

Invariance between subjects of brain wave representations of language

Patrick Suppes^{*†}, Bing Han^{**}, Julie Epelboim^{*}, and Zhong-Lin Lu[§]

^{*}Center for the Study of Language and Information and [†]Department of Engineering–Economic Systems and Operations Research, Stanford University, Stanford, CA 94305; and ^{**}Department of Psychology, University of Southern California, Los Angeles, CA 90089

Contributed by Patrick Suppes, August 18, 1999

In three experiments, electric brain waves of 19 subjects were recorded under several different experimental conditions for two purposes. One was to test how well we could recognize which sentence, from a set of 24 or 48 sentences, was being processed in the cortex. The other was to study the invariance of brain waves between subjects. As in our earlier work, 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. A least-squares criterion of fit between prototypes and test samples was used for classification. In all three experiments, averaging over subjects improved the recognition rates. The most significant finding was the following. When brain waves were averaged separately for two nonoverlapping groups of subjects, one for prototypes and the other for test samples, we were able to recognize correctly 90% of the brain waves generated by 48 different sentences about European geography.

The three experiments reported here extend our earlier work (1, 2) on brain wave recognition of words and sentences in three significant ways. First, more complex sentences requiring an evaluation of their truth or falsity were used as auditory or visual stimuli. Second, the number of different sentences presented to subjects was increased 4-fold to 48. Third, in several different experimental conditions, brain wave data were averaged across subjects. No such averaging was considered in our earlier studies. Electroencephalographic (EEG) recordings of brain waves were made in all three experiments. Relevant studies of brain processing of language, especially through EEG recordings, were reviewed in refs. 1 and 2.

Methods

For all subjects, EEG recordings were made in our laboratory using 22 model 12 Grass amplifiers and Neuroscan's SCAN 4 software (SCAN 3 was used for experiment I; Neuroscan Inc., Sterling, VA). Sensors were attached to the scalp of a subject according to the standard 10–20 EEG system, either as bipolar pairs, with the recorded measurement in millivolts being the potential difference between each such pair of sensors, or single sensors referenced to the left or right mastoid; the distribution of sensors on the scalp is shown in Fig. 3. For experiment I, the recording bandwidth was from 0.3 to 100 Hz with a sampling rate of 750 Hz for the syllable part and 384 Hz for the sentence part; for the other two experiments, the recording bandwidth was from 0.3 to 100 Hz with a sampling rate of 1,000 Hz. The length of recording of individual trials varied with the experiments, as described below.

A computer was used to present auditory stimuli (digitized speech at 22 kHz) to subjects via small loudspeakers. Visual stimuli were presented on a standard computer monitor.

Nineteen subjects participated in the experiments. We number the subjects consecutively with the nine subjects used in refs. 1 and 2 because we continue to apply new methods of analysis to our earlier data; S7 (from ref. 1) was also used as a subject. Subjects S7, 10, 12, 13, and 18–23 were used in experiment I, which took place in November, 1998; subjects S10–19 were used

in experiment II, which took place in January, 1999; and subjects S10–13, 16, 18, and 24–27 were used in experiment III, which took place in March, 1999, but S18 participated only in one of the two conditions. Eight of the subjects were female and eleven were male, ranging in age from 18 to 76 years. Two were left-handed, and two were not native English speakers.

In the first part of experiment I, with S7, S10, and S13 as subjects, on the basis of the results in refs. 1 and 2, only the following six bipolar pairs of sensors were attached to the scalp: C3–T5, C4–T6, T3–C3, T4–C4, T5–T3, and T6–T4. In the second part, an additional sensor, Cz, in the bipolar pair Cz–T5, was added to all nine subjects (S7 did not participate). In experiment II, S10–15 and S19 had the following 16 sensors attached to the scalp: Fp1, Fp2, F7, F3, F4, F8, T3, C3, C4, T4, T5, P3, Pz, P4, and T6. S16–18 had 22 bipolar pairs: Cz–Fz, Cz–F4, Cz–C4, Cz–P4, Cz–Pz, Cz–P3, Cz–C3, Cz–F3, Fz–Fp2, Fz–Fp1, F4–Fp2, F4–F8, C4–F8, C4–T4, C4–T6, P4–T6, P3–T5, C3–T5, C3–T3, C3–F7, F3–F7, and F3–Fp1. In experiment III, all subjects had the 22 bipolar pairs.

In most of our analyses, starting with trial 2, we averaged data for half of the total number of trials over every other trial, for each sentence or syllable and each EEG sensor. Using all of these even trials, this averaging created a prototype wave for each sentence or syllable. In similar fashion, the odd trials were used to produce a test wave for each sentence or syllable. This analysis was labeled E/O (even/odd). We then reversed the roles of prototypes and test samples by averaging odd trials for prototype and even trials for test. This analysis was labeled O/E. In ref. 1, we performed both analyses for all cases, and the two did not greatly differ. Here, as in ref. 2, the analyses were performed only for E/O. We imposed this limitation because of the larger data files for the sentences as compared to the isolated words studied in ref. 1.

The main additional methods of data analysis were these. First, we took the complete sequence of observations, for each prototype or test, and placed the sequence in the center of a sequence of 4,096 observations for the sentences or 2,048 observations for the syllables. We next filled the beginning part of the longer sequence by mirroring the beginning part of the centered prototype or test sequence, and we filled the ending part by mirroring the ending part of the center sequence. Using a Gaussian function, we then smoothed the two parts filled in by mirroring. The Gaussian function was put at the center of the whole prototype or test sequence with a standard deviation equal to half of the length of the prototype or test sequence. Each observation in the longer sequence that was beyond the centered prototype or test sequence was multiplied by the ratio of the value of the Gaussian function at this observation and the value of the Gaussian function at either end of the centered prototype or test sequence. After mirroring and smoothing, we applied a

Abbreviation: EEG, electroencephalography.

[†]To whom reprint requests should be addressed. E-mail: suppes@ockham.stanford.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Steps of data analysis

Step	
1	Average over trials, half for prototypes, other half for test samples
2	Mirror and smooth with a Gaussian function
3	Fast Fourier transform prototypes and test samples
4	Filter the results with a fourth-order Butterworth bandpass filter having parameters (L, H)
5	Inverse fast Fourier transform
6	Classify by least-squares criterion, for observations in a given temporal interval (<i>s</i> , <i>e</i>)
7	Repeat steps 4–6 with new parameters (<i>L'</i> , <i>H'</i> , <i>s'</i> , <i>e'</i>) from a set of values selected on past experience until the set is exhausted
8	Select best recognition performance and corresponding parameters

(discrete) fast Fourier transform to the whole sequence of 4,096 observations for the sentences or 2,048 observations for the syllables for each sensor. (The fast Fourier transform algorithm used restricted the number of observations to a power of 2.) We then filtered the result with a fourth-order Butterworth bandpass filter (3) as described in ref. 1. After the filtering, an inverse fast Fourier transform was applied to obtain the filtered wave form in the time domain.

As in refs. 1 and 2, the decision criterion for prediction was a standard 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 duration whose beginning and end were parameters to be estimated. We next squared this difference and then summed over the observations in this interval. 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.

Using this procedure of analysis, we estimated four parameters for each subject in each of the conditions in the three experiments. First, we estimated the low frequency (L) and the high frequency (H) of the optimal bandpass filter (“optimal” defined, as in refs. 1 and 2, in terms of correct recognition rate). Simultaneously, we estimated, again for the best recognition rate, the starting point (*s*), after the onset of the stimulus, and ending point (*e*) in ms of the sample sequence of observations used for recognition, with the same *s* and *e* for a given set of stimuli to be recognized. These methods of analysis are summarized in Table 1. The parameters *s* and *e* are omitted in the tables, because quite often the gradients were too flat to make the selection of *s* or *e* other than arbitrary within a couple of hundred milliseconds. Means only of *s* and *e* are given in *Results* for experiments II and III. Some typical recognition-rate surfaces are shown in ref. 2. In Tables 2 and 3 (see below), the best EEG sensor, or bipolar pair of sensors, is shown in the second column of data for each condition, and the optimal bandpass filter is shown in the third column. The first column of data in Tables 2 and 3 shows the recognition rates achieved, expressed in percent.

To test the hypothesis that it is averaging *per se*, not averaging across subjects as such, that improves recognition rate by eliminating noise, we constructed for each condition (experiments II and III) a “typical” subject (TypUS or TypBS, U for unipolar, B for bipolar), having 10 trials on each stimulus in each condition and each experiment by selecting the third and seventh trials from the five subjects who had the highest recognition rates, with a variation for the three bipolar subjects in experiment II.

Experiment I: Exploratory

Procedures. In experiment I, which was meant to be exploratory, the first part consisted of auditory presentation to three subjects of 8 syllables such as *pa* and *to*, and 24 syllable pairs, such as *paba*. The list of 32 stimuli was randomly presented 12 times in the first session, and, after a short break, 13 times in the second session, for a total of 800 trials. Subjects were instructed to listen carefully to the syllables, but no overt response was required. The interstimulus interval was 2,050 ms.

In the second part of experiment I, also exploratory, 24 different sentences were presented auditorily at a natural didactic pace in 10 randomized blocks, with all 24 sentences in each block. The 24 sentences were statements about commonly known geographic facts of Europe, with half of the sentences true, half false, half negative, and half positive, e.g., *The capital of Italy is Paris* and *London is not the largest city of France*. The possible forms of sentences were: X is W of Y, X is not W of Y, W of Y is X, W of Y is not X, X is Z of X, X is not Z of X, where $X \in \{\textit{Berlin, London, Moscow, Paris, Rome, Warsaw}\}$, $Y \in \{\textit{France, Germany, Italy, Poland, Russia}\}$, $W \in \{\textit{the capital, the largest city}\}$, and $Z \in \{\textit{north, south, east, west}\}$. The maximum length of the 24 sentences was 2.59 s, and the minimum length was 1.35 s. The mean length was 2.18 s with a standard deviation of 0.30 s. After a spoken sentence was heard by the subject, the subject was asked to orally respond “true” or “false,” with the sentences and responses being recorded in real time on tape. The interval between onset of successive sentences was 4,050 ms.

Results. Because this experiment was exploratory in terms both of stimuli and method of response, we summarize the results briefly, with an emphasis on the averaged results. For the first part, with 32 syllables as spoken stimuli, we used 416 of the trials to create 32 prototypes and the other 384 trials to create 32 test samples, for each of the three subjects. We were able to correctly identify 47% of the test samples for S7, 35% for S10, and 28% for S13. Averaging the data from all three subjects led to an improvement in recognition rate to 59%.

In the second part of the experiment, for the 24 spoken sentences, each sentence repeated 10 times, we obtained the following recognition results when we averaged for each subject every other of the 10 trials for each sentence as a prototype and the remaining half to form two test samples for each sentence. For recognition purposes, the best subject was S18, with a rate of 44% (21/48)—this subject was also the best in experiment II (see Table 2). The worst rate, at 21%, was from S19. Averaging the data over seven subjects (omitting two because of format problems) and having only one test sample, instead of two, to have more averaging, we obtained a recognition rate of 67% (16/24), which exceeded the best individual result by a surprising 23%.

Experiment II: Twenty-Four Sentences

Procedures. In experiment II, 24 sentences were presented visually, one word at a time in 10 randomized blocks, with the same timing sequence as the identical spoken sentences that were used in the second part of experiment I and as part of experiment III. In particular, the onset time of each visual word was the same as the onset time of the corresponding auditory word, measured from the beginning of the stimulus. The offset of the last word of a visually presented sentence was at the same time as its offset when presented auditorily. Only one visual word was shown at a time on the screen; the screen was not blanked out between words.

After a visual sentence was presented to the subject, the prompt “???” appeared on the screen for the subject to respond. Because of problems with the variability of the oral response in experiment I, in experiment II, the numerical keypad was used

Table 2. Experiment with 24 sentences

Subject	%	Sensor	Filter (Hz)	Correct answer,	Mean latency (ms)	
				%	Correct	Incorrect
S10	79	T5	3–8	70	1,027	1,046
S11	83	P3	2–6	93	1,019	2,125
S12	71	T6	4–10	98	753	1,353
S13	54	P3	1–6	95	600	907
S14	38	T5	3–14	90	584	835
S15	25	T5	2–21	52	1,344	1,636
S16	58	Cz–P3	1–9	91	819	1,245
S17	38	F4–Fp2	2–14	92	1,085	2,198
S18	100	P4–T6	0.5–10	94	787	1,004
S19	75	T5, T6	1–17	88	1,059	1,508
TypUS	38	T5	2–12			
TypBS	75	C4–T6	2–8			
AvgUS	88	T6	3–10			
AvgBS	100	C4–T6	2–14			
SepAS	83	C4	2–10			

All sentences were visual. TypUS, typical unipolar subject; TypBS, typical bipolar subject; AvgUS, averaged unipolar subjects; AvgBS, averaged bipolar subjects; SepAS, separately averaged subjects.

by subjects to respond, with 1 = true and 2 = false. The interstimulus interval was 4,050 ms. If a response was not received in time, the system gave the subject an additional 1 s. If no response was received 5 s after the onset of the sentence, the next sentence began and the absence of a response was flagged and stored as 0.

Results. In ref. 2, we studied the brain waves generated from hearing 12 three-word sentences. In the present experiment, following the exploratory effort in experiment I, we doubled the number of sentences to 24, increased their complexity and length, and required that subjects evaluate the truth or falsity of each visually presented sentence by pressing one of two keys on the computer keyboard. The percent correct evaluations for each subject and mean latencies of correct and incorrect key responses are shown in Table 2. Each of the sentences, such as *Berlin is not west of London* was easy for subjects to evaluate, but the pace of presentation was brisk enough to make all subjects answer incorrectly some of the time, the best percent being 98 and the worst 52. A similar individual range is noticeable in the mean latencies of response. On the other hand, in spite of these variations, all 10 subjects' mean incorrect response latencies were slower than those for correct responses.

The main results of this experiment, the brain wave recognition rates we obtained, are also shown in Table 2. By far the best-performing individual subject was S18 with 100%. In particular, 24 test samples were recognized correctly when compared to 24 prototypes, each prototype and test sample being the average of 5 trials. The wide range of success we had in recognition is shown by the lowest result of 25% for S15. On the other hand, for 5 of the 10 subjects we were able to recognize more than 70% of the test samples.

The sensors, bandpass filters, and optimal time intervals for individual subjects show the variations found in our earlier studies (1, 2), and also the predominance of optimal sensors being T5, and its dual, T6, along with F4, P3, P4, and C4. The unipolar mean *s* (starting time after onset of stimulus) of optimal time interval was 219 ms and the unipolar mean ending *e* was 2,515 ms (the means were averaged over the 7 unipolar subjects). The bipolar means, averaged over three subjects, were *s* = 150 ms and *e* = 2,833 ms.

In this experiment, we averaged together, first, the unipolar

group of seven subjects and then, separately, the three bipolar subjects. For the "typical" unipolar subject (two trials from the highest five subjects), we recognized 38% of the test samples, but when all seven were averaged together (each prototype and test sample being the average of 35 trials), we got 88%. In the bipolar case, for the "typical" subject, with four trials from S18 and three each from S16 and S17, we recognized 75%. When all three subjects' data were averaged (each prototype and test sample being the average of 15 trials), we got the same as for S18 alone, namely, 100% (24 correctly recognized out of 24 test samples, one test sample for each sentence).

Experiment III: Forty-Eight Sentences

Procedures. Experiment III, which used the true–false response procedures of experiment II, included two sessions on different days, one with spoken sentences and one with the same sentences presented visually, as in experiment II. Experiment III used the 24 sentences of experiment II, but extended the number to 48, presented in 10 randomized blocks, with all 48 sentences in each block. We added city and country names to increase the pool of possibilities, and added the forms Y is Z of Y and Y is not Z of Y. New city names were *Athens*, *Madrid*, and *Vienna*, and the new country names were *Austria*, *Greece*, and *Spain*. Of the 48 sentences, the maximum sentence length was 2.59 s, and the minimum length was 1.35 s. The mean length was 2.27 s, with standard deviation 0.31 s.

The corresponding visual presentation of the 48 sentences matched the timing of the spoken sentences, as described above for experiment II. The interstimulus interval and the use of the numerical keypad for responding were also the same as in experiment II. Of the nine subjects who were in both conditions, five began with the auditory condition and four with the visual one.

Results. This experiment differs from experiment II in two important respects. First, we doubled the number of different sentences to 48. Second, we ran each subject on both the auditory and visual conditions, with the exception of S18, who ran only on the auditory condition. The two conditions, as described earlier, were closely matched, with the words appearing one at a time in the visual condition on a temporal schedule that matched the pacing of the spoken speech for each sentence in the auditory condition. We note that reading one word at a time, without any eye movements, has been shown to be the most effective fast way to read for detailed content (4). What is surprising is that our

Table 3. Experiment with 48 sentences

Subject	Auditory			Visual		
	%	Sensor	Filter (Hz)	%	Sensor	Filter (Hz)
S10	23	C4–T4	2–6	56	C3–T5	1–15
S11	10	F4–F8	2–14	29	C3–T5	2–5
S12	15	Cz–P4	2–21	60	Cz–P4	1–9
S13	15	F3–F7	7–14	40	C3–T5	0.5–11
S16	15	C3–T5	2–21	38	P4–T6	2–6
S18	25	C3–T3	3–12			
S24	25	C3–T5	1–7	56	P4–T6	1–7
S25	17	F3–F7	5–6	35	C3–T3	1–8
S26	13	F3–F7	2–21	79	C4–T6	1–15
S27	19	C4–T4	2–4	38	Cz–P3	1–14
TypS	19	C3–T5	2–8	38	C3–T5	3–5
AvgS	58	C3–T5	2–7	98	C4–T6	2–6
SepAS	27	C3–T5	1–10	90	C4–T6	2–8
MixP	17	C3–T5	3–10	77	P4–T6	1–10

TypS, typical subject; AvgS, averaged subjects; SepAS, separately averaged subjects; MixP, mixed prototype.

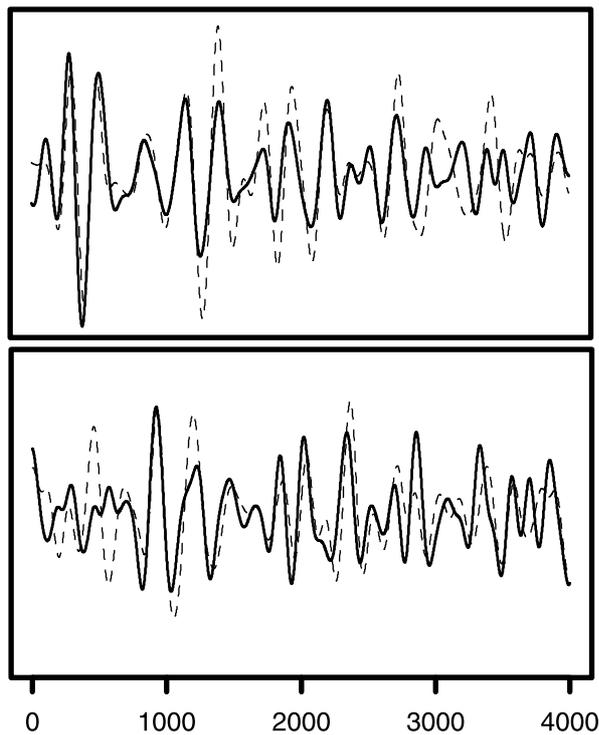


Fig. 1. Prototypes and test samples generated by two sentences. (Upper) The averaged, filtered brain waves for the sentence *The capital of Italy is Paris*. The solid curved line is the prototype, and the dotted line is the test sample. (Lower) The corresponding brain waves for the sentence *The largest city of Austria is not Warsaw*. Both halves were from the bipolar pair of sensors C4–T6. The x axis is measured in ms after the onset of the first word of the sentence.

recognition rate was so much better for the brain waves in the visual condition than in the auditory condition. As can be seen from Table 3, the results were better for all nine subjects, and even more surprising, the recognition-rate percent was at least twice as great for all nine subjects. We have no clear explanation of this highly significant difference. The corresponding comparison of the same two conditions in ref. 1 was not nearly so one-sided. One important difference was that in the present experiment, subjects were required to judge and respond overtly as to whether each sentence was true or false. In the earlier experiment (1), subjects were asked only to silently repeat each visual word displayed. We did try several additional methods of analysis on the data for the auditory condition, e.g., normalization of individual trials, but without any significant improvement.

The optimal filters and sensors for subjects are shown in Table 3, but not the optimal time intervals. For the auditory condition, the mean s , averaging over subjects, all of whom were bipolar, was 205 ms and the mean e was 2,710 ms. For the visual condition, the mean s was 217 ms and the mean e was 2,544 ms.

The comparison of “typical” subjects, as defined earlier, followed the same trend, with recognition rate of 19% for the auditory condition and 38% for the visual condition.

As before, we also averaged brain waves from all subjects together, using every other trial for prototypes and the remaining trials for test samples. For the auditory condition, this massive averaging produced a spectacular improvement, 58% correct recognition of brain waves, which is more than twice the percent achieved for any single subject. In the visual condition, there was clear improvement from such averaging, but necessarily not as much: 98% (47 of 48 test samples correctly recognized), as compared to 79% for the best single subject, S26.

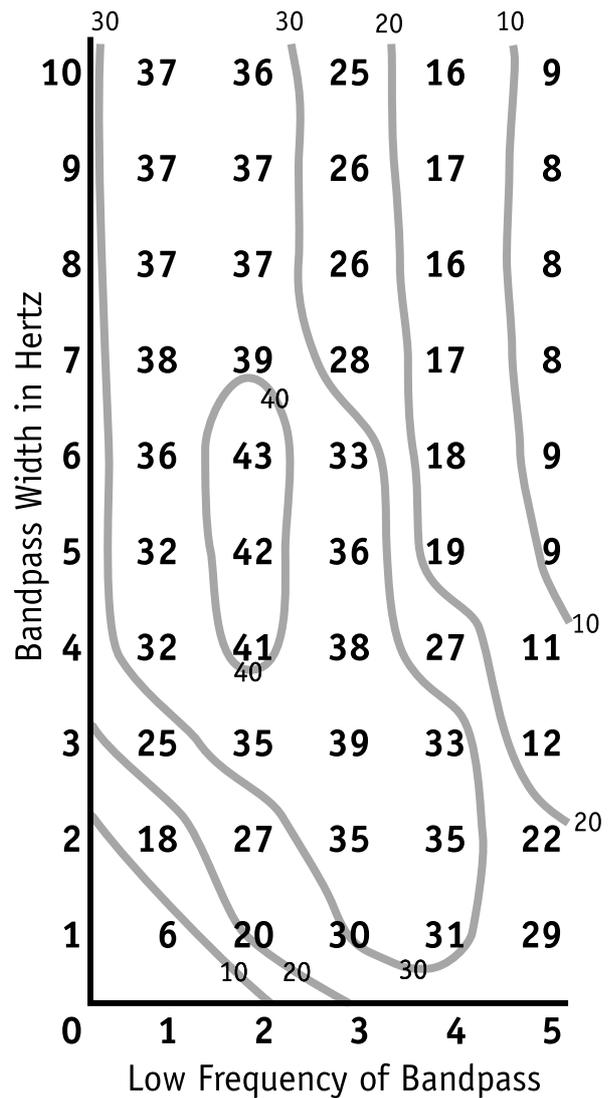


Fig. 2. Contour map of recognition-rate surface for bandpass filter parameters L and W for the separately averaged (SepAS) prototypes and test samples in experiment III. The x coordinate is the low frequency (L) in Hz and the y coordinate is the width (W) in Hz of a possible filter. The number plotted at a point on the map is the number of test samples, of 48 possible, correctly recognized with the Butterworth filter whose parameters are the coordinates of the point. The prediction at that point is that of the best bipolar pair of sensors, for that filter, so the surface shown optimizes the choice of bipolar pair for the parameters of a given filter. The single best prediction, 43 of 48, was made by the bipolar pair C4–T6.

Two other averaging methods were used here and are shown in Table 3. As would be expected, the Separate Average (SepAS) did not do quite as well. This is when one group of subjects was averaged for the prototype and the remaining group of subjects was averaged for the test samples. The results were 27% for the auditory condition and 90% for the visual condition, which is still a good result, even though less than the 98% we obtained when all subjects were averaged together (AvgS). This 90% is, in fact, our best result for invariance of sentences, because no subject contributed trials to both the prototypes and the test samples. In Fig. 1, we plot the filtered and separately averaged waves for prototypes and test samples for the two sentences *The capital of Italy is Paris* and *The largest city of Austria is not Warsaw*.

To exhibit aspects of this invariance result in more detail, in Fig. 2 we show the recognition-rate contour map of possible

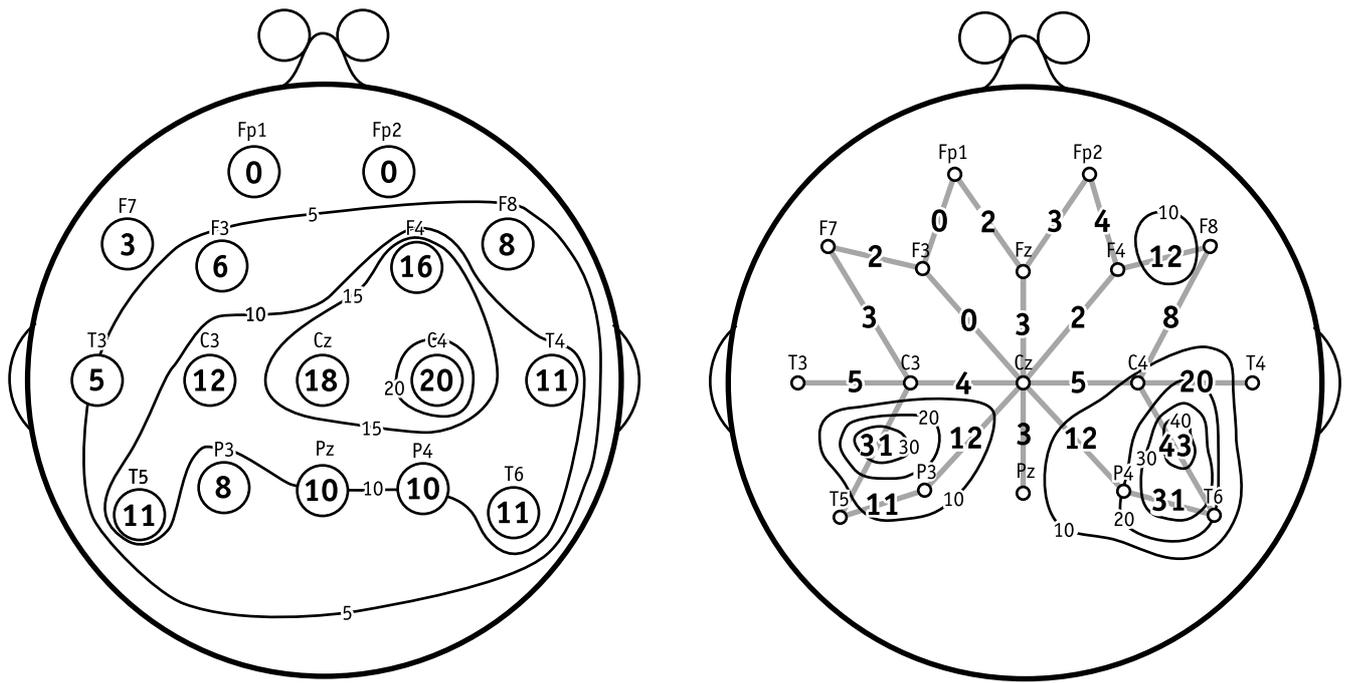


Fig. 3. Contour maps of recognition-rate surfaces for experiment II (Left) and experiment III (Right). Left shows 16 sensors of the 10–20 EEG system; Right shows 22 bipolar pairs of the 10–20 system. In both cases, the number plotted next to the location of a sensor (Left) or at the midpoint of the segment joining a bipolar pair (Right) is the number of test samples of sentences correctly recognized.

bandpass filters. The low frequency (L) of the filters is shown on the x axis and the width (W) of the possible filters is shown on the y axis. As can be seen, the best result is $L = 2$ Hz and $W = 6$ Hz, and so $H = L + W = 8$ Hz, the high frequency of the bandpass filter. The gradient around the peak of 43 of 48 sentences correctly recognized is smooth and gradual, but the gradient along the L axis (i.e., the x axis) is clearly steeper than along the W axis. The important point is that the smooth and systematic behavior of the gradients in all directions from the peak of 43 shows that this best result is not a statistical singularity of a chance nature. More generally, the regularity of the contours is sufficient evidence to decisively reject a null hypothesis of random fluctuations as a model of the recognition data.

In Fig. 3, we show recognition-rate contour maps for the EEG sensors, unipolar in the case of experiment II and bipolar in the case of experiment III. The map on the left is for the 24-sentence data of experiment II and that on the right is for the 48-sentence data of experiment III. The maps show similar patterns, with considerable symmetry between the left and right hemispheric recognition rates, rather like the purely bipolar results reported in ref. 2. The contour lines of the left are for recognition rates of 5, 10, 15, and 20; on the right, with twice as many sentences, for rates of 10, 20, 30, and 40. We note that the best result of 43/48 is, at 90%, greater than that for the 24-sentence case (20/24, or 83%), although one would expect, *a priori*, a better rate for the smaller set of sentences. In our view, the most plausible explanation is that the 24-sentence experiment was unipolar and the 48-sentence one was bipolar. This view is supported by the generally better results for the three bipolar subjects in experiment II compared with the results for the seven unipolar subjects. Details on this point can be found in Table 2.

Finally, we used recorded brain waves from both conditions to form a “mixed” prototype (MixP), half of the trials from the auditory condition and half of the trials from the visual condition. We then used the other half of each condition to form 48 test samples, one for each sentence. The results, as shown in Table 3, were 17% for the auditory condition and 77% for the

visual condition. Given the relatively poor results for classifying the auditory condition, even when averaged altogether, we were surprised at the 77% for the visual condition. It shows that the features of the brain waves recorded under the visual condition come through even in the presence of what is the apparently more noise-prone waves recorded under the auditory condition.

Discussion

Averaging and Invariance. Various results we obtained in the three experiments show that averaging EEG-recorded brain waves across subjects can lead to much improved recognition rates. The mere fact of this successful averaging across subjects supports the view that the actual forms of brain waves for words and sentences have an objective or invariant status, in the sense of being very similar across subjects. Although not much previously investigated, as far as we know, the result is not at all paradoxical, given how easily the same speaker or printed text is understood so quickly and easily by many different persons.

On the other hand, we are not claiming, and have not demonstrated, that averaging across many trials for a number of subjects is better than averaging across the same number of total trials for one individual. In fact, our averaging over a small number of trials and subjects to generate a “typical” subject did not support the idea that averaging across subjects *per se* is helpful. It seems likely the main effect is the total number of trials. It is also obvious why we did not push this idea further. In experiment III, with 48 sentence stimuli, it would have required another 18 experimental sessions by a single subject to match the total number of trials used in the two conditions.

The important conceptual point about averaging across subjects is that in a number of cases it works well. To confirm still further our findings on averaging, we went back to the data of ref. 1, the auditory condition with seven isolated words to be recognized. We averaged all of the data for S3, S4, and S5, which consisted of 300 trials for each word. Using half of the data for each word as a prototype and the other half as a test sample, we

correctly recognized 7 of 7 test samples (100%), which is better than any of the individual recognition rates reported in ref. 1.

Separate Averaging. Someone of a skeptical nature, ignoring the positive point that we can average across different subjects, might still say of our averaging results that in some sense it is not surprising, since each subject in the average contributes trials to both the prototype and test sample of a given type. It is in anticipation of this response that we also used what we have termed Separate Averaging, meaning that subjects averaged for the prototypes are not the same as those averaged for test samples in a given experiment. In particular, in experiment II, with 24 sentences as stimuli, we got an 83% recognition rate, and in experiment III, with 48 sentences as stimuli, a 90% recognition rate in the visual condition, when we used Separate Averaging. We emphasize again that no subject contributed data to both a prototype and a test sample. The 83% and 90% recognition rates are our most salient experimental results supporting the surprising objectivity of brain wave shapes across subjects.

William James (5) summarized well over 100 years ago the general view about perception still held today: “. . . whilst part of what we perceive comes through our senses from the object

before us, another part (and it may be the larger part) always comes out of our own heads.” Without denying the subjective aspects of much perception, in contrast to James’ emphasis, we point to our evidence that electric activity in the cortex, lagging a couple of hundred milliseconds, at the very most, behind the auditory or temporally paced visual words of the sentences we presented, has a very similar wave shape across subjects, especially when averaged and filtered to remove noise.

We recognize that, in averaging across individuals, we may eliminate not just pure noise but information significant for a particular individual. The presence of such additional information, e.g., associative ties to personal memories, is not inconsistent with our claim that words and sentences in and of themselves have an approximately invariant brain-wave representation across a large population of individuals using them.

We thank Dr. David Spiegel, Department of Psychiatry, Stanford Medical School, for loan of the Grass amplifiers and related EEG equipment used in the experiments reported here. We thank Paul Dimitre for producing the figures and Ann Gunderson for preparation of the manuscript. We have also received useful comments and suggestions for revision from George Sperling.

1. Suppes, P., Lu, Z.-L. & Han, B. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 14965–14969.
2. Suppes, P., Han, B. & Lu, Z.-L. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 15861–15866.

3. Oppenheim, A. V. & Schafer, R. W. (1975) *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ), pp. 211–218.
4. Rubin, G. S. & Turano, K. (1992) *Vision Res.* **32**, 895–902.
5. James, W. (1890) *Principles of Psychology* (Henry Holt, New York), Vol. 2, p. 103.