# Brain wave recognition of words

PATRICK SUPPES*†, ZHONG-LIN LU‡, AND BING HAN*

*Center for the Study of Language and Information, Stanford University, Stanford, CA 94305; and ‡Department of Psychology, University of Southern California, Los Angeles, CA 90089

**ABSTRACT** Electrical and magnetic brain waves of seven subjects under three experimental conditions were recorded for the purpose of recognizing which one of seven words was processed. The analysis consisted of averaging over trials to create prototypes and test samples, to both of which Fourier transforms were applied, followed by filtering and an inverse transformation to the time domain. The filters used were optimal predictive filters, selected for each subject and condition. Recognition rates, based on a least-squares criterion, varied widely, but all but one of 24 were significantly different from chance. The two best were above 90%. These results show that brain waves carry substantial information about the word being processed under experimental conditions of conscious awareness.

In the last two decades new methods of imaging brain activity [positron emission tomography, functional magnetic resonance imaging, and magnetoencephalography (MEG)] have augmented older methods such as electroencephalography (EEG) to dramatically increase our knowledge, especially about where in the brain different kinds of activity occur (1–5). Increased knowledge of temporal sequence of location of activity, as in orally naming a visual object, also has been substantial (6). On the other hand, aside from some success in simple mental state classification (7, 8) and EEG-based human computer communications (9, 10), the detailed analyses of how or what information is processed by the brain, at almost all levels, are still mostly lacking (11). Early attempts to classify averaged EEG waveforms associated with speech production date back to 1967 (12). However, it was later discovered (13) that scalp-recorded potentials preceding and accompanying speech primarily represent volume-conducted activity from musculature involved in speech production. In the current study, we were careful to rule out contributions from muscle movement in the auditory comprehension and internal speech conditions.

The research reported here is meant to be a definitely positive, even if limited, step toward giving such an analysis of brain-wave activity as imaged by EEG and MEG. The brief description of our approach is simple. We analyze brain waves to recognize the word being processed. The general methodological approach is similar to speech recognition, but almost all the details are different. In terms of performance, we are at a level comparable to that of speech recognition in its early days (14, 15).

## METHODS

For subjects S1-S5, EEG and MEG recordings were performed simultaneously in a magnetically shielded room in the Magnetic Source Imaging Laboratory (Biomagnetic Technology, San Diego) housed in Scripps Institute of Research.

Sixteen EEG sensors were used. Specifically, the sensors, referenced to the average of the left and the right mastoids, were attached to the scalp of a subject (F7, T3, T5, FP1, F3, C3, P3, Fz, Cz, FP2, F4, C4, P4, F8, T4, and T6); sensors FP1, Fz, and FP2 were not used with S1 and S2. Two electrooculogram sensors, referenced to each other, were used to monitor eye movement. The Magnes 2500 WH Magnetic Source Imaging System (Biomagnetic Technology) with 148 superconductive quantum interference device sensors was used to record magnetic field near the scalp. The sensor array is arranged like a helmet that covers the entire scalp of most of the subjects. The recording bandwidth was from 0.1 Hz to 200 Hz with a sampling rate of 678 Hz. For the first two subjects, 0.5-s prestimulus baseline was recorded, followed by 2.0-s recording after the onset of the stimulus. For the other three subjects, baseline recording was 0.3 s, followed by 1.2-s recording after the onset of the stimulus.

An Amiga 2000 computer was used to present the auditory stimuli (digitized speech at 22 kHz) to the subject via airplane earphones with long plastic tube leads. The duration of the stimulus was about 300 ms. Stimulus onset asynchrony varied from 2.5 to 2.7 s for the first two subjects, and from 1.5 to 1.7 s for the other three subjects. To reduce the alpha wave in this condition, a scenery picture was placed in front of the subject. He was asked to look at the picture during the recording.

The visual stimuli were generated on the Amiga computer outside of the magnetically shielded room and projected to the front visual field of the subject by using an optical mirror system. Each word was presented for 200 ms, preceded and followed by dynamic random noise masks. Stimulus onset asynchrony was the same as in the auditory comprehension condition.

For subjects S6 and S7, data were collected by using a 64-channel NeuroScan EEG system at the Palo Alto (CA) Department of Veterans Affairs Health Care System. Subjects wore a 64-channel Tin electrode cap that covered the entire scalp during the experiment. The electrodes were connected to two 32-channel SynAmps amplifiers with the linked ears as the reference electrode. The recording bandwidth was set from DC to 200 Hz. with a sampling rate of 500 Hz. The stimulus conditions were similar to those for S1 and S2.

Subjects S1-S5, five normal male native English speakers, aged 25 to 40 years, four right-handed and one left-handed, were run under two or three different conditions, with simultaneous 16-sensor EEG and 148-sensor MEG recordings of brain activity in each condition. The observations recorded were of electric (EEG) or magnetic (MEG) field amplitude every 1.47 ms for each sensor. Two additional male subjects, S6 and S7, the first a 30-year-old native speaker of Chinese but with excellent command of English and the second a 75-year-old native speaker of English, were run in one condition with 64-sensor EEG recordings every 2.0 ms.

Table 1.    Steps of data analysis

1. Normalize baseline for each trial and each sensor.
2. Average over trials for prototypes and test samples.
3. FFT prototypes and test samples.
4. Select optimal bandpass filter.
5. Inverse FFT.
6. Normalize baseline again.
7. Select best EEG and MEG sensors.

All seven subjects were recorded under the auditory comprehension condition of simply presenting randomly one of a small set of spoken words, 100 trials for each word; subjects were instructed to passively but carefully listen to the spoken words and try to comprehend them. Subjects S1-S5 were recorded in the internal speech condition of seeing on a computer screen a single word at a time, for 100 trials per word; in this condition a subject was asked to silently "say" the word immediately after seeing it. In both conditions we emphasized the importance of being aware of the word being processed. Finally, two subjects (S1 and S2) also were run in the normal speech condition, which was like the internal speech condition, except that the subject spoke aloud the word seen on the screen. For subjects S1, S2, S6, and S7, seven words (first, second, third, yes, no, right, and left) were used in all three conditions. For subjects S3-S5, two words (two and here) were added in the auditory comprehension condition, for a total of nine; and in the internal speech condition, two homonyms of two were added, to and too, and one of here, namely, hear, for a total of 12 words.

For each trial, we used the average of 204 observations before the onset of the stimulus as the baseline. After subtracting out the baseline from each trial, to eliminate some noise, we then averaged data, for each word and each EEG and MEG sensor, over every other trial, starting with trial 2, for a total of 50 trials. By using all of these even trials, this averaging created a prototype wave for each word in each condition, auditory comprehension, internal or normal speech. In similar fashion, five test wave forms, using 10 trials each, were produced for each word under each condition by dividing all the odd trials evenly into five groups and averaging within each group. This analysis was labeled E/O. We then reversed the roles of prototypes and test samples by averaging 50 odd trials for prototype and even trials for test. This analysis was labeled O/E.

The main additional methods of data analysis were the following. First, we applied a fast Fourier transform (FFT) to the 1,018 observations (204 observations before and 814 observations after the onset of the stimulus) for each sensor. We then filtered the result with a fourth-order Butterworth bandpass filter (16) selected optimally for each subject, as described in more detail below. After the filtering, an inverse

FFT was applied to obtain the filtered wave form in the time domain, whose baseline was then normalized again.

The decision criterion for prediction was a standard minimum least-squares one. We first computed the difference between the observed field amplitude of prototype and test sample, for each observation of each sensor after the onset of the stimulus, for a total duration of 1.2 s for S1-S5 and 0.84 s for S6 and S7. We next squared this difference and then summed over the observations. The measure of best fit between prototype and test sample for each sensor was the minimum sum of squares. In other words, a test sample was classified as matching best the prototype having the smallest sum of squares for this test sample. These seven steps of data analysis are shown in Table 1.

## RESULTS AND DISCUSSION

**Best Recognition Rates.** In Table 2 we show the predictions of the EEG individual sensors doing the best for each subject in the auditory comprehension and internal speech conditions, where the task was to correctly recognize, by a least-squares comparison with the prototypes, five test samples for each word. (We consider the MEG sensors later.) Each of the 35 test samples was constructed by averaging over 10 individual trials. On the left of Table 2 are the E/O results, and on the right the O/E results.

Several general conclusions emerge at once from Table 2. (*i*) The percentage range of test samples correctly recognized was very wide, from 34% to 97%. (*ii*) In 15 of the 24 predictions, the recognition rate was above 50 percent correct. The null hypothesis is that the sensors of a given set are statistically independent and correctly recognizing test samples of words at a chance level, which in the present case is 1/7. The extreme order statistic for the best sensor performance can easily be computed for this null hypothesis. The probability that among $n$ sensors the best performance is $k$ correct or more out of 35 test samples, is complementary to the joint probability that all of the $n$ sensors predict less than $k$ correct. So, given the statistical independence of the sensors, the exact probability of the best performance being $k$ correct or more is $1 - P(\mathbf{X} < k)^n$ for the binomial random variable $\mathbf{X}$, with 35 trials and probability correct of 1/7, where $n = 13$ for S1 and S2, $n = 16$ for S3-S5, and $n = 60$ for S6 and S7. Of the 24 recognition rates shown in Table 2, all but one is significant at the .01 level, 15 are significant at the $10^{-5}$ level and 8 at the $10^{-10}$ level. These significance levels support the claim that brainwaves carry substantial information about words of which subjects are made consciously aware.

**Selection of Optimal Filters.** Adaptive filters now are widely used in signal processing, but the standard literature assumes the desired signal representation is known (17). Unfortunately, we do not yet know the brain's representation of words. But we

Table 2.    Highest recognition rates for each subject and condition

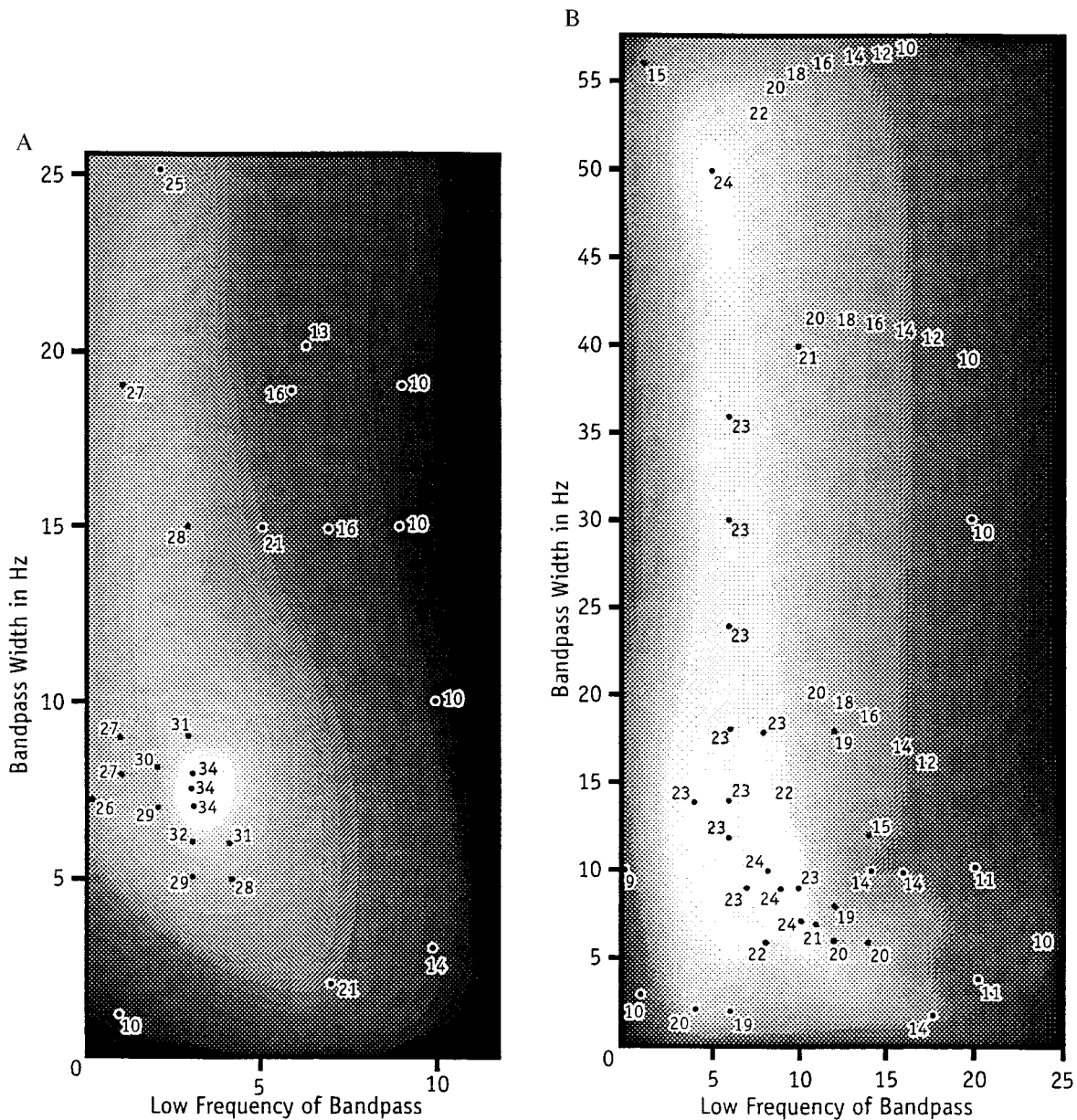| Subject | Auditory comprehension | | | | | | | | Internal speech | | | | | | | |
| | E/O | | | | O/E | | | | E/O | | | | O/E | | | |
| | Best filter (Hz) | | Best EEG sensors | | Best filter (Hz) | | Best EEG sensors | | Best filter (Hz) | | Best EEG sensors | | Best filter (Hz) | | Best EEG sensors | |
| | Low freq | High freq | % | Loc | Low freq | High freq | % | Loc | Low freq | High freq | % | Loc | Low freq | High freq | % | Loc |
| S1 | 2 | 9 | 51 | T3 | 3 | 11 | 51 | Cz | 1.5 | 8.5 | 86 | T6 | 1 | 10 | 91 | T6 |
| S2 | 3 | 13 | 66 | T4 | 4 | 9 | 57 | T4 | 5 | 16 | 51 | T4 | 5 | 16 | 54 | T6 |
| S3 | 3 | 11 | 97 | T3 | 3 | 11 | 77 | T3 C3 | 17 | 23 | 37 | FP2 | 5 | 15 | 37 | F4 |
| S4 | 1.5 | 6.5 | 63 | C3 | 1.5 | 6.5 | 49 | C3 Cz | 1.5 | 7 | 43 | T6 | 2 | 8 | 34 | T6 |
| S5 | 0.5 | 40 | 43 | F4 | 6 | 12 | 37 | C3 | 3 | 11 | 46 | C4 | 2 | 14 | 46 | T6 |
| S6 | 3 | 10 | 71 | C4A | 4 | 12 | 69 | C4A | | | | | | | | |
| S7 | 9 | 15 | 60 | CZA | 5 | 55 | 69 | C1 | | | | | | | | |

FIG. 1.   Shaded contour maps of recognition-rate surfaces. The x-coordinate is the low frequency in Hz of a filter and the y-coordinate is the width in Hz of the filter. The number plotted at a point on the map is the number of test samples of words correctly recognized out of a total of 35 such samples by the Butterworth filter with the coordinates of the point.

can find optimal filters by another strategy, that of optimizing the recognition rate for each subject in a given experimental condition. Because we do not have a detailed theory of the brain's representation of words, we find the optimal filter by intensive computation of the recognition rate for a large number of Butterworth bandpass filters with different cutoff frequencies. Given this set of computations for a given subject and experimental condition, we can draw a recognition-rate surface.

Two such surfaces are shown in Fig. 1. The lighter the shade of gray the higher the recognition rate, with the maximum being shown in white. The changes in recognition rate are slow enough and systematic enough to generate confidence in the approximate correctness of the surfaces, as may be seen from the observations printed on each surface. The surface shown on the left (Fig. 1*A*) is for our best predictive result, 34 of 35 test samples correctly recognized, for a recognition rate of 97 percent, but the general shape of the surface is typical of many others. The surface on the right (Fig. 1*B*), in contrast, is

unusual in the very long narrow plateau, which indicates relative insensitivity to the width of the best bandpass filter.

The variation of surface and of optimal recognition rate, even from E/O to O/E data for a given subject and condition, shows that at the margin the exact recognition rate is affected by small changes in the prototypes or test samples. On the other hand, the regularity of the recognition-rate surfaces shows that the exact bandpass filter selected as optimal can be approximated reasonably well in recognition performances by filters close by, or even at a distance but on the same contour.

**Location of Best EEG Sensors.** Table 2 shows for each of the 24 cases of prediction the EEG sensor whose data were the basis of the prediction. When applicable, ties are shown. As can be seen from the table, 12 of the 24 cases are T3-T6 in the standard EEG 10–20 system, where an occurrence of T3-T6 is counted as one, even when another sensor was tied for best recognition rate. Another nine, with ties counted as one, are either C3, C4, or CZ.

**MEG Data.** Because the recognition rates for EEG sensors were in general so much better than those for MEG sensors,

Table 3.    Confusion matrices for the seven words after FFT and filtering

| | Auditory comprehension | | | | | | | | Internal speech | | | | | | |
| | First | Second | Third | Yes | No | Right | Left | | First | Second | Third | Yes | No | Right | Left |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| First | 47.5 | 2 | 3 | 9.5 | 2.5 | 2.5 | 3 | First | 20 | 2 | 7 | 2.5 | 2.5 | 9 | 7 |
| Second | 7 | 45 | 1.5 | 5 | 5 | 4 | 2.5 | Second | 5 | 33.5 | 3 | 2 | 1 | 3.5 | 2 |
| Third | 2.5 | 2.5 | 53 | 6 | 0 | 4.5 | 1.5 | Third | 6.5 | 4.5 | 25.5 | 1 | 2 | 6 | 4.5 |
| Yes | 3.5 | 4.5 | 8 | 41.5 | 6 | 2.5 | 4 | Yes | 3.5 | 2.5 | 3 | 31 | 5 | 2 | 3 |
| No | 4.5 | 2 | 6 | 10.5 | 35.5 | 5.5 | 6 | No | 3 | .5 | 6 | 11 | 26.5 | 0 | 3 |
| Right | 1.5 | 1.5 | 7.5 | 7.5 | 10 | 35 | 7 | Right | 7 | 3.5 | 2 | 4 | 3.5 | 22 | 8 |
| Left | 2.5 | 1.5 | 2 | 3.5 | 7 | 9.5 | 44 | Left | 0.5 | 2.5 | 11 | 3 | 5 | 2.5 | 25.5 |

we have not exhibited the MEG data in a table. To avoid extensive additional computations in searching for the optimal MEG filters, we compared recognition rates for the best EEG and MEG sensors for each subject (S1-S5) and condition by using a constant filter with a bandpass from 1 to 20 Hz. The results for the 24 cases of prediction were: EEG > MEG in 15 cases, EEG = MEG in four cases, and MEG > EEG in five cases. Moreover, for the auditory comprehension and internal speech conditions, the best MEG sensor was only 49 percent correct (S2, Internal speech, E/O). Given these results we do not report further on the MEG data in this article.

**Electromyography (EMG) Data.** For S1-S5 and all conditions EMG sensors also were recorded, one only next to the mouth for S1 and S2, three for S3-S5, one on the nose, one next to the mouth, and one on the vocal chords. None were used in the predictions given in Table 2. Moreover, for the auditory comprehension and internal speech conditions none of the EMG sensors correctly recognized test samples for any subject strictly better than the best brain-wave sensors, for either EEG or MEG. In the case of the normal speech condition, which was used only with S1 and S2, in three of the eight cases EMG was better than the best of EEG or MEG, which is not surprising, although EEG predictions were quite good, ranging from 71% to 86%.

**Single-Trial Predictions.** There are also several other features of the recognition data we analyzed. Because of the excellent 91% correct brain-wave predictions for S1 in the internal speech O/E condition after FFT and optimal filtering (see Table 2), by using the original prototypes, we tried recognizing the word being processed for three blocks of individual trials, each block containing 105 trials. The three blocks had percents correct of 46, 39, and 43, respectively, which under the null hypothesis for this extreme order statistic has $P < 10^{-7}$ for each block. Because the ultimate recognition objective is to correctly predict what word is being processed on individual trials, i.e., individual occurrences of words, as in the case of speech recognition, we were encouraged by this result, although it is evident it is far from the best possible.

**Subject-Independent Predictions.** A natural question is if the prototypes of one subject, in a given experimental condition, can be used to recognize correctly the test samples of another subject in the same experimental condition. Surprisingly, within limits the answer is affirmative. By using the suboptimal uniform filter from 1 to 20 Hz, we summarize the results for three pairs of subjects and report here only on the best EEG sensor. In the auditory comprehension condition, when using S1 O/E prototypes, 16 S2 E/O test samples of 35 were classified correctly, and when reversed, 11 samples were

Table 4.    Confusion matrices for the homonyms after FFT and filtering

| | Internal speech | | | | | Internal speech | |
| | Too | Two | To | | | Here | Hear |
|---|---|---|---|---|---|---|---|
| Too | 11 | 10 | 9 | Here | | 20 | 10 |
| Two | 7 | 10 | 13 | Hear | | 13 | 17 |
| To | 8 | 4 | 18 | | | | |

classified correctly. When using S3 E/O prototypes, 11 S4 E/O test samples were classified correctly, and when reversed, the number was 12. Finally, in the internal speech condition, using S1 O/E prototypes, 16 S2 E/O samples were correctly classified, and when reversed, the number was 18. The three results of 16, 16, and 18 are significant at the 0.01 level.

**Confusion Matrices.** On the other hand, the confusion matrices, given in Table 3, for the seven words summed over all subjects, for the auditory comprehension and internal speech conditions, show clearly enough that no simple features, such as duration of the word being processed, will lead to reliable discrimination. (As is usual in such matrices, each row shows the frequency of classification of the test samples of a given word.)

With S3-S5 we also did a small experiment in the internal speech condition on homonyms. We presented the words to, too, and two and the words here and hear randomly among the seven words listed above and given to all subjects. The two confusion matrices are shown in Table 4. The null hypothesis of each cell having the same number of observations is rejected at $P \le .05$ by a $\chi^2$ test for each contingency table. The borderline character of this significance test suggests further investigation, for a strong rejection of the null hypothesis would show that more than purely phonetic processing took place.

**Conclusion.** Brain-wave recognition of words being processed is feasible in simple experimental conditions, but even in such conditions recognition results leave substantial room for improvement. The very best results obtained of 97% and 91% correct recognition are encouraging, but the wide variability in percent correct for different subjects and different experimental conditions indicates that the path to follow for continued improvement is not easily discerned. All the same, we are confident that the present brain-imaging technologies available are sufficient to permit continued scientific progress by us and others in recognizing and understanding what the brain is processing under conditions of conscious awareness.

**Open Questions.** We recognize that the research reported here suggests many more questions than it answers. The range of these questions is too great for us to canvass, but there are two that we think have special significance. First, how can we significantly improve our predictions by using as our primary representation of a word the spatiotemporal, continually changing brain image of a word? In other words, will "movies" of the brain processing be the source of the additional concepts we need for recognition? Also, will it be practical to use machine learning techniques to determine, among the many spatiotemporal features we can identify, those that are really crucial for correct prediction?

Second, it is clear from the research reported here that the global electric and magnetic brain waves carry significant information about words being processed. But is this global level adequate when there is no conscious awareness of the processing, or are additional necessary data only to be found in the relatively inaccessible individual neurons (or highly local collections of them) that are the sources of the fields, i.e., waves? The extensive research on motor control in both

humans and monkeys shows beyond much doubt that the relevant cognitive processing uses in an essential way time-varying populations of neurons (18), which makes detailed neuronal observations even more difficult. In fact, recent evidence from aphasic patients (19) suggests that different populations of neurons are used for storing regular and irregular past tenses of verbs.

1. Posner, M. I., Petersen, S. E., Fox, P. T. & Raichle, M. E. (1988) *Science* **240,** 1627–1631.
2. Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E.-J. & Shadlen, M. N. (1994) *Nature (London)* **369,** 525.
3. Williamson, S. J., Lu, Z.-L., Karron, D. & Kaufman, L. (1992) *Brain Topography* **4,** 169–180.
4. Hämäläiinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J. & Lounasmaa, O. V. (1995) *Rev. Mod. Phys.* **65,** 413–497.
5. Gevins, A. S. & Remond, A., eds. (1987) *Methods of Analysis of Brain Electrical and Magnetic Signals* (Elsevier, New York).
6. Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T. & Havinga, J. (1991) *Psychol. Rev.* **98,** 122–142.
7. Anderson, C. W., Devulapalli, S. V. & Stolz, E. A. (1995) *Scientific Programming* **4,** 171–183.
8. Jung, T. P., Makeig, S., Stensmo, M. & Sejnowski, T. J. (1997) *IEEE Trans. Biomed. Eng.* **44,** 60–69.
9. Wolpaw, J. R. & McFarland, D. J. (1994) *Electroenceph. Clin. Neurophysiol.* **90,** 444–449.
10. Farwell, L. A. & Donchin, E. (1988) *Electroenceph. Clin. Neurophysiol.* **70,** 510–523.
11. Bullock, T. H. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 1–6.
12. Schafer, E. W. P. (1967) *Nature (London)* **216,** 1338–1339.
13. Szirtes, J. & Vaughan, H. G. (1977) *Electroenceph. Clin. Neurophysiol.* **43,** 386–396.
14. Lindgren, N. (1965) *IEEE Spectrum* **2,** 114–136.
15. Reddy, D. R. (1976) *Proc. IEEE* **64,** 501–531.
16. Oppenheim, A. K. & Schafer, R. W. (1975) *Digital Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ), pp. 211–218.
17. Haykin, S. (1996) *Adaptive Filter Theory* (Prentice–Hall, Englewood Cliffs, NJ), 3rd Ed.
18. Georgopoulos, A. P. & Pellizzer, G. (1995) *Neuropsychologia* **33,** 1531–1547.
19. Marslen-Wilson, W. D. & Tyler, L. K. (1997) *Nature (London)* **387,** 592–594.