Unknown Word Detection based on Event-Related Brain Desynchronization Responses

Takafumi Sasakura, Sakriani Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura

Abstract The appearance of unknown words often disturbs communication. Most work on unknown words in spoken dialog systems deals with words that are uttered by the user, but which are not covered by the system's vocabulary. In this paper, we focus on detecting unknown words from the user side, in the case where the system utterance is unknown to the user. In particular, we develop a classifier based on Electroencephalography (EEG) signal from the user's brain waves, including the use of absolute power and Event-Related Desynchronization (ERD) features. The results show that we could detect the characteristics of brain waves at the time of unknown word perception significantly better than the chance rate.

1 Introduction

Skilled human communicators often adapt their language to suit the domain expertise of dialog partners. For example, a doctor will use technical medical terms when speaking to other medical professionals, and simpler terms when speaking to patients. Therefore, it is desirable to develop dialog systems that can adapt to user conditions in a similar way. The challenge is to provide the ability to detect miscommunication and dynamically generate user-adaptive utterance variations.

One of major problems that causes miscommunication is the appearance of the unknown words. Various techniques in the spoken language understanding component have been proposed in order to detect and handle user words that are not covered by the system lexicon [11, 2]. On the other hand, methods to detect and handle the system utterances that might be unknown to the user have not been widely explored. Our study focuses on the latter issue. The aim is to detect when the user does not know one of the terms output by the system. Through this, the system may be able to estimate the knowledge level of the user, and therefore have the capability to

T. Sasakura, S. Sakti, G. Neubig, T. Toda, S. Nakamura,
Nara Institute of Science and Technology (Japan),
e-mail: {sasakura.takafumi.sl4, ssakti, neubig, tomoki, s-nakamura}@is.naist.jp

adapt and express the content in words that match the user's vocabulary (i.e., use a known synonym or describe the unknown word in other words).

However, having awareness of the user's state and detecting the user's (lexical) domain knowledge is not straightforward. One could imagine several ways to do so, including extracting paralinguistic information such as gaze and face expressions or performing explicit confirmations in which the user is queried about their understanding. However, these paralinguistic signals are not guaranteed to occur every time an unknown word occurs, and explicit confirmation is burdensome for the user. In this work we take a different approach, looking directly into the user's mind through electrophysiological measurements of brain waves. Specifically, we present a new way of detecting user misunderstanding in the form of unknown words based on the user's Electroencephalography (EEG) signal.

2 EEG Event-Related Desynchronization

EEG is an electrophysiological measurement of the brain activity at the human scalp surface whereby voltage variations of cortical field potentials are imaged [7]. It records electrical signals generated by the brain through electrodes placed on different points on the scalp, and measures by comparing the voltage between two or more different sites. With regards to dialogue, Seshadrwe et al. [10] presented NeuroDialog, which uses an EEG based predictive model to detect system misrecognitions during live interaction. In this work, instead of system misunderstanding, we focus on detecting user misunderstanding in the form of unknown words.

When an unknown word is perceived, it is assumed that there is the matching process between the word and the memory. Sederberg et al. [9] found that during memory encoding of later recalled nouns, power of the specific frequency band was significantly higher than for not recalled nouns. Sauseng et al. [8] interpreted the result as memory matching between incoming visual information and stored (top-down) information. Klimesch [3] presented EEG oscillations in the alpha and theta bands that reflect cognitive and memory performance in particular. In this study, we classify EEG data using the EEG state in the time of the perception of the unknown word, and the change related to the event, which is called Event-Related Desynchronization (ERD) [6]. The ERD value is expressed as the ratio of the decrease in band power of the target epoch (P_t) as compared to a reference interval (P_r), which is selected by experimenter before the target epoch, by using the simple equation:

$$ERD = \frac{P_r - P_t}{P_r}. \tag{1}$$

We extract ERD values as features from the EEG data. The mean of each feature is normalized to 0 and the standard deviation to 1. To improve performance, features are selected with the parameter subset selection forward algorithm [1], which is shown below.

1. Initialize each subset to consist of one feature.
2. Calculate the score J for each subset. J given by

$$J = \frac{S_B}{S_W} \tag{2}$$

where

$$S_B = \sum_{j}^{L} N_j (m_j - M)^t (m_j - M) \tag{3}$$

$$S_W = \sum_{j}^{L} \sum_{i}^{N_j} (x_i - m_j)^t (x_i - m_j) \tag{4}$$

L is the number of classes, N is the number of subsets and N_j is the number of subsets of class j. M is the means of all subset vectors, m_j is the mean of the subset vectors of class j and x_i is a subset vector.

3. Select the features of the subset which had the highest score in the 2nd step.
4. Add a feature not included in the subset selected in 3rd step to the subset.
5. Repeat from 2nd to 4th step until the maximum score falls.

As a baseline, we also test a system that uses only the power of each frequency band as features.

3 Experimental Set-Up

3.1 Subjects and Stimuli Procedure

Six male Japanese-speaking subjects (23-24 years old in average) participated in the experiment. All participants were right-handed and had normal or corrected to normal vision. They don't have history of psychiatric or neurological illness or alcohol abuse, as well as no history of visual deficit. However, due to the effect of unnecessary components such as muscle artifacts, only five out of six EEG data were analyzed.

300 Japanese noun words of 4 mora were presented to the subjects as visual stimuli. A mora is a Japanese subsyllabic unit which provides the root of rhythm [5]. These words were constructed from Familiarity-controlled Word-lists 2007 Corpus (FW07) [4]. Originally, the list consists of words with four levels based on familiarity. In this study, we only used 150 words each from familiarity levels 1 and 4, which are the maximum and the minimum familiarity levels.

All subjects sat in a comfortable chair in a dark soundproofed room. The visual stimuli were presented on a 27 inch TV screen located 120 cm in front of them. The 300 words were presented visually at the center of the TV screen in white letters on a black background.

In this experiment, the subjects' task was to read a visually presented word and to answer a question about that stimulus. Fig. 1 illustrate the stimuli procedure, which consists of following steps: (1) 4000 ms of meaningless stimulus; (2) 1500 ms of a plus mark "+" as a warning signal; (3) 1500 ms of another meaningless stimulus; (4) 1500 ms of a stimulus word (in Katakana) which chosen at random; (5) Subjective evaluation, where the subject gives a mark whether a shown word was unknown

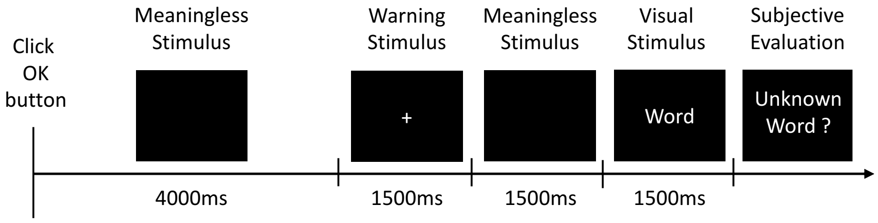

Fig. 1 An outline of the procedure of the experiment. When the values of ERD were calculated, the reference was during when the warning stimulus was presented, the target was during when the visual stimulus was presented.

or known/overlooked. After the subjects completed the evaluation, they may click the OK button to start the next trial. These trials were repeated 300 times for all subjects.

3.2 EEG Recording

We recorded EEG from 29 sites on the scalp using a BrainAmp made by the Brain Product company. The grounding electrode was placed on both earlobes, and the reference electrode to the apex of nose. To improve the signal to noise ratio, the impedance of each electrode was reduced to less than 5 $k\Omega$ using exclusive paste. EEG data was recorded with a sampling frequency of 1,000 Hz.

We then cut the high frequency components such as muscle artifacts using a low-pass filter less than 40Hz. Furthermore, trials including an amplitude more than 80 μV are excluded from the analysis. These procedures are called artifact reduction.

We extracted the target data, using EEG signals from successive 256 ms (256 points) time segments (windows or epochs) with 50% overlap for 1024 ms from the time when visual stimuli were shown. A Hamming window was applied to each time segment to attenuate the leakage effect. Power density of the spectral components was then calculated based on a fast Fourier transformation (FFT). Furthermore, to calculate the power change (ERD), the same processing was carried out for the EEG data starting 1024 ms from when the warning stimuli were presented. Using these values, the ERD value was calculated.

3.3 SVM Classifier

The data was classified into two classes labeled according to the subjective evaluation results and the data of two classes were balanced. The feature used for classifier are selected from the power or the ERD value per each frequency band for seven time windows of each 29 channels. Because the EEG characteristics vary among individuals, we chose to make a separate classifier for each individual. Af-

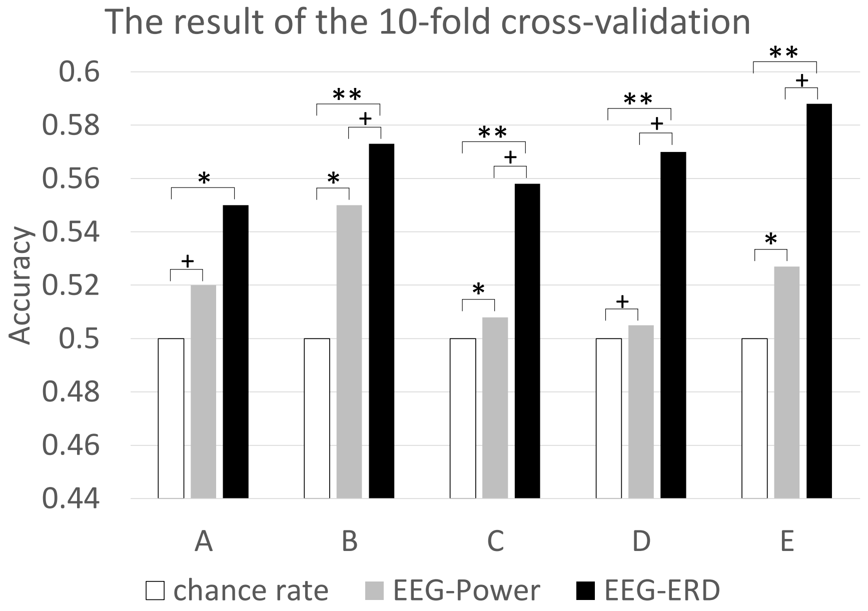

Fig. 2 Difference of the accuracy among the three kinds of features for the 5 subjects A-E. The bars marked $*$ have a significant difference compared with the accuracy of the chance rate, the bar marked $^+$ has a marginally significant difference. ($**p < 0.01$, $*p < 0.05$, $^+p < 0.10$, binomial test)

ter performing feature extraction for each subject, the selected features were used to train a classifier to distinguish between known or unknown words. Support Vector Machines (SVM) with the Radial Basis Function (RBF) Kernel were used for classification.

4 Results and Discussion

Fig. 2 shows the means of the accuracy of 10-fold cross-validation of 5 subjects. As a baseline, we use the chance rate. First, we apply a classifier simply on the absolute power of each frequency of EEG signals as features (denoted as “EEG-Power”). Second, we apply a classifier on ERD features (denoted as “EEG-ERD”). The results show that, EEG-Power only provided a significant difference for 3 subjects. In contrast, the accuracy of classification using ERD (“EEG-ERD”) was much higher than the chance rate in all subjects. Compared with EEG-Power, EEG-ERD provides a marginally significant difference over 4 subjects.

According to this result, we can see that it is important to capture the differential from the background signal. Because the accuracy of EEG-ERD is higher than EEG-Power in all subjects and there is marginally significant difference for most subjects, it is clear that the absolute value of power is not enough for prediction. Therefore, the user of differential features as in "EEG-ERD" has provides a better solution.

5 Conclusion

In this study, we detected unknown words from EEG when a subject perceived a word visually. As a result, both EEG-based classifiers (EEG-Power and EEG-ERD) showed a better performance than the chance rate. The best performance was obtained by classifier using ERD features (EEG-ERD).

Future work includes improvement of the performance of the classifier, experiments in an environment like a real conversation, and application to a multi-modal dialog system.

Acknowledgements Part of this work was supported by the Commissioned Research of National Institute of Information and Communications Technology (NICT) Japan, Microsoft CORE 10 Project, and JSPS KAKENHI Grant Number 26870371.

References

1. N. Hiruma, K. Sagara, Y. Tanaka, H. Takeichi, O. Yamashita, R. Hasegawa, T. Okabe, and T. Maeda. Brain communication : Theory and application. *IEICE Society Conference*, Vol. 94, No. 10, p. 926, 2011.
2. A. Kai, Y. Hirose, and S. Nakagawa. Dealing with out-of-vocabulary words and speech disfluencies in an n-gram based speech understanding system. In *ICSLP*, Vol. 2, pp. II–21, 1998.
3. W. Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, Vol. 29, No. 2, pp. 169–195, 1999.
4. T. Kondo, S. Amano, S. Sakamoto, and Y. Suzuki. Development of familiarity-controled word-lists (fw07). *IEICE Society Conference research report*, Vol. 107, No. 432, pp. 43–48, 2008.
5. T. Otake, G. Hatano, A. Cutler, and J. Mehler. Mora or syllable? speech segmentation in japanese. *Journal of Memory and Language*, Vol. 32, No. 2, pp. 258–278, 1993.
6. G. Pfurtscheller and A. Aranibar. Event-related cortical desynchronization detected by power measurements of scalp {EEG}. *Electroencephalography and Clinical Neurophysiology*, Vol. 42, No. 6, pp. 817 – 826, 1977.
7. S. Regel. *The comprehension of figurative language: electrophysiological evidence on the processing of irony*. PhD thesis, Universitätsbibliothek, 2009.
8. P. Sauseng, W. Klimesch, W. R. Gruber, and N. Birbaumer. Cross-frequency phase synchronization: a brain mechanism of memory matching and attention. *Neuroimage*, Vol. 40, No. 1, pp. 308–317, 2008.
9. P. B. Sederberg, M. J. Kahana, M. W. Howard, E. J. Donner, and J. R. Madsen. Theta and gamma oscillations during encoding predict subsequent recall. *The Journal of Neuroscience*, Vol. 23, No. 34, pp. 10809–10814, 2003.
10. S. Sridharan, Y.-N. Chen, K.-M. Chang, and A. I. Rudnicky. Neurodialog: an eeg-enabled spoken dialog interface. In *ICMI*, pp. 65–66. ACM, 2012.
11. S. R. Young. Detecting misrecognitions and out-of-vocabulary words. In *ICASSP*, Vol. 2, pp. II–21. IEEE, 1994.