Invariance of brain-wave representations of simple visual images and their names

Patrick Suppes†§, Bing Han†§, Julie Epelboim†, and Zhong-Lin Lu¶

†Center for the Study of Language and Information, Stanford University, Stanford, CA 94305-4115; §Department of Management Science and Engineering, Stanford University, Stanford, CA 94305-4023; and ¶Department of Psychology, University of Southern California, Los Angeles, CA 90089

In earlier work, we have reported on brain-wave representa-

tions of language. Initially we concentrated on being able to

ercognize correctly a single word being processed in the cortex

(1). We next focused on brain-wave recognition of sentences (2).

Most recently, we extended this work to a larger set of 48

sentences, presented as either spoken or printed text. The

important finding was that brain-wave recognition rate was

notably improved by averaging over subjects as well as trials (3).

The results provide surprisingly strong evidence of the invari-

ance between subjects of brain-wave representations of language

as first processed upon reaching the cortex. The brain-wave

representations we have studied are based on electroencepha-

lographic (EEG) recordings of auditory activity in the cortex. 

Review of related research is given in the references cited.

Using the methods of analysis developed in our earlier work,

the present study reports the findings of two new experiments

focused on the brain-wave representation of simple visual images

and their names. The images are patches of color or familiar

shapes such as circles and squares. We analyze the representa-

tions of the images and words separately, but our main focus is

on the comparison of the brain waves representing images with

those representing names of the images. The results support in

a quite direct way the solution proposed by Bishop Berkeley and

David Hume in the 18th century, to the long-standing problem of how the mind

represents simple abstract ideas.

Methods

For all subjects, electroencephalographic recordings were made in

our laboratory by using 15 or 22 model-12 Grass Instruments

(Quincy, MA) amplifiers and Neuroscan's SCAN4 software (Ster-

ling, 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. For both 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 experi-

ments, 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 com-

puter screen.

Fourteen subjects were used in the experiments. We numbered

the subjects consecutively with those used in refs. 1–3, because we

continue to apply new methods of analysis to our earlier data.

Subjects S10–19 participated in experiment I, which took place in

January 1999; S10–14, 16, 19, 25, and 28–30 participated in

experiment II, which took place in June and July 1999. S29

participated in two sessions on different days; in the later analysis

of experiment II, each session is counted as a subject, S29.1 and

S29.2. Nine of the subjects were female and five were male, ranging

in age from 23 to 54 years. One was left-handed, one was ambili-

dextrous, and three were not native English speakers.

In experiment I, S10–15 and S19 had the following 16 unipolar

sensors attached to the scalp: Fp1, Fp2, F7, F3, F4, F8, T3, C3,

Cz, C4, T4, T5, P3, Pz, P4, and T6. S16–18 had 22 bipolar pairs:

Cz-Fz, 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

II, all subjects had the following 15 bipolar pairs attached to the

scalp: Cz-C4, Cz-P4, Cz-Pz, Cz-P3, Cz-C3, F4-Fp2, F4-F8,

C4-F8, C4-T4, C4-T6, P4-T6, P3-T5, C3-T5, C3-T3, and F3-F7.

Using the methods of refs. 1–3, averaging half of the trials for

prototypes and the other half for test samples, then a fast Fourier

transform, followed by filtering, and an inverse transform to the

time domain, we estimated four parameters for each subject in

each of the conditions in the two experiments. First, we esti-

mated the low frequency and the high frequency of the optimal

bandpass filter (optimal defined, as in ref. 1, in terms of correct

recognition rate). Second, 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. The parameters s and e are

omitted in the tables of results of experiment I, because quite

often the gradients were too flat to make the selection of s or e

other than arbitrary within a couple of hundred ms. Some
detailed results for s and e are given for experiment II. Some

typical recognition-rate surfaces are shown in refs. 2 and 3.

In both experiments, we followed the methodology of our

earlier article (3) and averaged brain waves over subjects, as well

as subjects of brain-wave representations of simple patches of colors and

simple visual shapes and their names, the words blue, circle, etc. 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.

The most significant results were these. By averaging over differ-

ent subjects, as well as trials, we created prototypes from brain

waves evoked by simple visual images and test samples from brain

waves evoked by auditory or visual words naming the visual

images. We correctly recognized from 60% to 75% of the test-
samples of the images and words separately, but our main focus is

toward the solution proposed for visual shapes, by Bishop Berkeley and David Hume in the 18th century, to the long-standing problem of how the mind represents simple abstract ideas.
as trials, to achieve in many cases, but not all, better recognition results. The notation we use is: AvgUS, averaged over unipolar subjects; TypUS, typical unipolar subject (artificial subject made up of two trials from each of the best five individual unipolar subjects); AvgBS, averaged over bipolar subjects; AvgS, averaged over all subjects; SepAS, separately averaged subjects for prototypes and test samples; and TypS, typical subject (made up from two trials each from best five individual subjects).

**Experiment I: Visual Shapes and Their Names**

**Procedures.** In experiment I, 10 stimuli in each of four conditions were presented to subjects with the same interstimulus interval of 1,550 ms. Each stimulus was presented 10 times in random blocks of 10 trials, each block containing all 10 stimuli. In the two auditory conditions, one with a female voice and one with a male voice, the 10 stimulus words were: circle, square, line, arrow, dog, man, fish, cube, face, star. The duration of each auditory stimulus was about 400 ms. Words spoken by the two speakers were randomized together, for a total of 20 blocks. In the visual-word condition the same 10 words were presented visually for 500 ms on a computer screen. In the visual-image condition, the stimuli were stick drawings on the computer screen representing the 10 words and also were presented visually for 500 ms. The order of presentation to all the subjects was the same, first 200 auditory-word trials, then 100 visual-word trials, and finally 100 visual-image trials.

**Results.** The results for the four conditions are shown in Table 1. The first column of data for each condition in the table shows the recognition rate achieved, expressed in percent. The best EEG sensor, or bipolar pair of sensors, is shown in the second column of data for each condition, and in the third column the optimal bandpass filter in Hz. We note first that the recognition rate of 100%, the highest achieved in this experiment, was for averaged data (AvgUS, S10–15, S19) in the visual-image condition where subjects saw stick drawings of 10 familiar objects. In the visual-word condition the same 10 words were presented visually for 500 ms on a computer screen. In the visual-image condition, the stimuli were stick drawings on the computer screen representing the 10 words and also were presented visually for 500 ms. The order of presentation to all the subjects was the same, first 200 auditory-word trials, then 100 visual-word trials, and finally 100 visual-image trials.

![Figure 1](image_url)

**Fig. 1.** Comparison of brain waves generated by visual images and auditory or visual words, experiment I. (Upper) The averaged, filtered brain waves for the visual image of a circle as prototype (solid curved line) and for the visual word circle as test sample (dotted line). (Lower) The same comparison for the brain waves generated by the visual image of a square (prototype) and the spoken word square (female speaker) as test sample. Because of the slightly slower brain-wave response to the auditory word, the test-sample wave was moved 50 ms to the left. The x axis is measured in ms after the onset of the image or word.
Equally surprising, using the same visual-image brain-wave prototypes, but now using as test samples the average auditory-word brain waves (female voice), we also recognized six of the 10 test samples correctly. We show in Fig. 1 (Lower) how similar the two averaged brain waves from sensor F3 for square are, one generated by the visual image of a square and the other by the spoken word square.

For a comparison of brain waves generated by the visual image of a circle, but in different laboratories, we show in Fig. 2 the averaged wave from sensor T3 along with the T3 wave generated in connection with our earlier experiments (1, 2) conducted at the Scripps Institute of Research (La Jolla, CA). The institute data are averaged from three subjects shown a visual image of a circle. We note that the visual-image generated wave from T6 for a circle-line, line-square, square-triangle, and triangle-circle. The other eight of the pairs were for colors: four pairs of same colors, and four pairs of different colors: blue-yellow, green-blue, red-green and yellow-red. The randomization was restricted so that trials alternated in pair of colors and shapes.

Condition I of each session presented visual images of colors and shapes. For example, the color red was represented by a blank screen with red background, and a square shape was represented by a white line drawing of a square displayed on the screen against a black background. Condition I consisted of 15 blocks. Each block contained the same 16 pairs, randomized in different order in different blocks. Eight of the pairs were for colors: four pairs of same colors, and four pairs of different colors: blue-yellow, green-blue, red-green and yellow-red. The other eight of the pairs were for shapes: four pairs of same shapes, and four pairs of different shapes: circle-line, line-square, square-triangle, and triangle-circle. The randomization was restricted so that trials alternated in pair of colors and pair of shapes.

Condition II of each session presented visual words and auditory words. Instead of visual images, we used auditory words, blue, etc., to represent the colors and visual words, circle, etc., displayed on the screen to represent the shapes. The rest of the experimental setup was the same as in condition I, except that there were only 12 blocks of 16 trials each. Because of the way we represented colors and shapes, trials within each block at the beginning of the session and instructions specific to each condition right before its start. Before condition IV began, the subject also was given two representative examples of the stimuli to be presented.

Every trial in all conditions contained a pair of stimuli, presented in temporal sequence. For each stimulus in the pair, recording started 50 ms before the stimulus onset and lasted until 1,350 ms after the stimulus onset. Each stimulus itself lasted for 200 ms in the nonauditory cases and ranged from 275 ms to 421 ms for auditory stimuli. There was a 100-ms pause within each trial between the recordings of the two stimuli and another 1,100-ms pause after recording for the second stimulus, before the next trial started. After presentation of the second member of a pair, the subject used the numeric pad on the computer keyboard to respond “1” if the two stimuli in the pair were the same and “2” if they were different. Subjects were instructed that the same-different distinction was obvious and did not require a subtle perceptual discrimination. The length of each trial was 4 s in total. Interstimulus interval was 1,500 ms within a trial and 2,500 ms between onset of the second stimulus of a trial and onset of the first stimulus of the next trial. Trials were randomized within each condition. But all subjects were presented, in a given condition, with the same sequence of randomized stimuli.

There were four colors: blue, green, red, and yellow, and four shapes: circle, square, triangle, and line (at 135° angle, bottom to the left), as the contents of the stimuli.
alternated not only in pairs of colors and pairs of shapes, but also pairs of auditory words and pairs of visual words.

Condition III was very similar to condition II, except that auditory words were shape words and visual words were color words.

In condition IV, all pairs were auditory words and all pairs contained the two same words. The possible difference between the two words in a pair was only whether they were said by a female or a male speaker. Subjects were instructed to respond whether the voices in a pair were the same voice or two different voices. Eight pairs of same words (four colors and four shapes) were presented with each of the four possible combinations of speakers (female-female, male-male, female-male, male-female) in each block. Hence, we had six blocks of 32 trials each. As before, randomization within each block was done in such a way that trials alternated in pairs of color words and pairs of shape words.

Results. The results for conditions I, II, and III of experiment II are shown in Table 2. The percent recognition rates shown for both visual images and words are in terms of recognizing the eight visual images, four colors and four shapes, or their visual or auditory names. As in experiment I, for each subject, half of the trials were averaged to create eight brain-wave prototypes and the other half to create eight brain-wave test samples. The results for visual images were the best. Recognition of four visual images, four colors and four shapes, or their visual names, both in condition IV, are shown. The recognition rates for both AvgS and SepAS were 100% and 75% for TypS. The results for the auditory presentation of the eight names of the visual images in conditions II and III are shown in Table 2. The recognition results are in between those for visual images and their printed, i.e., visual, names, but close to the good results for the visual images. The recognition rate of the brain waves for nine subjects was 88% (only one error per subject), and for the remaining three subjects, 75% (two errors). The recognition rates for both AvgS and SepAS were 100% and 75% for TypS. Of the three types of averaging, SepAS, the averaging of separate subjects for prototypes and test samples, was, at 100%, for the auditory presentation, scientifically the most significant. It strongly supports the invariance results across subjects reported in ref. 3.

In Table 3, the recognition results for the brain waves generated by a female-voice and by a male-voice presentation of the eight names, both in condition IV, are shown. The recognition results are not as good as for the auditory presentation of conditions II and III. They are somewhat better for the male voices. Of the three types of averaging, SepAS, the averaging of separate subjects for prototypes and test samples, was, at 100%, for the auditory presentation, scientifically the most significant. It strongly supports the invariance results across subjects reported in ref. 3.

In Table 4, the cross-modality recognition results averaged over subjects, exp. II, are shown. The percent recognition rates shown for both visual images and words are in terms of recognizing the eight visual images, four colors and four shapes, or their visual or auditory names. As in experiment I, for each subject, half of the trials were averaged to create eight brain-wave prototypes and the other half to create eight brain-wave test samples. The results for visual images were the best. Recognition of four visual images, four colors and four shapes, or their visual names, both in condition IV, are shown. The recognition rates for both AvgS and SepAS were 100% and 75% for TypS.

In Table 5, the cross-modality recognition results with separate averaging (SepAS) of prototypes and test samples, exp. II, are shown. The percent recognition rates shown for both visual images and words are in terms of recognizing the eight visual images, four colors and four shapes, or their visual or auditory names. As in experiment I, for each subject, half of the trials were averaged to create eight brain-wave prototypes and the other half to create eight brain-wave test samples. The results for visual images were the best. Recognition of four visual images, four colors and four shapes, or their visual names, both in condition IV, are shown. The recognition rates for both AvgS and SepAS were 100% and 75% for TypS.

### Table 3. Female and male voices for auditory words as stimuli, exp. II

<table>
<thead>
<tr>
<th>Subject</th>
<th>%</th>
<th>Sensor</th>
<th>Filter</th>
<th>Time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>S10</td>
<td>75</td>
<td>P3-T5</td>
<td>3-5</td>
<td>90-550</td>
</tr>
<tr>
<td>S11</td>
<td>63</td>
<td>C4-T6</td>
<td>7-10</td>
<td>210-600</td>
</tr>
<tr>
<td>S12</td>
<td>75</td>
<td>C4-T4</td>
<td>10-14</td>
<td>60-1,250</td>
</tr>
<tr>
<td>S13</td>
<td>63</td>
<td>C4-T4</td>
<td>4-10</td>
<td>60-800</td>
</tr>
<tr>
<td>S14</td>
<td>75</td>
<td>F4-Fp2</td>
<td>2-16</td>
<td>90-650</td>
</tr>
<tr>
<td>S16</td>
<td>63</td>
<td>C4-T6</td>
<td>0.5-9</td>
<td>120-1,150</td>
</tr>
<tr>
<td>S19</td>
<td>75</td>
<td>C4-T4</td>
<td>3-11</td>
<td>300-800</td>
</tr>
<tr>
<td>S25</td>
<td>63</td>
<td>F4-Fp2</td>
<td>6-20</td>
<td>270-1,100</td>
</tr>
<tr>
<td>S28</td>
<td>63</td>
<td>C4-T6</td>
<td>0.5-7</td>
<td>270-1,300</td>
</tr>
<tr>
<td>S29.1</td>
<td>75</td>
<td>F4-Fp2</td>
<td>6-10</td>
<td>30-1,150</td>
</tr>
<tr>
<td>S29.2</td>
<td>63</td>
<td>C4-T6</td>
<td>0.5-8</td>
<td>150-800</td>
</tr>
<tr>
<td>S30</td>
<td>75</td>
<td>C4-T6</td>
<td>3-8</td>
<td>300-650</td>
</tr>
<tr>
<td>AvgS</td>
<td>88</td>
<td>C4-T6</td>
<td>2-8</td>
<td>240-600</td>
</tr>
<tr>
<td>SepAS</td>
<td>88</td>
<td>C4-T6</td>
<td>5-16</td>
<td>300-650</td>
</tr>
<tr>
<td>TypS</td>
<td>63</td>
<td>C4-T6</td>
<td>1-5</td>
<td>180-500</td>
</tr>
</tbody>
</table>

FVoice, female voice; MVoice, male voice.

### Table 4. Cross-modality recognition results averaged over subjects, exp. II

<table>
<thead>
<tr>
<th>%</th>
<th>Sensor</th>
<th>Filter</th>
<th>Time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI-VW</td>
<td>75</td>
<td>P4-T6</td>
<td>0.5-6</td>
</tr>
<tr>
<td>VI-AW</td>
<td>63</td>
<td>C4-T4</td>
<td>10-34</td>
</tr>
<tr>
<td>VI-AWF</td>
<td>75</td>
<td>C4-T6</td>
<td>6-20</td>
</tr>
<tr>
<td>VI-AWM</td>
<td>63</td>
<td>C4-T4</td>
<td>6-20</td>
</tr>
<tr>
<td>VW-AW</td>
<td>63</td>
<td>C4-T6</td>
<td>7-26</td>
</tr>
<tr>
<td>VW-AWF</td>
<td>75</td>
<td>P4-T6</td>
<td>6-21</td>
</tr>
<tr>
<td>VW-AWM</td>
<td>63</td>
<td>C3-T3</td>
<td>3-20</td>
</tr>
</tbody>
</table>

VI, visual image; VW, visual word; AW, auditory word, conditions II and III; AWF, female voice, condition IV; AWM, male voice, condition IV.

### Table 5. Cross-modality recognition results with separate averaging (SepAS) of prototypes and test samples, exp. II

<table>
<thead>
<tr>
<th>%</th>
<th>Sensor</th>
<th>Filter</th>
<th>Time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1-VW2</td>
<td>75</td>
<td>F3-F7</td>
<td>9-20</td>
</tr>
<tr>
<td>V2-VW1</td>
<td>63</td>
<td>Cz-C4</td>
<td>10-22</td>
</tr>
<tr>
<td>V1-AW2</td>
<td>75</td>
<td>C4-T4</td>
<td>6-14</td>
</tr>
<tr>
<td>V2-AW1</td>
<td>63</td>
<td>C3-T3</td>
<td>7-19</td>
</tr>
<tr>
<td>V1-AWF2</td>
<td>75</td>
<td>C3-T3</td>
<td>10-23</td>
</tr>
<tr>
<td>V2-AWF1</td>
<td>63</td>
<td>P4-T6</td>
<td>6-12</td>
</tr>
<tr>
<td>V1-AWM2</td>
<td>63</td>
<td>C4-T6</td>
<td>4-10</td>
</tr>
<tr>
<td>V2-AWM1</td>
<td>63</td>
<td>C4-T4</td>
<td>6-10</td>
</tr>
<tr>
<td>VW1-AW1</td>
<td>75</td>
<td>C3-T3</td>
<td>10-20</td>
</tr>
<tr>
<td>VW2-AW1</td>
<td>63</td>
<td>C3-T3</td>
<td>3-20</td>
</tr>
<tr>
<td>VW1-AWF2</td>
<td>63</td>
<td>F4-F8</td>
<td>2-6</td>
</tr>
<tr>
<td>VW2-AWF1</td>
<td>63</td>
<td>P4-T6</td>
<td>6-20</td>
</tr>
<tr>
<td>VW1-AWM2</td>
<td>75</td>
<td>C4-T4</td>
<td>2-15</td>
</tr>
<tr>
<td>VW2-AWM1</td>
<td>88</td>
<td>F4-F8</td>
<td>0.5-20</td>
</tr>
</tbody>
</table>

VI, visual image; VW, visual word; AW, auditory word; M, male; F, female.
voice than the female voice, perhaps because the male speaker’s voice was also the one heard in conditions II and III. We do not review in detail the results for individual subjects, which are shown in Table 3, but we note that for both voices, AvgS and SepAS were the same and at a good level of recognition, 88% (one error) for each of the four cases in the table.

We now turn to the two tables that summarize the results of experiment II that are most directly relevant to the focus of this article, invariant brain waves for visual images and their names. Table 4 summarizes the results when data for all the subjects were averaged together for each condition. So, for example, the first row of Table 4 is based on the averaged EEG brain-wave data for visual images to form eight prototypes and the corresponding data for visual words to form eight test samples. In the rest of Table 4, as in this example, the condition used for forming the prototypes is given first, and the condition for the test samples second. As can be seen, and as would be expected, the recognition results are not quite as good as those found in Tables 2 and 3, but are comparable to those for the visual-word condition alone in Table 2. More directly relevant is the fact that the cross-modality results are generally better than those obtained in experiment I. In particular, all 14 recognition percentages of Table 4 are better than the two rates of 60% reported for experiment I.

In Table 5, we show the SepAS cross-modality results for experiment II, with six subjects being used for the prototypes and a different six for the test samples. To give a complete analysis, we ran a given six subjects’ averaged brain waves as prototypes, for instance, from the visual-image condition, labeled in the first row of Table 5, VI1, vs. the other six subjects’ averaged brain waves, from the visual-word condition, labeled VW2, as test samples. The second row reverses the six subjects’ role of prototype and test sample, so that now the “other” six subjects’ EEG data, from the visual-image condition, labeled VI2, form the prototypes. The 14 rows of Table 5 give the complete set of cross-modal analyses, with the prototypes and test samples always averaged (SepAS) over disjoint sets of six subjects.

For the eight cases in Table 5 of visual images as prototypes, three of them are at a recognition rate of 75% (two errors) and the remaining five at 63% (three errors). These results are comparable to those in experiment I. In Fig. 3 we show six pairs of waves from the VII-AW2 condition, four for the colors blue, green, red, yellow, and two for the shapes line and triangle. Each pair consists of the average of brain waves generated by a visual image (solid line) and the average of brain waves generated by the corresponding auditory word (dotted line). Of the six pairs shown, only the pair for the image and word red was misclassified, i.e., not correctly recognized. The fits of the six pairs are not perfect, as is also the case for Figs. 1 and 2. On the other hand, and this is the point to be emphasized, the fits reflect five of six correct recognitions for the waves shown. Only the case of red was misclassified. Moreover, the quantitative least-squares measure of fit is actually lower, and therefore better, for the VII-AW2 cross-modality case, with separate subjects for

![Fig. 3. Comparison of averaged and filtered brain waves generated by visual images and auditory words, experiment II. The six brain-wave prototypes (solid curved lines) were generated by the four color stimuli, named in the top four panels, and by the two visual shapes line and triangle in the bottom two panels. The test samples (dotted lines) were generated by the spoken names of the six visual stimuli averaged over both speakers. No translations along the x axis to improve the fit were made. The x axis is measured in ms after the onset of the visual image or its spoken name.](www.pnas.org)
views are well supported by our results. After visual display of a patch of red or of a circle, the image is represented in the cortex by the brain wave of the word red or circle within a few hundred ms of the display and somewhat quicker than is the representation in the cortex of the spoken word red or circle. To the skeptical response that we do not really know it is the word red or circle that is being represented in the cortex, as opposed to the particular visual image, we respond that everything we have learned thus far about the one-dimensional temporal representation of words, presented either auditorily or visually, supports our inference, the spatial unidimensionality of the temporal representation used for recognition, above all. Perhaps just as important, the filtered brain waves representing the spoken color or shape words conform closely to the brain waves of the many other words whose brain waves we have identified in our earlier work. However, we emphasize, as we did in ref. 3, that the invariance we are observing between brain-wave representations of visual images and words is consistent with the existence of other significant information we have averaged and filtered out.

**Related Psychological Studies.** Various related psychological studies of memory support our conclusion as well. For example, when words are presented visually for immediate recall, the errors tend to be acoustic in character (7), or, if for longer storage, an auditory representation is used (8). More detailed results and a survey of many relevant experiments on the primacy of the auditory representation of words in memory are to be found in ref. 9. There is much evidence that the memory of purely visual images decays quickly, almost always less than 200 ms (10, 11), even though years of research have generated a lot of controversial results (12, 13) on visual sensory memory. On the other hand, the field seems to have reached some consensus that visual sensory memory consists of several components: those that can be masked and decay within 100–300 ms; and a limited-capacity, longer-lasting short-term memory (14–18), with some authors (12, 14, 18) attributing the limited capacity short-term memory to verbal memory. In contrast, short-term auditory memory lasts 2–5 sec (19, 20), so it is most efficient to represent simple visual images in memory by the auditory representation of their names or simple descriptions. The brain-wave experiments reported here support in an unusually direct way that this is indeed what Berkeley and Hume conjectured long ago, but for different reasons than the brevity of visual memory.

We thank David Spiegel, Department of Psychiatry, Stanford Medical School, for loan of the Grass amplifiers used in experiment I and Adrian Raine for loan of the Grass amplifiers used in experiment II. We thank Paul D'Imperio for producing the three figures and Ann Gunderson for preparation of the manuscript. We also received useful comments and suggestions for revision from George Sperling.

---

**Table 6. Comparison of least-squares fit for three conditions, exp. II.**

<table>
<thead>
<tr>
<th></th>
<th>Cross-modal</th>
<th>Auditory</th>
<th>Visual</th>
</tr>
</thead>
<tbody>
<tr>
<td>V11-AW2</td>
<td>39.2</td>
<td>292.9</td>
<td>97.0</td>
</tr>
<tr>
<td>Line</td>
<td>40.1</td>
<td>288.0</td>
<td>124.6</td>
</tr>
<tr>
<td>Triangle</td>
<td>31.3</td>
<td>212.2</td>
<td>71.0</td>
</tr>
<tr>
<td>Square</td>
<td>63.2*</td>
<td>158.9</td>
<td>80.1</td>
</tr>
<tr>
<td>Red</td>
<td>56.1*</td>
<td>67.7</td>
<td>113.1</td>
</tr>
<tr>
<td>Green</td>
<td>24.7</td>
<td>394.0</td>
<td>113.7*</td>
</tr>
<tr>
<td>Blue</td>
<td>24.8</td>
<td>260.6</td>
<td>121.3</td>
</tr>
<tr>
<td>Yellow</td>
<td>33.9</td>
<td>63.9</td>
<td>124.8</td>
</tr>
</tbody>
</table>

VI, visual image; AW, auditory word.

*Misclassification of brain waves.

---