2018 IEEE International Conference on Systems, Man, and Cybernetics

Online Detection of Real-World Faces in ECoG Signals

Christoph Kapeller
Guger Technologies OG
Graz, Austria
kapeller@gtec.at

Johannes Grünwald
Guger Technologies OG
Graz, Austria
gruenwald@gtec.at

Kyousuke Kamada
Department of Neurosurgery
Asahikawa Medical University
Asahikawa, Japan
kamady@asahikawa-med.ac.jp

Hiroshi Ogawa
Department of Neurosurgery
Asahikawa Medical University
Asahikawa, Japan
a040010@gmail.com

Shusei Fukuyama
Department of Neurosurgery
Asahikawa Medical University
Asahikawa, Japan
kyokui090208@gmail.com

Takahiro Sanada
Department of Neurosurgery
Asahikawa Medical University
Asahikawa, Japan
sanataka9103@gmail.com

Robert Prückl
Guger Technologies OG
Graz, Austria
prueckl@gtec.at

Christoph Guger
Guger Technologies OG
Graz, Austria
guger@gtec.at

***Abstract*—Previous neuroimaging studies have reported that the ventral temporal cortex (VTC) processes visual stimuli and thereby establish visual categories, which can be detected in electrophysiological signals such as electrocorticography (ECoG). However, most of the studies are based on visual stimulation through a computer. Thus, the degree to which those categories can be generalized is unclear under real-world conditions. This study extends the findings of a previous experiment, which aimed in real-time detection of visual perception, and investigated whether neural face and kanji categories obtained by computer stimuli can be confirmed in a real-world scenario. The real-time decoder accuracy and latency of two patients with epilepsy revealed that real-world faces and kanji can be detected with 79.9% and 28.4% accuracy, respectively, showing an average online detection latency of 447 ms with respect to presentation time. Hence, the VTC cortex elicits robust and similar responses to computer stimuli and real-world face, leading to a powerful brain-computer interface to track a person's attention in a real-world scenario.**

Keywords—BCI, ECoG, Broadband Gamma, Face, Kanji

I. Introduction

Online evaluation of cortical processes related to visual perception could contribute to more powerful and versatile human-computer interaction. Especially the recognition of faces is an innate and very well developed ability of the human brain [1] that can serve as field of research to help understanding attention or intention mechanisms. Particular regions on the ventral temporal cortex (VTC) were found that play an important role in processing information specifically related to faces [2], but also other categories like words [3]. Most notably, regions on the fusiform gyrus, including the so called fusiform face area (FFA) [4] and the anterior VTC [5], have been identified in functional magnetic resonance imaging (fMRI) studies. Electrophysiological markers related to face processing in electrocorticographic (ECoG) signals include face-specific evoked potentials [6] and broadband γ activity [7]. Several studies have presented decoding systems achieving discrimination performances of 90.4% for visual categories like presented faces and objects [8], and 96% for faces and houses [9]. However, most of these studies involved cued visual stimuli presented on a computer monitor with subsequent offline data processing. Recently, we presented an ECoG-based real-time face decoder that does not depend on artificial synchronization to stimulus presentation [10]. The system was calibrated with training data presented on a computer monitor and validated using unseen natural stimuli in a real-world environment. This successful experiment supports the hypothesis that the medium of the stimulus (i.e., picture on a screen vs. real face) is in fact irrelevant for face-related cortical processing. However, since [10] was only a single-subject study, we aim for more evidence by investigating further subjects to justify more general conclusions. To this end, here we extend and confirm the previous findings with another subject, whose implanted ECoG grids covered similar cortical locations on the VTC/FFA. The decoder was calibrated by pictures of faces and kanji-characters on a computer screen and validated by printed faces and kanji-characters as well as real-world faces. The decoder feedback was provided in real-time and without synchronization to stimulus presentation.

II. Methods

A. Subjects

Two patients with epilepsy (S1: 26y male aforementioned in [10]; S2: 22y male), undergoing surgical treatment at Asahikawa Medical University, volunteered to participate in this study. Each patient was temporarily implanted with subdural platinum electrodes to localize seizure foci prior to resective brain surgery. The electrodes over the ventral temporal cortex had a diameter of 1.5 mm with 5 mm spacing and are highlighted in Fig. 1. The study was approved by the institutional review board of Asahikawa Medical University. Both subjects gave informed consent prior to the experiment.

B. Data Acquisition

ECoG signals were recorded at the bedside with a DC-coupled g.HIamp biosignal amplifier (g.tec medical

European Union Eurostars project RAPIDMAPS 2020 (9273) and Horizon 2020 Graphene Flagship project Graphene Core1 (696656)), the Japanese government (No. 16H05434, JP15H01657 and 26670633).

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00030

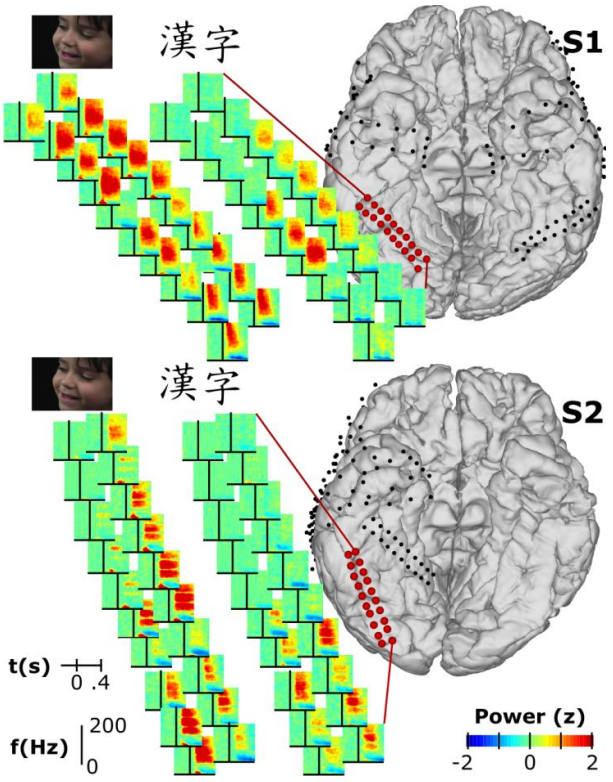

Fig. 1. Brain models with ECoG locations (black and red balls) of the two subjects after co-registration of pre-operative MRI and post-operative computerized tomography (CT) scans. ECoG locations highlighted by the red balls refer to the time-frequency plots, showing standardized ECoG responses in z-scores to presented faces and kanji images.

engineering GmbH, Austria) after neuro-monitoring and prior to resective surgery. Data were digitized with 24-bit resolution at 1,200 Hz, synchronized with stimulus presentation for the calibration runs, and stored using the *g.HIsys real-time processing library* (g.tec medical engineering GmbH, Austria).

C. Experimental Procedure

For calibration, subjects were asked to observe 120 stimuli of three types (40 stimuli each), namely, *Face* (colored and greys photos of faces), *Kanji* (images of kanji characters) and *Idle* (black screen) that were presented in randomized order with a presentation time of 400 ms each on the computer screen. After each stimulus a black screen was shown for a randomized duration between 2.0 s and 3.3 s. Each subject performed two calibration runs, leading to a total number of 240 trials for calibration. After that, the subjects participated in a real-world asynchronous experiment, including multiple printed kanji characters and photos of faces on paper, a mirror for the patients to see themselves, and a group of 2-4 persons to look at. After showing the printouts to the subjects and placing the mirror in front of them, the persons in the group successively appeared in front of the subject. A computer performed data processing and classification amongst Face, Kanji, and Idle in real time, and provided visual feedback in the form of a schematic face, a schematic kanji character, or a black screen. The whole experiment, including the feedback monitor, was recorded by a video camera at a rate of 30 frames per second for later synchronization with the ECoG data.

D. Decoder Calibration

The design of the decoder has been introduced in a previous visual categorization study [1]. A preceding offline calibration is required to enable real-time decoding according to Fig. 2. Thus, ECoG signals $\boldsymbol{x}[m]$ were initially high-pass (HP) filtered (Butterworth IIR, 4th order) to remove DC drifts for visual inspection. Then, a 110-140 Hz band-pass (BP) filter extracted broadband γ activity $\boldsymbol{x}_{filt}[m]$. Common spatial patterns (CSP) were computed from the filtered signal, improving the signal-to-noise ratio and reducing feature space dimensionality. Specifically, a set of three "one-versus-all" spatial filters was composed to create distinctive features for Face, Kanji and Idle. These filters were obtained within a window from 100 ms to 600 ms post-stimulus ECoG data. The four filters that contributed most to the discrimination task of each paired condition were used for classification. Hence, twelve feature channels remained for classification, each extracted by a corresponding spatial filter ($\boldsymbol{w}_{CSP,j}, j \in \{1,2,\dots,12\}$):

$$\boldsymbol{x}_{CSP,j}[m] = \boldsymbol{w}_{CSP,j}^{\mathrm{T}} \cdot \boldsymbol{x}_{filt}[m] \qquad (1)$$

Next, the variance $\boldsymbol{\sigma}_{\boldsymbol{x}_{CSP,j}}[n]$ was estimated from moving windows (500 ms length, 15 ms step size) for each feature channel $\boldsymbol{x}_{CSP,j}[m]$. These signals were log-transformed, yielding the normalized broadband γ power $\boldsymbol{y}_{CSP}[n]$. Finally,

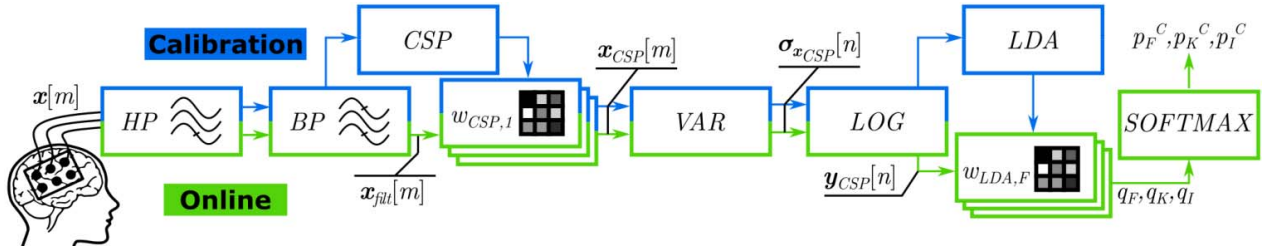

Fig. 2. Signal processing pipeline for calibration (blue boxes) and online decoding (green boxes). Calibration: ECoG signals $\boldsymbol{x}[m]$ were HP and BP filtered and submitted to a CSP analysis, computing a set of spatial filters ($\boldsymbol{w}_{CSP}$). Spatially filtered signals $\boldsymbol{x}_{CSP}[m]$ underwent variance estimation (VAR) and log-transformation (LOG), resulting in a normalized $\boldsymbol{y}_{CSP}[n]$. An LDA generated class-specific weights ($\boldsymbol{w}_{LDA}$). Online Decoding: Same steps as for calibration, but without CSP and LDA. The linear classifier ($\boldsymbol{w}_{LDA}$) weighted the features in $\boldsymbol{y}_{CSP}[n]$ and generated an LDA output score (q_F,q_K,q_I) for Face, Kanji and Idle. Finally, a Softmax function transformed the LDA output into complementary probabilities ($p_F{}^C$, $p_K{}^C$, $p_I{}^C$).

the three-class classification problem was solved via linear discriminant analysis (LDA), where each class was tested against the aggregated data of the remaining classes. This finally completed the calibration, i.e., the class-specific weights $\boldsymbol{w}_{LDA,i}$ with $i \in \{F, K, I\}$ denoting the respective class label.

E. Real-Time Decoding

For real-time decoding, the ECoG data were processed in frames of 16 samples, resulting in a processing rate of 75 Hz. In each processing step, data were HP- and BP-filtered according to Fig. 2, yielding $\boldsymbol{x}_{filt}[m]$. The calibrated spatial filters were applied according to Eq.(1) to get $\boldsymbol{x}_{CSP}[m]$ for subsequent variance estimation and log-transformation, yielding $\boldsymbol{y}_{CSP}[n]$. Then, $\boldsymbol{y}_{CSP}[n]$ were weighted by $\boldsymbol{w}_{LDA,i}$, resulting in the three LDA output scores q_i for Face, Kanji and Idle:

$$q_i = \boldsymbol{w}_{LDA,i}^{\mathrm{T}} \cdot \boldsymbol{y}_{CSP}[n] \quad (2)$$

Final output metric for decision making was the complementary probability that features belong to class i, which was derived from the LDA output by means of a Softmax function:

$$p_i{}^C = 1 - \frac{e^{q_i}}{\sum_{l \in \{F,K,I\}} e^{q_i}} \quad (3)$$

The decision criteria was then defined as follows: select the class that corresponds to the lowest $p_i{}^C$ if it undershoots the confidence threshold $p < 0.05$, otherwise set the output to Idle.

F. Decoding Performance Evaluation

As the decoder asynchronously outputs the result, it is not aligned to the stimulus presentation. In fact, it shows a delayed output with respect to the stimulus presentation. This delay originates, on the one hand, from the sliding window of 500 ms for variance estimation, and on the other hand from the natural visual processing time of the brain. This systematic offset was compensated by shifting the classifier output sequence, such that classification accuracy reached its maximum. Furthermore, the decoder performance was computed with and without subsampling of processed variance windows to compensate for unbalanced class occurrence (i.e., Idle occurred more often than Face and Kanji). Subsampling was performed 50 times, whereby each time the decoder accuracy was obtained from equally balanced number of samples of Face, Kanji and Idle. Additionally, the significance values of all decoding accuracies were obtained after 1000-fold bootstrapping of class labels.

III. Results

Decoding accuracy and latency for single-trial detection of recognized faces and kanji characters were assessed after calibration, where CSP and LDA weights were obtained from run 1 and tested by run 2. Additionally, the decoding accuracy was assessed for real-world faces and printed kanji characters in a real-world scenario with target persons around the subjects.

A. Decoder Validation

The classification output matched the presentation sequence best after shifting the classifier output back by 467 ms for S1 and 427 ms for S2. Spatial filters and classifier weights were obtained from the initial calibration run and then tested with the validation run. Table 1 and Table 2 show the decoder performance with and without subsampling, respectively.

TABLE I. Accuracy Without Subsampling

Subjects	Accuracy					
	Overall (%)	*Idle (%)*	*Face (%)*	*Kanji (%)*	*Random (%)*	*Sign. p<*
S1[a]	90.6	93.0	81.3	63.8	75.6	0.0005
S2[a]	92.7	92.9	98.3	83.1	73.6	0.0005

[a] Class occurrence for S1 and S2: 87.8% Idle, 6.1% Face and 6.1% Kanji

TABLE II. Accuracy With Subsampling

Subjects	Accuracy		
	Overall (%)	*Random (%)*	*Sign. p<*
S1	80.8	33.3	0.001
S2	92.4	33.3	0.001

B. Real-World Face Detection Performance

The classification output was first synchronized with the stimulus presentation in the video tapes and then shifted by the decoder latencies obtained in the decoder validation (i.e., 467 ms for S1 and 427 ms for S2). Table 3 and Table 4 show the decoder performance for real-world faces and printed kanji characters with and without subsampling, respectively. Time courses of the classifier outputs are shown in Fig. 3 together with the stimulus presentation times and exemplary photographs from the video tapes.

TABLE III. Accuracy Without Subsampling

Subjects	Accuracy					
	Overall (%)	*Idle (%)*	*Face (%)*	*Kanji (%)*	*Random (%)*	*Sign. p<*
S1[b]	82.8	84.7	72.4	52.9	69.7	0.0005
S2[c]	89.8	94.0	87.4	3.9	83.1	0.0005

[b] Class occurrence for S1: 87.3% Idle, 7.4% Face and 5.3% Kanji

[c] Class occurrence for S2: 91.2% Idle, 4.4% Face and 4.4% Kanji

TABLE IV. Accuracy With Subsampling

Subjects	Accuracy		
	Overall (%)	*Random (%)*	*Sign. p<*
S1	74.8	33.3	0.001
S2	61.3	33.3	0.001

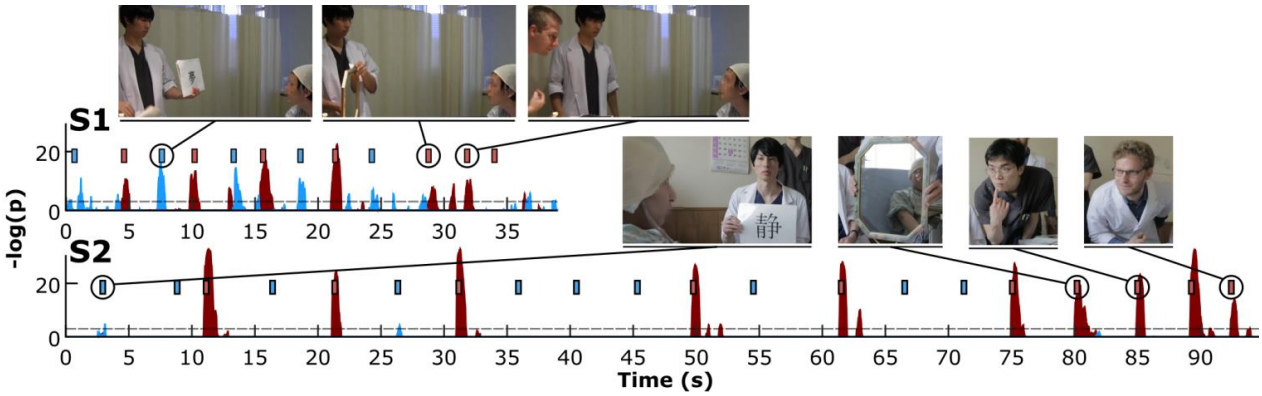

Fig. 3. Classification output –log(p) over time for Face (red) and Kanji (blue) in the real-world scenario. Small bars over the time courses indicate either Face (red bar) or Kanji (blue bar) stimuli. Dotted lines represent the significance border ($p<0.05$). Photographs show exemplary stimulus presentation events.

IV. Discussion

The VTC has been shown to be responsible for visual categorization, including face processing, mainly by neuroimaging studies, investigating electrophysiologic evoked or hemodynamic (i.e. fMRI) responses to visual stimuli on a presentation monitor. However, it is unclear whether visual categories established by population-level neural responses to computer stimuli can be generalized to real-world scenarios. The current study addressed this issue, showing that 72.4-87.4% of perceived real-world faces and 3.9-52.9% of printed kanji characters can be identified correctly. Hence, face detection was more accurate and turned out to be robust against paradigm changes. In fact, more electrodes on the VTC as shown in Fig. 1 responded to face stimuli than to kanji stimuli. Compared to the validation runs, the accuracy for the real-world faces only reduced from 89.8% (S1: 81.3%; S2: 98.3%) to 79.9% (S1: 72.4%; S2: 87.4%) on average. This indicates that neural activity during face processing can be reproduced even in complex scenarios and confirms the specific contribution to the visual processing on the ventral stream [1]. In comparison, the accuracy for detection of printed kanji dropped from 73.5% (S1: 63.8%; S2: 83.1%) to 28.4% (S1: 52.9%; S2: 3.9%). Notably, the decoder demonstrated a comparable performance for printed kanji in S1, whereas it was not able to confirm the decoding performance in S2. One explanation could be the time between calibration/validation runs and the real-world scenario. While all runs were recorded within one day for S1, seven days passed between calibration and real-world recording for S2. Hence, the CSP for Kanji may have altered and thus impaired the decoder. Especially, since the broadband γ activity was less distributed over the VTC in Fig. 1, the decoder is more sensitive to signal changes in individual ECoG channels and the activation topology in general. Another reason may be the bilateral electrode coverage in S1 compared to the right-sided coverage in S2, as the left fusiform gyrus is known to process visually presented words [3].

V. Conclusion

Spontaneous face perception can be robustly detected in real-time and works even across different types of stimuli, e.g., for untrained real-world faces. This can support future BCI applications with increased context awareness to track a person's attention.

References

[1] V.M., Reid, K. Dunn, R.J. Young, J. Amu, T. Donovan, and N. Reissland, "The human fetus preferentially engages with face-like visual stimuli," Current Biology, 27(12), 1825-1828, 2017.

[2] G. Schalk, C. Kapeller, C. Guger, H. Ogawa, S. Hiroshima, R. Lafer-Sousa, Z. M. Saygin, K. Kamada and N. Kanwisher, "Facephenes and rainbows: Causal evidence for functional and anatomical specificity of face and color processing in the human brain," Proc. Natl. Acad. Sci., p. 201713447, 2017.

[3] L. Cohen, S. Dehaene, L. Naccache, S. Lehéricy, G. Dehaene-Lambertz, M.A. Hénaff and F. Michel, "The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients," Brain J. Neurol., vol. 123 Pt2, pp. 291–307, 2000.

[4] N. Kanwisher, J. McDermott, and M.M. Chun, "The fusiform face area: a module in human extrastriate cortex specialized for face perception," J. Neurosci., vol. 17, 11, pp. 4302–4311, 1997.

[5] J.A. Collins and I.R. Olson, "Beyond the FFA: The role of the ventral anterior temporal lobes in face processing," Neuropsychologia, vol. 61, pp. 65–79, 2014.

[6] T. Allison, A. Puce, D.D. Spencer, and G. McCarthy, "Electrophysiological Studies of Human Face Perception. I: Potentials Generated in Occipitotemporal Cortex by Face and Non-face Stimuli," Cereb. Cortex, vol. 9, 5, pp. 415–430, 1999.

[7] A. D. Engell and G. McCarthy, "Face, eye, and body selective responses in fusiform gyrus and adjacent cortex: an intracranial EEG study," Front. Hum. Neurosci., vol. 8, 2014.

[8] E.M. Gerber, T. Golan, R.T. Knight, and L.Y. Deouell, "Persistent neural activity encoding real-time presence of visual stimuli decays along the ventral stream," bioRxiv, p. 088021, 2016.

[9] K.J. Miller, G. Schalk, D. Hermes, J.G. Ojemann, and R.P.N. Rao, "Spontaneous Decoding of the Timing and Content of Human Object Perception from Cortical Surface Recordings Reveals Complementary Information in the Event-Related Potential and Broadband Spectral Change," PLoS Comput Biol, vol. 12, 1, p. e1004660, 2016.

[10] C. Kapeller, H. Ogawa, G. Schalk, N. Kunii, W.G. Coon, J. Scharinger, C. Guger and K. Kamada, „Real-time detection and discrimination of visual perception using electrocorticographic signals," Journal of Neural Engineering, 15(3), 036001, 2018.