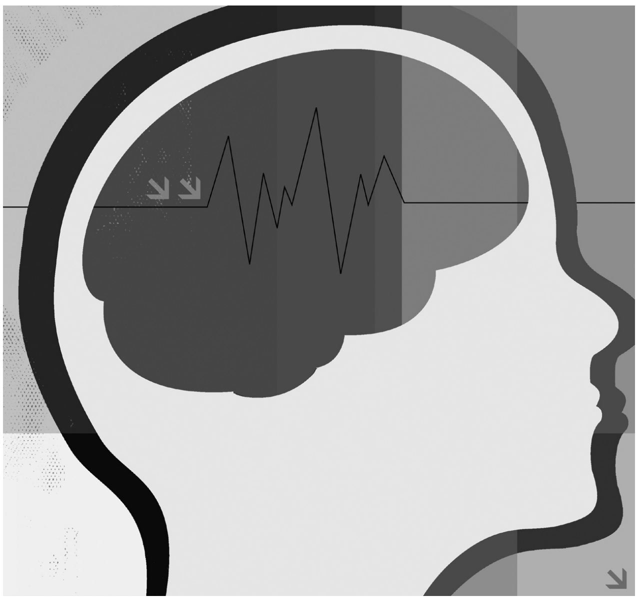
Medical Applications of EEG Wave Classification

Wanli Min and Gang Luo



Did you know your brain continuously emits electric waves, even while you sleep? Based on a sample of wave measurements, physicians specializing in sleep medicine can use statistical tools to classify your sleep pattern as normal or problematic.

Brain-computer interfaces (BCIs) now being developed can classify a disabled person's thinking based on wave measurements and automatically execute necessary instructions. This type of research is exciting, but conducting it requires knowledge of medicine, biology, statistics, physics, and computer science.

Electroencephalogram (EEG) is the recording of electrical activity through electrode sensors placed on the scalp. The electricity is recorded as waves that can be classified as normal or abnormal. Measuring EEG signals is not an intrusive procedure; it causes no pain and has been used routinely for several decades.

Different types of normal waves can indicate various states or activity levels. Abnormal waves can indicate medical problems. Two important applications of EEG wave classification are diagnosis of sleep disorders and construction of BCIs to assist disabled people with daily living tasks.

Medical Background

Sleep, which takes up roughly one-third of a person's life, is indispensable for health and well-being. Nevertheless, one-third of Americans suffer from a sleep problem. For example, one in five American adults has some degree of sleep apnea, which is a disorder characterized by pauses of 10 seconds or longer in breathing during sleep. A person with sleep apnea cannot self-diagnose the disorder. To make diagnoses for sleep disorders, physicians usually need to record patients' sleep patterns.

A typical sleep recording has multiple channels of EEG waves coming from the electrodes placed on the subject's head. Sample sleep recordings are shown in Figure 1. In the left panel, the waves from a healthy subject are stable at about zero and show relatively high variability and low correlation. In the right panel, the waves from a person with sleep difficulty show less variability and higher correlation. Sleep staging is the pattern recognition task of classifying the recordings into sleep stages continuously over time; the task is performed by a sleep stager. The stages include rapid-eye movement (REM) sleep, four levels of non-REM sleep, and being awake.

Sleep staging is crucial for the diagnosis and treatment of sleep disorders. It also relates closely to the study of brain function. In an intensive care unit, for example, EEG wave classification is used to continuously monitor patients' brain activities. For newborn infants at risk of developmental disabilities, sleep staging is used to assess brain maturation. Many other applications adapt the EEG wave classification techniques originally developed for sleep staging to their purposes. Besides being used to study human activities, sleep staging also has been used to study avian bird song systems and evolutionary theories about mammalian sleep.

To make many EEG-based applications practical enough for routine use, it is necessary for the wave classification to be accurate. The more accurately sleep stages are classified, the faster patterns can be recognized. Because different sleep disorders have different sleep stage patterns, more accurate sleep stage classification allows physicians specializing in such disorders to diagnose problems better and faster. In fact, such specialists often spend several years in residency programs for special training in recognizing sleep patterns before obtaining board certification in sleep medicine. Expediting sleep disorder diagnoses also can help reduce the costs, which have surged in recent years, of treating sleep problems.

Statistical Analysis of Sleep EEG Data

Among popular statistical methods for performing sleep staging are autoregression, Kullback-Leibler divergence-based

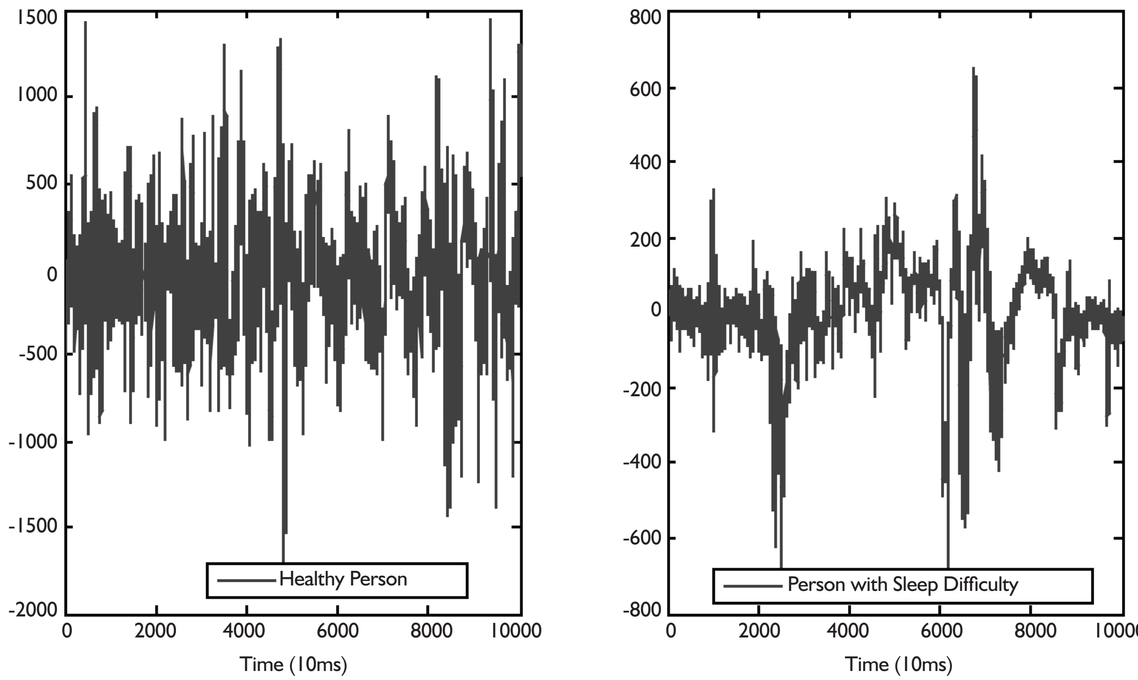


Figure 1. EEG signals from a healthy person and a person with sleep difficulty

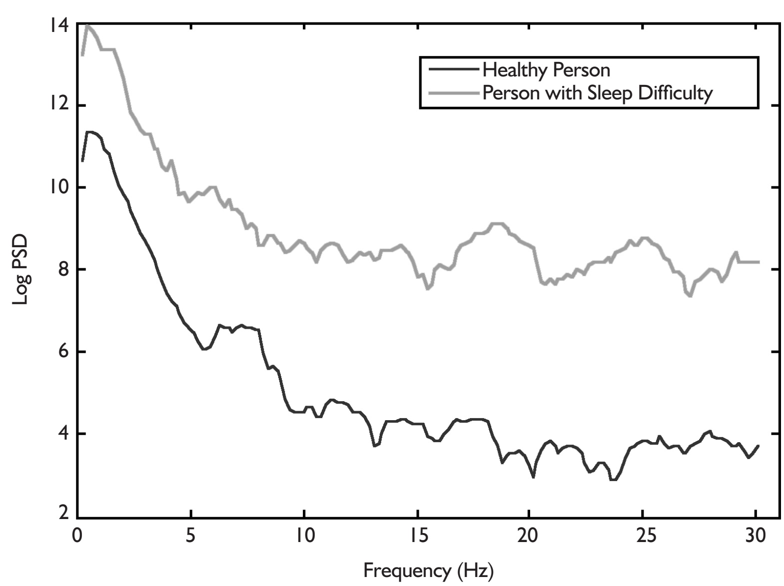


Figure 2. Logarithmic power spectral density of EEG signals from a healthy person and a person with sleep difficulty

nearest-neighbor classification, and statistical modeling using a hidden Markov model (HMM). These methods typically consist of two steps: signal processing, which is extracting useful feature variables from EEG signals at each time-stamp of the EEG recording, and statistical classification based on the extracted features.

Although some researchers have used nonlinear characteristics (e.g., fractal dimensionality), the prevailing technique for EEG signal processing is spectral analysis. Spectral analysis, which converts the original time series to the frequency domain, is a natural choice for EEG signal processing because EEG signals are often described by α, β, θ, and δ waves, whose

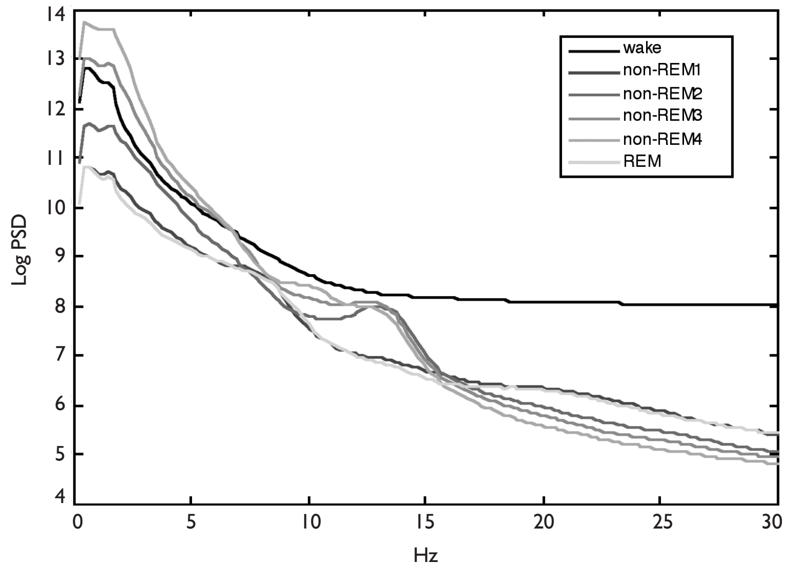


Figure 3. Stage-specific average logarithmic power spectral density of EEG signals from a healthy person

frequency ranges are 8–12 Hz, 12–30 Hz, 4–7 Hz, and 0–3 Hz, respectively.

As shown in figures 2 and 3, the frequency content of EEG signals is characterized by power spectral densities (PSDs). In Figure 2, the log PSD plot is substantially higher, especially at higher frequencies, for the subject with sleep difficulty. Figure 3 shows that one major difference between sleep stages in a healthy subject occurs at high frequencies, where the curve representing the wake stage is much higher than those representing the other stages.

During signal processing, a spectral density estimator is typically applied to each epoch—a time window of fixed length—of the raw EEG data. To reduce variance, adjacent epochs have overlapping segments. The size of the epoch depends largely on the sampling rate (number of measurements per second) of the EEG signal. On the one hand, each epoch needs to contain enough raw signals for any sampling-based spectral density estimation method to work well. On the other, the time window for each epoch cannot be too wide or classification will become difficult, as information from later sleep stages is mixed with information from the current stage in the extracted feature variables. In testing on a subject, for example, our sleep stager can achieve a classification accuracy of 80% when three-second epochs are used, but accuracy drops to 49% when 10-second epochs are used. Thus, the tradeoff needs to be evaluated statistically.

PSD is usually estimated through a periodogram using the fast Fourier transform. It is known that the raw periodogram estimator is biased for a process with continuous spectrum. To address this problem, we use kernel-type methods, including Parzen window, Bartlett window, and multi-taper.

In general, a sleep stager needs enough training data to achieve good classification accuracy. Training data includes both EEG signals and the corresponding sleep stages, which require time-consuming manual labeling. To obtain sufficient training data on a subject, a specially trained technician needs to spend several days, or weeks, on the labeling process. As a result, there is usually sufficient training data, D_{old}, on several old subjects, s_{old}, but limited training data, D_{new}, on a new subject, s_{new}.

For the sleep stager, extracting feature variables and classifying stages are equally important. If either is not done well, stager performance will deteriorate. Before performing EEG wave classification, it is helpful to quickly assess the quality of extracted feature variables. This assessment does not require correlation of feature variables extracted at different time stamps. Instead, it can treat these feature variable vectors $\overrightarrow{X_t}$ as independent because the major patterns of sleep stages are described by frequency components without referring to time correlation. For this purpose, it would be sufficient to perform a straightforward discriminant analysis by redefining sleep stages to a simple structure.

For instance, with one non-REM and one REM sleep stage, a reasonable classification accuracy such as 70% would suggest that the extracted feature variables have good quality. If the extracted feature variables cannot be classified reasonably accurately in the discriminant analysis, we should not continue to pursue the corresponding feature variable extraction method.

Another issue that needs to be considered is the number of frequency bands used (i.e., the dimensionality of feature variable vectors $\overrightarrow{X_t}$). If the amount of training data is extremely large, using more frequency bands usually leads to better discriminating power. However, the amount of training data is often limited, so this property no longer holds. Instead, it is desirable to confine $\overrightarrow{X_t}$ to cover the subset of all the frequency bands that have the most significant discriminating powers, as bands with low discriminating powers can interfere with parameter estimation of the statistical model.

Once feature variables are extracted, there are many approaches to classifying EEG waves. Among them, the HMM and its variants are widely used. HMM-style methods

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z	0	1
2	3	4	5	6	7	8
9	⊔	(	)	!	@	#
$	%	^	&	*	,	.

Figure 4. Matrix used in the P300 brain-computer interface

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	**W**	**X**	**Y**	**Z**	**0**	**1**
2	3	4	5	6	7	8
9	⊔	(	)	!	@	#
$	%	^	&	*	,	.

(a)

A	**B**	C	D	E	F	G
H	**I**	J	K	L	M	N
O	**P**	Q	R	S	T	U
V	**W**	X	Y	Z	0	1
2	**3**	4	5	6	7	8
9	**⊔**	(	)	!	@	#
$	**%**	^	&	*	,	.

(b)

Figure 5. Seven characters are intensified simultaneously. One row of characters is intensified in (a), and one column of characters is intensified in (b).

take into account an important aspect of sleep staging: the serial correlation of sleep stages across time. In contrast, discriminant analysis treats feature variable vectors $\overrightarrow{X_t}$ with different time stamps as completely independent. Although discriminant analysis captures the main effect, it misses the obviously important secondary effect of correlation across time so cannot achieve satisfactory classification accuracy. For example, it is unlikely that a transition from deep sleep (the third or fourth level of non-REM sleep) to REM sleep occurs at two consecutive time stamps.

HMM-style methods consider the serial correlation of sleep stages across time and model it through the transition probabilities of the hidden Markov chain. Nevertheless, they make unnecessary assumptions about the distribution of the feature variable vectors $\overrightarrow{X_t}$. (Typically, they assume $\overrightarrow{X_t}$ follows a Gaussian distribution, which can be far from true.)

Unlike HMM, a linear-chain conditional random field (CRF) directly models the probabilities of possible sleep stage sequence given an observed sequence of feature variable vectors, without making unnecessary independence assumptions about the observed vectors. Consequently, CRF overcomes the shortcoming of HMM, that it cannot represent multiple interacting features or long-range dependencies among the vectors observed. According to our experiments, the linear-chain CRF method performs much better than the HMM method for human sleep staging, improving average sleep-stager classification accuracy by about 8%. Similar results hold for bird sleep staging.

Brain-Computer Interfaces

The BCIs now being developed will facilitate the control of computers by people who are disabled. As they think about what they want the computer to do, their thinking will be classified based on their EEG waves and the computer will automatically execute the corresponding instructions. Accurate EEG wave classification is critical for computers to issue the correct instructions.

Among the various types of BCIs, the P300 BCI using EEG signals is one of the most promising because it is noninvasive,

1	2	3	4	5	6	7
3	5	1	6	7	2	4
2	6	4	3	1	7	5
6	1	7	2	4	5	3
5	4	6	7	3	1	2
4	7	5	1	2	3	6
7	3	2	5	6	4	1

Figure 6. A Latin square of order seven

A	B	C	D	E	F	G
H	I	**J**	K	L	M	N
O	P	Q	R	**S**	T	U
V	**W**	X	Y	Z	0	1
2	3	4	5	6	**7**	8
9	⊔	(	**)**	!	@	#
$	%	^	&	*	,	**.**

Figure 7. Seven characters are intensified simultaneously according to symbol ones (1s) in the Latin square in Figure 6.

easy to use, and portable. Additionally, the set-up cost is low. P300 refers to a neurally evoked potential component of EEG. The current P300 BCI communicates one symbol at a time and works as shown in figures 4 and 5. A matrix of characters or pictures is displayed on the computer screen. To communicate a desired character, the user focuses on the matrix cell containing it and counts the number of times it is intensified when a predetermined number of intensification rounds are performed. In each round, all the rows and columns of the matrix are intensified once in a random order—one row or column at a time. The row and column containing the desired character form the rare set (the target), and the other rows and columns form the frequent set (the nontargets). If the user is attending to the desired character, intensification of the target row or column should elicit a P300 response because it is a rare event in the context of all the other row or column intensifications. By detecting the P300 responses from the recorded EEG signals of the user, we can classify the target row and column whose intersection cell contains the classified character the user intends to communicate.

Experimental design is the term describing how characters are arranged and intensified. To maximize both the classification accuracy and communication speed of the P300 BCI system, an appropriate experimental design is needed to obtain strong P300 responses. Nevertheless, the existing experimental design is nonoptimal due to an undesirable effect caused by neighboring characters. Amyotrophic lateral sclerosis (ALS) patients, one of the most important user groups of BCI, have eye movement problems. When neighboring characters in a row or a column are intensified simultaneously, an ALS patient's attention can be distracted from the desired character, weakening the P300 response and reducing classification accuracy. To minimize this interference, it is better to intensify non-neighboring characters simultaneously. The larger the distances between simultaneously intensified characters, the less interference the ALS patient will receive.

One approach to intensifying non-neighboring characters simultaneously is to use the mathematical structure of the Latin square. A Latin square of order n is an $n \times n$ matrix based on a set of n symbols, so that each row and column contains each symbol exactly once. Without loss of generality, the symbols are assumed to be 1, 2, …, and n. Figure 6 shows an example of a Latin square of order seven.

If we intensify characters according to a Latin square (Figure 7), the simultaneously intensified characters will not be direct neighbors either horizontally or vertically. To ensure

EEG Classification

Mark L. Scheuer, Chief Medical Officer, Persyst Development Corporation

With respect to sleep, an improvement in diagnostic classification would result in some improvement in diagnostic accuracy, but improvement in this particular area would be uncertain and warrant study in actual use. The area in which it would help is in reducing the time physicians and high-level medical technologists require to establish a clear classification of the problem based on the EEG/polysomnography data. This would improve efficiency in a costly endeavor.

There are other areas in which improved classification would be useful, among them intensive care unit monitoring of brain activity and a host of other applications in which EEG signals provide a wealth of physiological information about a person (awake or sleep, sleeping well or not, brain functioning properly or not, a potential stroke in progress, a seizure occurring, a person with diabetes having lapsed into an obtunded state due to blood sugar problems, etc.). The EEG provides a direct window on brain function, one with high time resolution, great versatility, and reasonable spatial resolution.

Probably the main problem with EEG classification and interpretation is that the signal is complex and noisy. The noise, itself, is complex and easily confused with the actual cerebral signal. Humans who can read EEG well have generally spent many months to years learning how to do so.

the desired character can be uniquely determined within each round of intensification, we can resort to the concept of orthogonal Latin squares. Intuitively, Latin squares L_1 and L_2 of the same order n are orthogonal if the cells in L_1 containing the same symbol can be regarded as a conceptual row, the cells in L_2 containing the same symbol can be regarded as a conceptual column, and each conceptual row and column has a unique intersection cell.

For an $n \times n$ character matrix M, the following new experimental design can ensure unique character determination by mapping M to the superposition of L_1 on L_2. Whenever the experimental design intensifies the hth ($1 \le h \le n$) row of M, the new experimental design intensifies the characters in M corresponding to the hth conceptual row in L_1. Whenever the experimental design intensifies the kth ($1 \le k \le n$) column of M, the new experimental design intensifies the characters in M corresponding to the kth conceptual column in L_2. By detecting the P300 responses from the recorded EEG signals of the user, we can classify the target conceptual row and column whose unique intersection cell contains the classified character the user intends to communicate. If we expand nonsquare matrices into square matrices by adding dummy rows or columns, this method also works for nonsquare character matrices.

In general, given a positive integer n, we can obtain many pairs of orthogonal Latin squares of order n. The pair of orthogonal Latin squares used to communicate a character can vary from one character to another through random selection. This provides much flexibility and makes the character intensification pattern more unexpected by the user. As mentioned by Eric W. Sellers and colleagues in their *Biological Psychology* article, "A P300 Event-Related Potential Brain-Computer Interface (BCI): The Effects of Matrix Size and Interstimulus Interval on Performance," such unexpectedness can lead to stronger P300 responses and improve classification accuracy.

When choosing Latin squares, we can impose various distance constraints. One is that, in the Latin square, the distance between any pair of cells containing the same symbol is no smaller than a predetermined threshold t. Using that constraint, we can ensure that, at any time, the distance between any two simultaneously intensified characters is no smaller than t, which can reduce interference for ALS patients, lead to stronger P300 responses, and improve classification accuracy.

We emphasize that the orthogonality of Latin squares is a desired, but not mandatory, property. Moreover, the pair of Latin squares used can vary more frequently (e.g., from one round of intensification to another), even within the communication process of the same character, because multiple rounds of intensification are performed to communicate a character.

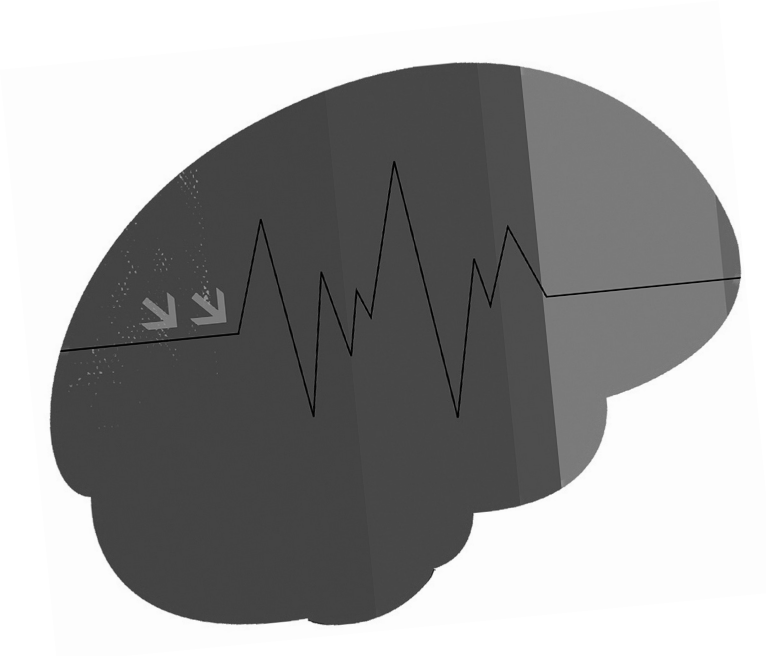
Within each round of intensification, a separate score reflecting the likelihood of being the desired character can be computed for each character in the matrix according to some classification algorithm. The final classification is performed by combining the scores in all rounds (i.e., using an aggregation or voting schema). If the two Latin squares used are nonorthogonal, we may not be able to uniquely determine the desired character in a single round of intensification. Nevertheless, since the pair used varies from one round of intensification to another, the combination of the scores of all the rounds can uniquely determine the desired character if all the Latin squares are chosen appropriately.

Future Directions for EEG Wave Classification Research

EEG wave classification research appears to be going in the following three promising directions:

1. *Subject adaptation for sleep staging*. When there is extensive training data, D_{old}, on several old subjects, s_{old}, but limited training data, D_{new}, on a new subject, s_{new}, it is not desirable to train the parameter vector Θ of the classifier using only D_{new}. Instead, subject adaptation needs to be performed to improve classification accuracy on s_{new}. The high-level idea of subject adaptation is to use the knowledge on Θ that is learned from D_{old} to obtain a prior distribution of Θ. Using D_{new} and Bayes' theorem, a posterior distribution of Θ can be computed to obtain a regulated estimate of Θ. In this way, even without any D_{new}, classification accuracy on s_{new} will be relatively acceptable. Moreover, accuracy increases with increases in D_{new}.

Because subject adaptation for sleep staging was proposed only recently, a few issues remain open. To ensure robustness, subject adaptation requires training data from many subjects with a wide variety of characteristics. This requires building a large, integrated, publicly available EEG sleep recording



database. Currently, such databases are owned by individual institutions and will not be released to the general public for at least several more years. One contains data from 6,400 subjects in the Sleep Heart Health Study. Moreover, different subjects (e.g., newborn babies vs. older people, healthy people vs. people with mental disorders) have different characteristics. When classifying sleep stages for a s_{new}, it is undesirable to train the sleep stager using data from subjects whose characteristics are dramatically different. However, training the stager using data from subjects that have similar characteristics requires categorizing all the subjects in the database into multiple clusters and having a mechanism to find clusters that match s_{new}. The stager should be trained on either the data from the matching clusters or on all the data in the database. In the latter case, discounts or corrections are applied to data from the nonmatching clusters.

2. *Automatic identification of candidates with possible sleep disorders.* A natural way to diagnose sleep disorders is to first use a sleep stager to classify patients' sleep stages and then let physicians check the patterns. However, this might not be the only way of using stagers to diagnose sleep disorders.

Suppose a sleep stager is trained using training data from healthy people and then asked to classify sleep stages for a new person. If the classification accuracy is low, we would suspect this new person of having different characteristics from healthy people and possibly a sleep disorder. It would be interesting to investigate how well this conjecture matches reality.

3. *Language modeling for brain-computer interface.* Besides diagnosing sleep disorders, EEG wave classification is useful for BCIs. In BCI, a person can think about a sentence one letter at a time. Individual letters are recognized by classifying the person's EEG wave. Then, all the recognized letters are concatenated into a sentence that is automatically entered into the computer.

The most straightforward way to implement this is to treat each letter as a state while using a method similar to the sleep staging method to classify EEG waves into individual letters. But this is not the best approach. The sentence the person thinks of is natural language, with its own frequency patterns and characteristics. For instance, some letters occur more often than others (e.g., e vs. z). Some letters are more likely to follow a specific letter than other letters. Certain word pairs are invalid in English. Some words are more likely to follow a specific word than other words. All such information can be used to make EEG wave classification more accurate.

One way to capture the information provided by the structure of the natural language is to use the language modeling method, which has been widely adopted in speech recognition, information retrieval, and machine translation. A language model is a way of assigning probability measures over strings drawn from a language. Each string W has the prior probability $P(W)$. Using the person's EEG waves and the Bayesian framework $P(W \mid X) \propto P(X \mid W)P(W)$, where X represents the observed sequence of EEG feature variable vectors, the posterior distribution for all possible strings can be computed and used to obtain the most likely string. According to past experiences in speech recognition, using the language modeling method can significantly improve classification accuracy.

The literature already contains models for many natural languages related to speech recognition, information retrieval, and machine translation. However, to use the language modeling method to support EEG-based BCI, a large, integrated EEG database with enough training data needs to be built. Current EEG databases for BCI are too small.

Further Reading

Lafferty, J.D., A. McCallum, and F.C. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth international conference on machine learning*. Williamstown.

Luo, G., and W. Min. 2007. Subject-adaptive real-time sleep stage classification based on conditional random field. In *Proceedings of the American Medical Informatics Association annual symposium*, 488–492. Chicago.

Okamoto, K., S. Hirai, M. Amari, T. Iizuka, M. Watanabe, N. Murakami, and M. Takatama. 1993. Oculomotor nuclear pathology in amyotrophic lateral sclerosis. *Acta neuropathologica* 85(5):458–462.

Priestley, M.B. 1981. *Spectral analysis and time series*. New York: Academic Press.

Rabiner, L.R., and B.H. Juang. 1993. *Fundamentals of speech recognition*. Englewood Cliffs: Prentice Hall.

Sellers, E.W., D.J. Krusienski, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw. 2006. A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and interstimulus interval on performance. *Biological Psychology* 73(3):242–252.

Thomson, D.J. 1982. Spectral estimation and harmonic analysis. In *Proceedings of the IEEE* 70(9):1055–1096.

Zhang, L., J. Samet, B. Caffo, and N.M. Punjabi. 2006. Cigarette smoking and nocturnal sleep architecture. *American Journal of Epidemiology* 164(6):529–537.