

IX. PROCESS ANALYSIS AND SYNTHESIS

Dr. M. V. Cerrillo
Prof. P. Elias

Prof. R. M. Fano
Prof. E. A. Guillemin
Prof. Y. W. Lee

Prof. C. E. Shannon
R. E. Wernikoff

1. Introduction

1.1 In the Quarterly Progress Reports of January 15, 1957, page 101, and April 15, 1957, page 100, the objectives of this research were described by means of some illustrative and particular, but idoneous, examples and problems.

1.2 We now intend to carry on the description of the work along similar lines. First, a new problem that is concerned with the question of "style" will be presented. Second, we shall consider methods of attack on and possible solutions for some problems that have already been introduced. Specifically, we shall discuss in detail the formal solution of the problem of rhythm separation in music. (See Quarterly Progress Report, April 15, 1957, p. 101.)

2. Problems Described in Terms of Envelopes. The Question of Style

2.1 There are several important problems that can be described in terms of the envelopes of certain oscillatory functions. In fact, there are processes whose complete characterization strongly depends upon the envelope functions, and the particular instantaneous behavior of the oscillatory function may or may not be of basic importance.

2.2 Problems of this nature may have interesting interpretations and implications. Some of these interpretations may be of primary importance in connection with the work of our group. For example, some questions concerning style can be answered, in a first approximation, in terms of envelopes.

2.3 To illustrate this connection, we shall produce a specific, but illuminating, problem. In this example we shall show that the description depends almost exclusively on the envelope function and has a negligible connection with the actual instantaneous oscillations. The reader must evaluate the meaning of the problem from its implications in the general theory of communication rather than from its particular value.

The problem we have in mind reads:

"Suppose that we have two recordings, A and B, of the same musical composition, say a symphonic number. Let us assume that:

- (a) Both recordings are made under the same conditions.
- (b) Both are played by the same orchestra but under the direction of two different conductors. The conductors will be designated A and B.
- (c) Both conductors respect the meter, key, and instrumentation suggested in a certain musical script, but otherwise they are free to impose their own interpretation as their peculiar style.

(IX. PROCESS ANALYSIS AND SYNTHESIS)

- (d) The style of each conductor is such that we can tell, just by listening to each recording, whether A's or B's recording is being played by a regular phonographic machine, without reading the description on the records.

Under these premises, we inquire: Is it possible to construct a system of musical reproduction so that when A's recording is played the output of the machine sounds, as a "first approximation," as if conductor B were leading the orchestra, and vice versa?"

To better describe our objective, we can reword the final statement to read: "a machine that changes – or transforms – the 'style' of conductor A into the 'style' of conductor B, and vice versa."

2.4 Problems of this nature are of fundamental importance in the general theory of communication. If we use a more common technical nomenclature, a problem of this kind is an advanced example of "patterns of recognition." Although we have propounded this question in connection with music, its implication in general problems of communication is quite obvious. For example, we can speak of similar questions in connection with painting, and so on.

2.5 There are two fundamental questions concerning this style-transformation problem:

1. How can we characterize each single style?
2. How can we define the transformations that carry over these changes?

A detailed discussion of this style-transformation problem will appear in a later report. It is convenient, now, to present a preliminary discussion concerning the meaning of some terms that we have used. In particular, we are concerned with the terms "envelope," "style," and "first approximation."

2.6 Consider the term "envelope." By envelope we shall mean (unless it is otherwise specified) the wave or graph obtained by the double rectification of an oscillatory function after a certain smoothing process is introduced. The amount of smoothing depends upon the smaller instantaneous-wave frequency of the modulation function the traces of which are still allowed to appear in this envelope function. (As we shall see, sometimes a 12-cps trace is allowed in connection with the problem of rhythm separation.)

2.7 Concerning the meaning of the term "style," we find ourselves in a delicate and peculiar situation. We must proceed cautiously. A few explanatory digressions are pertinent.

(a) We do not intend to use the word "style" in its general connotation because, with such a broad meaning, our problem would become so colossal that we would not know how to handle it. In beginning the study of this subject we must deal with the most elementary concepts and terms.

(b) In trying to assign a certain meaning to "style" that will fit our immediate objectives, we must remember that we are not confronted with a semantic problem or

(IX. PROCESS ANALYSIS AND SYNTHESIS)

a problem in aesthetics or of general philosophy. Our problem is a simple mathematical problem, and that is what makes the situation difficult.

(c) We intend to limit the concept of style to its most elementary meaning and even then we must point out that not all elementary styles may admit the mathematical characterization that we propose.

(d) Our goal is purely mathematical. In some restricted elementary cases, at least in the situation that we want to create, we can establish a certain kind of correlation between some elementary concepts of style and the envelope properties of the oscillatory functions that represent the music. In other words, a functional associated with the envelope function can be interpreted as a restrictive definition of style. By this, we mean that the descriptive properties of such functionals follow a certain parallelism with the usual descriptive characterization of certain elementary styles. In more precise mathematical terms, we can say: a functional described in terms of envelope functions produces a definition of restricted styles. The external and internal construction of such a functional will be discussed with sufficient deterministic detail later.

(e) Statements made in (d) do not imply that every possible elementary style has a functional representation in terms of envelopes alone. We shall restrict ourselves, at present, to the cases in which this is so.

(f) In introducing this elementary functional representation of styles we must not oversimplify matters and produce trivial situations. As we shall see in a later discussion, the particular cases of style in which these functional relations can be established are quite illuminating.

(g) In producing the functionals between envelopes that characterize our kind of "style" we must bear in mind that a pure abstract definition of these functionals may be undesirable. This abstract approach may produce obscure situations in which we can scarcely visualize what is happening. On the contrary, we must produce simple and lucid definitions that can be easily understood. The functional construction and the descriptive elements must also be simple. These descriptive simple elements must accept a clear interpretation or translation into the actual descriptive elements of the musical style in question. The functional must have a structure simple enough to show the connection between elements and the effect of any change in these elements upon the resulting change of the style described by the functional.

(h) We have considered a problem in terms of musical elements and thus produced a concept of style. But in producing it we must not limit ourselves to music. The functional involved must be applicable to other artistic media. In other words, by a simple change of meaning of the structural elements of "style" functionals, we can apply the functionals to other artistic fields – painting, for example.

Let us summarize the preceding ideas concerning the mathematical definition and characterization of our elementary concept of style.

(IX. PROCESS ANALYSIS AND SYNTHESIS)

1. By "style" we shall understand a mathematical functional between envelope functions and some of their analytical elements (for example, some derivatives and a few integrals of the envelope functions).

2. By "change of style" we shall understand functional transformations among the "style" functionals.

2.8 Let us consider the term "first approximation." Obviously, this term is associated with the idea of how well a certain style functional represents a certain style or a certain transformation between given styles. Once a style functional is established, the question concerning the meaning of "first approximation" is reduced mathematically to the question of the establishment of a set of tolerances. Even then, although it sounds quite simple, the evaluation of the error of a representation is by no means a simple matter.

It is quite clear that the closer we want to make the representation of an actual style by a functional the more difficult and intricate the problem becomes. Since we do not yet understand how to characterize styles mathematically, even for simple cases, it is prudent not to use strict tolerances in the beginning of the investigation. We need, instead, loose tolerances. But, how loose?

To answer this question we may invoke the concepts of "gross structure," "fine structure," and "resolution." Let us begin with "gross structure."

(a) Suppose that we have two phonographic turntables and an ordinary reproduction system. On the first turntable we place the record labeled "A," and on the second the record labeled "B." Now let us play the recordings alternately. By the assumption of section 2.3(d) we can tell which record is being played. Now let us pay attention to the most conspicuous differences which necessarily must exist between the two recordings.

If record A were played in the style-transformation machine, the music output would sound – by the hypotheses of our problem – just like record B would sound. But the resulting sound illusion need not be a perfect one in order to produce a recognizable feeling that the style transformation has been accomplished. There is a sort of minimum number of conspicuous differences that are enough to produce in us the feeling that the transformation is beginning to take place, without the necessity of imposing all of the other details that "exactly" characterize the style of conductor B. Let us put it in this way. At a distance, we can recognize Michelangelo's statue of David even if we cannot see the specific detail of the eyes, hair, marble grain, and so forth.

(b) When the minimum number of prominent features is introduced in such a way that a definite feeling of recognition is felt, then we say that the gross structure of the transformation is achieved. The remaining elements form the "fine structure." The ability to distinguish between more or less fine-structure elements is called "power of resolution."

(IX. PROCESS ANALYSIS AND SYNTHESIS)

Concerning our proposed problem of style transformation, we may say, from the discussion of the problem that will be presented in the Quarterly Progress Report of October 15, 1957, that the gross elements of the transformation are given: by contrast, by attack, by impact, and by orchestral group emphasis differences between conductors.

We summarize this discussion by asserting: By "first approximation" we shall mean that at least the minimal number of prominent elements that characterizes the gross structure have been introduced.

2.9 Research work tending to clarify problems concerning elaborate processes of recognition has been conducted by the author since the beginning of 1953. The initial experimental mathematical work began in the Research Laboratory of Electronics at the end of 1953. A few machines were designed and constructed in which these transformations of style are possible in a first approximation. The principle on which they are founded will be discussed later. From January 1954 to May 1955, mathematical research on questions of recognition, style, and so on, was carried out by the author in the Department of Mathematics of the National University of Mexico. The author's work has continued at M. I. T. since June 1955. At the same time, parallel research has been conducted in connection with painting, by means of photographic processes.

3. RHYTHM EXTRACTION IN A MUSICAL COMPOSITION

(For the statement of this problem and other basic considerations, see the Quarterly Progress Report of April 15, 1957, pp. 101-102.)

3.0 Rhythm is the fundamental structure in music that serves as support for its other basic elements – mainly, melody and harmony. The proposed problem of style transformation, as stated in section 2, is strongly related to the problem of rhythm separation. In other words, to solve the problem of style change, it is first necessary to solve the problem of rhythm separation. For this reason, we shall now begin a formal discussion of the problem of rhythm extraction, in order to prepare ground for a future discussion of the problem of style change.

3.1 For the convenience of the reader, we restate the problem of rhythm extraction. The problem reads:

"Construct a system with two-pair access terminals such that, when excited in one pair by an input function taken from an appropriate recording of a musical composition, having, for example, constant meter throughout, then the output of the system at the other pair of terminals reproduces the rhythmic elements – both quantitative and qualitative components – of the musical composition. The rhythm must come out as tapped sounds, or as a recorded graph of impulses, or as a sampling of the beat on an oscilloscope screen. Moreover, the separation has to be made just after a few bars, say two or three, of the meter have been played."

(IX. PROCESS ANALYSIS AND SYNTHESIS)

3.20 A few considerations on the nature of the problem.

(a) The solution of the problem implies the separation of the very-low-frequency components (of the order of cycles per minute) of the musical composition.

(b) The separation of these slow components has to be made in a time whose order of magnitude is comparable with the period of the signals to be extracted.

(c) The extracted rhythm must reappear not only in synchronism, but also in phase, with the original rhythm of the music.

Requirement b is very strong indeed. It eliminates almost automatically the use of procedures based on mathematical statistical methods. Simple reflections on properties of linear systems raise doubts concerning the existence of a solution under this strong requirement b. But the problem has a solution, as revealed by the fact that our brain produces a solution with extreme ease. Of course, we do not understand how the brain manages to solve the problem and of course we do not now attempt to find out. What we accept is the crude fact that somehow there exists a solution. That is the challenge. We must try to find a solution with the means at our disposal.

3.21 Three possibilities suggest themselves as convenient starting points.

A. To try to filter out the low periodicities by direct means.

B. To use the fact that the beating of the meter produces relative maxima and minima of the intensity of the sound.

C. To try to count the number of notes per bar and use proper rational divisors in order to obtain the actual rhythm.

Clearly, B and C have in common their dependence on maxima and minima of the sound. Thus we can consider them as one.

Consider possibility A. Some methods of solution that have a linear character are immediately suggested; for example:

1. Construct a classical extra-narrow-band lowpass filter with a mean frequency close to the low periodicities to be extracted. Simple reflection indicates that such a solution is impractical and would not work well because of the transient phenomena in the filter.

2. Construct electrical models of Besicovitch spaces and then form the operator in such a way that, when they operate upon the music function $f(t)$, then low-frequency components will come out. Since we require that solutions be produced in finite time, then the Besicovitch space must be of the Banach type. (Besicovitch spaces are the spaces of the almost periodic functions (1). Since we do not follow this method, we omit further detail.) A simple computation reveals that the time needed to ensure a fair probability that the solution will be the low-periodic component caused by the actual rhythm is very much longer than that required under condition b of section 3.20.

The method of solution that we are going to describe is based on possibility B

(IX. PROCESS ANALYSIS AND SYNTHESIS)

(or C). The presentation is rather formal and it is more elaborate than is needed. However, we follow this method because it outlines the congruence of the whole process more sharply. A machine based on these principles (but greatly simplified) has been designed by the author and has been operating since 1955.

3.22 Although it is obvious that there is a connection between the rhythm and some maxima and minima of the musical sound, the solution of the problem of rhythm extraction by using this idea is by no means simple. The main difficulty is that maxima and minima of the sound intensity come from other elements of the musical composition which may be more numerous and better defined than those coming from the direct rhythm. In fact, the extremal values (maxima and minima) of the rhythm are, in general, so badly scrambled with other sounds that it is impossible to detect them at all times by direct observation of the graph of the music. Interference and reinforcement phenomena between the waveforms of the different instruments may also cause this confusion.

In extracting the rhythm we must consider the basic steps that lead to a final solution:

First, separation of the maxima and minima that come directly from the rhythm.

Second, separation of the emphatic or qualitative elements associated with the rhythm.

Third, arrangement of the maxima and minima in the pattern – or mosaic – that corresponds to the rhythm.

Fourth, introduction into this pattern of the emphasis produced by the qualitative elements.

Fifth, phasing the mosaic with the actual rhythm of the music. The mosaic is in synchronism with the music, but not necessarily in phase. The phase shift is caused by the time required to perform the operations of extraction. A mechanism must be provided to produce this rephasing.

Sixth, sustaining the mosaic even if the music stops for a fraction or so of the meter (when such a stop is called for in the script). But the mosaic must stop if the musical sound stops for the duration of a few bars.

3.30 Now we shall give a formal account of a procedure that tends to extract the rhythm of the composition and has the features indicated above.

In order to simplify the presentation we shall use block diagrams, sketches, graphs, and symbols for illustrating the method. We shall follow, as far as possible, the flow of the signal in the operations listed in section 3.22.

3.31 First Step: Channel Separation

The direct transcript of the musical signal, which is denoted by $f(t)$, is separated

(IX. PROCESS ANALYSIS AND SYNTHESIS)

into several frequency-band channels. (In fact, the channel separation is obtained by time-domain considerations. For the purpose of rhythm extraction, frequency-band separation works just as well.) Two or three channels are enough, but we use five in preparation for the problem of style transformation.

A convenient channel-frequency division and bandwidth are indicated in the block diagram of Fig. IX-1. For rhythm extraction, this separation is not critical. It is used here for its application to the problem of style transformation.

The objectives of channel separation in connection with the extraction of rhythm are:

First, to increase the probability that certain maxima and minima belong to the rhythm patterns.

Second, to increase the probability associated with the qualitative elements of the rhythm patterns.

Third, to make possible the extraction of the rhythm during the first few bars of the meter.

3.32 Second Step: Envelope Extraction

The problem of rhythm separation, as solved by maxima and minima methods, is a typical envelope problem. To be more specific, the maxima and minima we have been talking about belong to or are associated with the envelope. We need three envelopes for each channel. Two are basic; the third is auxiliary.

The following operations must be performed with each channel-output function $f_k(t)$; $k = 1, 2, \dots, 5$. See Fig. IX-2. (The graphs of the $f_k(t)$ are merely symbolic sketches.) After $f_k(t)$ suffers double rectification into $\psi_k(t)$, then $\Psi_k(t)$ is smoothed into three components. In the first smoothing, frequencies in the envelope of 40 cps or higher do not show, but 12-cps frequencies appear in the envelope. In the second smoothing, 12 cps are practically smooth, but 4 cps, or so, show. The auxiliary envelope suffers a strong smoothing up to 10 cpm.

3.33 Third Step: Maxima-Minima Separation

By means of a differential dc amplifier and appropriate level adjustments, we form the function $\lambda_k(t) = \theta_k(t) - \eta_k(t)$. The positive part of this curve is associated with the relative maxima of the $\theta_k(t)$, and the negative part is associated with the relative minima. By a second rectification process, we obtain two new functions $M_k(t)$ and $m_k(t)$, which represent the separation of the relative maxima and minima of the function $\theta_k(t)$. The functions $M_k(t)$ and $m_k(t)$ have a simple interpretation. By a proper selection of the time constant of the smoothing circuit, we can associate $M_k(t)$ with the number of notes played and the function $m_k(t)$ with the number of silence intervals. This is not absolutely true; nevertheless it is approximately so.

The next step tends to emphasize this association. It may be noticed that the

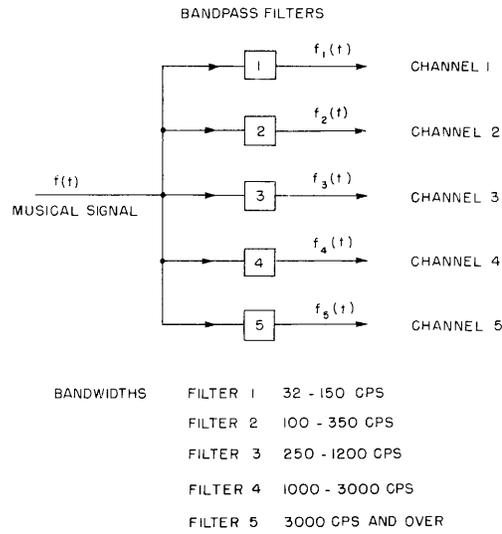


Fig. IX-1. Channel separation of the incoming musical signal. Separation of the music into channels will help to extract the rhythm of the music after the first few beats. These frequencies are not critical.

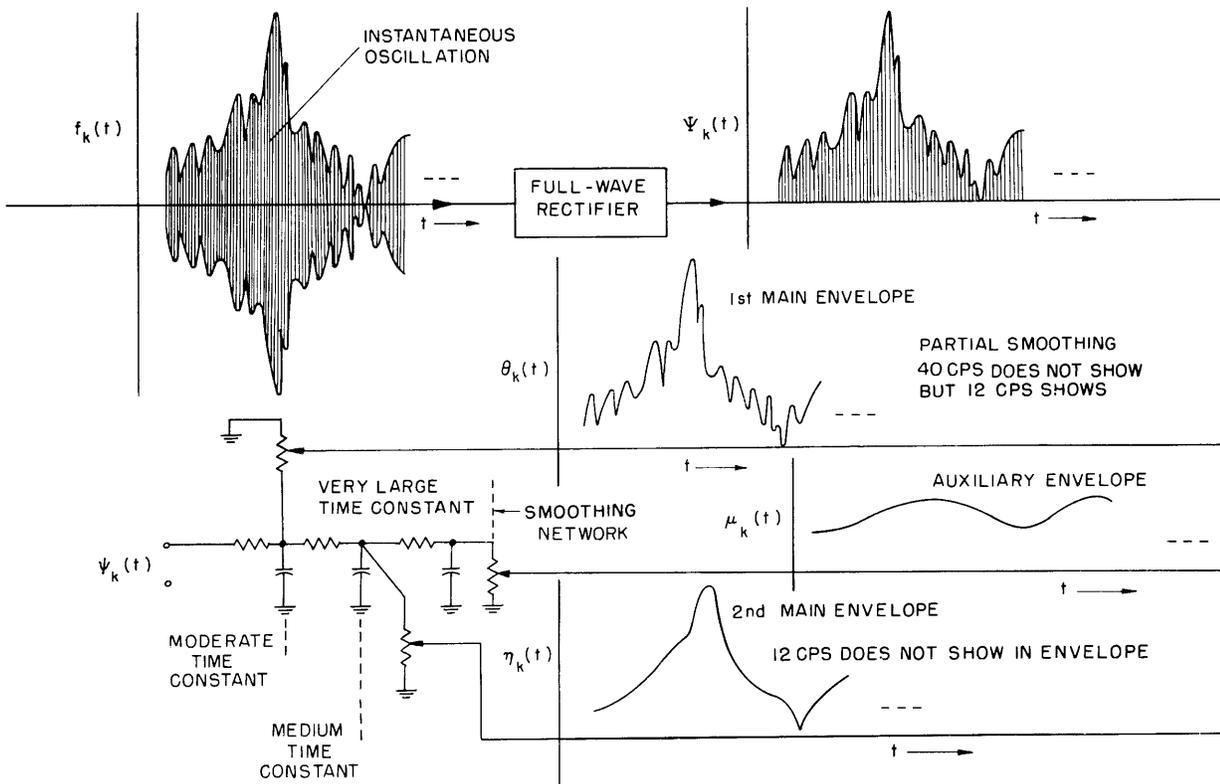


Fig. IX-2. Main and auxiliary envelopes. Since the time constants of the integrating circuits are different, there is a shift between the two main envelopes.

(IX. PROCESS ANALYSIS AND SYNTHESIS)

functions $\theta_k(t)$ and $\eta_k(t)$, which represent the two main envelopes, are shifted in time because of the different time constants of the corresponding smoothing circuits. The shift may offset our interpretation of the association of $M_k(t)$ and $m_k(t)$ as the number of notes played and the silent intervals of the notes. In practice, however, this time shift has little effect, provided that the smoothing is not excessive. See Fig. IX-3. Note that the envelope $\mu_k(t)$ will find later application.

3.34 Fourth Step: Emphasis and Slenderizing of $M_k(t)$ and $m_k(t)$

The graphs of the functions $M_k(t)$ and $m_k(t)$ are drawn with an exaggerated shape that is used to accent their characteristic features and make their interpretation more lucid. In general, these graphs, although they show a strong oscillating character, do not show the separation of peaks very well. The operations, which we now introduce, tend to:

- i. increase the contrast between the magnitudes of the peak values of $M_k(t)$; and the same for $m_k(t)$.
- ii. slenderize the peak form of the graphs.

These objectives are accomplished, for example, by a successive squaring process. Thus, we have

$$M_k(t) \rightarrow M_k^2(t) \rightarrow M_k^4(t) \dots \rightarrow M_k^{2n}(t)$$

$$m_k(t) \rightarrow m_k^2(t) \rightarrow m_k^4(t) \dots \rightarrow m_k^{2n}(t)$$

In practice, one or two squaring operations are required (particularly, when we select a convenient point of operation for the vacuum tubes of the squaring circuits).

Let us write

$$S_k(t) = M_k^{2n}(t)$$

$$s_k(t) = m_k^{2n}(t)$$

These functions are called the "note distribution" and the "silence distribution" spectra. Figure IX-4 shows graphs of this spectral distribution. From the effect of the successive squaring operation, the peaks become sharper. For the sake of simplicity, we use lines instead of peaks.

The reader may note that the functions $S_k(t)$ and $s_k(t)$ have a much stronger definitive association with the number of notes and intervals of silence than the original functions $M_k(t)$ and $m_k(t)$.

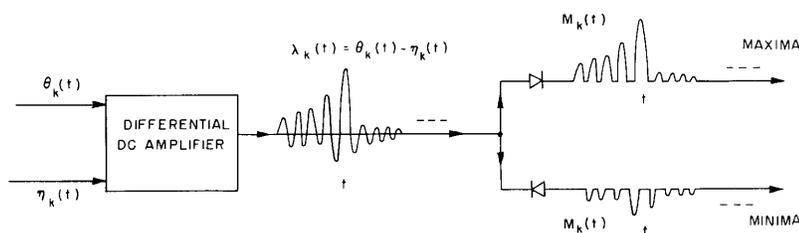


Fig. IX-3. The functions $M_k(t)$ and $m_k(t)$ associated with the number of notes played and the number of silence intervals.

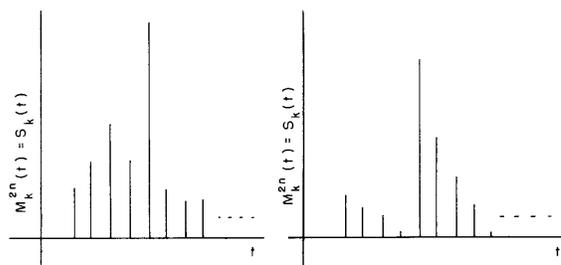


Fig. IX-4. Sound and silence distribution spectra. Symbolic representation.

3.35 Fifth Step: Preliminary Operation Tending to Separate the Quantitative and Qualitative Elements of Musical Rhythm. The Emphasis of Sound and Silence Subspectra

Clearly the different heights of the lines representing $S_k(t)$ and $s_k(t)$ can be associated with the emphasis and de-emphasis (or silence) of the musical sound. The separation of the quantitative and qualitative elements can be obtained by a simple process of clipping $S_k(t)$ and $s_k(t)$ at a certain noncritical level. We thus obtain the graphs of the functions $R_k(t)$, $r_k(t)$, $E_k(t)$, and $e_k(t)$. See Fig. IX-5. Note that the clipping levels for $S_k(t)$ and $s_k(t)$ are not necessarily the same, which is quite convenient.

3.40 Number of Notes Played. Number of Silence Intervals. Basic Emphatic Elements. Determination of the Bar Duration. Dividers and Time Signature. Basic Rhythm Extraction

We have obtained the extremal distribution spectra $R_k(t)$, $r_k(t)$, $E_k(t)$, $e_k(t)$ for each channel. The objective of this section is to determine the basic elements of the meter from these functions: bar duration, basic emphasis, time signature, and, finally, the rhythm. To achieve this we compare the aspects of the spectral functions that correspond to different channels, and also perform convenient operations, following, for the most part, Boolean algebra. There are three basic steps in our plan of attack.

First, the lines of the spectral functions $R_k(t)$, $r_k(t)$, $E_k(t)$, $e_k(t)$ are generated by

(IX. PROCESS ANALYSIS AND SYNTHESIS)

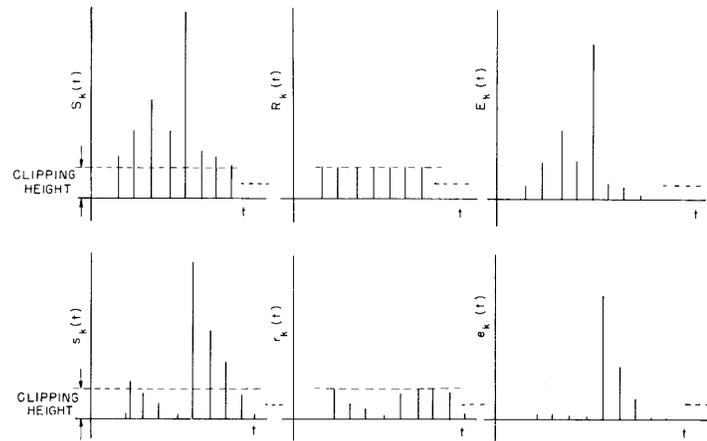


Fig. IX-5. Sound distribution spectra $R_k(t)$, sound emphasis spectra $E_k(t)$, silence distribution spectra $r_k(t)$, sound de-emphasis spectra $e_k(t)$.

the metric elements of the music as well as by the different degrees of loudness in the composition. Since we are interested in the reconstruction of the metric elements, we must perform appropriate operations which tend to measure the probability that a certain line of a spectral distribution is, or is not, attributable to the metric effect in the music. In other words, we must increase the probability that certain lines in $R_k(t)$, $r_k(t)$ are due to the quantitative elements of the meter, and certain lines in $E_k(t)$ and $e_k(t)$ are due to the qualitative elements of the meter. The reader can appreciate the necessity of this step by considering, for example, the function $R_k(t)$ in a long time interval that is equal to a large number of bar durations and realizing that the number of spectral lines that are present in successive subintervals, say each of one bar length, is not necessarily the same, either in number or in position. A similar situation is encountered with respect to $E_k(t)$, $r_k(t)$, and $e_k(t)$. If every spectral line came from the effect of the meter only, then graphs of these spectral lines would be periodic, having the bar length for the period.

Second, once the spectral lines are classified in increasing probability order, it is necessary to isolate the periodic structures described above.

Third, the methods that tend to achieve the objectives described in the first and second steps must include the condition that the required operation last not more than a few bars.

[Editor's note: This discussion will be continued in the Quarterly Progress Report of Oct. 15, 1957.]

M. V. Cerrillo

References

1. For definitions of Besicovitch spaces see: A. C. Zaenlen, *Linear Analysis* (Interscience Publishers, Inc., New York, 1953), pp. 103-108 for Banach type space; pp. 121-122 for Hilbert type space; cf. A. S. Besicovitch, *Almost Periodic Functions*, Cambridge Mathematical Tracts (Cambridge University Press, London, 1932; reprint Dover Publications, Inc., New York, 1955).