D. Gabor

Associative Holographic Memories

Abstract: Recently Longuet-Higgins modeled a temporal analogue of the property of holograms that allows a complete image to be constructed from only a portion of the hologram. In the present paper a more general analogue is discussed and two two-step transformations that imitate the recording-reconstruction sequence in holography are presented. The first transformation models the recall of an entire sequence from a fragment while the second is more like human memory in that it provides recall of only the part of the sequence that follows the keying fragment. Both models require only the three operations: shift, multiplication and addition.

This investigation was inspired by a note of H. C. Longuet-Higgins[1] on a "Holographic Model of Temporal Recall." Prof. Longuet-Higgins asked whether the interesting property of holograms that allows reconstructing a whole picture from a fragment of the original could be reproduced in the temporal domain. He answered his question by describing a dynamic model. This model consists of a series of narrow-band-pass filters that cover the frequency range of the sequence to be recalled. Each filter has an adjustable gain that is made proportional to the work done on it or, in other words, to the energy in its pass-band. The series of filters is thus changed into a "matched filter" which, when excited by a fragment of a sequence, will automatically continue the sequence to its recorded end.

This is indeed a temporal analogue of holography, though a somewhat restricted one, because holography allows not only reconstruction of the original picture from a fragment, but also reconstruction of another picture that has been associated with it in the recording. This leads to the somewhat more general question: By what two-step transformations can we imitate the properties of the recording-reconstruction sequence in holography? The answer is, in brief, that there must exist an infinite family of such paired transformations, but that there is one which recommends itself by its simplicity and this is the one I want to explain in somewhat more detail than I have done in two previous short notes.[2,3]

Consider first in a general way the *total* transformation achieved in holography from the recording to the reconstruction. In Fig. 1a there is a first plane x, y in which we have two complex wave amplitudes $a(x, y)$ and $b(x, y)$. We record the two together on a photographic plate in a second plane u, v where the total complex amplitude is $A(u, v) + B(u, v)$. Let the distance between the two planes be L. This arrangement is completely general in the usual Gaussian approximation of holographic theory. If there is some optical system between the two planes, we need only replace the distance L by the radius of curvature R of the spherical waves that issue from a point in the first plane.

The amplitude $A(u,v)$ is expressed by the Fresnel-Kirchhoff diffraction integral[4]; after the familiar small-angle and large-L approximations[4,5] are made and some irrelevant factors are dropped, this amplitude can be written as

$$A(u, v) = e^{\pi i(u^2+v^2)/\lambda L} \times \iint a'(x, y) e^{-2\pi i(xu+yv)/\lambda L}\, dx\, dy, \qquad (1)$$

where

$$a'(x, y) = a(x, y) e^{\pi i(x^2+y^2)/\lambda L}. \qquad (2)$$

A similar result describes the amplitude $B(u, v)$. The expression for A is the Fresnel transform of $a(x, y)$ and it differs from a Fourier transform only in having two Gaussian factors. Their effect is the same as if we had taken a Fourier transform, but with two thin lenses inserted, both with focal length L, each covering one plane.

When we make a hologram, we record $|A + B|^2$ and the amplitude transmission of the plate contains an interference term $A\bar{B} + \bar{A}B$, where the bar denotes complex conjugate. Of these, the first is the "twin wave," which does not interest us in the present context. It is the

The author is currently Staff Scientist at the CBS Laboratories, High Ridge Road, Stamford, Connecticut 06905 and Visiting Professor at the State University of New York, Stony Brook, Long Island, New York 11790.

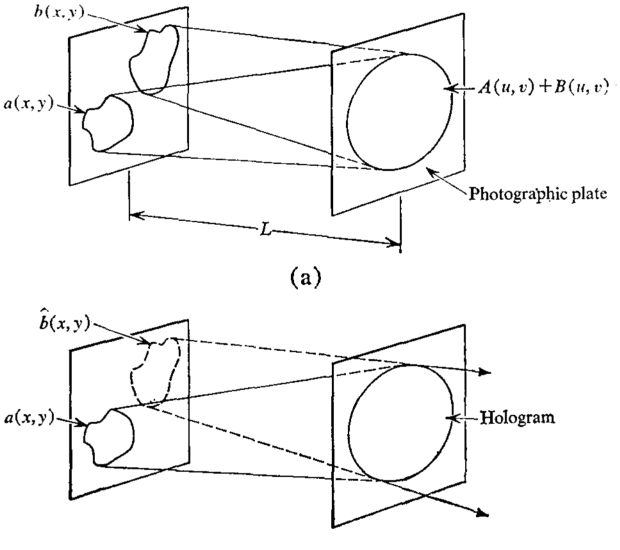


Figure 1 A complete holographic transformation: (a) recording; (b) reconstruction.

second term that, on illumination with the wave A, will give the reconstructed wave

$$A\bar{A}B = e^{\pi i(u^2+v^2)/\lambda L}\iint a' e^{-2\pi i(xu+yv)/\lambda L}\,dx\,dy$$

$$\times \iint \bar{a}' e^{2\pi i(xu+yv)/\lambda L}\,dx\,dy$$

$$\times \iint b' e^{-2\pi i(xu+yv)/\lambda L}\,dx\,dy. \qquad (3)$$

We can transform the product of the first two integral factors into the Fourier transform of an autocorrelogram. For simplicity, this is shown in one dimension. We suppress the y variables in Eq. (3), label the x variables as x, x' and x'' and make the substitutions $x' = x - \xi$ and $x'' = \eta - \xi$. The first two factors in Eq. (3) now appear as

$$\iint a'(x)\bar{a}'(x')e^{-2\pi iu(x-x')/\lambda L}\,dx\,dx'$$

$$= \int\left[\int a'(x)\bar{a}'(x-\xi)\,dx\right]e^{-2\pi iu\xi/\lambda L}\,d\xi$$

$$= \int a' \circledast \bar{a}' e^{-2\pi iu\xi/\lambda L}\,d\xi. \qquad (4)$$

The extension of Eq. (4) to include the third integral factor in Eq. (3) is

$$\int\left[\int (a' \circledast \bar{a}')b'(\eta - \xi)\,d\xi\right]e^{-2\pi iu\eta/\lambda L}\,d\eta$$

$$= \int [(a' \circledast \bar{a}') * b']e^{-2\pi iu\eta/\lambda L}\,d\eta. \qquad (5)$$

We can now interpret the general result $A\bar{A}B$ as the Fresnel transform of the convolution of $b(x, y)$ with the autocorrelogram of $a'(x, y)$. That is to say, the reconstruction $\hat{b}(x, y)$ (see Fig. 1b) differs from $b(x, y)$ by being redrawn, instead of with a delta-function, with a spread function that is the autocorrelogram not of the "key wave" $a(x, y)$ but of its modification $a'(x, y)$.

This theorem is usually proved for Fourier holography only and we may note in passing that the modification is not without interest. For instance, if a is a "slab" of width D, its autocorrelogram is a triangle of width $2D$. With the modifier [see Eq. (2)], however, the autocorrelogram in the slab case becomes

$$e^{-\pi i\xi^2/\lambda L}\,\frac{\sin[\pi(D-\xi)\xi/\lambda L]}{\pi\xi/\lambda L}.$$

This function has its first zero at $\xi_0 = \lambda L/D$ instead of at D; hence, if $\lambda L/D^2$ is a small number, there can be a gain in resolution. In other words Fresnel holograms *can* be less sensitive to the key wave than Fourier holograms. This can be an advantage in some applications, a disadvantage in others.

Disregarding now the difference between a and a' we see that the result of a complete holographic transformation is a convolution of the associated signal b with the autocorrelogram of the key wave a:

$$\hat{b} = b * (a \circledast \bar{a}) \qquad (6)$$

Whenever we have a key wave or key figure a which correlates sharply with itself, $\hat{b}$ will not differ substantially from b. This will be the case for all "noise-like" functions without a dc term and also for all figures composed of thin lines. Since written and printed letters consist of thin lines, this is the basis for character recognition by holography. Moreover, *parts* of such noise-like or complicated figures also correlate sharply with themselves and this is the reason why, in holography, a part is sufficient to recall the whole.†

A first example of such a two-step transformation was a succession of two *convolutions*[2]: To make a record, convolve the key sequence A with B. To recall B, convolve the record with A', where A' is a fragment of A. I now prefer a second method, which is a succession of two *correlations*.[3]

Let the sequences in question run from $t = 0$ to $t = T$, or at least be contained in this interval. In the *recording*

† For some years now this property of holograms has attracted the interest of neurophysiologists who were puzzled by the difficulty of locating the "engram" in the human or animal memory. As is well known, especially since the famous experiments of Lashley,[6] large parts of the brain can be destroyed without wiping out a learned pattern of behaviour. This has led to speculation that the brain may contain a holographic mechanism. (See, in particular, Ref. 7.) For my part I am inclined to believe that there exists an abstract, mathematical similarity, but I am rather skeptical regarding the existence of waves or even of tuned resonators in the brain, especially after having found that the transformation by Eq. (6) can be carried out without postulating any such intermediaries.

we form the cross-correlation

$$\phi(t') = \int_{t'}^{T} B(\tau)\, A(\tau - t')\, d\tau. \tag{7}$$

The integration limits are of no great importance, since $A(\tau - t')$ is zero earlier than $\tau = t'$ and $B(\tau)$ is zero after $\tau = T$. We can therefore replace these limits by wider ones.

In the *recall* we cross-correlate the key sequence A' with the record, and form the recall function

$$R(t) = \int_{0}^{t} A'(t - t')\phi(t')\, dt'. \tag{8}$$

Again the limits can be replaced by wider ones because A' does not start before zero and $\phi(t')$ does not extend beyond $t' = T$.

Substituting (7) into (8), extending the integration limits, and introducing the new variables $\xi = t' - \tau$ and $x = t - t'$, we obtain

$$\begin{aligned} R(t) &= \iint B(\tau)\, A(\tau - t')\, A'(t - t')\, d\tau\, dt' \\ &= \int B(t - \xi)\left[\int A(x - \xi)\, A'(x)\, dx\right] d\xi \\ &= B * (A \circledast A'), \end{aligned} \tag{9}$$

the convolution of B with the cross-correlation of A and A'. If now $A'(t) = A(t + t_0)$, i.e., the original key delayed by t_0, or at least a sufficiently long fragment of it, the cross-correlation $(A \circledast A')$ becomes a delta-like function which restores B, so that the recall function approximates $B(t + t_0)$.

These operations can be visualized by means of Fig. 2. They correspond to an integration over the time domain delineated by the lower trapezoid (the second model) of the triple product

$$B(\tau)\, A(\tau - t')\, A'(t - t')$$

for the case in which A' is only a fragment of A and starts at t_0. Imagine that the bar of the reading time t is first lowered to $t' = 0$ and then raised. It can be seen that from the instant at which the remembered fragment comes to an end, the recall mechanism is acting with full strength, recalling $B(t + t_0)$ with a constant factor, that is to say, with *full fidelity*.

Figure 2 shows also, in the upper trapezoid, the first model, which is a succession of two convolutions. I now prefer the second model, which is more "brain-like," in that it allows one to recall only the part of A that follows the remembered fragment A' and not the part before it. This is a well-known weakness of our memory, in particular when something has been memorized "by rote." If it is a matter of constructing an artificial memory, the

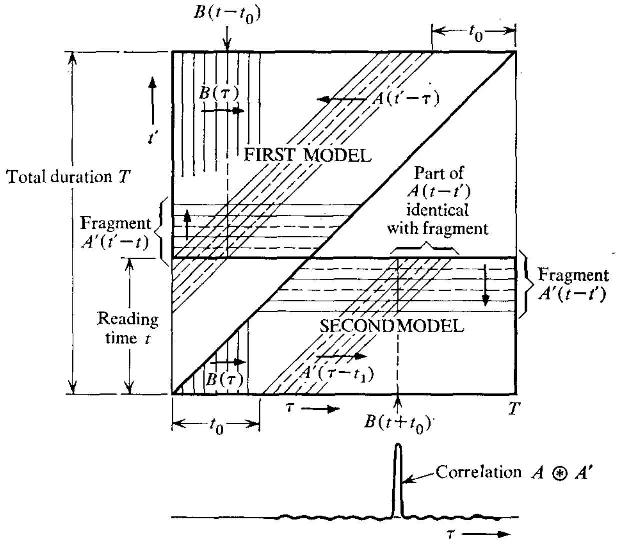


Figure 2 Holographic models of an associative memory. The first model, represented in the upper trapezoid, is two convolutions in series and the second model, in the lower trapezoid, is two correlations in series.

first model may well be preferable because it allows one to recall the whole sequence, wherever the fragment A' may be located.

The recording by Eq. (7) and the recall by Eq. (8) can be carried out by any system that can perform the three operations: shift, multiplication, and summation. There is no difficulty in setting them up on any digital computer. In principle it must be possible to realize them with McCulloch-Pitts neurons. It would be desirable for experts in neural networks to work out concrete realizations and to look for evidence of such structures in the nervous system. As I have insufficient knowledge in this field, I prefer to illustrate the operations with an optical model, shown in Fig. 3.

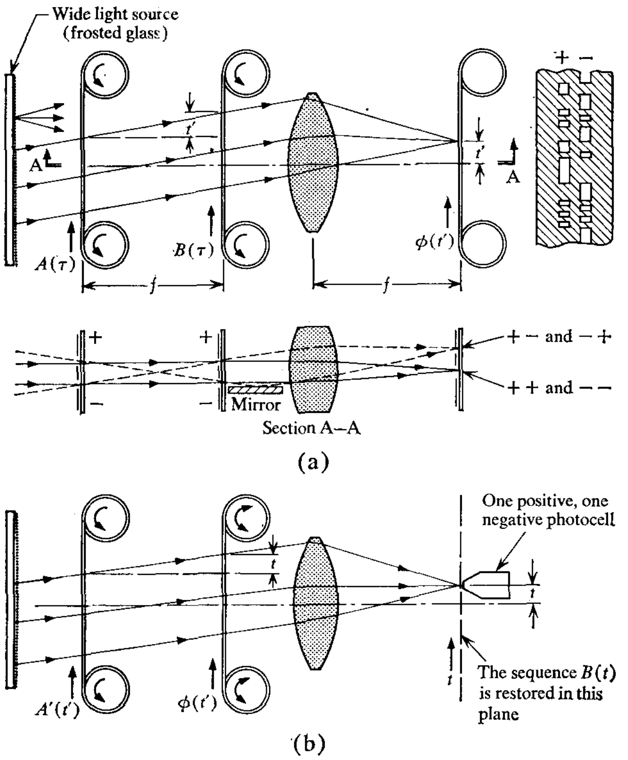
The simplest type of signal that is suitable for discrimination and recall by correlation or convolution methods is a binary sequence of $+1$ and -1, of equal occurrence in the mean. It is also convenient, but not necessary, to have a zero level to ensure that pauses do not give correlation.

In the optical model, which operates with incoherent light and a broad, diffuse source, it is necessary to arrange the plus and the minus parts of the sequences on two separate tracks, since there is no "negative light." If there is no zero level, there is a logical redundancy because the absence of a plus indicates a minus; hence the two tracks could be dispensed with in more sophisticated arrangements in which the dc level is subtracted automatically. Both the binary ($\pm$) and the ternary ($\pm$ and 0) codes for the associated signals A and B can be realized by means of black-or-white film, or even more simply by means of punched tape. In the optical correlator shown in Fig. 3a the truth-table of multiplication is realized by means of a mirror that throws the $+-$ combinations into the same track of the ϕ record as the $-+$ combinations, while the $++$ and $--$ combinations go automatically into the other track. To realize Eq. (8) one requires a recording film with a transmittance proportional to the light exposure.

The recall apparatus shown in Fig. 3b is essentially the same as the recorder. The read-out can be accomplished, for instance, with two photocells in opposition, one on the positive, the other on the negative track.

The discriminatory power of this process is remarkably, even surprisingly, high when applied to sequences of such length that statistical considerations can be applied. For example, assume that the correlator handles N bits at a time, while the fragment A' has $n = xN$ bits. Referring again to Fig. 2, we see that at the recall line (drawn with a thick trace) the recall signal has an amplitude $\pm xN$, plus

Figure 3 An associative memory with optical (incoherent) correlation. (a) A sequence $A(\tau)$ is to be associated with $B(\tau)$. Their cross-correlogram $\phi(t')$ is recorded on the third film strip. (b) The sequence $A'(t')$, which need not be more than a fragment of $A(t)$, is correlated with $\phi(t')$ and the sequence $B(t)$ is reconstructed. Note that in this system the negative track contains no information that is not already contained in the positive track. The two tracks are used only for convenience, since there is no "negative light."



some noise arising from contributions outside this line. Let us compare this with the signal arising from uncorrelated sequences. If we assume that N is a sufficiently large number, the probability of an excess of one sign over the other will have a Gaussian distribution. The cross-correlation between two such sequences is the excess of agreements over disagreements in the two sequences and it also has a Gaussian distribution with a probability for a value n of

$$p(n) = \sqrt{2/\pi N}\,\exp\left(-n^2/2N\right). \qquad (10)$$

The mean square of the spurious signal is $\langle n^2\rangle = N$; hence random sequences of length N will give spurious signals of the order of $N^{\frac{1}{2}}$. Comparing this value with the true signal n, one might think that this result does not give much safety when the fraction x is equal to or less than $N^{-\frac{1}{2}}$. However, the *true* signal of length $n = xN$ gives not one, but xN spikes, all *consistently* of value $\pm\, xN$, and the probability of such a sequence occurring by chance is

$$\left[1 - \operatorname{erf}\left(x\sqrt{N/2}\right)\right]^{xN}. \qquad (11)$$

This can be very small, even if x is of the order of $N^{-\frac{1}{2}}$. For instance, if $N = 100$ and $x = N^{-\frac{1}{2}} = 0.1$, the probability of 10 spurious spikes of value $\pm$ 10 is only about 10^{-5}.

The power of this system for recognizing short fragments of coded sequences can not only be good, but it can be too good. One may have to take precautions lest a large parallel store, on being given *one* word, offer the user *all* the long sentences that contain that word—like the *Thesaurus Linguae Latinae*,[8] for example.

Finally, I want to emphasize again that I do not suggest that processes such as those described here *are* present in the human or animal nervous system. I contend only that on present evidence the possibility cannot be excluded and that the hypothesis of their existence deserves careful examination.

References

1. H. C. Longuet-Higgins, *Nature* **217**, 104 (1968).
2. D. Gabor, *Nature* **217**, 584 (1968).
3. D. Gabor, *Nature* **217**, 1288 (1968).
4. M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, Inc., New York, 3rd ed., pp. 382 to 386 (1965).
5. J. A. Armstrong, *IBM J. Res. Develop.* **9**, 171 (1965).
6. K. S. Lashley, *Brain Mechanisms and Intelligence*, University of Chicago Press, Chicago (1929).
7. P. R. Westlake, "Towards a Theory of Brain Functioning: A Detailed Investigation of the Possibilities of Neural Holographic Processes," Doctoral Dissertation, Univ. of California (Los Angeles), 1968.
8. *Thesaurus Linguae Latinae*, B. G. Teubner Verlagsgesellschaft, Leipzig (1900 to 1963; in progress); the publishers of this dictionary, in Latin, plan to record, with representative quotations from each author, every word in the text of each Latin author down to the Antonines with a selection of important passages from the works of all writers to the seventh century.

Received October 8, 1968