On the effect of subliminal priming on subjective perception of images: a machine learning approach

Parmod Kumar[1], Faisal Mahmood[2], Dhanya Menoth Mohan[3], Ken Wong[4], Abhishek Agrawal[3], Mohamed Elgendi[5], Rohit Shukla[6], Justin Dauwels[3*], and Alice H.D. Chan[3*]

***Abstract*— The research presented in this article investigates the influence of subliminal prime words on peoples' judgment about images, through electroencephalograms (EEGs). In this cross domain priming paradigm, the participants are asked to rate how much they like the stimulus images, on a 7-point Likert scale, after being subliminally exposed to masked lexical prime words, with EEG recorded simultaneously. Statistical analysis tools are used to analyze the effect of priming on behavior, and machine learning techniques to infer the primes from EEGs. The experiment reveals strong effects of subliminal priming on the participants' explicit rating of images. The subjective judgment affected by the priming makes visible change in event-related potentials (ERPs); results show larger ERP amplitude for the negative primes compared with positive and neutral primes. In addition, Support Vector Machine (SVM) based classifiers are proposed to infer the prime types from the average ERPs, which yields a classification rate of 70%.**

I. INTRODUCTION

Priming is a nonconscious form of memory, which is concerned with perceptual identification of words and objects [1]. It refers to activating particular representation or association in memory just before carrying out an action or task. For example, if a person is exposed to the word 'lung-cancer', and later had a choice to smoke, there is a greater probability that the person choose not to smoke if the word is perceived negatively. The term 'subliminal' means 'below a limen' (sensory threshold). In other words, subliminal priming is not perceived by the conscious mind but has an influence on the unconscious mind.

In the following, we summarize prior conceptual frameworks relevant to priming. The Lexical Decision Task (LDT) is a widely used task in cognitive psychology, which has shown that lexical decision responses are made more quickly for semantic priming, i.e., when the target word is related to the previous word presented [2]. A similar study on the lexical priming [3] demonstrates the occurrence of N250 and N400 component at Pz channel. Holcomb and Grainger addressed similar effects in their masked repetition priming paradigm [4].

Recently, Event-Related Potential (ERP) based studies have received increased attention in assessing the effect of affective priming on rating emotional pictures. In an earlier study, pleasant and unpleasant pictures are associated with enlarged early posterior negativity (EPN) as well as late positive potential (LPP) amplitudes compared to neutral pictures [5]. The processing of emotional pictures influences the ERP amplitude at separate latencies, reportedly found to occur at short latencies of 100-300 ms [6], [7]. ERP studies with picture priming mismatch have shown that a late posterior component between 225 ms and 500 ms can differentiate between priming conditions [8].

In a previous cross-domain study [9], a similar task to LDT is employed, where the lexical prime is replaced with a masked picture. The results show similarity with the LDT based study. This suggests that semantic processing of information can be achieved through subliminal presentation of pictures to affect successive processing in the memory. In a subsequent study [10], masked lexical primes are presented before participants are asked to rate how much they like certain pictures on a scale of 1-4. This paradigm uses cross-domain priming (word primes combined with target images) and demonstrates that subliminal affective priming can affect decision making. A more positive late signal following positive primes is evoked at the right-hemisphere electrodes.

In the present study, a similar method as in [10] is devised aiming to examine the subliminal affective priming on image rating. The prime words used in this research exhibit positive, negative, and neutral connotations with respect to the pictures. We hypothesize that the prime will cause participants to respond differently to the pictures. We also expect to see a late ERP component, between 300-500 ms, in the posterior region, reflecting different priming conditions. In addition, this difference would be more pronounced in the right hemisphere than the left hemisphere. We also hypothesize that we would see the N400 effect at Pz reflecting the lexical prime effect.

The contributions of this paper are as follows. We will show that the present experimental results replicate the previous accounts of ERP effects in a similar paradigm, and

An abstract of this paper was submitted at the 35^{th} annual international conference of the IEEE Engineering in Medicine and Biology Society, Osaka, Japan, 2013.

* Corresponding author

[1]P.K., [2]F.M., [4]K.W., [5]M.E., and [6]R.S. were with the Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore 639798. P.K. is now with INRIA Sophia Antipolis−Méditerranée (e-mail:kumaparm@isu.edu). F.M. is now with the Okinawa Institute of Science and Technology, Okinawa, Japan 904-0495 (e-mail:faisal.mahmood@oist.jp). K.W. is now with Temasek Laboratories, National University of Singapore, 117411 Singapore (e-mail:kianwong@nus.edu.sg). M.E. is now with the University of Alberta Edmonton, AB, Canada (e-mail:moe.elgendi@gmail.com). R.S. is now with the University of Wisconsin-Madison, Madison, Wisconsin 53706 (email:rshukla3@wisc.edu).

D.M.M., A.A., and J.D. are with the Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore 639798 (e-mail: dhanya001@e.ntu.edu.sg, abhishekagrawal.bits@gmail.com, and jdauwels@ntu.edu.sg). A.H.D.C. is with the Nanyang Technological University, School of Humanities and Social Sciences, Singapore 639798 (e-mail:alice@ntu.edu.sg).

also support our hypothesis. We could see a strong interaction between the response scores and the prime types. It is also interesting to note that the ERPs are affected by the prime types at separate or overlapping latencies. It is possible to reveal what a person perceives when exposed to different subliminal affective stimuli by exploring the neural correlates of unconscious stimuli processing. The subjective mental state can be read out from a person's brain by analyzing the patterns of brain activity (ERP). We propose Support Vector Machines (SVMs) based pattern classifiers to brain read the participants' mental states (priming conditions). The proposed classifier utilizes ERP features in time, frequency, and time-frequency domains, to discriminate the three prime types. We demonstrate that the prime type can be inferred from the average ERP in a relatively reliable manner. We will investigate the performance and reliability of the proposed methods.

The remainder of this paper is organized as follows. In Section II, we describe the experimental setup and procedure. In Section III, we present the results for behavioral and ERP data and discuss our observations. Finally, we provide concluding remarks in Section IV.

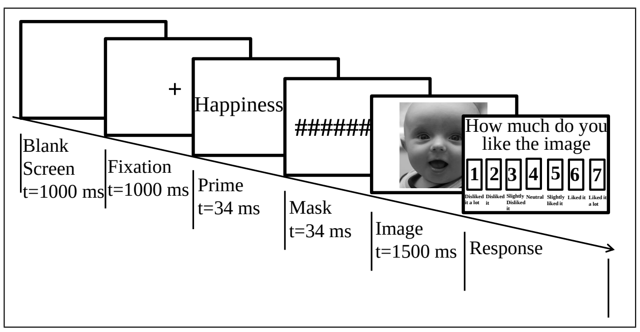

Fig. 1. Experimental sequence for a single trial consisting of blank screen, fixation mark, prime stimulus, mask, main stimulus and a response box.

II. METHOD

Here we describe the experimental setup, designed stimuli, and procedure.

A. Participants

Forty English speaking students (26 males and 14 females, in the age, ranging from 19 to 33, with a mean age of 22.3 years) participated in this study at Nanyang Technological University, Singapore. Thirty-nine participants were right handed, and one was left handed. All had normal or corrected-to-normal vision and all participants received a monetary remuneration for their participation.

B. Materials and Procedure

We use one fifty 'prime-word - image' pairs for the experiment based on the ratings obtained from the pilot studies. There were positive (juicy(prime word)-cherries(image)), negative (fungus-corn), and neutral (duck-jacket) 'prime word - image' pairs in equal number (50) (see Table I). Each participant performed 150 trials of the rating task

TABLE I

THREE TYPES OF 'PRIME-WORD - IMAGE' PAIRS USED IN THE EXPERIMENT ARE GIVEN FOR A REFERENCE.

'prime-word - image' pair		
Positive	Negative	Neutral
juicy - cherries	fungus - corn	duck - jacket
dazzling - bangles	bitter - cucumber	butter - curtain
appraised - certificate	inaccurate - hourglass	socks - camera
knowledge - bookcase	tsunami - earth	water - candy

with the designed sequence of events for each trial of the experiment (see Fig. 1). The start of each trial was triggered by presenting a blank screen for 1000 ms, followed by a fixation point, mark '+' at the center of the screen for 1000 ms. It helped to ensure that the participants were attending to the screen while showing the prime words. Each trial consisted of a presentation of prime-word using backward masking technique. The prime words were shown for a period of 34 ms followed by a mask "##########" for 34 ms. Then the associated pictures were presented for 1500 ms, and the participants were instructed to rate the extent to which they liked the image on a 7-point Likert scale, ranging from one (liked the least) to seven (liked the most). Simultaneously, the EEGs signal were recorded. The exposure time for prime and mask was determined from pilot studies and is in agreement with [11].

Visual stimuli were presented on an LCD monitor (Dell computer, resolution 800×600, refresh rate of 60 Hz, color depth of 16-bit) at a horizontal viewing distance of 60 cm. The EEG signal were collected at a sampling rate of 250 Hz by means of a 32-channel HydroCel GSN (HCGSN) sensor array by EGI, arranged according to 10-20 system [12].

Furthermore, we conduct a similar image rating experiment without any prime words prior to showing the target images. We conducted this experiment with 10 subjects, different from the subjects in the experiments with primes. The aim of this experiment is to measure and compare the participants' responses in primed and unprimed cases. This is to offer further evidences for our observations that the effects generated are purely due to the presence of primes.

III. RESULTS AND DISCUSSIONS

In this section, we provide results for (i) behavioral responses (Likert scores) analysis for the experiment with and without primes, and (ii) analysis of the ERP data for the experiment with primes. We discuss in detail the method to infer the primes from the average ERP data.

A. Behavioral Data Analysis

The behavioral data, i.e., responses on 7-point Likert scale for 40 subjects, are averaged across the trials with three conditions, corresponding to positive, negative, and neutral. These average response scores (dependent variables) are analyzed by means of paired-samples t-test and one-way repeated measure ANOVA test to examine the changes in the participants' responses over different priming conditions.

TABLE II

THE P-VALUES OBTAINED FROM PAIRED-SAMPLES T-TEST PERFORMED OVER THE AVERAGE RESPONSE SCORES CORRESPONDING TO POSITIVE-NEGATIVE (POS-NEG), POSITIVE-NEUTRAL (POS-NEU) AND NEGATIVE-NEUTRAL (NEG-NEU) PAIRS. * SIGNIFICANT AT P<0.05.

paired-samples t-test		
Prime pair	Experiment-with primes	Experiment-without primes
Pos-Neg	1.41E-09*	0.709
Pos-Neu	3.36E-12*	0.692
Neg-Neu	0.596	0.758

The test returns significant results (significant level chosen is p<0.05) for positive-negative (p=1.41E-09) and positive-neutral (p=3.36E-12) conditions, indicating a strong shift in the likeness judgment in the positively primed condition compared to the negative and neutral. The significance of negative primes on image rating is, however, not evident in the behavioral data (p=0.596 for Neg-Neu). By contrast, no such effects are observed in the behavioral data collected from the experiment conducted without primes (Experiment-without primes), i.e., p>0.05 (see Table II). The priming effect on behavior is further investigated and confirmed with the help of one-way repeated measure ANOVA with three conditions, corresponding to positive, negative, and neutral prime words (p=1.11E-16).

B. EEG Data Analysis

The recorded EEG data are filtered using a 0.1-30.0 Hz band pass filter and then referenced to the average of all the channels. Epochs for ERPs are collected at -1000 ms to 1500 ms around the image onset for each priming condition. The baseline is set to be -1000 ms to 0 ms. Independent Component Analysis (ICA) is applied to eliminate eye, muscle, and line noise artifacts. We analyze the most relevant segment of the ERP signal, i.e., from 0 ms to 500 ms after the stimulus onset.

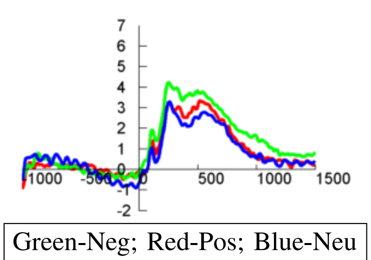

Fig. 2. Grand ERP average (average across all trials) at channel Pz for positive, negative, and neutral priming conditions.

1) A standard averaging approach: The grand ERP average at channel Pz reflects the difference in brain activity among the three priming conditions at an early and late latencies (see Fig. 2). We also notice the N400 effect at channel Pz indicating the effect of lexical priming. In addition, the average ERP signal reveals that the late posterior component between 225 ms and 500 ms can differentiate the prime types effectively in tasks with image-priming mismatch.

In order to compute the difference in activity among positive, negative, and neutral priming conditions, we average the ERP signals across the trials. Furthermore, a window based averaging technique with a window size of 25 ms is adopted. The average amplitudes (dependent variables) corresponding to different time windows are then assessed by means of one-way repeated measure ANOVA test with three conditions. The average ERP signals for three conditions show a significant difference among positive, negative, and neutral at different latencies (see Table III). The one-way repeated measures ANOVA test yields significant differences between positive-negative, positive-neutral, and negative-neutral conditions in occipital, lower temporal, and parietal lobes. The difference mainly is between positive-negative and positive-neutral conditions.

TABLE III

THE ONE-WAY REPEATED MEASURES ANOVA TEST RESULTS FOR THE AVERAGE ERPS WITH 25 MS ANALYSIS WINDOW THAT YIELD LOWEST P-VALUES.

	75-100 ms			400-425 ms		
Ch	P-Neg	P-Neu	N-Neu	P-Neg	P-Neu	N-Neu
O1	1.4E-05*	1.5E-04*	0.51	0.61	0.29	0.09
O2	4.2E-06*	3.9E-04*	0.45	0.97	0.62	0.58
T8	0.48	0.38	0.80	0.01*	0.01*	0.32
Pz	0.11	0.91	0.15	0.06	0.50	0.01*

2) Decoding ERPs: SVM as a classifier for different prime types:

a) SVM Formulation: We design SVM based binary class and multi-class classifiers to determine the prime types from the average ERP signal. SVMs are known to have good generalization performance, insensitivity to overtraining and to the curse-of-dimensionality. Moreover, a few hyper parameters are to be tuned, namely, the regularization constant C and the RBF width σ. The multi-class classification problem is formulated using One-vs-All strategy. We test SVM performance with a number of features from time, frequency, and time-frequency domains.

Let us consider a set of vectors of ERP features $\{\mathbf{x_i} \in \mathbb{R}^d\}$, and corresponding class label $\{\mathbf{y_i} \in \{1 =$positive, $2 =$negative, $3 =$neutral prime types$\}\}$ where i$= 1,...,N$. Each input $\mathbf{x_i}$ contains d input features. SVMs map the given input vectors from the input space to a high-dimensional feature space by a suitable kernel function, and try to build a discriminant hyperplane to separate the classes. We employ radial basis function (RBF) kernel, as it is highly effective in dealing with non-linear data. The optimal values of C ($C = 2e^{-1}$) and σ ($\sigma = 2^3$) are determined by cross validation.

We extract features such as the amplitudes of the ERP signal, power spectral density (PSD), relative band power, and discrete wavelet transform (DWT) coefficients, and employ linear local fisher discriminant analysis (LFDA) [13] to obtain relevant feature sets. We use leave-one-subject-out cross validation procedure to test the SVM classifier and measure the performance of the classifier using the confusion matrix. We express the classifier performance in terms of accuracy, which is calculated as the sum of correct classifications divided by the total number of classifications.

b) Feature Extraction: The acquired single trial ERPs are first averaged across the trials, and then across the time window of width 25 ms, with three priming conditions, to obtain the features in the time domain. In the frequency domain, we apply Fast Fourier Transform (FFT) on single trial ERPs to compute the power spectrum, and the relative power corresponding to different frequency bands. Furthermore, we average the relative power across the trials for different bands.

The Short-Time Fourier Transform (STFT) is applied to single-trial ERPs with a Hamming window of length 128 point with 50% overlapping. The FFT algorithm is then applied to each segment. The power spectral density (PSD) estimates of each segment corresponding to different frequency bands are extracted, and are used as an input to the classifier after performing dimensionality reduction.

DWT provides an optimal resolution in both the time and the frequency domains. The signal is decomposed by passing it through a half band high pass and low pass filters followed by subsampling. Different levels (2 to 7) of decomposition are applied on the average ERP signal. The approximate coefficients are then used to reconstruct the signal. It is observed that the reconstructed signal preserves the significant features of the raw ERPs up to level-3 decomposition. Hence, level-3 decomposition is chosen for further analyses. Several wavelet functions, such as Daubechies (db2, db4, and db8), Symlet (sym8), and Biorthogonal (Bior4.4) waves are tested to select the appropriate wavelet function. The approximate coefficients obtained from sub-band cA3 are used for classification.

TABLE IV

SVM CLASSIFIER PERFORMANCE FOR A COMBINATION OF FEATURES.

SVM	Accuracy (%)	Input Features
Pos-Neg	95.0	relative power at channel T7, DWT coefficients (sym8) at channel P4, amplitude at channel C4, and PSD values at channel Pz.
Pos-Neu	87.5	relative power at channel T8, DWT coefficients (db4) at channel P7, amplitude at channel T7, and PSD values at channel T7.
Neg-Neu	85.0	relative power at channel P4, DWT coefficients (sym8) at channel C4, amplitude at channel P4, and PSD values at channel T8.

c) Performance Evaluation: The features extracted from the ERP signal at different channels are fed to the LFDA model, and the resulting reduced sets of features are used for classification. Different combinations of features such as amplitude, relative power, DWT coefficients, and PSD values at parietal, central, and temporal lobes are submitted to the classifiers. The classifiers yield a performance rate of 95.0%, 87.5%, and 85.0% for Pos-Neg, Pos-Neu, and Neg-Neu binary classifiers respectively (see Table IV). In addition, the multi-class classifier trained on a combination of features such as amplitude at channel P7, relative power at channel O2, and DWT coefficients (db4) at channel P8 achieves a correct classification of 70% (accuracy). The multi-class SVM correctly classifies 80% of positive prime types, 72.5% of negative prime types, and only 57.5% of neutral prime types.

IV. CONCLUSIONS

This study investigates the effect of subliminal priming on subjects' responses to relatively natural and neutral images. Consistent with past research in subliminal priming, results of this research show a strong significant effect of priming on subjects' responses. We observed a significant shift in the likeness judgement towards the positively primed images and the effect of negative priming was also clearly visible in the ERP data collected. The proposed SVM based classifiers yielded a maximum accuracy of 95.0%, 87.5% and 85.0% for Pos-Neg, Pos-Neu, and Neg-Neu binary classifiers respectively. The results strongly suggest that ERPs encode different subliminal prime conditions. Significant correlation among several of the identified parameters points towards the possible practical use of ERPs to evaluate the effects of priming. Certainly more work need to be done in the future to look at the long term and short term effect of the priming on subjective and objective judgments of visual and audio stimuli.

REFERENCES

[1] Psychology Today. "What is Priming?". Internet: http/www.psychologytoday.com/basics/priming, April 4, 2014.

[2] S. Bentin, G. McCarthy, and C. C. Wood. "Event-related potentials, lexical decision and semantic priming". *Electroencephalography and clinical Neurophysiology*, Vol. 60(4), pp. 343-355, 1985.

[3] J. Morris, J. H. Porter, J. Grainger, and P. J. Holcomb. "Effects of lexical status and morphological complexity in masked priming: An ERP study". *Language and Congnitive Process*, Vol. 26 (4/5/6), pp. 558-599, 2011.

[4] J. Grainger and P. J. Holcomb. "Watching the Word Go by: On the Time-course of Component Processes in Visual Word Recognition". L*anguage and linguistics compass*, Vol. 3(1), pp. 128-156, 2009.

[5] H. T. Schupp, M. Junghöfer, A. I. Weike, and A. O. Hamm. "The selective processing of briefly presented affective pictures: An ERP analysis". *Psychophysiology*, Vol 41(3), pp. 441449, May 2004.

[6] J. K. Olofsson, S. Nordin, H. Sequeira, and J. Polich. "Affective picture processing: An integrative review of ERP findings". *Biological Psychology*, Vol. 77(3), pp. 247265, 2008.

[7] H. T. Schupp, B. N. Cuthbert, M. M. Bradley, J. T. Cacioppo, T. I. Ito, and P. J. Lang. "Affective picture processing: the late positive potential is modulated by motivational relevance". *Psychophysiology*, Vol. 37(2), pp. 257-261, 2000.

[8] W. B. McPherson and P. J. Holcomb. "An electrophysiological investigation of semantic priming with pictures of real objects". *Psychophysiology*, Vol. 36(1), pp. 53-65, 1999.

[9] R. Dell'Acqua and J. Grainger. "Unconscious semantic priming from pictures". *Cognition*, Vol. 73, pp. B1-B15, 1999.

[10] H. Gibbons. "Evaluative priming from subliminal emotional words: Insights from event-related potentials and individual differences related to anxiety". *Consciousness and Cognition*, Vol. 18, pp. 383-400, 2009.

[11] S. C. Draine and A. G. Greenwald. "Replicable unconscious semantic priming". *Journal of Experimental Psychology: General*, Vol. 127, pp. 286-303, 1998.

[12] H. H. Jasper. "The ten-twenty electrode system of the International Federation". *Electroencephalography and Clinical Neurophysiology*, Vol. 10, pp. 371-357, 1958.

[13] M. Sugiyama. "Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis". *Journal of Machine Learning Research*, Vol. 8, pp. 1027-1061, 2007.