OPEN ACCESS
International Journal of Neural Systems, Vol. 31, No. 7 (2021) 2150025 (16 pages)

DOI: 10.1142/S0129065721500258

Prediction of Speech Onset by Micro-Electrocorticography of the Human Brain

Emanuela Delfino[*,†,**], Aldo Pastore[*,†,††], Elena Zucchini[*,†,‡‡], Maria Francisca Porto Cruz[*,†,‡,§§], Tamara Ius[§,¶¶], Maria Vomero[¶,‖‖], Alessandro D'Ausilio[*,†,***], Antonino Casile[*,†††], Miran Skrap[§,‡‡‡], Thomas Stieglitz[‡,‖,§§§] and Luciano Fadiga[*,†,‖‖‖‖]

[*]*Center for Translational Neurophysiology*
Istituto Italiano di Tecnologia, Via Fossato di Mortara 17-19
Ferrara 44121, Italy

[†]*Section of Physiology, University of Ferrara*
Via Fossato di Mortara 17-19, Ferrara 44121, Italy

[‡]*Laboratory for Biomedical Microtechnology,*
Department of Microsystems Engineering (IMTEK)
University of Freiburg, Georges-Köhler-Allee 102
Freiburg im Breisgau 79110, Germany

[§]*Struttura Complessa di Neurochirurgia*
Azienda Ospedaliero-Universitaria Santa Maria
della Misericordia, Piazzale Santa Maria
della Misericordia 15, Udine 33100, Italy

[¶]*Bioelectronic Systems Laboratory*
Columbia University, 500 West 120th Street
New York, NY 10027, USA

[‖]*BrainLinks-BrainTools Center*
University of Freiburg, Georges-Köhler-Allee 80
Freiburg im Breisgau 79110, Germany

[**]*emanuela.delfino@unife.it*
[††]*aldo.pastore@iit.it*
[‡‡]*elena.zucchini@unife.it*
[§§]*mariafrancisca.portocruz@unife.it*
[¶¶]*tamara.ius@gmail.com*
[‖‖]*mv2803@columbia.edu*
[***]*alessandro.dausilio@unife.it*
[†††]*antonino.casile@iit.it*
[‡‡‡]*miran.skrap@gmail.com*
[§§§]*stieglitz@imtek.uni-freiburg.de*
[‖‖‖‖]*luciano.fadiga@iit.it*

Received 29 March 2021
Accepted 30 March 2021
Published Online 14 June 2021

[‖‖‖‖]Corresponding author.
[**,††]Co-first authors.

Recent technological advances show the feasibility of offline decoding speech from neuronal signals, paving the way to the development of chronically implanted speech brain computer interfaces (sBCI). Two key steps that still need to be addressed for the online deployment of sBCI are, on the one hand, the definition of relevant design parameters of the recording arrays, on the other hand, the identification of robust physiological markers of the patient's intention to speak, which can be used to online trigger the decoding process. To address these issues, we acutely recorded speech-related signals from the frontal cortex of two human patients undergoing awake neurosurgery for brain tumors using three different micro-electrocorticographic (μECoG) devices. First, we observed that, at the smallest investigated pitch (600 μm), neighboring channels are highly correlated, suggesting that more closely spaced electrodes would provide some redundant information. Second, we trained a classifier to recognize speech-related motor preparation from high-gamma oscillations (70–150 Hz), demonstrating that these neuronal signals can be used to reliably predict speech onset. Notably, our model generalized both across subjects and recording devices showing the robustness of its performance. These findings provide crucial information for the design of future online sBCI.

Keywords: μECoG; Broca; speech arrest; spatial resolution; speech preparation; BCI.

1. Introduction

Recent advances in neuroprosthetics demonstrated that intelligible speech can be offline synthesized from cortical activity.[1,2] While this represents an important stepping stone, it still leaves open crucial problems that need to be solved to develop speech brain computer interfaces (sBCI) which can be effectively implanted in patients and work continuously online.[1–6] Here, we addressed two of them.

The first problem is the need of developing devices that can chronically record brain signals in a reliable manner. Several techniques have been presented in the literature, which differ in their degree of invasiveness and spatiotemporal resolution. Starting from one of the least invasive methodologies, electroencephalography (EEG) probes electrical potential variations by using scalp electrodes. Neural oscillations are collected from large regions of the brain, making it an appropriate method for investigating communication within the brain during speech-related tasks[7] and a powerful clinical tool to recognize, among several others, speech and auditory deficits.[8] Nevertheless, when dealing with sBCI applications, electrocorticography (ECoG) — performed by placing grids of electrodes directly above the cortical surface — can be considered an excellent trade-off between several requirements. Indeed, this technique offers the advantage of recording neural activity from distributed brain areas with a spatiotemporal resolution inaccessible to non-invasive methodologies (e.g. EEG) and reduced invasiveness when compared to intracortical devices.[4,9–12] In the case of chronic recordings, as is the case for BCIs, the use of ultra-flexible micro-ECoG (μECoG) arrays, rather than traditional ECoG grids, has the further advantage of lowering the foreign body reaction, and thus improving long-term performances, as largely demonstrated in animal models.[13–20] Indeed, thin μECoG do conformably adhere to the brain surface with a curvature of a human brain without adding pressure to it.[13] Therefore, good signal-to-noise ratio of recorded signals occur. Not only low frequencies can be detected but even spike-like activity can be recorded with these arrays and electrode site diameters in the hundreds of micrometer range.[14]

Nonetheless, while μECoG arrays represent a promising strategy, several of their design parameters still need to be determined, of which, a crucial one, is the pitch distance between electrodes. Indeed, signal redundancy between neighboring electrodes increases as the electrode pitch decreases.[21,22] Thus, below a given threshold distance, signals would become highly correlated and the negligible additional information that they provide would not justify their additional design and manufacturing costs. Such threshold is presently unknown, and it needs to be estimated from experimental data, also considering the purpose of the BCI.[22]

The second problem we addressed here is that presently available speech-decoding devices are designed for an offline use. That is, they synthesize words and sentences from brain signals that are known to be collected during speech-related task.[1,2] On the contrary, in a prospective real-life online scenario, BCIs would be constantly exposed to the flow of neuronal activations with no additional information as to whether these signals are related to speech production or not, similarly to inner speech

settings.[23–26] Under these circumstances, the device would continuously attempt to convert patterns of neuronal activity into words, with conceivably high computational cost.[1,2]

In recent years, substantial efforts went into developing new strategies to both get closer to a natural speech scenario and optimize the decoding process.[23–31] Among the explored options, one includes the identification of speech-preparatory neural signals to reliably detect the speech onset. Previous studies reported that the most accurate speech onset/offset neuronal signals are typically found in the temporal cortex[27,28] raising the issue that they might be related to the auditory feedback of the subject's own voice. Unfortunately, such signals, while indeed highly correlated with speech onset/offset,[27,28,31] would not be available for a real-life sBCI deployment which implies the decoding of speech from patients that can no longer produce it.

Aiming to complement previous attempts in the field, one should investigate neuronal markers with two crucial characteristics: (1) ability of predicting the speech onset, and thus being able to provide sufficient time to trigger the decoding process, and (2) high correlation with speech preparation processes, and thus being available irrespective of the actual emission of speech. Similar to how a vocal cue is employed to start commonly used virtual assistants (e.g. Google Assistant, Alexa or Siri), such "neuronal cue" would serve the purpose of precisely identifying speech intentions and consequently trigger the initiation of the decoding process in time.

One suitable candidate region of the human cortex where to find physiological signals related to speech preparation is the speech arrest in Broca's area. Indeed, experimental evidence shows that direct electrical stimulation[32–34] during speech production induces the so-called speech arrest phenomenon, i.e. the complete interruption of ongoing speech[35] in absence of oro-facial movements and vocalizations.[36] This reversible functional arrest identifies Broca's area, which is known to be active prior to articulation rather than during spoken responses.[37] Specifically, results showed an increase of the high-gamma activity immediately before the speech onset or the peak of verbal response.[37,38]

In this study, we recorded neural activity from Broca's area using innovative dense μECoG grids (pitch distances = 600, 750 and 2500 μm) acutely implanted in two patients performing speech production tasks. We used signals recorded from the speech arrest area to provide a quantitative estimate of the correlation between electrodes, as a function of their distance. If there would be high correlations between adjacent electrodes this would be a sign for redundancy. If not, signals are independent and only one electrode was selected. This estimate, together with our time-frequency analysis, allowed us to identify the most appropriate frequency band in terms of spatial confinement and strict anticipatory nature with the respect to the speech onset, and thus to select the most robust physiological marker of speech preparation periods.

2. Materials and Methods

2.1. *Subjects*

Data were collected from two patients undergoing awake neurosurgery for tumor resection (low-grade glioma). The patients gave their informed consent, and the protocol was approved by the Ethics Committee of Azienda Ospedaliera Universitaria Santa Maria della Misericordia (Udine, Italy) after verification of the Italian Ministry of Health.

2.2. *Recordings*

Device specifications and recording setup were described in previous publications.[13,39–41] Briefly, three different epicortical arrays were used for the recordings (Fig. 1): the first array (hereinafter Epi) consisted of 64 channels arranged in an 8×8 square grid layout, with a pitch of 600 μm between contacts and a contact diameter of 140 μm; the second array (Multi Species Array; hereinafter MuSA) consisted of 16 channels arranged in a 4×4 square grid layout, with a pitch of 750 μm between contacts and a contact diameter of 100 μm; the third array (hereinafter EpiBig) consisted of 64 channels arranged in a rectangular grid, with a pitch of 2500 μm between contacts and a contact diameter of 200 μm. As required by the surgical procedures, the devices were sterilized before use.

The reference electrodes on the arrays were disconnected. Recordings were performed in a single-ended configuration by shorting the reference and ground contact and connecting them to the *dura mater*.

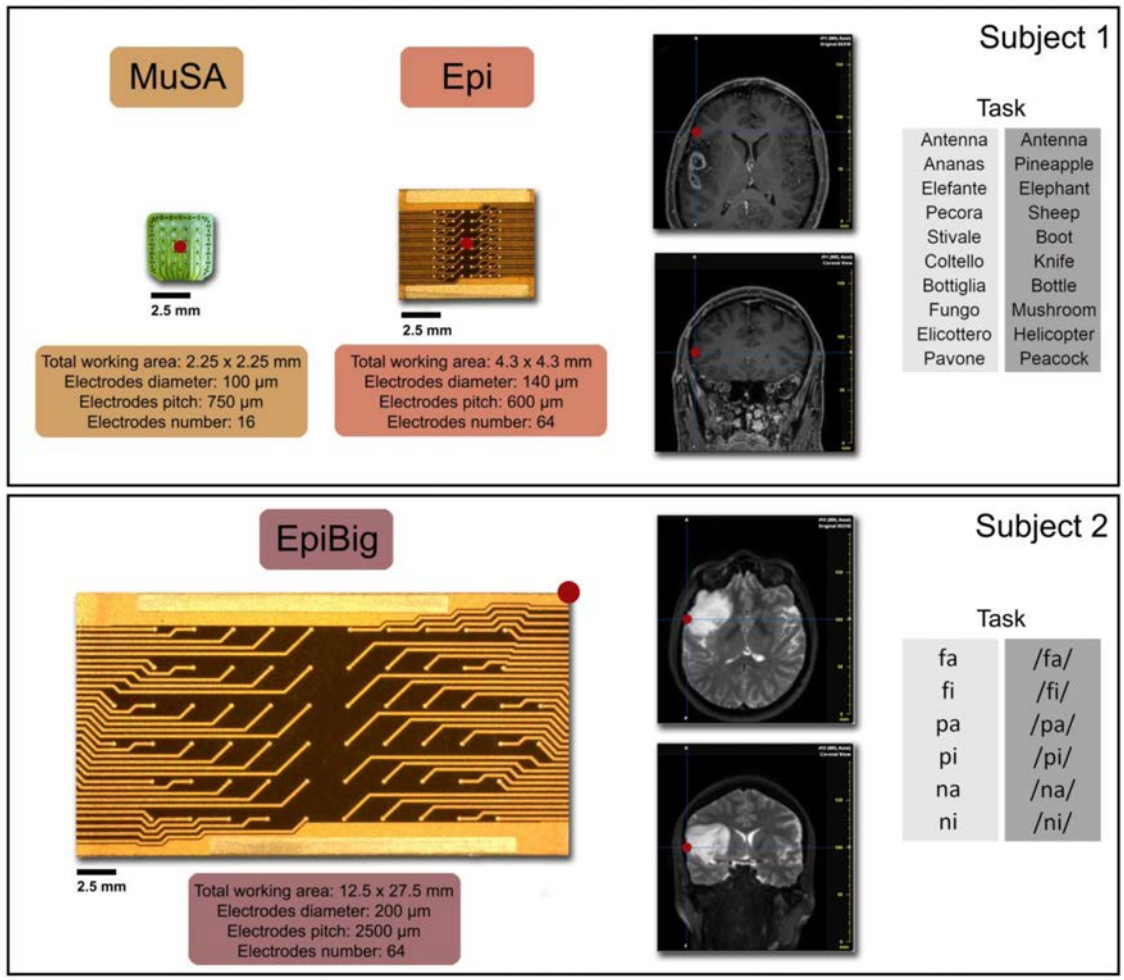

Fig. 1. (Color online) μECoG arrays layout and position over the cortex of subject1 (top) and subject2 (bottom). The top-left panel shows pictures of the Epi and the MuSA μECoG array. The top right panel shows and horizontal and coronal section of the patient's MRI scan. The center of each array (red dot in the left panel) was positioned over the speech arrest area (red dot in the right panel). The bottom-left panel shows a picture of the EpiBig μECoG array. The red dot localizes upper-right corner of the array superimposed to the MRI scan of the patient (horizontal plane and coronal plane). For both subjects, the speech production tasks are reported on the rightmost side of the panels.

The position of the μECoG arrays on the cortex was determined based on pre-surgical analyses and intra-operative procedures. Pre-surgical analyses included a functional Magnetic Resonance Imaging (fMRI) session while performing different speech production tasks. Intraoperative procedures consisted in identifying the position of the speech arrest area by means of electrical stimulation (IES). Briefly, using a neuronavigation system (Brainlab) and an IES probe, it was possible to map eloquent areas of the brain and visualize them superimposed to the fMRI scan of the patient. This procedure is typically conducted to identify the exact position of specific regions, such as the speech arrest, and evaluate their relative distance from the tumor. We used the same approach to collect the coordinates of the speech arrest area and of the position of the array once in place, which allowed us to align the MuSA and the Epi devices.

Neural signals were collected before the surgical procedure at a sampling frequency of 3051.8 Hz, while the voices were recorded at 24 kHz. The voice and the neural signals were recorded using the same data acquisition equipment; thus, they were automatically synchronized. Delays were therefore constant and identical in all trials with respect to

the technical equipment. The onset for each trial was identified manually from the spectrogram of the audio signals computed with the free software Audacity®.

2.3. *Tasks*

The first subject (53-year-old male, Italian native speaker, hereinafter subject1) performed two sessions of a naming task, the same conducted during the presurgical fMRI to identify eloquent areas. The task consisted in naming different images shown on a screen. Each session consisted of three blocks during which 10 pictures representing Italian nouns were presented. The order of the stimuli is shown in Fig. 1 and was not randomized across blocks because of limitations of the equipment available in the surgery room. During the first session, neuronal signals were recorded using the Epi array (Epi dataset, 30 trials) and during the second session data were collected using the MuSA array (MuSA dataset, 30 trials).

The second subject (41-year-old female, Italian native speaker, hereinafter referred as subject2) performed a phoneme production task (see Fig. 1). The task consisted in listening to different phonemes and in repeating them. In this task, differently from the naming one, the stimuli were randomized across blocks and neuronal signals were recorded with the EpiBig device (EpiBig dataset, 84 trials).

2.4. *Characterization of the signal redundancy across electrodes*

Data were analyzed in Matlab (version 9.5, Mathworks, Inc., Natick, MA) with the aim of characterizing the spatiotemporal dynamic of neural activity related to speech preparation. Ground-truth speech onset times were determined based on the subjects' voices recorded during the experiment. We focused our analysis on three frequency bands: beta (15–30 Hz), low-gamma (30–60 Hz), and high-gamma (70–150 Hz).[22] Signals were filtered in these three bands, by applying the Matlab function *filtfilt* to minimize phase distortion (8th order Butterworth). We also removed line noise by applying a notch filter at 50 Hz and its harmonics up to 200 Hz. Finally, the filtered data were segmented into trials, spanning from 500 ms before to 500 ms after speech onset, and analyzed as follows.

2.4.1. *Spatial correlation analysis*

We used a correlation analysis to quantify signal redundancy across electrodes. To this end, we computed the correlation coefficient of the filtered and segmented signals for each pairwise combination of electrodes and averaged it across trials. Then, we averaged the results across electrodes sharing the same distance. The correlation decay was computed from data recorded with the Epi matrix, as this probe possesses the smallest distance between electrodes (600 μm) and thus the highest spatial resolution.

2.4.2. *Spectrograms*

Spectrograms were computed using the Matlab function spectrogram setting a temporal window of 100 ms for low and high gamma, and 150 ms for the beta band. The overlap between windows was set to 90%. The frequency resolution was set to 1 Hz for low and high gamma bands, and to 0.5 Hz for the beta band. Power spectra were then averaged across trials.

2.5. *Prediction of speech onset*

To identify speech-preparation activities, we first segmented each recording session into N non-overlapping intervals, where N represents the number of words or phonemes (hereinafter vocalization), according to the task performed. Each interval ranged from 500 ms before an instance of speech onset to 500 ms before the subsequent one. It thus contained only one vocalization. For each interval, we extracted vectors of features from neuronal signals and labeled them either as preparation or non-preparation. Specifically, we labelled as "preparation" features extracted in the 500 ms preceding a speech onset event and "non-preparation" features extracted in all other time intervals (Fig. 2). For each channel, we then trained a support vector machine (SVM) to classify feature vectors based on their assigned labels. Figure 3 reports a diagram of the prediction procedure.

2.5.1. *Feature extraction*

Features for the SVM were extracted from the high-gamma range spectrograms. To this end, we first averaged spectrograms across frequency thus obtaining a single time-varying profile of the power spectral density (hereinafter Mean Power Profile, MPP)

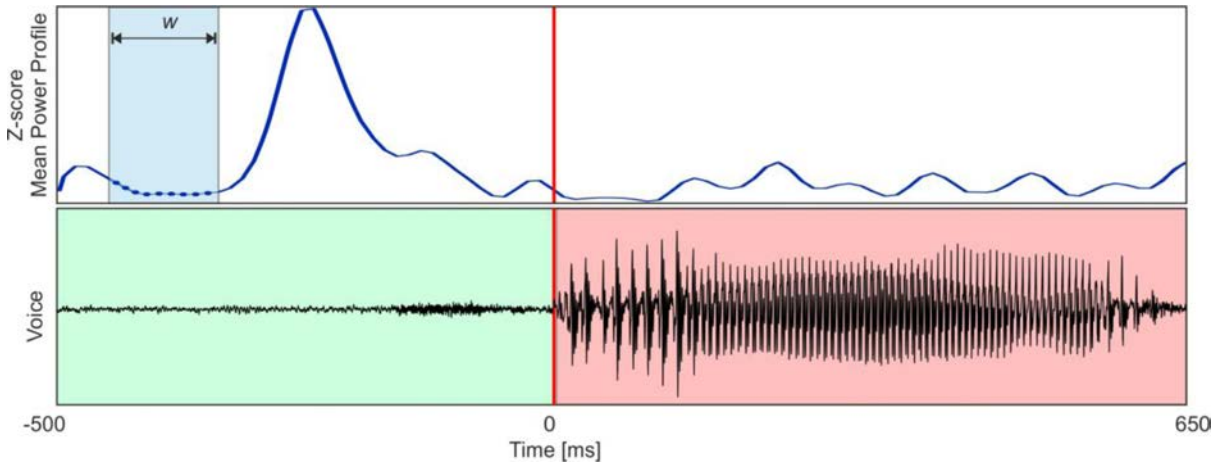

Fig. 2. (Color online) Graphical representation of the feature extraction and labeling procedure. Each consecutive and non-overlapping window (w) of the z-score Mean Power Profile (MPP, (blue line) was considered as an observation. Observations were labeled as preparation within 500 ms before the speech onset (vertical red line) were labeled as preparation (class 0, in green). Observations belonging to the vocalization interval and the following silence were labeled as non-preparation (class 1, in red).

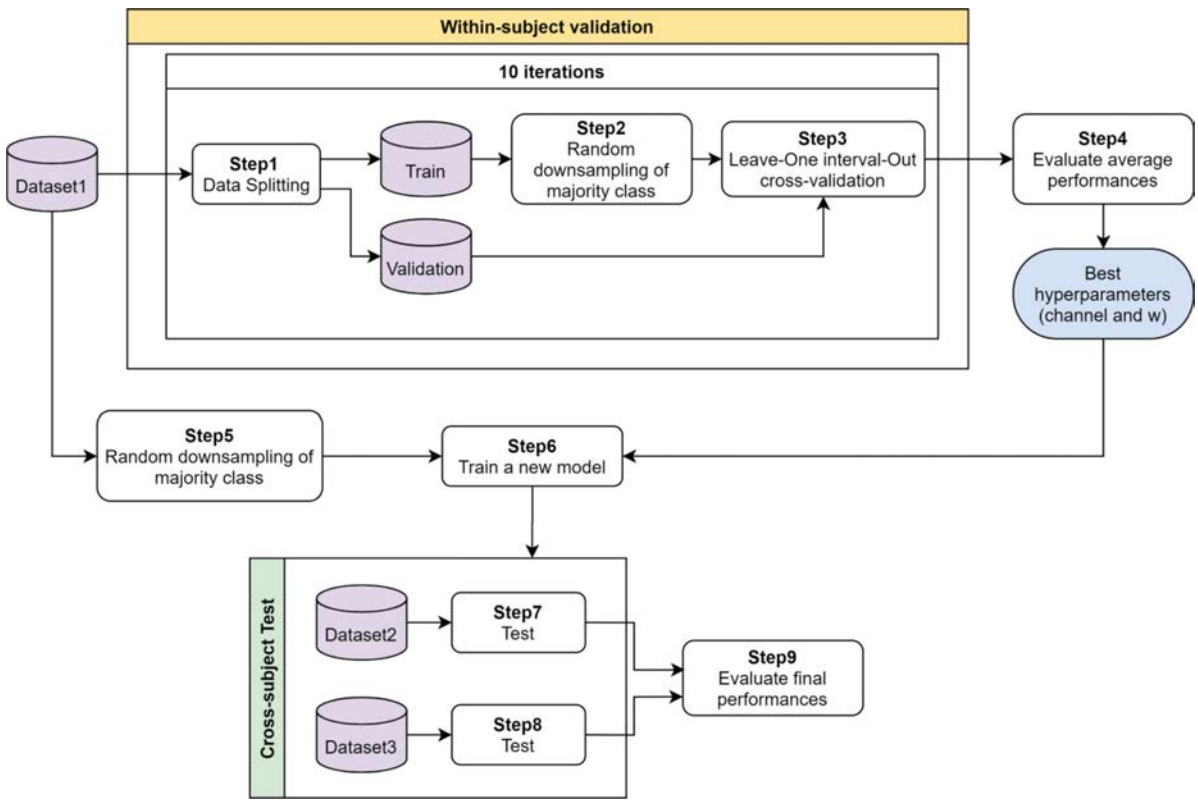

Fig. 3. (Color online) Training procedure of our classifier. (Top) Within-subject validation. (Step 1) Each recording session was segmented into N intervals, where N represents the number of vocalizations. Data where then split into train and validation set. (Step 2) Random down-sampling of the more represented class (i.e. "non-preparation") to train our classifier with balanced classes. (Step 3) Training of the classifier with all intervals except one. The left-out interval was used for validation. To test the robustness against the random downsampling, this procedure was iterated 10 times and performances were then averaged. This procedure yielded the optimal hyperparameters for our model. (Step 6) The classifier with the optimal hyperparameters was trained using the whole dataset and (Steps 7–8) tested using a cross-subject approach.

for each channel. Since the MPP is the average of the power spectrum values for each bin computed, the time resolution of the MPP is the same of the high-gamma spectrograms. Then, we segmented the data into intervals containing only one vocalization. Z-score normalization was applied to compare data recorded from different devices and subjects. Each feature vector consisted of w consecutive samples of the z-score MPP, with no overlap between consecutive vectors (see Fig. 2). Thus, w is the parameter that determines how many features are included in each observation. We tested for each dataset and channel independently different window lengths (w), specifically: 36, 60, 84, 108, 132, 156 milliseconds.

2.5.2. *Classification approach*

Considering each channel separately, we used a support-vector machine (SVM), which is a supervised learning method typically used for the classification of observations that cannot be linearly separable in their space. SVMs have been widely used in biomedical research to decode speech or its related features directly from neural signals.[25,27,29,42,43] Here, we trained a set of SVM models to classify observations as preparation or not-preparation using the Matlab function fitcsvm for a two-class (binary) problem. This function supports mapping the predictor data using kernel functions.

We used a Gaussian kernel (or Radial Basis Function, RBF), already employed for speech detection from ECoG signals,[27,28] with a fine kernel scale. The software divides all elements of the predictor matrix by the value of Kernel Scale and applies a Box Constraint that controls the maximum penalty imposed on margin-violating observations, which helps to prevent overfitting (regularization). Both values were set to 1 as default value.

In our experiments, speech periods were separated by longer intervals in which the patients remained silent. Our dataset contained thus a significantly greater number of "non-preparation" than "preparation" feature vectors. To reduce the skewness in our data and properly train our classifiers we randomly down-sampled them in order to get balanced classes (Fig. 3, Steps 2–5). Then we used a Leave-One interval-Out validation to select the optimal combination of window length w and most informative channel. Our classifier was trained using feature vectors belonging to all the intervals except one (Fig. 3, *Step* 3) and tested using all feature vectors belonging to the left-out interval. Since the non-preparation class was randomly down-sampled for the training, this procedure was repeated ten times for each left-out interval. Validation performances were obtained by averaging across the 10 randomizations (Fig. 3, *Step* 4).

2.5.3. *Performance evaluation*

To find the optimal value of the window length w, we assessed the performance of each model by means of F-score index. This index is defined as the harmonic mean of Precision and Recall and is specifically designed to deal with imbalanced datasets in which one label (i.e. non-preparation) is significantly more represented than the other (i.e. preparation).[44,45] To estimate an empirical chance level for the F-score, we used a Monte Carlo approach in which we trained our classifier on a data set with shuffled feature labels. The empirical chance level was defined as the average F-score across 10 shuffling (Fig. 3, *Steps* 4–9). For each dataset, we identified the best combination of window length w_o and channel number *cho* as that yielding the highest and above chance F-score.

2.5.4. *Cross-dataset model testing*

For the purpose of cross-dataset model testing, we first trained a new model on channel *cho* using window length w_o (Fig. 3, *Step* 6). We then tested this model on all channels of the other datasets. This procedure was repeated for all pairwise combinations of datasets (Fig. 3, *Steps* 7–8). The number of observations divided by class, for each dataset, before and after the dowsampling of the majority class ("non-preparation") are reported in Table 1.

Table 1. Number of observations divided by class, for each dataset, before (BDs) and after (ADs) dowsampling.

	Epi	EpiBig	MuSA
Preparation	150	320	441
Non-Preparation BDs	613	2494	1884
Non-Preparation ADs	150	320	441
Training set dimension	**300**	**640**	**882**

3. Results

3.1. *Characterization of the signal redundancy across electrodes*

As a first step, we studied the degree of signal redundancy between electrodes. Figures 4(a)–4(c) show the mean correlation coefficients computed from the signals recorded from the Epi array for the three frequency bands (beta: 15–30 Hz; low-gamma: 30–60 Hz; high-gamma: 70–150 Hz). We selected this probe as it has the narrowest pitch (0.6 mm). As expected, there was a clear trend in the spatial extent of the correlations. Specifically, the high-gamma band (Fig. 4(c)) exhibited correlations at a narrower spatial scale than the low-gamma band (Fig. 4(b)), whose spatial correlations were narrower than those in the beta band (Fig. 4(a)). To quantitatively study this trend, we computed the average correlation coefficients as a function of the distance between electrodes. Results in Fig. 4(d) show that (1) the correlation coefficient decreases with the increasing distance between electrodes and (2)

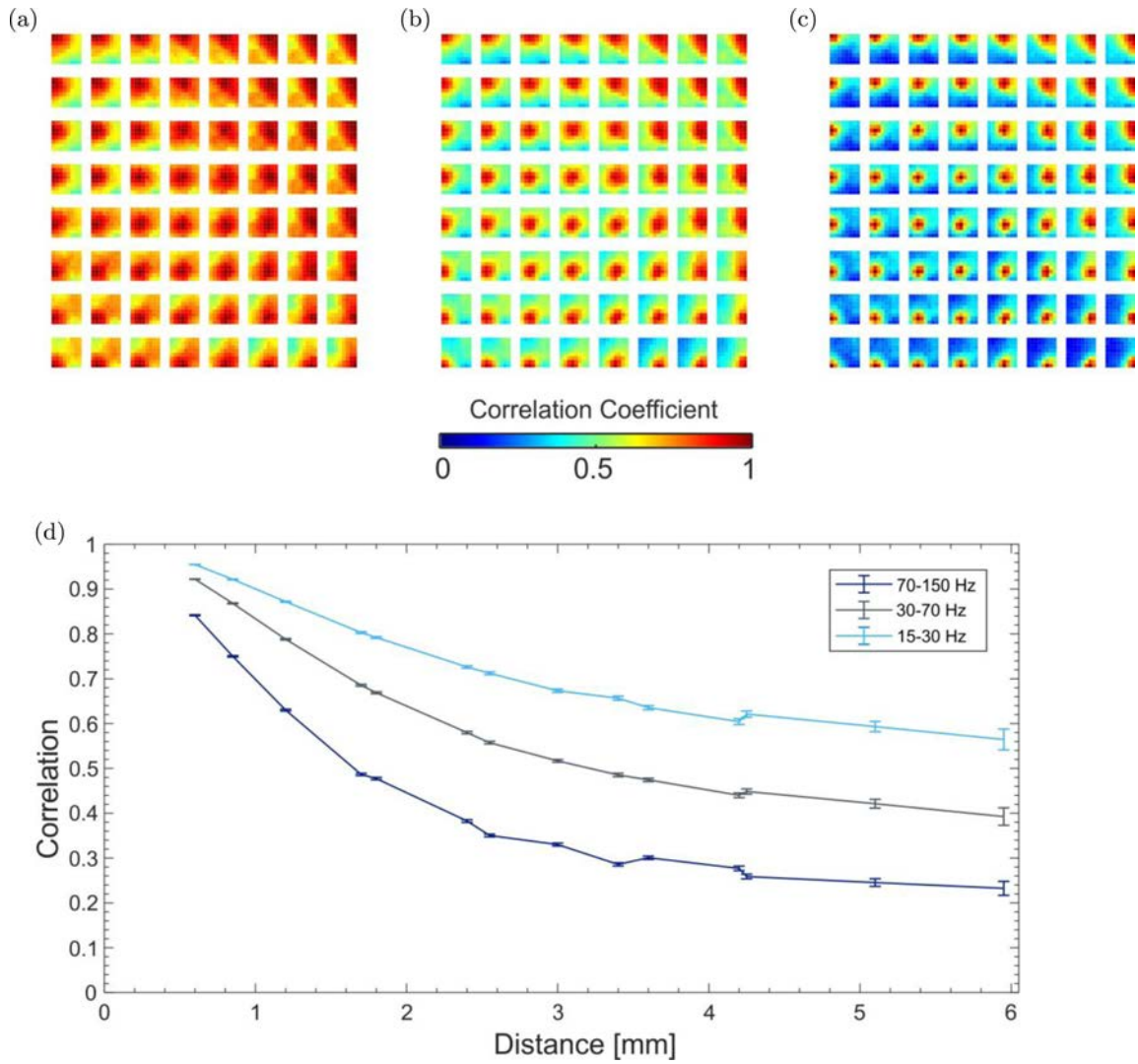

Fig. 4. (Color online) Characterization of the signal redundancy across electrodes performed on the Epi dataset. (a–c) Mean correlation maps of signals in the beta (15–30 Hz, panel (a)), low-gamma (30–60 Hz, panel (b)), and high-gamma (70–150 Hz, panel (c)) frequency bands obtained averaging across trials. Each square of the plot represents the correlation coefficients computed for the electrode in that position against all the others. (d) Correlation profiles (mean ± SE) obtained averaging the correlation coefficients of electrodes sharing the same distance for all the tested frequency bands (light blue for beta, grey for low-gamma, and dark blue for high-gamma).

higher frequencies consistently yield lower correlation coefficients. Notably, at the smallest considered pitch distance of 0.6 mm, signals were highly correlated, and thus redundant, in all three considered frequency bands (correlation coefficient > 0.8). This result suggests a lower bound for the pitch distance, as it shows that values smaller than 0.6 mm would provide low gain in the amount of information provided by nearby recorded signals.

3.2. *Prediction of speech onset*

We next performed a time-frequency analysis of the recorded signals. Figures 5(a)–5(b) show the average across trials of the high-gamma spectrograms, aligned to the speech onset event. Results were computed from signals recorded from the Epi and MuSA probes implanted in the same patient (Subject 1). Panels A and B show a clear, time-localized increase in power few hundreds of milliseconds before the speech onset event in a subset of neighboring electrodes (channels enclosed in the rectangular frame). Interestingly, the spatial locations of electrodes in the Epi and MuSA arrays exhibiting such anticipatory activity were overlapping (Fig. 5(c)). The increase in power observed in the low-gamma and beta bands was not as equally precise in both time and space (Fig. 6).

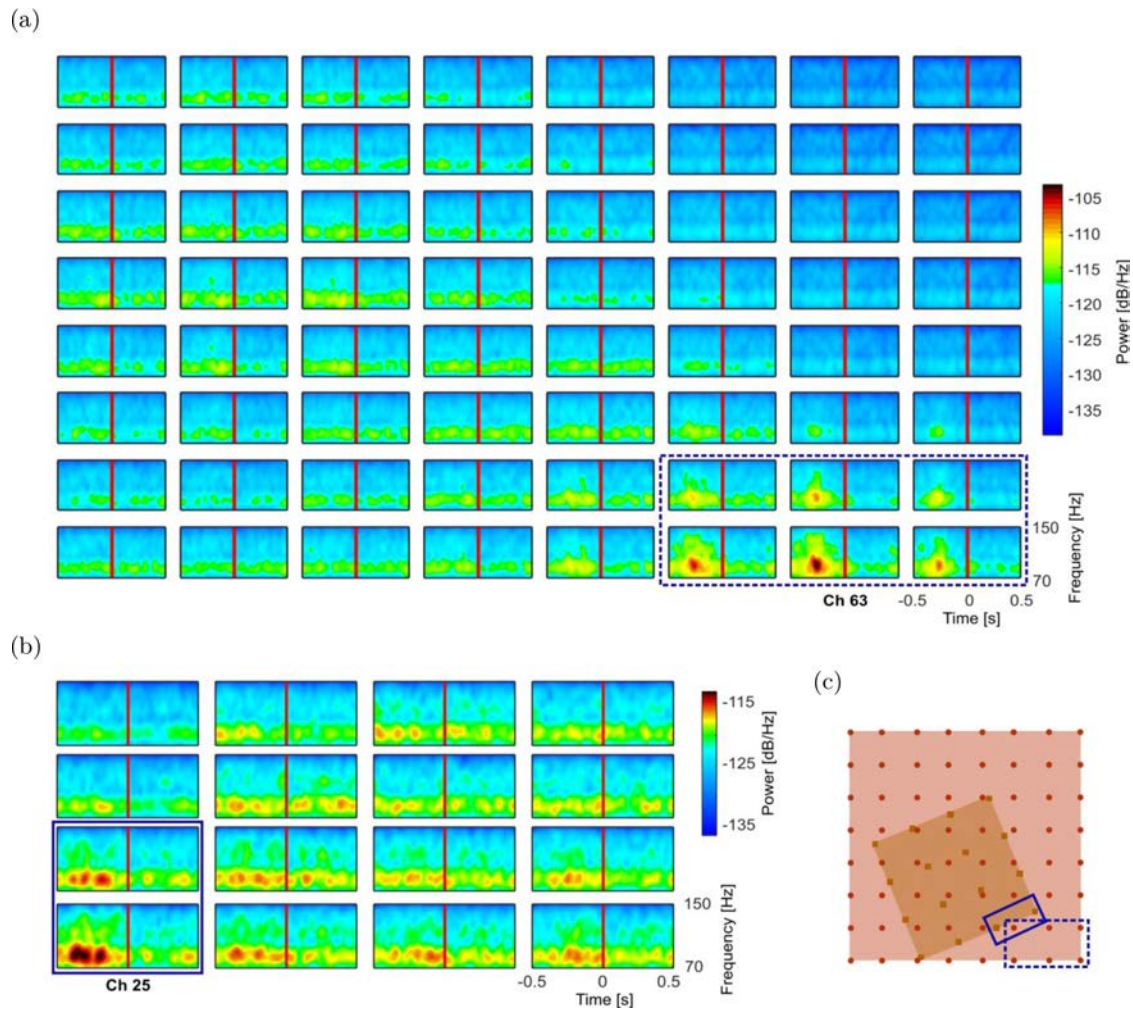

Fig. 5. (Color online) Mean spectrogram maps for the Epi (a) and the MuSA (b) arrays. Data are filtered in the high-gamma band (70–150 Hz) and averaged over trials. (c) Relative orientation on the cortex of the MuSA (light brown) and the Epi (red) devices. Blue rectangles refer to the electrodes highlighted on the spectrogram's plots (dashed line, Epi array; solid line, MuSA array).

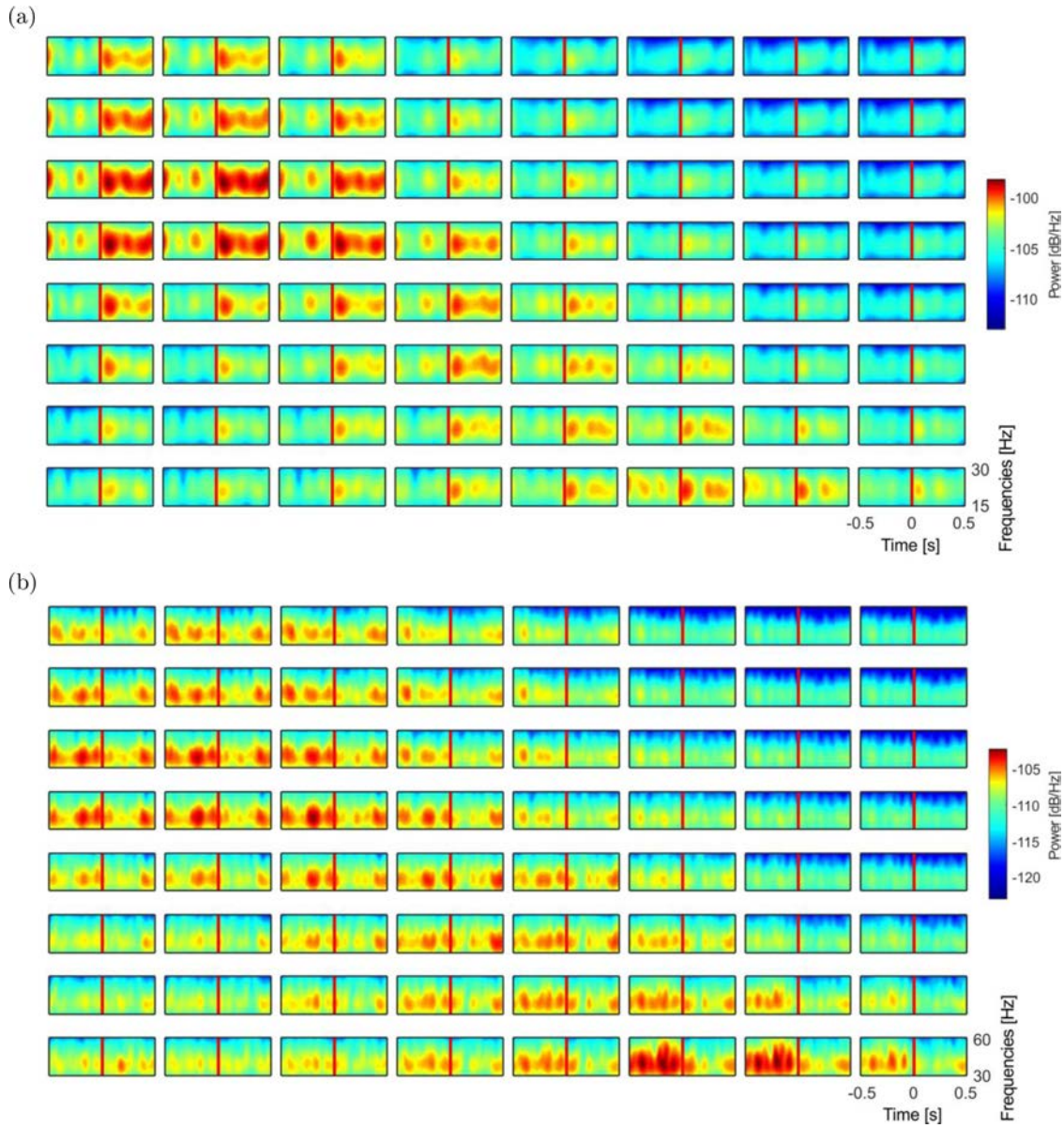

Fig. 6. (Color online) Mean spectrogram maps of the Epi array in the beta (15–30 Hz, panel (a)) and low-gamma (30–60 Hz, panel (b)) frequency bands. Data are averaged over trials aligned to the speech onset (vertical red line).

In this context, time-frequency and correlation analysis have been used to inform the feature selection process. Indeed, spectrograms were used to visualize the time alignment of the band-power increase, while the correlation maps provided a quantitative estimate of the spatial confinement of the signals in the different frequency bands. Results in Figs. 4–5 show that the high-gamma modulations were the only ones temporally confined prior to the speech onset and with high spatial specificity.

Consequently, we sought to investigate whether such increase in power was a reliable predictor of speech onset on a trial-by-trial basis. To this end, we trained a support vector machine (SVM) to classify a given time bin as belonging to a "preparation" or "non preparation" interval based on the spectral features of the signals. Model's hyperparameters were set by a leave-one-out approach and the classifier's performance was assessed by means of an F-score index (see Figs. 7, 8, and Sec. 2.5 for further

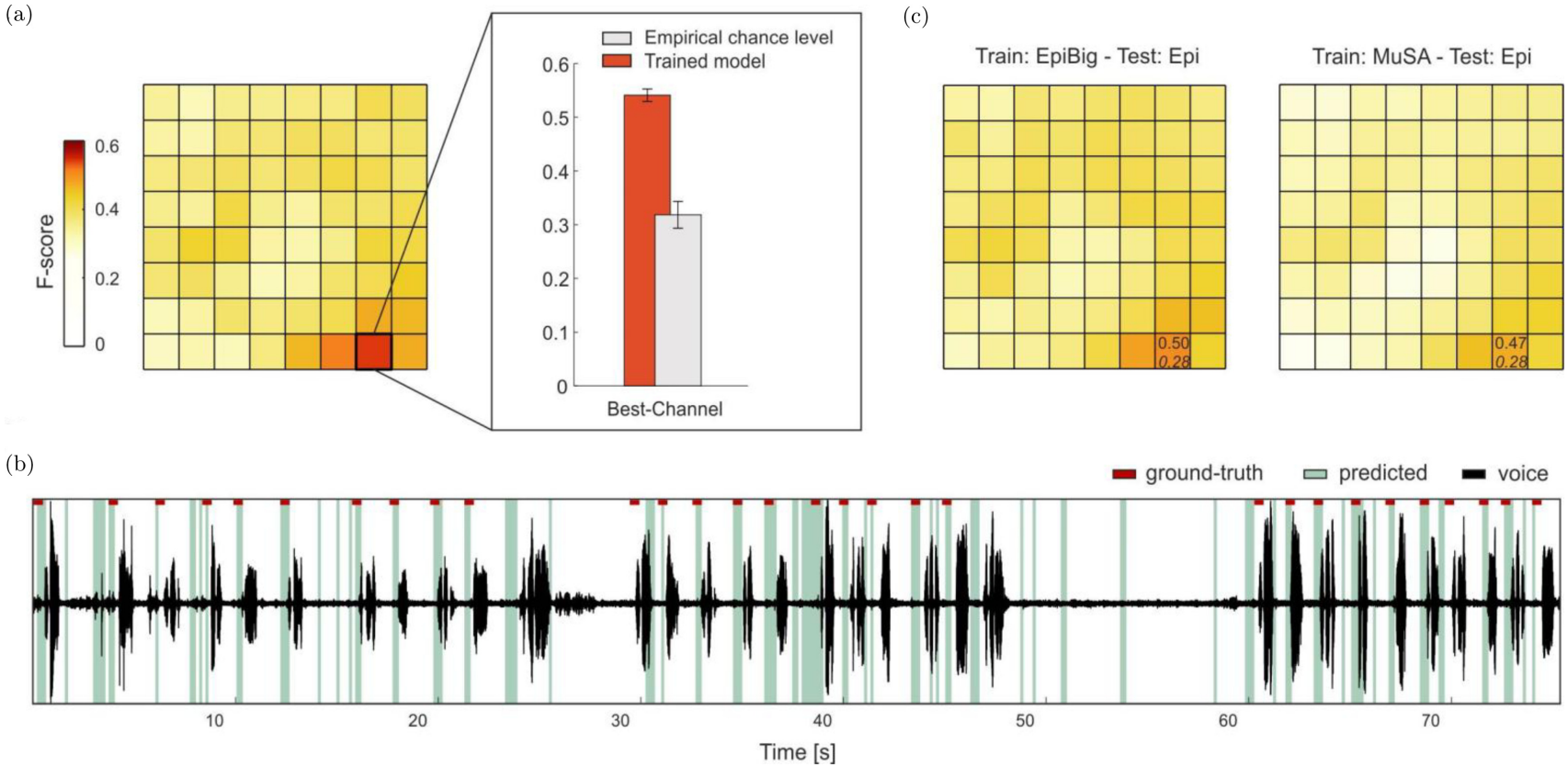

Fig. 7. (Color online) Prediction of speech onset. (a) On the left, the mean of 10 run F-score maps, obtained with the optimal window length tested for high-gamma MPP features of the Epi dataset (subject 1, naming task). On the right, the mean F-score of the best channel (red bar) with its standard deviation, is compared to the empirical chance level (grey bar). The non-random model resulted significantly higher (two-sided t-test, $P < 0.0001$) than the random one. (b) Predicted (light green bars) and ground-truth (red segments) speech preparation profiles are shown aligned with the voice of the subject (black signal). The reported predicted preparation intervals belong to the best channel of the Epi dataset. (c) The mean F-score maps, for the Epi dataset (subject1, naming task), obtained from the cross-dataset model testing; from left to right respectively the model were trained on the EpiBig (subject 2, phoneme task) and MuSA (subject 1, naming task) datasets. For the best channel, numeric values indicating the average F-score and the corresponding empirical chance level (italic) are reported. Interestingly, the device area where the models achieved the highest performances overlapped with the one resulting from the within-dataset validation, highlighting the robustness of the neural correlates decoded by the different models

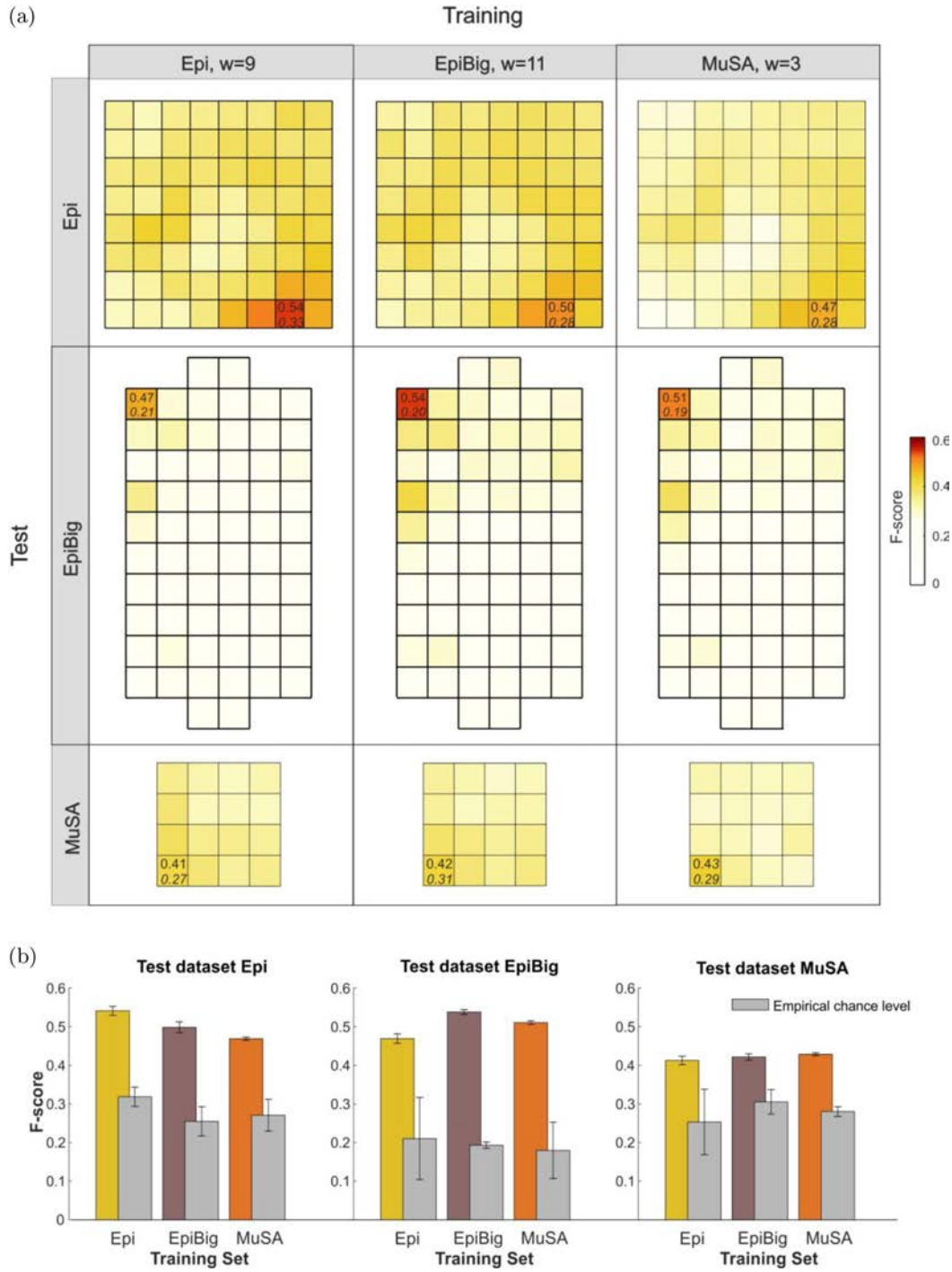

Fig. 8. (a) Average F-score maps obtained during cross-dataset model testing of each pairwise combination. Each training was performed considering for each dataset the best channel (cho) and window (wo) obtained during the hyperparameters optimization. Empirical chance level of the best channel is reported in italic. (b) Average F-scores of the best channels compared to the empirical chance level (gray bars). All the within-dataset models were significantly better than the randomized ones (diagonal terms, two-sided t-test, $P < 0.001$). All cross-dataset tests show significantly higher performances than the randomization test (off-diagonal terms, two-sided t-test, $P < 0.001$). Data are reported as mean $\pm$ SD.

details). To provide a quantitative comparison for our model's performance, we used a shuffling procedure to computationally estimate the chance level F-score (henceforth empirical chance level). The spatial distribution of the F-scores obtained when we trained and tested our classifier on the Epi dataset (Fig. 7(a)-left). Consistently with the results of Fig. 5(a), channels exhibiting a clear anticipatory increase in power in the high-gamma band also yielded classification performance significantly higher than the empirical chance level. This means that spectral features in the high-gamma band can reliably predict speech onset on a trial-by-trial basis as further demonstrated by the temporal prediction profile shown in Fig. 7(b). Here, the voice is aligned to segments detected as preparation by the best-channel classifier (in light green), as well as the ground-truth (in red).

Neuronal responses can be recorded with a variety of probes and in different subjects. Are the identified spectral features robust with respect to these factors? We investigated this issue by means of a cross-training approach in which we trained and tested our model on all pairwise combinations of datasets, respectively. Figure 7(c) shows the results obtained when our classifier was tested on the Epi dataset and trained on the EpiBig (left panel) and Musa (right panel) datasets (see Fig. 8 for all other combinations). Comparison of Figures 7(a) and 7(c) shows that, irrespective of the training dataset, our model yielded higher than the empirical chance level classification performance on electrodes at the same locations of the Epi probe (see Figs. 7(a)–7(c)). This result is particularly notable as the EpiBig and MuSA datasets used for training were recorded from two subjects and with two different types of probes. Taken together, the results of Fig. 7 show that activity in the high-gamma band is a reliable marker of speech onset that is robust across subjects and recording devices.

4. Discussion

In this study, we used dense μECoG arrays to record the epicortical signals from awake patients undergoing brain surgery. We first investigated the spatiotemporal specificity of the neural activity during speech production. Results in Fig. 4 show that, with an electrode pitch of 600 μm, the correlation between neighboring electrodes is greater than 0.8 in all investigated frequency bands. We next used a machine-learning based approach to show that high-gamma frequency signals (70–150 Hz) recorded from the speech arrest area are a reliable predictor of speech onset. These results are important in view of transitioning from offline to online speech brain computer interfaces (BCIs). A previous study reported the correlation profiles in epicortical recordings between electrode pairs with a pitch of 4 mm. The correlation trends showed increases with decreasing distance and that, at the minimum investigated electrode pitch, signals in the low-gamma and high-gamma bands are still largely uncorrelated, although differences and similarities might be affected by local anatomy, electrodes impedance, as well as the physical properties of the measured electric field.[22] These results suggested that arrays with electrode pitch smaller than 4 mm were promising solutions for increasing signal resolution at high frequencies. While valuable, this study provided however no lower bound for the electrode pitch. Here, we leveraged recent advances in array design[13,39–41] to experimentally assess, for the first time, the redundancy between the activities of submillimeter spaced electrodes in the beta, low-gamma, and high-gamma frequency bands. Results in Fig. 4 show that at a pitch distance of 600 μm the correlation coefficient between neighboring electrodes is greater than 0.8 in all investigated frequency bands. This result is of fundamental relevance for the design of future probes. Indeed, it suggests that this distance should be considered a lower bound for the pitch between electrodes as more densely spaced electrodes would accrue low additional information at the cost, however, of higher design, manufacturing and computational costs. In addition to being spatially confined, high-gamma neural modulations in Broca's area are known to be specifically elicited by language production.[35,38]

Results in Fig. 5 show that these modulations have a clear anticipatory nature, as they consistently increase few hundreds of milliseconds before speech onset. To deploy an effective *online* speech BCI, the detection of the speech onset is of crucial importance. Indeed, without such knowledge, an online speech BCI would constantly attempt to convert neuronal activations into words, even during periods of silence, with consequently higher error

rates than in a controlled task that would render the BCI practically useless. A reliable detector of speech preparation would allow instead to trigger the decoding procedure in advance, and only when neuronal signals are effectively related to speech encoding, and bypassing the auditory feedback.[27,28,31] Indeed, our findings provide neuronal markers that are predictive of speech onset, and highly correlated with speech preparation processes, thus present irrespective of the actual emission of speech. This predictive biomarker could play a key role in view of a real-time sBCI since it would allow to trigger the decoding process. Recent studies confirmed that speech can be offline synthesized starting from ECoG signals[1,2] but the deployment of analogous online models in clinical application has not been achieved yet.

Here, aiming to support the transition from offline to online speech BCIs, we trained a support-vector machine classifier to recognize speech-related motor preparation on a per-channel basis. The performances obtained during validation confirmed that the high-gamma activity was indeed well-suited (Figs. 7 and 8). More importantly, especially for translational applications, the spatial maps of the averaged F-scores were highly consistent when the classifier was tested on data recorded from different patients with different devices, executing different experimental tasks (Figs. 7 and 8). Indeed, the best-performing channels obtained during cross-dataset model testing were spatially coherent with those found during the within-dataset validation. This result demonstrates that the model was able to generalize both across different probes and patients. This is of critical relevance if we imagine that in real-life settings, patients would be using a chronically implanted BCI when they already have lost speech production abilities (i.e. no labeled training data would be available).[23]

In this study, we aimed to demonstrate the feasibility of speech onset detection in a clinical context. That is in a condition where few trials are typically available, preventing thus the use of complex models. Nevertheless, improvement and better tuning of the decoding algorithms are crucial points that should be continuously pursued. While significantly higher-than-chance performances have been already obtained with our per-channel paradigm, in the future it would be worth exploring the possibility of pooling single channel classifiers by means of "mixture of experts" approach. Indeed, although our correlation analysis indicates that most of the information is shared between neighboring channels (correlation coefficient is higher than 0.8 for the high-gamma band), a multi-channel paradigm could significantly improve the decoding performances. Future studies will also have to include more subjects and optimize the algorithm selection, potentially exploring more powerful machine learning approaches[46–50] and methods which are able to better deal with imbalanced datasets.[51]

5. Conclusion

To the best of our knowledge, this study is the first one using acutely implanted μECoG grids to investigate speech onset in Broca's area.

Some methodological advancements allowed us to find two novel and, in our view, important results. First, electrodes separated by shorter distances than 600 μm would likely provide, at least when data is analyzed in the frequency domain, a lot of redundant information so as not to justify their design and manufacturing costs. To establish whether 600 μm represents a lower bound for the electrode pitch or whether a multi-electrode approach would lead to better results, further investigations are necessary. Second, high-gamma oscillations represent a reliable signature of speech onset that is robust across both recording devices and subjects. These results provide critical information for the design of future real-time speech BCI that are suitable for chronic long-term implant.

References

1. G. K. Anumanchipalli, J. Chartier and E. F. Chang, Speech synthesis from neural decoding of spoken sentences, *Nature* **568** (2019) 493–498.
2. 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, *J. Neural Eng.* **16** (2019) 036019.
3. A. Ortiz-Rosario and H. Adeli, Brain-computer interface technologies: From signal to action, *Rev. Neurosci.* **24**(5) (2013) 537–552.
4. Q. Rabbani, G. Milsap and N. E. Crone, The potential for a speech brain–computer interface using chronic electrocorticography, *Neurotherapeutics* **16** (2019) 144–165.
5. S. Martin, J. D. R. Millán, R. T. Knight and B. N. Pasley, The use of intracranial recordings to

decode human language: Challenges and opportunities, *Brain Lang.* **193** (2019) 73–83.
6. N. J. Hill, D. Gupta, P. Brunner, A. Gunduz, M. A. Adamo, A. Ritaccio and G. Schalk, Recording human electrocorticographic (ECoG) signals for neuroscientific research and real-time functional cortical mapping, *J. Vis. Exp.* **64** (2012) 1–5.
7. Y. Zhu, J. Liu, T. Ristaniemi and F. Cong, Distinct patterns of functional connectivity during the comprehension of natural, narrative speech, *Int. J. Neural Syst.* **30**(3) (2020) 2050007.
8. M. Mozaffari Legha and H. Adeli, Visibility graph analysis of speech evoked auditory brainstem response in persistent developmental stuttering, *Neurosci. Lett.* **696** (2019) 28–32.
9. G. Hong and C. M. Lieber, Novel electrode technologies for neural recordings, *Nat. Rev. Neurosci.* **20** (2019) 330–345.
10. K. M. Szostak, L. Grand and T. G. Constandinou, Neural interfaces for intracortical recording: Requirements, fabrication methods, and characteristics, *Front. Neurosci.* **11** (2017) 665.
11. R. Chen, A. Canales and P. Anikeeva, Neural recording and modulation technologies, *Nat. Rev. Mater.* **2** (2017) 1–16.
12. G. Buzsáki, C. A. Anastassiou and C. Koch, The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes, *Nat. Rev. Neurosci.* **13** (2012) 407–420.
13. M. Vomero, M. F. Porto Cruz, E. Zucchini, F. Ciarpella, E. Delfino, S. Carli, C. Boehler, M. Asplund, D. Ricci, L. Fadiga and T. Stieglitz, Conformable polyimide-based μECoGs: Bringing the electrodes closer to the signal source, *Biomaterials* **255** (2020) 120178.
14. T. Bockhorst, F. Pieper, G. Engler, T. Stieglitz, E. Galindo-Leon and A. K. Engel, Synchrony surfacing: Epicortical recording of correlated action potentials, *Eur. J. Neurosci.* **48** (2018) 3583–3596.
15. L. Luan, X. Wei, Z. Zhao, J. J. Siegel, O. Potnis, C. A. Tuppen, S. Lin, S. Kazmi, R. A. Fowler, S. Holloway, A. K. Dunn, R. A. Chitwood and C. Xie, Ultraflexible nanoelectronic probes form reliable, glial scar–free neural integration, *Sci. Adv.* **3** (2017) e1601966.
16. A. Weltman, J. Yoo and E. Meng, Flexible, penetrating brain probes enabled by advances in polymer microfabrication, *Micromachines* **7** (2016) 180.
17. I. R. Minev *et al.*, Electronic dura mater for long-term multimodal neural interfaces, *Science* **347** (2015) 159–163.
18. J. Viventi *et al.*, Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity *in vivo*, *Nat. Neurosci.* **14** (2011) 1599–1605.
19. D.-H. Kim, J. Viventi, J. J. Amsden, J. Xiao, L. Vigeland, Y.-S. Kim, J. A. Blanco, B. Panilaitis, E. S. Frechette, D. Contreras, D. L. Kaplan, F. G. Omenetto, Y. Huang, K.-C. Hwang, M. R. Zakin, B. Litt and J. A. Rogers, Dissolvable films of silk fibroin for ultrathin conformal bio-integrated electronics, *Nat. Mater.* **9** (2010) 511.
20. R. Biran, D. C. Martin and P. A. Tresco, Neuronal cell loss accompanies the brain tissue response to chronically implanted silicon microelectrode arrays, *Exp. Neurol.* **195** (2005) 115–126.
21. N. Rogers, J. Hermiz, M. Ganji, E. Kaestner, K. Kılıç, L. Hossain, M. Thunemann, D. R. Cleary, B. S. Carter, D. Barba, A. Devor, E. Halgren, S. A. Dayeh and V. Gilja, Correlation structure in micro-ECoG recordings is described by spatially coherent components, *PLoS Comput. Biol.* **15**(2) (2019) e1006769.
22. L. Muller, L. S. Hamilton, E. Edwards, K. E. Bouchard and E. F. Chang, Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography, *J. Neural Eng.* **13** (2016) 56013.
23. S. Martin, I. Iturrate, J. del R. Millán, R. T. Knight and B. N. Pasley, Decoding inner speech using electrocorticography: Progress and challenges toward a speech prosthesis, *Front. Neurosci.* **12** (2018) 422.
24. A. R. Sereshkeh, R. Trott, A. Bricout and T. Chau, Online EEG classification of covert speech for brain–computer interfacing, *Int. J. Neural Syst.* **27** (2017) 1750033.
25. S. Martin, P. Brunner, I. Iturrate, J. del R. Millán, G. Schalk, R. T. Knight and B. N. Pasley, Word pair classification during imagined speech using direct brain recordings, *Sci. Rep.* **6** (2016) 25803.
26. S. Martin, P. Brunner, C. Holdgraf, H. Heinze, N. Crone, J. Rieger, G. Schalk, R. Knight and B. Pasley, Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front. Neuroeng.* **7** (2014) 14.
27. V. G. Kanas, I. Mporas, H. L. Benz, K. N. Sgarbas, A. Bezerianos and N. E. Crone, Joint spatial-spectral feature space clustering for speech activity detection from ECoG signals, *IEEE Trans. Biomed. Eng.* **61**(4) (2014) 1241–1250.
28. V. G. Kanas, I. Mporas, H. L. Benz, K. N. Sgarbas, A. Bezerianos and N. E. Crone, Real-time voice activity detection for ECoG-based speech brain machine interfaces, in 2014 19*th Int. Conf. Digital Signal Processing* (Hong Kong, China, 2014), pp. 862–865.
29. D. A. Moses, M. K. Leonard, J. G. Makin and E. F. Chang, Real-time decoding of question-and-answer speech dialogue using human cortical activity, *Nat. Commun.* **10**(1) (2019) 1–14.
30. F. H. Guenther, J. S. Brumberg, E. J. Wright, A. Nieto- Castanon, J. A. Tourville, M. Panko, R. Law, S. A. Siebert, J. L. Bartels, D. S. Andreasen, P. Ehirim, H. Mao and P. R. Kennedy, A wireless

brain-machine interface for real-time speech synthesis, *PLoS One* **4** (2009) e8218.
31. D. Dash, P. Ferrari, S. Dutta and J. Wang, NeuroVAD: Real-time voice activity detection from non-invasive neuromagnetic signals, *Sensors* **20** (2020) 2248.
32. E. F. Chang, J. D. Breshears, K. P. Raygor, D. Lau, A. M. Molinaro and M. S. Berger, Stereotactic probability and variability of speech arrest and anomia sites during stimulation mapping of the language dominant hemisphere, *J. Neurosurg.* **126** (2017) 114–121.
33. E. Mandonnet, S. Sarubbo and H. Duffau, Proposal of an optimized strategy for intraoperative testing of speech and language during awake mapping, *Neurosurg. Rev.* **40** (2017) 29–35.
34. M. C. Tate, G. Herbet, S. Moritz-Gasser, J. E. Tate and H. Duffau, Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited, *Brain* **137** (2014) 2773–2782.
35. V. Ferpozzi, L. Fornia, M. Montagna, C. Siodambro, Castellano, P. Borroni, M. Riva, M. Rossi, F. Pessina, L. Bello and G. Cerri, Broca's area as a pre-articulatory phonetic encoder: Gating the motor program, *Front. Hum. Neurosci.* **12** (2018) 64.
36. P. Gómez-Vilda, A. Gómez-Rodellar, J. M. Ferrández Vicente, J. Mekyska, D. Palacios-Alonso, V. Rodellar-Biarge, A. Álvarez-Marquina, I. Eliasova, M. Kostalova and I. Rektorova, Neuromechanical modeling of articulatory movements from surface electromyography and speech formants, *Int. J. Neural Syst.* **29**(2) (2019) 1850039.
37. A. Flinker, A. Korzeniewska, A. Y. Shestyuk, P. J. Franaszczuk, N. F. Dronkers, R. T. Knight and N. E. Crone, Redefining the role of Broca's area in speech, *Proc. Natl. Acad. Sci.* **112** (2015) 2871–2875.
38. X. Pei, E. C. Leuthardt, C. M. Gaona, P. Brunner, J. R. Wolpaw and G. Schalk, Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition, *Neuroimage* **54** (2011) 2960–2972.
39. I. Rembado, E. Castagnola, L. Turella, T. Ius, R. Budai, A. Ansaldo, G. N. Angotzi, F. Debertoldi, D. Ricci, M. Skrap and L. Fadiga, Independent component decomposition of human somatosensory evoked potentials recorded by micro-electrocorticography, *Int. J. Neural Syst.* **27** (2016) 1650052.
40. E. Castagnola, A. Ansaldo, E. Maggiolini, G. N. Angotzi, M. Skrap, D. Ricci and L. Fadiga, Biologically compatible neural interface to safely couple nanocoated electrodes to the surface of the brain, *ACS Nano* **7** (2013) 3887–3895.
41. E. Castagnola, A. Ansaldo, E. Maggiolini, T. Ius, M. Skrap, D. Ricci and L. Fadiga, Smaller, softer, lower-impedance electrodes for human neuroprosthesis: A pragmatic approach, *Front. Neuroeng.* **7** (2014) 8.
42. Z. Wang, A. Gunduz, P. Brunner, A. Ritaccio, Q. Ji and G. Schalk, Decoding onset and direction of movements using electrocorticographic (ECoG) signals in humans, *Front. Neuroeng.* **5** (2012) 15.
43. W. Wang, A. D. Degenhart, G. P. Sudre, D. A. Pomerleau and E. C. Tyler-Kabara, Decoding semantic information from human electrocorticographic (ECoG) signals, in 2011 *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 6294–6298.
44. Y. Sun, A. Wong and M. Kamel, Classification of imbalanced data: A review, *Int. J. Pattern Recognit. Artif. Intell.* **23** (2009) 687–719.
45. T. Saito and M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, *PLoS One* **10** (2015) e011843.
46. M. Ahmadlou and H. Adeli, Enhanced probabilistic neural network with local decision circles: A robust classifier, *Integr. Comput. Aided Eng.* **17**(3) (2010) 197–210.
47. M. H. Rafiei and H. Adeli, A new neural dynamic classification algorithm, *IEEE Trans. Neural Netw. Learn. Syst.* **28**(12) (2017) 3074–3083.
48. D. R. Pereira, M. A. Piteri, A. N. Souza, J. Papa and H. Adeli, FEMa: A finite element machine for fast learning, *Neural Comput. Appl.* **32**(10) (2020) 6393–6404.
49. K. M. Rokibul Alam, N. Siddique and H. Adeli, A dynamic ensemble learning algorithm for neural networks, *Neural Comput. Appl.* **32**(10) (2020) 6393–6404.
50. T. Hirschauer, H. Adeli and T. Buford, Computer-aided diagnosis of Parkinson's disease using an enhanced probabilistic neural network, *J. Med. Syst.* **39**(11) (2015) 1–12.
51. Manohar (2021). SMOTE (Synthetic Minority Over-Sampling Technique). Available at https://www.mathworks.com/matlabcentral/fileexchange/38830-smote-synthetic-minority-over-sampling-technique, MATLAB Central File Exchange [retrieved on February 22, 2021].