Nathan Rasmussen
Psych 7695.04
Autumn 2014

## Methods

**Oddball experiment.** Two visual and two auditory stimuli were used, X and O for visual and 200 Hz and 400 Hz tones for auditory. Subjects were presented with rare and common stimuli in eight blocks: four each for auditory and visual, covering four combinations of two factors. One factor was the (disclosed) assignment of the two stimuli to response keys; the other was the (undisclosed) assignment of the two stimuli to rare and common conditions. The sequence of the eight blocks was randomized per subject.

Response keys were selected so as to produce similar, minimal motor artifacts. The stimulus assigned to the common condition was presented 80% of the time, with the other 20% going to the rare stimulus. The sequence of presentations within each block was randomized per subject. Presentations were jittered in order to average out any signal that might carry over from one epoch to the next.

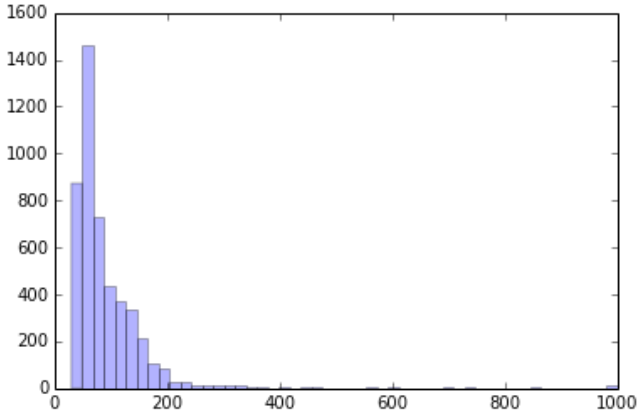
Subjects were instructed before each block as to the sensory modality and the response key assignment. They were given a break after each block. They also performed a practice block, with two stimuli at equal frequencies, before the first experimental block.

**Data collection and preprocessing.** Each participant was recorded on 32 channels at international 10/20 standard locations: Fp1, Fp2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8, O1, Oz, and O2, with a forehead ground. Data was rereferenced offline to the average of the mastoids (TP9 and TP10) and Cz was dropped. Data were high-pass filtered at 0.25 Hz to eliminate DC drifts and resampled to 250 Hz to eliminate ringing from the filter.

EEG recordings were matched to behavioral data using a series of sync pulses generated by the experiment. Generated pulses and recorded pulses were aligned or skipped, according to their spacing, and a line regressed through the aligned pairs to calculate offsets between the EEG data and the logged experimental events.

Eyeblinks were detected and corrected with WICA, a wavelet version of Independent Component Analysis. This algorithm first reduces the dimensionality of the channel data, producing components that capture covariance between channels. Components that weigh heavily on Fp1 and Fp2 are considered to potentially capture eyeblinks and are selected for cleaning after human inspection. Selected components are then decomposed spectrally to separate the high-amplitude, low-frequency signal of an eyeblink from other, potentially meaningful covariance. The components are reassembled without the blinks and the channel data reconstituted from the clean components.

Finally, remaining dirty data was excluded from analysis by dropping any event whose data had kurtosis < 5 on any channel, or a swing of more than 200 microvolts on any channel. High kurtosis is the kind of peaky signal that more likely comes from static electrical discharges than from neural dipoles diffused by a skull, and the threshold of 5 is standard practice. Large swings suggest jostled electrodes, and the threshold of 200 is chosen semi-empirically after viewing a histogram of voltage swings:




**Data analysis.** The underlying approach to identifying significant differences between EEG signals in different conditions is Student's t-test. The mean voltage is taken across events for each channel and each time point after the epoch, the difference between means for two conditions is calculated within each subject, and then t-values are calculated across subjects on these differences. High t-values suggest significant differences between the two conditions at the specified time point on the specified channel. However, this measure of significance, if applied naively to EEG data, involves a vast number of comparisons, making it difficult to set a meaningful significance threshold that anything will clear.

We therefore wrap the t-test procedure in two modifications. First, we enhance the truly meaningful differences with threshold-free cluster analysis, taking advantage of the high temporal and spatial autocorrelation of EEG data. This opens a gap in which we can place our significance threshold. Second, we determine where to place the threshold by empirically determining what kind of results do happen by chance.

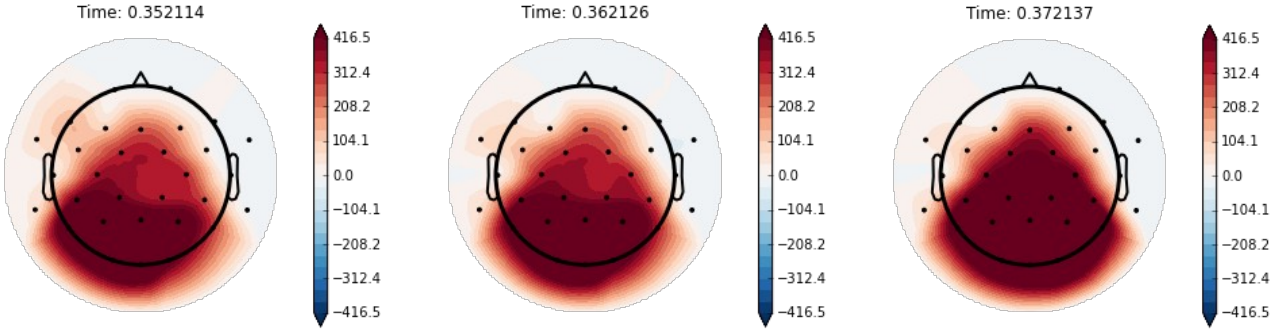
Cluster analysis augments the naïve t-values in proportion to the heights of their neighbors, which in this context means both adjacent time points and adjacent electrode locations. It proportionally rewards both the height to which all values in a contiguous region attain, and the breadth of that region. This sharpens tall peaks with adequately broad bases far more than it does isolated tall peaks or small wide hills, and this enhances precisely those differences that interest us.

This nonlinear rescaling of our data requires an appropriate significance threshold. We calculate this by simulating 500 non-significant experiments and observing the distribution of cluster-enhanced t-values. Specifically, in each of 500 cases we randomly permute the labels of the two conditions of interest in our data. Mean amplitudes calculated over permuted labels will therefore reflect only the effects of chance. After calculating t-values and clustering, we take the largest value emerging from each random permutation and set a threshold at the $97.5^{th}$ percentile of these 500 maxima. This is the value such that our procedure will deliver equal or larger results by chance 2.5% of the time, which brings us back to the usual definition of *p*-values. A comparable negative threshold can be found at the $2.5^{th}$ percentile. (I note that the distribution of clustered values is typically quite asymmetrical, so the two thresholds are probably not equal distances from zero. I am not sure if our software takes account of this.)

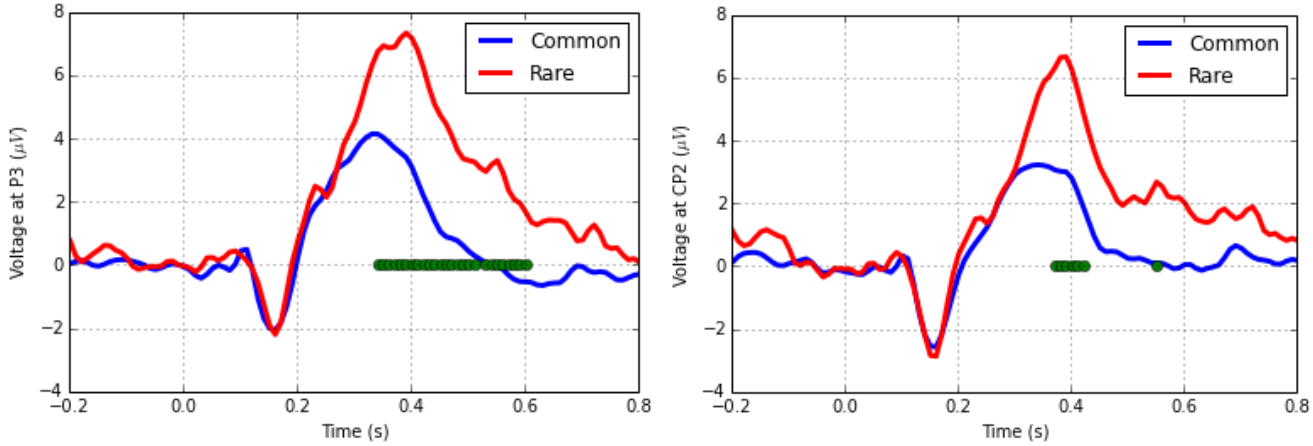
## Results

**Rare and common stimuli.** For the difference in means between rare and common

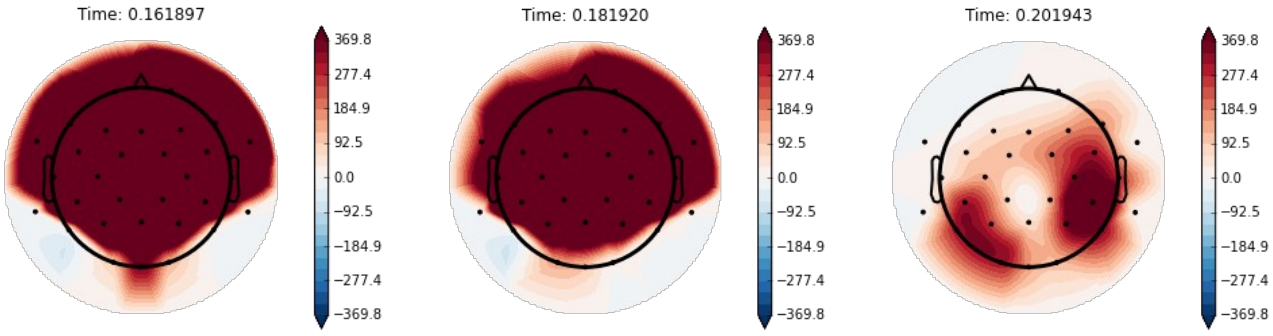
stimuli, clustered t-values rose to significance only from 352 to 372 ms after stimulus, and only in one contiguous region. At 352 ms, the center of this region is left parietal, extending forward to the central-parietal line, backward to the occipital, and rightward across the parietal. By 372 ms, the region has become less lateralized, and extends forward almost to the centerline, with a halo of values nearly as large extending further forward still.



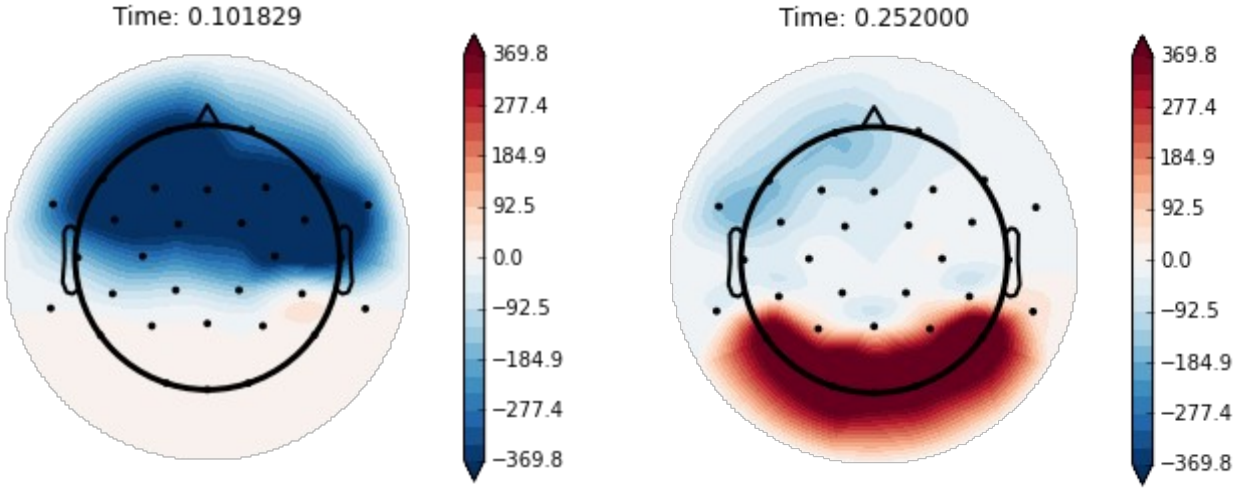

We will take the P3 electrode as typical of this distribution, and the CP2 to represent its fringe. With green dots to mark the times of significant clustered t-values, these are their grand average ERPs:



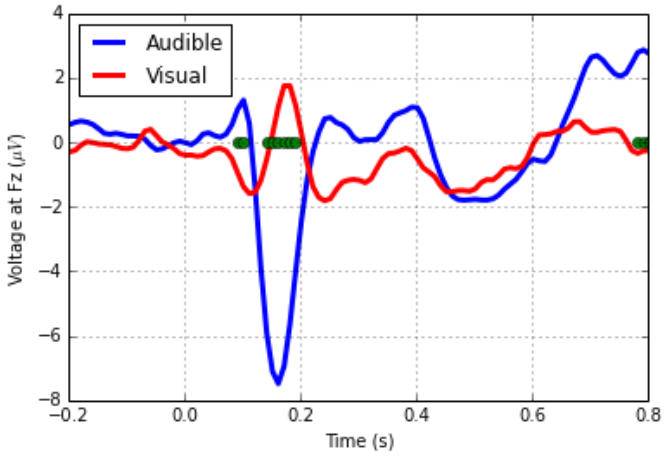

**Sensory modality.** We now investigate the differences between the visual and auditory stimuli, without regard for their frequency or rarity. For mean amplitude with visual stimuli minus mean amplitude with auditory stimuli, the clustered t-values were significant as early as 152 ms and as late as 202 ms. After this, presumably, processing is able to abstract away from modality to more remote concepts (such as common and rare). Significance obtains over a long list of electrodes, with the visual signal stronger in every case. Three snapshots summarize the differences: A widespread early difference, a slight collapse in the forward left, and then at 200 ms an abrupt collapse.

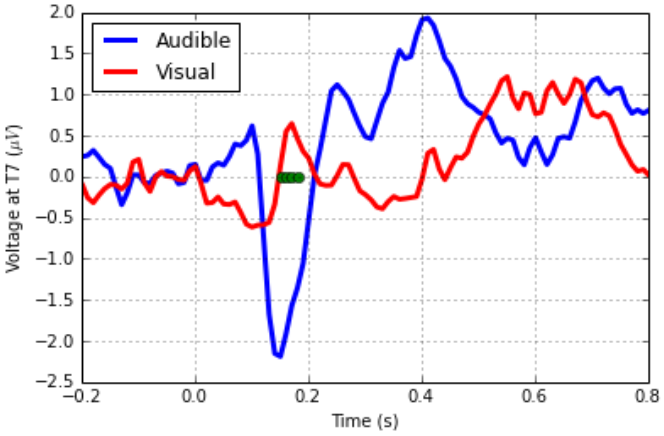


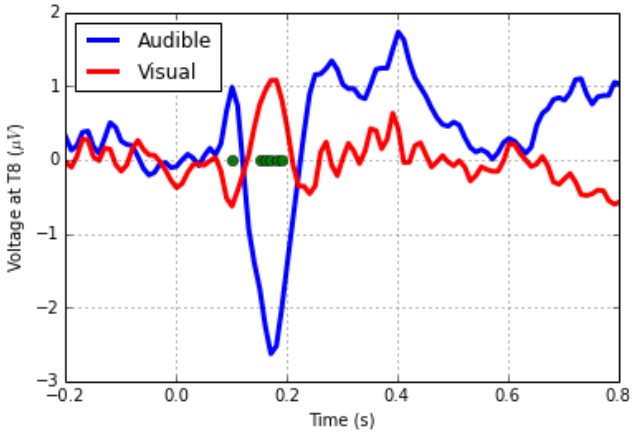

However, this picture is deceptive. It is not that visual processing simply produces a massively stronger signal than auditory. Rather, these significant differences show a frontal and lateral negative dipole in auditory processing, slightly preceding an occipital, positive one in visual processing. This is easily visible when looking around the neighborhood, i.e. 50 ms before and after the region of significance:

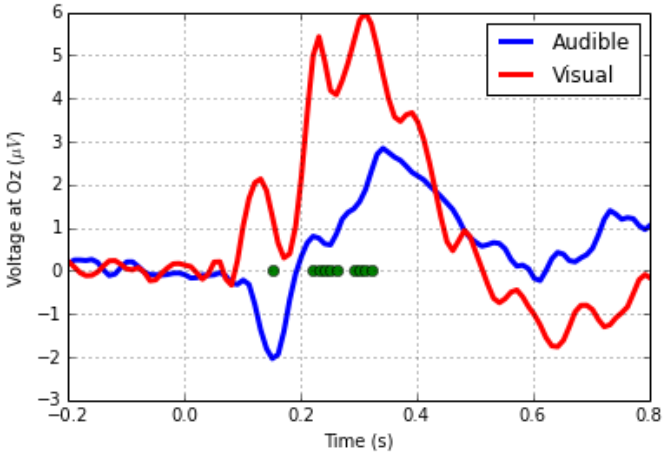



Here we see the sorts of signals we might expect: Auditory processing in the temporal region and forward (the prefrontal area presumably being involved in categorizing the stimuli and executing the button-press), and visual processing in the occipital lobe. Grand average ERPs from around the scalp tell the tale:




I would have liked to explore the different signals for correct and incorrect quick responses, where 'quick' is operationalized as each subject's first quartile of response times, hoping to catch a distinctive signal for either 'yeah, I got this' or 'oops'. Unfortunately, I was unable to figure out how to calculate percentiles over the RTs per subject and prepare an appropriate index vector. I'm sure NumPy and the RecArray can do this, I just did not find out how.

**Experimental design**

**Motivation.** Certain grammatical constructions, such as relative clauses, connect two words that may be separated by an arbitrary number of phrase and clause boundaries. These connections are known as unbounded dependencies. Naturally, maintaining an incomplete grammatical structure for an arbitrary length of time potentially poses a challenge to the human sentence processing mechanism.

Past studies have indeed shown behavioral and electrophysiological correlates of dependency length. As dependency length increases, accuracy on grammaticality judgement tasks drops, coherence between different brain regions rises, and the P600 signal of grammatical integration difficulty changes in amplitude (growing if the task requires subjects to attend to details of syntax, and shrinking if it does not). Some of these studies have used unbounded dependencies only, and others have also used bounded dependencies, but there is reason to think the two kinds of dependencies may be processed differently.

This experiment aims to identify the practical limits of unbounded dependency processing. We will use a grammaticality judgement task and incentivize subjects to attend to the grammar of difficult sentences by offering a bounty on correct answers.

**Stimuli.** Stimuli will be audio recordings, delivered at normal speaking rates (90 to 130 words per minute), of English sentences. Critical items will use three unbounded-dependency constructions, shown with the head and dependent bolded: Relative subject extraction is exemplified by *Their preferred **music**, which **is** called "trance," was invented in Germany.* Free relatives are exemplified by ***What** injures freshmen's health most **is** jumping into Mirror Lake.* Subject extraction from embedded clauses is exemplified by *The **lawyer**, who Kelly suspects **is** corrupt, didn't mind screwing the plaintiff over.*

Each critical item will be in one of two grammaticality conditions, created by altering agreement morphology at the end of the dependency: *Their preferred **music**, which **is/*are** called "trance," was invented in Germany.*

Each critical item will also be in one of three lengths, here summarized by placing optional material in parentheses: *Their preferred **music** (for chilling out (when the main dance floor is too stimulating)), which **is/*are** called "trance," was invented in Germany.* Each subject will encounter only one of these six forms of a given item.

Filler items will manipulate other aspects of grammar, including agreement in bounded dependencies, and will also appear in a range of lengths.

**Analysis.** In a critical item, let the epoch be the utterance of the word that determines grammaticality. With this epoch calculate a cross-subjects average ERP for each combination of grammaticality and length. (We will need 30-40 recordings for each of these.) Identify the peak amplitude of P600, and subtract this amplitude in the grammatical case from its counterpart in the ungrammatical case. Regress this difference against the length of the dependency (in words, not seconds) and check this regression for significance. Data from the recordings will also be usable for many analyses of greater sophistication, but this is a start.

**Predictions.** P600 amplitude is understood to reflect the difficulty of syntactic integration. It is increased both by ungrammaticality and by long dependencies, provided that subjects are motivated to attend to grammatical correctness (if they are not, it is *decreased* by long dependencies). With short dependencies, the effect of grammaticality will result in a large difference between peak amplitudes for the two conditions, but with long dependencies this effect will be washed out by the effect of length. The x-intercept of the regression line represents the dependency length at which the average subject can no longer hold on to grammatical

features well enough to recognize agreement.