A MATHEMATICAL MODEL FOR PERCEPTION APPLIED TO THE PERCEPTION OF PITCH

David Rothenberg
Inductive Inference, Inc.

Abstract

A mathematical model for perception which derives from a theory of efficient data representation in the central nervous system is described. A possibly infinite and continuous space, S (over which sensory stimuli range), is mapped into a finite space of discrete points, C (the indices on the classification of such stimuli). An ordering of pairs of points in $S \times S$ together with either the number of classifications required or a maximum tolerable error is given and assumed to derive from feedback (experience). The model chooses a "characteristic" finite subset, P, of the stimulus space, which defines a function and a range about each element of the subset such that the union of all such ranges is maximal and that the function provides a metric (on C) which preserves the given ordering. Restrictions on the choice of P derive from limitations in the information carrying capacity of the resulting classification system. A context-dependent "Gestalt" description of perception results in which extremely complex and varied phenomena can be perceived without proportionately large human memory. For each application specific distortions of perception in specified contexts are predicted. The system, in some aspects, resembles an hierarchical clustering scheme. It is hence useful for representing different patterns of satisfaction of several "features" in a pattern recognition scheme by a single set of integers with metric properties which reflect relevance to the task (i.e., an n-valued logic replace a two-valued logic).

1. Proper Mappings[1]

Consider S to be the retina of the eye. Let the pairs of points $(S \times S)$ be ordered by the endpoints of the projection of a rigid rod on the retina of the eye. Such projection would alter with rotation of the rod. Since we know the rod is rigid, all such projections would be classified as equivalent with respect to size. The projections of another rod, which experience has taught us is longer, would form a different classfication and an ordering relation would exist between the two classifications. Of course, in binocular vision each point in S would be a duple (each element of a duple from one endpoint of the projection on each retina) and $S \times S$ would be a pair of such duples.

A relation, $\leq$, called the initial ordering, is defined on $S \times S$ which is a preorder, i.e. transitive, connected and reflexive. Define $(x,y) \sim (z,w)$ to mean $(x,y) \leq (z,w) \wedge (z,w) \leq (x,y)$ and $(x,y) < (z,w)$ to mean $(x,y) \leq (z,w) \wedge \sim \left[(z,w) \leq (x,y)\right]$. Require that always $(x,y) \sim (y,x)$.

[1] See Section 10 for theoretical basis (which may be read before Section 1).

It is assumed that $S \times S$ is mapped into the code, C, by some function, which is dependent only upon the preordering of $S \times S$ (i.e. contains no statistical weights or other arbitrary parameters). Since S is very large (possible infinite in the mathematical model), it is assumed that human memory cannot contain a preordering of $S \times S$. The mapping into the code must therefore depend upon much less stored information. The first step in the strategy of the model is to find a finite subset P of S and a function, f, such that f is dependent only on the preordering of $P \times P$ and $f: P \times P \rightarrow C$. (Consider C to the the set, $\{1, \ldots, n\}$, of integers.)

f is specified in the following manner: let ab denote the pair $(a,b) \in P \times P$. We choose some pair, $\overline{xy} \in P \times P$, called the link size, such that iff $ab \leq \overline{xy}$, a and b are defined as adjacent. The choice of xy is determined by the desired cardinality of C or by a given tolerance — to be described. A sequence of elements in P, (a, b, c, d...) such that a is adjacent to b, b to c, c to d, etc., is called a chain. $f(a,b)$; $a,b \in P$, is now defined as the cardinality of (the number of elements in) the smallest chain connecting a and b, minus one (i.e., the number of edges). When P is chosen it is thus necessary for the model to know (remember) only which elements in P are adjacent in order to map $P \times P$ into C. (It may sometimes be necessary to add "ideal" points to P which are not in S in order that $P \times P$ be connected. This is analagous to the brain's "filling in" images in locally damaged portions of the retina.)

Different possible choices of $P \subset S$ and link size are classified according to which of the following stated properties are satisfied:

(a) $ab > cd \rightarrow f(a,b) \geq f(c,d)$

(b) $ab \sim cd \rightarrow f(a,b) = f(c,d)$.

If (a) is satisfied both f and P are defined as proper; if both (a) and (b) are satisfied f and P are called strictly proper; if f and P are not proper they are called improper. ab is called an ambiguous pair iff there exists a pair $cd \in P \times P$ such that $ab \sim cd$ and $f(a,b) \neq f(c,d)$. Both ab and cd are called contradictory iff $ab < cd$ and $f(a,b) > f(c,d)$.

For reasons which become obvious if (a) and (b) are examined,[2] it is hypothesized that proper sets, P, correspond to "Gestalts" or "reference frames". It is also easily seen that when P is strictly proper f provides a metric on P

$$f(a,b) + f(b,c) \geq f(a,c); \; f(a,b) = f(b,a); \; f(a,a) = 0$$

which preserves the preordering on $P \times P$.

2. Mapping from $P \times S$ into C

Let P be proper. We define a proper modification of P as an assignment to each

---

[1] Actually, this definition is weaker in most applications—see Sections 3 and 10.

[2] See Section 10.

$p_i \in P$ of a "neighborhood", i.e., set $R_i \subset S$ such that, if we define $g(p_i, x) = f\left(p_i, r_i(x)\right)$ where $(x \in R_i) \longleftrightarrow \left(r_i(x) = p_i\right)$, g is a proper mapping. (The definition of g makes sense only if the R_i are disjoint; this is so in all interesting cases (see Rothenberg, 1969).) A maximum proper modification is a proper modification, $R = \bigcup_i R_i$ which is properly contained in no other proper modification.

Clearly, we may begin with, say p_2, and choose R_2 such that it is maximal, then do the same for p_5, then p_1, etc. That is, the R_i may be successively maximized for elements of any permutation of the elements of P. Each such maximal R_i will constrain the remaining R_i, and we may obtain one or more different proper modifications, $R = \bigcup_i R_i$ for each permutation, $\alpha = i, j, k, \ell, \ldots$, of the elements of P.[1] For simplicity, consider the case where each proper modification, $R^{(\alpha)}$, is unique for that α.

We define the range, $\overline{R}_i$, of p_i as the maximal R_i when $\forall (j \neq i) R_j = \{p_j\}$. Let $\overline{R} = \bigcup_i \overline{R}_i$. It is easily shown that $\overline{R} = \bigcup_\alpha R^{(\alpha)}$. Similarly, we define $\underline{R} = \bigcap_\alpha R^{(\alpha)}$ such that $\underline{R}_i = \bigcap_\alpha R_i^{(\alpha)}$ (where i indexes the different R_i obtained in the order specified by α). $\underline{R}$ is called the blur of p_i and for all i, $\underline{R}_i \subseteq R_i \subseteq \overline{R}_i$. $\overline{R}$ and $\underline{R}$ are obviously more easily computed than all $R^{(\alpha)}$. Hence we observe that for any i, $\overline{R}_i \cup \left[\bigcup_{j \neq i} \underline{R}_j\right]$ is a proper modification (but not usually maximal).

Note that often, $\overline{R} \subsetneq S$. Actually, we adjust our methods so that $\overline{R} = S$ (see Sections 3 and 10). For techniques for mapping from $S \times S$ to C, see Rothenberg (1969).

Note that the space in which P is embedded need not be Euclidean and may have differing local "dimensionality" at each point (related to the number of chains passing through that point).

3. Tolerance

Notice that the cardinality of P is specified by a real pattern recognition task in one of two ways: either the size of the code (alphabet or number of classifications required) is specified (this corresponds to the length of the maximum chain), or a maximal tolerable confusion is specified together with a range over which such limitation applies. For example, let S be interpreted as cells in the retina of the eye. Suppose that in order to perform a particular task, discrimination of straight lines which differ in length by only one centimeter and which are observed at a fixed distance from the eye, is required. Then any pair of points on the retina which correspond to a projected distance of more than one centimeter cannot lie in the same R_i (or some necessary discriminations would fail). Such a tolerance, ϵ, may be incorporated into the system by weakening 2(a) and 2(b) (above) to accomodate reversals of ordering less than ϵ.[2] Then the R_i will usually overlap and may be considered "fuzzy sets".

---

[1] For methods of constructing proper modifications, see Rothenberg (1969).

[2] Relations between ϵ (which is an element of $S \times S$) and the link size exist so that S can be covered by $\overline{R}$. See Rothenberg (1969).

4. Sufficient Sets

Since each P may be a "Gestalt" or a "phonemic alphabet", problems arise when more than one "Gestalt" may be used (as in vision), or when a listener speaks more than one language. Minimal cues for the identification of the appropriate set, P (the "Gestalt" or "alphabet"), must be derived. These minimal cues are subsets of S which allow a unique identification of a particular $P \subset S$ from all possible P's available to the listener. From these a system of mapping P (as well as $P \times P$) into an "alphabet" is derived. (In vision this applies to the fixing of the position of an object in the visual field.) Consider a set $\{P_v\}$ (the set of learned "Gestalts"), where v indexes different $P_v \subset S$. (Here we assume that there exist no P_x and P_w in $\{P_v\}$ such that $P_x \subset P_w$.) We define a sufficient set for P_v, as a subset, Q, of P_v such that for all w, if $w \neq v$, Q is not a subset of P_w.

Consider a language whose phonemic alphabet (code) contains $\bar{n}$ distinct elements ("phonemes" or "letters"). How many distinct n-letter words can be formed using this alphabet? Of course, certain restrictions exist which limit the sequences of letters which can occur (e.g., no more than two consonants in a row or, as in Chinese, all words have only one syllable). The more distinct words that can be formed whose length is less than or equal to some maximal n, the more "efficient" the alphabet (code) may be said to be.

Consider all non-repeating sequences of all points (say $\bar{n}$) in P_v. There are $\bar{n}!$ such sequences. Let s_i be the number of elements in each sequence which must appear before a sufficient set is encountered. Then, $F(P_v)$ is defined as the average,

$$F(P_v) = \sum_{i=1}^{\bar{n}} s_i / \bar{n}! \quad .$$

$F(P_v)$ may be interpreted as the average number of elements in a non-repeating sequence of the $\bar{n}$ elements of P_v required to uniquely determine v. Efficiency, $E(P_v)$, is defined as $F(P_v)/\bar{n}$ and redundancy, $R(P_v)$, as $1 - E(P_v)$.

$E(P_v)$ may be interpreted as a measure of the assymmetry of P_v with respect to all rotations and translations of itself.

5. The Directed Graph

Suppose there exist P_x and P_w in $\{P_v\}$ such that $P_x \subset P_w$. All elements of $\{P_v\}$ may be arranged in a directed graph, G, in which a connection from P_x to P_y indicates that $P_x \subset P_y$ and where the P_w with the fewest elements are at the bottom of G. We now define a graph sufficient set for graph node, P_v, as a subset, H, of P_v such that for all w

$$H \subset P_w \longrightarrow P_v \subset P_w \quad .$$

If we utilize the hypothesis that classification procedures are as efficient as possible (i.e., whenever several possibilities exist, the lowest node on the graph will be used), we define a node sufficient set for P_v as a subset H, of P_v such that for all w

$$H \subset P_w \longrightarrow P_v \subset P_w \quad \underline{\text{or}} \quad P_w \subset P_v \quad .$$

We now define graph efficiency, E^G, and node efficiency, E^N, as before, using graph or node sufficient sets instead of sufficient sets.

To each set, H, which is a subset of an element of $\{P_v\}$ there corresponds a subset of $\{P_v\}$,

$$V(H) = \left\{ P_v \middle| H \subset P_v \right\} \quad .$$

Two sets, H_1 and H_2 are called graph equivalent iff $V(H_1) = V(H_2)$. To each H let there correspond a number

$$I(H) = \text{card}\{P_v\} - \text{card}\, V(H) \quad .$$

I(H) is the number of graph nodes of which H is not a subset, and is called the information value of H with respect to graph, G.

We now define the image distance, $\bar{I}(H_1, H_2)$, between two sets, H_1 and H_2 as

$$\bar{I}(H_1, H_2) = 1 - \frac{\text{card}\left(V(H_1) \cap V(H_2)\right)}{\text{card}\left(V(H_1) \cup V(H_2)\right)}$$

6. Application to Spoken Speech

Consider the recognition of random vowel sounds by a monolingual speaker of a natural language, say French. The set of random vowel sounds would correspond to S and the set of French vowel sounds to P (note: $P \subset S$). We now obtain our initial ordering on $S \times S$ by noise modulating the vowel sounds in S and noting the relative confusions of pairs of vowels. Our hypothesis states that there should exist a link size,[1] $\bar{x}\bar{y}$, such that P is proper (as defined) and such that the range, R_i, of each $p_i \in P$ corresponds to the set of random vowels confused with p_i. (P would correspond to points in the "vowel quadrilateral" used by linguists.)

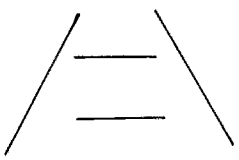
Suppose we have a multilingual speaker and we are presenting him with random syllables which include those in all of the languages he speaks, $\{P_v\}$. Now our graph sufficient sets should be the minimal cues necessary for him to choose a particular language, P_x, as a reference frame or "Gestalt" so that subsequent confusions are as predicted for P_x (as above). The amount of information supplied by a subset of vowels in a particular language is given by the information value of that subset and determines the probability of his choosing that

[1] Or an appropriate neighbornood system—see footnote, Sections 2 and 10.

language as a "Gestalt". Hence, we here have different predictions generated by the model for different sequences of stimuli (i.e., it is "context dependent").

7. Visual Illusions

Consider the familiar optical illusion whereby the moon appears larger on the horizon than when high in the sky. Note that buildings, trees and airplanes become smaller as they approach the horizon according to the laws of perspective. Because of its great distance from us, the moon does not. Hence our normal P derives from a metric involving contractions due to perspective. This P, however, is inappropriate for judging relative sizes of vastly distant objects. In similar fashion, familiar illusions due to perspective, such as the "railroad tie" illusion:



are easily explained.[1]

When applied to color vision, if the initial ordering is obtained by measuring confusions of colors on a particular photograph, if a separate P is computed for each of three primary colors and we then assume color mixing according to the values of C obtained by our mapping, we have a quantitative version of Edwin Land's "Theory of the Retinex".

Note also that if a chain (as here defined) is interpreted as a series of connected neurons, and if a set of neurons, all of which are connected to each other, is interpreted as a range, our function, g, which counts edges in a shortest chain, corresponds roughly to the length of time required for a stimulus to pass from one neuron clump (range) to a neuron end which is not in that clump. Analogies to "shadow enervation" on the retina also can be made.

8. Application to the Perception of Pitch

Here we assume a simple ordering on S (pitch or frequency) together with the preordering on $S \times S$ ("musical intervals"), and the following axiom limits dimensionality:

(a) $x < y < z \longrightarrow xy,\ yz < xz$

For Western music we may also assume that

(b) $xy < zw \longrightarrow \exists u\,(xy \sim zu)$ and $z < u < w$ or $w < u < z$

(c) $(x < y < z) \wedge (u < v < w) \wedge (xy \sim uv) \wedge (yz \sim vw) \longrightarrow xz \sim uw$

(c) specifies that equal musical intervals "added" to equal intervals are equal. (b) postulates the ability to interpolate a pitch between two other pitches, and permits us to obtain the preordering on $S \times S$ as follows: to compare zw with xy add u on the same side of z as w so

[1] See Stratton (1897), Thouless (1931) and von Senden (1932) pertaining to the absense of a stable metric in the visual cortex.

that $xy \sim zu$ and determine if u is internal to zw. If it is, $zw > xy$; if u is external, $zw < xy$; otherwise $xy \sim zw$. The u at which zu is perceived as equivalent to xy has been experimentally shown to depend upon the timbre of the tones, x, y, z and u, as does the entire preorder obtained in this manner.[1] Axiom (a) will accommodate all music (except "klangfarbenmusik", for which the more general model must be used.)

Now, let i in p_i range from $-\infty$ to $+\infty$ and index $p_i \in P$ such that the simple ordering on $P \subset S$ is obeyed. Then

$$f(p_i, p_j) = |i - j|$$

and if we define $\delta_{ij} = p_{i+j}p_i$, $\underline{\delta}_i = \min_j \delta_{ij}$ and $\bar{\delta}_i = \max_j \delta_{ij}$ it is easily shown that

$$\text{(d)} \quad \forall i(\bar{\delta}_i < \underline{\delta}_{i+1})$$

is a necessary and sufficient condition for a strictly proper mapping, and if "$\leq$" replaces "$<$" in the formula, for a proper mapping. Otherwise P is improper.

P has <u>period</u> n if for all i, j, $p_i p_k \sim p_{i+n} p_{k+n}$ and if n is the least positive integer satisfying the condition. In this case (e.g., octave equivalence), the positive integers suffice to rank the order of all pairs $\delta_{ij} \in P \times P$ according to the initial ordering. This rank order for $\delta_{ij} = p_{i+j} p_j$ is called α_{ij}, and it is now possible to define a matrix, $[\alpha_{ij}]$, called the <u>reduced matrix</u> of P:

$$\begin{matrix} \alpha_{1,1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{(n-1),1} & \cdots & \alpha_{(n-1),n} \end{matrix}$$

Now, if we define $\underline{\alpha}_i = \min_j \alpha_{i,j}$ and $\bar{\alpha}_i = \max_j \alpha_{i,j}$, we may replace (d) (above) by "$\forall_i(\bar{\alpha}_i < \underline{\alpha}_{i+1})$" which is applied only to elements of the reduced matrix.

It is now easily proved that, if P is proper and periodic, R_i (range) is an interval about p_i (i.e., $r_i(x)$ is proper), all R_i and R_{i+1} intersect at one point at most (all others are disjoint), if <u>every</u> two consecutive ranges intersect, then $\bigcup_i \bar{R}_i = S$. Simple methods for computing proper modifications, ranges, blurs, and a method for generating all proper P such that $P \subset S$ (S is finite) have been developed. Efficiency, E, as well as all other quantities and sets defined are easily computed or generated by operations on the reduced matrix.

Whenever there exists an $\alpha_{i,j} = \alpha_{i+1,k}$ we have an ambiguous pair (musical interval)

---

[1] See Shouten (1962), Evetts (1958), Licklider (1959), Pratt (1928) and Munsterberg (1892) for interactions between pitch and timbre.

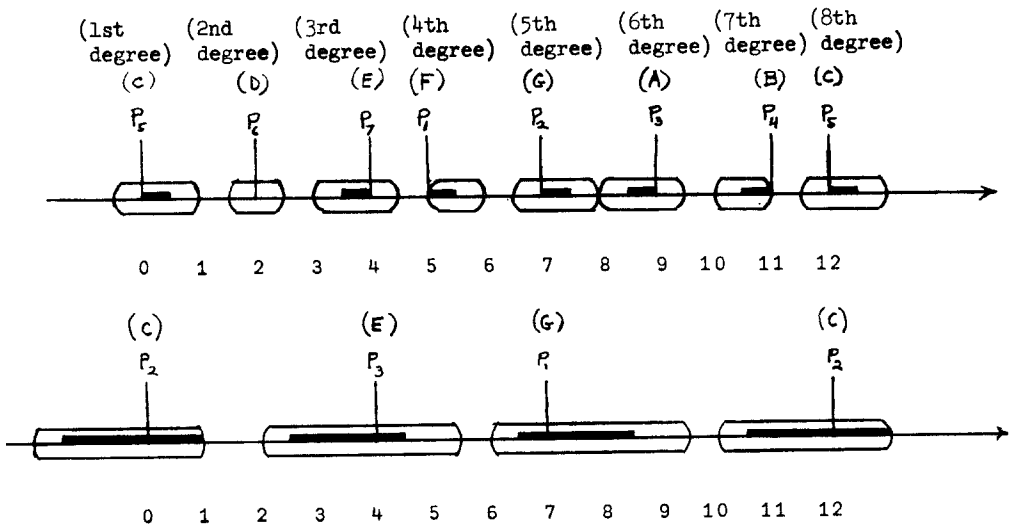
and if $\alpha_{ij} > \alpha_{i+1,k}$ we have contradictory pairs. In our interpretation P is a "musical scale" (and all its "modes") or a "chord" (and all its "inversions"). The reduced matrix where P is the "major scale" (as tuned in 12-tone equal temperament) is shown below and ambiguous α_{ij} are encircled:

$$[\alpha_{ij}] = \begin{bmatrix} 2 & 2 & 2 & 1 & 2 & 2 & 1 \\ 4 & 4 & 3 & 3 & 4 & 3 & 3 \\ \textcircled{6} & 5 & 5 & 5 & 5 & 5 & 5 \\ 7 & 7 & 7 & \textcircled{6} & 7 & 7 & 7 \\ 9 & 9 & 8 & 8 & 9 & 9 & 8 \\ 11 & 10 & 10 & 10 & 11 & 10 & 10 \end{bmatrix}$$

p_1p_4 and p_4p_1 are ambiguous

Note that the first column contains all musical intervals less than an octave which have the "fourth degree" of the major scale as an end point. Hence we note that the two tritones between the fourth and seventh degrees and seventh and fourth degrees are ambiguous pairs. We also define <u>stability</u>, $\bar{S}$, as the proportion of unambiguous pairs in P × P For the major scale $\bar{S}$ = 9524.

Below are two diagrams which show the ranges, $\bar{R}_i$, and blurs, $\underline{R}_i$, of each p_i in one octave of the "C major scale" (the first diagram) and in a "C major triad" (the second diagram). Each $\bar{R}_i$ is shown by the brackets enclosed on top and bottom, and each $\underline{R}_i$ by a darkened rod. The numbers under the diagram index the "semitones" in the octave starting with C = 0:



Since ambiguous pairs cannot be unambiguously classified except in reference to adjacent elements, it is an obvious rule of musical usage that an ambiguous musical interval (except when it occurs in a "chord" in which it is not ambiguous) must be approached or left by a "step" (i.e., it must "resolve" (move) to a tone adjacent to one of its component tones which, together with its other component tone, no longer forms an ambiguous pair (musical interval)). It is also predicted that a tone, $x \in S-P$ which is in the range R_i of some $p_i \in P$ must either replace or "resolve" (move) to that p_i. Hence the explanations of "auxilliary" and "altered" tones. "Root" or "tonality" corresponds to an element of P which temporarily is an endpoint of all pairs measured by f (or g). In cases where a drone or "ostinato" is used, it is not significant that P is proper, and improper P ("scales") are often used (to be discussed).

Note that a given P may have many $P_v \subset P$ which are proper. When P is the major scale, those P_v are "chords". Those proper P_v which are subsets of both the "major" and "minor" scales and their "alterations" comprise the "figured bass" system of Western classical music. Consider a string of proper sets, $P_1, P_2, P_3, \ldots, P_m$ such that for all $k < m$, there exists a proper modification of P_k, $R(P_k)$ such that $P_k \subset R(P_{k-1})$. (This is analogous to an hierarchical clustering.) If classification is always performed by the smallest proper set, the ranges of the points in that set will include the elements of the larger set next in the above string. Hence, if we temporarily use the "C major triad" as our "Gestalt" (i.e., P) when using other tones of the "C major scale", its ranges determine the traditional "rules of voice leading"; i.e., "B leads to C, D and F lead to E and A leads to G" (see diagram above). Similar traditional results obtain from application to the minor scales and to proper subsets of improper scales whose ranges include all tones in that improper scale. In the latter case a partition of scale tones into "principal" and "auxiliary" tones results. Similar analysis elucidates chromatic usage within the major-minor system.[1]

It is also predicted that those proper scales will occur in the music of different cultures which are maximum both in stability, $\bar{S}$, and efficiency, E, as computed when all "keys" (i.e., "transpositions") of the scale are in $\{P_v\}$. Appropriate computations have been made, and this has been shown to be the case[2] (note: the initial ordering is determined by the timbre of the instruments used in each culture). Restrictions on the tuning of scales such that the initial ordering is retained have been computed. These, in general, are more stringent in improper scales than proper scales, and are consistent with cross-cultural observations. In Java there exist two scale systems, "Slendro" and "Pelog", each containing a variety of "scales". It has been observed that all scales in the "Slendro" class are strictly proper and that all in the "Pelog" class are improper. In a study conducted with the

---

[1, 2] See Rothenberg (1969).

assistance of Mr. Surya Brata of the Ministry of Education and Culture, Jakarta, the uses of these scale systems were observed to be in accord with the predictions of this model. [1]

Notice that in a proper scale with high stability, the musical intervals (i.e., pairs in P × P) are, for the most part, unambiguously classified (measured). Hence sequences of similar interval patterns (i.e., "modal motivic sequences") are easily apprehended as similar, and their use would be anticipated. When efficiency is also high, the elements of P have their positions (absolute pitches) quickly fixed (relative to tones previously heard) and we would also expect the use of "tonics", i.e., "tonal music". When efficiency is low, it would not be expected that tonality would be used (as musical material). This indeed appears to be the case. Motivic sequences are used in proper scales, but tonality is avoided (or irrelevant) in those with low efficiency (high redundancy) such as the "twelve tone scale" (note the motivic properties of "tone row" use), the "whole tone scale", etc. When the major and minor scales as well as the whole tone scales are included in the directed graph, G, and the graph efficiency, $E^{G}(P_{v})$, is computed where P_{v} is the whole tone scale, this efficiency becomes high. This is consistent with the successful extensive tonal use of the whole tone scale only in conjunction with the major and minor scales by Debussy and Ravel.

When a scale is improper (or of low stability) many motivic (intervallic) similarities are not apparent (easily perceived). Therefore, it is important that the tones in the scale be quickly fixed, i.e., the scale "degrees" be identified) so that the partition between "principal" and "auxilliary" tones be clear (witness the customary use of a drone in Indian music). Hence high redundancy (low efficiency) is required. This appears to characterize all improper scales observed.

Note that in tonal music a "cadence" must fix both the scale, P, and its "tonic" (an element of P.) Hence it would be expected that it would contain a sufficient set, the tonic (and probably a tone a "fifth" above it so that it is reinforced by the resulting difference tone).[2] The cadence would contain as many proper subsets of P as possible, so as to reveal its harmonic substructure. All traditional cadences are accounted for in this manner, as well as cadences in "free twelve tone music" (when all proper subsets of the "chromatic scale" are included on the directed graph, G).

When all such proper subsets of the chromatic scale are included on the graph, information values and image distances assume significance. The proper subsets with high stability and with six or more elements (those with fewer usually function as "chords") which appear on the graph are all scales which can be formed by choosing an element of P from S and selecting the remaining tones of P from S in clockwise fashion in accordance with the

---

[1] See Rothenberg (1969), revised version, and also Kunst (1949) and Hood (1954, 1966).

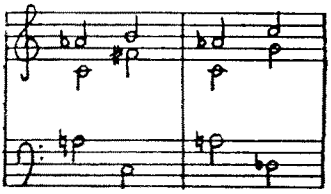
[2] See Helmholtz (1948).

following vectors (where "1" represents a distance of one "semitone" and "2" of a "whole tone", etc.). Each vector denotes a set of "keys" of a scale"

(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)	("twelve tone" scale)
(2, 1, 1, 2, 1, 1, 2, 1, 1)	(of borderline stability - used by Olivier Messiaen—see Messiaen (1944))
(2, 1, 2, 1, 2, 1, 2, 1)	("string of pearls")
(2, 2, 2, 2, 1, 2, 1)	("melodic minor")
(3, 1, 3, 1, 3, 1)	
(2, 2, 2, 2, 2)	("whole tone" scale)

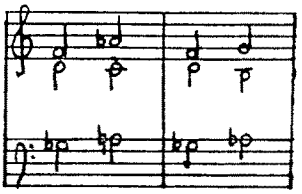
The information value of a given group of tones indicates the ease with which a P is chosen from the graph for its classification (by a listener). The three tone sets ("triads") of maximal information value are given by the vectors, (10, 1, 1) (e.g., C, $A^{\#}$, B), (8, 3, 1) (e.g., C, $A^{\flat}$, B), (8, 1, 3) (e.g., C, $G^{\#}$, A), (6, 5, 1) (e.g., C, $F^{\#}$, B) and (6, 1, 5) (e.g., C, $F^{\#}$, G). The frequent use of these "chords" in twentieth century music (and especially in "cadences") is well known; hence "minjor"[1] chords and chords in fourths. A startling example is Anton Webern's "Piano Variations" (opus 27). The entire composition consists of a succession of graph sufficient sets for the "twelve tone scale", although the "serial" technique of composition does not guarantee this.

Image distance is an extremely sensitive indicator of apparent differences between "chords" in twentieth century western music. In addition to obvious similarities due to the presence of common tones, the following subtle differences are revealed (the image distances are below the progressions):



A B A C
I = 2/5 I = 6/7

(A) = (8, 1, 3); on C
(B) = (6, 5, 1); on C
(C) = (6, 5, 1); on $D^{\flat}$



A B A C
I = 2/3 I = 6/7

(A) = (8, 3, 1); on $G^{\flat}$
(B) = (8, 1, 3); on C
(C) = (8, 1, 3); on B

The model is far more detailed than here presented and more complete in its musical application.

[1] That is, a combination of a minor and major triad (e.g., C, $E^{\flat}$, $E^{\natural}$, G).

10. The Development of New Materials

In addition to ethno-musicological testing, which has thus far strongly supported the theory, experiments and equipment have been designed for the testing of the application to the perception of pitch. These experiments in part resemble those designed for testing the perception of vowels in a natural language. Related to these is the use of the model for the generation of new musical materials.[1] The experimental equipment is appropriate for their musical exploitation. Many of the new musical materials generated are consistent with Western musical tradition.

Of particular interest is the generation of materials for "tone color" music (Klangfarbenmusik"), which resembles the application to spoken vowels: The composer is asked to supply a set of sounds he likes, S. He is then asked how large he wants his alphabet, P, to be. The initial ordering on $S \times S$ is obtained by direct inquiry as to perceived similarity or by means of confusion in the presence of noise (as in the case of vowels—no ordering on S exists here). All proper $P \subset S$ and the ranges of each of their elements are generated. These P's are chosen such that they have a proper modification, $\bar{R} = S$. The composer then chooses one of these P's. The properties of the perception of the elements of S are determined as in the case of pitch and their musical usage is similarly circumscribed.

Note that, while the technique known as "cluster analysis" has been successfully applied to the study of the perception of phonemes of natural languages, this tool is not adequate for our purpose here (i.e., to generate a musical "phonemic alphabet"—a synthetic alphabet of musical materials). The principal reason is that, in experiments involving the phonemes of a natural language, deviant phonemes tend to "cluster" about normative phonemes in the language as a consequence of the context of that language in the mind of the subject providing experimental data. However, in a language not known to the subject (e.g., a "musical alphabet" of novel sounds), no such effect can be expected. Experimental data produced by a subject before he is familiar with a language will almost certainly differ from data produced after he is fluent in the language. A cluster analysis of each such set will therefore produce different results. By contrast, the question we here consider is, "how can we select our "alphabet" of materials (P) in such fashion that the experimentally produced ordering (i.e., the "initial ordering") before the subject has learned the "alphabet" will be minimally altered in experiments performed after he has learned that "alphabet"?

---

[1] In particular, many novel "musical scales" employing "microtones" have been generated. See Rothenberg (1969).

It can be shown that this question is logically equivalent (if we assume that the hypotheses of this paper are valid) to the problem of selecting a "proper" set (P) from the domain (of stimuli), S, utilized in the experiments.

A similar application to the perception of patterns (in particular, textures) used in abstract animated films is planned.

11. Theoretical Basis:[1]

Let L be a (possibly infinite) set whose elements correspond to given stimuli or (alternatively) sensory receptors. Assume a set of atomic predicates which specify an ordering of pairs of elements of S (e.g., the ordering specifying the relative similarities of pairs of stimuli; i.e., "xy < zw" would indicate that x is "more similar" (or "closer") to y than z is to w). Suppose we weaken the ordering "<" by the introduction of some ϵ which is an element of S × S; i.e., we say that "$xy \underset{\epsilon}{\leq} zw$" if and only if xy exceeds zw by at least ϵ (precisely, $xy \underset{\epsilon}{\geq} zw \equiv xy > zw \wedge \forall v \left((xv \leq zw \rightarrow vy > \epsilon) \wedge (xy \leq zv \rightarrow wv > \epsilon)\right)$). We consider mapping the structure, $\langle S, \underset{\epsilon}{\geq} \rangle$ into a substructure, $\langle P, \underset{G}{\geq} \rangle$ where P is a finite discrete space (intended to correspond to a "Gestalt", "reference frame" or classification of stimuli) such that the following conditions hold (and define "$\underset{G}{\geq}$"):

P is a Freschet space (see Sierpinski, 1952, Chapter 1) wherein points are assigned to neighborhoods such that each point has a minimum neighborhood (i.e., no other neighborhood is properly contai ned therein) and such that P and the null set are the only closed sets.[2] We define points in the same minimal neighborhood as "adjacent" and define a metric, f(x,y), on P as the number of edges in the shortest path along adjacent points from x to y. If f(x,y) is greater than f(z,w) we denote this by "$xy \underset{G}{\geq} zw$". When ψ is a continuous mapping from $\langle S, \underset{\epsilon}{\geq} \rangle$ onto $\langle P, \underset{G}{\geq} \rangle$ such that if $x,y,z,w \in S$, $xy \underset{\epsilon}{\geq} zw \rightarrow \psi(x)\psi(y) \underset{G}{\geq} \psi(z)\psi(w)$, P is called a "basis" for S and ψ is called a "reduction mapping". Such a mapping from $\langle S, \underset{G}{\geq} \rangle$ onto a substructure $\langle P, \underset{G}{\geq} \rangle$ is of interest because it can be shown that any subset A of S which satisfies a formula of the second order predicate calculus (i.e., a "property" of the set, A, of stimuli) has an image $\psi(A)$ in P which satisfies a modified version of that same formula. This modification is accomplished by replacing universal quantifiers of the formula by "numerical quantifiers", $Q^{(k)}x$, which may be read "for k 100 percent of x" (similarly to the reading of "$\forall x$" as "for all x"), in the following manner:

---

[1] A brief mathematical treatment of this section can be found in another paper in this volume entitled "Predicate Calculus Feature Generation", Sections 14 and 17.

[2] This guarantees the connectedness of the space (Rothenberg, 1974).

Suppose we are given a formula, F, with one free set variable, and a set $B \subset S$ which satisfies F.[1] When $\psi(B)$ does not satisfy F, we consider the replacement of a single universal quantifier by a numerical quantifier, $Q^{(k)}$, where k is maximal such that $\psi(B)$ satisfies the modified formula, F'. This is done for all universal quantifiers, and unity subtracted from the largest value of k thus obtained is defined as the "degree of satisfaction of $\psi(B)$ with respect to F", $D(\psi(B), F)$. When $\psi(B)$ does satisfy F, we define $D(\psi(B), F)$ as the largest value of k over all replacements of a single existential quantifier by $Q^{(k)}$ such that $\psi(B)$ still satisfies the modified formula, F'. Note that when B is finite, the latter definition, if applied to B instead of $\psi(B)$, defines the degree of satisfaction of B with respect to F, D(B, F).

Intuitively (dealing with geometrical figures), if F is satisfied only by starlike[2] sets and both B and C are starlike, "$D(B, F) > D(C, F)$" says that "B is more starlike (i.e., more nearly convex) than C". If neither B nor C is starlike, the inequality is intuitively interpreted as "C is less starlike (i.e., more "hollow" at its "center") than B. Hence we define $D(B, F) - D(\psi(B), F)$ as the "degradation of B with respect to F by ψ", and $\psi(B)$ is called the "degraded image" of B.[3] A bound on such degradation may be computed by examination of the syntax of F and the smallest ϵ (in "$<_\epsilon$") such that ψ is a reduction mapping. Similarly, a bound, $b(F, \psi)$, on the maximum degradation over all B such that $B \subset S$ and B satisfies F may be computed (i.e., $b(F, \psi) = \max_B \left(D(B, F) - D(\psi(B), F)\right) \mid B \subset S \wedge B$ satisfies F)).

This latter quantity exposes the effect that ψ has on the satisfaction of F; i.e, it is a measure of the degree to which properties of subsets of our set of stimuli, S, are preserved by their images in our "reference frame", P.[4] If P is a musical scale" we hypothesize that those properties which enjoy minimal degradations (as defined above) are those which will define the relations used in the construction of musical forms. We also propose

---

[1] Note that the discussion applies if B is not a set, but a sequence of substitutions of elements (and/or subsets) of S for the free variables in a formula with many free variables.

[2] That is, there exists a point in B such that for every other point of B all elements between the two points are contained in B.

[3] Metaphorically, performing a "reduction mapping" is analogous to tearing a hologram into several parts. One of these parts corresponds to a "basis" of the mapping. The image produced from that part would correspond to the "degraded image" of the image produced by the entire hologram. A formula with one free set variable (as above) would be a property of an image which makes it recognizable (e.g., "it is spherical" or "it consists of a cube adjacent to a convex ball"). Such a property might be less (or possibly more) pronounced (e.g., "less spherical") in the image produced by a portion of the hologram than in the image produced by the entire hologram. The degree to which such a property (of a particular image) is pronounced is analogous to the "degree of satisfaction" of that property by the image. Intuitively a degraded image may be thought of as a "fuzzy" image of the original, much as a picture of a object on a television screen is a "fuzzy" image of that object.

[4] That is, it is a measure of specific kinds of information loss resulting from the utilization of ψ.

that the "rules of voice-leading" and "rules of harmony" of various cultures are chosen so as to restrict the degradation of the properties on which these relations are based. Although psychological experiments for direct verification of the predictions of the model have been designed, results are not yet available. However, the musical predictions of the model are in accord with Western musical practices and those of various Asian musical cultures which have been examined (Rothenberg, 1969, revised version).

For brevity, we here[1] eliminate ϵ from consideration, call the weakened reduction mappings which result "proper mappings" and confine our discussion to Western music. "Proper modifications" are preimages of points in the "basis" of a "proper mapping" and "contractory (and ambiguous) pairs" are those pairs of musical intervals (i.e., musical intervals are pairs of elements) which account for degradations by virtue of the inversion (or collapsing) of their order as a consequence of the mapping. "Stability" is a rough measure of the degradations of that class of formulae whose "models" (i.e., sets of substitutions for the free variables such that the formula is satisfied) are invariant with respect to translation[2] (and various other transformations as well) under the particular proper mapping (or corresponding basis, P) chosen. "Efficiency" similarly ranks the bases of various mappings according to the degradation of formulae whose models are not invariant with respect to translation. These various bases are considered as "musical alphabets" and models of negligibly degraded formulae are hypothesized to be the units of "musical form" (e.g., "motifs", etc.) when such alphabets are used in a musical composition.

References

Evetts, J.E. (1958), "The Subjective Pitch of a Complex Inharmonic Residue", unpublished report, Pembroke College, England.

Helmholtz, H.L.F. (1948), On the Sensations of Tone as a Physiological Basis for the Theory of Music (Translated by A.J. Ellis, 1885), New York; Peter Smith.

Hood, M. (1954), The Nuclear Theme as a Determinant of Patet in Javanese Music, Groningen, Djakarta: J.B. Wolters.

Hood, M. (1966), "Slendro and Pelog Redefined", Selected Reports, Institute of Ethnomusicology, University of California at Los Angeles.

Kunst, J. (1949), Music in Java, The Hague; Martinus Nijhoff.

Licklider, J.C.R. (1959), "Three Auditory Theories", in S. Koch (ed.), Psychology: A Study of a Science, New York: McGraw-Hill, pp. 41-144.

Messiaen, O. (1944), Technique de mon langage musical, Paris: Alphonse Leduc.

---

[1] For a treatment of ϵ, see Rothenberg (1969, 1974).

[2] That is, still satisfiy the same formula after translation. See Rothenberg (1974).

Munsterburg, H. (1892), "Vergleichen der Tondistanzen", Beitrage zur experimentelle Psychology, 4, 147-177.

Pratt, C. C. (1928), "Comparison of Tonal Distance", and "Bisection of Tonal Intervals Larger than an Octave", J. Experimental Psychology, 11, 77-87 and 17-36.

Rothenberg, D. (1969), "A Pattern Recognition Model Applied to the Perception of Pitch", Air Force Office of Scientific Research Technical Report, Dept. of Information Sciences. Revised version to appear as series of articles in Mathematical Systems Theory

Rothenberg, D. (1974), "Predicate Calculus Feature Generation", this volume.

Schouten, J. F., R. J. Ritsma and B. L. Cardozo (1962), "Pitch of the Residue", J. Acoustical Society of America, 34, 1418-1424.

Sierpinski, (1952), General Topology, Toronto: University of Toronto Press.

Stratton, G. M. (1897), "Vision Without Inversion of the Retinal Image", Psych. Rev., 4, 341-360; 463-481.

Thouless, R. H. (1931), "Phenomenal Regression to the Real Object", British J. Psychology, 21, 339-359.

von Senden, M. (1932), Raum- und Gestaltauffassung beioperierten Blindgeborenen vor und nach der Operation, Leipzig: Barth.