudoutterances than for vocalizations as well as for congruent trials than for incongruent trials. We noted another main effect of prime-target congruency [$F(1,35) = 9.92$, $p = 0.003$].

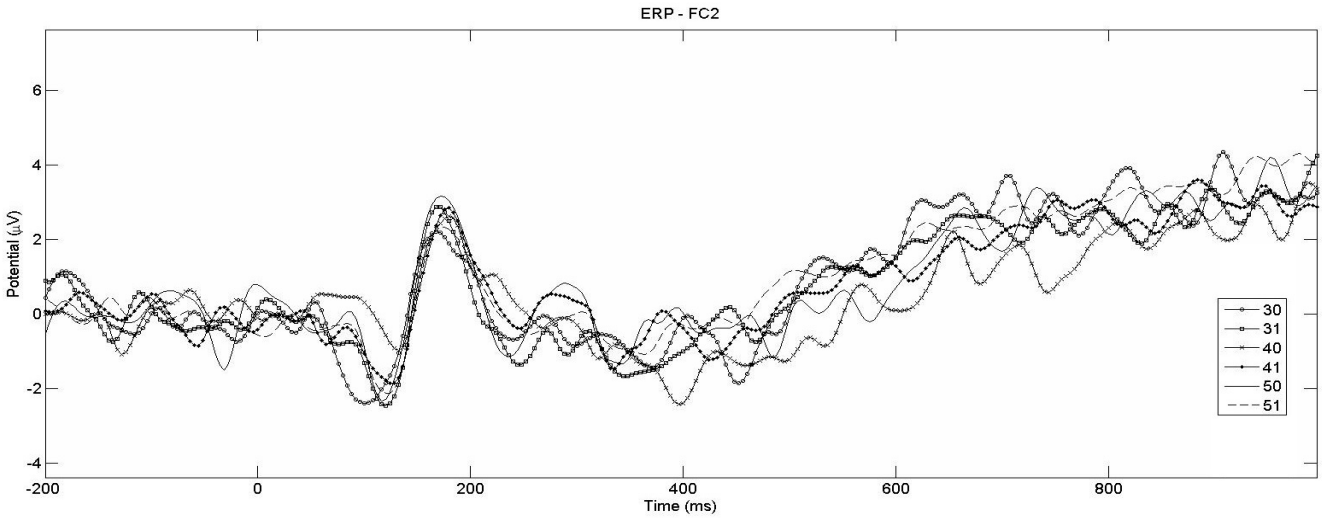

Fig. 2a Vocalizations – Averaged ERPs for FC2 electrode

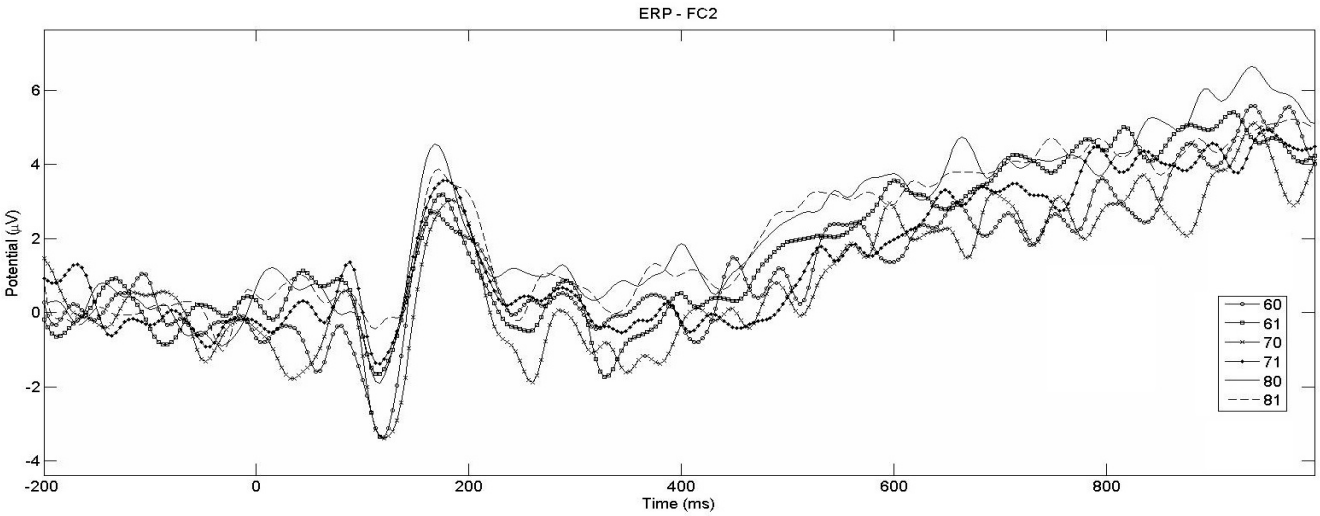

Fig. 2b Pseudoutterances – Averaged ERPs for FC2 electrode

Temporal window of 150 msec – 250 msec. The analysis for the temporal window 150 msec – 250 msec revealed a significant main effect of prime type [$F(1,35)$, $p = 0.027$]. Overall, we observed more negative ERP amplitudes for pseudoutterances than for vocalizations as well as for congruent trials than for incongruent trials. We also noted another main effect of prime-target congruency [$F(1,35) = 17.66$, $p < 0.001$]. Furthermore, a grouping factor – region of interest (ROI) – was introduced into the statistical analysis. We saw a significant interaction of prime type, prime-target congruency, and ROI [$F(6,35) = 2.60$, $p = 0.03$]. This indicated the possibility that ERP amplitudes were modulated as a function of both variables in combination in certain regions. Further analysis of specific ROIs revealed an interaction between prime type and prime-target congruency in left frontal electrodes [$F(1,5) = 15.56$, $p = 0.01$] and right frontal electrodes [$F(1,5) = 7.17$, $p = 0.04$]. In both regions, we observed significantly more negative ERP amplitudes in incongruent trials than in congruent trials for vocalizations. However, we observed significantly more negative ERP amplitudes in congruent trials than in incongruent trials for pseudoutterances. The effects in the left and right frontal regions are identical, so the FC2 electrode is used to illustrate them in Fig. 2.

Table 1 indicates the legend code to condition conversion for increased readability. A significant interaction between emotion and prime type was observed when ROI was used as a grouping factor [$F(12,70) = 1.98$, $p = 0.039$]. Further investigation showed that there was an interaction between emotion and prime type in left frontal electrodes [$F(2,10) = 4.30$, $p = 0.045$] and right frontal electrodes [$F(2,10) = 5.50$, $p = 0.024$]. In both regions, sadness was seen to elicit significantly more positive ERP amplitudes than anger and happiness for vocalizations, while it elicited significantly more negative ERP amplitudes than anger and happiness for pseudoutterances.

Code	Condition
30	Congruent Angry Vocalization
31	Incongruent Angry Vocalization
40	Congruent Sad Vocalization
41	Incongruent Sad Vocalization
50	Congruent Happy Vocalization
51	Incongruent Happy Vocalization
60	Congruent Angry Pseudoutterance
61	Incongruent Angry Pseudoutterance
70	Congruent Sad Pseudoutterance
71	Incongruent Sad Pseudoutterance
80	Congruent Happy Pseudoutterance
81	Incongruent Happy Pseudoutterance

Table 1
Code – Condition Conversion

Discussions

In the present study, we used analysis of ERP amplitudes to investigate the influence of auditory prime type on the judgment of emotional facial expressions. A previous study by Paulmann & Pell (9) investigated the effect of prime length and observed that medium length primes (400 msec or more) elicited a normal response in prime-target congruency in that incongruent trials elicited more negative ERP amplitudes than congruent trials. For this reason, this study made use of prime stimuli that were at least 400 msec in length. The main focus was to understand whether or not vocalizations and pseudoutterances convey emotional meaning to the same extent by comparing both the ERP amplitudes elicited by subsequent judgment of emotional facial expressions as well as behavioural results.

Implications of Behavioural Results

Behavioural results demonstrated that there was a significant difference in the accuracy of emotional facial expression judgment between the vocalizations and pseudoutterances as primes. As reported, vocalizations allowed for more accurate judgment overall – an average accuracy of 88% – than pseudoutterances – an average accuracy of 84%. It is speculated that a reason for this difference is the very nature of the facial expression stimuli. Fig. 1 provides an example of the typical facial stimuli used in the experiment; upon inspection, it becomes apparent that the facial expressions resemble expressions one might have in a real-life conversation after having conveyed an emotion in a fashion that closely resembles a vocalization. It is possible that the relation between the vocalization as a prime and the contextually similar facial expression allowed for more accurate responses by the participant.

Unconventional N400 Congruency Results

Studies have shown that prime-target incongruency results in more negative ERP amplitudes than congruent trials when primes of 400 msec are used (5, 10). As previously mentionedthe primes used in this study were at least 400 msec long, quite often reaching up to 2000 msec in length. We used long primes in hopes that the conventional N400 pattern of congruency would be observed. This seemed to be a reasonable deduction based on the fact that emotional speech tone has been seen to influence the early judgment of facial expressions (1, 2). However, just the opposite was observed for pseudoutterances; *congruent* trials elicited more negative ERP amplitudes than incongruent trials. This was unanticipated, and the reasons behind the phenomenon can only be speculated. It is well known that emotional information from an auditory prime is integrated early after hearing it (11). Perhaps excessive length of primes caused disintegration in the priming effect by the time the participant was required to make a judgment of the facial expression.

Another interesting effect observed was that pseudoutterances elicited greater negative ERP amplitudes than vocalizations. It is possible that vocalizations more effectively conveyed emotion, or similarly to the behavioural effect, provided emotional information that the participant found more relevant to the target faces. This would have allowed an association to be made more quickly between the prime and emotions expressed by the facial expressions when preceded by vocalizations. Absolute values for ERP amplitudes were not analyzed - however, it is likely that pseudoutterances elicited normal ERP amplitudes and that vocalizations simply elicited a relatively more positive response.

Negative ERP Amplitudes in the 150 msec – 250 msec Window

As with the 440 msec – 540 msec temporal window, the 150 msec-250 msec window showed greater negativity in trials with pseudoutterances as primes than for vocalizations. Moreover, the effect of congruency was also similar: there was greater negativity observed for congruent versus incongruent trials. This temporal window differed from the 440 msec – 540 msec window in that significant interactions were observed in the left and right frontal regions between type and congruency, and between emotion and type. The left frontal region is commonly associated with the processing of positive emotions while the right frontal region is commonly associated with the processing of negative emotions (12). This was not considered in this study; but based on the fact that the emotions of the stimuli were anger, sadness, and happiness, the observed ERPs on both sides of the brain are not surprising results. As for the particular interactions, the interaction between prime type and prime-target congruency showed greater negativity in incongruent trials than in congruent trials for vocalizations, but the opposite effect for pseudoutterances. As previously mentioned, the emotional meaning of an auditory stimulus is processed very early on. Differences in pattern between vocalizations and pseudoutterances – where vocalizations show the conventional greater negativity in incongruent trials – may be the result of the lengthier pseudoutterances favouring controlled processing. On the other hand, vocalizations tended to be less than 1000 msec in length, which may have favoured automatic processing and thus showed the expected pattern in ERP amplitudes. There was also interaction between the emotion and type of prime; greater negativity was seen for a sad prime in pseudoutterances than in vocalizations. This reiterates the idea that there may have been a weaker association between the pseudoutterance and facial expression simply due to the nature of the two prime types. Vocalizations may allow for a stronger association between the emotion of the prime and the facial expression seen.

Further Investigations

Future studies should seek to understand the ability of vocalizations to convey stronger emotional meaning than pseudoutterances as well as why full-duration pseudoutterances elicited greater negativities for congruent trials. Once these factors are understood, it would be interesting to conduct similar experiments on a greater range of emotions. This could be taken further by classifying them into positive and negative emotions in order to study further levels of interaction - for example whether positive vocalizations elicit greater or lower negativities than pseudoutterances.

Conclusions

These findings indicate that, from a behavioural standpoint, facial expressions preceded by vocalization primes are more accurately recognized than when preceded by pseudoutterance primes. This is likely due to the relations between facial expressions and vocalizations. As expected, vocalizations were shown to elicit a greater negativity for incongruent trials than for congruent trials. Interestingly, pseudoutterances caused a reversed effect. The conventional priming effect of pseudoutterances may have been absent in the averaged ERP data as a result of their long duration, as compared to previous studies. Future studies should consider capping the length of the pseudoutterances and vocalizations to see if the conventional N400 response is restored.

References

[1] G. Pourtois *et al. NeuroReport.* **11**, 1329-1333 (2000).
[2] B. De Gelder, J. Vroomen, *Cognition & Emotion.* **14**, 289-311 (2000).
[3] C. Brown, P. Hagoort, *Journal of Cognitive Neuroscience.* **5**, 34-44 (1993).
[4] M.D. Pell, *Brain & Cognition.* **48**, 499-504 (2002).
[5] M.D. Pell et al. *Journal of Phonetics.* **37**, 417-435 (2009).
[6] P. Belin, S. Fillion-Bilodeau, *Behavior Research Methods.* **40**, 531-539 (2008).
[7] D. Grandjean *et al. Nature Neuroscience.* **8**, 145-146 (2005).
[8] S. Paulmann, S.A. Kotz, *Brain & Language.* **105**, 59-69 (2008b).
[9] S. Paulmann, M.D. Pell, *Cognitive, Affective, & Behavioral Neuroscience.* **10**, 230-242 (2010).
[10] Q. Zhang, A. Lawson, C. Guo, Y. Jiang, *Brain Research Bulletin.* **71**, 316-323 (2006).
[11] S. Paulmann, S.A. Kotz, *NeuroReport.* **19**, 209-213 (2008a).
[12] R. Graham, R. Cabeza, *Cognitive Neuroscience and Neuropsychology.* **12**, 1-4 (2000).