Separatum

EXPERIENTIA VOL. VII/4, 1951 - p. 128

BIRKHÄUSER PUBLISHERS, BASEL/SWITZERLAND

B Delgutte

A Duplex Theory of Pitch Perception[1]

By J. C. R. LICKLIDER[2], Cambridge, Mass.

Theories of pitch perception have shared the presupposition that pitch is a unitary attribute of auditory experience. It is by no means entirely clear, however, that such is the case. In some musical circles, in the older psychological literature[3], and in recent papers on absolute pitch[4], pitch is held to be duplex in nature. Two pitch-like qualities are distinguished. They are given various pairs of names by various authors: tone height and tone chroma, ordinary pitch and chroma, pitch and quality, pitch and tonality, etc. If such a distinction is warranted—and several considerations suggest that it is—the part of auditory theory that concerns the perception of pitch is in need of modification.

The stimulus basis for pitch is also duplex: On the one hand we have frequency, on the other hand, periodicity. That frequency and period are reciprocally related is not sufficient reason for throwing one away and examining only the other, for with each one is associated a method of analysis. Of the two methods, one—frequency analysis performed by an array of band-pass filters—has been incorporated into auditory theory. The cochlea is almost universally regarded as being an extended wave filter that distributes oscillations of different frequencies to different places. The possibility that the other method, autocorrelational analysis, plays a role in the auditory process has been neglected[5].

Autocorrelational analysis is an analysis, carried out entirely within the time domain, that yields the same information as the power spectrum which is obtained through analysis in the frequency domain. WIENER's famous theorum[6] shows that the autocorrelation function and the power spectrum of a wave are a FOURIER transform pair. The attractiveness of autocorrelational analysis *per se* therefore lies not in revealing anything that cannot be found through frequency analysis; it lies in the fact that the operations involved in carrying out the autocorrelational analysis are quite different from those involved in making the frequency analysis.

[1] The preparation of this paper was supported by a contract between Massachusetts Institute of Technology and the Air Force Research Laboratories, Cambridge, Mass.

[2] Acoustics Laboratory, Massachusetts Institute of Technology Cambridge, Massachusetts.

[3] G. RÉVÉSZ, *Zur Grundlegung der Tonpsychologie* (Viet & Co., Leipzig 1913). – MAX F. MEYER, Psychol. Bull. *11*, 349 (1914).

[4] A. BACHEM, J. Acoust. Soc. Amer. *9*, 146 (1937).

[5] The suggestion has been made by R. M. FANO of the Research Laboratory of Electronics and by L. A. DE ROSA of the Federal Telecommunication Laboratories, Inc., that the cochlea may operate more nearly as an autocorrelator than as a filter. This suggestion is quite different from, and in fact quite contrary to the hypothesis proposed in the present paper.

[6] N. WIENER, Acta Math. *55*, 117 (1930).

The essence of the duplex theory of pitch perception is that the auditory system employs both frequency analysis and autocorrelational analysis. The frequency analysis is performed by the cochlea, the autocorrelational analysis by the neural part of the system. The latter is therefore an analysis not of the acoustic stimulus itself but of the trains of nerve impulses into which the action of the cochlea transforms the stimulus. This point is important because the highly nonlinear process of neural excitation intervenes between the two analyses.

The neural mechanism of analysis

In so far as the frequency analysis is concerned, the duplex theory follows the resonance-place theory of HELMHOLTZ[1] and the space-time pattern theory of FLETCHER[2]. If we designate the lengthwise dimension of the uncoiled cochlea as the x-dimension, we can describe the cochlear frequency analysis by saying that the cochlea transforms the stimulus time function $f(t)$ into a running spectrum $F(t, x)$, position x being the neural correlate of stimulus frequency. Thus the cochlea does, in a rather different way and perhaps with less resolution in frequency, essentially the same thing as the Sound Spectrograph developed by the Bell Telephone Laboratories[3]. The running spectrum, a spatial array of time functions, is transmitted brainward by neurons of the auditory nerve. Neurons terminating at x_b near the base of the cochlea act as a group to carry $F(t, x_b)$, while those terminating at x_a near the apex carry $F(t, x_a)$, and the others in between handle other parts. Each $F(t, x_i)$ is an integral over the behaviors of many neurons. The contribution of an individual neuron, say tha jth neuron in group i, is $N_{ij}(t)$, a function that has either the value 0 (quiescent) or 1 (firing).

It will facilitate the description of the autocorrelational analysis if we think of it, at first, as being performed upon the individual functions $N_{ij}(t)$. The running autocorrelation function of $N_{ij}(t)$ is defined[4] (see appendix) as

[1] H. VON HELMHOLTZ, *Sensations of tone* (English translation by A. J. ELLIS, London, 1895, of *Die Lehre von den Tonempfindungen*, 2nd English ed., Longmans, Green, & Co., London, 1885).

[2] H. FLETCHER, J. Acoust. Soc. Amer. *1*, 311 (1930).

[3] J. C. STEINBERG and N. R. FRENCH, J. Acoust. Soc. Amer. *9*, 146 (1946); and a series of articles by members of the Bell Telephone Laboratories in the same issue. Also R. K. POTTER, G. A. KOPP, and HARRIET C. GREEN, *Visible speech* (D. Van Nostrand Co., Inc., New York, 1947).

[4] R. M. FANO, J. Acoust. Soc. Amer. (in press). Also K. N. STEVENS, J. Acoust. Soc. Amer. (in press). These papers are, respectively, on the mathematical relations between running autocorrelation functions and power spectra and on autocorrelation functions of speech sounds.

$$\varphi_{ij}(t, \tau) = \overline{N_{ij}(t)\, N_{ij}\,(t-\tau)} \qquad (1)$$

in which τ is the variable interval by which $N_{ij}(t)$ is delayed to produce $N_{ij}\,(t-\tau)$ and the overline designates a running integral over the more or less recent past of t. Therefore $\varphi_{ij}(t, \tau)$ is simply a running accumulation of recent values of the product of $N_{ij}(t)$ and the same function delayed by τ. It provides a progressive description of the periodicity of the discharges of neuron ij.

The nervous system is nicely set up to perform running autocorrelational analysis. The operations specified by expression (1) are to delay the input function $N_{ij}(t)$ by a variable interval τ, to multiply the delayed function $N_{ij}\,(t-\tau)$ by the original function $N_{ij}(t)$, and to determine the running integral $\varphi_{ij}(t, \tau)$ of the product. A chain of neurons makes an excellent delay line. The spatial aspect of synaptic summation provides approximate multiplication. And the temporal aspect of synaptic summation gives us running integration. The theory postulates, therefore, that the lower centers of the auditory system perform the three operations.

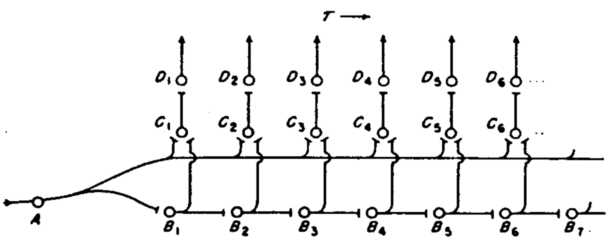

Fig. 1. – *Basic schema of neuronal autocorrelator.* A is the input neuron, B_1, B_2, B_3, ... is a delay chain. The original signal and the delayed signal are multiplied when A and B_k feed C_k, and a running integral of the product is obtained at the synapse between C_k and D_k, where excitation accumulates whenever C_k discharges and dissipates itself at a rate proportional to the amount accumulated. Since these operations correspond to the definition of running autocorrelation, the excitatory states at D_1, D_2, D_3, ... provide a display of the running autocorrelation function of the input time function, the temporal course of the discharges of A.

The basic neuronal connections are shown in Fig. 1. The state of neuron A is $N_A(t)$. If each synaptic delay in the chain B is $\Delta\tau$, the chain gives us $N_A(t)$ under various delays. The state of neuron B_k, for example, is $N_A\,(t - k\Delta\tau)$. Assuming, with MCCULLOCH and PITTS[1], that both A and B_k must fire almost simultaneously to make C_k fire, we have

$$N_{C_k}\,(t) = N_A(t)\, N_A\,(t - k\Delta\tau). \qquad (2)$$

Neuron C_k impinges upon neuron D_k. Excitation is built up at D_k by the discharges of C_k and dissipates itself spontaneously, perhaps at a rate proportional to the amount accumulated. The excitation at D_k at a particular instant t is therefore

$$E_{D_k}(t) = \overline{N_{C_k}(t)}. \qquad (3)$$

Substituting in (3) the value of $N_{C_k}\,(t)$ from (2), we obtain

$$E_{D_k}(t) = E(t, k\Delta\tau) = \overline{N_A(t)\, N_A\,(t - k\Delta\tau)}. \qquad (4)$$

Since expression (4) is equivalent to the definition (1) in every respect save for the substitution of $k\Delta\tau$ for τ, $E(t, k\Delta\tau)$ is an approximation of the running autocorrelation function of $N_A(t)$. Viewing the arrangement of Fig. 1 as a neuronal autocorrelator, we therefore identify the lengthwise dimension of the delay chain with τ, and we regard the time-varying excitation at D_1, D_2, .., D_k.., as a spatial display of successive cross-sections of the autocorrelation function[1].

We must now take into account the facts that sensory systems have many neurons roughly in parallel and that the important physiological quantities appear to be averages or integrals over sets of neurons. We therefore think of neuron A as but one of many neighbors; it is, let us say, the jth neuron in the ith group terminating along the basilar membrane of the cochlea. It is the behavior $F(t, x_i)$ of the group, not $N_{ij}(t) = N_A(t)$ of the individual neuron, that is subjected to autocorrelational analysis. The arrangement of Fig. 1 must be modified by introducing many neurons in parallel with A. Each has its own delay chain, but the outputs of several delay chains are fed to C_1, C_2, .., C_k.. Finally, many sets like C_1, C_2, .., C_k. impinge upon D_1, D_2, .., D_k.. We substitute for the assumption that both A and B_k must discharge to fire C_k the more plausible assumption that the accumulation of excitation at C_k must reach a certain level, which may vary secularly. Other neurons in parallel with C_k have other thresholds. These are distributed in such a way that the behavior of the group of neurons in parallel with C_k is related nonlinearly to the input, which is approximately $F(t, x_i)$ plus $F\,(t - k\Delta\tau,\ x_i)$. The nonlinearity gives rise to an output that is a rough approximation of the arithmetic product of the two components of the input.

In order to obtain the running integral of the product, we follow in principle the description given in the preceding paragraphs. However, as soon as we substitute groups of neurons for individuals, we may redefine the output of the neuronal autocorrelator in terms of the actual behavior of groups of neurons, which is more directly observable than the excitatory state defined as the output in expression (4). To do this, we need only assume that the behavior of the

[1] W. S. MCCULLOCH and W. PITTS, Bull. Math. Biophys. *5*, 115 (1943).

[1] For the sake of simplicity, we assume here that delay in time is proportional to distance traversed in the neural tissue. If the tissue should prove to be non-homogeneous, we should designate the lengthwise dimension of the delay chain as $y = y(\tau)$. This would parallel our procedure in handling frequency, which is not transformed linearly into the x-dimension: we labeled the spatial dimension $x = x(\omega)$ instead of ω itself.

group of neurons in parallel with D_k is roughly proportional to E_{D_k}. As soon as we take that step, however, we see that the chain of neurons $D_1, D_2, .., D_k$. (introduced in the first place only to facilitate the description by separating the operations) can be eliminated. We can give the synapses at $C_1, C_2, .., C_k$. responsibility for both multiplication and integration. The former depends upon the nonlinearity of the synaptic relay, the latter upon its sluggishness. Let us therefore amend the notation and call the behavior of the group of C-neurons (i.e. the output of the system) $\varphi(t, \tau)$.

We have thus far an arrangement of neurons that determines approximately the running autocorrelation function of $F(t, x_i)$. The arrangement has a simple network of neurons as its basic schema. However, it is not in essence a digital machine. Since the basic operations involve integrations over sets of neurons, the discontinuity of the discharges of the individual neurons is smoothed over and does not appear in $\varphi(t, \tau)$. For the same reason, exact replication of the neuronal arrangement shown in Fig. 1 is not required. In fact, a certain amount of statistical variation of microstructure is quite as desirable for the functioning of the mechanism as it is bound to occur in neural tissue.

Bringing together the cochlear frequency analysis and the neuronal autocorrelation, we note that our discussion of the latter has given us the autocorrelation function of a single channel of the cochlear output. The cochlear frequency analysis transforms $f(t)$ into $F(t, x)$, of which $F(t, x_i)$ is but one part, separated spatially from the others. Our $\varphi(t, \tau)$ describes only the signal

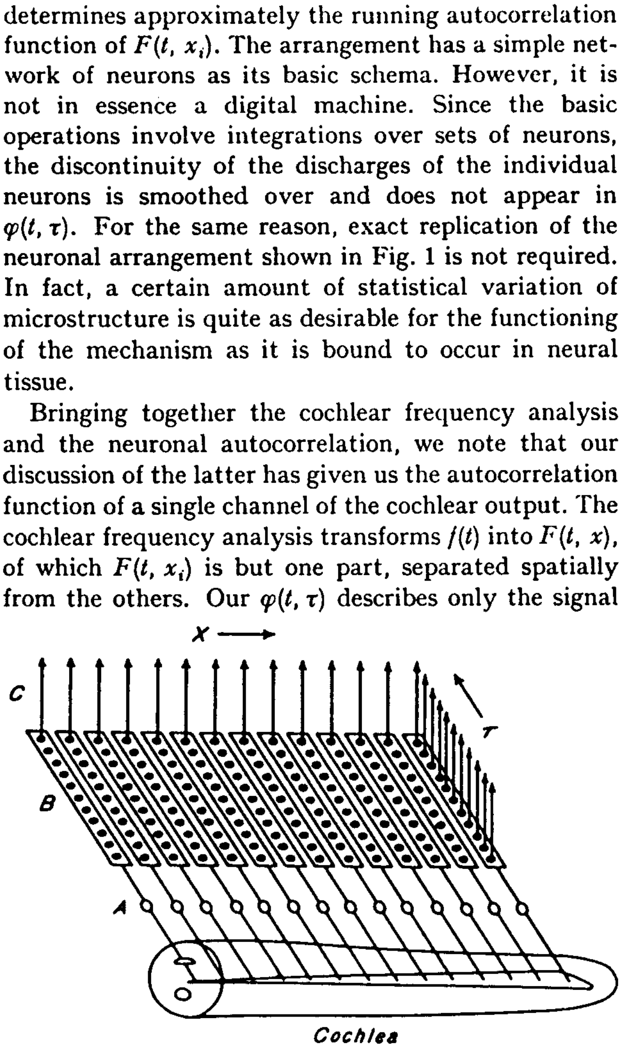

Fig. 2. – *Schematic diagram of overall analyzer.* At the bottom is the uncoiled cochlea. Its lengthwise dimension and the corresponding dimension in the neural tissue above it is designated the x-dimension. The cochlea performs a crude frequency analysis of the stimulus time function, distributing different frequency bands to different x-positions. In the process of exciting the neurons of the auditory nerve, the outputs of the cochlear filters are rectified and smoothed. The resulting signals are carried by the groups of neurons A to the autocorrelators B, whose delay- or τ-dimension is orthogonal to x. The outputs of the autocorrelators are fed to higher centers over the matrix of channels C, a cross-section through which is called the (x, τ)-plane. (Output arrows arise from all the dots; some are omitted in the diagram to avoid confusion.) The time-varying distribution of activity in the (x, τ)-plane provides a progressive analysis of the acoustic stimulus, first in frequency and then in periodicity.

in the ith channel; it is $\varphi(t, \tau, x_i)$. We must think of the neural arrangement, therefore, as extended in two spatial dimensions. The one corresponding to frequency is the x-dimension, or the dimension of the nervous tissue into which the lengthwise dimension of the cochlea projects. The whole arrangement for determining autocorrelation functions is replicated in the x-dimension. The τ-dimension is functionally orthogonal to the x-dimension, and we can think of it, at least for convenience of graphical representation, as being spatially orthogonal, also. The over-all system, then, yields a representation of the stimulus $f(t)$ in two spatial dimensions and time, a running autocorrelation $\varphi(t, \tau, x)$ of the components in each of many frequency bands. The arrangement is shown schematically in Fig. 2.

Relations between theory and observation

The duplex theory accounts immediately for two observations that cause ordinary place theories great difficulty. These are the observations of Miller and Taylor[1] on the pitch of interrupted white noise and of Schouten[2] on the residue phenomenon. Miller and Taylor found that their listeners could match with an oscillator tone the pitch of random fluctuation noise that was chopped into segments (on half the time, off half the time) at rates between about 40 and 250 per second. According to the duplex theory, both the tone and the interrupted noise produce activity in the same stria of the (x, τ)-plane. The distributions of activity set up by a 100-c.p.s. sinusoid and white noise interrupted 100 times per second are shown schematically in Fig. 3, A and B (see Fig. 3).

The basis for the pitch match is evident.

The acoustic stimulus that gives rise to Schouten's effect consists of the high-frequency harmonics of a frequency in the interval 30 to 300 c.p.s. Schouten's listeners reported that the high-frequency sound had about the same pitch as a (low-frequency) sinusoid of the same fundamental period. Repeating Schouten's work with a spectrum consisting of lines at 4,000, 4,100, 4,200, ..., Rosenblith[3] found that many of his listeners made the same judgment: they matched the high-frequency sound in pitch with a sinusoid of about 100 c.p.s. However, some insisted that the pitch of the sound was quite high[4]. The distribution of activity in the (x, τ)-plane, shown in Fig. 3C, shows that both reports are reasonable. There is simply a disagreement among the listeners about which of the attributes—the one based on periodicity in τ or the one based on position in x—is meant by "pitch".

[1] G. A. Miller and W. G. Taylor, J. Acoust. Soc. Amer. *20* 171 (1948).

[2] J. F. Schouten, Philips Tech. Rev. *5*, 226 (1940).

[3] W. A. Rosenblith, Progress Report II (PNM-6) of the Psycho-Acoustic Laboratory, Harvard University (1947).

[4] A similar division of judgment of pitch was reported by H. Davis at the June, 1950, meeting of the Acoustical Society of America. Davis' acoustic stimulus was a carrier of about 2000 c.p.s. modulated at 123 c.p.s.

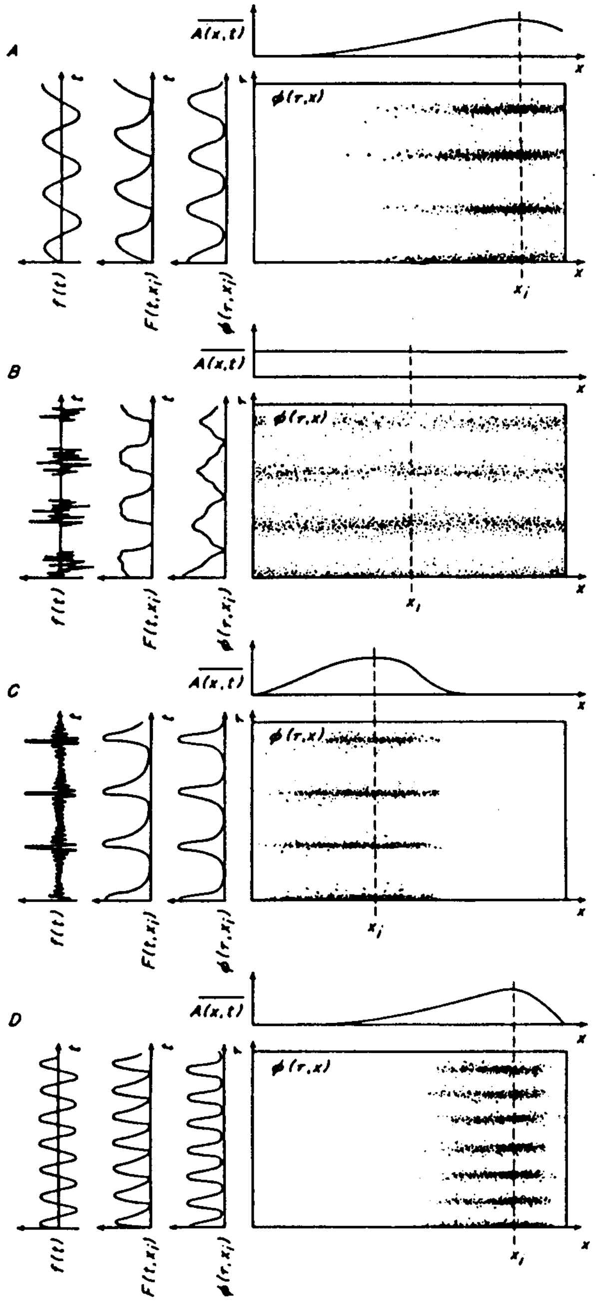

Fig. 3. – *Schematic illustrations of duplex analysis. A* represents the analysis of a 100-c.p.s. sinusoid, *B* of white noise interrupted 100 times per second, *C* of a set of high-frequency harmonics of 100 c.p.s., and *D* of a 200-c.p.s. sinusoid. At the left in each plot are shown the stimulus waveform $f(t)$, the waveform $F(t, x_i)$ of the signal carried by the first-order neurons (acting as a group) at x_i, and the autocorrelation function $\varphi(\tau, x_i)$ of $F(t, x_i)$. At the top of each plot is the distribution of activity along the length of the cochlea: $\overline{A(x,t)}$ is the root-mean-square of the instantaneous amplitudes of oscillation $A(x, t)$ at various positions along the cochlear partition[15]. $F(t, x,$ results from the rectification and smoothing of $A(x, t)$. The density of stippling in the rectangle represents $\varphi(\tau, x)$, the autocorrelation functions of the signals in the various x-channels. (Since the signals are in steady state, the t-dimension is omitted here.) Note that the first three (x, τ)-plots are similar in the τ- but not in the x-dimension. This corresponds to the fact that they are subjectively similar in one pitch-like attribute but not in another. *D* is somewhat similar to *A* in the τ-dimension: the odd-numbered maxima of $\varphi(\tau, x)$ in *D* coincide with the maxima in *A*. This corresponds to the subjective uniqueness of the octave relation.

The octave relation, the musical third, fourth, and other consonant intervals are understandable on essentially the same basis. When the frequencies of two sounds, either sinusoidal or complex, bear to each other the ratio of two small integers, their autocorrelation functions have common peaks. The 200-c.p.s. sinusoid in Fig. 3*D* gives rise to maximal activity wherever the 100 c.p.s. sinusoid of Fig. 3*A* does, and also in the strips that are half-way between. Furthermore, making use of the phenomenon illustrated in Fig. 3*C*, we note that the fundamental components of complex sounds need not be energetically present in the acoustic stimulus. The non-linearity of the neural excitation process introduces a component at the fundamental frequency before the autocorrelational analysis occurs. This fact explains the "case of the missing fundamental"[1].

The duplex theory also accounts for the subjective difference between the difference tone, heard when two moderately strong primary tones are presented to the ear, and a sinusoid of the difference frequency. It may account for the distinction between the two pitch-like attributes made by the listeners with "absolute pitch" who place a tone in the right region of the scale on the basis of ordinary pitch and then fix the note precisely with the aid of chroma. And it may account for some of the differences that have been noted between low-frequency and high-frequency hearing. The autocorrelational analysis must operate only for frequencies (frequencies of modulation *or* frequencies energetically present in the stimulus) that can be represented by volleys in the first-order neurons. Although there is evidence[2] that the volley principle operates up to 3,000 or 4,000 c.p.s., the effect at those frequencies is very weak, and it is safer to restrict the autocorrelational analysis to 1,000 c.p.s. or less. (It is unlikely that synapses provide delays of less than a millisecond, though of course axonal delays of almost any smaller duration may be postulated.)

A final comment concerns the plausibility of the autocorrelational schema from the neurological point of view. One of the essential features of the schema is division of the input into two channels, one with and one without built-in delay. In his histological investigation of the cochlear nucleus, LORENTE DE NÓ[3] found that the first-order auditory fibers branch, one division taking a one-synapse route to the next relay station, the other passing into a region of dense ramifications and thicket-like synaptic connections. Furthermore, recent work of GALAMBOS, ROSENBLITH, and ROSENZWEIG[4] shows that it is entirely reasonable to postulate

[1] S. S. STEVENS and H. DAVIS, *Hearing, its psychology and physiology* (John Wiley and Sons, Inc., New York, 1938). – J. F. SCHOUTEN, Proc. K. Ned. Akad. Wet. *43*, 356 (1940).

[2] E. G. WEVER, *Theories of hearing* (John Wiley and Sons, Inc., New York, 1949).

[3] R. LORENTE DE NÓ, Laryngoscope *43*, 1 (1933).

[4] R. GALAMBOS, W. A. ROSENBLITH, and M. R. ROSENZWEIG, Periodic Status Report IX (PNM-18) of the Psycho-Acoustic Laboratory, Harvard University (1949).

Time constant ≈ 2-3 msec

for the cochlear nucleus the delay of 1/30 sec. that is required if the autocorrelator is to operate as low as 30 cps. These considerations, together with the fact that the autocorrelator should be as near the cochlea as possible so that it may operate upon the signal before temporal resolution is lost, suggest that the cochlear nucleus may be the site. It is perhaps best, however, not to commit the theory at the present time to a definite statement about the location of the mechanism.

APPENDIX

Autocorrelation

The running autocorrelation function defined roughly by expression (1) in the text is a generalization of the function known to mathematicians[1] as the unnormalized autocorrelation function

$$\varphi(\tau) = \lim_{\mu \to \infty} \frac{1}{\mu} \int_{-\mu/2}^{\mu/2} f(t)\, f(t+\tau)\, dt. \qquad (5)$$

$\varphi(\tau)$ is the average over all time of the product of the original time function $f(t)$ and the same function advanced by τ. The generalization is achieved by relaxing the requirement, which of course cannot be met in practice, that the average extend over all time. In general, we know nothing about the futures of the messages we receive. Certainly, the auditory system operates only upon the present and the not-extremely-far-distant past of the acoustic stimulus. We therefore take a running average (or, what amounts to the same thing, a running integral) instead of the average over all time. We also reverse the sign of the τ, so that we delay the signal instead of advancing it. This, again, avoids operating upon the future. The reversal of sign makes no difference to $\varphi(\tau)$ because, as defined in (5), it is an even function, symmetrical about $\tau = 0$. However, the substitution of the running integral for the all-time average requires explication.

As shown in Fig. 4, the first two operations in the determination of the running autocorrelation may be thought of as the same (except for the reversal of the sign of τ) as the first two operations in the determination of the function defined in (5). The function to be analyzed, $f(t)$, is shown as the heavy line in *A*. The first step is to delay $f(t)$ by a variable interval τ. In the figure, delaying $f(t)$ by τ_j (i.e. by one particular value of the variable τ) yields $f(t - \tau_j)$, which is shown as a dashed line. The second step is to multiply the original function by the delayed function. The product $\Pi(t, \tau_j) = f(t)\, f(t - \tau_j)$ is shown in *B*. We should of course have a set of such product functions, one for each value of the variable τ, instead of the single one shown in the figure.

[1] N. WIENER, *Extrapolation, interpolation, and smoothing of stationary time series* (John Wiley and Sons, Inc., New York, 1949).

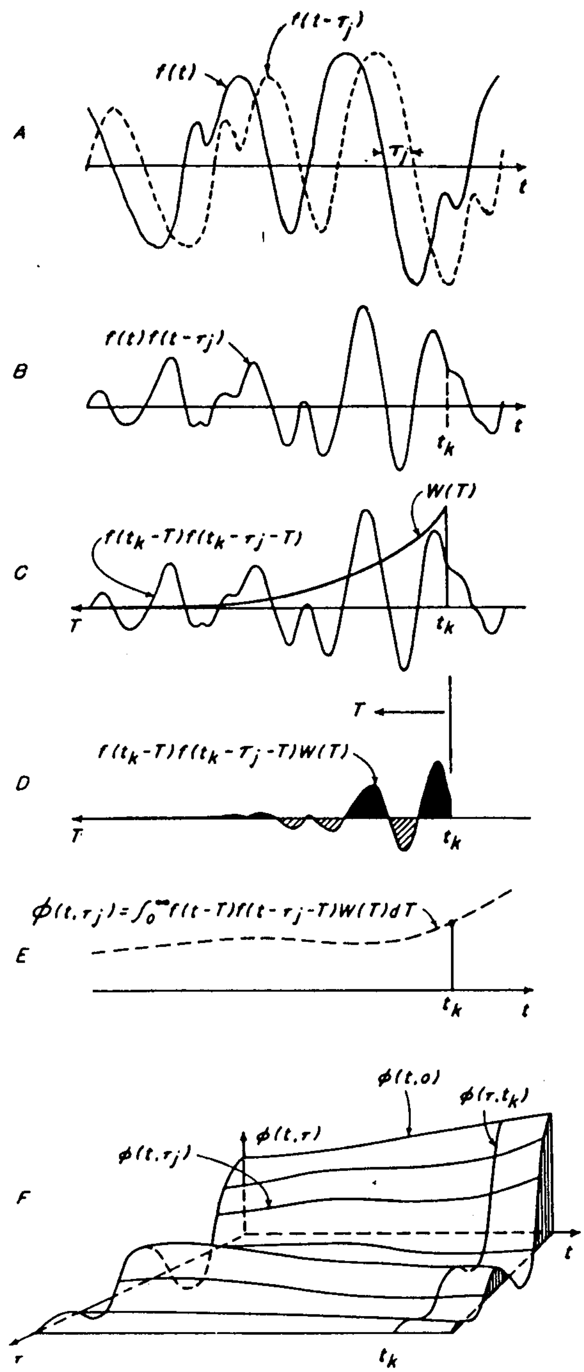

Fig. 4. – *Steps in the determination of the running autocorrelation function.* *A* shows the original function $f(t)$ and the delayed function $f(t - \tau_j)$. Their product is shown in *B*. In *C*, the instant t_k is taken as the present, and the product is re-expressed as a function of *T*, the distance into the past of t_k. The weighting function $W(T)$ determines the strength of the contribution of each past ordinate of the product to the autocorrelation coefficient. In order to let time *t* flow from left to right, the scale of *T* is oriented in the reverse direction as indicated by the arrows in *C* and *D*. The curve of *D* is the weighted product. It is integrated, the solid areas being considered positive and the cross-hatched areas negative, and the integral is plotted above t_k in *E*. Letting the present flow along in time and repeating the process for other values of time than t_k, we obtain the dashed curve in *E*. And, finally, we replicate the whole analysis for other values of the delay than τ_j. This generates the surface $\varphi(t, \tau)$ shown in *F*. The surface is the running autocorrelation function.

Thus far, we have proceeded as instructed by (5), but as we take the third step, we depart from it, substituting running integration for the average over all time. The running integral is the convolution of $\Pi(t, \tau_i)$ and a weighting function that specifies how strongly each past value of the product contributes to the accumulation at the present instant t_k. As in *C*, therefore, we rewrite the product as a function, not of t, but of distance T into the past of t_k. The choice of the weighting function $W(T)$ is to a considerable extent arbitrary: we might define any number of running autocorrelation functions, one for each possible $W(T)$. For the purpose of the duplex theory, a declining exponential with a time constant of 2 or 3 milliseconds is a reasonable choice. [$W(T)$ need not be defined for negative values of T.] The autocorrelation coefficient $\varphi(t_k, \tau_j)$ is then the integral of the weighted product over the entire past of t_k, but the distant past receives so little weight that it is effectively ignored. The result of weighting the product is shown in *D*. The integral of the weighted product is shown by the dot above t_k in *E*.

The procedure illustrated in *A* through *E* of Fig. 4 must of course be repeated for other values of t and τ. Repeating it for all t yields the function $\varphi(t, \tau_j)$ represented by the dashed line in *E* and shown again as a contour in *F*. Then repeating it for other values of τ generates the surface

$$\varphi(t, \tau) = \int_0^\infty f(t - T)\, f(t - \tau - T)\, W(T)\, dT. \qquad (6)$$

This is the running autocorrelation function for which expression (1) is a short-hand definition. The overline in (1) specifies that the operation

$$\int_0^\infty [\qquad\qquad]\, W(T)\, dT \qquad (7)$$

is applied to the product $f(t)\, f(t - \tau)$ after $t - T$ has been substituted for t.

Interpretation of the autocorrelation function is often facilitated, especially in cases in which we either naturally or through force of habit think in terms of frequency, by use of the WIENER theorem:

$$\left.\begin{aligned} \Phi(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(\tau) \cos \omega\tau\, d\tau \\ \varphi(\tau) &= \int_{-\infty}^{\infty} \Phi(\omega) \cos \omega\tau\, d\omega \end{aligned}\right\} \qquad (8)$$

Expression (8) tells us, for example, that the autocorrelation function of any sinusoid is a cosine function, the period in τ being the same as the period in t. Taking advantage of the fact that, if signals are superposed, the autocorrelation function of the sum may be found by superposing the autocorrelation functions of the individual signals, we can obtain the autocorrelation function of any signal that has a line spectrum by superposing cosine functions. Other aids to intuition are useful if the signal to be analyzed has a continuous spectrum. It is by no means necessary, however, to make the mental detour through the frequency domain. Some signals lend themselves more naturally to autocorrelational than to spectral analysis. The easiest way to find the spectra of telegraph and teletype messages, for example, is to determine the autocorrelation function first and then to take its Fourier transform.

The foregoing comments about the relation between the autocorrelation function and the power spectrum refer to $\varphi(\tau)$ and $\Phi(\omega)$, both of which involve integration over all time. Fortunately, analogous statements can be made about the running autocorrelation function, defined in (6), and the running power spectrum, measured with band-pass filters. FANO[1] has shown that, if the weighting function $W(T)$ in (6) is a declining exponential, and if the filters employ certain simple arrangements of resistances, capacitances, and inductances, the FOURIER transform relation extends to the running autocorrelation function and the running power density spectrum.

A final comment concerns the distinction between normalized and unnormalized autocorrelation functions. Often, the distinction is not made explicit. In communication engineering, the function defined in expression (5) is usually called "the autocorrelation function" without qualification. The signal $f(t)$ may have any average power; it may include a d-c component. "Autocorrelations" greater than unity may therefore arise. In other fields, especially those in which "correlation" means PEARSON product-moment correlation, it is natural to normalize $f(t)$ before operating upon it. The normalization eliminates the d-c component (sets the mean at 0) and adjusts the average power to unity (sets the variance at 1). The magnitude of the coefficient of autocorrelation is restricted to the interval $-1 \leqq \varphi \leqq 1$ by the normalization. For the autocorrelation function based on the average over all time [expression (5)], the normalization changes only the zero point and the scale factor; it leaves the shape unaltered. For the running autocorrelation function [expression (6)], however, the distinction is fundamental. In Fig. 4*F*, for example, the contour $\varphi(t, 0)$ would be a straight line at $\varphi = 1$ if the function were normalized. As the figure stands, unnormalized, $\varphi(t, 0)$ is the running average power (or squared amplitude) of $f(t)$.

Zusammenfassung

Es wird eine neue Theorie der Tonhöhenempfindung beschrieben. Diese hat unter anderem die Aufgabe, Beobachtungen zu erklären, in welchen die Tonhöhe (pitch) als eine *zweifache* Eigenschaft (duplex attribute) der Gehörsempfindung erscheint. In den meisten früheren

[1] R. M. FANO, *op. cit.*

Theorien wurde die Tonhöhe als einfache und einheitliche Eigenschaft betrachtet. Die Theorie setzt voraus, daß das Gehörorgan eine Doppelanalyse, und zwar eine Frequenzanalyse und eine Autokorrelationsanalyse vornimmt. Wir können uns vorstellen, daß diese Analysen in der folgenden Weise vor sich gehen:

Die Schnecke wirkt wie eine Anzahl von breiten Bandfiltern, die den Reiz auf viele Frequenzbänder verteilen. Diese Bänder sind räumlich voneinander getrennt, sie liegen einer einzigen Dimension entlang, nämlich der x-Dimension des Nervengewebes. In jedem Frequenzband wird der Ton einer weiteren Analyse unterzogen, und zwar durch eine Gruppe von Neuronen, die als ein Autokorrelator wirken. Dieser Autokorrelator analysiert in einer zweiten räumlichen Dimension, der τ-Dimension. Die Theorie erklärt gewisse psychophysische Beobachtungen, welche sich durch gewöhnliche Ortstheorien (Einortstheorien) nicht leicht erklären lassen. Rosenblith berichtet zum Beispiel, daß eine Impulsfolge, die nur aus den hochfrequenten Obertönen von 100 Hertz besteht, zwei «tonale» Qualitäten hat: die eine ist hoch, die andere tief. Diese Zweiartigkeit der Tonhöhenempfindung kann in der folgenden Weise aufgefaßt werden: Die eine Qualität ist verknüpft mit der Energiekonzentration in einem besonderen Frequenzband oder, vom Gesichtspunkt der Nervendimensionen aus gesehen, mit einem Aktivitätsort entlang der x-Dimension. Die andere Qualität ist mit der Reizperiodizität verknüpft und daher in dieser Theorie mit Aktivität in einem oder in mehreren Punkten entlang der τ-Dimension.