EEG classification of imagined syllable rhythm using Hilbert spectrum methods

Siyi Deng, Ramesh Srinivasan, Tom Lappas & Michael D'Zmura*

*Corresponding Author

Address

Department of Cognitive Sciences
University of California, Irvine
SSPA 3151
Irvine, California 92697-5100
USA

Email

mdzmura@uci.edu

Abstract

We conducted an experiment to determine whether the rhythm with which imagined syllables are produced may be decoded from EEG recordings. High density EEG data were recorded for seven subjects while they produced in imagination one of two syllables in one of three different rhythms. We used a modified Second Order Blind Identification (SOBI) algorithm to remove artefact signals and reduce data dimensionality. The algorithm uses the consistent temporal structure along multi-trial EEG data to blindly decompose the original recordings. For the four primary SOBI components, joint temporal and spectral features were extracted from the Hilbert spectra (HS) obtained by a Hilbert-Huang Transformation (HHT). Hilbert spectra provide more accurate time-spectral representations of non-stationary data than do conventional techniques like short-time Fourier spectrograms and wavelet scalograms. Classification of the three rhythms yields promising results for inter-trial transfer, with performance for all subjects significantly greater than chance. For comparison, we tested classification performance of three averaging-based methods, using features in the temporal, spectral, and time-frequency domains, respectively, and the results are inferior to that of the SOBI-HHT-based method. The results suggest that the rhythmic structure of imagined syllable production can be detected in non-invasive brain recordings, and provide a step toward the development of an EEG-based system for communicating imagined speech.

Submitted to

Journal of Neural Engineering

1. Introduction

Recent advances in EEG technology have caused a surge of interest in the development of a brain-computer interfaces (BCI). BCI research often focuses on finding a substitute for a broken mind-body chain that can help paralyzed patients move and communicate. Much research has been directed toward building a brain-computer interface which would allow its user to avoid the need for motor output by allowing direct cortical control of prostheses or machines (*e.g.*, Donchin *et al*. 2000, Birbaumer *et al* 2000, Craelius 2002, Blankertz *et al* 2004). Research on classifying mind states using various forms of brain-imaging data has contributed to these advances. For example, scientists have successfully predicted the properties of unseen visual stimuli using fMRI signals from primary visual cortex (Haynes and Rees 2005). EEG signals recorded during imagined movement may also be classified (Ince *et al* 2006, Wang and Makeig 2009) and used for real-time control (Pfurtscheller *et al* 2006, Leeb *et al* 2007, Zhao *et al* 2009). Such results confirm that the human cortex generates signals observable extracranially which may be used to infer the internal state of the subject.

Research on decoding verbal thoughts from brain signals can be dated back to the 1960s, when attempts were made to transmit letters using Morse code extracted from EEG (Dewan 1967). BCI text generation based on the P300 oddball paradigm has been a particularly successful approach to communication based on brain signals alone (Wolpaw *et al* 2000, Schalk *et al* 2004). There are also potential non-clinical applications of BCIs for communication. In situations where it may be difficult or inappropriate to communicate vocally or gesturally, a BCI system for automatic imagined speech recognition could be particularly helpful.

A variety of neuroimaging studies show that there are large-scale cortical networks involved in speech signal processing and in language production (Bokde *et al* 2001, Weiss and Mueller 2003, Vigneau *et al* 2006, Hickok and Poeppel 2007). The dynamics of these networks can be traced by brain imaging techniques like EEG (Indefrey and Levelt 2004) and MEG (Ahissar *et al* 2001, Luo and Poeppel 2007). With an eye toward future development of an EEG-based BCI to communicate imagined speech, we conducted an experiment to determine whether the rhythm with which imagined syllables are produced may be decoded from EEG recordings.

EEG signal analysis for BCI applications typically uses filters or Fourier transforms to isolate signatures in the time-frequency plane that may be used to classify cognitive events reliably. In this paper, we use *Hilbert Spectra* to explore EEG signals in the time-frequency domain (Huang *et al* 1998). Hilbert spectra are well-tailored to the analysis of nonstationary time-series. They provide a compact time-frequency representation of EEG signals and can surpass in precision traditional time-frequency representation techniques like short-time Fourier transform spectrograms and wavelet scalograms. We demonstrate that Hilbert spectral analysis can be used to identify features that may be used to classify imagined speech rhythms. A preliminary report of these data analyzed with conventional spectral analysis methods has been presented elsewhere (D'Zmura *et al* 2009b).

2. Methods

We describe here the EEG experiment on imagined speech and then turn to the Hilbert spectral analysis applied to these data. The analysis comprises seven steps. 1). A variant of the SOBI algorithm (Belchourani *et al* 1997) is used to determine linear combinations of channels (EEG electrodes at various scalp locations) which provide, in the least-square-error sense, an optimal decomposition of the raw data into a set of mutually orthogonal components. 2). An Empirical Mode Decomposition on the time-varying waveforms, aggregated across space, is used to break these down into Intrinsic Mode Functions (Huang *et al* 1998). 3). The instantaneous frequency and amplitude of each Intrinsic Mode Function are estimated to determine the Hilbert spectra. 4). The resulting Hilbert spectra contributing energy in the 3-20Hz frequency band are normalized by the PSD of the baseline interval. 5). The Hilbert spectra are binarized by setting a threshold on spectral amplitude. 6). Suprathreshold, proximal, time-frequency bins are aggregated to determine time-frequency "spots of interest". 7). Trials are classified predictively using various features of the resulting "spots of interest".

2.1 EEG Experiment

Seven subjects participated in the experiments. They sat in a dimly lighted room and were instructed to minimize body and eye movements while keeping their eyes open. Their task was to imagine speaking one of two syllables, /ba/ or /ku/, in one of three rhythms (see Figure 1). The first rhythm has the time structure {| 1.5 | 1.5 |}, the second has the structure {| 0.75 | 0.375 | 0.375 | 0.75 | 0.375 | 0.375 |}, and the third has the structure {| 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |}; the vertical lines "|" represent the expected times of imagined syllable production onset, while the numbers indicate the time intervals in seconds between imagined syllables. Each experimental trial started with a period during which the syllable and rhythm were cued; the cue period was followed by a period during which the subject imagined saying the cued syllable in the cued rhythm without any vocalization (D'Zmura *et al* 2009a).

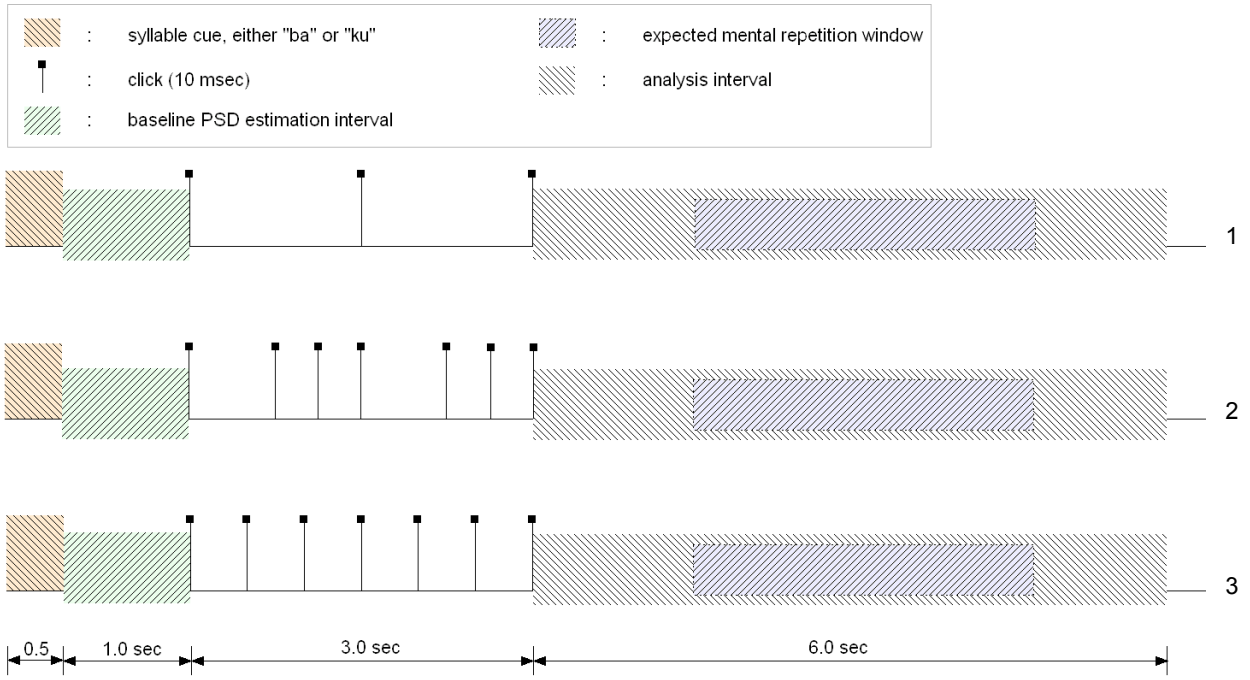


Figure 1. Trial time-course for the three rhythms. Syllable and rhythm cues were presented through earphones during an initial cue period of duration 4.5 sec. The syllable cue, either "ba" or "ku", was presented during the initial 0.5 sec. EEG recorded during the following 1.0 sec of silence was used as the baseline power spectral density estimator in offline analysis. During the following 3.0 sec period, a rhythmic cue was provided by clicks occurring at times appropriate to rhythm 1 (top), rhythm 2 (middle) or rhythm 3 (bottom). Subjects imagined speaking the cued syllable in the cued rhythm during the following production period of duration 6.0 sec. Refer to text for details.

Each trial lasted 10.5 seconds. During the first 4.5 seconds, the cues to syllable and rhythm were presented through earphones. The syllable cue, a spoken "ba" or "ku", was presented during the initial 0.5 sec of each trial. Clicks presented during the next 3.0 sec interval provided the cue to the rhythm. The cue period was followed by an imagined speech production period of duration 6.0 sec. Subjects were instructed to imagine the cued syllable using the cued rhythm and tempo identical to that indicated by the cue period clicks.

Syllable and rhythm cues were presented using STAX electro-static earphones. These headphones do not interfere with EEG recordings. EEG was recorded simultaneously using a 128-channel Geodesic Sensor Net (Electrical Geodesics Inc.) in combination with an ANT-128 (Advanced Neuro Technology) amplifier. The EEG was sampled at 1024 Hz and online average referenced. The presented cue signals were accompanied by sinusoidal markers which were fed directly into the EEG amplifier to mark the onset times of cue and imagined speech production periods of each trial.

Each subject performed 120 trials for each of three rhythms and two syllables, for a total of 720 trials recorded per subject. During data analysis, we performed classification on the data for all six conditions and for just three rhythms; the latter analysis was conducted by pooling trials of two syllables of the same rhythm to get 240 trials per rhythm. In our primary analysis, data from the 6.0 sec long imagined speech

periods were analyzed to determine whether one can decode imagined speech rhythm from EEG signals. In a second analysis, we used EEG data from the 3.0 sec long click cue period of each trial to find out whether the cue sensory signals can be decoded from EEG.

2.2 Blind source separation using SOBI

For each subject, we used a variant of Second-Order Blind Identification (SOBI) algorithm (Belchourani *et al* 1997) to decompose the raw EEG data into a set of mutually orthogonal components. The classical SOBI algorithm can be described as follows. Consider an n-channel EEG time series x with mean zero. Each sample is assumed to be an instantaneous linear mixture of n unknown sources s, where the mixing occurs through some unknown linear mixing matrix A:

$$x = As \tag{1}$$

This can be written equivalently as:

$$s = A^{-1}x \tag{2}$$

In order to calculate the unknown separation matrix A^{-1}, the data is first pre-whitened by finding a whitening matrix W so that the auto-correlation of x is normalized to identity:

$$z = Wx \tag{3}$$

$$R_{z,0} = E[(Wx)(Wx)^T] = WE[xx^T]W^T = WR_{x,0}W^T = I \tag{4}$$

In Eqns. 3 and 4, z is the whitened time series, $R_{z,0}$ is the (auto) correlation matrix of z at delay 0, $E[]$ denotes the expectation operator, and I is an Identity matrix of the same size as the autocorrelation.

The whitening matrix W can be found by diagonalizing $R_{x,0}$ as follows:

$$R_{x,0} = PDP^T \tag{5}$$

in which D is a diagonal matrix, whereby

$$W = D^{-1/2}P^T \tag{6}$$

After pre-whitening, a set of cross-correlation matrices $R_{z,\tau}$ is calculated for certain time delay values τ. In choosing time delay values for this study, we followed the suggestion of Tang and colleagues (2005) that the delays span a wide range, up to and including the Nyquist frequency. In this study we use the values of

$$\tau = \{2,4,10,20,40,80,120,250,400,512\} \tag{7}$$

These numbers of samples may be converted to delays (sec) using the sampling rate 1024 samples/sec. The key procedure of SOBI is to find a unitary transformation V that jointly diagonalizes $R_{z,\tau}$ by minimizing the criterion (Cardoso 1998)

$$\sum_{\tau}\sum_{i\neq j}\left(V^{T}R_{z,\tau}V\right)_{i,j}^{2} \tag{8}$$

The estimate of the separation matrix A^{-1} is thus provided by the product of the whitening matrix W (Eqn. 6) and the unitary transformation matrix V that provides joint-diagonalization (Eqn.8):

$$A^{-1} = VW \tag{9}$$

A problem which limits the application of classical SOBI is that the transformation basis of recovered components is not consistent across trials. By assuming a stationary distribution for the neural basis, we can extend classical SOBI to multi-trial recording by determining a single, approximate, pre-whitening solution and by jointly diagonalizing the cross-correlation matrices for the same set of time delays across trials. First, the across-trial pre-whiten solution is computed as

$$E\left[R_{z,0}\right] = WE\left[R_{x,0}\right]W^{T} = I \tag{10}$$

Second, the within-trial cross-correlation term $R_{z,\tau}$ is changed to the across-trial expectation $E\left[R_{z,\tau}\right]$ for each time delay τ; the joint diagonalization is then performed on the expectations. Finally, after estimating the separation matrix, SOBI components are estimated separately for each trial by projecting the raw data onto the SOBI space.

Four of the across-trial SOBI components were chosen for each subject. We ranked the full set of components by energy and then selected, for each subject, the first four components with spectral content compatible with EEG signals and neurophysiologically-interpretable topography, consistent with dipole sources in the brain (Nunez and Srinivasan 2006). The spatial distribution of the mixing weights for the four components used for one subject in further analysis is shown in Figure 2. Selecting for further analysis these four SOBI components not only greatly reduces the dimensionality of the data but also screens out most artefact signals of non-cortical origin.

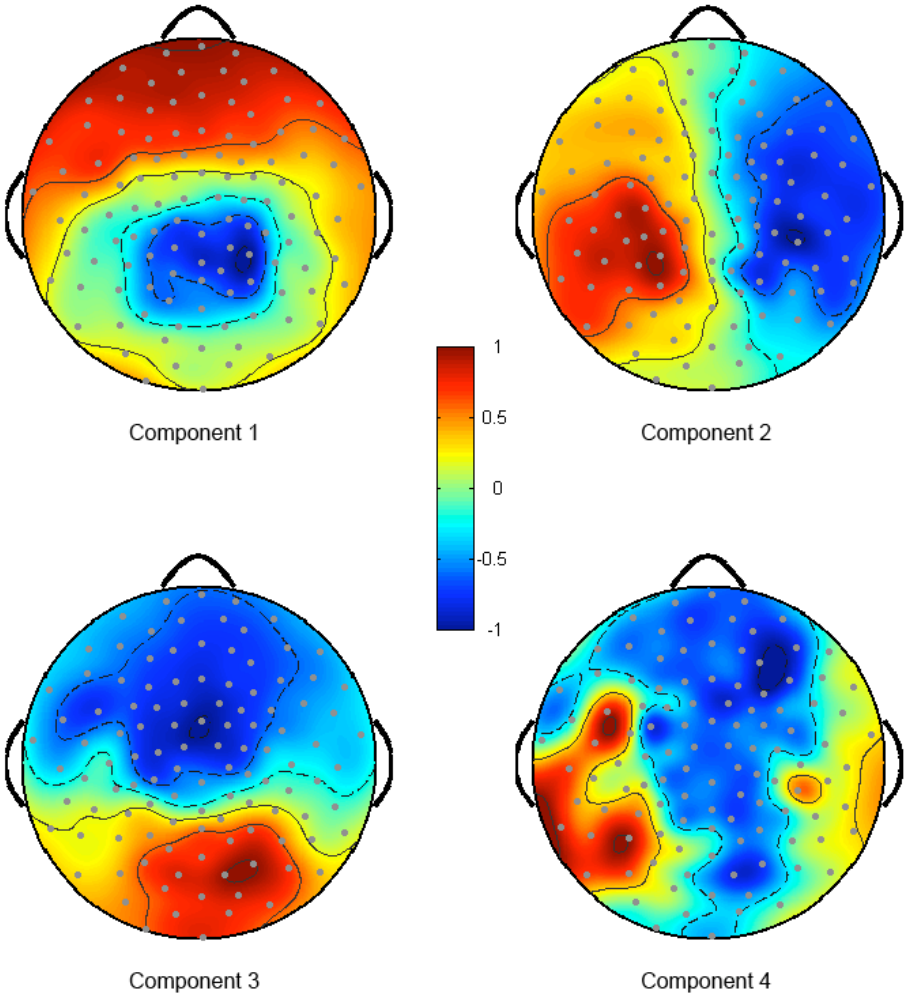


Figure 2. Cortical distributions of the four SOBI components of the mixing weights for subject 6. Values are normalized to the range of [-1 1]. Dark blue represents a mixing weight of -1 and dark red +1; mixing weight polarity is arbitrary.

2.3 Hilbert Spectrum

The Hilbert Spectrum (HS) is an emerging method of joint time-frequency representation of signals. Unlike classical time-frequency analysis methods, HS does not employ a constant or linear time *vs.* frequency resolution window. It adaptively tracks the evolution of the time-frequency basis in the original signal and can provide much detailed information at arbitrary time-frequency scales (Huang *et al* 1998). The fundamental step in calculating a HS is the Empirical Mode Decomposition (EMD). The EMD expresses a signal x in terms of a finite (and usually small) set of mathematically well-defined components c, called Intrinsic Mode Functions (IMFs), and a residue term:

$$x = \sum_{k=1}^{n} c_k + r \tag{11}$$

In practice, each IMF can be treated as a new time series and several IMFs can be superimposed to generate new signals. Figure 3 illustrates the IMFs (bottom) of a raw EEG waveform (top), with time series at left and their corresponding power spectra at right.

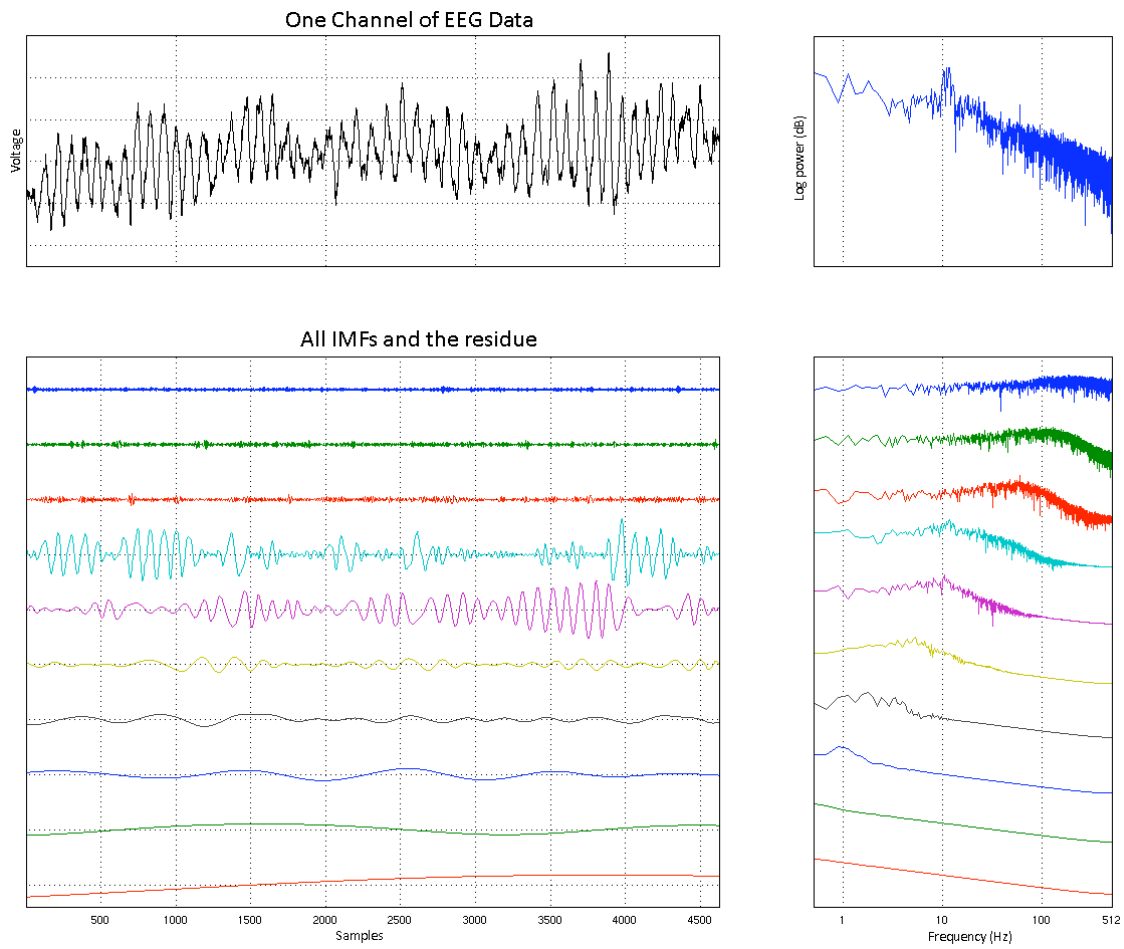


Figure 3. Example of the EMD and corresponding spectra for 4.5 seconds of EEG data for one electrode. The top left panel shows the time-course of the raw EEG. The bottom left panel shows all IMFs from the EMD, ranked downward with successive order, and the residue, which is the bottom-most component. The corresponding power spectra are shown on the right panels, in logarithmic scale. The spectra of IMFs overlap and differ from band pass filtered spectra.

The intrinsic mode function spectra, shown on the right of Figure 3, indicate that IMFs may be considered roughly as outputs of a set of band limited filters, but that the choice of bandwidths is adapted to the intrinsic modes of the original signal. Although IMFs of higher order (toward the bottom of Figure 3) generally have relatively greater energy in successively lower frequency bands, there is a considerable amount of spectral content overlap among IMFs.

Once IMFs are calculated, the instantaneous amplitude a and phase φ for each IMF could be computed from their analytical representations:

$$y = c + i\hat{c} = ae^{i\varphi} \tag{12}$$

$$a = |y| = \sqrt{c^2 + \hat{c}^2} \tag{13}$$

$$\varphi = \tan^{-1}\left(\frac{\hat{c}}{c}\right) \tag{14}$$

where $\hat{c}$ is the Hilbert transform of the IMF. The instantaneous frequency ω is defined as the rate at which phase changes with time:

$$\omega = \frac{d\varphi}{dt} \tag{15}$$

The original signal x can be expressed in a manner similar to a Fourier transform as:

$$x = Real \sum_{k=1}^{n} a_k e^{i\int \omega_k dt} + r \tag{16}$$

Finally, the Hilbert Spectrum $H(\omega, t)$ can be visualized by plotting the instantaneous amplitude a (or power) at the instantaneous frequency ω as a function of time t. A major difference between HS and classical time-frequency plots like the short-time Fourier transform is that the instantaneous functions in HS keep continuous traces of frequency change, resulting in a precise but sparse representation. The HS, spectrogram and wavelet scalogram of real EEG data are shown in Figure 4. The HS gives the most economic time-frequency representation without any compromise of resolution in either temporal or spectral basis. In addition, only HS retains the information of how trend frequency changes in the original time series. For example, if one compares all three time-frequency plots over the 3500th to 4000th sample region, one sees an increase of power in the alpha band around sample 3900. Yet only HS suggests clearly that the observed change is accompanied by the adjustment of central frequency toward the low alpha band of certain signal components (primarily by the blue, topmost, component shown in the HS plot at the bottom of Figure 4).

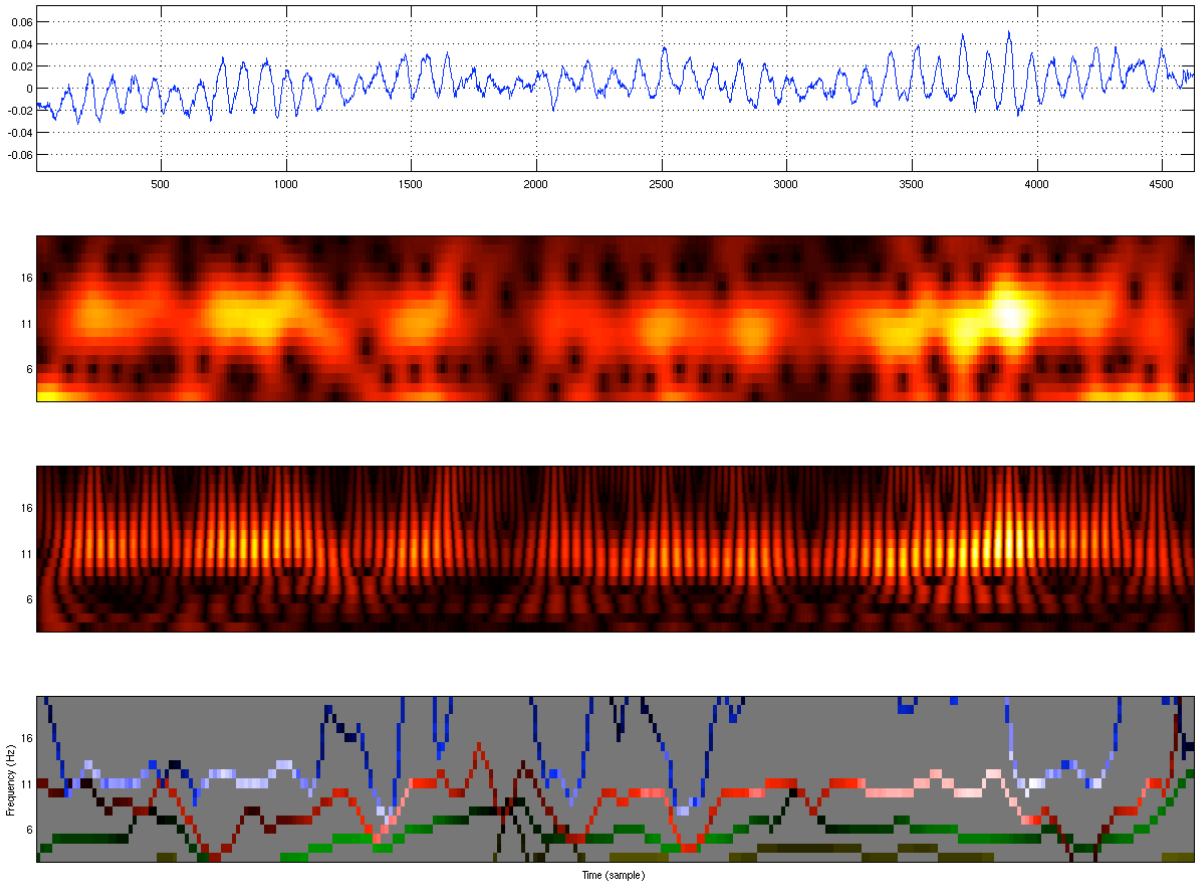


Figure. 4 Comparison of the spectrogram, wavelet scalogram and Hilbert spectrum of a same time series. Top Row: Original signal from SOBI output. Second row: Short time Fourier Transform spectrogram. Third row: Wavelet Scalogram, using a Morlet mother wavelet. Bottom row: Hilbert Spectrum, projected onto a [3-20] Hz x 384 time-point grid. Instantaneous frequency trends from different IMFs are coded with different colours but the intensity scale is kept the same (See figure 5 for the intensity scale).

2.4 Spectral density normalization

The spectral power distribution in the EEG is not uniform. A considerable amount of energy clusters around 8 ~ 14 Hz, or the alpha-band. Yet previous work suggests that useful information may be tracked in other frequency bands. For example, in Luo and Poppel's study (2007), signals following the envelope of the speech acoustic waveform were detected in theta and alpha bands (3 ~ 14 Hz). The beta band (16 ~ 30) has also been suggested to carry cognitive-related information. In a regular time-frequency plot, energy in bands other than the alpha is usually dwarfed by the alpha peaks. This suggests that one use normalized spectral densities if activities in different frequency bands are to be evaluated equally. In this study, we used the one second of data directly following the syllable cue as the baseline estimator for each trial (see Figure 1 and Figure 5). Baseline power spectral density between 3 to 20 Hz was

calculated by applying a FFT to the data and then fitting a spline along the frequency bins to determine a smoothly-fitting function $M(\omega)$. Next, the Hilbert Spectrum $H(\omega, t)$ of that trial was normalized using $M(\omega)$ along each time frame over the frequency window of 3 ~ 20 Hz to provide the spectral-density normalized Hilbert spectrum $\bar{H}$:

$$\bar{H} = \frac{H(\omega,t)}{M(\omega)} \tag{17}$$

2.5 Energy normalization

The total energy in $\bar{H}$ can vary dramatically across trials, partly from the task and especially in a way which depends on the alertness of the subject. Such energy variation can be problematic when setting thresholds on multiple trials. To ensure comparable spectral energy across different trials, we further normalized the total energy in each HS by the first quartile Q_1 of its power distribution:

$$\bar{\bar{H}} = \frac{\bar{H}(\omega,t)}{Q_1} \tag{18}$$

We use Q_1 instead of the median because the power distribution in EEG is skewed (Weineke *et al* 1980) and we expect Q_1 to be a more robust estimator of the spontaneous EEG activities.

2.6 Binary transform and feature extraction

After $\bar{\bar{H}}$ is obtained for all trials for a subject, a threshold is selected in order to find "spots of interest" (SOI). The threshold level was chosen so that any $\bar{\bar{H}}$ contains least one SOI and at most eight SOIs. Typically the threshold is about 50% of the range of $\bar{\bar{H}}$. After thresholding, $\bar{\bar{H}}$ is reduced to a binary representation. The binarization may result in many proximal yet disconnected suprathreshold time-frequency bins. A screening procedure was used to aggregate these proximal suprathreshold time-frequency bins. The procedure has three criteria. First, any pair of bins differing by 2Hz or less were grouped. Second, any pair of bins differing by 160 msec or less were grouped. Finally, isolated and small time-frequency bins or aggregates spanning less than 2Hz or less than 160 msec were dismissed form further analysis. The resulting time-frequency bin aggregates are the spots of interest shown in the bottom panel of figure 5. This figure shows the original HS (Fig 5.A); EEG power is distributed primarily around 12 Hz. After normalizing using the spectral density (Fig 5.E), energy at lower frequencies becomes more prominent (fig 5.B). Energy normalization using the first quartile (Fig 5.F) produces the fully normalized HS (Fig 5.C). The binary transform effected by applying an amplitude threshold picks up various suprathreshold time-frequency bins. Those proximal yet disconnected bins are grouped into fewer regions, the SOI, using the screening criteria (Fig 5.D).

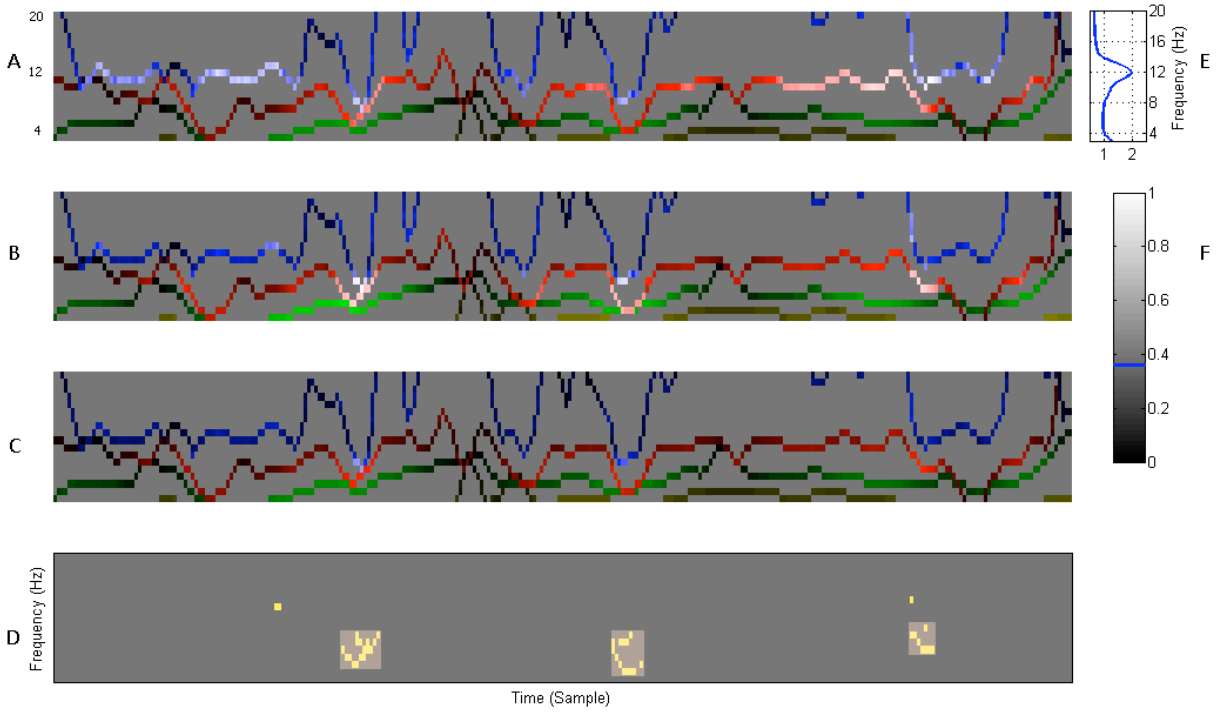


Figure 5. Preprocessing of the Hilbert Spectrum. A. Original Hilbert Spectrum H . B. H is Normalization by the baseline spectrum density (see E) to provide $\bar{H}$. The alpha band is generally suppressed by this normalization and lower frequency content emerges. C. $\bar{H}$ is normalized further by the first quartile of its power distribution to obtain $\bar{\bar{H}}$. D. $\bar{\bar{H}}$ is binarized by applying an amplitude threshold. Suprathreshold time-frequency bins are aggregated using screening criteria, detailed in the text, to provide "spots of interest"; three are shown in the bottom panel; screening criteria eliminate the two isolated suprathreshold bins from further consideration. E. The Spectral power distribution used for normalization, estimated for each trial from the 1 sec interval following the syllable cue. F. The intensity scale for A B and C. The binarization threshold is shown in blue on this scale. See text for details.

After the binary transformation and screening, the following time-frequency properties of the remaining SOIs were evaluated for each spectrum to generate the feature vector for data classification (see Figure 6): F1) the number of SOIs; F2) the average time distance between each two neighbouring SOIs; F3) the average time span of SOIs; F4) the average central frequency of SOIs, and F5) the average frequency span of SOIs. 20 featural data per trial (5 features by 4 SOBI components) were used for classification.

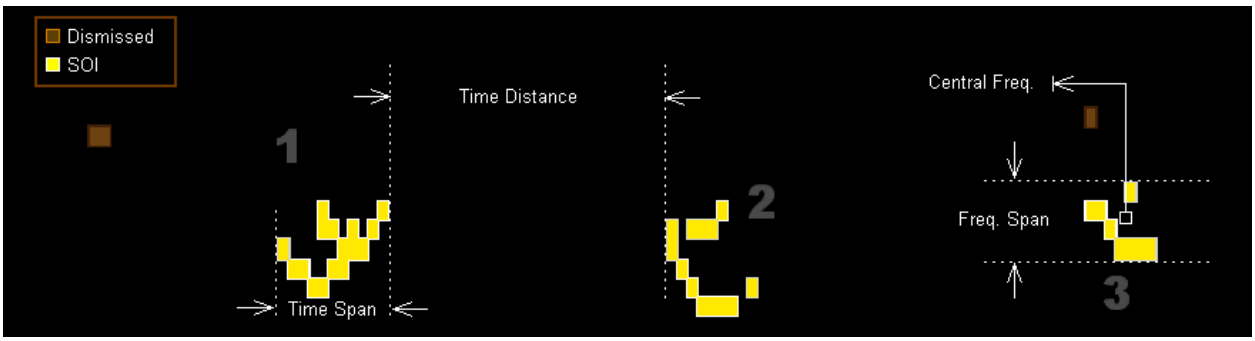


Figure 6. Diagram of feature selection. After screening out lesser SOIs, five features were extracted from each binary image: F1) the number of SOIs; F2) the average time distance between neighbouring SOIs; F3) the average time span (duration) of SOIs; F4) the average central frequency of SOIs, and F5) the average frequency span of SOIs. SOIs within 2 Hz frequency distances and 160 ms time distances were regrouped. Small isolated SOIs were dismissed.

2.7 Classification

Trials were randomly partitioned into a testing set and a training set. We used a fixed number of trials in the testing set, 150, for the 6-condition classification, and 300 for the 3-rhythm classification. All remaining trials were used for training; the exact number varies among subjects. We used a Bayesian classifier based on multiclass linear discriminant analysis (LDA), which projects linearly the feature vectors to positions which maximally separate different groups. This procedure can be described in terms of finding a weight matrix W which maximizes the following criteria:

$$L(w) = \frac{w^T C_B w}{w^T C_W w} \tag{19}$$

where C_B and C_W corresponds to the between-class and within-class covariance matrices, respectively. We examined the coefficients for the four different SOBI components as well as for five different time-frequency features in the weight matrix W to evaluate their contribution to the overall classification.

2.8 Classification using features based on temporal, spectral and time-frequency domain averages

We also analyzed the imagined speech data using conventional averaging algorithms with features obtained in three domains: temporal, spectral and joint time-frequency.

The first analysis is a straightforward multivariate linear classification based on EEG temporal waveforms. The raw EEG recordings were first band-pass filtered to remove line noise. They were then segmented by trials and conditions using the same procedures as described above. The filtered and segmented data were then decimated to 1/16th of the original size to avoid rank deficiency in covariance matrix calculations. The resulting data were partitioned randomly using the same criterion as in the SOBI-HHT method (408 trials for training, 300 trials for testing), and linear discriminant analysis (LDA) classification was performed separately for each EEG channel.

The second analysis is similar to the first one; the sole difference is that it is performed in the frequency domain rather than in the time domain. The raw EEG recordings were Fast-Fourier Transformed after having removed any linear tendency. The resulting complex-valued Fourier coefficients were partitioned and supplied to the LDA classifier in much the same way as described above.

For the third analysis, we performed a time-frequency transformation of EEG data using short-time Fourier transform methods (STFT). We computed spectrograms by convolving cosine- and sine-phase Gabor functions at 18 center frequencies with imagination-period EEG data. The center frequencies were spaced logarithmically in the interval 4 ~ 28 Hz. The Gaussian window standard deviations were decreased as center frequency increased to maintain an identical number of cycles within the Gaussian window. The spectrograms computed for the imagination periods had too many entries for covariance matrices to be inverted, a step required for LDA classification. We thus used a matched filter approach to classification. For each of the three rhythm classes we first computed class-average spectrograms, one per rhythm. The class-average spectrograms are not generally orthogonal to one another. One can derive orthogonal filters based on the class-average spectrograms by unraveling the three spectrograms to produce 3 vectors. The pseudoinverse of the matrix comprising these three vectors itself comprises three vectors (the matched filters) with desirable properties: the inner product of the matched filter for class i and the class-average spectrogram for class j is equal to 1 if $i = j$ and 0 otherwise.

As with the other classification methods, matched filters were computed for training trial data and applied to test trials to determine classification performance. The matched filter which provides the largest response to a given test trial provides this method's guess as to which class the trial belongs to. The class-average spectrograms and corresponding matched filters are computed separately for each electrode. Matched-filter responses from each electrode may be aggregated across the scalp to provide a single best guess by summing corresponding matched-filter responses across electrodes and determining which matched-filter class provides the largest response.

3. Results

In what follows we first report the results of using HHT-based method to classify imagined speech rhythm based on EEG signals. We then examine whether more traditional averaging-based methods work as well, and find that two of these (using features in the temporal and spectral domains, respectively) failed to produce significant classification, while the third one (using features in the time-frequency domain) produces results that can potentially rival the HHT method. Finally, we use the SOBI-HHT method to classify EEG data from the 3-second click cue period of each trial; the method works on rhythms specified by sensory input as well as imagined rhythms.

3.1 HHT-based decoding of imagined speech rhythm

Spot of interest features found for the four SOBI components help to classify imagined speech rhythm but fail to classify both rhythm and syllable. The latter result is shown by the performance of the six-condition classification, which seeks to identify the rhythm and syllable cued on a particular trial from EEG recordings of imagined speech. Table 1 summarizes these results. The average correct classification level across the seven subjects is 22.10%, which is not significant (p-value = 0.04) against

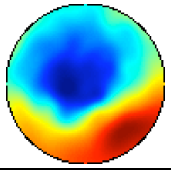
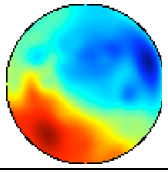
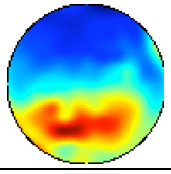
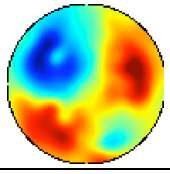
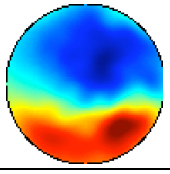
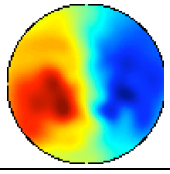
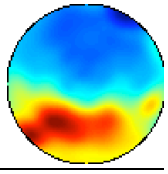
chance level (16.67%) at $\alpha = 0.001$, assuming a binomial mass function ($n = 150$, $p = 1/6$). While classification rates for two subjects (S4, S6) are significant at $\alpha = 0.01$, the corresponding classification performance levels (24.67% and 26.0%, respectively) are barely informative.

Results for the three-rhythm classification are shown in Table 2. The mean performance among all subjects, using all components of the SOI feature-SOBI data vector, is 58.05%. Assuming a binomial mass of $n = 300$, $p = 1/3$, then the significance level $\alpha = 0.001$ corresponds to 42% (126/300), so that the results of all seven subjects are significantly greater than chance. The best classification performance is found for S6, with a classification performance of 218/300 or 72.67%. Table 2 shows also, for each subject, which SOI feature and which SOBI component (shown topographically) contribute most to successful predictive classification. The SOI feature F3, the average time span of SOIs, is the predominant contributor to rhythm classification in five of seven subjects; the normalized classification weight for F3 ranges from 16% to 31% across subjects.

Table 1. Summary of the six-condition classification results for all seven subjects S1-S7. The available trials are partitioned into training and testing sets and the classification performance is the percentage of correctly classified testing trial.

	S1	S2	S3	S4	S5	S6	S7
# Training	201	238	201	201	201	204	195
# Testing	150	150	150	150	150	150	150
Accuracy	20.66%	22.67%	22.00%	24.67%	19.33%	26.00%	19.33%

Table 2. Summary of three-rhythm classification results for all seven subjects. The first three rows shows the partitioning of trials into training and testing sets and the classification performance. The last three rows show the time-frequency attribution, the normalized classification weights for the most significant data point in the feature vector as well as the topographic SOBI component distribution for each subject. Feature categories are listed in the caption to figure 6.

	S1	S2	S3	S4	S5	S6	S7
# Training	402	276	402	402	402	408	390
# Testing	300	300	300	300	300	300	300
Accuracy	54.00%	61.67%	55.00%	63.33%	48.33%	72.67%	51.33%
Category	F3	F4	F3	F2	F3	F3	F3
Weight	16.80%	18.07%	30.82%	16.88%	17.79%	20.16%	24.91%
Topography							

Another interesting observation is that the best SOI features, when viewed across subjects, are almost all features in the time domain. These features include F2, the average time distance between neighbouring SOIs, and F3, the average SOI duration. For only one subject is a frequency-domain SOI feature most informative; feature F4, the SOI central frequency, is most informative for S2. From the topographic distribution of the most highly weighted data point, we see that five out of seven contrast anterior and posterior regions of cortex. However, for the best performing subject, S6, the contrast is between left and right hemispheres. In general, the classification benefits from a feature vector containing more features and components; the most highly weighted feature usually carries less than 30% of total classification power. An example of the distribution of classification weights by SOBI components and time-frequency attributions is shown in Figure 7. This figure shows that the SOI duration (F3) and inter-SOI interval (F2) contribute most to the overall discriminability among imagined speech rhythms for best-performing subject S6. Specifically, the value of F3 gradually decreases from rhythm 1 to 3.

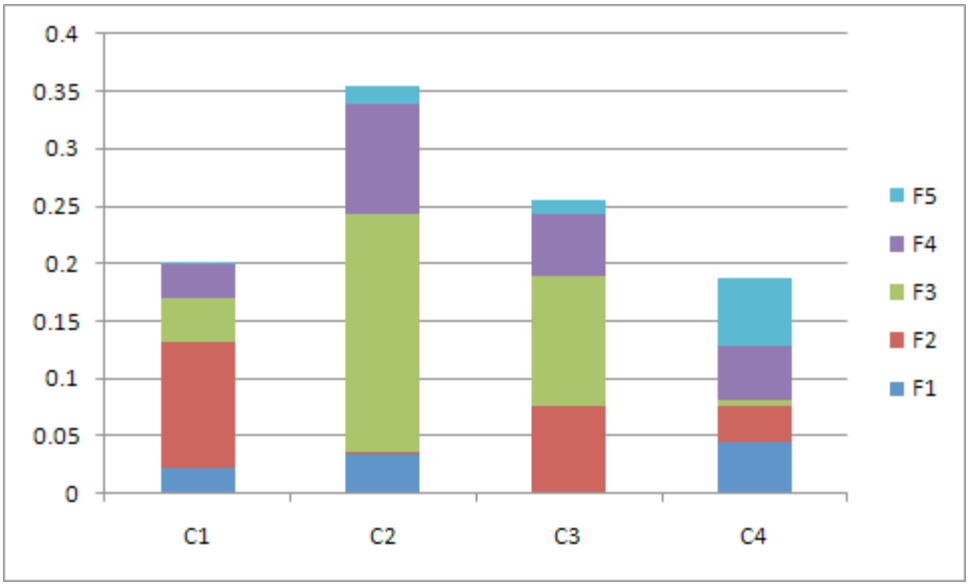


Figure 7. Linear classifier weight for subject 6. The height of each bar section reflects the weight of the corresponding SOBI component.

3.2 Averaging-based classification of imagined speech rhythm

As a comparison to the HHT-based classification results, we analyzed the imagined speech data obtained from our best performing subject S6 using conventional averaging algorithms in three domains: temporal, spectral and time-frequency.

The results of classification based on time-domain averages are summarized numerically in the first row of Table 3 ("Temporal LDA"); channel classification rates for S6 are shown topographically in the leftmost panel of Figure 8. The classification performance for this method is only 39% (*vs*. the chance-level performance of 33%).

The results of classification based on frequency-domain averages are shown in the second row of Table 3 ("Spectral LDA"); channel classification rates are shown in the middle panel of Figure 8. The resulting classification performance 40% is not much better than for the time domain.

The third analysis involved the application of matched filters based on class-average spectrograms. The results for subject S6 are shown in the third row ("Spectrogram matched filter") of Table 3. The resulting classification performance 67% is statistically highly significant and begins to rival the classification rate of 73% found using the SOBI-HHT algorithm.

Table 3. Averaging-based classification results.

	Accuracy	Significance
Temporal LDA	38.67%	0.0226
Spectral LDA	40.33%	0.0047
Spectrogram matched filter	66.50%	0.0000

3.3 HHT-based decoding of heard rhythm

Conceptually, this analysis may be conducted independently of the imagined rhythm classification by performing a separate SOBI procedure to generate spatial mixings specific to EEG recorded during the click cue periods. Yet such treatment can complicate the problem and make it difficult to compare results from click cue and imagination periods. We thus used the same set of SOBI mixings for the click cue periods as in the imagination periods. While this choice may lead to sub-optimal classification results, the results are likely similar to those obtained from independent analyses of the two period types, because the most informative SOBI spatial mixings tend to contrast large-scale patterns. The HHT and supra-thresholding procedures were performed independently for click cue and imagination periods, while the threshold criteria and partitioning of trials for the click cue periods were kept identical to those for the imagination periods.

The results are summarized in Table 4. All seven subjects show higher classification rates for the click cue period data than for the imagination data (see also Table 2). The linear classifier weights of S6 are shown in Figure 8. The click cue period classifier still uses more information in the time domain than in the frequency domain. Yet compared to results for the imagination periods (Fig. 7), one finds that the relative importance of temporal features differs; the most informative feature is the average time distance (F2) followed closely by the number of SOIs (F1).

Table 4. Classification result based on click cue period

	Accuracy	Category	Weight
S1	68.00%	F1	18.57%
S2	71.67%	F1	19.45%
S3	59.67%	F3	23.35%
S4	66.33%	F2	34.83%
S5	60.67%	F3	27.44%
S6	73.67%	F2	22.05%
S7	54.67%	F2	30.14%

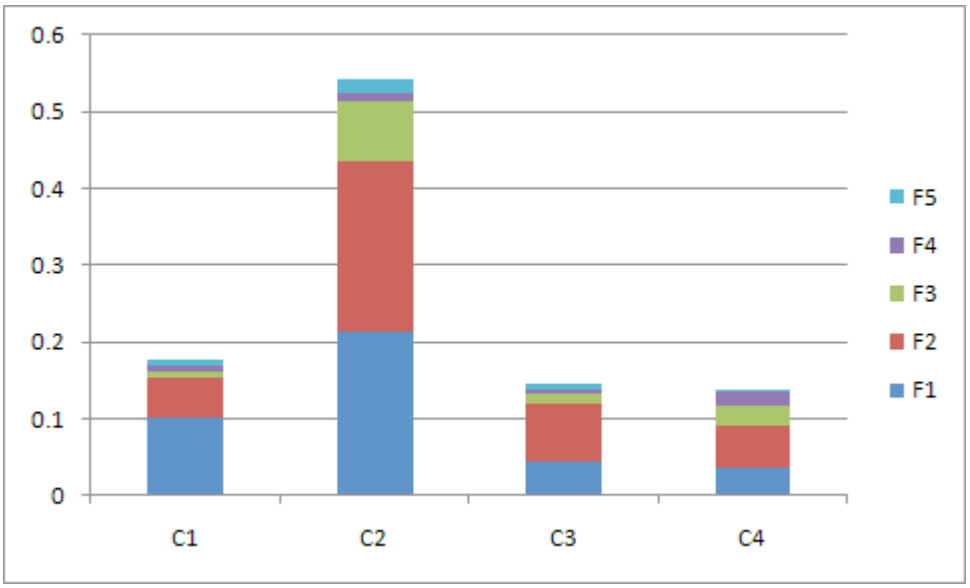


Figure 8. Click cue period linear classifier weights for S6

4. Discussion

The problem of recognizing imagined speech via brain signals is similar to the problem of recognizing speech automatically at a cocktail party, except that the party is now more clamorous and our recording equipment is set behind a thick wall of bone and tissues. In situations like this, a human listener will typically try his or her best to focus on the vocal source by tracking its acoustic properties, like pitch and volume fluctuations, and worry about meaning only after successfully segregating the source stream from the noise and performing an initial segmentation. For much the same reason, we consider the

classification of imagined speech rhythm an indispensable step toward EEG-based recognition of imagined speech.

4.1 Hilbert Spectra captures more information than conventional time-frequency methods

The fact that brain signals are generally not stationary calls for analytic methods that use representations which are richer than those available in time or frequency domains alone and account for the increasingly wide application of joint time-frequency analysis techniques (Vialatte *et al.* 2007). The Hilbert-Huang Transform is an emerging time-frequency decomposition method that offers several advantages over short-time Fourier transform and wavelet decompositions. First, it does not project data onto a constant basis; rather, it adjusts the frequency band of interest adaptively based on the signal envelopes. Second, the time and frequency resolutions of HHT are also adaptive, so that this method potentially provides far more accurate estimates over regions of interest (Huang and Shen 2005). These properties are particularly useful in our problem, because discriminating among different rhythmic conditions depends on distinguishing changes in power and frequency at intervals substantially shorter than one sec.

The results of traditional averaging-based methods show that classification benefits from the use of both temporal and spectral information; the two classifications based on content in just a single domain are not significant at $\alpha = 0.001$ level. In contrast, the performance of the spectrogram matched filter method, although not as high as that with the SOBI-HHT algorithm, is very promising. One possible reason for the performance difference between the spectrogram matched filter method and the SOBI-HHT method is that the matched filters are based on class-average spectrograms and, so, do not take into account possible phase shifts in the underlying signals. Class-average spectograms are blurred by such phase shifts. This is not true for the SOBI-HHT; its suprathresholding makes it robust in the face of such shifts. There could also be an effect of difference in resolution between STFT and HHT. Finally, the SOBI step in our method congregates more informative channels to outperform any single one.

The computational complexity of HHT is orders of magnitude greater than that of classical spectral analysis methods such as FFT. This remains a major hindrance to the implementation of HHT in real-time analysis. With the development of faster computers and better algorithms, we expect that this problem will diminish. Yet we investigated also a spectrogram matched filter method, which is based on class-average spectrograms. This method classified imagined speech rhythm in these experiments well, if not as well as the HHT method. Its major advantage is a far lesser computational complexity. We expect that HHT method classification performance can be improved. For example, the thresholds that were used to binarize the Hilbert spectra were determined *ad hoc*. We expect better classification performance if these parameters are determined quantitatively for each subject using a larger set of training data.

4.2 SOBI provides automatic spatial mixing and noise reduction.

Due to the superficial positions of EEG electrodes on the scalp, the true signals of neuronal origin are almost always contaminated by noise of various sorts. Blind-source separation techniques have proven very effective at removing statistically uncorrelated noise and at boosting SNR to facilitate further analyses (Tang *et al* 2005, Cichocki and Thawonmas 2000). In this study, the SOBI procedure is crucial to the success of the further joint time-frequency decomposition because it acts not only to reduce

artefacts but also to automatically generate specific spatial patterns which help contrast different cortical regions (Dornhege *et al* 2006). We use SOBI instead of other PCA and ICA algorithms because SOBI takes account of the time-variant nature of our signal. However, for the purpose of reducing artefact, ICA is likely also sufficient as a preprocessing step before joint time-frequency decomposition of the data. We extended the classical SOBI algorithm to process multi-channel multi-trial recordings, and although the current analysis is completely offline, we expect that the SOBI filters generated from training sessions can be readily used during online classifications as well.

The SOBI algorithm provides an automatic scheme to mix all channels in a way that maximizes cross-lag inter-component separation (Delorme *et al.* 2007). In most cases, the performance of single channels will fall behind that of the SOBI components. Yet, just as is found with most data rotation methods, the automatically-generated spatial patterns may vary across subjects, which can make them difficult to interpret. The most significant of the automatically-generated weight maps usually vary slowly across space, as is shown by the topographies in Figure 2 and in Table 2. For example, the most significant component for four of our subjects, S1, S3, S5, and S7, contrasts the occipital and lower temporal-parietal regions with mid-frontal regions. Contrast between left and right temporal-parietal channels can also produce valuable results, as shown in S6. Conceivably, this specific pattern (component 2 of Figure 2) contrasts responses of auditory origins, and is consistent with ERP and BCI literatures (Osman *et al.* 2006; Kruif *et al.* 2007). In addition, although the distributions of SOBI component vary across subjects, our experience is that on the same subject the pattern is generally consistent across trials within the same recording session. For instance, we compared the patterns contrasting the occipital and frontal regions for different trials of the same subject, and the correlations are in the range of 0.8 ~ 0.9. Such consistency ensures that for the same subject the SOBI analysis only needs to be performed once and the resulting loading patterns may potentially be used in future recordings or BCI applications.

4.3 A step toward EEG-based classification of imagined speech.

Our results suggest that information extracted from time-frequency representations of EEG data can be used to predict imagined speech rhythm. The temporal information is more informative; in six of seven subjects, we found that the most significant type of feature is in the temporal domain. This finding confirms one's intuition that brain activity during imagined speech should bear some similarity to the temporal structure of the speech itself. Among temporal features, the average time span of SOIs proved to be the most informative one. For the subject providing the greatest accuracy in predictive classification, increasing the rate at which syllables are produced is correlated with decreasing SOI duration. This result suggests that certain features of EEG signals recorded during imagined speech production carry information concerning the temporal structure of the underlying neuronal dynamics, and that this structure is highly consistent with that of its overt, vocalized, counterpart. Given enough temporal resolution and signal gain, we expect to be able eventually to use this isomorphism to recognize more complex rhythms in a real-time fashion and to provide concrete steps toward the development of EEG-based communication of imagined speech.

Previous research shows that sensory responses to heard speech may be decoded from both MEG (Numminen *et al.* 1999; Houde *et al.* 2002; Luo & Poeppel 2007) and EEG (Deng & Srinivasan 2010). We used the SOBI-HHT method to classify EEG generated during the click trains which were used at the

beginning of each trial to cue the rhythm with which imagined speech was then generated. EEG responses to these heard clicks could be classified using the HHT method with greater accuracy than EEG responses produced during imagined speech. The features which proved most informative for classifying click train rhythm differed from those most informative for classifying imagined speech (of the same rhythms). The change in feature weight distribution is likely to be closely related to the phase locking of responses evoked by rhythmic sensory inputs (Snyder and Large 2005; Zanto *et al.* 2006). During imagination periods, such evoked responses are absent. Instead, the endogenously driven modulation of perception induces less precisely time-locked responses which are characterized better by their overall time spans.

The classification of implicit thoughts has a self-contradictory nature, since objective verification that subjects are performing the desired tasks in some manner independent of the brain-imaging data is not forthcoming. Although we obtained 72.67% classification performance of imagined speech rhythm for our best subject, there is some extent to which we do not know whether this number represents the classification of genuine EEG due to imagination or represents classification based on interference factors like artefacts or sustained evoked responses from cueing. Nevertheless, researchers should not be daunted by these technical or methodological issues, as they may only be answered by further exploration in this field.

5. Acknowledgements

We thank David Poeppel for proposing that such an experiment be performed and Samuel Thorpe for his help with the experiment. This work was supported by ARO 54228-LS-MUR.

6. References

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H and Merzenich M M 2001 Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Nat. Acad. Sci. USA* **98**(23), 13367--13372.

Belouchrani A, Abed-Meraim K, Cardoso J F and Moulines E 1997 A blind source separation technique using second-order statistics. *IEEE J. SP* **45**(2), 434--444.

Birbaumer N, Kübler A, Ghanayim N, Hinterberger T, Perelmouter J, Kaiser J, Iversen I, Kotchoubey B, Neumann N and Flor H 2000 The thought translation device (TTD) for completely paralyzed patients. *IEEE Trans. Rehabil. Eng.* **8**(2), 190--193.

Blankertz B, Müller K R, Curio G, Vaughan T M, Schalk G, Wolpaw J R, Schlögl A, Neuper C, Pfurtscheller G, Hinterberger T, Schröder M and Birbaumer N 2004 The BCI Competition 2003: progress and perspectives in detection and discrimination of EEG single trials. *IEEE Trans. Biomed. Eng.* **51**(6), 1044--1051.

Bokde A L, Tagamets M A, Friedman R B and Horwitz B 2001 Functional interactions of the inferior frontal cortex during the processing of words and word-like stimuli. *Neuron* **30**(2), 609--617.

Cardoso J F 1998 Blind signal separation: statistical principles. *IEEE J. Proc.* **86**(10), 2009--2025.

Cichocki A and Thawonmas R 2000 On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics. *Neural processing letters*, 12(1), 91-.

Craelius W 2002 The bionic man: restoring mobility. *Science* **295**(5557), 1018--1021.

Dewan E M 1967 Occipital alpha rhythm eye position and lens accommodation. *Nature* **214**(5092), 975--977.

Deng S and Srinivasan R 2010 Semantic and acoustic analysis of speech by functional networks with distinct time scales. Manuscript submitted for publication.

Delorme A, Sejnowski T and Makeig S 2007 Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis; *Neuro Image*, 34(4), 1443.

Donchin E, Spencer K M and Wijesinghe R 2000 The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Trans. Rehabil. Eng.* **8**(2), 174--179.

Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G and Müller K R 2006 Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Trans. Biomed. Eng.* **53**(11), 2274--2281.

D'Zmura M, Deng S, Lappas T, Thorpe S and Srinivasan R 2009a Toward EEG sensing of imagined speech. Jacko J A (Ed.) *Human-Computer Iteraction, Part I, HCII 2009, LNCS 5610*. Berlin: Springer, 40--48.

D'Zmura M, Deng S, Lappas T, Thorpe S and Srinivasan R 2009b EEG classification of imagined syllable rhythm using Hilbert spectrum methods. Chicago, IL: Society for Neuroscience, 2009.

Haynes J D and Rees G 2005 Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* **8**(5), 686--691.

Hickok G and Poeppel D 2007 The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**(5), 393--402.

Houde J F, Nagarajan S S, Sekihara K and Merzenich M M 2002 Modulation of the auditory cortex during speech: An MEG study, *Journal of Cognitive Neuroscience* **14**, 1125--1138.

Huang N E and Shen S S P 2005 *Hilbert-Huang transform and its applications*. London : World Scientific.

Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C and Liu H H 1998 The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proc. Math. Phys. & Eng. Sci*, 1971(454) 903--995

Ince N F, Arica S and Tewfik A 2006 Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time-frequency tilings. J. Neural Eng. 3(3), 235--244.

Indefrey P and Levelt W J M 2004 The spatial and temporal signatures of word production components. Cognition 92, 101--144.

Kruif B J, Schaefer R and Desain P 2007 Classification of imagined beats for use in a Brain Computer Interface; 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 678.

Leeb R, Lee F, Keinrath C, Scherer R, Bischof H and Pfurtscheller G 2007 Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment. IEEE Trans. Neural Systems and *Rehabilitation Engineering* **15**, 473--482.

Luo H and Poeppel D 2007 Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* **54**(6), 1001--1010.

Nunez P L and Srinivasan R 2006 *Electric fields of the brain: the neurophysics of EEG.* Oxford University Press.

Numminen, J, Salmelin, R and Hari, R 1999 Subject's own speach reduces reactivity of the human auditory cortex. *Neuroscience Letters* 265, 119--122.

Osman A, Albert R and Ridderinkof K R 2006 The beat goes on: Rhythmic modulation of cortical potentials by imagined tapping; *Journal of Experimental Psychology: Human Perception and Performance* **32**(4), 986--1005.

Pfurtscheller G, Leeb R, Keinrath C, Friedman D, Neuper C, Guger C and Slater M 2006 Walking from thought. *Brain Research* 2006, 145--152.

Schalk G, McFarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* **51**(6), 1034--1043.

Snyder J S and Large E W 2005 Gamma band activity reflects the metric structure of rhythmic tone sequences; *Cognitive brain research* **24**(1), 117--126.

Tang A C, Liu J Y and Sutherland M T 2005 Recovery of correlated neuronal sources from EEG: the good and bad ways of using SOBI. *Neuroimage* **28**(2), 507--519.

Tang A C, Sutherland M T and McKinney C J 2005 Validation of SOBI components from high-density EEG. *Neuroimage* **25**(2), 539--553.

Vialatte F B, Martin C, Dubois R, Haddad J, Quenet B, Gervais R and Dreyfus G 2007 A machine learning approach to the analysis of time-frequency maps, and its application to neural dynamics. *Neural Network* **20**(2), 194--209.

Vigneau M, Beaucousin V, Hervé P Y, Duffau H, Crivello F, Houdé O, Mazoyer B and Tzourio-Mazoyer N 2006 Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* **30**(4), 1414--1432.

Wang Y and Makeig S 2009 Predicting intended movement direction using EEG from human posterior parietal cortex. Schmorrow, D.D. *et al*. (Eds): *Augmented Cognition*, *HCII 2009, LNAI 56380.* Berlin: Springer, 437--446.

Weiss S and Mueller H M 2003 The contribution of EEG coherence to the investigation of language. *Brain Lang.* **85**(2), 325--343.

Weineke G H, Deinema C H A, Spoelstra P, Storm van Leeuwen W and Versteeg H, 1980 Normative spectral data on alpha rhythm in males adults, *Electroencephalogr. Clin. Neurophysiol*. **49**, 636--645.

Wolpaw J R, McFarland D J and Vaughan T M 2000 Brain-computer interface research at the Wadsworth Center. *IEEE Trans. Rehabil. Eng.* **8**(2), 222--226.

Willson J A and Williams J C 2009 Massively Parallel Signal Processing using the Graphics Processing Unit for Real-Time Brain–Computer Interface Feature Extraction. *Frontiers in Neuroengineering* Volume 2.

Zanto T P, Syder J S and Edward W L 2006 Neural correlates of rhythmic expectancy; *Advances in cognitive psychology*, **2**(2), 221--231.

Zhao Q, Zhang L and Cichocki A 2009 EEG-based asynchronous BCI control of a car in 3D virtual reality environments. *Chinese Science Bulleting* **54**, 78--87.