Style as Information
Author(s): Joseph E. Youngblood
Source: *Journal of Music Theory*, Vol. 2, No. 1 (Apr., 1958), pp. 24-35
Published by: Duke University Press on behalf of the Yale University Department of Music
Stable URL: http://www.jstor.org/stable/842928
Accessed: 18-11-2015 05:45 UTC

STYLE AS INFORMATION

by

Joseph E. Youngblood
Indiana University

The purpose of this paper is to explore the usefulness of information theory as a method of identifying musical style.[1] The particular stylistic aspect considered is melody, although any number of other aspects could have been chosen.

The first part of the paper will be concerned with describing information theory and relating it to musical practice. The most fruitful applications of information theory have probably been in the field of communication[2] and in the field of language[3]. The basic assumptions

---

1. Since the completion of the present paper, the writer has encountered Richard C. Pinkerton's article "Information Theory and Melody," (Scientific American, Vol. 194, 2:77-87). The treatment of the data in the two papers is virtually the same, although there is a basic difference of approach between them. Mr. Pinkerton's work began as an inquiry into the melodic characteristics of nursery tunes, and information theory was chosen, presumably from among several possible statistical methods, as an analytical procedure. The present paper, on the other hand, began as an inquiry into the usefulness of information theory as an analytical device for music, and melody was selected, again from among several possible choices, as the aspect upon which to try it. Mr. Pinkerton's article apparently presupposes a fairly high degree of statistical and mathematical sophistication on the part of his readers, and a rather large part of the paper is given over to discussing ways in which nursery tunes (as well as Shostakovitch tunes) might be mechanically created. Mr. Pinkerton's article is recommended to those readers interested in the subject, since it presents an analysis of a third style, to which the analyses of the two styles in this paper may be compared.

2. Especially Claude E. Shannon and Warren Weaver, The Mathematical Theory of Communication (Urbana, University of Illinois Press, 1949).

3. Charles F. Hockett, in his review of Shannon and Weaver in Language, XXIX, 69-93, suggests several applications. See also E. Colin Cherry, ed., Information Theory (New York, Academic Press, Inc.; London, Butterworths Scientific Publications; 1956), especially pp. 149-153 (D. A. Bell and Alan S. C. Ross, "Negative Entropy of Welsh Words") and pp. 154-170 (Wilhelm Fucks, "Mathematical Theory of Word Formation"); Charles E. Osgood and Thomas A. Sebeok, eds., "Psycholinguistics," International Journal of American Linguistics, Memoir 10 (Baltimore, Waverley Press, 1954), pp. 35-49 (by Kellogg Wilson) and pp. 103-110 (by John B. Carroll); E. Colin Cherry, Morris Halle, and Roman Jakobson, "Toward the Logical Description of Languages in Their Phonemic Aspect," Language, XXIX, 34-46.

underlying this paper are that both music and language are communication systems and that a technique which has proved useful in describing the latter should be applicable to the former.[4] Communication, in the engineering sense, refers to the problems involved in reproducing at one point a message selected at another.[5] The message may be one end of a telephone conversation or a Beethoven symphony, inasmuch as both are sequences selected from or generated by a finite set of symbols — the speech sounds of English or the tones of our musical system. The aspect of the problem which is of particular interest here, however, is not the mechanical one of telephone lines or room acoustics but rather the system of symbols itself and the constraints which it imposes on possible messages.

A composer, setting forth to write a symphony, finds himself confronted by twelve tonal contrasts which may be combined in an infinite number of ways.[6] Because of aesthetic, traditional, or technical pressures, a composer rarely uses more than a small fraction of these possible combinations; he reveals himself as predisposed to certain combinations and quite uninterested in others. This paper will attempt to determine the extent of the restrictions under which certain composers worked; it will not attempt to explain why certain combinations were favored and others eschewed.

In information theory, information refers to the freedom of choice which a composer has in working with his materials or to the degree of uncertainty which a listener feels in responding to the results of a composer's tonal choices. Low information means high predictability and, conversely, high predictability results in low information. Thus, when we read English, we know that virtually every occurence of q is followed by an occurence of u. Writers of English feel a high degree of constraint in their use of q since it virtually imposes upon them the use of u as the next letter. Readers of English, on the other hand, receive virtually no information from the use of u following a q since this use of u was highly predictable. That this predictability is not 100 % is demonstrated by the previous sentence in which q was used without a succeeding u. Similarly, if the leading-tone were always followed by the tonic, we would predict the latter always from an occurence of the former. We cannot say that q is always followed by u, nor can we say that the leading-tone is always followed by tonic. But we can say that the probability of both is quite high in normal English language situations and in traditional harmonic styles respectively.

If information is freedom of choice, then it must be a function of the probability of what might follow. Although the frequency with which

---

4. Shannon and Weaver, op. cit., p. 95, suggest that this is true, as does George P. Springer, "Language and Music: Parallels and Divergencies," For Roman Jakobson (The Hague, Mouton & Co., 1956), p. 508.

5. Shannon and Weaver, op. cit., p. 3.

6. The possibilities are infinite, since there are no restrictions on the length of the combinations. The possible two-tone combinations are certainly finite, and are equal to 12^2, or 144.

the leading-tone is followed by tones other than the tonic tells us something about the way in which it is used, it would seem useful to have a way of measuring this restraint and comparing it with the freedom with which, for example, the dominant is used. Mathematics has developed a method for measuring information which might also work for the analysis of musical styles.

If one is confronted with a situation in which he must choose between two messages (or events), and either choice is equally likely, then it is arbitrarily said that the information associated with this situation is unity or one[7]. We have seen above that messages may be equated to sequences of musical tones. Let us take a very elementary example. A man is confronted with two drums, a large one (O) and a small one (o). This is the situation which is characterized by one unit, or bit, of information. The reason for the use of this as a measure of information will become apparent as we proceed. If there is nothing to influence the man otherwise, the probability of his striking either drum is, obviously, 50-50 or 1/2 or 0.5, and the possible results are O or o. If he be allowed to strike the drums twice, then the possible messages which he might emit (or the events which might occur) would be doubled, since he might perform any of the following sequences: OO, Oo, oO, oo. The information, or freedom of choice, or degree of uncertainty, associated with this source (in this case, the drums) would also be doubled and would be 2 bits, and the probability of any sequence would be 0.25. This is equivalent to a situation in which the man is given four drums but is allowed to strike only once — 2 bits of information, and probability 0.25 for each drum. Finally, if the man be allowed to strike at the two drums three times, the original information should be trebled. But we see that the possible outcomes of such a performance are eight: OOO, OOo, Ooo, OoO, oOo, oOO, ooO, ooo. It now becomes apparent that the measure of information (number of bits) is progressing on a logarithmic scale; that is, the amount of information in a situation in which all of the results are equally likely is equal to the logarithm (to the base 2 in this case) of the number of choices. The possible number of messages in the above examples involving two drums were 2, 4, and 8, and $\log_2 2 = 1$; $\log_2 4 = 2$; $\log_2 8 = 3$.

Information, or uncertainty, can be increased in two ways: by an increase in the length of the message (effected by increasing the number of strokes) or by an increase in the number of symbols in the source (in this case, the number of drums). We should now be able to assume that sixty-four different combinations would be possible with six strokes; or, which amounts to the same thing, a man with sixty-four drums, of which he may strike only one, would confront a listener with six times as much uncertainty as to the outcome of the action as a man with only two drums. Of course, if he has only one drum, there is no uncertainty, no freedom of choice, and consequently no information, since we know in advance what the message is going to be, and there is nothing to be gained, at least in a communication sense, by his striking it at all.

7. Shannon and Weaver, op. cit., p. 100.

In the situations thus far mentioned, the person beating on the drums was completely free to choose the drums he hit, and it was assumed that he was not biased toward any particular ones. In the situations of primary interest to musicians this is probably not the case, since it is probably true that a composer's style, be it melodic, harmonic, or rhythmic, is recognizable because of its emphasis on certain progressions or other patterns. The next step, therefore, is to investigate the measurement of information in situations in which the choices are not equiprobable but which occur with certain probabilities deduced from the frequencies observed over a period of time.

A source which produces a sequence of symbols (which may be musical tones symbolized by notes) according to certain probabilities is called a stochastic process.[8] For a composer's style to be anything but stochastic would necessitate wholesale variations in it from time to time, in which case it would probably be unrecognizable as a style owing to its very lack of homogeneity. Later we shall observe the frequencies with which three composers use the twelve tones of the chromatic scale and the variations which they show in their usages (see Table I, page 32). One type of stochastic process, the Markov chain, is of particular relevance to our problem. A Markov chain is a sequence of events in time in which each event has a calculable probability. Furthermore, this probability is the result not only of the frequency of the event in the set of events under consideration but also of the effect of those events which immediately preceded it on the event of the moment. The English language situation cited earlier is an example of a Markov process. The letter u normally has a relatively low probability in written English. However, in a sequence of letters involving a q its probability increases to very nearly 100 % at that moment.

A random sampling of Beethoven's music would reveal that, on the average, a certain number of the tones used are tonics; but we would learn more about Beethoven's style if, regarding it as a Markov process, we determined what fraction of these tonics is preceded by leading-tones and what fraction is preceded by other scale steps. A table showing the frequency with which each pitch is followed by each other pitch (see Table II, page 33) is called a "matrix of first-order transition probabilities."

The freedom of choice in a situation wherein each of four drums has a probability of 0.25 is markedly different from another situation in which one drum has a probability of 0.5, a second of 0.25, and the third and fourth of 0.125 each. It will be observed that the sum of the probabilities of a set of choices is always equal to one. Reason tells us that freedom of choice, and therefore information, is greatest when all choices are equal, i.e., when all probabilities are equal to the

8. This is the operational definition of "stochastic process" given by Shannon and Weaver, op. cit., pp. 10, 102. A more general description of the process is given by W. Ross Ashby, An Introduction to Cybernetics (New York, John Wiley & Sons, 1956) p. 162, in which it is contrasted with other, non-stochastic, processes.

reciprocal of the number of choices, and that information must decrease as any one of the probabilities increases toward one, since the other probabilities will, at the same time, decrease toward zero. When probability "1" is reached by any choice, all information, freedom of choice, and uncertainty vanishes, and a situation is reached equivalent to that in which a drummer has only one drum. Obviously, the formula for measuring information in bits as given above (= $\log_2$ of the number of choices) is incomplete, since the number of drums remains constant, whether the probabilities of their being struck are equal or not.

In order to compensate for the inequality of probabilities, Shannon[9] has introduced the quantity of entropy (H), which, for our purposes, is equivalent to information. Entropy is found by multiplying the logarithm of each probability by that probability and summing the results.[10] Thus, in the first example above, the calculation would be (assuming logs to the base 2):

$$(1)\ H = -(0.25 \log 0.25 + 0.25 \log 0.25 + 0.25 \log 0.25 + 0.25 \log 0.25) = 2$$

which is the same result as achieved by the former means. In the second case, the calculation would be:

$$(2)\ H = -(0.5 \log 0.5 + 0.25 \log 0.25 + 0.125 \log 0.125 + 0.125 \log 0.125) = 1.75$$

which is, as predicted in the previous paragraph, smaller than the amount when the probabilities are equal.

It is generally helpful to compare the entropy of a source with that which it would be if all the choices were equal; their ratio is the relative entropy (H_r). Quantity (1) above represents the maximal amount of information achievable with four drums; the ratio of 1.75 to 2 is 0.875, which means that in the second situation the drummer is 87.5% as free as he possibly could be to choose between the drums that he is using. And what of the other 12.5%? This figure is the redundancy (R) or the restraint on the drummer — that portion of the message which is predetermined by the source.

It is important to avoid attaching a pejorative connotation to the word "redundancy," since it is merely a descriptive term in communi-

9. Shannon and Weaver, op. cit., p. 20.

10. This sentence may be expressed mathematically, as

$$H = -(P_1 \log P_1 + P_2 \log P_2 + \ldots + P_n \log P_n)$$

or

$$H = -\sum P_i \log P_i$$

in which P_i represents each of the probabilities in turn. Since probabilities are always numbers less than 1, their logarithms are always negative. The minus sign is introduced by Shannon in order to make the total positive.

cation theory, a measurement of the assurance that the message will be received. An extreme example of linguistic redundancy is shown by the Latin phrase "per Jesum Christum Filium tuum Dominum nostrum," in which the accusative case is marked six times; whereas, if it be assumed that it is necessary to mark words for accusative at all (an assumption which the English language does not make), one marking should be sufficient for communication purposes. It is not necessary to inquire into the reasons for any particular set of restraints being preferred over any others — the pattern of choices might be influenced by the placement of the drums, the aesthetic sense of the drummer, or the presence of a weak head on one drum; it is necessary here only to note and to measure these redundancies. The familiar rhythmic example, <u>boom di-dee boom boom</u> ____ ____, can be completed by virtually anyone, as can the progression, I II^6_5 V_{87} __, and, in this sense, the final elements of both are redundant.

Even though given the preceding chords most musicians can accurately guess the last chord of a Beethoven symphony, few of us would approve of a conductor's stopping an orchestra on the penultimate chord in an effort to cut out at least some of the redundancy inherent in Beethoven's style. One of the things which a piece of music communicates to a musician is information about the style. A certain amount of redundancy is necessary for the transmission of this information. Rather than seeking to eliminate this redundancy, we are interested in finding and measuring it.

The approach described above has been applied in a preliminary way to an analysis of melody. A corpus of twenty songs was analyzed: eight songs from <u>Die Schöne Müllerin</u> by Schubert, six arias from <u>St. Paul</u> by Mendelssohn, and six songs from <u>Frauen-Liebe und Leben</u> by Schumann. The frequency of each of the twelve tones was recorded and the first-order transition probabilities were computed. From these data, the entropy, relative entropy, and redundancy in each category were computed, according to the formulas given above.

The study was limited to melodies in major keys. Once the overall tonality was determined, the whole of each piece was analyzed in that key, that is, if a melody were determined to be in the key of F, F remained the tonic (and was listed as tone I) throughout the piece, regardless of any modulations which might occur. All melodies were considered as proceeding from an enharmonic system of twelve tones,[11] so that the tonic was always I, the raised tonic or lowered supertonic always II, the supertonic, III, etc. The aspect of rhythm was not considered in this study nor were the aspects of phrasing or word-painting.

It might be suspected <u>a priori</u> that Mendelssohn would tend to be the most restricted composer of the three, and this suspicion is borne

11. There may be several factors which influence a composer's choice between enharmonic equivalents, but one of these factors is probably not pitch.

out by the fact that his works show the highest degree of redundancy. Although the P_i[12] redundancy is not much greater than Schumann's, Table I shows that Mendelssohn uses the chromatic degrees of the scale much more sparingly than does Schumann. Schubert, while using the chromatic tones more frequently than Schumann, appears to be more rigid in his usage of them when they are considered as units, as in the frequencies of P_{ij} (see Table II). Furthermore, a great deal of Schubert's apparent chromaticism is a result of modulation, and combinations which occur frequently in the tonic key seem also to be common in other tonal areas into which he moves. Movement from the tonic to the supertonic, which in the tonic key is indicated by I to III, occurs frequently in the key of the lowered submediant, where it is shown as IX to XI. The chromaticism in the Schumann examples, on the other hand, is as likely to be a result of decoration as it is of modulation, as shown by the slightly more random distribution of the two-tone combinations.

In spite of these small variations, however, the overall figures for the three composers seem quite similar.[13] This is probably due in part to the fact that the three are virtual contemporaries and as such cannot be expected to show dramatic divergencies in style. That the three are different is also doubtlessly true, and an analysis of the three from the harmonic point of view, which could also be done by the techniques described in this paper, would probably bring out these differences more clearly.

For purposes of comparison, this same type of analysis was carried out on four selections of Gregorian chant.[14] The four selections were all assigned to Mode I in the Liber Usualis; the steps were numbered diatonically, starting with the final (D) as I and proceeding through the sub-final as VII.[15] The results of this analysis are given in Tables III and IV, page 34.

---

12. In the tables accompanying this paper, P_i refers to the probability of any tone occurring; P_{ij} to any two-tone combination, and $P_{i(j)}$ to a tone (i) being followed by a tone (j). H = entropy; H_r = relative entropy; R = redundancy, expressed as a percentage. In Table II and Table IV, the rows involve tones (i); the columns involve tones (j).

13. The real significance of these deviations cannot, of course, be determined until a range is established through the analysis of much more music.

14. The selections analyzed are the Gloria, Sanctus, and Agnus Dei from the first Mass for Solemn Feasts (Liber Usualis, pp. 19-22), and the Kyrie from the Mass Orbis Factor for Sundays throughout the Year (Liber Usualis, p. 46).

15. Although both B and B^b occur as the sixth degree of Mode I, they can be considered alternate forms (or conditioned variants) of the same structural pitch-area. A linguistic analogy is provided by the initial sounds of the English words "key" and "cool." The two are different from an articulative point of view, since the former is produced with the lips spread and the tongue forward, whereas the latter is produced with the lips rounded and the tongue drawn back. Structurally, however, the two are conditioned variants of the same sound. Cf.

Viewed as being the product of a seven-tone diatonic system, the chant shows a high degree of entropy. Although the final and dominant appear most frequently, the remaining tones are more evenly distributed than the chromatic degrees in the first group of composers. It is important to keep in mind that these figures refer only to chants in the first mode from the Ordinary of the Mass, even as the previous figures referred only to songs in a major key; analysis of the psalm-tones would doubtlessly have given different results as would the analysis of piano works in a minor key.

It is also interesting to note how these figures appear when the chant is viewed as progressing from a twelve- rather than a seven-tone system. The writer does not feel that this is completely unjustified, since modern listeners are prepared to respond to twelve divisions of the octave, and consequently, maximum uncertainty is for them represented by $\log_2 12$ and not by $\log_2 7$. The redundancy shown in Table V, page 35, is on the whole higher than that of the romantic composers considered earlier, but even then the average redundancy in the first-order transition probabilities is still slightly lower than the average for Mendelssohn.

As was indicated earlier, some means of including the effects of such obviously important factors as rhythm, position in phrase, modulation, and word-painting would have to be found before this system could be accurate even for melody alone. It has not been tested on the more difficult, but possibly more fruitful, aspect of harmony. Most musicians can at present either intuitively or on the basis of certain vague generalizations identify at least five or six historical styles. It seems, however, that it would be useful to find a means of identifying and quantifying the characteristic features of a style, as well as measuring the differences between styles, if for no other reason than to provide a basis for understanding and evaluating contemporary music.

Charles F. Hockett, "A Manual of Phonology," International Journal of American Linguistics, Memoir 11 (Baltimore, Waverley Press, 1955), p. 35. That the distinction between B and B^b ultimately became structurally relevant in Western music seems incidental to a consideration of Gregorian chant.

Table I

Tonal Frequency in Melodies of
Schubert, Mendelssohn, and Schumann

Tones	Schubert	Mendelssohn	Schumann
I	182	103	215
II	7	4	16
III	168	84	148
IV	23	6	13
V	124	84	144
VI	83	52	66
VII	16	7	22
VIII	203	104	208
IX	30	7	5
X	78	68	118
XI	29	8	23
XII	82	50	88
	1025	577	1066
	$H = 3.127$	$H = 3.03$	$H = 3.05$
	$H_r = 0.87$	$H_r = 0.846$	$H_r = 0.85$
	$R = 13\%$	$R = 15.4\%$	$R = 15\%$

Matrix of First-order Transition Probabilities for
Schubert (St), Mendelssohn (Mn), and Schumann (Sn).

Tone (j)		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	H	H_r	R %
I	St	29	0	27	5	33	3	1	26	1	7	10	40	2.89	0.807	19.3
	Mn	29	0	12	1	6	6	1	11	2	7	2	26	2.81	0.785	21.5
	Sn	86	1	33	0	24	9	1	18	2	6	8	27	2.59	0.72	28
II	St	5	1	0	0	1	0	0	0	0	0	0	0	1.15	0.321	67.9
	Mn	0	1	3	0	0	0	0	0	0	0	0	0	0.81	0.226	77.4
	Sn	1	4	7	0	2	0	0	0	0	1	0	1	2.15	0.60	40
III	St	57	1	38	4	15	13	2	13	0	7	0	18	2.68	0.750	25.0
	Mn	23	3	24	2	13	5	0	5	0	4	0	5	2.69	0.752	24.8
	Sn	31	3	50	3	26	4	0	10	0	11	1	9	2.65	0.74	26
IV	St	3	2	11	2	0	0	0	3	1	0	1	0	2.28	0.638	36.2
	Mn	0	0	2	2	1	0	0	1	0	0	0	0	1.92	0.536	46.4
	Sn	0	0	1	0	10	0	0	0	1	1	0	0	1.15	0.32	68
V	St	24	0	41	0	13	4	1	36	0	5	0	0	2.25	0.629	37.1
	Mn	14	0	25	1	17	16	1	5	1	2	0	2	2.60	0.726	27.4
	Sn	32	1	37	0	34	5	5	18	0	12	0	0	2.54	0.71	29
VI	St	1	0	24	8	36	2	0	6	1	5	0	0	2.17	0.606	39.4
	Mn	1	0	2	0	26	14	0	5	0	2	0	2	1.98	0.553	44.7
	Sn	2	4	3	0	20	24	0	9	0	4	0	0	2.29	0.64	36
VII	St	1	0	0	0	1	0	7	7	0	0	0	0	1.54	0.430	57.0
	Mn	0	0	0	0	3	0	0	2	0	2	0	0	1.56	0.346	65.4
	Sn	1	0	0	0	1	4	4	10	0	0	0	2	2.19	0.61	39
VIII	St	33	2	12	3	22	39	6	58	7	5	2	14	2.92	0.816	18.4
	Mn	20	0	6	0	6	10	6	41	0	13	0	2	2.51	0.701	29.9
	Sn	31	0	7	8	16	18	9	89	1	17	0	12	2.64	0.74	26
IX	St	3	0	0	1	1	0	0	10	5	4	6	0	2.47	0.690	31.0
	Mn	1	0	0	0	2	0	0	0	0	4	0	0	1.38	0.385	61.5
	Sn	1	0	0	0	0	0	0	2	0	2	0	0	1.52	0.42	58
X	St	3	0	6	0	0	18	0	19	4	25	0	3	2.38	0.665	33.5
	Mn	7	0	7	0	4	0	0	25	3	14	2	6	2.57	0.718	28.2
	Sn	8	2	7	0	3	1	3	36	1	35	3	19	2.58	0.72	28
XI	St	2	1	0	1	0	1	0	1	10	3	9	1	2.49	0.696	30.4
	Mn	0	0	0	0	0	0	0	0	0	4	4	0	1.00	0.279	72.1
	Sn	1	0	2	1	0	0	0	1	1	8	9	0	2.16	0.60	40
XII	St	28	0	9	0	1	2	0	20	0	15	2	5	2.41	0.673	32.7
	Mn	10	0	3	0	5	2	0	9	0	16	0	7	2.41	0.673	32.7
	Sn	28	0	2	1	9	0	1	8	0	21	2	16	2.51	0.70	30
											Average		St			35.7
													Mn			44.3
													Sn			39
											P_{ij}		St	5.37	0.75	25
													Mn	5.34	0.745	25.5
													Sn	5.52	0.772	22.8

Table III

Tonal Frequency for Gregorian Chant Considered as a Seven-tone System.

Tone	Frequency
I	156
II	84
III	92
IV	98
V	110
VI	41
VII	77
	658
H	2.720
H_r	0.969
R	3.1 %

Table IV

Matrix of First-order Transition Probabilities for Gregorian Chant Considered as a Seven-tone System.

Tone	I	II	III	IV	V	VI	VII	H	H_r	R %
I	50	21	17	11	9	1	43	2.355	0.858	14.2
II	40	9	24	8	0	0	3	1.867	0.665	33.5
III	22	37	3	25	5	0	0	1.923	0.685	31.5
IV	2	0	41	10	40	4	1	1.758	0.626	37.4
V	14	0	4	35	23	22	12	2.362	0.841	15.9
VI	0	0	0	0	24	6	11	1.366	0.487	51.3
VII	28	16	4	4	5	14	6	2.436	0.868	13.2
					Average					28.1
					P_{ij}			4.695	0.836	16.4

Table V

Redundancy of Gregorian Chant
Considered as a Twelve-tone System
Compared with the Average Redundancy of
Schubert, Mendelssohn, and Schumann.

	Gregorian Chant	Romantic Composers
R_i	24.1 %	14.5 %
R_{ij}	34.4 %	24.4%
$R_{i(j)}$	43.9 % (average)	39.3 %