Toward A database of intracranial electrophysiology during natural language presentation

Toward a Database of Intracranial Electrophysiology during Natural Language Presentation

Erik Kaestner[1], Adam Milton Morgan[2], Joseph Snider[3], Meilin Zhan[4], Xi Jiang[1], Roger Levy[4], Victor S. Ferreira[2], Thomas Thesen[5], Eric Halgren[1,6]

[1]Department of Neurosciences, University of California at San Diego, La Jolla, California

[2]Department of Psychology, University of California at San Diego, La Jolla, California

[3]Institute for Neural Computation, University of California at San Diego, La Jolla, California

[4]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

[5]Department of Neurology, New York University Comprehensive Epilepsy Center, New York, New York

[6]Department of Radiology, University of California at San Diego, La Jolla, California

Abstract

Intracranial electrophysiology (iEEG) studies using cognitive tasks contribute to the understanding of the neural basis of language. However, though iEEG is recorded continuously during clinical treatment, due to patient considerations task time is limited. To increase the usefulness of iEEG recordings for language study, we provided patients with a tablet pre-loaded with media filled with natural language, wirelessly synchronized to clinical iEEG. This iEEG data collected and time-locked to natural language presentation is particularly applicable for studying the neural basis of combining words into larger contexts. We validate this approach with pilot analyses involving words heard during a movie, tagging syntactic properties and verb contextual probabilities. Event-related averages of high-frequency power (70-170Hz) identified bilateral perisylvian electrodes with differential responses to syntactic class and a linear regression identified activity associated with contextual probabilities, demonstrating the usefulness of aligning media to iEEG. We imagine future multi-site collaborations building an 'intracranial neurolinguistic corpus'.

Introduction

Intracranial electrophysiology (iEEG) studies using cognitive tasks have provided important novel contributions as well as corroborations in understanding the link between human cognition and neural activity. For language, this has occurred across multiple levels of linguistic study such as sub-lexical phonetics (Leonard, Bouchard, Tang, & Chang, 2015; Mesgarani, Cheung, Johnson, & Chang, 2014) and orthography(Gaillard et al., 2006; Thesen et al., 2012), lexico-semantic processing (Chan et al., 2011; Halgren et al., 2015; Nobre, Allison, & McCarthy, 1994), and syntax (Nelson et al., 2017). However, though iEEG during clinical treatment is recorded 24 hours-a-day for ~7 days for each patient, due to patient and clinical considerations task time is often severely limited. For language, this can mean research projects can stretch over many years to collect enough data to investigate a single linguistic question. Despite this enormous effort on the part of the patients, researchers, and clinical team

this dataset may not be useful again for additional questions. To demonstrate a natural language approach to increase the usefulness of iEEG recordings for language study, we provided patients with a tablet pre-loaded with digital media containing an abundance of natural language (movies, podcasts, etc.) which wirelessly synchronizes media presentation and iEEG data with high temporal precision. Once collected, the language presented can be characterized in a multitude of ways to characterize language comprehension at every level of linguistic study.

Here we validate this approach by reporting a vertical slice of what a fully scaled-up effort would look like. First we detail the unobtrusive, wireless syncing system between the media a patient experiences while using our tablet and the recorded clinical iEEG data. Modern tablet computers are powerful enough to record with the temporal precision needed for iEEG studies, portable enough to remain with patients for days, and simple enough to encourage enjoyable patient use. When the patient chooses to watch a video, listen to a podcast or audiobook, or play a game the synchronization system will facilitate later neural data analysis. Second, we describe the approach used to label the natural language stimuli being presented. Third, we report traditional stimulus-locked analyses to verify that stimuli are accurately tagged and that our synchronization method works. Finally, we illustrate how potential approaches for making use of these datasets in a post-hoc manner would proceed, focusing in our example on contextual probability. We approximated contextual probability of verbs by using a probabilistic model (Levy, 2008). Specifically, we used a smoothed trigram language model trained on a movie dialog corpus that estimated the probability of each word given the preceding two words.

Together, these aims highlight the approach advocated herein. First, the collection of electrophysiology collected intracranially, time-locked with high temporal precision to natural language presentation. Second, the usefulness of this data for later efforts to answer nuanced and complex linguistic questions.

Methods

Syncing Equipment

Figure 1A is a schematic of the synchronizing triggers which are wirelessly sent from the tablet to the clinical recording equipment. To remain unobtrusive in the clinical environment, we designed a temporal synchronization scheme for the tablet that was wireless and had sub-millisecond precision. The wireless feature was vital for ease-of-use in the clinical environment where patient care has absolute priority. The tablet had both BlueTooth and WiFi wireless technologies embedded, but neither could reach sub-millisecond precision in their timing, so we turned to the ZigBee (IEEE 802.15.4) wireless protocol designed for the Internet of Things and specifically including remote control of devices where latency is intolerable. ZigBee is not currently built into commercially available tablets, but it is implemented in a small off-the-shelf USB dongle made by Digi International (XStick XU-A11) that is fully compatible with the tablet. To feed information into the clinical amplifiers, we used an Arduino Fio V3 (SparkFun DEV-11520) that easily integrates a Digi International XBee Series 1 module (XB24-API-001). The Fio V3 has an Atmega 32U4 microcontroller that is compatible with the Arduino programming environment. The digital out pins on the Fio were connected to pins 2-9 on a DB25 connector compatible with the clinical amplifiers.

On the software side, the ZigBee appeared as a standard serial port, e.g. COM4, and was easily accessed from the tablet, including by NeuroBehavioral Systems Presentation software and Python. We used these interfaces to directly measure the latency of the wireless connection between the XStick dongle (on the tablet) and the XBee Series 1 module (connected to the clinical amplifier). As a reference we used another Arduino board (Teensy 3.1) with a wired connection to the tablet that was previously verified to have negligible latency. To measure the latency, we programmed the tablet to send simultaneous signals to both the wired reference board and the wireless ZigBee connection. The reference board also listened for a pulse from the DB25 output of the wireless receiver. The difference between these times was the latency of the system at 9.8(+/- 0.1) ms (variability reported for this and the following section

is standard error). This is well within the required 1ms precision, and the consistent latency can easily be subtracted.

Given this latency, we are then in a position to align the tablet's clock with the clock in the neural recording. We do this with a random interval scheme. A short python script always runs in the background whenever the participant is logged into the tablet. Every few seconds the script sends a pulse to the neural recording amplifier and records the pulse-time with respect to the tablet's own high precision clock. The time between pulses is randomly selected from 1, 2, 3, 4, or 5 seconds. The random sequence has maximum entropy of all possible sequences which makes it the most robust to random drop out or the occasional pulse that might be delayed. The pulses from the two clocks are aligned offline using robust fitting techniques. This whole process adds less than a millisecond of variability.

Media Presentation

Movies, TV shows, podcasts, audiobooks, and games were pre-loaded on the Surface tablet computer supplied to the patients. To play movies on the tablet and synchronize the participant's experience with their neural recording, we modified the open source VLC media player (VLC 2.2, original source available at https://www.videolan.org/ and modified source at https://github.com/oldstylejoe/vlc-timed2/tree/timing_mod). Internally, VLC uses time stamps to align its video and audio streams. Each of the streams are rendered by separate processes that tend to drift over time. As audio drifts ahead (behind) of the video frame, the markers of the two streams deviate and video frames are skipped or repeated. This makes for a good viewing experience, but can cause problems with alignment of the patient's audiovisual experience with the electrophysiology since a video frame comes about every 40ms (24Hz). To compensate for this, we modified the rendering code at the level of the Microsoft direct3d driver to record which frame is being displayed and when the frame was presented. The frame number is the time (in microseconds) since the beginning of the movie that the audio or video frame was recorded. The presentation time is recorded using the tablet's internal high precision clock. A limited

interface allowed the participant to open movies, play/pause them, and seek locations, while we could still record exactly what was on the screen and playing on the speakers at each instant.

The timing of the A/V was thoroughly tested using previously developed techniques (Snider, Plank, Lee, & Poizner, 2013). The most important quantities are the A/V latency: the time from when an event (sound or video) is requested to when it is rendered. For the audio, we first measured the latency of the tablet's microphone by playing a beep at a precisely known time, and we found a latency of 10.01(+/- 0.1)ms. This was as expected since audio recordings require very precise temporal alignment. We then programmed the tablet to play a loud sound at regular intervals so we could record the time between when the sound was requested and when the speakers made noise: 69(+/- 0.7)ms. To characterize the video, we used a sub-millisecond latency wired Arduino that could flash an LED at specific times. We then programmed the tablet to simultaneously flash the LED and change the screen. To measure the interval between these two, we used a high speed camera (420Hz) and recorded 100 repetitions. The screen had about a 16ms on/off time, i.e. it completely turned from black to white within 1 frame, and a clearly visible double buffering with 1 frame of no response (the back buffer populating), 1 frame of updating the screen, and then the image is fully updated.

Language Timing Annotation

To move from data collection to analysis of stimulus-locked neural data we obtained the timing of relevant stimuli and their identity, such as words and phonemes during their media experience. We used a freely available tool to semi-automate temporal tagging, FAVE-align from the University of Pennsylvania Linguistics Lab (Rosenfelder, Fruehwald, Evanini, & Yuan, 2011). Figure 1B at the top demonstrates how words and phonemes are extracted from the movie audio file. FAVE-align takes both audio file and transcript as input to determine stimulus beginning and ending times. For movies, the transcript is available in the form of the closed captions, which can be extracted from the movie file and fed into FAVE-align. Transcripts are also available for audiobooks, which by nature have transcripts, and podcasts, some of which

have transcripts available, providing the necessary ground truth input. With FAVE-align, it was necessary to go through and confirm the onset times were accurate. As voice recognition and annotation is an area of great progress over the recent past and foreseeable future, more powerful tools will become available to ensure even less work is necessary for the temporal annotation of audio media.

Patients

Data from two patients who had both watched the same movie, *Zoolander,* were selected for analyses. These patients had long-standing pharmaco-resistant seizures and participated after fully informed consent monitored by institutional review boards. WADA tests determined that both patients were left-hemisphere language dominant and both patients were judged to have average general intellectual function. Patient A (seizure onset age: 2; surgery age: 31) had bilateral electrode strips and Patient B (seizure onset age: 21; surgery age: 33) had a right-sided electrode grid and strips.

Electrophysiological Data Pre-Processing

Data were preprocessed using MATLAB (MathWorks), the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011), and custom scripts. We used an average subtraction reference for each patient, followed by a bandstop around line-noise and its harmonics (60,120,180Hz) to reduce noise. Analysis focused on power in the high-gamma frequency range (here defined as 70-170Hz) as it correlates with local neuronal population activity (Ray, Crone, Niebur, Franaszczuk, & Hsiao, 2008) and performs well in mapping neuronal activity in sensory (Steinschneider et al., 2011), motor (Darvas et al., 2010), and language (Crone et al., 2001; Mesgarani et al., 2014; Thesen et al., 2012) cortex with excellent spatial and temporal precision. To obtain high-gamma power, epochs were transformed from the time domain to the time-frequency domain using the complex Morlet wavelet transform from 70-170Hz in 10Hz increments. Constant temporal and frequency resolution across target frequencies were obtained by adjusting the wavelet widths according to the target frequency. The wavelet widths

increase linearly from 14 to 38 as frequency increased from 70 to 170Hz, resulting in a constant temporal resolution with a standard deviation of 16ms and frequency resolution of 10 Hz. For each epoch, spectral power was calculated from the wavelet spectra, normalized by the inverse square frequency to adjust for the rapid drop-off in the EEG power spectrum with frequency, and averaged from 70 to 170 Hz, excluding line noise harmonics. This data was smoothed by a moving window exactly matching the temporal characteristics of the wavelet. Data were then high-passed at 0.5 Hz to remove any offset and long slow drifts over the course of the several hours a movie takes.

Data Analysis: Stimulus-Locked Averages

The first analysis approach uses standard event-related average analyses to validate our ability to time-lock to language events perceived while listening to language-rich media. First, we compared the auditory words with a separate class of visual events. The visual events, referred to as “shot-changes”, were defined as when the proportion of displayed pixels which changed between frames was above a threshold. To show electrode preference for either auditory or visual stimuli, epochs were created for all perceived auditory words and visual shot-change events. The epochs were taken from 500ms prior to stimulus onset to 1500ms post stimulus onset with the onset of the stimulus, either visual or auditory, as the 0ms point. Epochs judged to contain artifacts were identified by outlier values in amplitude and variance, visually inspected, and removed from further analysis. Significant differences between visual and auditory stimuli were identified using a two-stage statistical procedure.

The first stage was to determine if the auditory and/or visual stimuli evoked a significant increase in high-gamma power at each electrode. This was accomplished with a 1-sample t-test, separately comparing auditory trials and visual trials timepoint-by-timepoint to a population mean of 0, corrected for temporal false-discovery rate at $p<.05$ (Benjamini & Hochberg, 1995).

The second stage was a one-way ANOVA, corrected using a bootstrapped shuffling method (Maris & Oostenveld, 2007), run timepoint-by-timepoint for high-gamma power between

aligned language and shot-change epochs to characterize electrode preference. For Patient A, 5159 auditory word-epochs were created and 2115 visual shot-change epochs were created. For Patient B, 5163 auditory word-epochs were created and 2227 visual shot-change epochs were created. The slight differences in trial numbers were caused by watching slightly different amounts of the movie. An electrode was judged to be responding preferentially to shot-changes if that electrode both responded above baseline to visual stimuli and if shot-change epochs were significantly greater than language epochs during a period of 100-to-400ms post stimulus onset. An electrode was judged to be preferentially responding to language if that electrode both responded above baseline to language stimuli and if language epochs were significantly greater than shot-change epochs during the period of 100-to-400ms post-stimulus onset.

Next, we sought to show that the electrodes with an auditory preference were sensitive to linguistic phenomenon instead of just representing a non-specific auditory sensory response. For this we broke the auditory epochs into content and function words to see if we could replicate the well-known phenomenon of differential neural processing based on broad syntactic class (Halgren et al., 2002, 2015). For Patient A there were 2642 Content-word epochs and 2517 Function-word epochs. For Patient B there were 2637 Content-word epochs and 2526 Function-word epochs. An electrode was judged to be syntactically selective if it fulfilled the two criteria establishing a language-preferring electrode and if it additionally showed a significant difference between function and content words during the period of 100-to-400ms post-stimulus onset using the same 1-way ANOVA and shuffling method from above.

Data Analysis: Example of Post-Hoc Language Characterizations using Contextual Probability

There is a long history of electrophysiological work on predictability in language comprehension, most notably focusing on the ease or difficulty of integrating a word's lexico-semantic identity into the ongoing larger cognitive context (Marinković, 2004). This often takes the form of contrasting unexpected words such as "dog" in contexts such as "I take my coffee with cream and dog" which elicit an increased negativity relative to more expected words, such

as “sugar” in the same context (Kutas, Hillyard, & others, 1980). Here we use a measure of the word contextual probability (i.e. word predictability) in ongoing naturalistic language processing, instead of experimenter-created stimulus sets. We trained a trigram language model with modified Kneser-Ney smoothing (Chen & Goodman, 1999) on Cornell Movie-Dialogs Corpus (Danescu-Niculescu-Mizil & Lee, 2011) using the SRI Language Modeling toolkit (Stolcke, 2002). This type of language model has previously been shown to be a good model for predictability effects in behavioral and neural language comprehension data (Smith & Levy, 2013) and using language models has been used for N400-like responses in the past (Parviz, Johnson, Johnson, & Brock, 2011). Using this language model, we estimated contextual probability as the log probability of each word given the preceding two words.

To examine the sensitivity of our neural data to contextual probability, we regressed high-gamma power on the logs of these frequencies for the verbs. We included various covariates which were intended to disentangle various types of predictive processing. Some of these were manually coded, such as whether the word was a function word (e.g., auxiliaries such as “be”) or a content word (e.g., “copying”). For others, we used data from a norming task in which we asked 60 participants to write the first sentence that came to mind when presented with each unique verb in the movie. Based on these norming data, we calculated and included the following values as covariates in the model: the log of the number of distinct syntactic frames associated with each verb, the log of the frequency with which each verb appeared as intransitive; the log of the frequency with which each verb appeared as transitive; the log of the frequency with which each verb appeared with a direct object; and the log of the frequency with which each verb appeared with an indirect object. We also included interaction terms between these last three covariates. In total 10 covariates were used: Function/Content, Log number of syntactic frames, log frequency of intransitivity, log frequency of transitivity, log frequency of direct object, log frequency indirect object, along with the three-way interaction and all 2-way interactions of the latter three covariates.

We performed 4256 linear regressions (28 time windows by 99 electrodes for Patient A and 28 time windows by 53 electrodes for Patient B) of the average HGP data in 25ms windows starting 200ms before word onset and ending 500ms after word onset for all electrodes. Each regression included 893 verb "trials." We performed FDR-corrections on the p-values of the F-statistics (11 and 875 degrees of freedom) of each linear regression. Below, we only report findings when the corrected model p-value was less than .05 and when the uncorrected p-value associated with contextual probability in that model was less than .05. To even more stringently avoid the possibility of reporting false positives, we also exclude results from electrodes which do not show significant activity in more than two of the 25ms time windows.

Results

Stimulus-Locked Averages

Figure 2 displays locations and example average waveforms of electrodes displaying stimulus-locked effects between auditory words and visual shot-changes. We wanted to establish that electrodes displayed differentiable activity between two stimulus modalities, visual shot changes and auditory words, using traditional stimulus-locked averages. Of the 118 electrodes measured, 12 were found to preferentially respond to auditory stimuli (~10%) and 38 were found to preferentially respond to visual stimuli (~32%). The language responsive electrodes were all clustered in perisylvian areas (e.g. 7 in the STG, 3 in Rolandic cortex, 2 in supramarginal cortex). The shot-change responsive electrodes were mainly clustered in ventral visual areas (e.g. 12 in the fusiform, 6 in ventral prefrontal, 7 in lateral temporal areas, 6 in medial temporal areas). Of all these selective electrodes, only a single electrode in the caudal STG was found to be both visual and auditory preferential; it first preferred auditory words and later preferred visual shot changes.

Figure 3A displays locations and example average waveforms of electrodes displaying stimulus-locked effects between function and content words. We wanted to establish that the auditory responses found were sensitive to language characteristics and did not just reflect non-

specific auditory sensory responses. To do this we compared content and function words. We found 7 of the 12 language selective electrodes exhibited a distinction between content and function words in the first 400ms after word occurrence. Content-word-preferring electrodes were found in posterior perisylvian areas (caudal STG with 3 electrodes and supramarginal with 2 electrodes) and function-word-preferring electrodes were slightly more distributed (caudal STG, middle STG, and inferior postcentral all with 1 electrode). Each patient had at least one content-preferring electrode and one function-preferring electrode, representing a double-dissociation between broad syntactic class for both patients. Figure 3A in the top right displays one electrode from each subject from the same area in the right hemisphere (caudal STG) which both show a preference for content words over function words, demonstrating the reproducibility across patients of these findings.

Insert Figure 3 about here

Language Probabilities and Verb Transitivity

Figure 3B displays results from our regression model between high-gamma power and contextual probability. In Patient A, we found significant activity associated with contextual probability measures in left middle STG beginning at 125ms after word onset and persisting until 225ms. Later, between 325 and 350ms as well as between 475 and 500ms, we found significant activity associated with contextual probability in left caudal STG. In Patient B, for whom we had only right hemisphere coverage, we found no activity associated with contextual probability. Since both Patient A & Patient B had electrodes responding preferentially to language in the right perisylvian areas this suggests, very tenuously since the results come from only two patients and only one movie, contextual probability effects may be localized to the left (language-dominant) hemisphere.

Discussion

Here we demonstrated from beginning-to-end how a comprehensive effort using iEEG to study natural language heard during an audiovisual media experience would progress. First, we

described the physical and software platform that was created to allow for a patient to unobtrusively experience natural media which was precisely synchronized with recorded clinical neural data. Second, we identified exact timing of words in a movie, tagged broad syntactic properties, and found word contextual probabilities. Third, we reported event-related analyses which confirmed we were able to accurately time-lock to words heard during the continuous media experience. Finally, we showed with an analysis of contextual probability effects on the neural data how experimenters could in a post-hoc manner take the identified sequence of heard language and characterize this language based on questions of scientific interest. At a most basic level, perisylvian electrodes were responsive to spoken dialogue and ventral-temporal electrodes were responsive to shot changes, identifying separable networks during processing of a movie, with the language network revealing sensitivity to broad syntactic class and contextual probability.

The power of this method for post-hoc analysis is suggested by our example focused on linguistic prediction as indexed by contextual probability. The results from this analysis revealed significant activity associated with contextual probability in the left STG. Of interest were the differing time courses of this relationship in the middle-STG, which began at ~125ms, versus the caudal STG, which began at ~325ms. Understanding the relationship of language comprehension and word predictability, particularly how the bottom-up sensory encoding experience interacts with top-down predictions (Bastos et al., 2012), is important in understanding how we can understand speech heard at ~5 words a second. As an example, in visual language it has been demonstrated that whether a function word predicts the following word will be a function word or a content word causes differentiable activity as early as ~60ms (Halgren et al., 2015). A more thorough study of this question with more patients, more words, and with additional electrode coverage would allow for in depth characterization of the interaction of predictability and language comprehension across the cortex. Examining the time

course of early responses, which in this example were modulated by the contextual probability of an incoming word, are possible only with the temporal precision of electrophysiology.

The introductory analyses presented here are of course just one part of the larger context of natural language. The annotations were performed with a mix of freely available software and custom built modeling. As the data accumulated in the manner we describe grows, because of the tight time-locking between media and neural data it will remain viable to approach it with novel and more complex annotation approaches as advances in computing, modeling, and analysis improve. Indeed, as the data accumulates and analyses are performed, the opportunities presented to link neural data and linguistic phenomenon may drive analyses in media annotation to answer questions that are not possible with current approaches. As the media experienced by the patient remains available for additional annotation, characterizing that media in new ways and extracting new variables for analysis is limited only by hypothesis generation.

The number of possible linguistic questions regarding the auditory media patients will hear is vast. However, we envision a fully scaled-up effort not just as an increase in the number of patients enrolled and the amount of language heard, but also in a more thorough characterization of the visual context in which that language is experienced. For example, how does the visual context affect language processing, such as the presence or absence onscreen an object or person being discussed? How about an understanding of different word expectation probabilities based on the physical appearance of the speaker? Also of importance is the related questions regarding whether video cues for their speech, i.e. lip movements, are present or absent.

The study of in-depth hypotheses regarding the full visual semantic context will require automated tools for dense labeling of the semantic entities and contexts that appear in each movie frame. An image can be annotated with the names of objects that appear in it as well as with the category of the scene that it represents. For example, an image that could be labeled

as an “outdoors city” scene might also be correctly described as containing ‘pedestrians’, ‘buildings’ and ‘cars’. The ability of computer vision algorithms to identify objects in an image has improved substantially in the recent years. Deep learning methods using convolutional neural networks (Lecun, Bottou, Bengio, & Haffner, 1998) have shown a remarkable performance on classification (Krizhevsky, Sutskever, & Hinton, 2012) of a wide range of object classes. Today, it is possible to classify images with respect to vocabularies of thousands of objects using off the shelf software (Jia et al., 2014). There have been recent advances in the problems of object localization and scene classification (Girshick, Donahue, Darrell, & Malik, 2014; Sermanet et al., 2013; Zhou, Lapedriza, Xiao, Torralba, & Oliva, 2014). Combining this dense labeling with eye-tacking software would allow for examinations of the interrelation of visual attention and language processing.

The last potentiality we mention is to move beyond just labeling media experienced, which emulates natural language, into labeling the truly natural language of conversation which the patient engages in during their stay. This would be an increase in necessary coordination with the patient and the clinical team to ensure patient comfort and experimenter conformance with patient privacy guidelines. But even if only with a subset of patients, the possibility of obtaining conversational natural language with iEEG would be an important corroboration and extension of the natural language findings from the experienced media. This would be accomplished using the tablet’s recording capabilities to capture the conversations in which the patient takes part. Another possibility is to use clinically recorded audio and video, which is increasingly high-definition, in patient rooms. Algorithms exist which automatically de-identify video by detecting and obscuring faces and auditory recordings of conversations can have word times tagged and identifying nouns removed so that only the time-stamp and word and phoneme identity are uploaded to a publicly-available database. Likely, there will be patients who are fine with engaging with the tablet but not with being recorded, which is why having an

adaptable and modular set-up will ensure that the most data is collected while keeping the experience for the patient and clinical team seamless and safe.

With ‘Big Data’ becoming less of a buzzword and more of a reality, we imagine multi-site collaborations together building an ‘intracranial neurolinguistic corpus’ of millions of words presented in a natural context and aligned to hundreds of channels of iEEG for each patient. Big databases of sub-lexical and lexical stimulus characteristics (Medler & Binder, 2005; Vaden, Halpin, & Hickok, 2009), stimulus meaning (Fellbaum, 2005), behavior (Balota et al., 2007; Pexman, Heard, Lloyd, & Yap, 2017), and extracranial electrophysiological data (Dufau, Grainger, Midgley, & Holcomb, 2015) have revealed nuances and precision in findings which would be impossible for traditional smaller studies. Indeed, large-scale multi-site studies using iEEG and cognitive tasks are already underway for neural prosthetics (Jacobs et al., 2016; Solis, 2017). By bringing a natural language approach in an unobtrusive, patient- and clinical-friendly way, researchers can collaboratively bring the same scalable benefits to iEEG language research.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Chan, A. M., Baker, J. M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., … Halgren, E. (2011). First-Pass Selectivity for Semantic Categories in Human Anteroventral Temporal Lobe. *Journal of Neuroscience*, *31*(49), 18119–18129. https://doi.org/10.1523/JNEUROSCI.3122-11.2011

Chen, S. F., & Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, *13*(4), 359–394.

Crone, Ne., Hao, L., Hart, J., Boatman, D., Lesser, R. P., Irizarry, R., & Gordon, B. (2001). Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology*, *57*(11), 2045–2053.

Danescu-Niculescu-Mizil, C., & Lee, L. (2011). Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 76–87). Association for Computational Linguistics.

Darvas, F., Scherer, R., Ojemann, J. G., Rao, R. P., Miller, K. J., & Sorensen, L. B. (2010). High gamma mapping using EEG. *Neuroimage*, *49*(1), 930–938.

Dufau, S., Grainger, J., Midgley, K. J., & Holcomb, P. J. (2015). A thousand words are worth a picture: Snapshots of printed-word processing in an event-related potential megastudy. *Psychological Science*, *26*(12), 1887–1897.

Fellbaum, C. (2005). WordNet and wordnets.

Gaillard, R., Naccache, L., Pinel, P., Cl?menceau, S., Volle, E., Hasboun, D., … Cohen, L. (2006). Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading. *Neuron*, *50*(2), 191–204. https://doi.org/10.1016/j.neuron.2006.03.031

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587).

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like Magnetoencephalography Responses Modulated by

Semantic Context, Word Frequency, and Lexical Class in Sentences. *NeuroImage*, *17*(3), 1101–1116. https://doi.org/10.1006/nimg.2002.1268

Halgren, E., Kaestner, E., Marinkovic, K., Cash, S. S., Wang, C., Schomer, D. L., … Ulbert, I. (2015). Laminar profile of spontaneous and evoked theta: rhythmic modulation of cortical processing during word integration. *Neuropsychologia*, *76*, 108–124.

Jacobs, J., Miller, J., Lee, S. A., Coffey, T., Watrous, A. J., Sperling, M. R., … Lega, B. (2016). Direct electrical stimulation of the human entorhinal region and hippocampus impairs memory. *Neuron*, *92*(5), 983–990.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., … Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675–678). ACM.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Kutas, M., Hillyard, S. A., & others. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Leonard, M. K., Bouchard, K. E., Tang, C., & Chang, E. F. (2015). Dynamic Encoding of Speech Sequence Probability in Human Temporal Cortex. *The Journal of Neuroscience*, *35*(18), 7203–7214. https://doi.org/10.1523/JNEUROSCI.4100-14.2015

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Marinković, K. (2004). Spatiotemporal Dynamics of Word Processing in the Human Cortex. *The Neuroscientist*, *10*(2), 142–152. https://doi.org/10.1177/1073858403261018

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Medler, D. A., & Binder, J. R. (2005). *MCWord: An on-line orthographic database of the English language*.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., … Pallier, C. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 201701590.

Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature*, *372*(6503), 260–263. https://doi.org/10.1038/372260a0

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1.

Parviz, M., Johnson, M., Johnson, B., & Brock, J. (2011). Using language models and Latent Semantic Analysis to characterise the N400m neural response. In *Proceedings of the Australasian Language Technology Association Workshop 2011* (pp. 38–46).

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: concrete/abstract decision data for 10,000 English words. *Behavior Research Methods*, *49*(2), 407–417. https://doi.org/10.3758/s13428-016-0720-6

Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J., & Hsiao, S. S. (2008). Neural Correlates of High-Gamma Oscillations (60-200 Hz) in Macaque Local Field Potentials and Their Potential Implications in Electrocorticography. *Journal of Neuroscience*, *28*(45), 11526–11536. https://doi.org/10.1523/JNEUROSCI.2848-08.2008

Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). FAVE (forced alignment and vowel extraction) program suite. *URL Http://Fave. Ling. Upenn. Edu*.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *ArXiv Preprint ArXiv:1312.6229*.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Snider, J., Plank, M., Lee, D., & Poizner, H. (2013). Simultaneous neural and movement recording in large-scale immersive virtual environments. *IEEE Transactions on Biomedical Circuits and Systems*, *7*(5), 713–721.

Solis, M. (2017). Committing to Memory: Memory Prosthetics Show Promise in Helping Those with Neurodegenerative Disorders. *IEEE Pulse*, *8*(1), 33–37.

Steinschneider, M., Nourski, K. V., Kawasaki, H., Oya, H., Brugge, J. F., & Howard, M. A. (2011). Intracranial Study of Speech-Elicited Activity on the Human Posterolateral Superior Temporal Gyrus. *Cerebral Cortex*, *21*(10), 2332–2347. https://doi.org/10.1093/cercor/bhr014

Stolcke, A. (2002). SRILM-an extensible language modeling toolkit. In *Seventh international conference on spoken language processing*.

Thesen, T., McDonald, C. R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., … Halgren, E. (2012). Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nature Communications*, *3*, 1284. https://doi.org/10.1038/ncomms2220

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine phonotactic online dictionary, Version 2.0.[Data file]. *Available from Www. Iphod. Com*.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. In *Advances in neural information processing systems* (pp. 487–495).

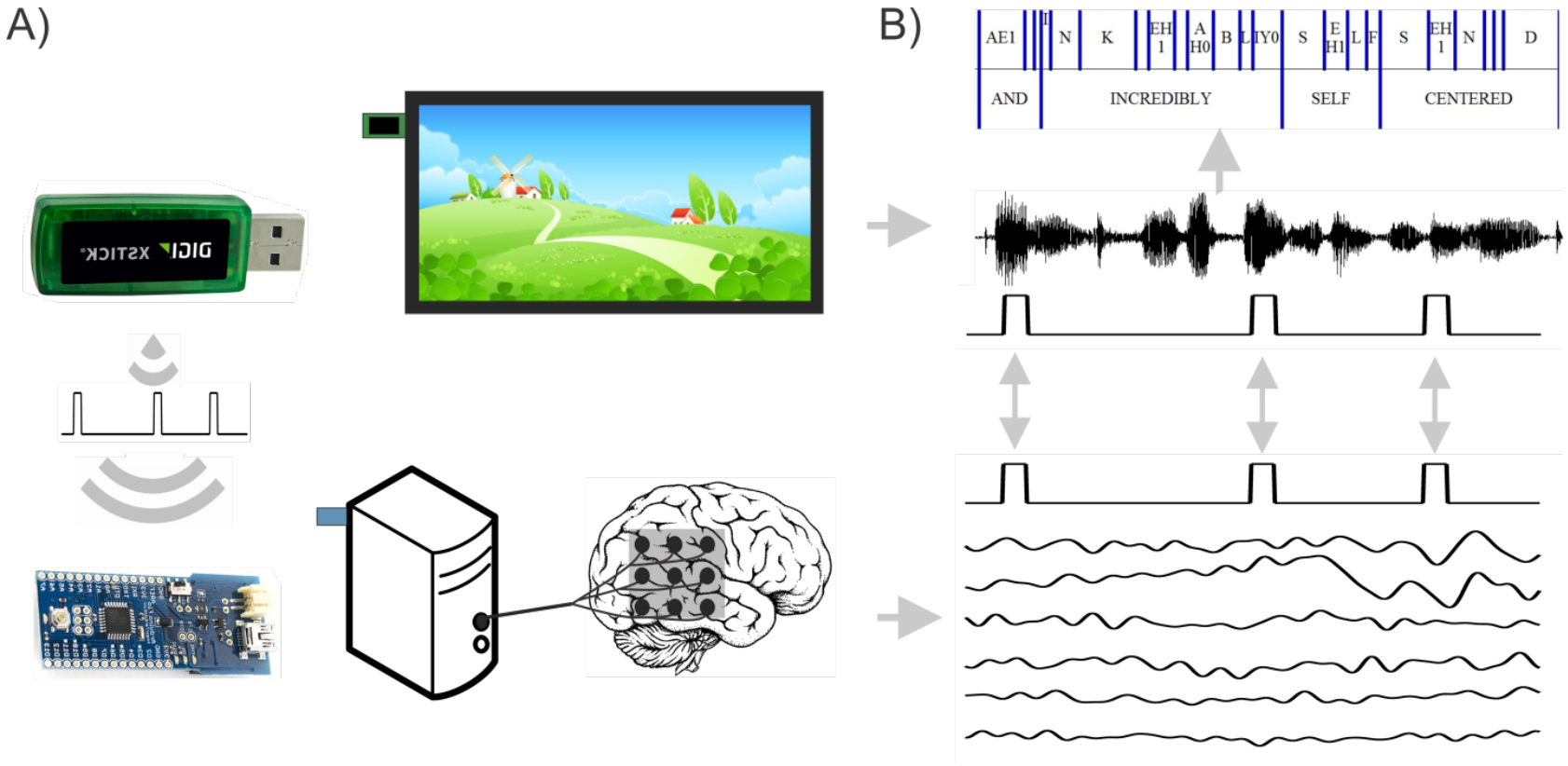
Figure Captions

Figure 1. Unobtrusive Media Presentation, Synchronization with Intrancranial Electrophysiology, and Language Annotation. A) The tablet (top) is pre-loaded with natural media language (movies, TV shows, audiobooks, and podcasts) and given to the patient for use throughout their stay. Media presented on the tablet is synchronized with wirelessly-sent synchronizing pulses to the recording computer (bottom) which is recording the clinical electrophysiology. The pulses are sent with an XStick (green USB) to the receiving Arduino Fio V3 (blue Arduino) which is plugged into a spare channel in the recording system. B) The audio file of the movie is extracted (middle) and aligned precisely with the electrophysiological data (bottom) by the shared synchronzation pulses in both data. The audio file is then annotated for language, here done by FAVE-align (top), so that the annotated words and phonemes are then time-locked with respect to the electrophysiological data.

Figure 2. Stimulus-Locked Averages from 2 Patients Distinguish Auditory Language from Visual Shot-Change Responses. Displayed on the average brain are electrodes from two patients which demonstrated a stimulus-specific response to either words (light blue), shot-changes (orange), or were not-responsive to either stimulus type (white). The shot-change specific electrodes are found in typically visual areas such as the ventral and inferior occipito-temporal areas. The language-specific electrodes are found in expected perisylvian areas. The four example wave-forms show the differences between language and shot-change high-gamma power averages, with two electrodes from each patient displayed. The three vertical black bars on each plot signify 0, 200, and 400ms and the horizontal black bar signifies a high-gamma power of 0. The blue bar at the bottom of each plot displays periods of significant difference between language and shot-change stimuli, corrected for false-discovery rate.

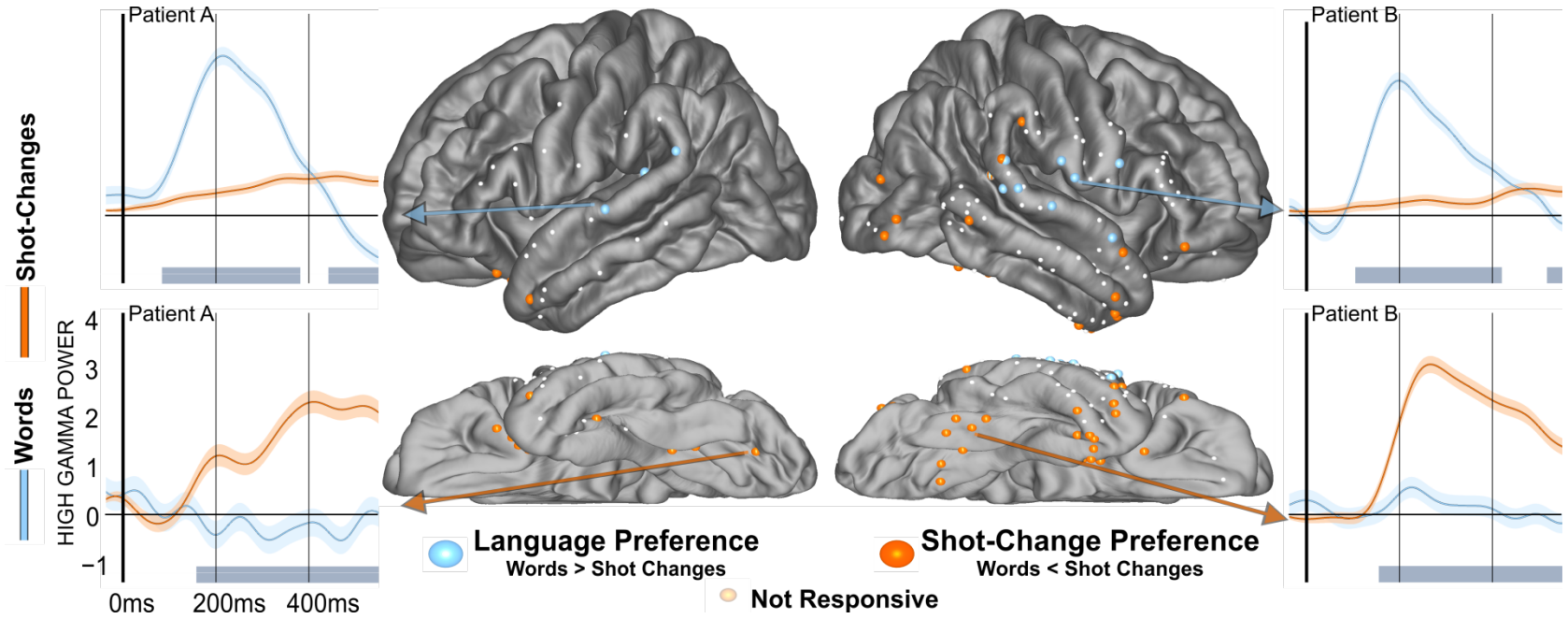
Figure 3. Language-Responsive Electrodes Display Sensitivity to Broad Syntactic Class and Contextual Probability. A) Displayed on the average brain are electrodes from two patients which displayed a preference for either function words (red violet) or content words (purple). The six example waveforms show that within the language-specific electrodes from Figure 2, many (7 out of 12) demonstrated significant differences based on broad syntactic class. For each plot the three vertical black bars signify 0, 200, and 400ms from word onset and the horizontal black bar signifies a high-gamma power of 0. The purple bar at the bottom of each plot displays periods of significant difference between function and content words, corrected for false-discovery rate. B) Two electrodes, circled in the top plot, showed significant activity associated with contextual probability. The electrode in the middle STG (blue-green circle) was significantly associated starting at ~125ms while the electrode in the caudal STG (blue circle) was significantly associated starting at ~325ms.

Figure 1. Unobtrusive Media Presentation, Synchronization with Intrancranial Electrophysiology, and Language Annotation



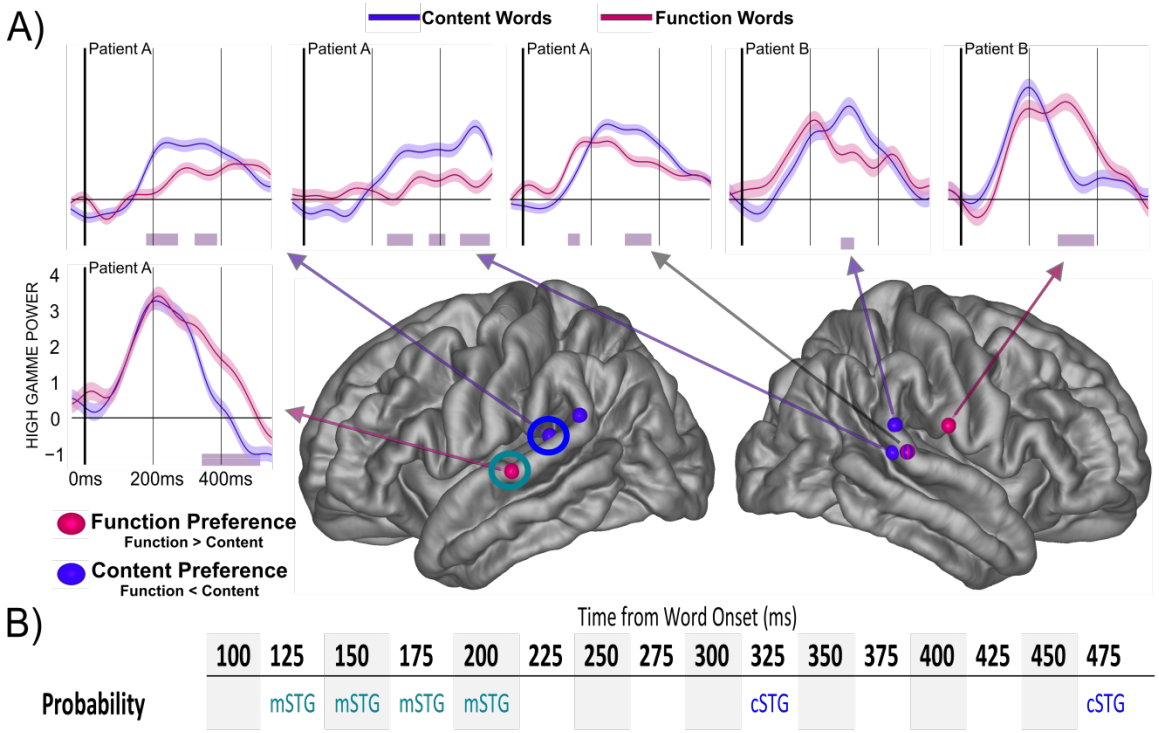
A) The tablet (top) is pre-loaded with natural media language (movies, TV shows, audiobooks, and podcasts) and given to the patient for use throughout their stay. Media presented on the tablet is synchronized with wirelessly-sent synchronizing pulses to the recording computer (bottom) which is recording the clinical electrophysiology. The pulses are sent with an XStick (green USB) to the receiving Arduino Fio V3 (blue Arduino) which is plugged into a spare channel in the recording system. B) The audio file of the movie is extracted (middle) and aligned precisely with the electrophysiological data (bottom) by the shared synchronzation pulses in both data. The audio file is then annotated for language, here done by FAVE-align (top), so that the annotated words and phonemes are then time-locked with respect to the electrophysiological data.

Figure 2. Stimulus-Locked Averages from 2 Patients Distinguish Auditory Language from Visual Shot-Change Responses.



Displayed on the average brain are electrodes from two patients which demonstrated a stimulus-specific response to either words (light blue), shot-changes (orange), or were not-responsive to either stimulus type (white). The shot-change specific electrodes are found in typically visual areas such as the ventral and inferior occipito-temporal areas. The language-specific electrodes are found in expected perisylvian areas. The four example wave-forms show the differences between language and shot-change high-gamma power averages, with two electrodes from each patient displayed. The three vertical black bars on each plot signify 0, 200, and 400ms and the horizontal black bar signifies a high-gamma power of 0. The blue bar at the bottom of each plot displays periods of significant difference between language and shot-change stimuli, corrected for false-discovery rate.

Figure 3. Language-Responsive Electrodes Display Sensitivity to Broad Syntactic Class and Contextual Probability



A) Displayed on the average brain are electrodes from two patients which displayed a preference for either function words (red violet) or content words (purple). The six example waveforms show that within the language-specific electrodes from Figure 2, many (7 out of 12) demonstrated significant differences based on broad syntactic class. For each plot the three vertical black bars signify 0, 200, and 400ms from word onset and the horizontal black bar signifies a high-gamma power of 0. The purple bar at the bottom of each plot displays periods of significant difference between function and content words, corrected for false-discovery rate. B) Two electrodes, circled in the top plot, showed significant activity associated with contextual probability. The electrode in the middle STG (blue-green circle) was significantly associated starting at ~125ms while the electrode in the caudal STG (blue circle) was significantly associated starting at ~325ms.