10.3217/978-3-99161-014-4-092

WAVELET PACKET DECOMPOSITION TO EXTRACT FREQUENCY FEATURES FROM SPEECH IMAGERY

A. Tates[1], A. Matran-Fernandez[1], S. Halder[1], I. Daly[1]

[1]Brain-Computer Interfaces and Neural Engineering Lab, School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

E-mail: at18157@essex.ac.uk

ABSTRACT: Speech Imagery (SI) is considered an intuitive paradigm for Brain-Computer Interface designs in particular for communication applications. In this work, we use Electroencephalography (EEG) for offline SI decoding. We recorded covert speech from 17 participants. We tested two types of wavelet decomposition techniques. Specifically, we considered coefficients from 6 decomposition levels with Discrete Wavelet Transform (DWT) and multiple 2 Hz spaced packets with Wavelet Packet Decomposition (WPD), we computed different statistical features from such coefficients to form vector inputs for our binary-class classification approach. We approached the issue of feature/sample gap by using the Maximum Relevance and Minimum Redundancy (MRMR) feature selector algorithm to select the most informative features. We achieved a mean accuracy of 76.6% ± 16 and demonstrated the potential of WPD to extract narrow-band features, and how its refined representation outperforms DWT in SI decoding.

INTRODUCTION

Speech Imagery (SI) has become an attractive paradigm due to its intuitiveness [1, 2]. The Brain-Computer Interface (BCI) user is prompted to covertly say or repeat a speech unit (e.g., a letter, word, or phrase). With accurate classification of such tasks, a user can convey different messages or commands, e.g., to change an application state. One potential application of SI-based BCIs is as an assistive technology to restore communication for people who have lost the ability to speak. Researchers have approached SI-based BCI designs using different speech units as vowels [3, 4], syllables [5, 6] or words [7, 8] and were able to achieve higher than chance decoding accuracies suggesting the potential use of this paradigm.

To classify the speech unit from recorded EEG signal, informative features need to be extracted, EEG dynamics are known for their non-stationarity therefore a need for techniques that capture time and frequency domain information[9].

The widely known Fast Fourier Transform (FFT) has been applied to extract SI frequency information, Bajestani et.al (2022) [10] used FFT coefficients to classify between tasks with higher than-chance accuracy. Modified forms of FFT have also shown promising results when extracting SI features, the Discrete Gabor transform was applied by Jahangiri et.al (2018) [6] where the coefficients helped identify the relevance of the gamma band (> 60 Hz). Mel Frequency Cepstral Coefficients initially used for audio decomposition were used as EEG features, and showed classifiable properties between SI tasks [11, 12]. These FFT-based methods represent well-frequency information but omit time domain features which may also be important for SI decoding.

Wavelet Decomposition is a method proven useful in extracting both, time and frequency domain features [13, 14], in particular, Discrete Wavelet Transform (DWT) has been used as a feature extraction technique in SI approaches [15, 16]. DWT decomposes the signal with a transformation analogous to high and low-pass filtering. However, it may not be optimal for accessing specific frequency ranges as the obtained decomposition levels are derived from the low-pass filtered version of the scaled signal [17]. Additionally, Wavelet Packet Decomposition (WPD) performs a more detailed representation as the decomposition levels derive from the low and high-pass filtered versions of the signal resulting in a representation with more frequency ranges to access [18].

The issue of participant-dependent frequency variability is known in the area of EEG decoding, as the prominent frequencies elicited from imagery tasks tend to change for individuals, the appropriate selection of frequency information would lead to better classification results [19], thus we investigated the use of WPD to find participant-specific frequency ranges and compare its performance with features from fixed frequency ranges from DWT.

Due to the relatively small number of SI samples in comparison with the large number of features obtained from the wavelet decomposition levels, a dimensionality reduction step is needed to select a reduced number of features for optimal performance of a machine learning classifier. We investigated the capabilities of the Maximum Relevance Minimum Redundancy (MRMR) feature selection algorithm as it has proven useful for selecting informative features from large feature sets [20].

We have chosen the two phonetically distinct monosyllabic words ‘left’ and ‘right’ for our SI experiments and emphasized the participants to focus on the inner pronunciation rather than its meaning.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-092

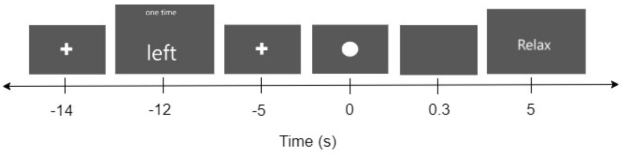

Figure 1: Timeline of the experimental protocol.

MATERIALS AND METHODS

Participants: Seventeen right-handed able-bodied participants (9 female) between the ages of 20 and 35 ($\mu = 25.65, \sigma = 8.3$) were recruited from the student population of the University of Essex. Participants received a compensation voucher worth £10 (GBP) for their time. All volunteers read, understood and signed the consent form based on the recommendations of the Ethical Committee of the University of Essex in January 2023 (Reference Number ETH2223-0220).

EEG Instrumentation: EEG was recorded using a 64-channel Biosemi Active-Two system. Electrode placement was done via the international 10-20 system, plus one electrode close to the pterion after each eyebrow for electrooculography (EOG) and one electrode behind each ear on the mastoids for electromyography (EMG) recording. Data was recorded at a sampling rate of 2048 Hz unaffected by hardware cut-off.

Experimental Protocol: Participants were seated in a comfortable chair facing a 52-inch screen. A graphical user interface developed with PsychToolbox 9.0 [21] in Matlab R2022 was used to display the prompts over a plain grey screen. We used a stimulus masking approach where we first showed the imagery prompt and then had a visual cue presented as a circle in the middle of the screen that remained for 300 ms, having a flash-like effect. See Figure 1 for the timeline of the experiment.

Participants were asked to perform the speech imagery of the words 'left' and 'right' as soon as they saw the cue stimulus.

We first presented a fixation cross for 2 seconds followed by the imagery prompt for 6 seconds and a time-variant (1.5–2 s) fixation cross before the cue. We cued our participants with the described flash stimulus and proceeded to leave a plain screen for 5 seconds until the 'relax' prompt was shown.

Signal Analysis: Raw EEG data were first downsampled to 1024 Hz from the original 2048 Hz, we then applied a notch filter (zero-phase, Hamming window FIR) at a cutoff frequency of 50 Hz and its harmonics at 100 Hz and 150 Hz to reduce the power line noise. We divided the data into regularly spaced epochs from $t_1 = -2$ s to $t_2 = 8$ s with respect to stimulus onset ($t = 0$), 25 trials per class were initially recorded. Channels were visually inspected and rejected when they looked overly noisy with respect to their neighbours. Epochs were visually inspected to reject those with bad movement artifacts. Between 4 to 7 epochs and 6 to 10 channels were dropped for each participant. Common Average Referencing (CAR) was then applied after to improve the signal-to-noise ratio.

In order to remove EOG and EMG artifacts, the signal components were estimated using Independent Component Analysis (ICA) with the Picard algorithm [22] to select and discard components encompassing evident eye blinks, lateral ocular movements or muscular artifacts based on their spatial or temporal locations and frequency distributions. Between one and four components were removed for each participant. The remaining components were used to reconstruct the data.

Feature Extraction: We used a 1.2-second-long post-stimulus window and tested the signal decomposition algorithms WPD, and DWT. We used Daubechies wavelet (db4) as the mother wavelet as it has been widely used for EEG approaches [8, 23].

DWT is known for its ability to represent time and frequency information [14, 24], it assumes that a signal is a linear combination of a particular set of wavelet functions, and these functions are scaled and shifted versions of a mother wavelet [17] WPD is a more refined version of wavelet decomposition which solves the scaling limitation of DWT as the decomposition happens on both detail and approximation coefficients at each level generating a larger frequency space, thus for the 6th WPD decomposition level, 64 packets of coefficients would be obtained[18].

To decompose the signal we first applied a low-pass filter (zero-phase, Hamming window FIR) at 128 Hz cut-off. For DWT we considered 6 levels of detail coefficients, D1(64–128 Hz), D2(32–64 Hz), D3(16–32 Hz), D4 (8–16 Hz), D5 (4–8 Hz), D6 (2–4 Hz) and one of approximation coefficients A6 (0–2 Hz). For WPD, we considered the 2 Hz step packets at the 6th decomposition level. These packets encompassed frequencies from 4–30 Hz and 70–128 Hz. We selected Alpha and Beta as previous approaches found informative features in such bands [10, 11, 25, 26] and also accounted for frequencies higher than 70 to explore the gamma band, also known to be relevant in SI-related activity [6, 27]. We did not consider frequencies between 30–70 Hz for WPD to narrow the number of options to select from therefore reducing computation cost.

We computed the next statistical and wavelet features from each level/packet: mean value, standard deviation, root mean square, slope, kurtosis, energy, entropy, mean absolute difference, negative turnings, positive turnings and wave centroid.

To find the most informative features we tested the classification performance of each statistical feature from each level/packet on a one-to-one basis, then combined the features with the top 3 classification accuracies to check for performance improvement.

Feature Selection: Each feature-level/packet combination formed a feature vector of shape channelsx1, as the average number of epochs per class was 21 ±4, we aimed for an ideal features vector shape of 10x1. To reduce the vector dimensionality we used the Minimum Redundancy Maximum Relevance (MRMR) method on every run of our cross-validation procedure. MRMR aims to maximize the relevance of features to the target variable while minimizing the redundancy among selected features [20], it uses a relevance score based on mutual information and a redundancy score based on Pearson correlation.

Classification: For each participant, we had an average of 21 trials ±4 per class. We evaluated the 2-class classification performance of our model with the median accuracy from a 6-fold cross-validation. We repeated the cross-validation 15 times, with a different seed at the time, and used the median score of each repetition to better estimate the model's performance.

Linear discriminant analysis has been widely used in BCIs. As large dimensionalities and overfitting are common problems in BCI, regularized LDA has been found to be useful for small training sample settings, we used the shrunken version of LDA [28], which adds a penalty term to the loss function, using the scikit-learn library [29] and the 'auto' shrinkage parameter that finds an optimal value based on the lowest error.

RESULTS

To get the most informative frequency ranges from the SI-EEG data, we recursively tested 11 statistical features computed from WPD packet and DWT level coefficients, we found the best-performing setting for each decomposition modality and participant based on the classification accuracy. We then reported the obtained accuracies and compared the results as seen in Figure 2. The use of features from multiple narrowed frequency intervals with WPD achieved 13% higher accuracy than the limited levels of decomposition from DWT, with ($p < 0.01$) from a two-sample test. WPD scores are above the 99% confidence interval, computed based on the trial number per class [30], marked for the black horizontal lines, while most of DWT results lay below this interval.

The MRMR algorithm for feature selection was shown to be useful in reducing dimensionality while retaining informative features. We have counted the occurrence of selection of each WPD packet over cross-validation folds and present them as a channel-feature heatmap in Figure 3. We observe that some relevant channels involve locations that may be reflecting speech processing-related areas as left-central channels C1, C5, frontal-temporal channel FT7 or temporal channel T7 one of the most selected by our feature selection process. However, informative features spread across different regions, as with Fp2 chosen along different frequencies or P8 highly relevant on the frequency band (26–28 Hz). Even if neural dynamics are considered to be produced by left hemisphere dominant processes [31–33], SI-relevant features from EEG appeared to spread around different regions depending on individuals.

Similar to frequency domain features, relevant information seems to be spread along all the tested bands above 10 Hz with particular highlights on bands at 26–28 Hz, 78–80 Hz and 122–124 Hz.

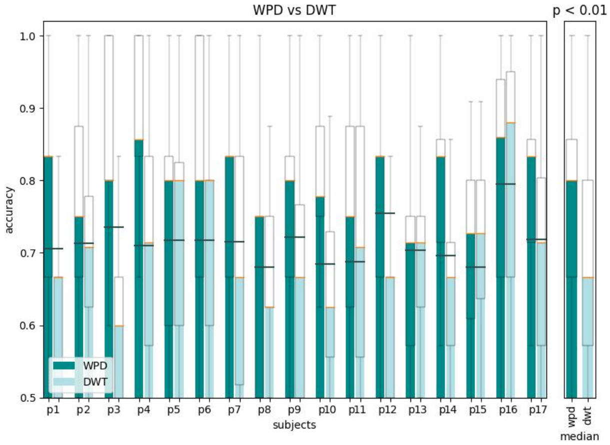

Figure 2: Comparison of obtained accuracies between the DWT and WPD decomposition methods across repeated 6-fold cross-validation, black horizontal lines represent the 99% confidence interval.

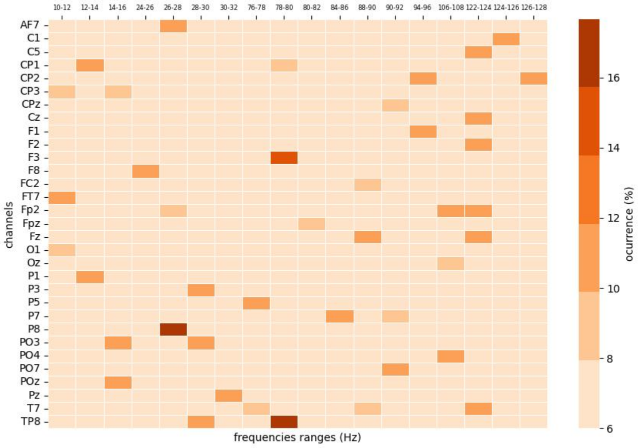

Figure 3: Heatmap of occurrences of channels vs features from the MRMR algorithm on WPD features.

During our analysis, we checked for the statistical features with discriminative properties between the SI classes on DWT and WPD coefficients, we counted the number of times that each feature gave a higher-than-chance result, Figure 4 shows the occurrence of significant results from each feature, where the slope of the coefficients, appeared as the most discriminative property from these wavelet representations.

DISCUSSION

Research into the Speech Imagery paradigm is gaining traction, different experiments and designs prove that SI can be classified from EEG signals [3, 6, 7].

The Motor Imagery (MI) paradigm, whose event-related desynchronization (synchronization) is well known to have a predominant range of frequencies (Alpha and Beta) and location in the central Motor Cortex [34, 35]. In contrast, the SI-related potentials are not fully under-

10.3217/978-3-99161-014-4-092

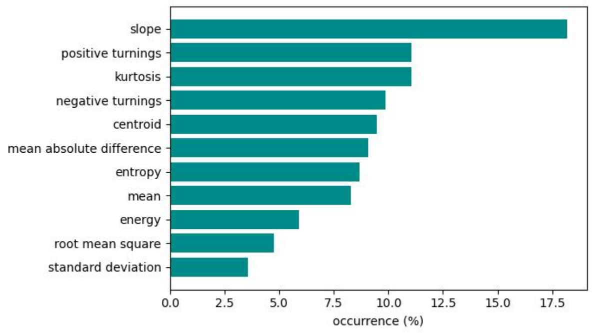

Figure 4: Occurrences of statistical features obtained from both DWT and WPD.

stood [36].

Speech Imagery involves more complex processing for the brain than MI [37], as a Language process, the brain regions known to be active during speech processing may be active during SI, the literature suggests that SI activity has a left hemisphere dominant processing, that involves different brain regions. Some commonly mentioned regions are the temporal-parietal junction that has been related to a memory and semantic decoding step, the frontal-temporal regions possibly handling syllabification and premotor and motor regions for the activity related to the somatosensory SI experiences [33, 38, 39].

Out of the most relevant features by the MRMR selector, we find channels located on regions that may be influenced by areas known to be active during language production, FT7 around Broca's area, C1 in the Motor cortex, T7 and C5 around the superior temporal region, P5 and P3 in the temporal parietal junction [33, 40]. However, the encountered relevant features are not restricted to these areas and are spread around different regions, as features from channels P8, TP8 or Fp2.

Studies of SI with Electrocorticography (ECOG) and EEG have found that this imagery paradigm involves broad-frequency dynamics and highlights the important contribution from the gamma band (> 60 Hz) [6, 38, 41]. Our results suggest that many informative features come from the narrow frequency ranges between 26–28 Hz, 78–80 Hz, or 122–124 Hz. It can also be noticed that relevant features appeared to be chosen nearly continuously in the Gamma range between 76–108 Hz but no features were significantly chosen between 96–106 Hz. We have tested WPD frequencies laying on Alpha, Beta and high Gamma bands and found relevant information is spread along different frequencies. Therefore we suggest that future SI analysis should consider a broad spectrum of the frequency domain.

The issue of participant-dependant frequency variability in SI from EEG data was demonstrated in our comparison between the two wavelet decomposition strategies. The general decomposition levels extracted with DWT in most of the cases did not lead to significant classification performance, however selecting participant-specific narrow frequency bands with the use of WPD significantly improved the classification accuracy, as shown in Figure 2. To test a decoding pipeline with multiple WPD configurations can be computationally expensive due to the total amount of available packets. In this work, we have pointed out some frequency ranges which combinations could be the starting testing point for future SI-related work.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-092

REFERENCES

[1] Wang L, Liu X, Liang Z, Yang Z, Hu X. Analysis and classification of hybrid bci based on motor imagery and speech imagery. Measurement. 2019;147:106842.

[2] Wang L, Zhang X, Zhang Y. Extending motor imagery by speech imagery for brain-computer interface. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013, 7056–7059.

[3] DaSalla CS, Kambara H, Sato M, Koike Y. Single-trial classification of vowel speech imagery using common spatial patterns. Neural Networks. 2009;22(9):1334–1339.

[4] Chengaiyan S, Anandan K. Effect of functional and effective brain connectivity in identifying vowels from articulation imagery procedures. Cognitive Processing. 2022;23(4):593–618.

[5] D'Zmura M, Deng S, Lappas T, Thorpe S, Srinivasan R. Toward eeg sensing of imagined speech. In: Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, 40–48.

[6] Jahangiri A, Sepulveda F. The relative contribution of high-gamma linguistic processing stages of word production, and motor imagery of articulation in class separability of covert speech tasks in eeg data. Journal of Medical Systems. 2018;43(2).

[7] Kaongoen N, Choi J, Jo S. Speech-imagery-based brain–computer interface system using ear-eeg. Journal of Neural Engineering. 2021;18(1):016023.

[8] Torres-García AA, Reyes-García CA, Villaseñor-Pineda L, García-Aguilar G. Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification. Expert Systems with Applications. 2016;59:1–12.

[9] Lotte F *et al.* A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. Journal of Neural Engineering. 2018;15(3):031005.

[10] Asghari Bejestani MR, Mohammad Khani GR, Nafisi VR, Darakeh F. Eeg-based multiword imagined speech classification for persian words. BioMed Research International. 2022;2022:1–20.

[11] Riaz A, Akhtar S, Iftikhar S, Khan AA, Salman A. Inter comparison of classification techniques for vowel speech imagery using eeg sensors. In: The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014). 2014, 712–717.

[12] Cooney C, Folli R, Coyle D. Mel frequency cepstral coefficients enhance imagined speech decoding accuracy from eeg. In: 2018 29th Irish Signals and Systems Conference (ISSC). 2018, 1–7.

[13] Adeli H, Zhou Z, Dadmehr N. Analysis of EEG records in an epileptic patient using wavelet transform. Journal of Neuroscience Methods. 2003;123(1):69–87.

[14] SUBASI A. Eeg signal classification using wavelet feature extraction and a mixture of expert model. Expert Systems with Applications. 2007;32(4):1084–1093.

[15] Mahapatra NC, Bhuyan P. Multiclass classification of imagined speech vowels and words of electroencephalography signals using deep learning. Advances in Human-Computer Interaction. 2022;2022:1–10.

[16] Sree RA, Kavitha A. Vowel classification from imagined speech using sub-band eeg frequencies and deep belief networks. In: 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN). 2017, 1–4.

[17] Mallat S. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989;11(7):674–693.

[18] Xue JZ, Zhang H, Zheng CX, Yan XG. Wavelet packet transform for feature extraction of eeg during mental tasks. In: Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693). 2003, 360–363 Vol.1.

[19] Ang KK, Chin ZY, Zhang H, Guan C. Filter bank common spatial pattern (fbcsp) in brain-computer interface. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). 2008, 2390–2397.

[20] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;27(8):1226–1238.

[21] DH B. The psychophysics toolbox. Spatial vision. 1997;10(4).

[22] Ablin P, Cardoso JF, Gramfort A. Faster independent component analysis by preconditioning with hessian approximations. IEEE Transactions on Signal Processing. 2018;66(15):4040–4049.

[23] Agarwal P, Kumar S. Electroencephalography based imagined alphabets classification using spatial and time-domain features. International Journal of Imaging Systems and Technology. 2021;32(1):111–122.

[24] Gramfort A *et al.* MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience. 2013;7(267):1–13.

[25] Idrees BM, Farooq O. Vowel classification using wavelet decomposition during speech imagery. In: 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN). 2016, 636–640.

[26] Biswas S, Sinha R. Wavelet filterbank-based EEG rhythm-specific spatial features for covert speech classification. IET Signal Processing. 2021;16(1):92–105.

[27] Nitta T *et al.* Linguistic representation of vowels in speech imagery eeg. Frontiers in Human Neuroscience. 2023;17.

[28] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of erp components — a tutorial. NeuroImage. 2011;56(2):814–825.

[29] Pedregosa F *et al.* Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;12:2825–2830.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-092

[30] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random? a closer look on bci results. International Journal of Bioelectromagnetism. 2008;10(1):52–55.

[31] Si X, Li S, Xiang S, Yu J, Ming D. Imagined speech increases the hemodynamic response and functional connectivity of the dorsal motor cortex. Journal of Neural Engineering. 2021;18(5):056048.

[32] Stephan F, Saalbach H, Rossi S. The brain differentially prepares inner and overt speech production: Electrophysiological and vascular evidence. Brain Sciences. 2020;10(3):148.

[33] Tian X, Zarate JM, Poeppel D. Mental imagery of speech implicates two mechanisms of perceptual reactivation. Cortex. 2016;77:1–12.

[34] Korostenskaja M *et al.* Characterization of cortical motor function and imagery-related cortical activity: Potential application for prehabilitation. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2017, 3014–3019.

[35] Jahangiri A, Chau JM, Achanccaray DR, Sepulveda F. Covert speech vs. motor imagery: A comparative study of class separability in identical environments. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018, 2020–2023.

[36] Kim HJ, Lee MH, Lee M. A bci based smart home system combined with event-related potentials and speech imagery task. In: 2020 8th International Winter Conference on Brain-Computer Interface (BCI). 2020, 1–6.

[37] Kraft E, Gulyás B, Pöppel E. Neural correlates of thinking. In: On Thinking. Springer Berlin Heidelberg, 2009, 3–11.

[38] Martin S *et al.* Decoding spectrotemporal features of overt and covert speech from the human cortex. Frontiers in Neuroengineering. 2014;7.

[39] Hickok G. The dual stream model of speech and language processing. In: Aphasia. Elsevier, 2022, 57–69.

[40] Broca's region. Oxford University PressNew York (2006).

[41] Martin S *et al.* Word pair classification during imagined speech using direct brain recordings. Scientific Reports. 2016;6(1).

Published by
Verlag der Technischen Universität Graz