INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Decoding imagined, heard, and spoken speech: classification and regression of EEG using a 14-channel dry-contact mobile headset

Jonathan Clayton[1], *Scott Wellington*[12], *Cassia Valentini-Botinhao*[12], *Oliver Watts*[12]

[1]The University of Edinburgh
[2]SpeakUnique Limited

jclayton558@gmail.com, scott@speakunique.co.uk, cvbotinh@inf.ed.ac.uk, oliver@speakunique.co.uk

Abstract

We investigate the use of a 14-channel, mobile EEG device in the decoding of heard, imagined, and articulated English phones from brainwave data. To this end we introduce a dataset that fills a current gap in the range of available open-access EEG datasets for speech processing with lightweight, affordable EEG devices made for the consumer market. We investigate the effectiveness of two classification models and a regression model for reconstructing spectral features of the original speech signal. We report that our classification performance is almost on a par with similar findings that use EEG data collected with research-grade devices. We conclude that commercial-grade devices can be used as speech-decoding BCIs with minimal signal processing.
Index Terms: EEG, brain-computer interfaces, imagined speech, neural decoding, stimulus reconstruction

1. Introduction

Brain-computer interface (BCI) research is a promising avenue for the development of voice output communication aids (VOCAs), for use by individuals with impaired speech resulting from conditions such as motor neurone disease. VOCAs incorporating BCIs, such as P300 spelling systems, have shown great improvements in recent decades [1]. For individuals with advanced phonatory function decline, such systems can be essential for effective communication [2]. However, these systems are limited by the slow rate of text entry they allow and the high level of concentration they require of the user. Some promising research, using non-invasive or invasive techniques, has shown the feasibility of leveraging the EEG brainwave signal directly to extract auditory, articulatory or phonetic features. For example, research published in the last year alone (2019) has reported exciting results for decoding features of speech from EEG of auditory stimuli [3], of overt spoken or mimed utterances [4, 5], and of covert imagined speech [6].

Researchers attempting to use EEG signals to decode speech will quickly discover that few purpose-built datasets have been recorded and released under open licences (e.g. [7, 8, 9]). Furthermore, these data are typically recorded with scientific-grade EEG devices whose cost, bulk, and setup requirements may limit their utility in everyday scenarios. In this study, we collect a dataset using a wearable, commercial-grade device which offers lower fidelity with fewer electrodes, but which is more practical for everyday use.

The present paper has three primary aims:

1. To evaluate the effectiveness of "lightweight" EEG devices for speech decoding, by comparing classification performance against data from a research-grade device.
2. To investigate different machine learning models for speech classification and regression tasks with EEG.
3. To present a new liberally licensed corpus of speech-evoked EEG recordings, together with benchmark results and code.[1]

2. The FEIS dataset

The FEIS (Fourteen-channel EEG for Imagined Speech) dataset [10], comprises EEG recordings of 21 English-speaking participants recorded with a lightweight, 14-channel mobile headset with dry electrodes (the Emotiv EPOC+) [11]. Recordings are time-aligned with phone stimuli, consisting of three stimulus types: heard, spoken internally (imagined) and spoken overtly.

Data collection methodology is adapted from that of the Kara One dataset [7]. The Kara One dataset is described below (section 3) and compared point-by-point with FEIS.

2.1. Participants

21 participants were recruited at the University of Edinburgh. Participants are either native or near-native speakers of English (CEFR level $\geq$ C1), with no known neurological disorders. Three participants are left-handed, one ambidextrous, and the remaining 17 right-handed (FEIS metadata available at [10]).

2.2. Stimuli

Sixteen English phonemes were chosen to represent a balanced categorical spread of binary phonological features ([±nasal], [±back], [±voice], *etc*). These are shown in Table 1.

Table 1: *Phoneme types in the FEIS dataset*

A. *Consonants*

	Labial	Alveolar	Postalveolar/ Velar
Plosive (-voice)	/p/	/t/	/k/
Fricative (-voice)	/f/	/s/	/ʃ/
Fricative (+voice)	/v/	/z/	/ʒ/
Nasal (+voice)	/m/	/n/	/ŋ/

B. *Vowels*

	Front	Back
High	/i/	/u/
Low	/æ/	/ɔ/

2.3. Recording procedure

High-quality audio of the phonemes listed in Table 1 was recorded in the participants' own voices. A single instance of each of the 16 phones was recorded at 44.1 kHz with a cardioid microphone. We used audio processing software to convert these single-phone prompts into stimuli consisting of five

[1]Available at: https://doi.org/10.5281/zenodo.3369178

http://dx.doi.org/10.21437/Interspeech.2020-2745

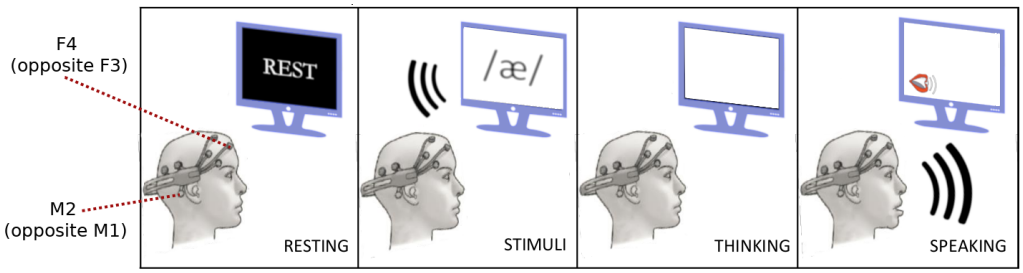

Figure 1: *Illustration of the recording procedure described in section 3.3. Participants listen to five repetitions of a phone (recorded in their own voice), then imagine speaking the phone five times (with the same rhythm), then overtly speak the phone five times.*

repetitions of each phone. For plosives (e.g. /p/), participants were instructed to form a neutral release (resulting in e.g. [pə]).

Participants carried out the experiment alone, sitting in a comfortable chair in front of a laptop screen, inside a hemi-anechoic chamber. Our intention behind these choices in methodology is to mitigate contamination from brainwave components resulting from unexpected audio or visual stimuli (such as the P300 event-related potentials (ERPs) [12].

The Emotiv EPOC+ is a mobile headset with semi-flexible sensor "arms" which allow for universal fitting, within a fixed configuration. While this allows ease of use, it means that electrode positions are inconsistent relative to the international 10-20 montage system [13], due to participants' different head sizes. For reasonable consistency, we ensure F3/F4 sensors are 20mm above each subject's eyebrows, and M1/M2 dummy electrodes placed on the mastoid process (see Figure 1).

The EEG recordings consist of 160 trials, comprising 6 phonemes $\times$ 10 repetitions, randomised to maintain participant attention. Each trial has four 5-second "epochs", as illustrated in Figure 1. First, a **"resting"** epoch, in which participants are shown the word "REST" on screen, and attempt to clear their mind (resting state measurement can be used for task-specific feature extraction, and also reduces cognitive load). Next, a **"stimuli"** epoch, in which participants are played their own vocalisation of a single phone looped five times, and shown a corresponding IPA representation (which participants were familiar with). Next, a **"thinking"** epoch, in which participants are presented with a blank screen, and imagine repeating the phone, but without any articulator movement. Finally, a **"speaking"** epoch, in which participants are prompted with an image of a mouth to then vocalise the phone. In each of the two latter epochs, subjects imagine/speak the phone five times in a steady rhythm, imitating the recording played in the stimuli epoch.

2.4. Noise removal

The built-in software of the Emotiv EPOC+ performs notch-filtering at 50 Hz and 60 Hz to remove powerline noise [11]. No signal preprocessing was carried out to remove physiological artifacts (such as blinks or saccades). Often, an independent component analysis (ICA) pipeline is used to remove such artifacts from the data [14]; however, since the Emotiv EPOC+ lacks the ocular channels typically used to isolate noise components through correlation, this was not carried out. Future work could perform ICA on FEIS by using ICA solutions from datasets collected using other devices (as described in [15]).

3. Previous Publicly-Available Datasets

The methodology followed in the design and collection of FEIS was based on that of the Kara One dataset [7]. Kara One contains multimodal recordings of speech (heard, imagined, and spoken). This includes audio recordings, EEG recordings, and recordings of facial movements. The prompts used in the Kara One dataset include English phones and single syllables.

The Kara One EEG recordings were made at a sampling frequency of 1000 Hz using a 64-channel Neuroscan Quik-Cap. There were 14 participants in the study, and around 30 to 40 minutes of recordings were taken for each participant.

The Kara One researchers used support vector machine (SVM) models for binary phonological classification tasks, reporting accuracy results ranging from 18.08% to 79.16% (above chance). Other researchers have reported improved results using deep neural models [16, 17].

Table 2: *A comparison of our dataset (FEIS) to open-access dataset Kara One, on which this study is modelled.*

	Kara One [7]	**FEIS (this dataset)**
EEG Device	64 channels	14 channels
Sampling frequency	1000 Hz	256 Hz
No. of participants	14	21
Recording duration (per participant)	30 to 40 minutes	60 minutes
Prompts used in trials	11 syllables/ phonemes	16 phonemes

4. Model Building

In preliminary experiments [18] [19], we tested both subject-dependent (test and training data are from the same subject) and leave-one-out subject-independent models. In all conditions, subject-dependent models demonstrated better accuracy and performance. All models presented here are therefore subject-dependent.

We employ a 80/10/10 split for training, testing and validation sets, with the exception of the support vector machine (SVM) model, where it was convenient to use an 80/20 training/test split with 5-fold cross-validation.

4.1. Model selection

The classification/regression task pipeline is summarized in Figure 2. First, we optionally perform ICA or other artifact removal procedures, although this was not carried out in this study (see Section 2.4). For each speech type (heard, imagined, spoken), the data is split into five-second epochs. Features are extracted from these chunks and are used to train a classifier using the corresponding phone class labels, and a regression

Table 3: *Parameters of machine learning models trialled*

	Layers	Loss function	Hyperparameters
SVM (Classification)	N/A	Hinge Loss	C ∈ [1:1000] gamma ∈ [0.001:0.000001]
CNN (Classification)	1 x [Conv2D, Conv2D, BatchNorm, ELU, MaxPool] 3 x [Dropout, Conv2D, BatchNorm, ELU, MaxPool] 1 x [Dense, MaxPool]	Cross Entropy	No. conv filters* ∈ [25, 50, 100, 200] Pooling/filter length ∈ [5, 10, 20, 40] Stride length (pool/filter) ∈ [3, 6, 9, 12]
Dense Network + DAE (Regression)	**EEG → Bottleneck:** 6 x [Dense, LeakyRELU] **Vocoder Feats. Autoencoder:** 3 x [GaussianNoise, Dense, LeakyRELU] (encoder) 3 x [Dense, LeakyRELU] (decoder)	Mean Squared Error	Adam: $\eta = 1 \times 10^{-4}; \lambda = 1 \times 10^{-5}$ Early stopping tolerance: 1×10^{-10} Additive Noise: $\mu = 0; \sigma = 1 \times 10^{-6}$

model using features extracted from the raw audio stimuli that were presented/imagined/spoken.

For the classification task, two types of models are tested; support vector machines (SVMs) and Convolutional Neural Networks (CNNs). SVMs are a good baseline, with above-chance performance on the Kara One dataset [7]. The models were trained to predict binary phone classes from the EEG signal (±consonant, ±/u/, ±voice).

For the regression task of predicting vocoder features from EEG signal data we employ two different models, a 7-layer densely-connected neural network and a stacked Denoising AutoEncoder (DAE) [20]. The latter model is trained on the VCTK speech corpus [21] to autoencode WORLD vocoder features [22]. The former model is trained on the FEIS dataset to predict DAE encoded features from the corresponding EEG signal. The DAE features are extracted by running the trained DAE model encoder on WORLD features extracted from the audio data.

4.2. Feature selection for classification

For our SVM model, we divide each 5 second "epoch" into 500ms windows with 250ms overlap, and we compute 28 statistical features (84, after adding delta and delta-delta features) per window and per electrode. Hence, we compute a feature vector of length 14 electrodes × 19 windows × 84 features = 22344. Of our 28 features, 19 were previously used by Zhao and Rudzicz [7]. A full list of features is given in [10].

For our CNN model, we use 2000ms overlapping windows of the unprocessed EEG signal. As the recording's sampling frequency is 256 Hz, the 2D input arrays have a dimensionality of 512×14 (frames×electrodes).

4.3. Signal processing for regression

We bandpass filter the EEG signal into five frequency bands that may encode functionally distinct neural oscillations [23] [24]: delta (<4 Hz), theta (4–7 Hz), alpha (8–15 Hz), beta (16–31 Hz) and gamma (32-70 Hz). We then extract the envelope in each band using the Hilbert transform. Following Akbari *et al.* [3] we extract another 8 envelopes from bands between 70-150 Hz, in 10 Hz increments (70-80 Hz, 80-90 Hz,*etc*). We then take the mean of these 8 high-gamma envelopes, for a total of 6 sets (δ, θ, α, β, γ and high-γ).

This creates an 84-value array per sample of data (6 sets × 14 channels), which our dense network uses as input to decode into 256-value DAE bottleneck feature vectors. For 16 kHz audio, WORLD is configured to sample at a rate of 200 Hz. We therefore resample our 256 Hz EEG signal to 200 Hz, using a low-pass FIR filter to prevent spectral aliasing.

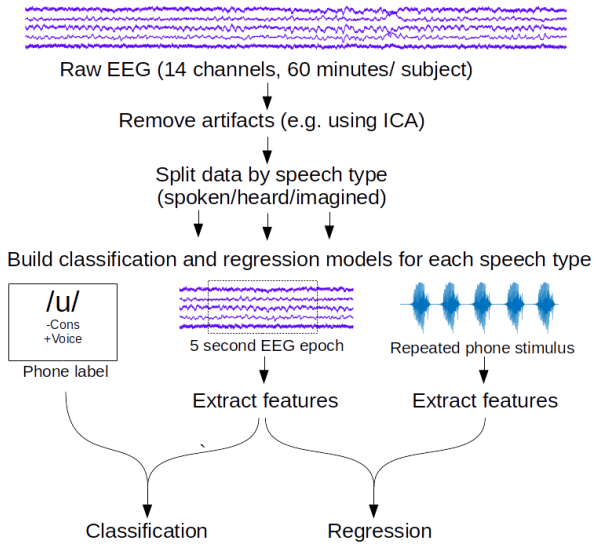

Figure 2: *Illustration of our model-building procedure (see text for details).*

4.4. Model implementation

Table 3 shows the details of the model architectures, as well as any hyperparameters which were tuned for by grid search. Following [7], to reduce the number of features for our SVM model, Pearson's r is used to inform an N best list, N ∈ [5:100], of most correlated features (averaged over 14 correlations-per-electrode) selected for the feature being classified (e.g. ± Consonant). We use a grid search to tune N, in addition to the hyperparameters in Table 3 (tuning for F-score).

Following Heilmeyer *et al.* [25], we employ the Deep4Net architecture [26] for our CNN model, trained for 30 epochs with the ADAM optimizer [27], with dropout set to 0.5.

For our regression model, the participants' own audio and those within the VCTK corpus are downsampled to 16 kHz before being translated into WORLD feature vectors to train the DAE. The DAE decoders perform regression, predicting 516-value WORLD vectors (513-value spectral envelope + f_0 + aperiodicity + excitation) from the bottleneck features generated by the EEG → bottleneck network.

4.5. Evaluation Metrics

For binary classification we evaluate using percentage accuracy, since we have ensured that in each condition training and test sets contain an equal number of tokens of each class. To eval-

uate regression models, we examine log spectral distortion (a measure of distance between original and decoded spectra) and signal-to-noise ratio, calculated (where P is average power) as $\text{SNR}_{\text{dB}} = 10\log_{10}(P_{\text{Signal}}) - 10\log_{10}(P_{\text{Noise}})$.

5. Results and Discussion

5.1. Comparison of Classification Models

Table 4 indicates that the better-performing model of the two classifiers we compared was the SVM, which consistently achieved above-chance accuracy, while the CNN did not. The reasons for this require further investigation, but may be due to improved feature selection for the SVM, or the more complex CNN model requiring more training data to be effective [28].

Table 4: *A comparison of the performance of our SVM and CNN on the ± Consonant classification task (FEIS dataset, averaged across 5 subject-dependent models)*

	Percentage Accuracy (std. dev)		
Task	**Hearing**	**Thinking**	**Speaking**
SVM	64.0 (16.5)	69.0 (13.2)	63.7 (21.2)
CNN	51.2 (5.7)	49.0 (6.1)	49.4 (6.3)

5.2. Classification results on Kara One and FEIS

The results in Figure 3 compare averages across five subject-dependent models built for randomly-selected subjects. We show that for most tasks using the Kara One dataset, collected with a 64-channel headset sampled at 1000 Hz, resulted in only slight improvements in classification accuracy over the FEIS dataset, collected with a 14-channel headset sampled at 256 Hz. We use the SVM model, as it was shown to give better results.

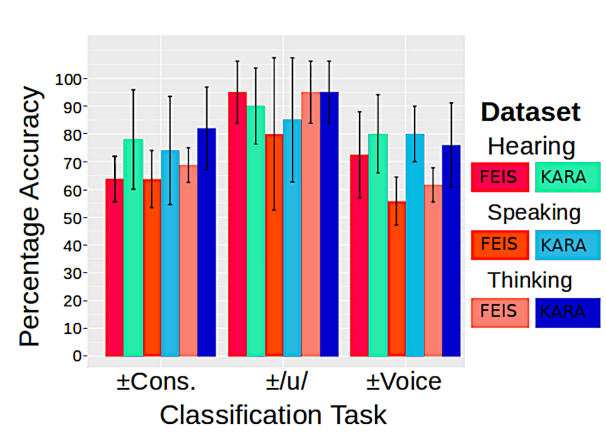

Figure 3: *A comparison of classifier accuracies on the FEIS dataset (presented in this paper) and the Kara One dataset. Error bars show one standard deviation.*

5.3. Regression results

Log spectral distortion results achieved by the best model constructed are shown in Table 5. The average SNR for this model is -1.62dB (standard deviation 0.91). An example spectrogram generated by this model is shown in Figure 4. While the decoded audio is not intelligible, it is nevertheless possible to discern speech from non-speech.

In future studies, we may modify our regression architecture to include convolutional layers. Our dense feedforward architecture relies on spectral information extracted during the preprocessing stage. CNN models, in contrast, may be able to learn to extract relevant spectral information [26]. CNNs have additional advantages for EEG data. Since they often achieve better results than dense feedforward nets with the same number of learnable parameters [29], we may be able to use a larger input, (corresponding a broader temporal context) without unduly increasing the size of the model. Additionally, there is evidence that EEG signals are hierarchical in the temporal domain (e.g. [30], [31]); CNNs are well-suited to extracting higher-level features from hierarchically-structured inputs [32]. However, since the CNN model trialled on the classification task performed poorly, more investigation is clearly required.

Table 5: *Log-spectral distortion measures on the test dataset, organized by articulation and speech type*

	Mean Log Spectral Distortion (std. dev)			
	Hearing	**Speaking**	**Thinking**	**Average**
Vowels	1.86 (0.17)	1.82 (0.20)	1.91 (0.21)	1.87 (0.20)
Nasals	1.69 (0.06)	1.62 (0.01)	1.71 (0.04)	1.67 (0.06)
Fricatives	2.37 (0.23)	2.37 (0.19)	2.39 (0.24)	2.33 (0.22)
Plosives	2.66 (0.13)	2.61 (0.12)	2.63 (0.12)	2.53 (0.24)
Average	2.17 (0.39)	2.14 (0.40)	2.19 (0.39)	2.16 (0.40)

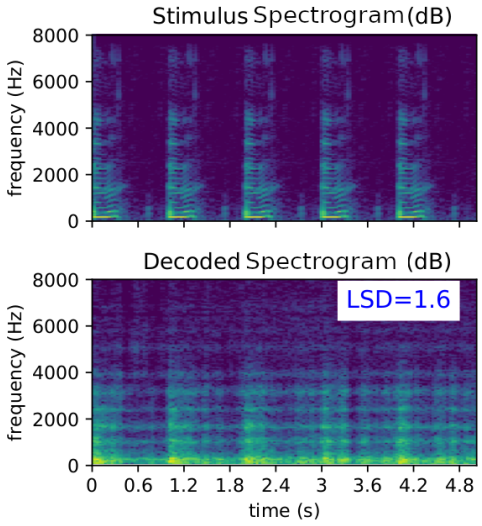

Figure 4: *A spectrogram of the original stimulus (above) and decoded EEG (below), showing 5 repetitions of vocalised /u/.*

6. Conclusions

Preliminary results using the FEIS dataset indicate that "lightweight", mobile EEG devices can obtain data that encode speech processing similar to research-grade devices with higher electrode density and fidelity, as with the Kara One dataset. These data are sufficient to carry out binary classification tasks based on short recordings of speech sounds with a greater-than-chance accuracy. We have also shown some suggestive results that the FEIS data can be used to synthesize an approximation of the original spectral envelope, and anticipate that better decoding performance may be achievable with deep neural architectures employing the latest optimization techniques.

7. References

[1] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nature Reviews Neurology*, vol. 12, no. 9, p. 513, 2016.

[2] E. W. Sellers, T. M. Vaughan, and J. R. Wolpaw, "A brain-computer interface for long-term independent home use," *Amyotrophic lateral sclerosis*, vol. 11, no. 5, pp. 449–455, 2010.

[3] H. Akbari, B. Khalighinejad, J. L. Herrero, A. D. Mehta, and N. Mesgarani, "Towards reconstructing intelligible speech from the human auditory cortex," *Scientific reports*, vol. 9, no. 1, p. 874, 2019.

[4] M. Angrick, C. Herff, E. Mugler, M. C. Tate, M. W. Slutzky, D. J. Krusienski, and T. Schultz, "Speech synthesis from ECoG using densely connected 3D convolutional neural networks," *Journal of neural engineering*, vol. 16, no. 3, p. 036019, 2019.

[5] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, pp. 493–498, 2019.

[6] P. Saha and S. Fels, "Hierarchical deep feature learning for decoding imagined speech from EEG," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 10 019–10 020.

[7] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 992–996.

[8] M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, "Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech," *Current Biology*, vol. 28, no. 5, pp. 803–809, 2018.

[9] J.-M. Schoffelen, R. Oostenveld, N. H. Lam, J. Uddén, A. Hultén, and P. Hagoort, "A 204-subject multimodal neuroimaging dataset to study language processing," *Scientific data*, vol. 6, no. 1, pp. 1–13, 2019.

[10] S. Wellington and J. Clayton, "Fourteen-channel EEG with imagined speech (FEIS) dataset," Aug 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3369179

[11] Emotiv EPOC+. [Online]. Available: https://www.emotiv.com/epoc/

[12] T. W. Picton, "The P300 wave of the human event-related potential," *Journal of clinical neurophysiology*, vol. 9, no. 4, pp. 456–479, 1992.

[13] G. H. Klem, H. O. Lüders, H. Jasper, C. Elger *et al.*, "The ten-twenty electrode system of the international federation," *Electroencephalogr Clin Neurophysiol*, vol. 52, no. 3, pp. 3–6, 1999.

[14] T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts: comparison between ICA and PCA," in *Neural Networks for Signal Processing VIII. Proceedings of the 1998 IEEE Signal Processing Society Workshop (Cat. No. 98TH8378)*. IEEE, 1998, pp. 63–72.

[15] F. C. Viola, J. Thorne, B. Edmonds, T. Schneider, T. Eichele, and S. Debener, "Semi-automatic identification of independent components representing EEG artifact," *Clinical Neurophysiology*, vol. 120, no. 5, pp. 868–877, 2009.

[16] C. Cooney, F. Raffaella, and D. Coyle, "Classification of imagined spoken word-pairs using convolutional neural networks," in *The 8th Graz BCI Conference, 2019*, 2019.

[17] S. Martin, I. Iturrate, P. Brunner, J. d. R. Millán, G. Schalk, R. T. Knight, and B. N. Pasley, "Individual word classification during imagined speech using intracranial recordings," in *Brain-Computer Interface Research*. Springer, 2019, pp. 83–91.

[18] J. Clayton, "Towards phone classification from imagined speech using a lightweight EEG brain-computer interface," University of Edinburgh, Edinburgh, UK, 2019. M.Sc. dissertation. [Online]. Available: https://doi.org/10.5281/zenodo.3369178

[19] S. Wellington, "An investigation into the possibilities and limitations of decoding heard, imagined and spoken phonemes using a low-density, mobile EEG headset," University of Edinburgh, Edinburgh, UK, 2019. M.Sc. dissertation. [Online]. Available: https://doi.org/10.5281/zenodo.3369178

[20] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of machine learning research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[21] C. Veaux, J. Yamagishi, and K. MacDonald, "CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit, [sound]. University of Edinburgh. The Centre for Speech Technology Research (CSTR)," 2017. [Online]. Available: https://doi.org/10.7488/ds/1994

[22] M. Morise, F. Yokomori, and K. Ozawa, "WORLD: a vocoder-based high-quality speech synthesis system for real-time applications," *IEICE TRANSACTIONS on Information and Systems*, vol. 99, no. 7, pp. 1877–1884, 2016.

[23] A. G. Lewis, J.-M. Schoffelen, H. Schriefers, and M. Bastiaansen, "A predictive coding perspective on beta oscillations during sentence-level language comprehension," *Frontiers in human neuroscience*, vol. 10, p. 85, 2016.

[24] C. Spironelli and A. Angrilli, "Developmental aspects of language lateralization in delta, theta, alpha and beta eeg bands," *Biological psychology*, vol. 85, no. 2, pp. 258–267, 2010.

[25] F. A. Heilmeyer, R. T. Schirrmeister, L. D. Fiederer, M. Volker, J. Behncke, and T. Ball, "A large-scale evaluation framework for eeg deep learning architectures," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2018, pp. 1039–1045.

[26] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[29] G. Urban, K. J. Geras, S. E. Kahou, O. Aslan, S. Wang, R. Caruana, A. Mohamed, M. Philipose, and M. Richardson, "Do deep convolutional nets really need to be deep and convolutional?" *arXiv preprint arXiv:1603.05691*, 2016.

[30] K. Kirihara, A. J. Rissling, N. R. Swerdlow, D. L. Braff, and G. A. Light, "Hierarchical organization of gamma and theta oscillatory dynamics in schizophrenia," *Biological psychiatry*, vol. 71, no. 10, pp. 873–880, 2012.

[31] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. Da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

[32] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 609–616.