



Published in final edited form as:

J Neural Eng. 2013 February ; 10(1): 016010. doi:10.1088/1741-2560/10/1/016010.

Decoding stimulus identity from multi-unit activity and local field potentials along the ventral auditory stream in the awake primate: implications for cortical neural prostheses

Elliot Smith¹, Spencer Kellis², Paul House³, and Bradley Greger^{1,*}

¹Department of Bioengineering University of Utah, Salt Lake City, UT

²Department of Electrical and Computer Engineering University of Utah, Salt Lake City, UT

³Department of Neurosurgery, University of Utah, Salt Lake City, UT

Abstract

Objective—Hierarchical processing of auditory sensory information is believed to occur in two streams: a ventral stream responsible for stimulus identity and a dorsal stream responsible for processing spatial elements of a stimulus. The objective of the current study is to examine neural coding in this processing stream in the context of understanding the possibility for an auditory cortical neural prosthesis.

Approach—We examined the selectivity for species-specific primate vocalizations in the ventral auditory processing stream by applying a statistical classifier to neural data recorded from microelectrode arrays. Multi-unit activity (MUA) and local field potential (LFP) data recorded simultaneously from AI and rostral PBr were decoded on a trial-by-trial basis.

Main Results—While decode performance in AI was well above chance, mean performance in PBr did not deviate >15% from chance level. Mean performance levels were similar for MUA and LFP decodes. Increasing the spectral and temporal resolution improved decode performance; while inter-electrode spacing could be as large as 1.14 mm without degrading decode performance. Significance: These results serve as preliminary guidance for a human auditory cortical neural prosthesis; instructing interface implementation, microstimulation patterns, and anatomical placement.

1. Introduction

The cochlear implant is the most widely used neural prosthesis. This device artificially reproduces hearing in a deaf person by stimulating the cochlea with pulses of electric current. Some deaf patients, however, don't have an auditory nerve that extends to the cochlea because of recession of the auditory nerve over time, cochlear ossification, or cranial nerve tumors. These patients require an alternative to the cochlear implant. Two such devices, which interface with subcortical nuclei, are currently being tested [1, 2]. Another possible avenue for eliciting auditory perception from stimulation of the nervous system could be through the auditory cortex.

Microstimulation of primary auditory cortex (AI) has been shown to elicit the correct behavioral responses in rats and cats trained to detect and discriminate pure tones [3, 4]. Stimulation of human auditory cortex has most commonly suppressed sound perception[5].

*bradley.greger@utah.edu.

Current address for Spencer Kellis: Division of Biology, Caltech, Pasadena, CA 91125

However, in the absence of sound stimuli, stimulation of Heschl's gyrus has elicited sound percepts [6]. Similar stimulation of human superior temporal gyrus (STG) with large, low-impedance electrodes has produced a variety of sensory percepts, most of them complex and holistic [7-9]. For example, one patient, when stimulated on the medial temporal lobe, reported hearing music; another, when stimulated on the STG, heard "a mother calling her little boy" [8]. Other patients have reported hearing "buzz," "hum," "knocking," "crickets," and "wavering" when stimulated on the STG[7]. Whether stimulation on smaller, penetrating electrodes could elicit more consistent perceptions remains to be tested. Yet, understanding how auditory information is encoded in these early auditory cortical areas will provide guidance on neural prosthetic design and use.

Visual cortex is organized into hierarchical processing streams. Spatial elements of visual stimuli are processed in a dorsal stream, while the identity of visual stimuli is processed in a ventral stream [10, 11]. Neuroanatomical and lesion studies suggest a similar separation of spatial location and stimulus identity pathways in the auditory system [12-17]. Neuroanatomical studies of the rostral belt and parabelt (PBr) support hierarchical organization in a ventral stream that extends along the STG, as these regions share reciprocal connections with ventrolateral prefrontal cortical areas (vIPFC) that are highly responsive to species-specific vocalizations [18]. PBr, in particular, has reciprocal connections with vIPFC as well as adjacent belt areas, superior temporal sulcus, and areas further rostral on the superior temporal plane (STP) [19, 20].

Whereas there is compelling neurophysiological evidence for a spatial processing stream in audition [21], physiological studies of a stimulus identity auditory stream have failed to develop a clear picture of how acoustic information is transformed through the cortical circuit. Much like inferotemporal cortex for vision, increasingly rostral areas on the STP showed increasingly sparse representations of natural stimuli and monkey vocalizations [22]. Much of the evidence for a ventral stream stems from a study showing increased selectivity for monkey calls at more rostral locations in belt auditory cortex and increased selectivity for the location of a sound at more caudal locations [23]. Other studies have shown less selectivity via robust responses to vocalizations across auditory cortical areas, and imaging has shown areas further anterior on the STP to be selective for vocalizations [24]. Most auditory cortical data has been collected from core and belt regions, with sparse sampling along the STP and no recordings yet in PBr. There is need for further study of auditory cortex on the STG, especially in the context of a neural prosthesis, as the ventral auditory stream is likely important for constructing auditory percepts.

Decoding stimulus identity from neural responses is one way to examine stimulus selectivity along the processing stream. This methodology, however, has yielded conflicting results. Auditory research using linear classifiers on action potential (AP) data has shown less selectivity in successive areas along the processing stream. The linear pattern discriminator (LPD) is a commonly used decoding technique in auditory research [25]. Using LPD on MUA data in awake macaques, a set of four vocalizations was decoded with approximately 90% accuracy in each of two core areas and three belt areas [26]; however, further along the ventral stream, neurons in vIPFC had lower decode performance than Belt neurons [27]. Decode performance of vocalizations in vIPFC was higher when using linear functions of the probabilistic output of a hidden Markov model, as opposed to linear functions of the spectrotemporal elements of the monkey vocalizations, as in LPD [28]. Why decode performance decreases along the ventral auditory processing stream remains an open question that has implications for neural prosthesis development. Greater understanding of how neural coding of stimulus identity at successive stages in an auditory identity processing stream will provide guidance on how to implement a cortically based auditory prosthesis.

Incorporating knowledge of neural coding into stimulation can improve prosthetic performance [29]. Therefore, as a first step towards designing a neural prosthesis based upon cortical microstimulation, we decoded neural signals recorded on chronically implanted microelectrode arrays during species-specific vocalizations. The impact of spatial, temporal, and frequency parameters on decoding performance were examined. Here we examine selectivity of MUA and local field potentials (LFP) for stimulus identity in an awake, behaving primate to better understand the possibility for an auditory cortical neural prosthesis. This study used macaque vocalization stimuli to examine discriminability of auditory stimulus identity in two cortical areas in the ventral processing stream. While there was no direct auditory cortex stimulation in this study, we show that decoding stimulus identity from auditory cortex has implications for basic auditory cortical physiology and provides guidance on the implementation of a cortically based auditory neural prosthesis.

2. Research methods

2.1 Research subject

The neural data examined in this study were recorded from 192 electrodes (96 per electrode array) in two cortical areas in one male rhesus macaque over a period of five months. All experiments were performed according to NIH guidelines for animal care and use and with approval from the University of Utah Institutional Animal Care and Use Committee.

2.2 Micro-electrode array implantation

Penetrating microelectrode arrays (Blackrock Microsystems, LLC, Salt Lake City, UT) were implanted in AI and rostral PBr (figure 1). The monkey's temporalis muscle was detached at the top of the skull and retracted, and a craniotomy was made to expose the superior temporal and inferior frontal lobes. The parietal lobe was carefully dissected away from the temporal lobe, exposing a small area on the STP. The parietal lobe was retracted from the surface of the STP, and the AI microelectrode array was inserted by hand at about 5 mm rostral to the inter-aural axis in stereotaxic coordinates, and from 3 to 5 mm within the lateral fissure. The PBr array was then implanted with a pneumatic insertion device on the surface of the STG [30]. We were unable to use the pneumatic insertion device to insert the lateral fissure array, as the parietal lobe prevented perpendicular access to the medial superior temporal plane. The craniotomy was sealed and temporalis muscle reattached.

2.3 Experimental paradigm

Seven macaque vocalization exemplars [31] were delivered free field through a piezoelectric speaker (ES1, Tucker Davis Technologies, Alachua, FL) randomly. There was at least 1 second between each stimulus presentation. Stimuli were presented while the monkey sat in an acoustically, optically, and electromagnetically shielded chamber (Acoustic Systems, Austin, TX).

Task control software was custom-built using LabVIEW (National Instruments, Austin, TX) and run in real time on a National Instruments PXI-embedded computer system. Digital markers and analog waveforms of auditory stimuli were recorded synchronously with the neural data.

2.4 Data collection and preprocessing

Neural data were recorded with a Cerebus System (Blackrock Microsystems). Electrophysiological signals were high-pass (0.3 Hz, one pole) and low-pass (7500 Hz, three pole) filtered and pseudo-differentially amplified with a gain of 5000 \times . High-frequency MUA data were digitally filtered (eight pole, high-pass 250 Hz, low-pass 7.5 kHz) and sampled at 30 kHz. Individual spikes were detected offline using t-distribution E-M sorting

[32]. Any MUA waveform that exceeded -3.5 RMS times the high pass filtered voltage signal was included as part of the MUA signal for the electrode on which the waveform was recorded. Large motion artifact waveforms, and waveforms with a shorter inter-spike interval than 1 millisecond were excluded from any analyses. The broadband data were recorded at 30 kHz and later filtered and down-sampled to 2 kHz for analysis of the LFP.

LFPs were recorded on 96 electrodes for 144 trials per class in AI (1008 total trials) and on 95 electrodes for 112 trials per class in PBr (784 total trials) over multiple days. To common average re-reference our LFP data, the average voltage from all electrodes for each trial was subtracted from the trial-by-trial response of each individual electrode. Spectrograms were calculated using multitapered analysis with a time-bandwidth product of 5, 9 leading tapers, 200-msec windows, and 10-msec step sizes. Spectrograms with either 77 or 304 frequency bins ranging from 0 to 300 Hz were calculated, to examine the effect of frequency resolution on decode performance.

MUA analysis utilized the same trials used for LFP data (1008 in AI and 784 in PBr). For MUA data, only the electrodes with a significant difference in change in the average firing rate (Wilcoxon rank-sum test, $P < 0.05$) between 500 milliseconds before the beginning of the vocalization and 500 milliseconds after the beginning of the vocalization were included in the analysis (65 electrodes in PBr and 62 electrodes in AI). Multitapered spectrograms and peri-stimulus time histograms (PSTHs) were generated using an open source neural analysis package [33]. Single-trial PSTHs and spectrograms were generated separately for four durations after the beginning of each vocalization (200, 400, 800, and 1600 msec). PSTHs were calculated using Gaussian kernels of five widths (5, 10, 25, 50, and 100 msec). These kernel widths constitute the five temporal resolutions we used in examining the effect of temporal resolution on decode performance.

2.5 Feature selection and processing

We extended the method of Kellis et al. [34] that simultaneously incorporates features representing dynamics in time, space, and frequency to apply to both MUA data and LFP data. We use the term *class* to refer to one type of conspecific vocalization (e.g., “grunt”) and the term *trial* to refer to one instance of a vocalization being played for the monkey.

To select training features for LFP data, two-dimensional spectrograms were calculated for a subset of seven trials from each class and each recording channel (figure 2). These multidimensional data were unwrapped to produce a two-dimensional matrix in which each row contained all the time and frequency features from all channels for a single trial. The feature matrix was z-scored, orthogonalized using PCA, and projected into the principal component space using a sufficient number of leading principal components to retain 95% of the variance in the data. A subset of seven trials from each class, directly following the training trials in time, was selected for testing the classifier, and spectrograms were calculated for each class and each recording channel (142 training and testing sets in AI and 110 training and testing sets in PBr). These data were unwrapped into a two-dimensional matrix, z-scored, and projected into the principal component space calculated during training.

As with the LFP-based decode, seven trials from each class were selected for training and seven trials following the training trials in time for each class were used as testing data for MUA data (142 training and testing sets in AI and 110 training and testing sets in PBr). For both training and testing, MUA data were collected into a large two-dimensional matrix where each row represented all firing rate data from all channels for a given trial. The data were then z-scored, orthogonalized, and projected into the principal component space using the aforementioned process for LFP data.

2.6 Evaluation

Using features derived from either MUA or LFP, data were classified on a trial-by-trial basis using linear discriminant analysis (LDA) [35] (supplementary figure 1). All possible combinations of two through seven classes were evaluated. Only one outcome exists for a combination of seven classes; otherwise, mean, median, standard deviation, interquartile range (IQR), and the 95% confidence interval were computed. We also evaluated performance for each single electrode on the AI microelectrode array and for varying numbers of electrodes (4, 7, 14, 24, 48, and 95 electrodes for LFP; and 4, 8, 16, 31, and 62 electrodes for MUA). These electrodes were chosen by successively removing twice as many electrodes between those used for the decode, while maintaining an even spatial sampling. Maps of electrodes we used for decoding in these conditions are shown in figure 4(d) and 4(f). These combinations, as well as class-by-class comparisons, were evaluated for the best performing durations after data onset (for MUA and LFP data) and best temporal resolutions (for MUA data).

Classification accuracy was measured against the level of chance, which was defined as equal likelihood for all classes. Chance performance therefore ranged from 50.0% for two classes to 14.3% for seven classes. Consistent classification above the level of chance indicated that the decode was finding and operating on relevant features from what could otherwise be stochastic physiological data.

2.7 Information theoretic analysis

We examined the mutual information between the vocalizations and the responses in AI and PBr. Probability distributions $p(s)$, $p(r)$, and $p(s,r)$ were taken from confusion matrices of pair-wise classification frequencies for all MUA- and LFP-based decodes, where $p(s)$ is the sum across one dimension divided by the total number of trials, $p(r)$ is the sum across the other dimension divided by the total number of trials, and $p(s,r)$ is the diagonal divided by the total number of trials. The equation

$$I(S;R) = \sum_{s \in S} \sum_{r \in R} p(s,r) \log_2 \left(\frac{p(s,r)}{p(s)p(r)} \right)$$

was then evaluated for these probability distributions. This analysis was performed for every pair-wise classification for all neurons for MUA-based decodes and all electrodes for LFP-based decodes. Mutual information was also calculated between stimulus and response for the five different kernel widths used for MUA-based decodes.

3. Results

3.1 Decode performance in AI relative to PBr for both MUAs and LFPs

Decode performance in AI was above chance, whereas decode performance in PBr for both MUAs and LFPs was near chance. For MUA-based decodes, the best decode performance in AI was $93.7 \pm 10.5\%$ (mean \pm std) (IQR: 7.9%; 95% confidence interval: 65.9% μ 100%) for pair-wise comparisons (all pair-wise comparisons are shown in Supplemental figure 2) and 73.2% for seven classes. The best MUA-based decode performance in PBr was $51.6 \pm 16.6\%$ (IQR: 17.7%; 95% confidence interval: 25.8% μ 90.8%) for pair-wise comparisons and 28.6% for seven classes. For LFP-based decodes, best performance in AI ranged from $83.5 \pm 14.1\%$ (IQR: 19.9%; 95% confidence interval 41.3% μ 97.8%) for two classes to 57.3% for seven classes. The best LFP-based decode performance in PBr was $53.8 \pm 09.3\%$ (IQR: 12.1%; 95% confidence interval: 36.8% μ 65.2%) for two classes to 21.2% for seven classes (figure 3).

3.2 Temporal and spectral classification dynamics

Since vocalizations were different lengths (mean length = 0.53 ± 0.37 seconds), we felt it necessary to examine the spectral and temporal parameters that may affect decoding vocalizations from AI and PBr. To examine temporal aspects of stimulus decoding, we applied the decode to data of different durations and temporal resolutions. Each temporal resolution for single-trial PSTHs is represented by a different hue line in figure 4(a). In AI, MUA-based decode performance improved with increased temporal resolution (figure 4(a)). The 5-msec temporal resolution outperformed other temporal resolutions (mean performance over data durations = $59.8 \pm 6.6\%$). MUA-based decode results were consistently near chance in PBr, regardless of data duration or kernel width (figure 4(a)). The best performing MUA-based decode in PBr was for the 1600-millisecond data duration and both the 50 and 25 -millisecond temporal resolutions (28.5 ± 0.1 for 50-millisecond temporal resolution and 28.5 ± 0.04 for 25-millisecond temporal resolution).

Temporal dynamics of LFP-based decodes were similar to temporal dynamics of MUA-based decodes. In AI, LFP-based decodes performed best with the 800-millisecond data duration and in PBr decodes performed best with the 400-millisecond data duration ($47.3 \pm 0.02\%$) and $22.4\% \pm 0.1\%$, respectively). To examine influence of spectral resolution on classification results, we decoded vocalizations using spectrograms with two different spectral resolutions. Increasing frequency resolution in the spectrograms improved LFP-based decode performance in both AI and PBr (figure 4(b)). While PBr results remained near chance overall, the 400-msec data duration for increased frequency resolution reached $22.6 \pm 0.01\%$ performance. In AI, higher frequency resolution also improved performance.

3.3 Spatial classification dynamics

Since decode performance was above chance in AI, we further explored decodability on the AI microelectrode array to better understand the topography for interfacing with AI. The decodability of each electrode on the array was examined by executing the LFP and MUA - based decodes for each individual electrode. Best electrode performance for seven classes ranged from 47.8% (one electrode) to below 15% (each of 46 electrodes) for LFP-based decodes and from 48.6% (one electrode) to below 15% (each of 31 electrodes). The majority of electrodes performing better than chance were on the medial half of the electrode array for LFP-based decodes (figure 4(c)). MUA performance was more variable across the array (figure 4(e)).

To examine the spatial scale of information processing in AI we explored LFP and MUA - based decode performance for fewer, and more sparsely spaced, electrodes on the electrode array. For LFP-based decodes, classifier performance was similar to performance using all electrodes for all electrode densities of 0.88 electrodes/mm² and above (14 electrodes). Performance dropped for 7 electrodes (0.44 electrodes/mm²) to 34% and again for 4 electrodes to 23% (0.25 electrodes/mm²) (figure 4(d)). Therefore, the minimum spacing required to maintain decode performance in AI is the inverse of the electrode density, (1/0.88 electrodes/mm²) or 1.14 mm²/electrode.

3.4 Mutual information

To examine information content over all neural responses recorded in AI and PBr, we calculated the mutual information between stimulus and response in the two cortical areas. Mutual information between stimulus and response was calculated for all MUA-based decodes for each temporal resolution. Decode performance in PBr was consistent among temporal resolutions and data durations for MUA-based decodes (figure 5(a)). Mutual information increased in AI for longer data durations and higher temporal resolutions (figure 5(a)).

To compare information content in MUA- and LFP-based decodes, mutual information between stimulus and response was then calculated for the best-performing temporal resolution for MUA-based decodes (5-msec temporal resolution) and best frequency resolution LFP-based decodes (1024-point) in both AI and PBr. We found that responses in PBr contained little information about the stimulus for all data durations (0.65 ± 0.11 bits for LFP-based decodes and 0.76 ± 0.02 bits for MUA-based decodes). AI decodes contained more information for longer data durations and contained more information than PBr decodes overall (figure 5(a)). MUA responses in AI contained more information than LFP responses in AI (0.56 ± 0.13 difference in bits across data durations).

4. Discussion

We have investigated the nature of LFPs and MUAs from AI and PBr in the context of conspecific vocalization stimulus to explore the potential for interfacing an auditory neural prosthesis directly with the neocortex. We observed that AI provided greater discriminability between vocalizations than PBr using linear statistical classification. We found that the decode performance was dynamic in space, time, and frequency. Increased temporal resolution improved MUA-based decode performance, and greater spectral resolution improved LFP-based decode performance. From this primary result, we suggest that information coding in AI relies on precise dynamics in both time and frequency domains. Spatial analyses estimated a lower limit of electrode spacing at 1.14 mm^2 for electrophysiological interface with AI. While increased spectral resolution improved LFP-based decode performance in PBr, most of the PBr decodes remained near chance. The process of implantation and data acquisition undertaken for this work illustrates the potential for chronic electrophysiological interface with awake, behaving primate AI over a period of months, similar to the type of interface that would be used for a cortical neural prosthesis.

4.1 Temporal precision of MUA responses

We found that increased temporal precision yields superior decode results in AI. We also found the highest information content for temporal resolutions of less than 10 milliseconds. Two of the three auditory linear pattern discriminator AP decode studies in macaques have shown increased performance for higher temporal precision (using binned spikes rather than Gaussian kernels) [26, 36]. These results also are in accord with decodes of marmoset vocalizations, where the greatest mutual information from primary auditory cortical neurons utilized bins smaller than 10 milliseconds [25], and with previous studies of temporal integration in AI in the marmoset and macaque showing high spike timing reliability and short latencies in AI [37, 38]. We can therefore conclude that high temporal precision is an important feature of stimulus coding in AI.

4.2 Effect of Duration on decode performance

We have shown that increased data durations in AI improved decode results and provided increased information content up to the 800-millisecond data duration, however information about the stimulus increases in AI up to 1600 milliseconds for both MUA and LFP. This suggests that linear spectrotemporal encoding in AI relies on the temporal precision of firing rates, as opposed to vocalization duration. The best MUA-based decode performance in PBr, however, utilized the 400-millisecond data duration. Whether this means that PBr could be utilizing a firing rate-based coding schema remains to be determined. Effects of data duration may also be a product of working memory capacity, as opposed to auditory processing.

One possible limitation of this study holds true for all auditory decodes of natural stimuli that vary in length: The varying length of vocalizations could be a potential source of

artificial discriminability among vocalizations. Trial-averaged spectrograms from both AI and PBr showed a broadband burst of activity coincident with the vocalizations' end, which could contribute to principal component reconstruction and therefore skew the results. However, the raw data suggest that this is not the case. PBr spectrograms demonstrated bursts of low frequency power at the beginning of the vocalizations, and high frequency power at the end of the vocalizations, and therefore represented the duration of the vocalization in the raw responses. If these indicators of vocalization length accounted for artificial discriminability between classes, decodes in PBr would perform much better than the current results suggest. PBr showed very little decodability despite the presence of the spectral information about vocalization length. The AI spectrograms show far more structure during the call, which may account for more overall between-class variance and temporal precision than the features associated with the end of vocalizations. This overall variability is likely what PCA is operating on in the decode.

4.3 Frequency resolution of LFP responses

Previous LFP decodes have focused on motor and visual modalities using similar classification methods as the current study [39, 40]. These and other LFP decodes selected features on the basis of defined neural bands. For example, spectral data near the gamma range were the most useful for decoding rat limb movements (70-120 Hz) [39], bi-stable visual perceptions (50-70 Hz) [40], and high-frequency LFP (100–400 Hz) provided the most information about monkey limb position [41]. Our study used frequencies between zero and 300 Hertz as features in the decode. We show that increasing the frequency resolution of the spectrograms used in the decode from 400 to 1024 points improves decode performance. Performance was increased in both AI and PBr when frequency resolution was increased.

A recent study reconstructed auditory stimuli that human patients heard using LFP recorded from the surface of the posterior STG with high-density micro-electrocorticographic (μ ECoG) electrodes as well as standard clinical ECoG electrodes [42]. This study found using high gamma band (75-150 Hz) LFP produces the highest accuracy in reconstructing auditory stimuli. In addition, stimulus reconstruction accuracy was improved using a nonlinear model, which was based on spectrotemporal modulations, when compared with a linear spectrogram. This result indicates that encoding in PBr may utilize nonlinear encoding schemes, or PBr may represent spectrotemporal modulations in a stimulus more than linear changes in spectrotemporal features of a stimulus. This type of encoding scheme, as well as sparse representations in PBr, or the possibility that vocalizations are not the optimal stimuli for PBr, could account for why PBr decodes were close to chance.

4.4 Spatial scale and electrode density

The LFP-based decode results over electrode densities provide evidence for a lower limit on functional electrode density for electrophysiological recording of about one electrode per square millimeter (inter-electrode spacing of 1.14 mm). This inter-electrode spacing is larger than those utilized in many penetrating electrode arrays, yet smaller than ECoG inter-electrode spacing (\sim 1 cm) and the reported lower limit on spacing is smaller than measurements of LFP activation in acoustic space [45]. The lower limit for electrode spacing found in the current study is almost double the optimal spacing predicted by spatial spectral models, which report a best electrode spacing of 0.6 mm [46]. The reported spacing is, however, on par with μ ECoG electrodes, which show independent processing at 1- and 1.4-mm electrode spacing [34].

We show that MUA-based decode performance increases as more channels are added to the decode, at a higher density. Whereas classification accuracy for MUA-based decodes over

all electrodes is almost 30% higher than the best individual electrode, LFP-based decodes over all electrodes is near the best electrode performance. This result indicates that variability over electrodes is adding to decode performance to a greater extent in the MUA-based decode than in the LFP-based decode. The smaller spatial scale of information from neurons provides more channels of information.

Electrical micro-stimulation of the cerebral cortex is likely to act at the scale of LFP and cortical columns, rather than the scale of individual neurons. Recent work has shown that stimulation thresholds in visual cortex are lower when more channels are used for stimulation, and that it was possible to evoke spatially distinct visual percepts with micro-stimulation of $\sim 1 \text{ mm}^2$ of primary visual cortex [43, 44]. While the spatial resolution results presented here are inadequate for determining perceptual resolution of electrical micro-stimulation of auditory cortex, these results provide the first fixed geometry electrode array study of primate AI and provide some insight into characterizing spatial elements of an electrical interface with AI. How spatial and temporal parameters of micro-stimulation of auditory cortex affect perceptual discrimination remains an open question that needs to be examined through micro-stimulation of auditory cortex.

4.5 Implications for an auditory cortical neural prosthesis

Although the cochlear implant is a successful neural prosthesis for the treatment of deafness, there is some need for an auditory neural prosthesis that can bypass the auditory nerve. There are at least a hundred cases per year in the United States of cochlear nerves being destroyed by neuromas brought on by neurofibromatosis type II. These patients would be excellent candidates for a cortical auditory neural prosthesis. In addition to showing functionality of the type of neural interface that could be used in a stimulating cortical prosthetic, the current study provides insight into the physiology of the auditory ventral stream that is useful for design and use of a neural prosthesis. Through exploration of cortical coding we address several aspects of implementing an auditory cortical prosthesis.

While microstimulation of AI in trained rats has been shown to elicit behavioural responses similar to those elicited by tones [3], whether this stimulation evokes auditory perception that is useful in constructing an auditory object remains to be determined. Intracortical microstimulation that allowed for discrimination behaviour in rats used a 200-Hz stimulation rate [3]. The current study suggests that high temporal precision is an important feature of coding in AI. Rapidly dynamic temporal stimulation patterns may therefore prove more successful for encoding information in an auditory cortical neural prosthesis.

In addition to potentially meeting the needs of a patient group without an auditory nerve, an additional benefit is the possibility for achieving increased frequency resolution through interface with core auditory cortex. This concept is derived from the large accessible area of core auditory cortex (approximately 24 mm^2 in humans and 20 mm^2 in macaques) [47]. It is possible to fit several 96-electrode microelectrode arrays on the human core auditory cortex, thereby oversampling the tonotopic map in two core areas (AI and rostral core) [47]. Between the lower limit on electrode density determined by electrophysiology in this study and the actual density of the electrode arrays we used, 112 to 768 electrodes could be fit on macaque core auditory cortex.

Because of vascularization and cortical folding, application of penetrating electrode arrays in the human auditory cortex may face challenges, yet methods exist to implant depth electrodes with microelectrode contacts similar to those in a μECoG , which can take advantage of the full tonotopic map in humans [48]. Early human visual prostheses utilized surface electrode grids to evoke phosphenes from stimulation of primary visual cortex [49] and macroelectrodes on STG to evoke varied auditory perceptions [8]. Stimulating with

high-density μ ECoG in primary cortical areas may be a functional alternative to using penetrating electrode arrays, which may have complex interactions with the cerebral cortex [43]. A recent study, which could have particular application to primate AI, provided proof of principle for stimulating the brain with a high-density μ ECoG inside a sulcus [50].

There has been a great deal of work done on implantable neural prostheses for the cochlear nucleus, as well as the inferior colliculus, to address the need to bypass the auditory nerve. The potential benefits of bypassing these areas and stimulating the cortex directly are surgical ease, patient safety, and increased frequency resolution. This study suggests that high, dynamic stimulation rates in AI could be a feasible solution for dramatically increasing the channel counts and frequency resolution with a cortical auditory neural prosthesis.

5. Conclusions

We have examined auditory stimulus coding at early stages along the hierarchical processing stream in neocortex in order to assess the possibility for a cortically based auditory neural prosthesis. We report a lower limit on electrode spacing for electrophysiological recording interface with AI, encoding in AI may improve with higher-resolution temporal and spectral information, and linear spectrotemporal coding for stimulus identity is higher in AI than PBr. These results together serve as a design input for a human auditory cortical neural prosthesis, and provide guidance on interface design, microstimulation parameters, and anatomical placement.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

An NIH NIDCD 5 T32 DC008553-02 training grant and University of Utah startup funds supported this work.

References

- [1]. Schwartz MS, Otto SR, Shannon RV, Hitselberger WE, Brackmann DE. Auditory brainstem implants. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*. 2008; vol. 5(1):128–36. [PubMed: 18164492]
- [2]. Lim HH, Lenarz M, Lenarz T. Auditory midbrain implant: A review. *Trends in Amplification*. 2009; vol. 13(3):149–80. [PubMed: 19762428]
- [3]. Otto KJ, Rousche PJ, Kipke DR. Cortical microstimulation in auditory cortex of rat elicits best-frequency dependent behaviors. *J. Neural Eng*. 2005; vol. 2(2):42–51. [PubMed: 15928411]
- [4]. Rousche PJ, Normann RA. Chronic intracortical microstimulation (icms) of cat sensory cortex using the utah intracortical electrode array. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*. 1999; vol. 7(1):56–68. [PubMed: 10188608]
- [5]. Sinha SR, Crone NE, Fotta R, Lenz F, Boatman DF. Transient unilateral hearing loss induced by electrocortical stimulation. *Neurology*. 2005; vol. 64(2):383–385. [PubMed: 15668450]
- [6]. Fenoy AJ, Severson MA, Volkov IO, Brugge JF, Howard MA. Hearing suppression induced by electrical stimulation of human auditory cortex. *Brain research*. 2006; vol. 1118(1):75–83. [PubMed: 16979144]
- [7]. Dobelle WH, Stensaas SS, Mladejovsky MG, Smith JB. A prosthesis for the deaf based on cortical stimulation. *The Annals of otology, rhinology, and laryngology*. 1973; vol. 82(4):445–463.
- [8]. Penfield W. Some mechanisms of consciousness discovered during electrical stimulation of the brain. *PNAS*. 1958:51–66. [PubMed: 16590173]

- [9]. Selimbeyoglu A, Parvizi J. Electrical stimulation of the human brain: Perceptual and behavioral phenomena reported in the old and new literature. *Front. Hum. Neurosci.* 2010; vol. 4(1):1–11.
- [10]. Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex* (New York, NY : 1991). 1991; vol. 1(1):1–47.
- [11]. Mishkin M, Ungerleider LG. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research.* 1982; vol. 6(1):57–77. [PubMed: 7126325]
- [12]. Mishkin M, Ungerleider LG. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav Brain Res.* 1982; vol. 6(1):57–77. [PubMed: 7126325]
- [13]. Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex.* 1991; vol. 1(1):1–47. [PubMed: 1822724]
- [14]. Hackett TA. Information flow in the auditory cortical network. *Hear Res.* 2011; vol. 271(1-2): 133–46. [PubMed: 20116421]
- [15]. Winer JA, Lee CC. The distributed auditory cortex. *Hear Res.* 2007; vol. 229(1-2):3–13. [PubMed: 17329049]
- [16]. Hefner HE, Heffner RS. Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by japanese macaques. *Journal of Neurophysiology.* 1986; vol. 56(3):683–701.
- [17]. Turken AU, Dronkers NF. The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Front. Syst. Neurosci.* 2011; vol. 5:1–20. [PubMed: 21347218]
- [18]. Romanski LM, Bates JF, Goldman-Rakic PS. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 1999; vol. 403(2):141–57. [PubMed: 9886040]
- [19]. Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci.* 1999; vol. 2(12):1131–6. [PubMed: 10570492]
- [20]. Hackett TA, Stepniewska I, Kaas JH. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* 1998; vol. 394(4):475–95. [PubMed: 9590556]
- [21]. Woods TM, Lopez SE, Long JH, Rahman JE, Recanzone GH. Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. *Journal of Neurophysiology.* 2006; vol. 96(6):3323–3337. [PubMed: 16943318]
- [22]. Kikuchi Y, Horwitz B, Mishkin M. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *Journal of Neuroscience.* 2010; vol. 30(39):13021–13030. [PubMed: 20881120]
- [23]. Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science.* 2001; vol. 292(5515):290–3. [PubMed: 11303104]
- [24]. Petkov CI, Kayser C, Studel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. *Nature Neuroscience.* 2008; vol. 11(3):367–374.
- [25]. Schnupp JWH, Hall TM, Kokelaar RF, Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *Journal of Neuroscience.* 2006; vol. 26(18): 4785–95. [PubMed: 16672651]
- [26]. Recanzone GH. Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. *Journal of Neuroscience.* 2008; vol. 28(49):13184–13193. [PubMed: 19052209]
- [27]. Russ BE, Ackelson AL, Baker AE, Cohen YE. Coding of auditory-stimulus identity in the auditory non- spatial processing stream. *Journal of Neurophysiology.* 2008; vol. 99(1):87–95. [PubMed: 18003874]
- [28]. Averbach BB, Romanski LM. Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex. *J Neurosci.* 2006; vol. 26(43):11023–33. [PubMed: 17065444]
- [29]. Nirenberg S, Pandarinath C. Retinal prosthetic strategy with the capacity to restore normal vision. *Proceedings of the National Academy of Sciences of the United States of America.* 2012

- [30]. Rousche PJ, Normann RA. A method for pneumatically inserting an array of penetrating electrodes into cortical tissue. *Annals of Biomedical Engineering*. 1992; vol. 20(4):413–422. [PubMed: 1510293]
- [31]. Rauschecker JP, Tian B, Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*. 1995; vol. 268(5207):111–4. [PubMed: 7701330]
- [32]. Shoham S, Fellows MR, Normann RA. Robust, automatic spike sorting using mixtures of multivariate t-distributions. *J Neurosci Methods*. 2003; vol. 127(2):111–22. [PubMed: 12906941]
- [33]. Bokil H, Andrews P, Kulkarni JE, Mehta S, Mitra PP. Chronux: A platform for analyzing neural signals. *J Neurosci Methods*. 2010; vol. 192(1):146–51. [PubMed: 20637804]
- [34]. Kellis S, Miller K, Thomson K, Brown R, House P, Greger B. Decoding spoken words using local field potentials recorded from the cortical surface. *J. Neural Eng.* 2010; vol. 7(5):056007. [PubMed: 20811093]
- [35]. Krzanowski WJ. *Principles of multivariate analysis*. 1990; vol. 3:xxii+563.
- [36]. Russ BE, Ackelson AL, Baker AE, Cohen YE. Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *J Neurophysiol*. 2008; vol. 99(1):87–95. [PubMed: 18003874]
- [37]. Bendor D, Wang X. Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. *Journal of Neurophysiology*. 2008; vol. 100(2): 888–906. [PubMed: 18525020]
- [38]. Scott BH, Malone BJ, Semple MN. Transformation of temporal processing across auditory cortex of awake macaques. *Journal of Neurophysiology*. 2010:1–75.
- [39]. Slutzky MW, Jordan LR, Lindberg EW, Lindsay KE, Miller LE. Decoding the rat forelimb movement direction from epidural and intracortical field potentials. *J. Neural Eng.* 2011; vol. 8(3):036013. [PubMed: 21508491]
- [40]. Wang Z, Logothetis NK, Liang H. Decoding a bistable percept with integrated time-frequency representation of single-trial local field potential. *J. Neural Eng.* 2008; vol. 5(4):433–42. [PubMed: 18971518]
- [41]. Zhuang J, Truccolo W, Vargas-Irwin C, Donoghue JP. Decoding 3-d reach and grasp kinematics from high-frequency local field potentials in primate primary motor cortex. *IEEE transactions on bio-medical engineering*. 2010; vol. 57(7):1774–1784. [PubMed: 20403782]
- [42]. Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF. Reconstructing speech from human auditory cortex. *PLoS biology*. 2012; vol. 10(1):e1001251. [PubMed: 22303281]
- [43]. Torab K, Davis TS, Warren DJ, House PA, Normann RA, Greger B. Multiple factors may influence the performance of a visual prosthesis based on intracortical microstimulation: Nonhuman primate behavioural experimentation. *J. Neural Eng.* 2011; vol. 8(3):035001. [PubMed: 21593550]
- [44]. Tyler Davis RP, House Paul, Bagley Elias, Wendelken Suzanne, Normann Richard A, Greger Bradley. Spatial and temporal characteristics of v1 microstimulation during chronic implantation of a microelectrode array in a behaving macaque. *Journal of Neural Engineering*. 2012
- [45]. Kajikawa Y, Schroeder CE. How local is the local field potential? *Neuron*. 2011; vol. 72(5):847–858. [PubMed: 22153379]
- [46]. Slutzky MW, Jordan LR, Krieg T, Chen M, Mogul DJ, Miller LE. Optimal spacing of surface electrode arrays for brain-machine interface applications. *J Neural Eng.* 2010; vol. 7(2):26004. [PubMed: 20197598]
- [47]. Hackett TA, Preuss TM, Kaas JH. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* 2001; vol. 441(3):197–222. [PubMed: 11745645]
- [48]. Reddy CG, Dahdaleh NS, Albert G, Chen F, Hansen D, Nourski K, Kawasaki H, Oya H, Howard MA. A method for placing heschl gyrus depth electrodes. *J Neurosurg*. 2010; vol. 112(6):1301–7. [PubMed: 19663547]
- [49]. Dobelle WH, Turkel J, Henderson DC, Evans JR. Mapping the representation of the visual field by electrical stimulation of human visual cortex. *Am J Ophthalmol*. 1979; vol. 88(4):727–35. [PubMed: 507145]

- [50]. Matsuo T, Kawasaki K, Osada T, Sawahata H, Suzuki T, Shibata M, Miyakawa N, Nakahara K, Iijima A, Sato N, Kawai K, Saito N, Hasegawa I. Intracal electrocorticography in macaque monkeys with minimally invasive neurosurgical protocols. *Front. Syst. Neurosci.* 2011; vol. 5:1–9.

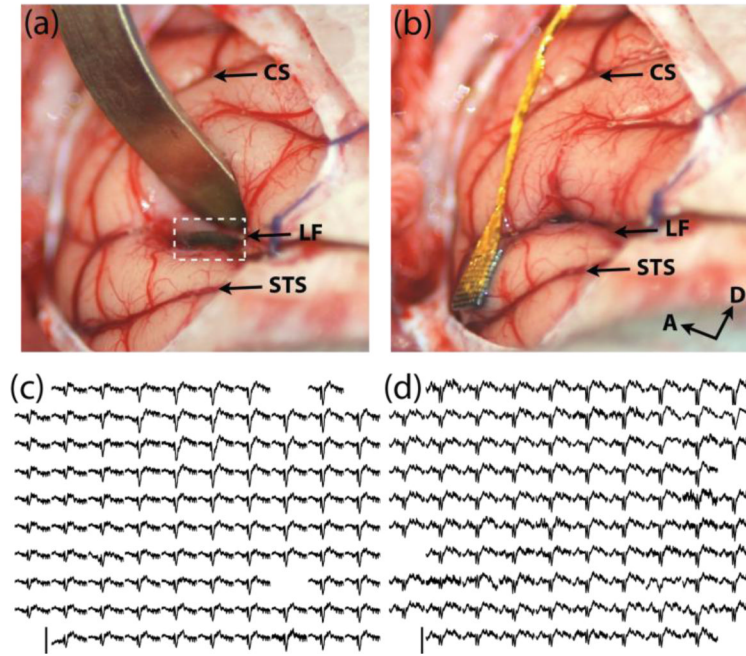


Figure 1.

Microelectrode array implantation in AI and PBr looking from the caudolateral perspective. (a) Retraction of the parietal lobe to expose the microelectrode array implanted several millimeters within the lateral fissure (white dashed box) (b) The implanted microelectrode array in PBr after AI array implantation. Acronyms: CS: Central Sulcus, LF: Lateral Fissure, STS: Superior Temporal Sulcus, A: Anterior, D: Dorsal. Averaged evoked potentials across the array footprint in AI (c) and PBr (d) in response to 15 presentations of the *harmonic arch* vocalization. Each evoked potential represents 1.8 seconds of averaged LFP data starting 0.2 seconds before stimulus onset. Amplitude scale bar extends from -500 to 300 microvolts in (c) and -300 to 200 microvolts in (d).

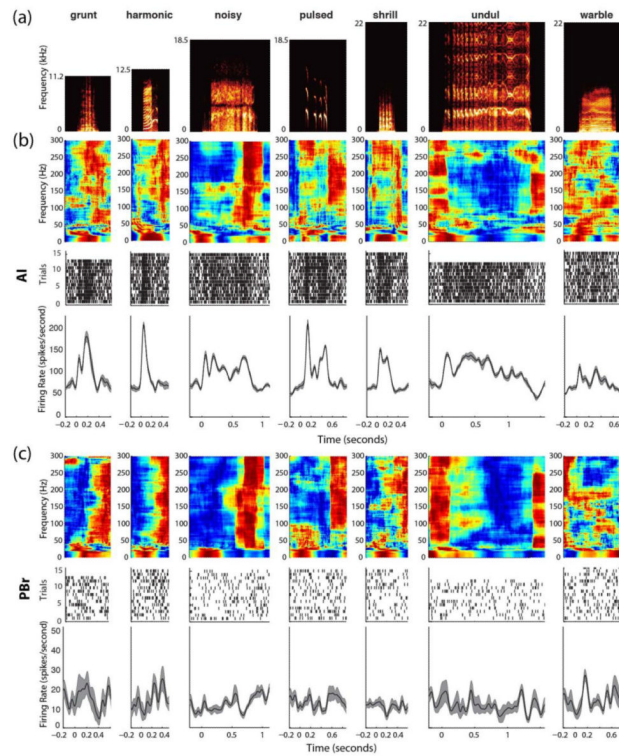


Figure 2. Neural data used for classification for each vocalization. (a) Spectrograms of sound stimuli for each vocalization. (b) Trial-averaged spectrograms of LFP responses to vocalizations for a single channel and trial-averaged PSTHs for Multiunit responses for AI. (c) Trial-averaged spectrograms of LFP responses to vocalizations for a single channel and trial-averaged PSTHs for MUA responses for PBr. All PSTH kernel widths are 25 msec.

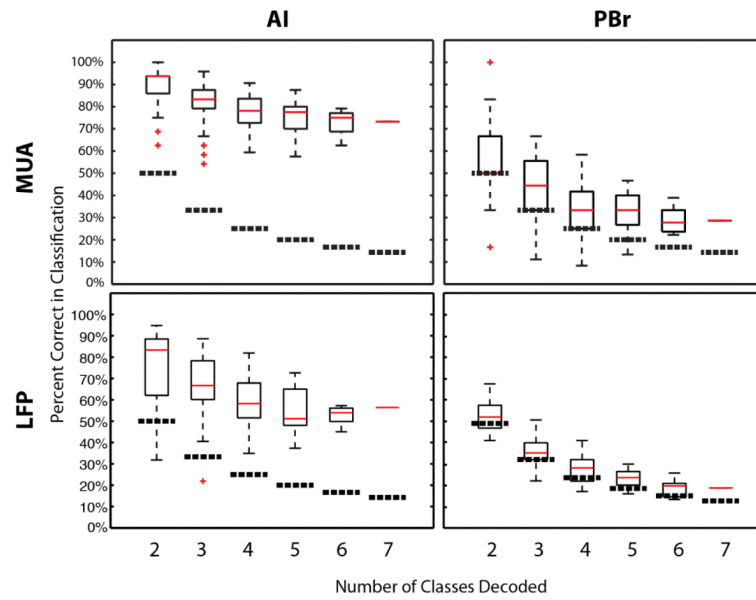


Figure 3.

Best classifier performance for MUA and LFP-based decodes. The top two panels show box plots for MUA-based decode performance for AI and PBr, while the bottom two show LFP-based decode performance. Mean performance is indicated by red lines across combinations of vocalizations for two through seven classes. Red crosses indicate outliers. Thick dotted lines indicate chance performance across vocalization combinations.

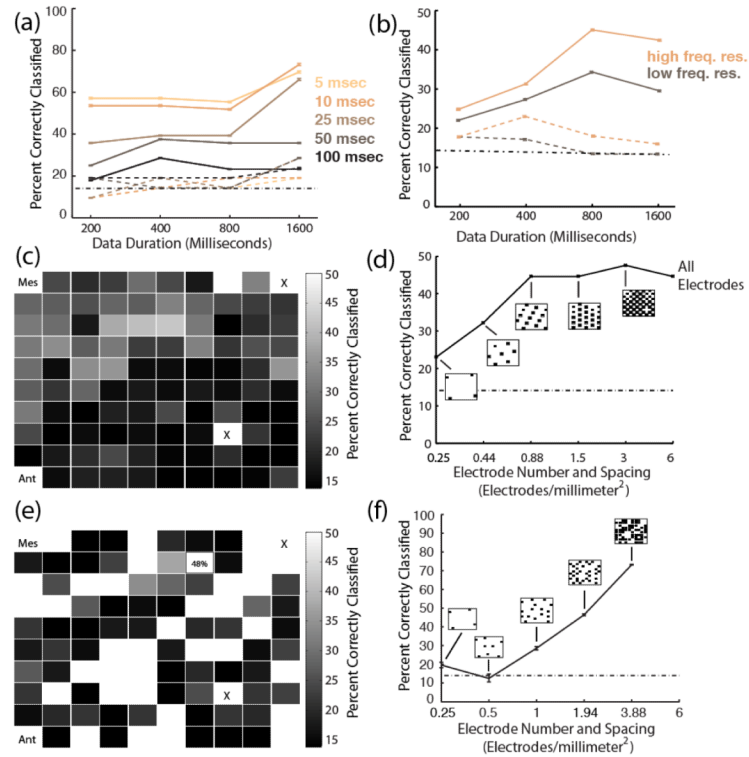


Figure 4.

Decode performance over time and space. (a) Seven-class performance for MUA-based decodes. Solid lines show performance for AI and dotted lines show performance for PBr. (b) Seven-class performance for LFP-based decodes. Solid lines show performance for AI and dotted lines show performance for PBr. Brown lines show performance for spectrograms with the lowest frequency resolution. Copper lines show performance for spectrograms with higher frequency resolution. Standard error bars are shown in (a), (b), (d), and (f); which are only slightly larger than the line width. The dash-dotted line in (a), (b), (d), and (f) indicates the level of chance. (c) LFP-based decode performance level for each electrode for seven classes is superimposed onto a map of the AI microelectrode array. “Ant” and “Mes” indicate the corners of the array that are pointing in the anterior direction (Ant) and towards the midline (Mes) on the STP. “X” indicates electrodes that were not connected. (d) LFP-based decode performance for different electrode densities across the array. Seven-class performance is shown for different electrode configurations, which are shown in black on the footprint of the AI electrode array for each electrode density tested. (e) Performance level for each MUA channel superimposed onto a map of electrodes with significant responses. “Ant” and “Mes” indicate the corners of the array that are pointing in the anterior direction and towards the midline on the STP. “X” indicates electrodes that were not connected. (f) MUA-based decode performance for different electrode densities. Seven-class performance is shown for different electrode configurations, which are shown in black on the footprint of the AI electrode array for each electrode density tested.

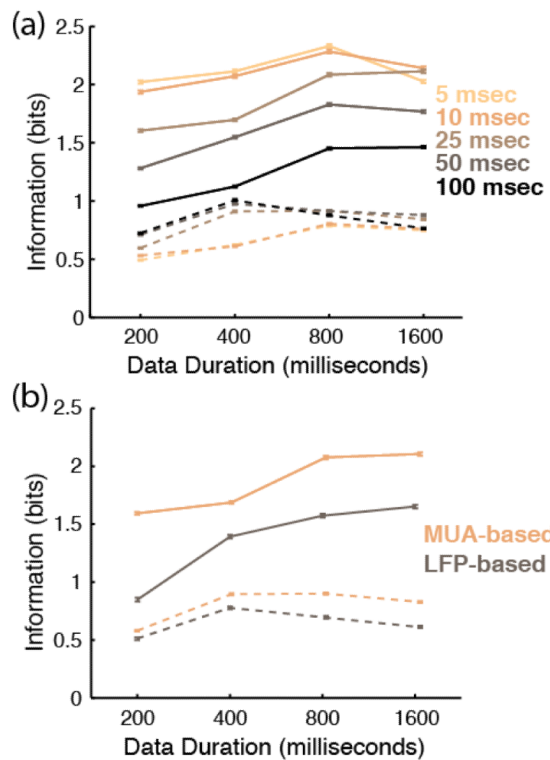


Figure 5. Information theoretic analysis. Solid lines show stimulus-response information content across data durations for AI. Dotted lines show stimulus-response information content across data durations for PBr. (a) Information content for five different kernel widths for the MUA-based decode. Different kernel widths are color coded. (b) Information content over data durations using the best performing kernel width and frequency resolution determined from Figure 4. Copper lines show information content across data durations for all MUA-based decodes. Brown lines show information content across data durations for all LFP-based decodes.