CORRELATION FUNCTIONS AND SPECTRA OF MUSIC AND SPEECH

ILHAN UYGUR

TECHNICAL REPORT NO. 250
JUNE 15, 1953

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS
The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research), and the Air Force (Air Materiel Command), under Signal Corps Contract DA36-039 sc-100, Project 8-102B-0; Department of the Army Project 3-99-10-022.
A history of the study of music and speech in the field of communication is given. The basic ideas and tools in the statistical theory of communication connected to music and speech are discussed. The electronic technique for performing the necessary mathematical operations to obtain the statistical parameters called correlation functions which yield to spectra are described, with emphasis on the delay problem. The experimental results of a study of music and speech with these methods are presented.
I. Introduction

Since the physical characteristics of sound affect its intelligibility and its psychological properties, technical studies in music and speech are of great importance to both engineers and psychologists. These characteristics have to be known so that we can produce, transmit, and reproduce the sound in the most acceptable way.

Since the beginning of the century, many studies have been made dealing with measurements on single notes or vowels (1-22). A summary of earlier studies can be found in D. C. Miller's book "Science of Musical Sounds". In 1930, a very important experimental study was made at the Bell Telephone Laboratories by Sivian, Dunn, and White (23, 24). They were interested in making measurements on actual musical selections and speech, rather than on single notes or vowels, to obtain an average picture of the selection as well as the distribution of amplitudes in magnitude and frequency. They used an apparatus in which speech or music spectra were divided into thirteen bands of frequencies and the power in each of them was measured. These measurements resulted in a set of curves which have become standard reference data in acoustical engineering literature for the absolute amplitudes and spectra of speech and music.

Since then, the rapid growth of electronic technique has provided very powerful tools for the measurement, recording, and reproduction of sound; tools that have revolutionized acoustic technique. In the last decade another useful tool, although perhaps not so obvious, has been discovered in the realm of statistical communication theory.

Earlier studies of sound did not have the advantages of these tools. In this study, the writer will try to show how he has attempted to apply methods of statistical communication theory to the study of music and speech.

Communication can be defined as "any form of transmission of information." The information which is to be transmitted cannot be considered as a known function of time because it would be completely specified by its amplitude and phase spectrum or by one complete period; and once this is known, the continuation of its transmission would not convey any new information. Thus, when we have a flow of information, it has to be a random function of time, that is, a statistical phenomenon. This analysis shows one that all communication problems are statistical in nature. This idea was developed by N. Wiener (25-27), who is among the many contributors in recent years to the development of the new theory of communication, in which the methods and the techniques of statisticians have been applied.

The branch of the statistical theory which is applicable to communication problems is the theory of random processes. In this branch of the theory, the most useful and complete statistical parameters are probability distributions and correlation functions. The probability distributions are more inclusive than the correlation functions.
in the sense that the correlation functions can be derived from the probability distributions, but not vice versa. However, in many practical cases, it is more economical, in time, to determine the correlation functions experimentally.

The correlation function \( \phi(\tau) \) of a random time function \( f(t) \) may be expressed \(^{28}\) as:

\[
\phi(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} f(t) f(t + \tau) \, dt
\]  

which is a time average and requires shifting, multiplying, and averaging over the infinite interval. In the case of stationary random process, that is, when the long-time average statistical properties of the correlation functions are independent of the time at which they are determined, the time average is equal to the ensemble average (ergodic hypothesis):

\[
\phi(\tau) = \iint a \, b \, P_2(a, b; \tau) \, da \, db
\]

where \( a \) and \( b \) are possible amplitudes of \( f(t) \); and \( P_2(a, b; \tau) \) is the joint probability of occurrence of \( a \) and \( b \) separated by a time \( \tau \). Hence a correlation function can be defined for a stationary random process as well as for a single function.

Since only one member of the ensemble is required for computation, a convenient approximate method of evaluating \( \phi(\tau) \) experimentally is to average a large number of products of pairs of samples of \( f(t) \):

\[
\phi(\tau) = \frac{1}{N} \sum_{n=1}^{N} a_n b_n(\tau)
\]

where \( N \) is a large number, and \( a_n \) and \( b_n \) are samples of \( f(t) \) separated by the interval \( \tau \).

A very important property of correlation functions, and it is of basic importance in this study, is the Wiener-Khintchine theorem, which states that the power density spectrum \( \Phi(\omega) \) and the correlation function \( \phi(\tau) \) are Fourier transforms of each other. That is

\[
\phi(\tau) = \int_{-\infty}^{\infty} \Phi(\omega) e^{i\omega \tau} \, d\omega
\]

*In the literature, this is called autocorrelation function. When the delayed \( f(t) \) is different from the undelayed \( f(t) \), it is called crosscorrelation function:

\[
\phi_{12}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} f_1(t) f_2(t + \tau) \, dt
\]
These relations indicate immediately that knowing the correlation function of a time function is in every sense equivalent to knowing the power density spectrum.

The well-known Fourier theories of spectrum analysis are not applicable, as they stand, to random functions such as music and speech. It is possible to obtain the spectra of random functions only with an extension of the Fourier integral theory (32). Then the power density spectrum can be written (29) as:

\[
\Phi(\omega) = \frac{1}{2\pi} \lim_{T \to \infty} \int_{-T}^{T} f_T(t)e^{-j\omega t} dt
\]

where \(f_T(t)\) is a section of duration \(2T\) of the random function \(f(t)\). This is a very tedious task even when one takes advantage of short cuts (28). Therefore, in cases in which random functions are involved, the indirect determination of the power spectrum through the use of correlation functions is frequently the most convenient procedure.

The practical application of statistical theory of communication is a result of the tremendous development of electronic methods during World War II and the development of Wiener's prediction and filter theory in 1942. By then, electronic techniques were sufficiently advanced so that this new theory could be applied to the study of some important communication problems such as improvement of signal-to-noise ratio, bandwidth reduction, the lowering of power requirements for the transmission of given messages, and so on. These problems were either very difficult or impossible to handle with the conventional theory of communication. Statistical studies started in different groups (30-33). While one group at the Bell Laboratories was working principally on the theoretical side of these problems, the M.I.T. Research Laboratory of Electronics formed a group to work on their experimental aspects. One of the earliest experimental devices was an analog correlator (34) which proved the feasibility of high-speed electronic computation of correlation functions. But the requirement for great accuracy and stability, and especially for very long storage, indicated the need for digital techniques. The outcome of this was a digital correlator (35). That was followed by a short-time correlator for speech waves (36), another analog correlator (37), and a five-channel analog correlator (38). At the same time, experimental studies were progressing on the problem of measuring various probability distributions of random functions (39).

Today, one can find many people applying the methods and concepts of mathematical statistics to the study of the basic principles of communication.

Statistics may be characterized briefly as the science of reduction and analysis of observational materials. Any scientific treatment of a given material demands the introduction of a certain order into the material dealt with. Order demands classification; therefore, any given science is faced by the problem of classifying the available material.
according to some principle. The question then arises, What shall this principle be? There is no one definite principle available a priori that would enable one to make a classification suitable for every purpose.

Many people in the past brought order to the field of music and speech by studying it from the angle of power spectra, amplitude distributions, zero crossings, and the like. In this study, the writer tried, with the aid of statistical tools, to bring some order to two questions which are of interest to communication engineers.

The questions are these:

1. When a musically trained person listens to a sample of music, he can usually recognize whether it belongs to the classical period or to the romantic period, and even identify its composer. In other words, musicians have a classification for types of composition. One asks, Is it possible to make a music classification of the same kind or of a different kind with the aid of one of its statistical parameters, the correlation function in this case?

2. A study of speech has been made at the M.I.T. Research Laboratory of Electronics by using the statistical technique (40). The correlation functions of a male and a female voice reading different passages from a current magazine have been calculated. These functions have different shapes. Our question is, How different would these curves be if the same person read samples from different languages, or from different kinds of literature (prose or poetry) in the same language, or if different persons read the same literature?

The results are presented in the next section.
Fig. 1
Correlation curve of orchestra playing a rhythmic polyphonic selection (Bach, Brandenburg Concerto No. 6 in B Flat Major, final movement).

Fig. 2
Correlation curve of orchestra playing a rhythmic classic selection (Beethoven, Quartet in C Sharp Minor, first movement).

Fig. 3
Correlation curve of orchestra playing a rhythmic classic selection (Beethoven, Piano Concerto in G Major, first movement).
Correlation curve of orchestra playing a rhythmic classic selection (Mozart, Piano Concerto in E Flat, final movement).

Correlation curve of orchestra playing a melodic classic selection (Mozart, Eine Kleine Nachtmusik, second movement).
Fig. 6
Correlation curve of piano playing a rhythmic polyphonic selection
(Bach, Piano Concerto in F Minor, final movement).

Fig. 7
Correlation curve of piano playing a rhythmic classic selection
(Beethoven, Sonata in F Minor, first movement).
Fig. 8
Correlation curve of piano playing a rhythmic romantic selection (Chopin, Polonaise in A Major).

Fig. 9
Correlation curve of piano playing a melodic polyphonic selection (Bach, Prelude No. 22 for Well-Tempered Clavier).
Fig. 10
Correlation curve of piano playing a melodic classic selection (Beethoven, Sonata in C Minor, second movement).

Fig. 11
Correlation curve of piano playing a melodic romantic selection (Chopin, Nocturne in D Flat).
Fig. 12
Correlation curve of violin playing a rhythmic polyphonic selection
(Bach, Violin Concerto in E Major, first movement).

Fig. 13
Correlation curve of violin playing a rhythmic classic selection
(Beethoven, Sonata in G, first movement).
Fig. 14
Correlation curve of violin playing a melodic polyphonic selection (Bach, Violin Concerto in A Minor, second movement).

Fig. 15
Correlation curve of violin playing a melodic classic selection (Beethoven, Sonata in C Minor, second movement).
II. Results

There are many ways that one can plan a study of this nature. The procedure that appealed the most to the writer was the following.

For music, three periods were chosen: polyphonic, classic, and romantic. A typical composer from each period was taken, and a rhythmic and a melodic composition from each composer were selected. The recordings of these selections were obtained under the same conditions. Then they were analyzed by obtaining their correlation functions.

For speech, three readers were taken; the first with a monotone voice, the second with a foreign accent, and the third with a regular voice and good accent. They all read separately the same prose and the same poetry in English. Then the third reader read passages from two different classes of languages; in this instance, German and Russian. Recordings of these readings were made under the same conditions. Then they were analyzed by obtaining their correlation functions.

This study of music started with orchestral selections. Since records are recorded under differing conditions, that is, in different halls with different reverberation times, with different microphones and amplifiers of different frequency responses, with different musicians and conductors, it is obvious that records could not have been used for this study. Therefore, it was decided that if the concerts of the Boston Symphony Orchestra playing at Symphony Hall, Boston, Massachusetts, were recorded with the same equipment, the factors which affected the recordings would be almost the same. Some of the selections from those recordings were analyzed. The results are the correlation functions shown in Figs. 1-5, all of different character. The study of these results showed that the field of the study had to be narrowed down; for even though the recordings were made from the same hall with the same equipment, the seating of the orchestra was not the same on different days for different compositions. The size of the orchestra was also different for different works. In order to obtain a much closer control of the environmental factors, it was decided to narrow down the field of the study to individual instruments and artists. The instruments chosen were piano and violin. The selection of rhythmic and melodic compositions for each period was made in such a way that if a person with no musical training listened to all the rhythmic compositions from different periods, he would not see much difference between them. In other words, the selections would appear to be of the same character and period. The correlation functions of the piano selections are shown in Figs. 6-11. They are, in order, polyphonic, classic, and romantic periods. The rhythmic compositions contain more variations in their correlation functions than the melodic selections. If these variations have some periodicities, the time interval of the periodicities corresponds to the power peaks in the frequency domain. For example, the periodicity of approximately 6 msec of Fig. 10 shows a power peak around 160 cps in the frequency spectrum (see Fig. 32). There is also a tendency to more variations in the compositions as the periods go from
polyphonic to romantic. Certainly from a limited study of this kind, one cannot come to definite conclusions. It is easy to see that it will not be simple to recognize the composer or the period of a musical piece by looking at its correlation function. But at least a correlator will be able to tell, as will a musically untrained person, whether a composition has a songlike tempo or a rhythmic and lively one. Maybe the correlators of the future with built-in memory and capacity to learn will be able to distinguish Bach from Beethoven.

In the right corner of each graph of the piano selections appears a histogram showing the relative frequency of the pitch, calculated on the basis of four units to a whole note. This calculation was derived directly from the score of the selection analyzed. The writer called these histograms "distribution curves". As will be explained later in Figs. 32 and 33, they show the relation between the music and its frequency spectra.

To eliminate completely the effect of any echo from the walls, the recordings of violin and speech selections were made in the anechoic chamber of the M.I.T. Acoustics Laboratory. Only selections from polyphonic and classic periods were analyzed for violin. Their correlation functions are shown in Figs. 12-15. They show the same characteristics as those of the piano selections; that is, the rhythmic compositions have more variations.

The last part of the study was devoted to speech. Speech correlation functions are shown in Figs. 16-23. The first reader is the one with a monotone voice (Figs. 16 and 17). The second reader has a foreign accent (Figs. 18 and 19). The third one is a professor of modern languages reading English poetry (Fig. 20), English prose (Fig. 21), German (Fig. 22), and Russian (Fig. 23). The comparison of these curves shows that the correlation functions or their Fourier transforms, the power density spectra, of speech waves depend more on the reader than on the language used or the text read. To find speech waves of this character is of interest, but not surprising. Since the main physical characteristic of speech, its quality, changes with the resonant pitches of the throat and mouth cavities, and these characteristics differ from person to person, the comparison of the curves of the first reader (the monotone voice) with the others shows clearly how differently pitched voices appear in the correlation functions. The vowel sounds, which are the low-pitched components of speech, carry most of the energy.

Intelligibility, on the other hand, is largely due to the high-pitched consonants added by the tongue, teeth, and lips. Difference in languages is formed by a few characteristic parameters of each language. For instance, an English-speaking person has to differentiate between the sounds "t" and "th". To a Turkish-speaking person, however, that would not be noticeable. To him, the difference is great between "u" and "ü". Another difference in languages is found in the sequence of sounds which are mostly influenced by consonants. Therefore the curves obtained in this study apparently show the influence of low-pitched vowel sounds of individuals. For the interest of linguists, a further study in this field would be to remove the consonants from a speech sample tape and then study it to see how much difference there is from normal speech. A statistical study of just
the consonants of a speech sample would also be of interest in the study of languages.

The next problem was to obtain the power density spectra of the selections in which the greatest number of technical people would be interested. As was shown in the first section, theory simply states that once the correlation function is known, its Fourier transform automatically gives its power density spectrum. In this study, the writer was confronted with some difficulties. Due to technical limitations of the available equipment, which will be explained in section 3, the correlation functions could be calculated only over a limited range. This fact introduced difficulties to the calculation of their power density spectra. To see the difficulties more clearly, let us examine Fig. 24.

The part of the correlation function obtained experimentally is shown by \( f(t) \); the unknown end of the function, by \( g(t) \); and the assumed zero level of the function (the square of the mean value), by \( h \). Since the computing circuits of the machine do not handle negative numbers, the input to the computer is biased to make it all positive. We call this bias \( k \). And the correlation function obtained from the computer will be

\[
D_1 = \sum_{i=1}^{N} \left[ f_i(t) + K \left( f_i(t + \tau) + K \right) \right]
\]

\[
D_1 = \sum_{i=1}^{N} f_i(t) f_i(t + \tau) + K \sum_{i=1}^{N} f_i(t) + K \sum_{i=1}^{N} f_i(t + \tau) + NK^2
\]

\[
D_1 = N\phi(\tau) + 2NK \overline{f(t)} + NK^2
\]

which is \( N \) (number of samples) times the correlation function with a constant \( [2NK \overline{f(t)} + NK^2] \) added to it.

To find the zero level of the correlation function as \( \tau \) goes to infinity, let us correlate* the function \( f(t) \) with a constant \( K \) (for simplicity, this \( K \) is taken the same as the bias \( K \)):

\[
D_2 = \sum_{i=1}^{N} \left[ f_i(t) tk \right] K
\]

\[
D_2 = \sum_{i=1}^{N} f_i(t)K + NK^2
\]

\[
D_2 = KN \overline{f(t)} + NK^2
\]

\[
N \overline{f(t)} = \frac{D_2}{K} - NK
\]

*Credit to Dr. A. Fleisher, of the Meteorology Department, Massachusetts Institute of Technology.
Then $\left[ N \overline{f(t)} \right]^2$ would give the value of the zero level of the correlation function as the output of the computer.

If we subtract Eq. 8 from Eq. 7, we obtain

$$D_1 - D_2 = N\phi(\tau) + KN \overline{f(t)}$$

From Eq. 7 we have

$$\phi(\tau) = \frac{D_1}{N} - K^2 - 2K \overline{f(t)}$$

Equation 8 gives us

$$\overline{f(t)} = \frac{1}{K} \left( \frac{D_2}{N} - K^2 \right)$$

Then we obtain the actual zero level of the correlation function:

$$\phi(\tau) - \overline{f(t)}^2 = \frac{D_1}{N} - K^2 - 2 \left( \frac{D_2}{N} - K^2 \right) - \frac{1}{K^2} \left( \frac{D_2}{N} - K^2 \right)^2 \quad (9)$$

By itself,* the Fourier transform of $f(t)$

$$\int_0^{+\infty} f(t) e^{-j\omega t} dt \quad (10)$$

would not exist because of the constant $h$ superimposed on it. The expression

$$\int_0^{+\infty} \left[ f(t) - h \right] e^{-j\omega t} dt \quad (11)$$

would be finite and the transform would exist if the value of constant $h$ were calculated. Even then, the transform would not represent the exact spectrum because, as shown in section 3.3, $g(t)$ is not known; that is, there is no way of finding out how the function approaches zero level. For this reason different arbitrary values of $h$ and $g(t)$ have been tried to show the effect of these factors on the spectra. Figures 25-27 show the power spectrum of Fig. 10 (Beethoven - melodic) for different values of $h$ without the extrapolation $g(t)$. As the zero level is raised, a constant value is subtracted from the correlation function which has as its Fourier transform a $(\sin x/x)$ curve. The comparison of these curves yields the conclusion that the last curve (Fig. 27) is the one with the most correct zero level, for it is known that there is not much power around very low frequencies. Figures 28-31 show the spectrum of the same composition for different $h$ and $g(t)$. It is obvious that the extrapolation of the correlation functions to zero

*The transform is taken from 0 to $\infty$ because the correlation functions are symmetrical about the zero axis.
level smooths out the power density spectrum. Figures 30 and 31 are probably the best approximations of the power spectrum of the composition studied.

The lower portions of Figs. 32 and 33 show the spectra of melodic and rhythmic piano compositions of Beethoven obtained through a different machine with the conditions of Fig. 27. The results of two machines, the first being mechanical and the second electronic, check very closely. Distribution curves obtained from the music are drawn on the same figures to the scale of the spectrum to facilitate comparison between them. One should not forget that in calculating the distribution curves only the fundamental frequencies of the notes written in the music are considered; the power density spectra contain all the harmonics of the fundamentals. A study made in Japan (41) on music and speech shows that the waveform of the signal usually is rich in harmonic content. The distribution curves calculated directly from the music, however, should give a good first-order approximation of the power spectrum.
Fig. 16
Correlation curve of speech:
Reader I reading English poetry.

Fig. 17
Correlation curve of speech:
Reader I reading English prose.

Fig. 18
Correlation curve of speech:
Reader II reading English poetry.
Fig. 19
Correlation curve of speech:
Reader II reading English prose.

Fig. 20
Correlation curve of speech:
Reader III reading English poetry.

Fig. 21
Correlation curve of speech:
Reader III reading English prose.
Fig. 22
Correlation curve of speech:
Reader III reading German prose.

Fig. 23
Correlation curve of speech:
Reader III reading Russian prose.

Fig. 24
Fig. 25
Normalized correlation curve of the melodic classic piano selection in Fig. 10.

Fig. 26
Normalized correlation curve of the melodic classic piano selection in Fig. 10.
Normalized correlation curves of the melodic classic piano selection in Fig. 10.

Fig. 27

Fig. 28

Fig. 29

Normalized correlation curve of the melodic classic piano selection in Fig. 10.
Fig. 30
Normalized correlation curve of the melodic classic piano selection in Fig. 10.

Fig. 31
Normalized correlation curve of the melodic classic piano selection in Fig. 10.
Fig. 32
Melodic classic piano selection (Fig. 10).

Fig. 33
Rhythmic classic piano selection (Fig. 7).
III. Techniques

3.1 Apparatus Used

Analog Correlator

A correlator is an electronic machine which evaluates the correlation functions by performing the operation of the equation

$$\phi(\tau) = \frac{1}{N} \sum_{n=1}^{N} a_n b_n(\tau)$$  \hspace{1cm} (3)

The analog correlator (37) of the Research Laboratory of Electronics was used in the first part of this study. The data are fed into this correlator in the form of a voltage. An amplitude sample, $a$, is taken from the input, which is stored as a charge on a capacitor. At a time, $\tau$, later, a sample, $b$, is taken and stored on a separate capacitor (see Fig. 34). Equation 3 indicates that the amplitudes of the input wave during each sampling period are to be multiplied. This is done by generating pulses of heights proportional to amplitude $a$, and of widths proportional to amplitude $b$. The area of the rectangle formed in this way ($a \times b$) is stored on an RC integrator. Then those samples are discarded and a new pair of samples is taken. The sampling and multiplying process is repeated with an interval, $\tau$, and each time the product obtained is added to the cumulative sum in the integrator. After $N$ such products have been obtained, the sum is recorded and the integrator is discharged. The sum recorded represents the value of the correlation function for the value of $\tau$ under consideration. By changing the value of $\tau$, we can obtain as many points on the correlation function as we desire.

The original Miller feedback integrator of the analog correlator had a time constant of approximately 100 sec. This is large enough, as compared to the time consumed in computing one point of the correlation curve (about 16 sec for 16,000 pairs of samples). But it was not free from drift, and this would mean a difficult future for the type of data for which the machine was going to be used. Therefore an integrator of the bootstrap type, with a time constant of 200 sec, using a unity-gain feedback amplifier, was substituted for the original integrator. Its operation is described in reference (38). The new integrator was made free from the small drift that it had by replacing carbon resistors with precision resistors, protecting it from draft by placing it in an insulated box.
putting batteries on grids and filaments, and using cooled and well-regulated power supplies. Actual data obtained from the integrator for a DC input are shown in Fig. 35. Each point is the equivalent of the cumulative sum of 16,000 pairs of samples integrated in 16 sec. A constant amount of voltage is subtracted from the product of each pair of samples. In this way, the recorder registers only the most significant part of the cumulative voltage, so that the fluctuations in the different cumulative voltages for different values of $\tau$ can be seen easily. In this case, there is no fluctuation because of the DC input. A picture of the analog correlator is shown in Fig. 36.

Digital Correlator

The digital correlator (35) runs on exactly the same principles as those of the analog correlator; but for great accuracy and stability, and especially for very long storage, the binary system is used for storing, multiplying, and integrating. With this system the only error sources are the sampling circuits and the circuit which generates binary digital numbers from the amplitude-modulated samples. These errors are minimized by careful design of the circuits used. The drift problem is reduced to a negligible amount by using a special feedback drift-compensating circuit. A picture of the digital correlator, which contains more than a thousand miniature tubes, is shown in Fig. 37. A comparison of this picture with the one of the analog correlator gives the reader an idea of how a machine grows up when more accuracy and flexibility is asked from it.

Delay Problem

To extend the study to the lower frequency range of music and speech, a larger delay of $\tau$ was necessary than either the analog or the digital correlator could offer. This increased delay could be obtained either by adding new units to the existing correlators to expand their delay limit or, externally, by using a separate delay unit and then cross-correlating the delayed and the undelayed function by keeping the delay unit of the correlator at $\tau = 0$.

The internal delay was obtained by adding a new memory bank, designed by A. J. Lephakis of the Research Laboratory of Electronics, to the digital correlator. This bank uses static magnetic memory units (42) which operate essentially as a magnetic trigger pair. They do not require vacuum tubes for maintaining position. A new
magnetic material called "Deltamax", having almost a rectangular hysteresis loop, provides information storage. It also provides the trigger-pair action which depends on whether the core material is represented by a point on the top of the hysteresis loop or on the bottom of the loop. The memory bank delays 10 digit binary numbers which correspond to the amplitudes sampled at channel A of the correlator. (See Fig. 38.) Each of the 10 delay channels consists of 200 static magnetic memory units, connected in a circuit of the shift-register type. Two units are associated with each of the 100 levels. The shift pulses are obtained from the correlator timing circuit, and occur at the correlator sampling rate T (see Fig. 34). Each shift pulse causes the contents of the delay channels to be transferred by one level. During the shifting process all stored pulses are sent to the crossbar relays, which may be positioned to apply the pulses from any one of the 100 levels to the output circuit. Stored pulses are lost when shifted out of the last level. The value of τ for which a correlation point is computed is equal to the delay between the pair of samples corresponding to the A and B numbers fed to the multipliers. Circuits in the correlator are capable of providing delays of from 0.1 μsec to approximately 5 msec. The memory band provides discrete delays equal to N correlator sampling periods, where N = 1, 2, ..., 100.

In the digital correlator, it takes about 1400 μsec to generate binary digital numbers from the amplitude-modulated samples, to multiply, and to integrate them. This necessitates starting the minimum sampling period from 2 msec. Therefore even with the combination correlator’s own delay plus the memory bank, which gives the delay in steps of sampling rate T of the correlator, there is still some gap to be filled in the time domain.

The first idea for obtaining a continuous large delay was to use a twin-track recorder and give an external delay to one of the tracks (50). The scheme is shown in Fig. 39. The twin-track recorder records on track A through head A, and on track B through head B. The position of the pulley C at which the recording is made is a zero-delay position between the tracks. Turning the threaded rod G clockwise will raise the pulley C and give delay between the tracks (track B is the delayed one) which can be different for various sizes of threads and different speeds of the tape. A picture of the device which was developed from this idea is shown in Fig. 40, connected to a twin-track magnecorder. A rotary solenoid D rotates a constant angle when energized. The shaft E of the solenoid is connected by gears to the threaded rod. Thus when the rotary solenoid is energized by the same pulse which resets the correlator at the end of the integration period, the pulley C is automatically raised by a definite amount which provides the delay. Different gear combinations give different steps, and discrete delays in small intervals up to seconds can be obtained.

There are some objections to this device. At the speed of 15 inches per sec, the tape going around the small pulleys F shown in Fig. 39 wears out too fast. These pulleys, of approximately 1/8 inch in diameter, were placed there to assure that the magnetic face of the tape would make a good contact with the heads. Another criticism is
that there is too much tension for the driving mechanism, because the peculiar path of the tape introduces some distortion to the supposedly constant speed of the tape. This distortion comes mostly from the slippage in the clutch mechanism. This could be avoided by adjusting the pressure of the clutch mechanism. Another error is introduced by the stretching of the tape because of the large pull on it.

In this study, it was thought that these errors could not be considered negligible, especially in the high frequencies of music. In low frequencies, where the tape could be run at a slower rate, most of these errors would be very small. The device could be improved by replacing the small F pulleys with pulleys of larger diameter to smooth out the track of the tape. That, at the same time, would improve the speed of the tape, because of the decreased pull on it. The stretching problem of the tape could be solved by using DuPont's "Mylar" plastic tape, which has a strong coating free from stretching.

The last source of delay was a variable time delay system (44) of the M.I.T. Acoustics Laboratory. The main interest of this system is a rotating drum coated with magnetic material. The input signal is recorded simultaneously on two tracks of the drum by techniques similar to those of conventional magnetic recording, and then reproduced a fraction of a revolution later. The relative time delay between the reproduced signals is varied by changing the angular spacing between the recording head and the reproducing head of one channel. This delay drum is much more accurate than the threaded-rod device because the error problems are not so great.

a) A mechanical driving system is developed with a peak value of flutter as small as less than 0.02 percent, which is much better than any tape-recording driving mechanism.

b) The use of the solid drum coated by a uniform magnetic material, obtained by spraying with a dispersion of iron oxide, solves the problem of the tape stretch.

Figure 41 shows the drum and its supports. The drum is made 6 inches wide in order to allow for the addition of other recording tracks in the future, which could be a good use for a multichannel correlator. The bearings are placed at one end so that the inside surface of the drum can also be used. The entire drum is machined from Duralumin and the outer surface (8 inches in diameter) runs concentric to within less than \( \pm 0.1 \) mil. This accuracy is needed in order to reduce the amplitude modulation of the signal during recording and reproducing processes. Recording, reproducing, and erasing heads are spaced a very short distance from the surface of the drum. If they were run in contact, the surface speed chosen would cause excessive wear on both heads and magnetic medium. Spacing the heads from the surface of the drum also reduces the mechanical load on drum drive to that of bearing friction alone, and makes a coupling system to the driving mechanism more effective in reducing fluctuation in angular velocity, and thus in delay \( \tau \). The over-all frequency response of the system is flat within \( \pm 2 \) db from 100 cps to 10 kc/sec with a 45-db signal-to-noise ratio. The relative time delay is continuously variable from -15 msec to 190 msec and is calibrated with an accuracy of \( \pm 0.2 \) percent, or 10 \( \mu \)sec, whichever is larger.
Figure 42 shows the whole system with its power supply and amplifiers.

Computer for Fourier Transforms

The Fourier transforms of two of the correlation functions of music were taken by a computer (45) to find the power density spectra of these selections.

The computer uses digital inputs and outputs in form of tapes punched in binary code. The internal operations are in part digital, in part analog. The computer uses the method of successive approximations based on a fundamental variational principle to evaluate integral transformation of the Fourier integral. The major mathematical function performed by the machine is that of transforming on an n vector with an n × n matrix.

The basic errors in this computer come from the following:
  a) Interpolation error:
     Error committed by interpolating the function space of a problem with a vector space. (See ref. 45, sec. 6.2)
  b) Round-off error:
     Error that occurs from computation with a limited number of significant figures upon data defined to a limited number of significant figures.
  c) Machine errors.

The Fourier Synthesizer

The electromechanical synthesizer of the M.I.T. Instrumentation Laboratory (46) was also used for the transformation of the correlation curves from the time domain to the frequency domain. This machine consists of an assembly of twenty-four synchros. The rotors of the synchros are geared together in the ratio 1:2:3:... The output from each synchro is the voltage induced across a stator winding. The stators can be set in any angular position to account for the phase angle of the complex exponential term that it is representing. In this case, they are oriented so that with the phase-angle setting at zero, the voltage induced across each synchro secondary is proportional to \cos \alpha, where \alpha is the angular position of the rotor. Then this induced voltage is attenuated by means of a high-precision potentiometer to give the desired magnitude of the complex exponential term. The attenuated voltages from all the synchros are added electrically. The resulting voltage is amplified, rectified, and impressed on the galvanometer of a Sanborn recorder of which the current is proportional to the sum of the cosine voltages. Since both the synthesizer and recorder are driven at constant speed, the timing line of the recording paper can be calibrated in terms of angular frequency \omega, and the transform is obtained as continuous functions of frequency.

3.2 Experimental Procedure

The first problem was to obtain permanent records of the music and speech which were going to be studied. For the orchestral music, the performances of the Boston Symphony Orchestra in Symphony Hall, Boston, Massachusetts, were recorded through
frequency modulation station WGBH. To obtain good recordings with minimum distortion, an Ampex tape recorder (47) was used. The next problem was to take passages from the recorded selections and correlate them. As pointed out in the previous section, the internal delays of the correlators were too short, and the delay intervals of the new memory bank were too great. Since the most reliable external delay unit was the magnetically coated drum, it was employed to solve the delay problem.

The following procedure was used to obtain the delay.

The output of the Ampex recorder was connected to the input of the delay drum. The outputs of the two heads of the drum, the undelayed and the delayed, were connected to a twin-track magnecorder. Then the delay of the drum was set to zero, the sample passage was played on the Ampex magnetic recorder, and was recorded on the twin-track recorder after going through the delay drum. At the end of the sample passage, the tapes were stopped on both recorders, the sample tape was rewound to the original starting point, a delay step of half a millisecond was given to the delay head on the drum, and the above procedure was repeated. In this way, rolls of two-track magnetic tapes were obtained with delay intervals of multiples of half a millisecond on each track. The last delay of each roll of tape was repeated at the beginning of the next roll to correct the recording levels if there was any difference in the gain of the magnetic materials of different rolls.

A continuous-loop method was tried for the rewinding of the sample tape, as shown in Fig. 43. It was found, however, that when a loop of tape longer than 15 feet or 20 feet was used, the tape (which has one side coated with conducting oxide material, and the other side with insulating plastic) built up a functional electric charge between layers of the tape. The charging action resulted in a force which pulled each layer of the tape to the next one. The attraction between the layers of tape produced an uneven pull on the drive mechanism of the recorder and in some cases broke the tape. The only remedy for this was to apply a graphite solution on the plastic side of the tape to make it conducting. This artifice was not satisfactory because of a distortion of the body of the tape. Therefore the continuous-loop method was not used. A new tape of the Minnesota Mining Company, which has an aluminum-sprayed plastic back, might be an answer to this problem, but it is not yet available in the market.

Once the rolls of the tape with delays were prepared, they were fed into the analog correlator with its delay set at $\Delta \tau = 0$. The undelayed track of the tape was fed to the A channel and the delayed track to the B channel of the correlator. In this way, voltage coming to the B channel had delay intervals of half a millisecond, and the results obtained from the correlator at the end of each integration period represented points of 0.5-msec interval on the correlation curve. After the preliminary curves were obtained on the analog correlator, the data were fed to the digital correlator for the final results.

A block diagram of the whole procedure is shown in Fig. 44. After many experimental runs, it was found that the sample selections could be reduced to about 2.5 minutes without losing the statistical characteristics of the whole sample. For the final curves,
these selections were run through the digital correlator and about 60,000 pairs of samples were taken for each point on the curve. After an average of ten runs was computed for each tape, the results obtained in the binary digital form were decoded and averaged. Each point on the correlation functions, therefore, represents the average of about half a million pairs of samples. Delay points zero and 2 msec apart were checked for each curve with the memory bank of the correlator to be sure that everything was in order during the recordings of the delayed tapes.

Fourier transforms of two of these correlation curves were taken with two different machines. The transforms in Figs. 25-31 were taken with the mechanical Fourier synthesizer, explained in section 3.1. To obtain the power spectrum, the correlation functions were quantized in 24 points. The maximum point being unity, they were normalized. Then the potentiometer of each synchro was set for the value of each of these points and the frequency spectrum was plotted by a Sanborn recorder.

The transforms in Figs. 32 and 33 were taken by the electronic machine that is explained in section 3.1. In this case, the correlation functions were quantized in 40 points, and the curves were normalized with a maximum value of 250. Then these values were coded to the binary digits and punched on a roll of paper. The plots obtained after decoding the binary digital output of the machine gave the power density spectrum of the selections of music studied.

The recordings of piano, violin, and speech were made with the same high-quality recorders (48) and a tuned ribbon microphone (49). The piano recording was made in a room measuring 15 feet \times 21 feet, with heavy curtains on large windows. The tail of the concert piano was closed during the recordings; the microphone was placed about 6 feet from the piano. The recordings of violin and speakers were made in the anechoic chamber of the M.I.T. Acoustics Laboratory. In this case, the microphone was placed about 3 feet from the subject. The purpose of this distance was to insure that the sound waves striking the ribbon would be essentially plane, and the bass boosting effect of the microphone when responding to spherical waves would be eliminated.

3.3 Discussion of the Errors

The errors in the measurements can be attributed to two sources: theoretical errors and errors from the equipment used.

The finite time average used in the measurement of the correlation functions, which are parameters of random time functions and are defined mathematically by infinite time and ensemble averages, contributes some statistical errors. The dependence of these errors on the averaging interval and on the finite number of samples has been well analyzed by Davenport and Middleton (50). An expression for the expected mean-square error of a sample mean and its variations with changes in total sampling time and spacing between samples is also well discussed by Costas (51). When measuring the mean or average value of a random time function of finite length by periodic sampling techniques, some error is to be expected. Davenport showed how the approximate idea
of the behavior of the variance of a correlation function, periodically sampled for a
time T, may be obtained by arbitrarily assuming some idealized form for the corre-
lation function studied. In the case of music and speech with rather peculiar correlation
functions, that procedure would be difficult. If we had used independent samples, the
variability of the correlation function computed from N such samples would be reduced
by $N^{-1/2}$.

The correlation functions obtained in this study came down to zero around 30 msec.
It seems reasonable to assume the independence achieved after this time. Therefore, if
we assume that we have been sampling the random function every 30 msec for 2 min, we
would have $2 \text{ min}/30 \text{ msec} = 4000$ pairs of independent samples. The variability would
be reduced to $N^{-1/2} = (4000)^{-1/2} \approx 1.6$ percent.

One cannot know what the absolute error in the measurement of a correlation function
is, unless he knows more about the statistics of the signal than the correlation function.
Davenport's latest report "Correlator Errors Due to Finite Observation Intervals" (Tech-
nical Report No. 191, Research Laboratory of Electronics, M.I.T.) shows, in Eq. 75,
the improvement of signal-to-noise ratio for N independent samples. To apply this
equation one must know:

$$R(0) = \frac{f(t)}{f(t)^2}$$

which is a fourth-order statistic of the signal.

The technical errors were caused mostly by the physical characteristics of the equip-
ment. It was found that the microphone has a boost of less than 3 db above 55 cps with
spacing used in recordings, as explained in the last section (50). The recorders have
frequency responses with pre-emphasis in recording process and post-emphasis in play-
back process. The combination of these two characteristics yields a response of $\pm 2$ db
from 50 cps to 15 kc/sec at the tape speed of 15 inches per second (48). The total har-
monic distortion generated in record-playback cycle, including tape and recording head,
has been computed and been found to be 2 percent (48). The magnetic delay drum has
a frequency response flat within $\pm 2$ db from 100 cps to 10 kc/sec (44). The anechoic
chamber where the recordings were made had a lower cut-off frequency of 70 cps (50).
This cut-off frequency is defined as the frequency below which more than one percent of
the energy incident upon the walls is reflected. It was the combination of the responses
of this equipment that brought the writer to the conclusion that he should not try to anal-
yze anything below approximately 75 cps if dependable results were to be obtained.

The noise problem in magnetic tapes studied by Wooldridge (52) was not considered
a problem in this study. Since the wideband noise that is caused by a large number of
small magnetic irregularities in the tape has an exponential correlation function that
comes down to zero level in a few microseconds, it contributes a negligible error to the
correlation functions obtained. There is a sharp upper limit and a lower limit to the
recording amplitudes which can be used without high distortion. Therefore the recordings
were made carefully within those limits, except in a very few cases of overshoot.
The errors which could be caused in the correlator by a faulty component were overcome by running the correlator with a constant input between each point and checking the results. Since the machine is a digital correlator, the only amplitude drift problem is in the sampling circuits. The errors of measurement in those circuits have been reduced by using separate compensators in each channel. (See Fig. 38.) The compensators act to keep the median value of the measured samples at a constant level. They provide a correction in the number-generating circuits which tends to stabilize the median value of the binary numbers in the digital parts of the machine.

The effects of synchronization of the input voltage periodicities with the sampling period of the correlator can be another source of error. If the sampling takes place in synchronism with a periodic input signal, it can be seen that the first samples of each pair of samples will always have the same value, depending only on the phase of sampling. In this case, the compensators will adjust the number-generating portions of the machine to produce the constant median number for the constant magnitude of sample. If the input samples change to another constant level, the compensators adjust again so that the same median number results and no change appears in the correlation curve. The compensators, having a time constant of about 4 sec, make such an adjustment in a few seconds. These effects of synchronization, however, did not introduce errors to the functions obtained in this study, because the input signal was not a pure periodic function of the sampling rate or of any harmonic of the sampling rate.

The characteristics of the various items of the equipment as described above were considered adequate for the purposes of this study.
Fig. 39

Fig. 40
Fig. 43

Fig. 44

-37-
IV. Conclusion

As noted in the introduction of this work, Sivian, Dunn, and White of the Bell Laboratories classified the music spectra according to instruments and the speech spectra according to male and female voices. A statistical study (53) done at the M.I.T. Research Laboratory of Electronics on the distribution of filtered speech amplitudes placed the speech results of Dunn and White in considerable doubt. Later, in the same laboratory, the results of another statistical study (38) based on the samples of the writer showed that when correlation functions of speech and music are obtained over 8-sec time intervals, successive curves are different. The results are shown in Fig. 45. The points are taken 400 μsec apart, the first one being $\tau = 0$. Some 4000 pairs of samples are taken during an 8-sec integration time. The time interval between successive curves is 20 sec. The integration time used for the results of Sivian, Dunn, and White was 15 sec. The writer found in this study that music or speech samples analyzed for less than two minutes were not sufficiently representative of the statistical characteristics of that particular piece of music or speech. In other words, the spectra obtained for different sections of that music or speech would be different. Furthermore, this study shows that it is difficult to classify the power spectrum of music according to the instrument, as was done by Sivian, Dunn, and White, for different types of music, even

![Fig. 45a](image1)

Correlation curves of male voice.

![Fig. 45b](image2)

Correlation curves of rhythmic classic orchestral selection.
on the same instrument and under the same conditions, give different spectra. It was also found that the spectrum of speech depends mostly on the reader's voice quality. Certainly a female voice, having a higher pitch than a male voice, would have power at higher frequencies. However, this is a classification of a gross character.

The studies of late years show clearly that it is a very difficult problem to transmit only the intelligence of a message and eliminate its redundancy. Through operational techniques it is possible to eliminate part of the redundancy from transmission. This, however, is usually accompanied by a small loss of information. The future will show how much progress can be achieved in this direction. Whatever results are obtained, it is certain that it will never enable us to grasp the problem in its totality any more than human intelligence will ever rise into the sphere of ideal spirit. However, we have the right to believe that the progress is real and not aimless. It must be achieved by filling the gaps. The writer will be happy if this study will fill even the smallest gap.

Acknowledgment

The writer wishes to express his gratitude for the inspiration and advice of Prof. J. Ruina, of Brown University, under whose supervision this study was carried out.

A study of this nature usually reflects the ideas of many people through the exchange of views and discussing the problems involved. Therefore a list of such persons would be long. However, the writer would like to thank Prof. Y. W. Lee for his encouragement throughout the course of the study, and express deep appreciation to the Director of the M.I.T. Research Laboratory of Electronics, Prof. J. B. Wiesner, for placing at the writer's disposal the facilities of the laboratory during this study.

The writer is indebted to Mr. L. G. Kraft, Mr. K. Goff, Mr. J. Kessler, Mr. J. Levin, and Mr. M. Stone for technical assistance. He would also like to acknowledge the assistance of his artists, Miss D. Bacon and Mrs. E. Fishman, and of his speakers, Prof. M. Halle, Mr. G. Constantine, and Mr. J. Graham.
References

1. H. Fletcher: Usefull Numerical Constants of Speech and Hearing, Bell System Tech. J., July 1925
4. H. M. Browning: Characteristic Partials of the Violin, Phil. Mag., Nov. 1926
References (continued)


34. T. P. Cheatham, Jr.: An Electronic Correlator, Technical Report No. 122, Research Laboratory of Electronics, M.I.T., 1951


43. P. E. Green, Jr.: Communication Research, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1951


47. Ampex Series 400 Magnetic Recorder Manual, Redwood City, California


49. Radio Corp. of America: Instructions for Velocity Microphone, Type 44-BX, RCA Victor Div., Camden, New Jersey

References (continued)

