MACHINE LEARNING FOR LEARNING HOW THE BRAIN RECOGNIZES SPEECH AND LANGUAGE

Janet M.Baker[1], Alex M. Chan[2,3], Ksenija Marinkovic[4], Eric Halgren[4], Sydney Cash[2]

[1]Saras Institute, Newton, MA, USA
[2]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
[3]Harvard-MIT Division of Health, Science, and Technology, Medical Engineering and Medical Physics, Cambridge, MA, USA
[4]Department of Radiology, University of California, San Diego, La Jolla, CA, USA
janet_baker(atsign)email.com, amchan(atsign)mit.edu, xenia(atsign)ucsd.edu, ehalgren(atsign)ucsd.edu, scash(atsign)partners.org

ABSTRACT

Over the past several decades, automatic speech recognition has made great progress through the application of statistics and machine learning, combined with perceptual and structural knowledge about speech and language, as well as its variability. This paper reviews some recent work that applies some of these approaches to cortical processing of speech and language in the human brain to better understand how it functions. Specific experiments demonstrate feasibility for the discrimination of small sets of words (83% on 10 spoken words) and semantic categories (76% on 2 categories). This speech and language information is broadly distributed both spatially and temporally across the brain.

Index Terms: speech recognition, semantics, machine learning, brain, magnetoencephalography, electroencephalography, support vector machines

1. INTRODUCTION

This paper briefly reviews the time course of some extracranial and intracranial events in the brain responding to diverse speech and language stimuli. Machine learning techniques are applied comparatively to both magnetoencephalography (MEG) and electroencephalography (EEG) data to demonstrate the existence of semantic category and individual word discrimination information. There are a great many unknowns about the stages of processing speech and language in the brain. Nonetheless, a set of consistent processing stages have been well characterized. Some intriguing comparisons can be made to state-of-the-art automatic speech recognition systems. We also look at some practical issues in extending this approach, and potential applications in the future.

2. CORTICAL SPEECH PROCESSING

A variety of different kinds of processing are used to record brain activity. Different types are routinely used to extract different kinds of information. Each of these has unique advantages and disadvantages. Taken together, multiple recording modes are usually complementary. This section looks at the results of several types to illustrate cortical responses to speech and language stimuli.

As is seen in the examples following, expectancy plays a major role in how the brain responds to stimuli. Responses are typically observed to stimuli across a range of levels, from lexical to syntactic and semantic.

2.1. Intracranial Recordings

Intracranial recordings can be obtained from implanted multicontact depth electrodes, typically in epileptic patients who are being clinically evaluated in advance of surgery. Local field potentials are obtained from normal brain tissue as well as from damaged brain tissue to guide surgical excision of affected areas while minimizing disruption to unaffected areas, especially significant language areas. Magnetic resonance imaging (MRI) is performed to confirm subject brain anatomy and precise electrode placement. When presented with visual stimuli of words and sentences requiring appropriate word completion, subjects are observed to generate electrical activation patterns which demonstrate a sequence of responses corresponding to different language processing stages [1]. Consistent triphasic responses were obtained, localized within Broca's area in the ~200 msec, ~320 msec, and ~450 msec timeframes following the onset of presentation. Recordings outside of Broca's areas demonstrated different patterns of responses. Subjects were instructed to focus on thinking silently about simply reading or appropriately inflecting a set of nouns, and verbs, and sometimes to press a key in response. Subjects reading individual words ("Read [*rock*]") evoke a lexical response ~200 msec from stimulus onset. Less frequent words elicit a greater amplitude response. Words of different syllable length are not distinguished by the observed response. Subjects asked to inflect grammatically correct forms of nouns or verbs in a sentence final position (e.g. "Yesterday they [*walk*]"), evidenced significant responses 320 msec from stimulus onset. At about ~450 msec, the phonological and phonetic encoding processes required for articulation are completed. For this last phase, the evoked response is greater for more complex polysyllabic utterances.

2.2 Electroencephalography (EEG)

Electroencephalography (EEG) is a noninvasive recording technique which measures summed electrical activity. Event-related potential (ERPs), are averaged EEG recordings which are typically time-locked to a specific stimulus event.

In 1980, Kutas and Hillyard [2] reported that when subjects are presented with semantically anomalous or ambiguous sentences, they generate significant ERP responses with a negative component in the 400 msec time frame following the onset of such a triggering event. Subjects were presented with a series of 7-word sentences where the final word was 1) semantically typical, or 2) moderately or 3) extremely unusual. Typical samples of semantically meaningful sentences include "He took a drink from a glass" or He spread his bread with butter". A moderately unusual sentence might be "He took a sip from a waterfall." A more

extreme case could be "He spread his warm bread with socks" or "He took a sip from the transmitter".

Another quite robust ERP positive voltage marker is the P300 which is typically associated with detection of "odd ball" phenomena (responds to unusual events, not just language-related) observed by the subject. This response is currently used by some video game players to activate game events. Also disabled patients have been taught to exercise control or make computer menu choices by triggering a P300 response. It typically is strongest over the parietal lobes though it may be generated by other parts of the brain.

Syntactic and other anomalies ("garden path sentences") can often be found to elicit a positive ERP component known as the P600. Research conducted by Thierry et al [3] examines the elicitation of P600 in subjects reading Shakespeare. Shakespeare is well-known not only for creating many thousands of novel words in the English language (e.g. eyeballs, tongue-tied, manager, etc.) but also for using words in unusual ways. He frequently made use of "functional shifts" whereby one word class is used as another in semantically meaningful constructions. Examples include "He childed as I fathered", "Strong wines thick my thoughts", and "You said you would companion me." Subjects reading these constructions typically produced P600 responses, flagging the syntactic violations, but not evoking the N400 semantic violations marker. Although frequently attributed to the detection of syntactic errors, it is also believed that the P600 may arise from a reanalysis or reprocessing/repair stage of language processing more generally. Linguistically mediated ERPs may also occur well beyond 600 msec from stimulus onset as individuals engage in yet further processing stages to interpret jokes.

Several of these ERP responses are also elicited in other domains as well as with speech and language. Subjects responding to unexpected arithmetic terms in a sequence [4], also elicit ERP P600 responses. Out-of-key chords in a musical chord progression can give rise to a late negativity at about 500 msec [5], [6], as well as a P600 [7] Out-of-key chords also can reliably give rise to Early Right Anterior Negativity (ERAN) responses in the 160-180 msec timeframe, for both musicians and nonmusicians [8]. There has been much discussion over how linguistic and music processing might or might not be dependent on some of the same neural substrates. Interestingly, an interaction between linguistic and music processing has been demonstrated where priming with either a sentence or musical excerpt followed by a semantically unrelated word, will produce an N400 response [9].

2.3 Magnetoencephalography (MEG)

Magnetoencephalography is a recording technique that measures changes in the magnetic field due to changes in electrical activations of the brain. Although similar in many regards to EEG, MEG samples the electrical currents differently and thus each provides information complementary to the other. Because the MEG signals are so extremely weak, heavy magnetic room shielding and very sensitive cryogenic detectors are required. In the experiments reported here, a helmet containing 3 (1 magnetometer plus 2 gradiometers) sensors at each of 102 uniformly spaced sensor locations (~2-3 cm apart) are embedded within a spherical dewar helmet that covers the subject's head, excluding their face. EEG signals can be recorded simultaneously with MEG. Both extracranial EEG and MEG signals, as well as intracranial EEG (iEEG) recordings, have the advantage of high

temporal resolution (msec) in comparison with other imaging techniques (secs or mins) including fMRI, positron emission tomography (PET), and single photon emission tomography (SPECT). These other imaging techniques have an advantage in providing much higher spatial resolution. Neither EEG nor MEG have a unique inverse solution to identify the precise areas from which their signals are derived. Nonetheless, utilizing various constraints can help reduce spatial ambiguities. Employing multiple complementary techniques enables the collection of better information than for any single technique.

3. EXPERIMENTAL PROCEDURES

3.1 Subjects, Data, and Speech/Language Tasks

Subjects consisted of 9 healthy, right-handed, native English-speaking males between 22 and 30 years of age. MRI scans were obtained and confirmed the absence of any apparent structural brain abnormalities. This data, experimental procedures, and prior modality-specific analysis have previously been reported in detail by Marinkovic, et al [10].

MEG data was recorded from a 306-channel Electa Neuromag Vectorview System with sampling at 600 HZ and filtering from .1 to 200Hz. A Polhemus 3Space Isotrack II system was used to determine the locations of the magnetic coils relative to critical anatomic points for precise coregistration with MRI images. A 64-channel EEG cap was utilized to obtain simultaneous EEG recordings, also sampled at 600 Hz with filtering from .1 to 200 Hz.

Two speech and language tasks were employed here to explore subject responses, separately to auditory (SA) and visual (SV) stimuli. Two sets of disjoint word lists were prepared, approximately balanced for word frequency, syllable, and letter counts. One set was used for the SA visual presentation and the other for the SV auditory presentation in two sessions about 4 months apart. Each set of words was evenly split between animals and objects, and within each category, the choice of animals or objects was further evenly divided between those that were smaller or larger than a foot in length. For the SA task, the individual words were recorded by a single male speaker and were digitally normalized for amplitude and duration (500 msec). Trial presentations of these audio recordings were made to subjects binaurally through plastic tubes, with one noun presented every 2.2 sec. For the SV task, words were presented as white letters on a black background projected in front of the subject, at 2 sec. intervals. At each SA and SV session, subjects were presented with a succession of trials, half of which were "repeat" words (10 representative words presented multiple times) and the balance of which were novel words, each presented only once.

For each SA or SV trial, the subject was asked to make a "size judgment" to determine if the animal or object noun presented was larger than 1 foot in any dimension (e.g. bobcat, crib) or smaller than 1 foot (e.g. buckle, amoeba). If and only if the observed noun, designated as a "target word", was larger than 1 foot, the subject was instructed to press a button with their left hand. Words smaller than 1 foot (in any dimension) were designated as "non-target words". Accuracy in performing this task was assessed for each subject.

During the SA and SV sessions, the repeat and novel words were randomized. A total of 390 trials were presented for the SV task, and 780 trials for the SA task. Within each set of the 10 repeat

Figure 1: Schema showing amplitude-based feature extraction used to train nonlinear SVMs to determine decision boundaries in both binary and multiclass sets.

Figure 2: Decode Accuracy for EEG, MEG, EEG+MEG
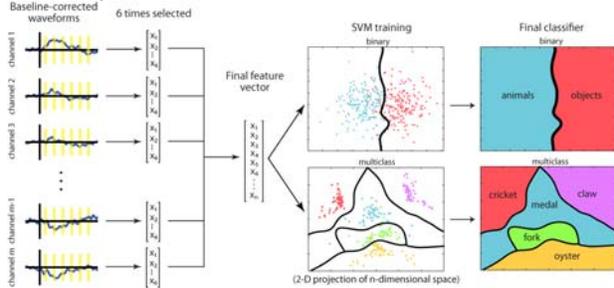
words for the SA and SV tasks, respectively, half of the words are animals and half are objects. Each set of the 10 repeat words for the SA and SV tasks contain 5 large or target words and 5 small or non-target words. The 10 SA repeat words are cricket, oyster, claw, fork, medal, serpent, lion, shirt, flag, and shelf. The SV repeat words are feather, mussel, lipstick, nametag, pistol, dinosaur, python, steer, banjo, and suitcase.

4. DATA ANALYSIS AND RESULTS

For both the EEG and MEG data, ERPs and event-related fields (ERFs) were computed. Individual channel signals were band-pass filtered from 1 to 30 Hz prior to determining averaged time-lock analyses. EKG, gross movement and other artifacts were automatically removed. A 500 msec pre-stimulus baseline was determined and used in computing corrected ERPs and ERFs. Custom MATLAB code incorporating routines from EEGLAB and Field Trip toolboxes (http://fieldtrip.fcdonders.nl) was used to compute timelocked averages aligned with stimulus onset times.

For each trial, amplitude features were sampled at 6 designated time points from each channel of these ERPs and ERFs, and concatenated, to create large feature vectors. The six time points chosen for decoding animals vs. objects, were 200, 300, 400, 500, 600, and 700 msec post-stimulus and for decoding individual words were 250, 300, 350, 400, 450, and 500 msec poststimulus (see Figure 1). These feature vectors thereby capture spatiotemporal characteristics in the cortical data for each trial.

Support vector machines (SVMs) were chosen to classify and discriminate between the different classes of categories (e.g. animal vs. object), and between individual words. This type of machine learning algorithm has been successfully employed in dealing with high dimensional data, to generate both linear and nonlinear decision boundaries. The SVMs employed here are MATLAB routines as implemented by Joachims [11] and Crammer and Singer [12]. Standard cross-validation techniques, always maintaining independent training and test sets, were used to evaluate classification accuracy. Between one trial (for individual words) to thirty trials (for animals vs. objects) were omitted in each cross-validation round before training the SVM on the remaining trials.

Figure 1 illustrates schematically the feature vector composition from a single trial, and the results of training a nonlinear SVM to to classify new words, based determine the boundaries between the two semantic classes (animals vs, objects) or between individual words. These boundaries are used during test to classify new words
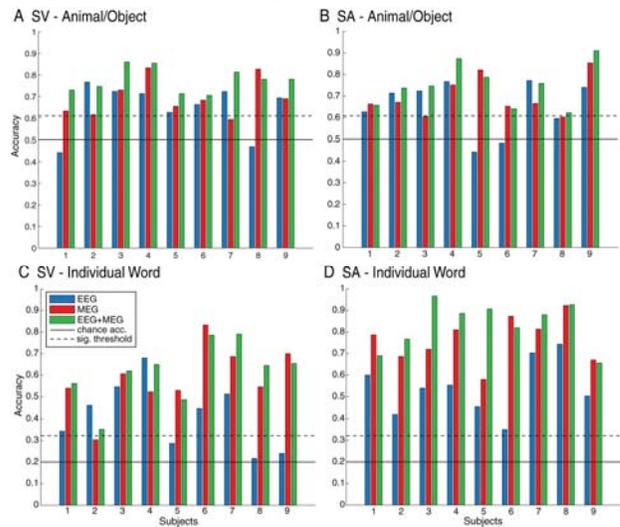
based on their feature vectors. Only novel or non-repeated words were used for distinguishing between animals and objects. In order to maintain larger sample sizes for training purposes, only repeat words were used in the tests to discriminate between individual words. Significance thresholds were determined by shuffling target labels and performing 1000 cross-validation tests.

For each of the 9 subjects, SVMs were trained separately on EEG features, MEG features, and both EEG and MEG features, 1) to discriminate between animals and objects, using only novel words, and 2) to discriminate between 5 individual words (separate sets of target and non-target words), using only repeat word data. Since subjects responding to target words were required to make a motor action pressing a button, it was decided to separate out the target and non-target word sets for the individual word decoding tests to avoid possible differential volitional motor effects between the two types of data. To minimize early auditory and visual effects, the time frames over which 6 evenly spaced points were chosen for the SVM features, was 200-700 msec, and for the individual words, 250-500 msec. **Figure 2** shows the decoding accuracy for these discrimination tests. Horizontal lines indicate chance accuracy (solid line) and significance thresholds (dashed line) for each test. For all SA and SV tests, statistical significance in discriminating between animals and objects, or discriminating between 5 individual words, is demonstrated for all 9 subjects when using combined EEG and MEG data. The discrimination accuracy benefit of combining both EEG and MEG data in training the SVMs, demonstrates their complementary characteristics. In computing averaged contributions across subjects for these two types of data and comparing the benefit of the combined data over either type alone, it is evident that the MEG data is making the greater contribution, in all 4 discrimination tests shown. When MEG SVM weights and data are projected onto 2-D topographic representations of the head, it is evident that both category and individual word discrimination information, for both the SV and the SA tasks, is quite broadly distributed both spatially and temporally [13]. Dynamic discrimination information, although prominently left-lateralized,, is also clearly evident bilaterally.

Using the SVMs trained separately, with EEG, MEG, and combined EEG and MEG data, to discriminate between the 10 individual repeat words for each of the SA and SV tasks, confusion matrices were generated. The highest accuracies are obtained by

the MEG SA data, followed by EEG SA, MEG SV, and then EEG SV. The confusion matrices obtained from MEG data alone are comparable to those combining EEG and MEG data. An analysis of variance (ANOVA) of the specific word errors by semantic category, either animals vs. objects or target vs, non-target words, demonstrates a striking dichotomy with confusion rates between words within a class, exceeding those between classes. For the SV target/non-target word discrimination test, the confusion rate between class mean$\pm$ s.e. = 0.0472+0.027 was much lower than the within class confusion rate = 0.125+0.045 (p<0.00001). For the SV animal/object test, the between class confusion = 0.074+0.037 was again lower than the within class confusion = 0.092+0.043 (p<0.005). Likewise for the SA target/non-target test, the between class confusion rate = 0.038+0.028 was significantly less than the within class confusion = 0.067+067+0.036 (p<0.00001). The average SA animal/object between class confusion test = 0.045+0.031 also was exceeded by the within class confusion = 0.058+ 0.034 (p<0.05). These results suggest the inclusion of semantic information in the cortical data recorded, and contributing to individual word discrimination. Additional details on data analysis and more related experiments on this data are available from Chan et al [14]. Extensions of this work using SVMs to decode intracranial EEG data also demonstrate significant individual word and semantic category discrimination [16]. Another way to use SVMs is to window the data and decode the individual frames to ascertain how much discrimination information is available in different time frames. In speech processing, signal processing information is typically windowed with overlapping frames (e.g. 20 msec frames overlapped by 10 msec). Here windows of 150 msec are computed, overlapped by 120 msec for the SA MEG individual 10 word discrimination task.
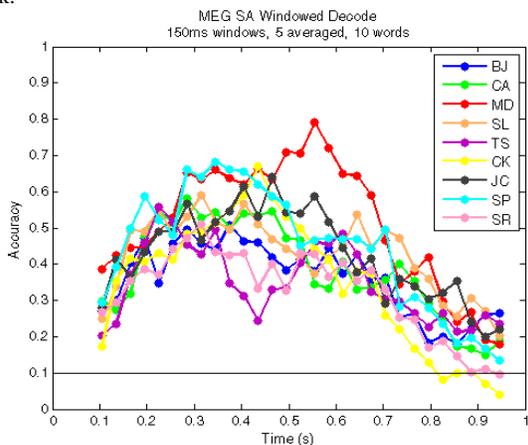
Figure 3 shows the results for all 9 subjects for the MEG SA data, throughout the 1second timeframe following stimulus onset. A horizontal line at 10% indicates chance performance. Discrimination accuracy for the individual 150 msec frames rapidly increases from about 100 msec, and by 300 msec averages about 50% across all the subjects. After about 600 msec, accuracy starts sharply declining, but remains well above chance even beyond 900 msec. Although broadly similar to one another, with multimodal peaks between about 200 and 750 msec,, the individual subject discrimination curves also evidence individual variability between subjects. This plot indicates a significant amount of discrimination information exists within these limited windows. Similar characteristics are also readily apparent in even shorter windows (e.g. 50 msec.). The evidence of this succession of information-rich windowed data frames, especially in conjunction with the broad spatial distribution of discrimination information

observed, all suggests that the cortical processing of speech and language data is part of a very robust, highly redundant system.

Figure 4 illustrates another perspective on how dynamic discrimination information grows cumulatively. This plot, also based on the SA task windowed MEG data (50 msec windows overlapped by 20 msec) for discriminating 10 individual words, shows the time course of accumulating discrimination information from stimulus onset. Already by 100 msec, discrimination accuracy exceeds chance for all subjects, by 200 msec, the average discrimination accuracy is about 35%, by 300 msec it is about 50%. By 600 msec, the average discrimination level is above 70%. At this point for most subjects, the discrimination response curves asymptote, and then start slowly declining after about 900 msec.
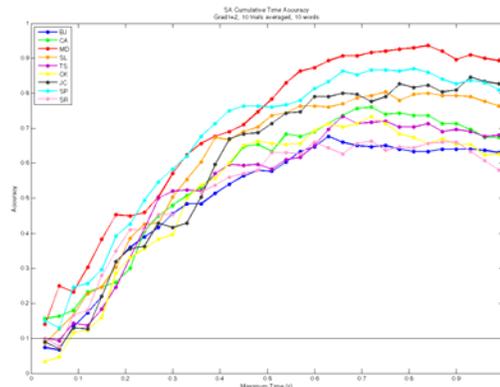
Figure 5 is identical to Figure 4 but has the addition of several vertical lines superimposed on it. These lines are positioned at ~200, ~320, and ~450 msec (corresponding to sequential stages of lexical, grammatical, and phonological information localized within Broca's area [1]), as well as the P300, the N400, and the P600. These markers, and the manipulation of them, may prove useful in helping us design future experiments to help unravel more information about how the brain processes speech and language, including across different modalities. Given that these time markers are observed from different contexts (incl. speech and music), with auditory as well as visual stimuli, this figure is not intended to illustrate a specific model. But rather it is intended for reflecting and being mindful of the parallel and serial processes that are transpiring during the time course of individual word recognition and other speech and language processing tasks.

Sequence of Some Observed Localized and Distributed Speech/Language and Music Processing Events
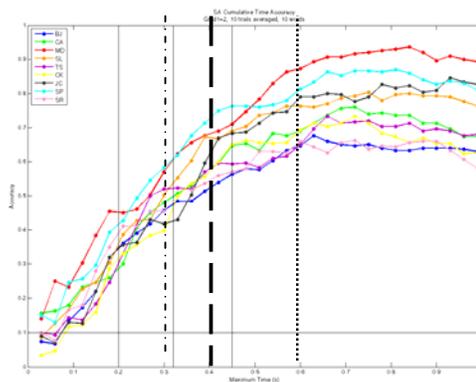
5. DISCUSSION OF COMPUTER VS CORTICAL PROCESSING OF SPEECH AND LANGUAGE

Though long used for speech and other pattern recognition challenges, the applications of machine learning techniques to cortical signal analysis have been somewhat limited. Some recent machine learning research by Mitchell et al [16] demonstrates the efficacy of this approach by successfully making predictions of fMRI brain activations to individual words based on their semantic associations.

Speech/Language Processing Components
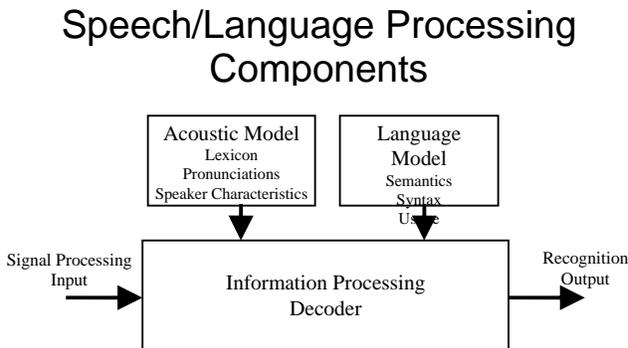
Figure 6: Basic system components for recognition of speech and language

When we compare what we know about cortical processing of speech and language to state-of-the-art stochastic processing based speech recognition systems, we see some intriguing similarities and contrasts. Although these systems are designed and engineered very differently, they exhibit some similar behaviors. Both types of systems process data dynamically in real-time as it becomes available. Biological systems "learn" characteristics of speech and language by example, extrapolation, and adaptation.

In automatic speech recognition systems, the major component acoustic and language models are represented probabilistically within network representations Speech recognition systems are "trained" by aligning examplars (training samples) with an underlying computable network representation. New incoming data is then decoded relative to these existing models, to determine the information they contain, by assessing the most likely state sequence observed in these underlying models. These models can also be refined and improved through adaptation. For both types of systems, biological or computational models are constructed, establishing a set of expectations around acoustic-phonetic representations, lexical components, the syntax and semantics of word usage and associations, speaker characteristics, environmental/channel noise, etc.

As we have seen in the preceding examples with the P300, N400, P600, etc., cortical responses to infrequent or unusual input can result in larger activation responses, at the lexical, semantic, and syntactic levels. Stochastic speech recognition systems typically respond to unusual or incongruous data by expanding or deepening their search through their network state spaces to explore more possibilities. This can be done at acoustic-phonetic levels as well as at semantic/syntactic levels. If recognition scores matching input speech data become poor; that is, fall below some predetermined threshold, many systems will then backtrack to

pursue other possibilities. Search parameters typically can be modified to meet memory space, time and accuracy requirements. In the event of ambiguities, both computer and cortical systems can take longer to process available data and make a determination. It appears that both types of systems may be able to reprocess and/or repair incorrect prior decisions.

One of the intriguing similarities of automatic speech recognition and cortical speech recognition is the presence in both, of a "Fast" or "Rapid Match" system which produces significant recognition results before the subject has heard the entire word being spoken! It is very evident in the data shown here in **Figures 3, 4, and 5**. The speech recordings prepared for the SA task, and presented to the 9 subjects, were slowly spoken, averaging about 500 msec. in duration. Inspection of the SVM dynamic decoding results in each of these figures, clearly shows high levels of discrimination well before 500 msec have elapsed from the onset of the stimulus. That the cortical system is able to do this, is again indicative of the rich information content and redundancy of the underlying speech signal.

Although we don't know how this is accomplished in the brain, there are a number of techniques for achieving this in automatic recognition systems. One of these methods [17], proceeds by combining, averaging, and smoothing several frames of data, and then comparing the similarity of these against typical similar sounding "word start groups" for each of the words being evaluated. Words that have dissimilar word start groups are thereby quickly eliminated from further consideration. Although more errors can be introduced by using these approximations, in practice, problems attributable to these approximations are minor, especially as compared with the significant savings in time and memory thereby achieved. As an aside, when this faster than real-time recognition was observed in the 1980s (even without Fast Match) for small vocabularies (e.g. 10 digits, etc.), it was soon discovered that to avoid user upset, it was important for the user interface to introduce a delay to wait until the speaker had finished speaking, and knew they were finished, before reporting the results to the speaker. Otherwise users became frustrated, felt they had been tricked, that the system was somehow reading their mind (!), etc. or just became confused.

So despite the drastic differences between computer and cortical processing systems, there are still some striking similarities!

For both types of systems, machine learning techniques can be applied to conduct rigorous quantitative research. It is hoped that this may prove to be a useful tool and probe 1) to better understand the structure and dynamics of speech and language understanding in the brain itself, and 2) to help form a robust basis for future productive applications. Better understanding how the brain processes speech and language could lead to improvements in automatic speech processing systems, which though useful in many regards, are still much inferior to cortical systems.

For understanding cortical processes, machine learning techniques can be applied both to localized responses, (e.g. intracranial studies) as well as distributed responses (e.g. EEG, MEG studies). Combining these approaches with complementary imaging (e.g. fMRI, etc.) and other studies enables a better spatiotemporal characterization, and a more complete and accurate picture to emerge of the underlying cortical processes. It is hoped that with improved understanding, opportunities will open up for improved medical therapies, prosthetic devices, educational applications, etc.

6. REFERENCES

[1] Sahin, N.T, Pinker, S., Cash, S. S., Schomer, D., Halgren, E., "Sequential Processing of Lexical, Grammatical, and Phonological Information Within Broca's Area", Science, Vol. 326: 445-449, 16 October, 2009.

[2] Kutas M. and Hillyard, S.A., "Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity", Science, Vol.207, No. 4427, 203-205, 11 Jan,, 1980.

[3] Thierry, G., Martin, C., Gonzalez-Diaz, V., Rezzaie, R., Roberts, N., and Davis, PM., "Event-related potential characterization of the Shakespearean functional shift in narrative sentence structure", Neuroimage 40:923-931, 2008.

[4] Nunez-Pena, M.I. and Honrubia-Serrano, M.L., "P600 related to rule violation in an arithmetic task", Cognitive Brain Research, Vol. 18, Issue 2, 130-141, January, 2004.

[5] Johnston, P.A., *Brain Physiology and Music Cognition*, Ph.D. Dissertation, Univ. of California, San Diego, 1994.

[6] Koelsch, S., Gunter, T.C., Friederici, A.D., and Schroger, E., "Brain indices of music processing: 'Non-musicians' are musical", Journal of Cognitive Neuroscience, 12:520-541, 2000.

[7] Patel, A. D, Gibson, E., Ratner, J., Besson, M., and Holcomb, P., "Processing syntactic relations in language and music: An event-related potential study", Journal of Cognitive Neuroscience, 10:717-733, 1998.

[8] Koelsch, S., Jentschke, S., Sammler, D., and Mietchen, D., "Untangling syntactic and sensory processing in language and music; an ERP study of music perception", Psychphysiology, 44:476-490, 2007.

[9] Koelsch, S., Kasper, E, Sammler, D., Schulze, K., Gunter, T., and Friederici, A., "Music, language and meaning: brain signatures of semantic processing", Nature Neuroscience, Vol. 7, No. 3., 302-307, March, 2004.

[10] Marinkovic, K., Dhond, R., Dale, AM., Glessner, M, Carr, V., and Halgren, E., "Spatiotemporal Dynamics of Modality-Specific and Supramodal Word Processing", Neuron, Vol. 38, 487-497, 8 May, 2003.

[11] Joachims, T. "Making Large-Scale SVM Practical" in Scholkopf, B., Burges, C. J. C. and Smola, A. J., eds., Advances in Kernal Methods-Support Vector Learning, Cambridge: MIT Press, 1999.

[12] Crammer, K., and Singer, Y., "On the algorithmic implementation of multiclass kernal-based vector machines", Journal of Machine Learning Research, 2:265-292, 2002.

[13] Baker, J.M., "Cortical Processing of Speech and Language Information" presentation for the Plenary Panel on Language and Cognition, Interspeech 2009 Conference, Brighton, UK, 7-10 Sept., 2009.

[14] Chan, A.M., Halgren, E., Marinkovic, K., and Cash, S. "Decoding word and category-specific spatiotemporal representations from MEG and EEG", paper in submission, 2010.

[15] Chan, A.M, Halgren, E., Marinkovic, K., Baker, J.M., Carlson, C., Devinsky, O., Doyle, D., Kazniecky, R., Thesen, T., Wang, C., Schomer, D., Eskandar, E., and Cash, S., "Decoding multiscale word and category-specific spatiotemporal representations from intracranial EEG", poster presentation at Fourteenth International Conference on Cognitive and Neural Systems, at Boston University, Boston, MA, 19-22 May, 2010.

[16] Mitchell, T.M., Shinkareva, S. V., Carlson, A., Chang, K. M., MaLave, V. L., Mason, R. A., Just, M. A., "Predicting Brain Activity Associated with Meanings of Nouns" , Science, Vol. 320, 1191-1195, 30 May, 2008.

[17] Gillick, L. and Roth, R., "A Rapid Match Algorithm for Continuous Speech Recognition", in the Proceedings of the DARPA Speech and Natural Language Workshop, held at Hidden Valley, PA, pp 170-172, June, 1990.