

DOCUMENT RESUME

ED 065 442

24

SO 004 683

AUTHOR Heller, Jack J.; Campbell, Warren C.
TITLE Computer Analysis of the Auditory Characteristics of Musical Performance. Final Report.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-9-0546
PUB DATE May 72
GRANT OEG-0-9-160546-4439-010
NOTE 131p.
EDRS PRICE MF-\$0.65 HC-\$6.58
DESCRIPTORS Analog Computers; *Auditory Perception; Aural Learning; *Computer Assisted Instruction; Digital Computers; *Listening Comprehension; *Music; Musical Instruments; *Music Education; Research; Research Methodology

ABSTRACT

The purpose of this research was to perform computer analysis and modification of complex musical tones and to develop models of perceptual and learning processes in music. Analysis of the physical attributes of sound (frequency, intensity, and harmonic content, versus time) provided necessary information about the musical parameters of intonation, vibrato, dynamics, and rhythm. A general purpose digital computer and appropriate analog devices were utilized to analyze and synthesize complex musical tones. The procedures included the transformation of audio tapes of music to digital tapes via a high speed analog-to-digital converter system. The significance of the research is based on the belief that: 1) objectifying certain parameters of musical performance will have a direct bearing on behavioral goals and methods of music education; and, 2) an understanding of the total problem of human information processing requires a detailed investigation of structured non-verbal stimuli in the auditory mode. Two supplementary investigations are included: Computer Analysis of Musical Performance, by Warren C. Campbell (Appendix I) and The Effects of the Attack Transient on Aural Recognition of Instrumental Timbres, by Ralph C. Thayer, Jr. (Appendix II). Appendix III, Computer Analysis System Software, by Jack Owens, describes a library of computer programs used to perform certain basic mathematical analyses of the digitized musical performances. A related document is ED 058 745. (Author/JMB)

ED 065442

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to:

SO

EM

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

ED 065442

U. S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIG-
INATING IT. POINTS OF VIEW OR OPIN-
IONS STATED DO NOT NECESSARILY
REPRESENT OFFICIAL OFFICE OF EDU-
CATION POSITION OR POLICY.

FINAL REPORT
Project No. 9-0546A
Grant No. OEG-0-9-160546-4439 (010)

COMPUTER ANALYSIS OF THE AUDITORY
CHARACTERISTICS OF MUSICAL PERFORMANCE

Jack J. Heller
Warren C. Campbell
The University of Connecticut
Storrs, Connecticut

MAY, 1972

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION AND WELFARE

Office of Education
Bureau of Research

CONTENTS

	Page
LIST OF TABLES	iv
LIST OF FIGURES	v
SUMMARY	vi
I. INTRODUCTION	1
Music Analysis	1
Development of Models	4
II. RESEARCH DIRECTIONS: EXPERIMENTS AND MODELS	7
A. Performance Adjudication and Analysis	7
Statement of Purpose	9
Statement of Problem	9
Method of Solving the Problem	9
Significance	10
Evaluation of Student Achievement	11
Application to Computer Assisted Instruction	12
Methodology	13
Psychometrics and Statistics	13
Pattern Recognition	14
Musical Acoustics and Special Equipment	14
Summary	16
B. Performance Modification and Listener Response	17
Introduction	17
Related Literature	18
Procedures	20
Results	20
C. Information Processing and Musical Perception	22
D. Systems for Performance Analysis and Modification	30
The Computer Analysis System	30
Tone Line Writer	32
Tone Line Reader	32
Harmonic Synthesizer	34
REFERENCES	36
APPENDIX I: COMPUTER ANALYSIS OF MUSICAL PERFORMANCE	41
Purpose and Problem	41
Procedures	41
Performance Selection and Preparation	42
Criterion Scores	44
Predictor Variables	45
Analysis	50
Summary	51

CONTENTS (continued)

APPENDIX I (continued)	Page
Results	51
Analysis of the Criterion Variable	51
Analysis of the Predictor Variables	58
Multiple Regression Analysis	58
Summary	66
Discussion and Conclusions	68
Criterion Scores	68
Standards for Evaluating Computer Performance	68
Predictor Simulation of Criterion Scores	70
Generalizability	71
Future Research	71
Summary	72
References	74
Appendix A: Information Given to Judges	75
Appendix B: Data Normalization and Reduction	78
APPENDIX II: THE EFFECT OF THE ATTACK TRANSIENT ON AURAL RECOGNITION OF INSTRUMENTAL TIMBRES	80
Introduction	80
Problem	81
Limitations	81
Hypotheses	82
Procedures	82
Test	83
Scoring	85
Results	85
Summary of Results	93
Conclusions	94
References	96
APPENDIX III: COMPUTER ANALYSIS SYSTEM SOFTWARE	103
Programming Considerations	103
Implementation	104
APPENDIX IV: SCANNER/SYNTHESIZER INTERFACE	115
APPENDIX V: HARMONIC SYNTHESIZER	118

LIST OF TABLES

Page

APPENDIX I: COMPUTER ANALYSIS OF MUSICAL PERFORMANCE

Table 1.	Characteristic Features, Calculated for Each Tone	46
Table 2.	Predictors Used for Subset Calculations	50
Table 3.	Interjudge Correlations	52
Table 4.	Correlation of Each Judge With Average of Other Judges' Scores	55
Table 5.	Summary of Inter-Judge Correlations, Inter-Group Correlations	55
Table 6.	Regrade Correlations for Three Judges	57
Table 7.	Intercorrelations of Performance Judging Categories	57
Table 8.	Correlations of Predictors with Criterion Scores, Set #1	59
Table 9.	Correlations of Predictors with Criterion Scores, Set #2	60
Table 10.	Computer Simulation of Human Judgements	61
Table 11.	Criterion of Simulated Scores with Criterion for Performance Subsets	63
Table 12.	Cross-validation of Random Subsets	65
Table 13.	Cross-validation of Natural Subsets	67
Table 14.	Criterion Correlation Ranges	69
Table 15.	Homogeneity and Reliability for Sixteen Judges	69

APPENDIX II: THE EFFECT OF THE ATTACK TRANSIENT ON AURAL
RECOGNITION OF INSTRUMENTAL TIMBRES

Table 1.	Means and Standard Deviations	86
Table 2.	Means Converted to Percentages	86
Table 3.	Results of Statistical Tests of Significance	87
Table 4.	Statistical Tests Between Groups	89
Table 5.	Correct Responses (In Percents) For Group C	89
Table 6.	Response Errors for Group C	90
Table 7.	Tables of Errors	97
Table 8.	Tables of Percentages	100

LIST OF FIGURES

Page

II. RESEARCH DIRECTIONS: EXPERIMENTS AND MODELS

Figure 1. Performance Analysis 8
Figure 2. Learning Windows for Speech and Music 29
Figure 3. Brain Processing Models 29
Figure 4. Computer Analysis System 31
Figure 5. Tone Line Writer 33
Figure 6. Tone Line Reader 35

APPENDIX I: COMPUTER ANALYSIS OF MUSICAL PERFORMANCE

Figure 1. Overview of Procedures 43
Figure 2. Sample Chart Recorder Output, Soprano 49
Figure 3. Sample Chart Recorder Output, Alto 49

APPENDIX III: COMPUTER ANALYSIS SYSTEM SOFTWARE

Figure 1. Program Configuration and Logic for Use
with the 2250 Display Console102
Figure 2. Obtaining the Wavelength from a Resonance106
Figure 3. Updating Disk Data Records108

APPENDIX IV: SCANNER/SYNTHESIZER INTERFACE

Figure 1. Scanner/Synthesizer Interface (Schematic)114
Figure 2. Scanner/Synthesizer Interface (Operational)117

APPENDIX V: HARMONIC SYNTHESIZER

Figure 1. Schematic Diagram #1119
Figure 2. Schematic Diagram #2119
Figure 3. Schematic Diagram #3119
Figure 4. Schematic Diagram #4119
Figure 5. Data Samples for f_1 through f_{11} 123
Figure 6. Data Samples Before and After Filtering124
Figure 7. Distortion at Breaks124

SUMMARY

Investigations in speech have shown the complex nature of the processing which enables us to decode the acoustic signal of spoken language. Constraints on natural language are imposed by both the production and receiving mechanisms. The perception of musical sounds is obviously subject to some of the same constraints, since the same receptors and, possibly, some of the same processing mechanisms are used for music as for natural language.

The ability to identify and classify significant perceptual parameters of musical performance, and to state the effects of these parameters in behavioral terms is a prerequisite to the establishment of effective training procedures and meaningful behavioral goals for music education programs. Knowledge of these parameters will enable objective models of musical performance to be developed and verified.

The purpose of this research was to perform computer analysis and modification of complex musical tones and to develop models of perceptual and learning processes in music. Useful models of aural perception in music have been verified by comparing the responses of a computer implementation with the responses of appropriate human listeners. Analysis of the physical attributes of sound (frequency, intensity, and harmonic content, versus time) provided necessary information about the musical parameters of intonation, vibrato, dynamics, and rhythm. Information has been provided regarding the relative importance of attack and steady-state portions of tones for specific instruments. Further data provided information regarding brain hemisphere dominance for tasks of identifying instrument attacks when two different instruments are presented dichotically.

In order to analyze and synthesize complex musical tones a general purpose digital computer and appropriate analog devices were utilized. The procedures included the transformation of audio tapes of music to digital tapes (numerical data) via a high speed analog-to-digital converter system. Multivariate statistical techniques were used to provide improved psychometric capability in the perceptual domain.

The significance of this research is based on the belief that 1) objectifying certain parameters of musical performance will have a direct bearing on behavioral goals and methods in music education, and 2) an understanding of the total problem of human information processing requires a detailed investigation of structured non-verbal stimuli in the auditory mode.

I. Introduction

The purpose of this research was to perform computer oriented analysis and synthesis of complex musical tones, and to develop models of perceptual and learning processes in music. This effort is motivated by the belief that the ability to identify and classify significant perceptual parameters of musical performance, and to state the effects of these parameters in behavioral terms is a prerequisite to the establishment of effective training procedures and pertinent behavioral goals for public school and college music programs. Knowledge of these parameters should enable objective models of musical performance to be developed and verified. This is, of course, not a new idea. One of the earliest proponents of this point of view is Carl E. Seashore (Psychology of Music, McGraw Hill, 1938). Additional impetus is given to this work by related experimentation in linguistics and psychology which point out that an understanding of the total problem of human information processing requires a detailed investigation of structured non-verbal stimuli in the auditory mode.

Music Analysis

The auditory characteristics of musical performance can be objectively examined in terms of the physical characteristics of the complex acoustic wave generated by the performer. For this research, the fact that there are visual aspects and other contingencies that may influence listener response to a performance is specifically not considered. This restricted situation is comparable to an individual listening to a tape or disc recording: the sound pattern can be completely represented at any of several points in the auditory channel as the variation of a single parameter plotted against time. Traditionally, if the vibration is reasonably repetitive, this information is reduced by introducing the concepts of frequency, intensity, harmonic content, and spectral distribution of noise, as functions of time. Their use allows for significant data reductions without serious loss of information of importance in the analysis of music.

Every perceptually significant nuance of the auditory component of musical performance, no matter how complex, can be described as a function of these attributes of sound waves. This statement should not be read as minimizing the extreme complexity of the subtleties and contextual dependencies involved in artistic performance.

Many laymen and musicians believe that attempts to analyze artistic performance are futile. They contend that individual difference is an overriding characteristic of artistic performance. The importance of individual difference is not denied, particularly when considered as the hallmark that permits performer identification. However, there is a need to understand the constituent elements and limits of variability common to all artistic performance, as a prerequisite to any discussion of the effect on the listener of individual difference.

Musicians probably will agree, that before one can produce an artistic performance, there are certain technical demands of performance which must be satisfied. Control of these technical aspects, such as intonation, dynamics, tone quality, rhythm, and attack, is a prerequisite to performance at a professional level.

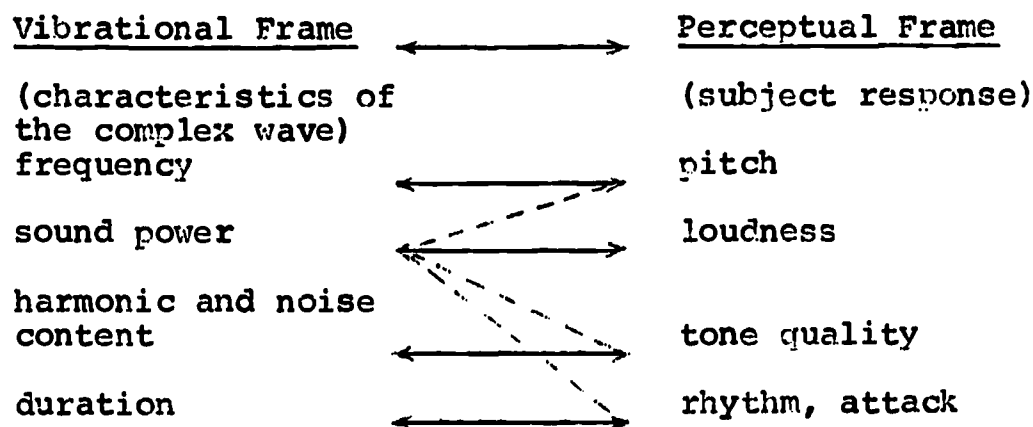
An analogy can be drawn between the musician and the literary writer. The latter cannot create a literary masterpiece until he has learned to control language. The writer who achieves optimal creative production through written communication must first master the language and make it subject to his control and manipulation. It is the same in the creative production of music. The artists who achieve a high level of performance in music will emerge from that group of musicians who have refined the technical aspects of performance.

This project was in no way an attempt to mechanize the creative process or to place restraints upon those persons who are capable of making unique contributions to music. Rather, the attempt was to identify the perceptually significant aspects of music by analyzing model performances that exemplify a high level of musicianship (as judged by professionals), and by comparing these to performances that fall short of this level.

The basic analysis may be thought of as a

mapping of auditory characteristics from the perceptual to the vibrational frames of reference, and the reverse. For each perceptual event, a vibrational counterpart or correlate is sought which can be described in terms of the parameters indicated below:

AUDITORY MAPPING



The relationships indicated by the solid lines between the vibrational and perceptual parameters are first-order effects. Functional relationships and difference limens were established for some of these main effects by the early experimenters in audiology. For example, the relationship between sound power or intensity and loudness can be roughly represented by converting power level readings in watts to the decibel scale:

$$\text{loudness (db)} = 10 \log_{10} \frac{P}{P_0}$$

where P is the sound power, and P_0 is a reference level. An approximation to musical pitch can be obtained by converting frequency to a measure representing pitch in the tempered scale:

$$\text{tempered pitch (semitones)} = 12 \log_2 \frac{f}{f_0}$$

where f_0 is a reference pitch of zero, and the octave above f_0 ($f = 2f_0$) is represented as "12" on the semitone scale. The same procedures can be followed for approximating musical pitch to other scales (Pythagorean, Just, etc.).

These transformations are attempts to represent the perceptual domain in terms of mappings from the vibrational domain. They are "first-order" approximations because they take into account the most important perceptual - vibrational links. However, the nuances of musical performance require that second- and perhaps third-order effects be represented in the mapping if prediction of the musically significant responses to a performance are to be achieved.

A known second-order effect is the change in pitch introduced by an intensity change in a tone held at constant frequency. Second order effects can be categorized as the predictable changes in one perceptual dimension due to changes in the vibrational correlate of another perceptual dimension. Seashore calls them the "normal illusions" of auditory perception. In terms of the diagram of auditory mapping, these effects would be indicated by the dashed lines connecting different levels (only those leading from "sound power" are shown). While many of these effects have been investigated for pure tones and static situations, their importance and application in a musical context have not, in most cases, been delineated.

Development of Models

Investigations in speech have shown the complex nature of the processing which enables us to decode the acoustic signal of spoken language. Constraints on natural language are imposed by both the production and receiving mechanisms. The perception of musical sounds is obviously subject to some of the same constraints, since the same receptors and, possibly, some of the same processing mechanisms are used for music as for natural language.

Perceptual parameters such as pitch, duration and tone quality, that are a regular part of the musician's vocabulary, are basic to an understanding of language processes (Lieberman, 1967). While the ear is the common receptor for speech, music and noise, the possibility has been raised that more than one mode of processing may be operating, that a speech/non-speech dichotomy may exist. The utility of basic brain-function

models can be better evaluated when data from research in music/speech perception are presented. For example, "hemisphere dominance" appears to be a uniquely human characteristic, which is a vital part of the human linguistic capability. One of the features associated with this phenomenon is the differential processing of auditory signals, depending on the type of signal input.

Music of all types consists of highly organized sets of complex auditory relationships which are processed by humans. An understanding of the feature extraction and organizing function of the brain for non-verbal auditory phenomena such as music may provide clues to the building and refining of models of human information processing. The pattern recognition and processing capability in language has been studied extensively. However, studies in linguistics and psychology have not developed testable models of information processing which isolate the unique characteristics of music.

While the evidence suggests that the auditory processing mode is dependent upon the type of input, the status of music in this apparent dichotomy has not been established. If there are two modes for processing auditory signals, the possibility exists that a perceptual disability, such as autism, may selectively interfere with only one of these processing modes. Knowledge of this condition may allow for the development of early diagnostic tests for the severity of the disability, and at the same time indicate the possibility of new communication techniques to improve the compensatory use of intact processing modes.

Knowledge of auditory processing modes is basic to an understanding of the acquisition of linguistic and musical skills by the normal child. The acquisition of reading skills and the development of tonal memory are basic educational problems certain to be affected by a knowledge of processing modes. In the course of normal development, the auditory processing mechanism apparently undergoes important changes up to about the age of five (Kimura, 1967). Beyond this age, at least with respect to auditory pattern recognition, we may all be "perceptually handicapped". The learning of basic sounds of language must occur before a

critical age for the discrimination to be fully integrated into the language function. On the basis of results reported by Kimura (1964), it is expected that this will also be true for music. Detailing the differences between the perception of speech and non-speech has begun to lead to specific models for the understanding of this developmental change.

II. RESEARCH DIRECTIONS: EXPERIMENTS AND MODELS

A. Performance Adjudication and Analysis

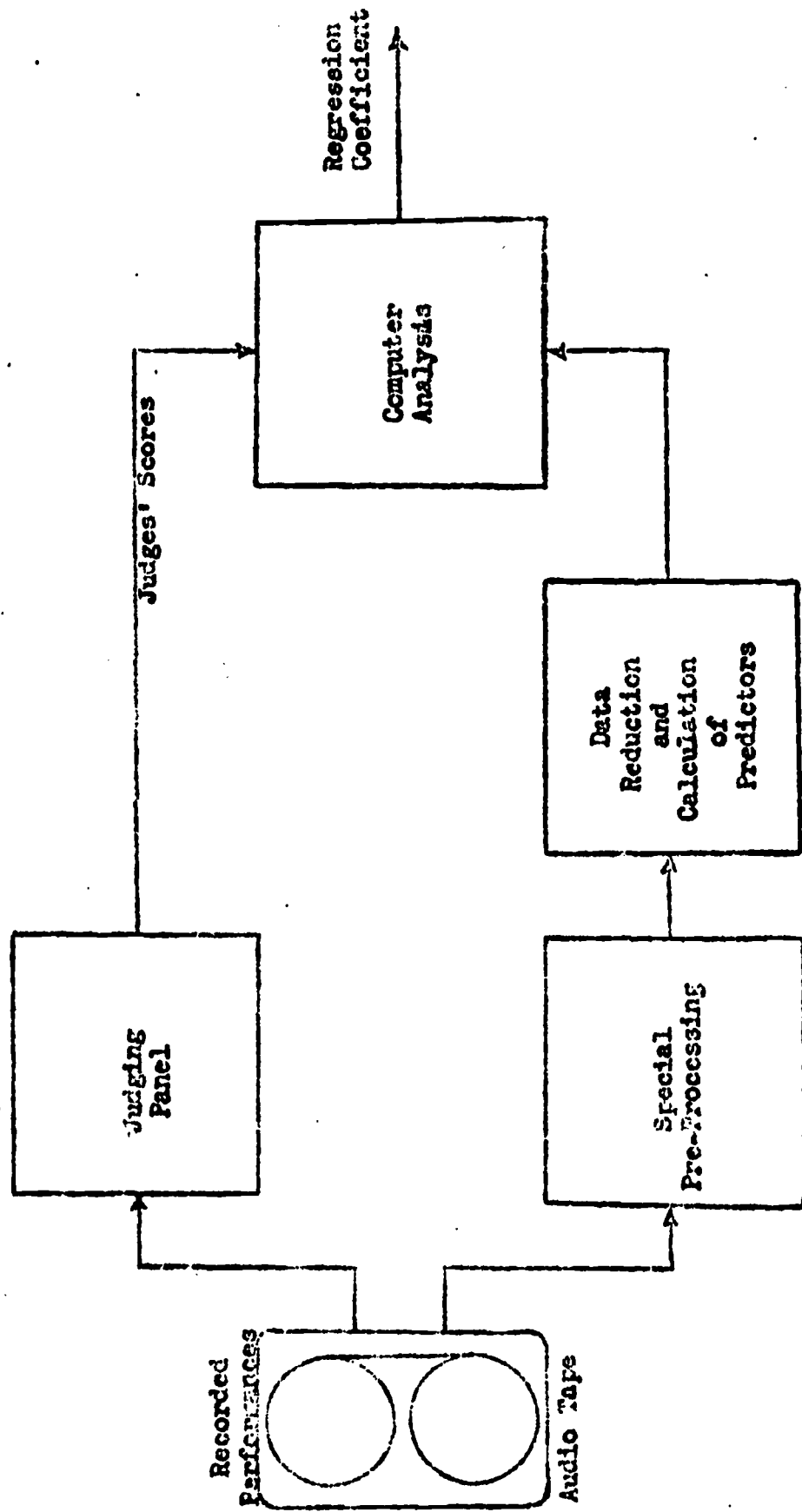
Evaluation of student achievement is a central problem at all levels of Education. Grading student output is a major chore for most teachers, and is often the one in which they find the least satisfaction. The problem is a particularly difficult one in music education, where the musical performance is the focus for student achievement. Experienced music teachers and performance judges contacted in the course of this project were skeptical of their own ability to maintain objectivity and uniformity under many of the conditions encountered in practical adjudication situations. No studies were found which dealt directly with the problem of the reliability of this type of evaluation of musical performance. Information is required concerning both the regrade reliability of a single judge, and the inter-judge agreement for groups of judges when dealing with various types of musical performances.

Studies of this type have a long history in the area of the grading of student essays. For example, studies by Findlayson (1951) and Phillips (1948) show a self-correlation among graders of between .60 and .70. That is, when a teacher is asked to regrade a set of essays after a time-lapse, the agreement between the two sets of grades is only 36 to 49 percent better than would be expected on the basis of chance alone. Interjudge correlations are in general even lower than this. Page and Paulus (1968) report inter-judge correlations ranging from .43 to .59 obtained by comparing the grades given by experienced judges on a set of student essays. An effort to improve the reliability of essay grading in an effective and practical way was reported by Page (1966) in an article entitled "The Imminence of Grading Essays by Computer."

The approach taken by Page uses a simulation of the human grading process through the use of a digital computer. A computer analysis of essay features that can be given an actuarial representation is used to define the score for a given essay. The averaged scores from several human judges is used as a criterion, since it has a higher

Figure 1.

PERFORMANCE ANALYSIS



reliability than the score given by a single judge, when the judges are all from the population of interest. The effectiveness of this approach, reported by Page and Paulus (1968), provided the stimulus for a similar application to the grading of musical performances, which is the subject of this section of the report. A pivotal premise for both studies is the possibility of defining variables which are useful correlates for the subjective variables influencing the human judge as he rates a given example of student output.

Statement of Purpose

This investigation is an attempt to add to the body of knowledge related to the following questions:

1. Are there operationally definable (i.e., objective, measureable variables for which musical performance limits can be established?
2. If so, what are they and where are the limits of acceptability for various performance situations?
3. Can a knowledge of such variables be made useful in solving the problems of music education?

Statement of Problem

Given a set of scores assigned by a group of competent human judges to a set of student musical performances, are there objective features of the recorded performance which can be used, with suitable computer analysis, to predict the averaged judges' score?

Method of Solving the Problem

The following steps were taken in this investigation: (See Figure 1.)

1. A sample of 62 recorded performances was selected from the 1967 Connecticut All-State music auditions.
2. A group of seven competent judges was hired to grade each performance on five categories. The

judges' averaged score in each category for each performance served as the criterion measure for computer grading.

3. Six objective features of the performances thought to be potentially useful for predicting performance quality were selected on an a priori basis. Specialized machine and hand techniques were developed to process the performances to extract the features. Values of the features associated with each note of each performance were punched on computer cards.

4. The digital computer was employed to process the feature values for normalization, recovery of initial data, and data reduction. Thirteen predictor values for each performance were derived from the original six features.

5. Two different performance standards were used in deriving the predictors, and the results compared.

6. The thirteen predictor variables were entered into a multiple regression program on the digital computer, and weightings were assigned to each variable in order to maximize the correlation between the predicted and the criterion grade.

7. The correlations were empirically cross-validated on subsets of the original sample in order to determine how well this program could be expected to predict scores for a new sample of performances.

Significance

The investigation of objective features of performance to determine their usefulness in predicting judges' responses is of importance in at least two major areas of music education, (i) the evaluation of student achievement and (ii) the development of practice and diagnostic sessions in which computer generated feedback supplements regular tutorial procedures (i.e., Computer Assisted Instruction).

Evaluation of Student Achievement

The evaluation of student musical performance by human judges would appear, from the results of this investigation, to be at least as reliable as the evaluation of student prose. Further confirmation of this result is required. The computer, given a valid regression equation, will produce an identical evaluation each time a particular performance is submitted and will apply ratings without the biases introduced into human judging by fatigue, effects of performance juxtaposition, knowledge of student personality, sex, race and appearance.

A successful simulation of the judging of musical performances by human auditioners has an immediate application to the problem of auditioning large numbers of student musicians for placement in schools and musical programs. Many states have "All-State" band, chorus and orchestra festivals to give experience and encouragement to music students. The process of auditioning hundreds of these students and to give each a fair chance to demonstrate his ability is a very difficult one. Computer grading may well provide the answer to this problem.

A direct comparison of student achievement or evaluation techniques between groups separated in either place or time has not been practical in the past. For example, it might be desirable to know how this year's students would be rated by last year's judges. The judges, as a group, may no longer be available. However, if an adequate sample of their scoring is available, a valid computer simulation would provide an accurate representation of their collective opinion on this year's performances, or on any other set of recorded performances.

This procedure is not limited to a particular standard of judgement. With judges representing several schools of thought or stylistic preference, categorization of performers with regard to stylistic suitability could be done effectively and efficiently. For example, a choral conductor might select an ideal grouping for a main and an echo choir, or, choose the voices most appropriate for a madrigal group, without the necessity

of individual auditions each time a new partitioning of the chorus is required.

Application to Computer Assisted Instruction

Kuhn and Allvin (1967), reporting on work by Spohn (1959) and Carlsen (1963) in the area of programmed instruction state (p. 2):

"As these two examples indicate, programming techniques for music instruction have been successfully undertaken as long as essentially verbal or symbolic response modes have been used, such as multiple choice, true-false, comparison, notation, and so forth. Typical response modes in such studies are marking, pointing, using a typewriter, and so forth.

Obviously, a new, direct, and essentially musical response mode is needed."

These two authors go on to describe an experiment in CAI in which a musical response mode is an essential part. Student vocal response was analysed with a "pitch extractor", and the resulting information was fed into an IBM 1620 computer for processing. Evaluation was made in terms of deviation limits from the true pitch, using equal temperament. Students were able to select limits of either a two percent or four percent range.

Deihl and Radocy (1969) report a similar effort, using an IBM 1500 Instructional System. Here, however, student response was off-line, but mention is made of the possibility of future on-line interaction.

These examples indicate the directions being taken in the application of CAI to music education. However, none of these studies comes to grips with a fundamental problem common to all attempts at computer interaction with the performer: the relationship between the acoustic variables being measured, and the subjective responses of competent listeners. Training in the control of an acoustic variable does not necessarily result in the musical control of a subjective variable.

The approach used in this investigation provides a test of the relationship between the acoustic parameters chosen, and the subjective responses indicated by the judges. Once a stable relationship has been discovered, a training procedure based on manipulation of the acoustic parameter can be developed, and the results can be anticipated with some assurance of success.

Methodology

The simulation techniques used in this experiment were patterned after those used in Project Essay Grade (Page and Paulus, 1968), an investigation supported by the U.S. Department of Health, Education and Welfare (Project No. 6-1318). Differences which occur between this project and "Project Essay Grade" are those imposed by the different mediums being explored, rather than any difference in approach. These appear primarily in the area of feature extraction. For the essay grading project, the data, consisting of hand-written student prose, was punched as literally as possible into IBM cards. Features were extracted by a computer analysis of the text: spelling, punctuation, word length, sentence length, etc. In the music project, features are extracted from a recording on magnetic tape of the performance to be analysed. For the essay grading project, judgements in five categories were predicted for 256 essays using thirty predictors. In the music project, judgements in five categories were predicted for 62 vocal performances using thirteen predictors.

Psychometrics and Statistics

The basic conceptual framework for evaluation and prediction in this investigation is found within the disciplines of psychometrics and statistics, as presented in many fundamental texts such as those by Hays (1963), Kelley (1948) and Winer (1963). Rozeboom (1966) was particularly useful in his discussion of homogeneity, reliability and validity. The programs for correlation were based on Cooley and Lohnes (1962). The step-wise program from the System/360 scientific subroutine package (IBM, 1968) was used for the multiple regression analysis.

Pattern Recognition

In a general sense, a simulation which involves responses to an input data set is a problem in the field of pattern recognition. In this case a pattern must be detected in the acoustical signal which can be used to predict the response of musical judges to that signal. A presentation of basic concepts in pattern recognition may be found in Sebestyen (1962) and in Nilsson (1965). Statistical models are an important part of pattern recognition procedures, with greater emphasis on decision making than is usually found in statistical usage. The extraction of features from which decision boundaries can be constructed is another fundamental process in pattern recognition, but there are no general algorithms for extracting useful features from a data base. Feature definition must be based on the attempts of previous investigations to find measurable aspects of the data field. Features which are useful for prediction are often specific to a given application. To be useful, features which are based on a nominal or ordinal scale must correlate, either separately or in combination, with the criterion. Nominal features must be evaluated using categorical decision processes. For this investigation, the features extracted were all given interval scale interpretations, and were based in part on the findings of studies in musical acoustics and the psychology of music, discussed in the next section.

Musical Acoustics and Special Equipment

The development of techniques for operationally defining important acoustic variables owes its greatest advancement in the 19th century to H.L.F. Helmholtz (1877). While others in this and earlier centuries laid the foundations for vibrational phenomena (Young (1784), Rayleigh (1877), Fourier (1822), etc.), Helmholtz explored the perceptual problems as well. Of the many recent publications in musical acoustics, books by Benade (1966) and C.A. Taylor (1965) have been particularly helpful in the present investigation.

At the turn of the century, the career of

Carl E. Seashore, who dominated the field until World War II, was just beginning. A frequently stated objection to many experiments in the psychology of music is that they are conducted in such a way as to remove them from the reality of musical performance. Studies in perception using isolated tones, pure tones and other convenient simplifications do not always generalize to real performance situations. Seashore tried to avoid this problem by working primarily from "live" as opposed to laboratory situations. The same attempt at relevance is made in this project.

In one of his first papers, A Voice Tonoscope (1902); Seashore described a device that was able to present visually, in graph form, the fundamental frequency of a live performance as a function of time. This device made it possible to measure systematically, variables thought to be related to musical perception, and to test hypotheses about them.

The equipment used to process the acoustic signal in this investigation, a modern version of the tonoscope, was planned and tested by J. Heller and W. Campbell. Its use in a related application, visual pitch matching, is described in a report by J. Heller (1969). A standard Frequency Modulation Sub-carrier Discriminator (Electro-Mechanical Research #287A-01), normally used in telemetry applications, was used as a frequency meter. It was modified by the manufacturer to operate over a pitch range of one octave from slightly above middle C to the next octave (266.7 Hz to 533.3 Hz, or $400 \text{ Hz} \pm 33\%$). Using a Schmidt trigger, the discriminator produces a DC output proportional to the input frequency, over the input range: $v = k(f - f')$ where v = output voltage, f = frequency of input signal, f' = center frequency, in this case 400 Hz, and k is the proportionality constant. The output voltage was presented as a function of time, on paper charts using an Esterline angus Speed Servo Recorder/S601 S.

A Sony Model TC-5600 Tape Recorder, with variable speed control, was used to adjust the performance pitch range to the acceptance range of the frequency meter. By noting the actual starting pitch for each performance, and entering these data into the computer program, all pitch and

time values were restored to correspond to the original performance speed. These calculations are detailed in Appendix B.

The choice of features which are related, at least indirectly, to the perceptually significant aspects of the acoustic signal is an essential part of this approach. Clues to the nature of these features can be found in the many sources in the psychology of music. Seashore (1939) provides a detailed account of the relationships between the objective and perceptual variables in his chapters on Pitch, Loudness, Duration and Timbre. He also enumerates (Seashore, 1938, p. 28) a list of basic principles in the psychology of music which are taken, with only minor changes, as fundamental concepts for this investigation. More recent studies (J.D. Harris, (1952), J.C.R. Licklider (1956), Robinson and Dadson (1956)) have provided details and further clarification of the illusions of hearing referred to by Seashore.

Summary

The first experiment completed under the grant was a computer simulation of an adjudication of musical performance. In this investigation, a panel of seven competent musical judges was asked to audition a set of sixty-two short vocal performances by students, which were recorded on audio tape. The judges were to respond by scoring each performance on a five point scale in several categories, such as intonation, dynamics, etc. The average of the judges' scores was found to be stable, and was the criterion for a computer simulation of the judges' response, using multiple regression analysis. The prediction of the judges' response to each performance was based on frequency, intensity and duration measurements of the performance.

This approach provides a means of examining the relationship between performance competence, as rated by experienced judges, and the vibrational characteristics of performance. Specific vibrational patterns can be tentatively identified as correlates of a particular response from the panel of judges. Synthesis of these vibrational patterns can then be used to test the possibility of a cause and effect relationship between the

pattern and the response. The results of this experiment were very encouraging, particularly in view of the highly simplified (linear) predictive model employed.

Some idea of the simulation capability achieved in this first attempt can be seen in the following table. The numbers indicated are typical correlation values. A correlation of unity means perfect agreement, and a value of zero indicates a random relationship.

<u>Score Comparison</u>	<u>Correlation</u>
Between two groups of judges (group averages)	.85
Between computer scores & judges' average	.65 (shrunk mult-r)
Between one judge and the group average	.65
Between two judges	.40

From this it can be seen that the computer comes as close to predicting the group average as the typical experienced judge (.65). The correlation between any two judges (.40) is considerably lower than this value. Refinements in technique are expected to further improve computer prediction of judges' scores. A complete report of this experiment appears in Appendix I.

B. Performance Modification and Listener Response

Introduction

A second experiment completed during the grant period was designed to determine the effect of the attack transient upon the recognition of instrumental timbres. This study approached the problem by mechanically replacing the attack of one instrument by the attack of another, to determine if the listener is influenced in his attempt to recognize instruments more by the attack or by the steady-state portion of the tone. This controlled modification of natural sounds was used to test hypotheses regarding differences between instrumental

timbre recognition for normal, altered, and "no-attack" tones.

Related Literature

Several previous studies of timbre have recognized the importance of the attack transient in timbre recognition. Seashore (1938) alludes to a study by Lewis and Cowan (1937) who mechanically replaced even vocal releases (decay) with gliding attacks. "The musically acceptable glide at the beginning of the tone became utterly intolerable when placed at the end of a tone." Seashore does not investigate the effects of attack or decay on sonance (tone quality), but states, "This experiment opens a very fertile field for the investigation of reasons for adaptive or habitual hearing."

Nolle and Boner (1941) found, in investigating the initial transients of organ pipes, that "these initial transients seem to be important in determining the subjective character of pipe organ music and in differentiating between the subjective character of pipe organ music and of music produced by electronic instruments."

In synthesizing instrumental tones, studies by Fletcher, Blackham, and Stratton (1962), Strong and Clark (1967), and Risset and Mathews (1969) have stated that the attack plays an important role in recognition of the instrument.

On the basis of analyses of the harmonic content of various wind instruments, with one result being that little substantial difference was found to exist between the harmonic content of certain instruments, Saunders (1946) concluded, "These facts lend support to the idea which is often expressed that an oboe and a violin would be indistinguishable if one were prevented from hearing the beginnings or the endings of the tones."

Richardson (1954), in an investigation of transients by spectrum analysis, states, "In spite of their evanescent nature, the view is now held that it is these transients which enable the listener to distinguish the sounds of different musical instruments or between two of the same class. The transient is indeed part of the 'formant'

of an instrument, and ought to be exhibited as a characteristic alongside the steady-state spectrum."

These studies have all indicated, to one degree or another, that some importance may be attached to the attack as an indicator of timbre. However, all of these studies were performed under conditions in which the evaluators were aware of what timbre was being considered. Under such circumstances an evaluator's judgment would tend to be based on an ideal concept of the particular tone being judged, rather than a mere identification of timbre.

A more valid approach to measuring the importance of attack to timbre recognition would be to present subjects with unidentified tones, with and without attacks, to be identified by the subjects. This method has been utilized in studies by Berger (1964), and Saldanha and Corso (1964).

Berger presented 30 subjects with the tones of 10 different wind instruments. The tones were presented in several forms: unaltered, played backwards, attack and decay removed, and all harmonics except the fundamental filtered out. The results showed 59 per cent of the unaltered tones identified correctly and 35 per cent of the tones minus attack and decay identified correctly. The tones played backwards and the filtered tones were correctly identified 42 per cent and 18 per cent of the time, respectively.

Saldanha and Corso, in a similar study, presented 20 subjects with the tones of 10 string and wind instruments, unaltered, and with various alterations. The results showed 41 to 44 per cent of the unaltered tones correctly identified (the two figures are a result of longer and shorter steady-states), and 32 per cent of the tones with no attack identified correctly. Both of these studies indicated a strong relation between attack and timbre identification, but no statistical analyses of the results are provided to confirm this fact, nor is the effect of the quality of the attack investigated.

In spite of the work which has been done in this field, one finds that the generally accepted definition of timbre remains as some exclusive

function of the harmonic content. Lundin (1967) discusses timbre strictly in terms of the steady-state portion of the tone, and Neilson (1970) describes timbre as the "relationship of the decibel strength in the fundamental of a given tone to that of various overtones."

The present study attempts to demonstrate whether or not the initial transient is an integral part of timbre recognition, and, if it is, to investigate what effects changes in the quality of the attack have on the recognition of timbre.

Procedures

A stimulus tape was prepared which included 120 tones. Four instruments, (flute, oboe, clarinet, and trumpet) each playing three pitches, (d', c'', gb'') were recorded. Each of these 12 tones was modified by replacing the normal attack by the attack of the other three instruments on the same pitch. A second modification of these 12 initial tones was made by eliminating the attack portion of each tone. This produced 48 tones. The original 12 tones (with no modification) was also included in the final tape. Each of these 60 tones were recorded a second time and placed on the final tape in a random sequence.

The stimulus tape was administered to three groups of subjects, high school instrumentalists (n=57), college students enrolled in an introduction to music history course (non-music majors, n=43), and college music majors including several music faculty and professional musicians (n=38).

Results

The flute and clarinet steady-state tones were identified correctly (82% and 79% respectively) more than the oboe and trumpet steady-state tones (70% and 77%). The identification of the flute was slightly less accurate than that of the clarinet when preceded by other attacks, but was more accurate when preceded by its own or no attack. The attack portions of flute and clarinet, however, are identified quite differently

from each other. The flute attack was least often correctly identified and the clarinet attack most often correctly identified. This indicates that the flute provides very strong identification information in its steady-state, but very little in its attack.

The oboe was the least recognizable instrument when presented without modification (70% correct). The trumpet was recognized somewhat better than the oboe (77% correct). With the removal of the attack, the trumpet dropped well below the oboe in degree of recognizability. With the addition of attacks of the other instruments oboe identification scores dropped considerably (from 63% to 51%), the trumpet scores only slightly (from 45% to 42%), nevertheless, the trumpet remained the least accurately identified tone quality. The oboe steady-state was most influenced by the attacks of other instruments, while the oboe attack had the least influence on the steady-states of other instruments.

The trumpet attack provided a great deal of identification information in combination with the trumpet steady-state. While in combination with other steady-states, the trumpet attack did not provide as much identification information as the clarinet attack, it provided more information than flute and oboe attacks. The trumpet steady-state by itself or in combination with other attacks was easily confused or identified as the attacking instrument, and the trumpet attack with other steady-states did not provide as much information as one might anticipate, in light of its effect on the trumpet steady-state.

The overall results of this study show that identification of timbre becomes less accurate (statistically significant beyond the .01 level) as the tone progresses from normal, to no-attack, to altered. That is, the attack affects aural recognition of timbre. (See Appendix II for a detailed discussion, and for tables of results.)

C. Information Processing and Musical Perception

It is a commonly held notion that speech and music occupy different levels of brain function. Speech perception is considered to be a highly ordered process, involving complex decoding and pattern recognition techniques. Music, by contrast is often considered to be a visceral activity, involving the lower brain centers, the emotions, and the rhythms of body movement and heart-beat.

This research is an initial attempt to test an entirely different model, in which music and speech are considered as analogous functions. That is, that musical perception is also a highly ordered process, subject to some of the same constraints and peculiarities that are a basic part of speech perception.

Man is the only mammal whose brain shows evidence of a strong functional asymmetry between the cortical hemispheres (Sperry, 1964). This unique adaptation, which reduces the redundancy and increases the functional capability of the brain, is closely related to the development of language. There is very strong evidence that between the ages of 1 and 6 in the human child, the language function is (normally) taken over by the left hemisphere, and that some musical and higher order visual processes are performed in the right hemisphere.

Evidence for the functional asymmetry of the human cerebral hemispheres comes primarily from four separate areas of investigation:

1. Tests of functional limitations in brain damaged patients for whom the location and extent of damage is known (Luria, 1970; Mountcastle, 1962).
2. Electro-encephalographic studies (Cohn, 1971).
3. Tests of epileptic patients who have undergone a sectioning of the corpus callosum (the inter-connecting structure between the cerebral hemispheres). (Gazzaniga, 1967).
4. Experiments on normal subjects using the technique of "dichotic presentation" (presentation of simultaneous but different stimuli to each

ear). The success of this technique in determining the locus of processing for a given input stimulus, is based upon the hypothesis that the representation for the stimulus presented to each ear is greater for the contralateral than for the ipsilateral hemisphere. (Shankweiler, Studdert-Kennedy, 1967; Kimura, 1961; Knox and Kimura, 1970). Many experiments with speech stimuli have been conducted using the dichotic technique, but only two have involved musical stimuli (Kimura, 1964; Shankweiler, 1966) presenting different melodies to each ear, with a left ear recall preference.

Because of the very limited nature of the investigations specific to music, and the many statements linking music variously with speech and non-speech (Warren, 1971) a pilot study was instituted in order to establish a basis for tentative brain function models in music. Shankweiler and Studdert-Kennedy (1967) showed a right ear superiority for consonants, but none for vowels. They stated "in view of Kimura's finding (1964) of a left ear advantage for musical melody recognition, as against a right ear advantage for spoken digits, the neutral status of steady state vowels, midway, as it were, between speech and music, is perhaps not surprising." (p. 60, Shankweiler and Studdert-Kennedy, 1967).

Melody recognition is of course only one of a large number of responses available to a moderately "literate" listener. It is possible, using plausibility arguments, to construct a comprehensive multi-level analogy between speech and music, from the phoneme level to the breath group and sentence. The analogy is, however, only an interesting exercise if it is not found to be in some way predictive of similar perceptual processes. The corresponding structural elements in music then, by implication, involve cues for decoding and defining the structure generating a particular musical "message".

The phoneme is the building block for natural language, with classification categories of vowels, consonants, semi-vowels, liquids, etc. In speech, the consonant has been shown to require a complex decoding process involving primarily the left

hemisphere (Broadbent and Gregory, 1964). Spectral envelope changes occurring within 20 to 60 milliseconds provide the primary identification cues (Liberman, et al., 1967). The musical analogy to the consonant/vowel/consonant sequence is the attack/steady-state/decay sequence for an instrumental tone. In speech, the consonant environment conditions the perception of vowels, and consonants provide a much greater reduction of uncertainty than do vowel sounds. A number of studies have shown that the attack portion of an instrument tone also has these attributes (Stumpf, 1926; Tenney, 1965; Risset and Mathews, 1969; Strong and Clark, 1967a, 1967b).

Since the phoneme is the fundamental speech segment, features such as length, stress and tone that span more than one segment are designated "suprasegmental". These suprasegmental features have direct counterparts in the musical phrase: stress, rubato, portamento, articulation, dynamic contour, timbre modification.

The musical counterpart to syntax has been extensively documented under the general category of melodic and harmonic structure.

In order to test the analogy it seemed appropriate to start at the most basic level, that of the phoneme. The method of dichotic presentation provides a clear operational test of the music/speech analogy. Are instrument attacks more readily identified when presented to the right ear, with the implication of left hemisphere processing?

To answer this question, six musical instruments (trumpet, violin, guitar, oboe, clarinet, flute) were recorded while producing a concert "middle-C" (nominal frequency, 261.6 Hz). These instruments were chosen to represent the major classes of sound production in music, with the exception of percussion. The violin was bowed and the guitar plucked. The flute was replaced with a small bottle tuned to 261.6 Hz, because of the very long attack transient for this low note in the flute's range.

The "attack" portion of each tone was defined by the experimenter on the basis of oscilloscope traces. The durations chosen were:

trumpet, 45 ms; violin, 43 ms; guitar, 33 ms; oboe, 47 ms; clarinet, 55 ms; "flute", 65 ms. The tones were recorded a second time with the steady-state portion of the original instrument replaced by an electronically produced triangular wave at approximately the same pitch and phase (at the transition point) as the original signal. Channel switching was accomplished using voltage controlled amplifiers, envelope generators and followers, trigger delays and oscillators from a Moog synthesizer.

The six hybrid sounds were then aligned and recorded on two separate tape tracks, so that all 15 pairings of dissimilar instruments were represented. A random sequence of 30 items (each stimulus presented twice, with channels reversed on the second presentation) was prepared, along with a single channel training sequence. Four subjects were taught to label the sounds reliably (the composite tones could be discriminated, but were not initially identifiable). They then listened twice to the random sequence of 30 stimulus pairs presented dichotically. For the

second presentation, the earphones were reversed, to correct for any imbalance between the two electronic channels. Subjects were asked to identify both instruments in a stimulus pair, guessing if necessary. The scoring was based on responses for which only one instrument was correct. All subjects responded at well above the chance level: the average number of presentations (out of sixty) for which at least one instrument was correct was 52.

All subjects showed a right-ear advantage for identifying instrument attacks. The right ear average was 24.25 correct responses, compared to a 14.25 average for the left ear (difference significant at $p < .05$). This result lays the foundation for further experiments to detail the processing analogy between speech and music.

The research in speech over the last 20 years has progressed steadily on the problem of human information processing. The study of the perception of music has received little attention from this point of view, and comparative music/speech studies are difficult to find (Slawson, 1963; Kimura, 1964). This neglect is

possibly due to the caricature of music, fostered even by musicians as a "visceral", "non-intellectual" activity.

It is our contention that music is as complex a pattern recognition activity, even for the casual listener, as is speech. In fact, it seems reasonable to suppose that the structure of music may provide a more useful map of the brain's symbol manipulation process than speech, since fewer constraints (cultural, articulatory, referential) are imposed upon the development of musical patterns and modes of production. Implicit in this surmise is the idea that both speech and music map to subsets of a generating structure which is more general than the structures delimited by "speech" and "music". In addition, this general structure, while very large in terms of storage, and very flexible in terms of pattern organization, cannot be considered infinite. In 1963, H. Bremermann showed that the fundamental coarseness of matter does not allow it to transmit more than 1.6×10^{47} bits per gram per second. Ashby has pointed out that, for combinatorial interactions, even a very moderate-appearing situation, such as a square screen of 20 by 20 lamps, provides enormously more possible patterns than those that could be processed by any device the size of the brain, over a lifetime. The number of discriminable patterns or pattern classifications must be very small indeed, compared to these fundamental limits.

In speech the recognizable articulatory patterns are limited to about 60 phonemes for all the world's languages and only a subset of these is used in any one language. In addition, their identification is not fixed, but is, to some extent, context dependent. In music, the "phonemic" elements are tones of distinct pitch, timbre, loudness and duration. Sequences and combinations of listener "enjoyment" or appreciation is based on the ability to recognize the sequence of tones as belonging to a class of patterns previously delimited (in memory) by implicit rules of musical grammar. It should be emphasized that for the general listener, this recognition may not be recoverable; that is, he cannot specify the basis upon which his decision is made.

This is analogous to the language auditor who responds "Da" when "Ba" has been presented, but has no awareness of the distinctive features which he uses to discriminate "Ba" from "Da".

Referring to a higher linguistic level, Langacker (1968, p. 234) states:

When the child has learned to talk, when he has mastered his native tongue, he is in possession of an abstract system of rules that specify an unbounded class of well-formed sentences. He is not conscious that he possesses this system in the sense that its structural patterns have been imposed on his psychological processes, so that these patterns are a factor in determining the course of his verbal activity. Learning to talk, like learning to ride a bicycle, involves the mastery of a set of principles; it involves the addition of structure to the body of psychological skill or competence that shapes our mentally directed behavior. These rules are thus no more accessible to conscious inspection than the rules for keeping one's balance while riding a bike. We talk and we keep our balance on bicycles, but in neither case do we know, at the level of consciousness, precisely what the guiding principles are.

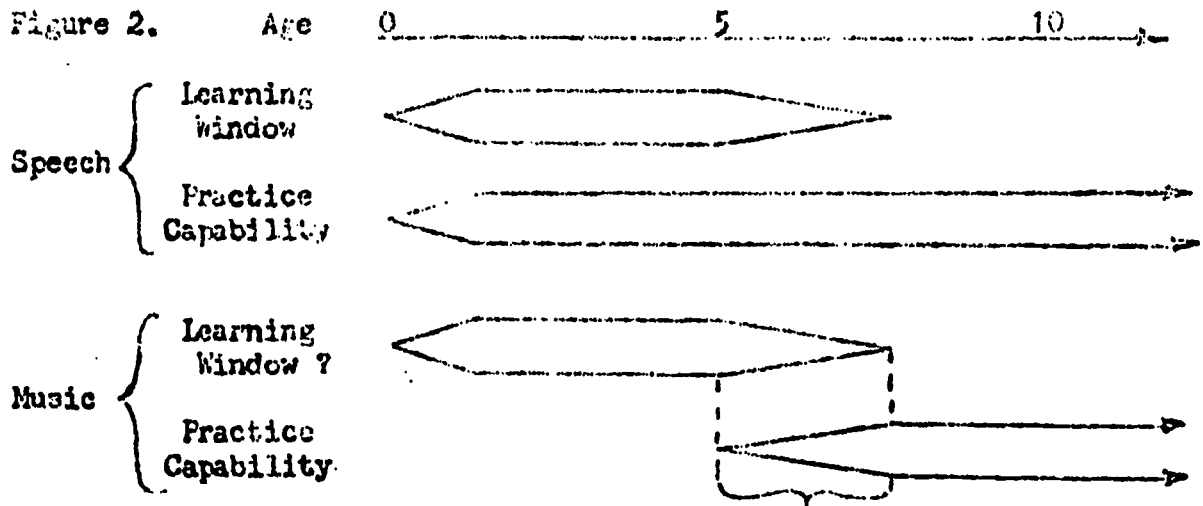
It is clear that for language the developmental period from birth to about six years old is a very critical one, in the sense that sound patterns not learned during this period are only learned with great difficulty, if at all, at a later age. A similar argument applies to syntax, although at this level, usable *ex post facto* rules can be devised to approximate the structure "discovered" by every normal child before the age of five.

This period (to approximately age 6) may provide a "learning window" for auditory pattern classification that is closed when the hemispheric asymmetry development is completed. In the case of language, participation in language production undoubtedly provides a necessary step in the formation of a complete language generating and decoding system. Sussman (1972) states "the

versatility of the human speech production system and an increasing body of evidence suggests that speech is controlled by an intricate closed-loop feedback system. To bring about feedback control of the speech musculature, the higher neural centers should be kept constantly aware of (a) the spatial position, (b) the direction of movement, and (c) the rate of movement of the articulators. This review describes the feedback mechanisms existing within the tongue that can mediate such dynamic space-time information."

Traditional means of musical production depend upon control systems (finger, hand, and arm muscles; breath control contrary to that needed for speech) which are not sufficiently developed in the young child to allow his exploitation of musical patterns, to the degree that he explores speech. The "learning window" idea provides testable hypotheses concerning musical "talent" (musical facility demonstrated in production). There may, for example, be a critical overlap between the "learning window" and the means for exploring, and extemporizing musical patterns. The anecdotal immaturity of many professional musicians may have a close relationship to their musical ability: a delayed "closing" of the learning window may cause sufficient overlap with the developing muscle control to provide the necessary feedback contingencies (see Figure 2).

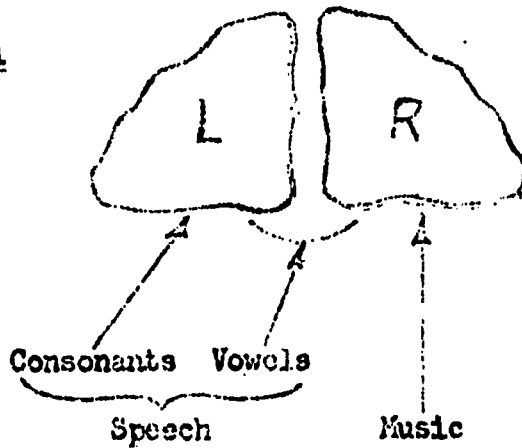
This model appears to have broad implications for many areas of human intellectual development. The role of musical pattern recognition in this model is dependent upon its possible inclusion under speech or speech-like processes. If the experiments described in this report show conclusively that some non-speech sounds are decoded in the left-hemisphere (and if, in addition they are found to produce categorical decisions: Liberman, et al., 1962) then it is unlikely that left hemisphere processing is completely limited by a finite store of feature extractors possessing some invariant relationship to a linguistic category. It is more likely that a general classification cue, such as a critical rate of change in the spectral envelope, determines the processing site. A synthesized set of sounds with variations of temporal and spectral attack parameters is being prepared in an attempt to isolate a processing cue which is not exclusively linguistic.



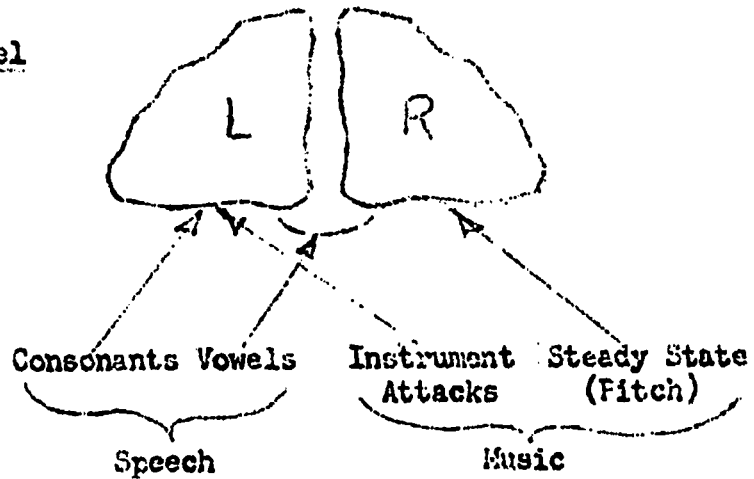
Hypothesis: A critical overlap for organizing auditory pattern and nuance structure for music?

Figure 3.

Previous Model



Proposed Model



In view of our findings taken in conjunction with those of Kimura (1967) and others, the music/speech analogy may have implications for reading problems, such as congenital dyslexia, where anecdotal accounts of unusual musical ability (non-notational) have up until now been considered as primarily compensatory adaptations. The problems of autistic children appear to be closely connected with an inability to encode language ("Autism: A Deficiency in Context-Dependent Processes?", Pribram, 1970). Anecdotes of musical precocity in some autistic children, and the role music plays in some treatment centers such as Benhaven in New Haven, Conn., indicate a close, complex relationship between speech and music for this disability. It seems to be of basic importance, however, to investigate the scope of left hemisphere auditory processing, particularly when it appears to extend beyond the range of stimuli considered in previous investigations. (See Figure 3.)

D. Systems for Performance Analysis and Modification

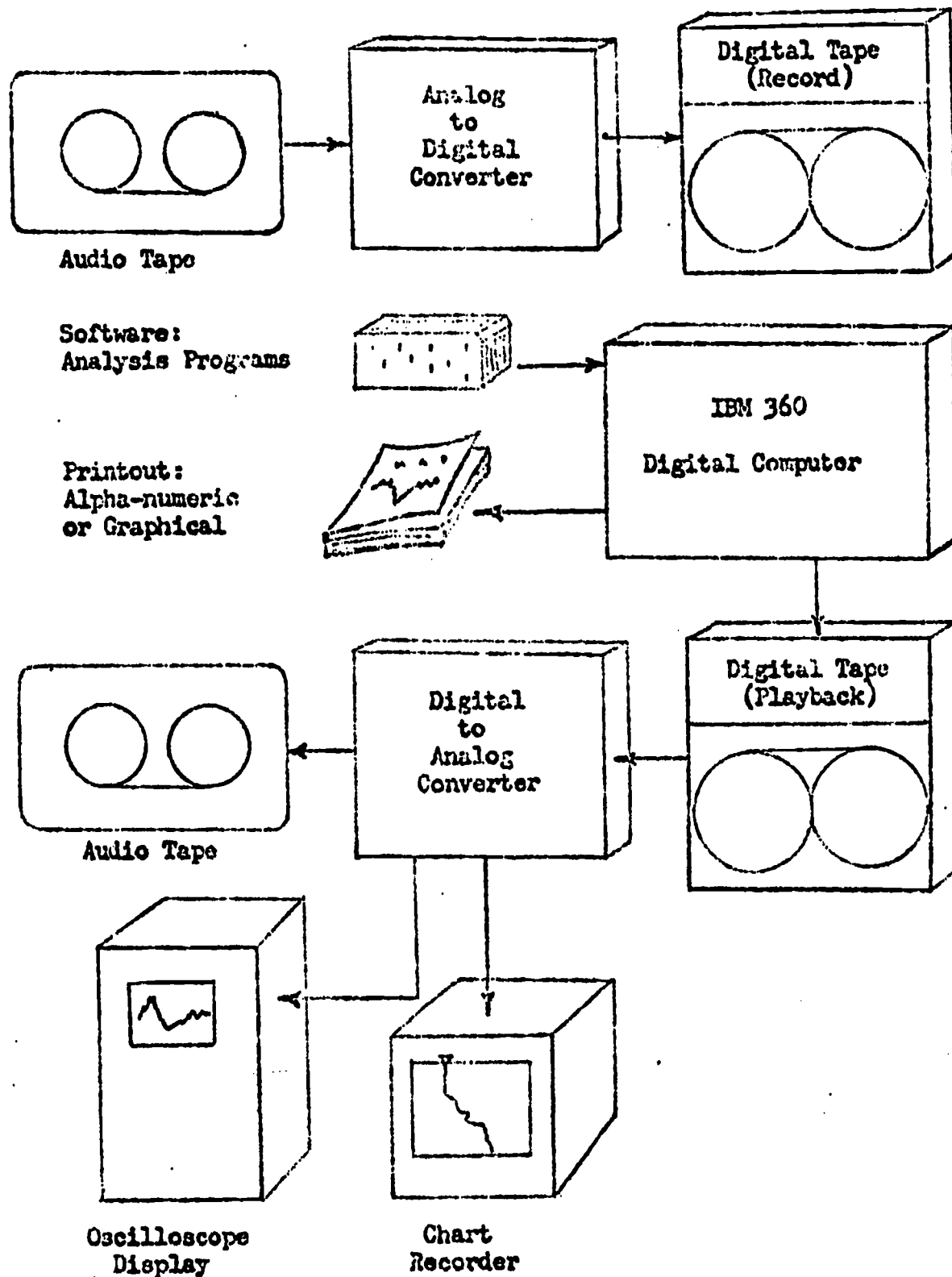
1. The Computer Analysis System

A large capacity digital computer provides a wide range of processing capabilities and is a necessity for reducing large amounts of data to useful descriptors. In order to prepare auditory material for digital processing, an analog to digital converter must be used. The analog (continuously varying) signal from an audio tape is converted to digital (discrete) values and is stored on magnetic tape in a form suitable to the input requirements of an IBM 360 digital computer. Two input formats are available:

- a. Five bits + sign, at 20K words per second, which provides a 30 db dynamic range up to 10K Hz.
- b. Nine bits + sign, at 10K words per second, which provides a 54 db dynamic range up to 5K Hz.

Duration, frequency, intensity, harmonic and

Figure 4.



COMPUTER ANALYSIS SYSTEM

noise analysis can readily be accomplished with the appropriate computer programs. In addition, programs for performance modification open many possibilities for investigation, since performances on traditional instruments can be altered slightly for controlled effects without destroying the naturalness of the processed sound. Figure 4 shows the system schematically, and Appendix III describes the software developed for the system.

2. Tone Line Writer

The tone line writer produces a visual record on graph paper of the fundamental frequency and amplitude contours of a recorded performance, over a three octave range. (133 Hz to 1067 Hz; approximately C below Middle C to the soprano high C). A modified FM Subcarrier Discriminator (a standard telemetry device) serves as an accurate frequency meter. An octave switch, designed specifically for this system, automatically sets the proper range, substantially reducing the problems associated with tones containing strong second and third harmonics.

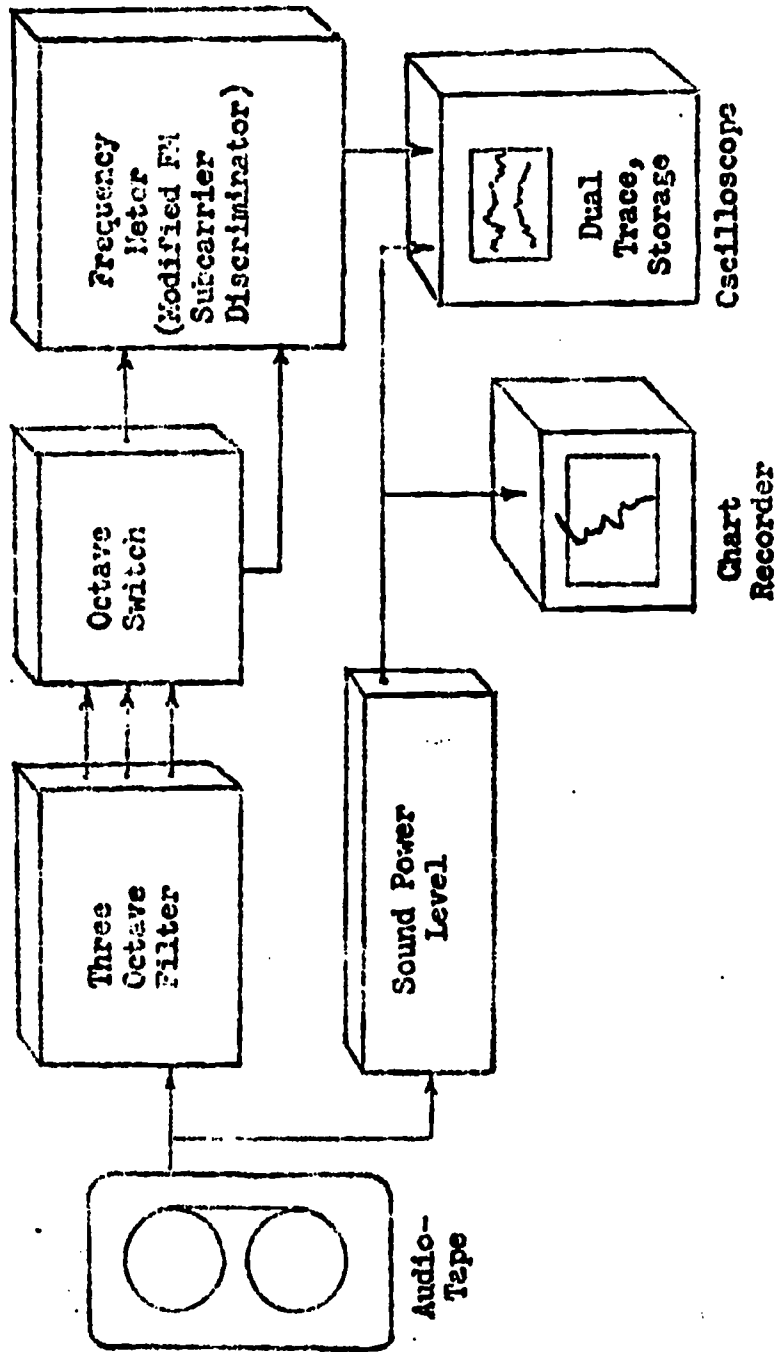
The graphs produced by this system allow visual inspection and measurement of the durations of tones and silences, and the vibrational correlates of vibrato (pitch modulation), tremelo (loudness modulation) and other vibrational characteristics. Figure 5 is a schematic of the tone line writer.

3. Tone Line Reader

The tone line reader accepts hand drawn charts representing pitch and loudness contours, and converts them to an auditory output by means of a voltage controlled synthesizer, such as the Moog. The optical scanner or "chart reader" is a modified document transmitter built by Graphic Transmission Systems, Inc. A two level signal from the scanner indicates the presence or absence of a line on the chart. The sample and hold unit determines the voltage appropriate for a line-indicating-pulse at each point in the scan, and then presents that voltage, when there is a pulse, to the proper sound control unit.

Chart paper is fed to the scanner at one inch per second, and sixty scans are made each second.

Figure 5.



TOE LINE WRITER

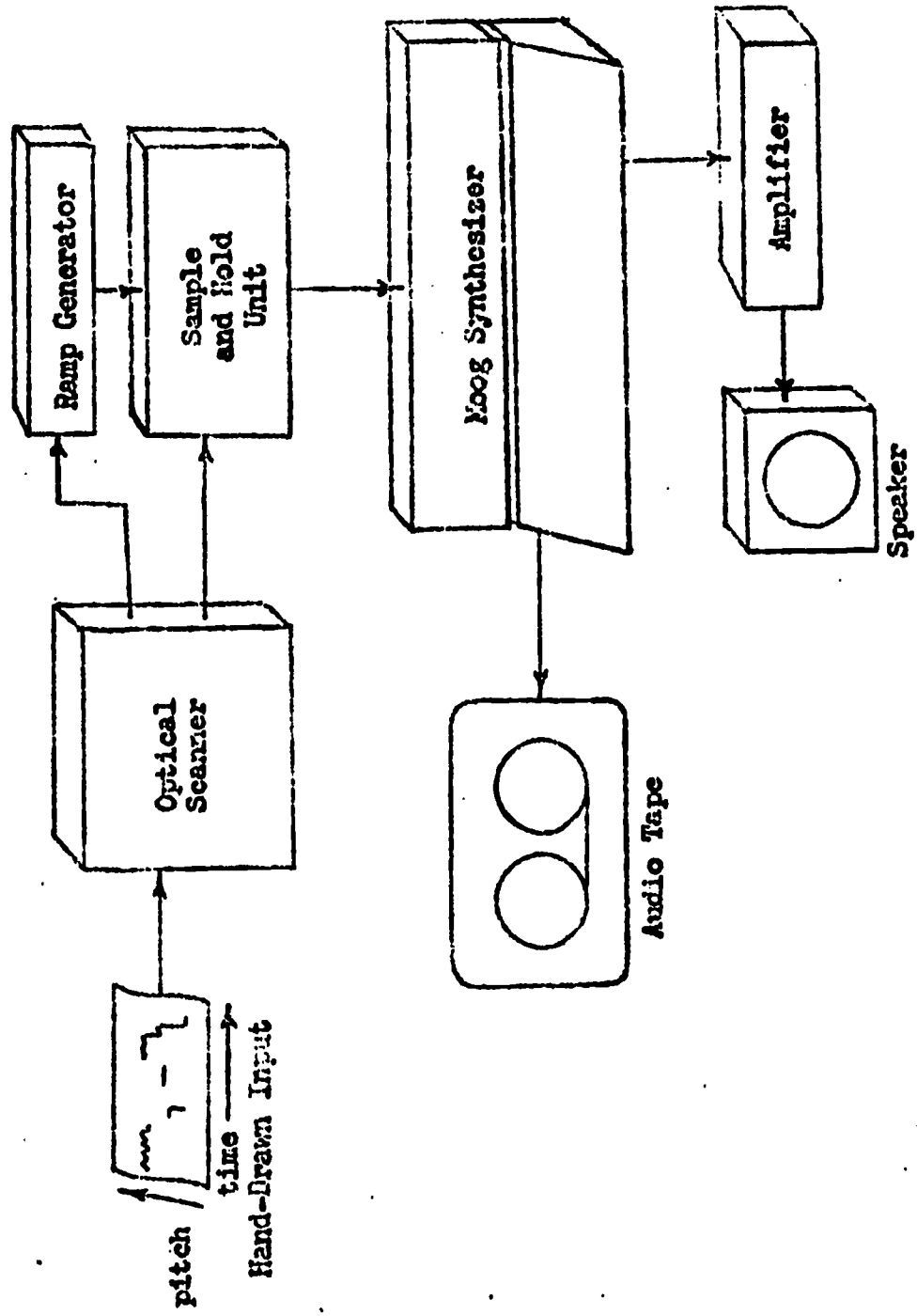
Both pitch and loudness indications can be scaled at will within the limits of the 8 1/2 inch scan width. The system is diagramed in Figure 6. The electronic interface, which converts the output of the optical scanner to a suitable control voltage, is described in Appendix IV.

4. Harmonic Synthesizer

An harmonic synthesizer with both phase and harmonic amplitude control would provide an invaluable tool for investigating the timbre changes necessary and sufficient for a particular listener response to a musical sequence of tones. Basic research on the pitch analysing processes used by the ear, the appearance of subjective tones, and the tonal characteristics of instruments would be greatly facilitated by such a device.

In view of this, a graduate student in the University of Connecticut Electrical Engineering Department was engaged to provide a prototype design for the device, and to simulate the output of the device using a digital computer. The results of this investigation are presented in Appendix V. The cost of building the synthesizer was outside the limits set by the present funding.

Figure 6.



TONE LINE READER

REFERENCES

- Benade, A. (1960) Horns, Strings and Harmony. Garden City: Doubleday.
- Berger, Kenneth W. (1964) Some Factors in the Recognition of Timbre. The Journal of the Acoustical Society of America. 36, 1888-91.
- Bremermann, H. J. (1963) Optimisation through evolution and recombination. In Self-Organizing Systems, ed. Yovits. Washington, D.C.: Spartan Bks. p. 93.
- Broadbent, D.E. & Gregory, M. (1964) Accuracy of recognition for speech presented to the right and left ears. Q.J. exp. Psychol. 16, 359-60.
- Carlsen, J. (1963) Programmed learning in melodic dictation. J. of Research in Music Education, 12 (Summer), 139-148.
- Cohn, R. (1971) Differential Cerebral processing of noise and verbal stimuli. Science, 172, 599-601.
- Cooley, W. and Lohnes, P. (1962) Multivariate procedures for the behavioral sciences. New York: Wiley.
- Deihl, N. and Radocy, R. (1969) Computer assisted instruction: Potential for instrumental music education. Council for Research in Music Education, Bulletin No. 15, Winter.
- Findlayson, D.S. (1951) The Reliability of the marking of essays. British Journal of Educational Psychology, 21, 216-234.
- Fletcher, Harvey. Loudness, Pitch and the Timbre of Musical Tones and Their Relation to the Intensity, the Frequency, and the Overtone Structure. The Journal of the Acoustical Society of America, 6 (1934), 59-69.
- Fourier, J. (1882) Theorie Analytique de la Chaleur.
- Gazzaniga, M.S. & Sperry, R.W. (1967) Language after section of the cerebral commissures. Brain, 90, 131-148.
- Harris, J.D. (1952) Pitch Discrimination. J. Acoustical Society of America, 24, 750-755.

- Hays, W.L. (1963) *Statistics for Psychologists*. New York: Holt, Rinehart and Winston.
- Heller, J. (1969) Electronic graphs of musical performance: a pilot study in perception and learning. *J. Research in Mus. Educ.* 17, No. 2, 202-216.
- Helmholtz, H.L.F. (1877) *On the Sensations of Tone*. Re-issued: New York: Dover, 1954.
- Kelley, T.L. (1947) *Fundamental Statistics*. Cambridge: Harvard University Press.
- Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychology*, 15, 166-171.
- (1964) Left-right differences in the perception of melodies. *Q.J. of exp. Psychology*, 16, 355-360.
- (1967) Functional asymmetry of the brain in dichotic listening. *Cortex*, 3, 163-178.
- Knox, C. & Kimura, D. (1970) Cerebral processing of non-verbal sounds in boys and girls. *Neuropsychologia*, 8, 227-237.
- Kuhn, W.E. and Allvin, R.A. (1967) Computer-assisted teaching: a new approach to research in music. *J. of Research in Music Education*, 15 (Winter), 305-315.
- Langacker, R.W. (1968) *Language and Its Structure*. New York: Harcourt, Brace & World, Inc.
- Lewis, Don, and Cowan, Milton (1937) Pitch Variations arising from Certain Types of Frequency Modulation. *JASA.* 9, 79.
- Liberman, Cooper, Harris & MacNeilage (1962) A motor theory of speech perception. In, *Proc. of the speech communication seminar*. Stockholm: Royal Institute of Technology.
- Liberman, Cooper, Shankweiler, & Studdert-Kennedy (1967) Perception of the speech code. *Psych. Review*, 74, 431-461.
- Licklider, J.C.R. (1956) *Auditory Frequency Analysis*. In: Cherry, C., Ed. *Information Theory*. London: Butterworth, 253-268.

- Lieberman, Philip (1967) Intonation, Perception, and Language. Cambridge: MIT Press.
- Lundin, Robert W. (1967) An Objective Psychology of Music. 2nd ed. New York: Ronald Press Co.
- Luria, A.R. (1970) Traumatic Apasia. The Hague: Mouton Press.
- Mountcastle, V.B. (1962) Interhemispheric Relations and Cerebral Dominance. Baltimore: Johns Hopkins Press.
- Neilson, James (1970) Timbre and Color. The Instrumentalist, April, pp. 39-41.
- Nilsson, N.J. (1965) Learning Machines. New York: McGraw Hill.
- Nolle, A.W., and Boner, C.P. (1941) The Initial Transients of Organ Pipes. The Journal of the Acoustical Society of America, 13, 149-55.
- Page, E.B. (1966) The imminence of ... grading essays by computer. Phi Delta Kappan, January, 238-243.
- Page, E.B. and Paulus, D.H. (1968) The Analysis of Essays by Computer. Final Report, HEW Project No. 6-1318, Storrs: The University of Connecticut.
- Phillips, G.E. (1948) The marking of children's essays. Forum of Education, 1, 19-29.
- Pribham, K.H. (1970) Autism: A deficiency in context-dependent process? In, Proceed. of 1970 Conf. of Nat. Soc. for Autistic Children. HEW.
- Rayleigh, Lord. (1877) The Theory of Sound. Re-issued: New York: Dover, 1937.
- Richardson, E.G. (1964) The Transient Tones of Wind Instruments. The Journal of the Acoustical Society of America, 26, 960-62.
- Risset, J. & Mathews, M.V. (1969) Analysis of Musical-instrument tones. Physics Today, 22, 23-30.
- Robinson, D.W. and Dadson, R.S. (1956) A re-determination of the equal-loudness relations for pure tones. Brit. J. Appl. Physics, 7, 166-181.

- Rozeboom, W.W. (1966) Foundations of the Theory of Prediction. Homewood, Ill.: The Dorsey Press.
- Saldanha, E.L., and Corso, John F. (1964) Timbre Cues and the Identification of Musical Instruments. The Journal of the Acoustical Society of America, 36, 2021-26.
- Saunders, F.A. (1946) Analyses of the Tones of a Few Wind Instruments. The Journal of the Acoustical Society of America, 18, 395-401.
- Seashore, C.E. (1902) A voice tonoscope. Iowa State Psychology, III, 1-17.
- Seashore, C.E. (1938) Psychology of Music. New York: McGraw Hill.
- Sebestyen, G.S. (1962) Decision-Making Processes in Pattern Recognition. New York: Macmillan.
- Shankweiler, D. (1966) Effects of temporal-lobe damage on perception of dichotically presented melodies. J. Comp. Physiol. Psychol. 62, 115-119.
- Shankweiler, D. & Studdert-Kennedy, M. (1967) Identification of consonants and vowels presented to left and right ears. Q. J. of exp. Psychol. 19, 59-63.
- Slawson, A.W. (1968) Vowel Quality and Musical Timbre as Functions of Spectrum Envelope and Fundamental Frequency. JASA, 43, 87-101.
- Sperry, R.W. (1964) The great cerebral commissure. Scientific American, 212, #1, 42-52 (January).
- Spohn, C. (1963) An exploration in the use of recorded teaching to develop aural comprehension in college music classes. J. of Research in Music Education, 11, (Fall) 91-98.
- Strong, William, and Clark, Melville (1967) A Synthesis of Wind-Instrument Tones. JASA, 41, 39-52.
- _____ (1967) Perturbations of synthetic orchestral wind-instrument tones. JASA, 41, 277-285.
- Stumpf, C. Die Sprachlaute, Berlin (1926).

- Sussman, H.M. (1972) What the tongue tells the brain.
Psychological Bulletin, 77, 262-272.
- Taylor, C.A. (1965) The Physics of Musical Sounds. New
York: Elsevier.
- Tenney, J.C. (1965) Gravesaner Blatter, 26, 106.
- Warren, R.M. & Obusek, C. (1971) Speech perception and
phonemic restorations. Perception and Psychophysics,
9, 358-363.
- Winer, B.J. (1962) Statistical Principles in Experimental
Design. New York: McGraw-Hill.
- Young, T. (1784) The Principal Phenomena of Sounds.

Appendix I
COMPUTER ANALYSIS OF MUSICAL PERFORMANCE
Warren C. Campbell

PURPOSE AND PROBLEM

This investigation is an attempt to add to the body of knowledge related to the following questions:

1. Are there operationally definably (i.e., objective, measurable) variables for which musical performance limits can be established?
2. If so, what are they and where are the limits of acceptability for various performance situations?
3. Can a knowledge of such variables be made useful in solving the problems of Music Education?

Given a set of scores assigned by a group of competent human judges to a set of student musical performances, are there objective features of the recorded performance which can be used, with suitable computer analysis, to predict the averaged judges' score?

PROCEDURES

Figure 1 provides an overview of the procedures. Two paths can be traced through the diagram, one for the human judging, the other for the computer simulation. Note that the simulation requires a representative set of the judges' scores in order to predict the judges' responses on other samples from the same population.

As shown in the computer branch of the diagram, aspects of the recorded performances were transferred to paper tape, using the equipment described in the previous chapter. The frequency meter produced a chart of performance frequency versus time, and the envelope follower presented the output voltage envelope (signal amplitude versus time).

Hand processing of the paper charts, described in this section, was used to extract the feature values for each note. Normalization and conversion of the chart values, and the subsequent reduction of the feature values to a small set of predictors were accomplished using special programs on an IBM 360 computer.

Multiple regression analysis, using these predictors

and based on a set of judges' scores produced a prediction equation which was then cross-validated on subsets of the sample. The cross validation, accomplished by correlating the predicted scores with the actual judges' scores for the subset is shown schematically at the bottom of Figure 1.

PERFORMANCE SELECTION AND PREPARATION

Fifty audio tapes, consisting of vocal and instrumental performances given at the 1968 Connecticut All-State Auditions, were auditioned to find a suitable data base for the present investigation. Of these, it was found that only the female vocal performances had a pitch range that could be processed satisfactorily using the available equipment.

A preliminary study was conducted to determine if a stable criterion could be established for this kind of short vocal performance, and to test the judging categories, instructions and format of the judging form. Reliabilities for eleven judges in this preliminary setting ranged from .70 to .89, and were encouraging enough to warrant the present investigation.

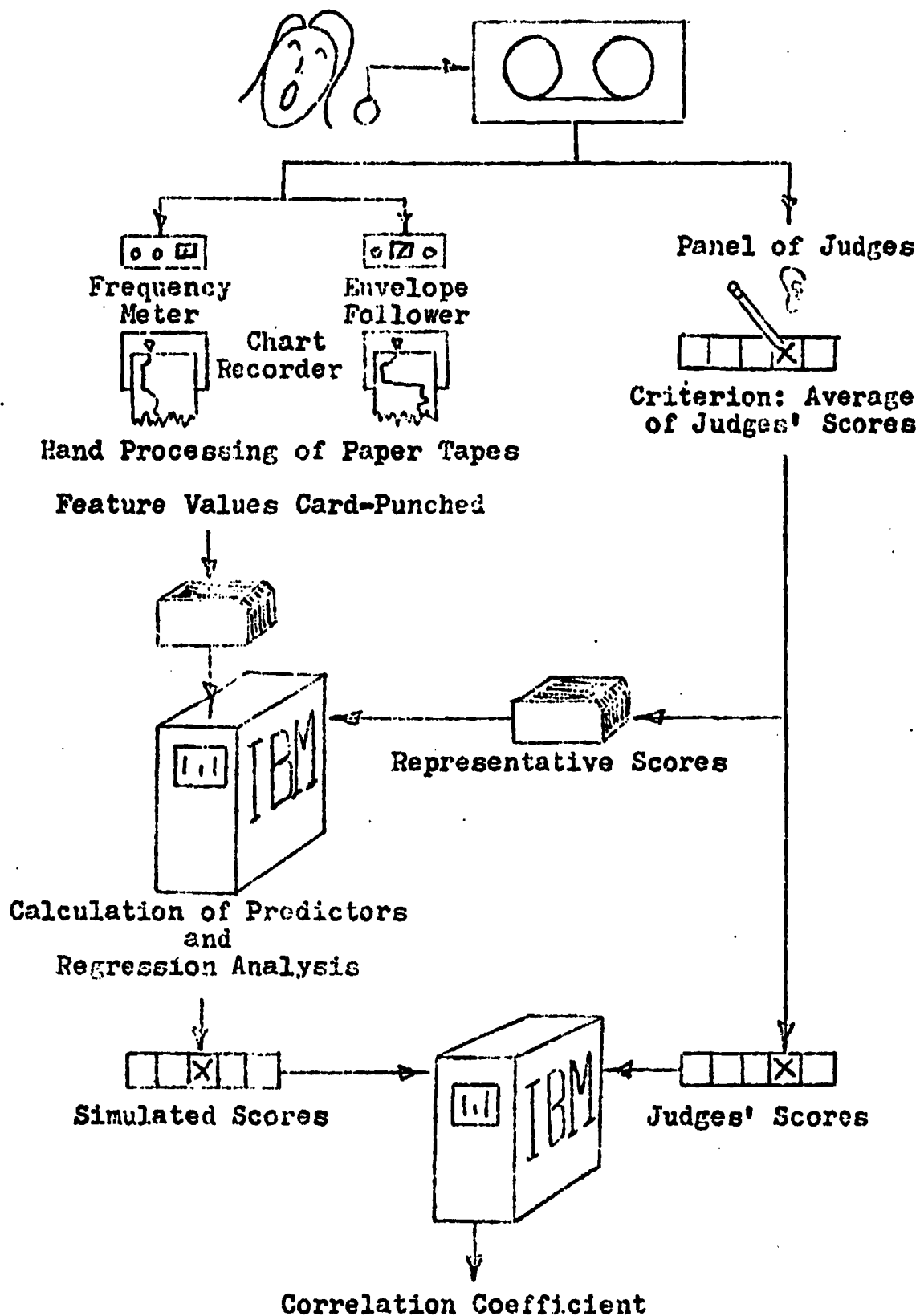
With the information available from this preliminary study, a set of sixty-two vocal performances by sopranos and altos were taken from the original audition tapes, and were used to prepare an analysis tape and a judging tape.

Performers auditioning were allowed to choose either the aria "If With All Your Hearts" from Mendelssohn's "Elijah", or "He Shall Feed His Flock" from Handel's "The Messiah". The portions of these selections used for performance judging are shown in Appendix A. Thirteen performers sang the Mendelssohn aria. These auditions were divided, so that each phrase starting with "If With All..." was treated as a separate performance. These twenty-six performances were presented on the judging tape so that the judges could not readily connect the two segments sung by one person. The remaining thirty-six performers sang the Handel aria, bringing the total number of performances to sixty-two. The duration of the Mendelssohn performances ranged from seven to twenty-two seconds. The duration of the Handel segment ranged from thirteen to forty-two seconds.

The judging tape was made by copying the original taped performance at its nominal recording speed, preceded

Figure 1.
Overview of Procedures

Vocal Performances on Audio Tape



by a performance number, and followed by a ten second silence. The running time of the judging tape was approximately 45 minutes. Seven copies of the original judging tape were made, so that each judge could audition the tape at his own rate and convenience.

The analysis tape was made by copying the original tapes on a variable speed recorder (Sony, Model TC-5600). By this means, the frequency range of the original performance was shifted to coincide with the range of the FM Subcarrier Discriminator.

CRITERION SCORES

Selection of Judges

The judges chosen for the panel were required to be experienced music teachers, some at the high school and some at the college level. Nine people fitting this description were contacted. Of the nine, seven were able to complete the judging task; of the other two, one was too busy, and the other found the task too difficult to do to his satisfaction. All the judges found the assignment to be an arduous one, because of the narrow range of abilities represented in the performance sample.

Directions to the Judges

The performance judges were asked for scores on a five point scale for intonation, vibrato, rhythm, dynamics, and an "overall" rating. The average of the first four categories was also calculated and used as an additional criterion value. The instructions given to the judges and the judging form are reproduced in Appendix A. The judging form contains five columns, one for each category. Each column was subdivided with headings "A" through "E", with "A" representing the best performances. The judge had only to check the letter category for each performance under each of five columns. Numerical results were card punched from these sheets with A=1, B=2, etc. Judges who completed the task were paid a fee of twenty dollars.

Each judge played the tape through at least twice. Several judges commented that the dynamics and rhythm categories were the most difficult to grade, and suggested that a diagnostic grading scheme might make their job easier.

Calculation of Criterion Scores

The grades assigned by the judges are essentially ordinal. When numerical values are substituted for the letter grades, and then averaged to provide a criterion score, an

interval scale is implied. The tacit assumption is made that the "distance" between the "A" and "B" categories is the same as the "distance" between all other adjacent categories, and that, for example, an average score of 1.5 is "halfway" between "A" and "B".

In addition, there is an assumption about the assignment of categories by the judge when the composite score used is the arithmetic mean: the scores must be considered as measurements which differ in a random fashion from a hypothetical "true" score, which represents the performance. It is only under this assumption that the average is a meaningful estimate of the performance score. If, for example, two schools of thought were represented in the panel of judges, with systematic differences between them, no single number would be a reasonable representation of the judges' opinion.

Stability of the Criterion Scores

The stability of the criterion scores can be estimated from the interjudge correlations, if the assumptions made in the previous section hold. The inter-judge correlations and a reliability estimate were calculated for the seven faculty judges. In addition, the correlation of each judges' scores with the average of the other six judges' scores was calculated as an indication of his agreement with the group opinion.

In order to test the reliability estimate, a class of nine graduate music students was asked to audition the same audio tape heard by the seven faculty judges. The average of the graduate students' scores was compared to the average of the faculty scores by calculating a correlation for each of the grading categories.

Three of the faculty judges consented to regrade the sample on a "one pass" basis, so that self-correlations could be compared with the inter-judge and inter-group correlations.

Considerable inter-category correlation was expected, based on the results obtained by Page and Paulus (1968). The stability of the less clearly defined categories was expected to benefit from this "halo" effect.

PREDICTOR VARIABLES

Selection of Features

The computer simulation of a criterion score is a problem in pattern recognition. The features of the pattern

upon which recognition is to be based are crucial to the success of a simulation. The selection, in this case, of the physical variables to be used as features depends on an assumption of a correlational relationship between the physical variables and the subjective responses of the judges.

The characteristic features chosen are listed in Table 1. These six values were calculated from measurements taken on each tone of the performances. It was expected that, while considerable intercorrelation would occur, the pitch features would be the primary intonation predictors, the power level would be the primary dynamics predictor, and that the duration features would be the primary rhythm feature. Because of the tedious hand processing necessary, no features were extracted specifically for vibrato.

Table 1

Characteristic Features, Calculated for Each Tone

1. Initial Pitch	(cents)
2. Middle Pitch	(cents)
3. Final Pitch	(cents)
4. Sound Power Level	(decibels)
5. Duration of Tone	(seconds)
6. Duration of Break or Glissando between Tones	(seconds)

Feature Extraction

Delays in equipment expected for machine processing made it necessary to employ hand processing techniques to extract the feature values from the recordings.

The paper tapes produced by the frequency meter and the rectifier were first segmented to define the onset and release of each tone sung. Tonal and non-tonal segments could then be measured along the time axis. The length of these segments represented the duration of a particular note sung by the performer, and the break or glissando leading to the next note. Because of noise and the decay characteristics of the system, no distinction was made between glissandi and breaks. Machine segmenting, while not a trivial problem, will undoubtedly provide more consistent results with much greater flexibility.

Figures 2 and 3 show excerpts from two performances as represented on the paper tapes, with vertical lines indicating the segmenting.

After segmenting, average pitch lines were drawn for each tonal segment by estimating the center of the vibrato envelope. Three ordinate values were tabulated from each tonal segment, at the beginning, middle, and end of the average pitch line. The maximum amplitude value for each tonal segment was also tabulated. These values are also indicated in the diagrams. Data relating directly to vibrato characteristics was not used. Because of the large variations in vibrato within single performances, this data was left for machine reduction at a later time.

The tabulation of these data resulted in an N by 6 data matrix for each of the sixty-two performances, where N is the number of tones scored for each performance. The initial frequency of the original taped performance was used to calculate the change in tape speed, and to establish the original frequency of all the performance tones, since they would be affected in the same proportion by the slowing or speeding of the audio tape. The tabulated data matrix for each performance was transferred to punched cards, which were then proof-read to assure their accuracy.

Data Normalization

Since the speed of the original recorded performance had been altered to fit the frequency window of the conversion equipment, the original frequencies and durations were recovered from the data matrix using calibration factors determined during the processing of the auditory signal. The original frequencies were then normalized to a cents scale, referenced to the starting pitch chosen by the singer. This made it possible to compare pitch deviations on a scale which is independent of starting frequency. A pitch change of one cent is defined as the change from f_0 to f when $f=f_0 \times 2^{1/1200}$. This frequency change represents a pitch change of 1/100 of a semi-tone of the tempered scale. Visual estimation of chart lines led to confidence limits of approximately ± 4 cents on each reading, a limitation which will be eliminated when machine readout is available. The sound level chart readings (output volts) were converted to the decibel scale, referenced to the minimum signal produced by the singer. Tonal and non-tonal durations were converted to proportions of the performance duration, so that differences in tempo would not affect the predictors, but only relative differences in tone durations. These normalizations did not, of course, provide any data reduction,

but only served to facilitate the next step, which was to compare each of the performances to a standard performance.

Data Reduction and Calculation of Predictors

The approach taken for data reduction was to establish a set of standard values for each tone of the musical selection. Deviations from the standard were then calculated for each performance, and the average deviations and mean square deviations from the standard were determined for each of the six features. These twelve numbers were the basic predictor set used in the multiple regression equation. A thirteenth predictor, the ratio of total tonal duration to total non-tonal duration, was added when it was found to have low correlation with the most effective of the other twelve predictors.

The data reduction program was run twice, using two different sets of standard values. The first set was from an operational standard: the best performance in each selection category (Mendelssohn 1, Mendelssohn 2, and Handel). The second set of standard values was obtained by using a literal transformation of the notation of the musical score: the three pitch values for each tone were identical, having the value, in cents, appropriate to the score indication for that tone; the duration values were proportional to the note durations in the printed score; non-tonal durations were zero, except where rests were indicated.

Since no score indications are given, sound power level was established by rule for the literal standard. Appropriate sound power levels appear to be, to a first approximation, pitch dependent. Therefore, decibel levels for the literal standard were made proportional to the score value for pitch referenced to the lowest tone in the sequence.

The primary differences between the operational and the literal standards can be listed as follows:

1. Pitch - The initial and final pitches in the operational standard vary considerably from the nominal pitch of the literal standard, especially on transition tones.

2. Duration - Tonal durations deviate from the nominal values, and short breaks appear between almost all tones of the operational standard. The duration ratio, since it is not referenced to a standard, is the same for both sets of predictors.

The programs used for data normalization and reduction are listed in Appendix B.

Figure 2.

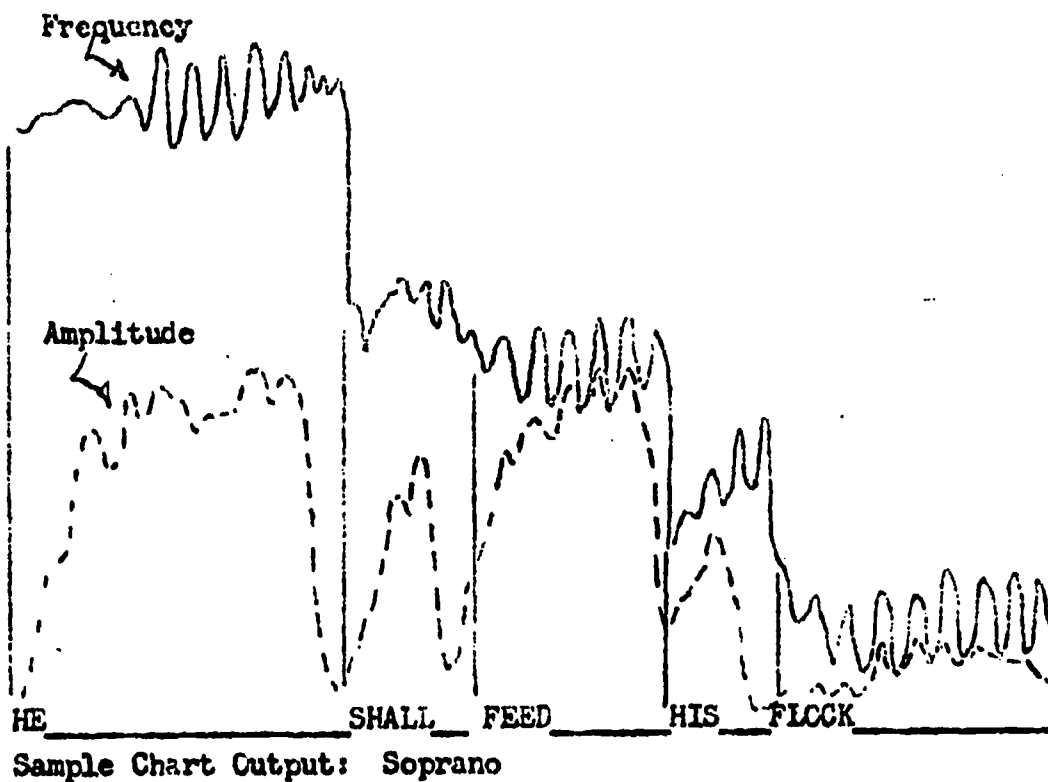
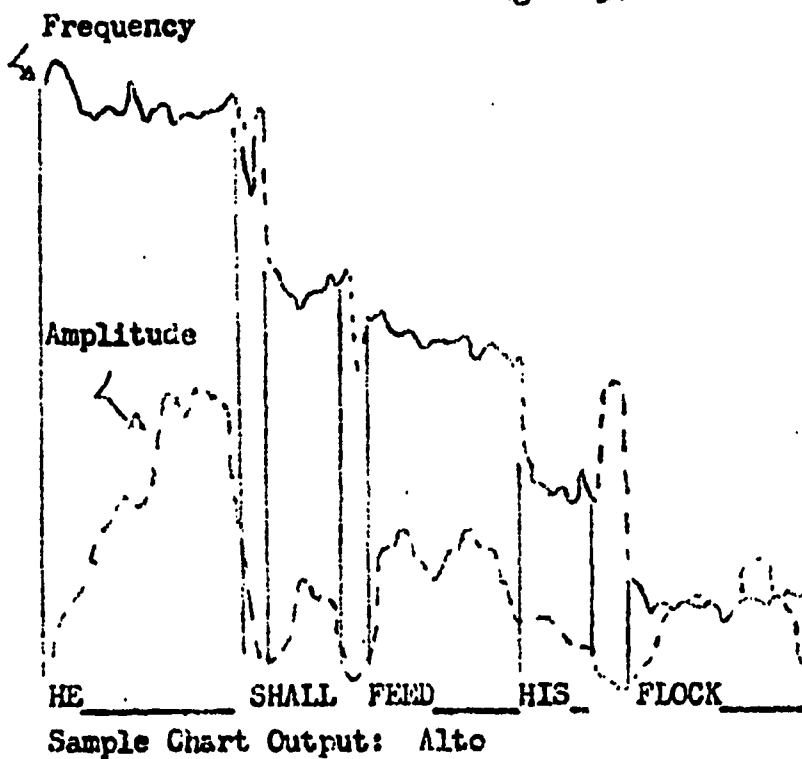


Figure 3.



ANALYSIS

Multiple Regression

A step-wise multiple regression analysis was run for both sets of predictors with the average of the seven faculty judges scores as the criterion. Each set of predictors was tested on the complete set of 62 performances, as well as on two partitions of the total set: a random split of 31 performances each, designated "31A" and "31B", and the natural subsets delineated by the Handel and Mendelssohn selections, designated respectively "36" and "26" in the tables, indicating the number of performances in each subset.

The use of all thirteen predictors in the linear regression equation for any subset increases the error variance accounted for, and therefore decreases the generalizability of the resulting predictor weights. In order to maintain a more nearly constant performance-predictor ratio, a selection of six predictors was made, and a new multiple regression equation using only these six predictors was calculated for each subset. The predictors were chosen by ranking, for each judging category, the predictors in the order in which they contributed to the reduction of the sum of squares. The six finally used were selected from those of the original thirteen which appeared most often in the upper half of the ranking for each of the six judging categories. The predictors used for the subsets are listed in Table 2.

There is no guarantee that these are the best combinations of predictors to select, since any one of them may be reducing primarily error variance. In that case, there would be little generalizability to other samples. However, if a set of predictors and b-weights, generated from one subset of the sample, satisfactorily predicts the scores from another subset, then these predictors may be expected to work on another sample from the same population.

Table 2

Predictors Used For Subset Calculations

<u>Operational Standard</u>	<u>Literal Standard</u>
1. Duration Ratio	1. Duration Ratio
2. Tonal Duration	2. Non-tonal Duration
3. Non-tonal Duration	3. Non-tonal Duration, Sq.
4. Non-tonal Duration, Sq.	4. Middle Pitch
5. Middle Pitch	5. Final Pitch
6. Final Pitch, Sq.	6. Sound Power, Sq.

- 50 ..

Cross Validation

Because of the small sample size (62 performances) cross-validation on any subsets of the original sample using all thirteen predictors would give a very poor indication of the generalizability of the regression coefficients. Therefore only the six most "generally useful" predictors from each predictor set, listed in Table 2, were used in the cross-validation. A step-wise multiple regression analysis was used to determine the b-weights for one subset of the performance sample. These linear coefficients were then used to calculate estimates of the judges' scores for the remaining performances in the sample.

SUMMARY

The procedures indicated in this section have been based on the program followed by Page and Paulus as presented in The Analysis of Essays by Computer (1968).

After selecting a suitable group of audio-taped vocal performances, a panel of judges was given the task of grading them in five different categories. The judges' scores were averaged to provide a criterion score. Acoustic features were extracted from the audio-tape, and this data was normalized, and reduced to thirteen predictor variables for each performance, using an IBM 360 digital computer. The predictors were used in a multiple regression analysis to simulate the criterion score. Cross-validation was accomplished by partitioning the sample, and predicting the scores for one subset using the regression coefficients from the other subset.

The major departure from the procedures followed by Page and Paulus is found in the feature extraction and data reduction procedures. The essay, composed of symbols compatible with machine input, does not require the extensive pre-processing necessary to translate aspects of the acoustic signal into a machine readable format.

RESULTS

ANALYSIS OF THE CRITERION VARIABLE

Inter-judge Correlations

A comparison was made between each judge and his peers for each grading category. The results of this comparison, the inter-judge correlations, are presented in Table 3.

If the average of the judges' scores is assumed to be the

Table 3
 Inter-judge Correlations
 Seven Faculty Judges
 N=62

Intonation

Judge	1	2	3	4	5	6	7
1	1.00	.48	.37	.54	.24	.57	.47
2		1.00	.50	.62	.35	.37	.55
3			1.00	.53	.31	.59	.49
4				1.00	.42	.48	.36
5					1.00	.42	.31
6						1.00	.42
7							1.00

Vibrato

Judge	1	2	3	4	5	6	7
1	1.00	.39	.03	.53	.40	.48	.39
2		1.00	.19	.50	.31	.47	.28
3			1.00	-.05	.04	-.13	-.10
4				1.00	.29	.64	.33
5					1.00	.46	.36
6						1.00	.46
7							1.00

Rhythm

Judge	1	2	3	4	5	6	7
1	1.00	.41	.49	.24	.21	.39	.28
2		1.00	.35	.19	.21	.22	.27
3			1.00	.29	.07	.37	.23
4				1.00	.48	.35	.16
5					1.00	.20	.22
6						1.00	.10
7							1.00

Table 3 (Continued)
 Inter-judge Correlations
 Seven Faculty Judges
 N=62

Dynamics

Judge	1	2	3	4	5	6	7
1	1.00	.38	.19	.17	.28	.45	.50
2		1.00	-.17	.24	.52	.18	.39
3			1.00	.00	.02	.19	.18
4				1.00	.14	.37	.18
5					1.00	.13	.32
6						1.00	.21
7							1.00

Overall

Judge	1	2	3	4	5	6	7
1	1.00	.57	.29	.68	.42	.58	.47
2		1.00	.42	.48	.43	.55	.42
3			1.00	.34	.20	.42	.31
4				1.00	.30	.60	.40
5					1.00	.34	.41
6						1.00	.55
7							1.00

Avg. of 1-4

Judge	1	2	3	4	5	6	7
1	1.00	.54	.36	.53	.36	.61	.49
2		1.00	.32	.56	.49	.48	.43
3			1.00	.33	.25	.47	.31
4				1.00	.46	.69	.36
5					1.00	.56	.40
6						1.00	.48
7							1.00

best estimate of the correct grade for the performance, then the correlation of each judge with the average of all the other judges can be used as an indication of judging accuracy. These correlations are shown in Table 4. By eliminating each judge from the average to which he is compared, the problem of spurious correlation (Benson, 1965) is avoided.

On the basis of the lowest correlation with the average for four out of the six categories, Judge #3 may be considered the maverick of the group. The validity of this designation was reinforced when the same judge regraded the sample as a member of the student judging group, and had the lowest correlation with the average of this group in five of the six categories.

A summary of the data presented in Table 3 can be made by calculating a typical inter-judge correlation for each category. This calculation is based on Rozeboom (1966, p. 320): the "homogeneity" of the judges is defined as

$$\text{Homogeneity} = \frac{\text{Average Proper Covariance between Judges}}{\text{Average Variance}}$$

A reliability value can be calculated using an equation which is analogous to the Spearman-Brown "prophesy" formula, but which is based on the homogeneity. This reliability estimate, designated "alpha", is given by Rozeboom (1966, p. 412) as: $\text{Alpha} = \frac{n (\text{Homogeneity})}{1 + (n-1) \text{Homogeneity}}$, where "n" is, in this case, the number of judges participating.

Table 5 shows the homogeneity and alpha values for each judging category for both the seven judge faculty group and the nine judge student group. Since alpha is a prediction of the inter-group correlation, based on the assumption of uncorrelated measurement error (deviations from the average), a correlation of the group averages will provide a test of the assumption. The inter-group correlations are given in the last column of Table 5.

The inter-group correlations shown are in the same range as the predicted reliability of the group average, justifying tentative acceptance of the assumption of uncorrelated deviations. The pattern across judging category is also very consistent: there is considerably more inter-judge agreement with regard to the "Intonation" and "Overall" categories than there is for the "Vibrato", "Rhythm" and "Dynamics" categories.

Intra-judge Correlations

A comparison of intra and inter-judge consistency is of

Table 4
Correlation of Each Judge
with Average of Other Judges' Scores

<u>Judging Category</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
1. Intonation	.59	.67	.63	.68	.44	.66	.59
2. Vibrato	.58	.56	.00	.57	.48	.63	.43
3. Rhythm	.55	.44	.47	.46	.36	.42	.34
4. Dynamics	.55	.53	.09	.28	.43	.40	.53
5. Overall	.70	.66	.43	.63	.47	.71	.58
6. Avg. of 1-4	.65	.65	.43	.65	.55	.77	.55

Table 5
Summary of Inter-judge Correlations,
Inter-group Correlations
(Seven Faculty, Nine Student Judges)

<u>Judging Category</u>	<u>Homogeneity</u>		<u>Reliability</u>		<u>Inter-group Correlations</u>
	<u>N=7</u>	<u>N=9</u>	<u>N=7</u>	<u>N=9</u>	
1. Intonation	.43	.40	.84	.86	.84
2. Vibrato	.28	.29	.73	.79	.86
3. Rhythm	.27	.30	.72	.80	.78
4. Dynamics	.24	.24	.69	.74	.73
5. Overall	.41	.39	.83	.85	.86
6. Avg. of 1-4	.41	.43	.83	.87	.88

interest to check the possibility of systematic differences between judges. If the judge's self-consistency is significantly greater than that between judges, some modification of the random error assumption may be necessary.

The three judges who regraded the performance set were Judges #1, #3 and #4. Particular interest was attached to the self-consistency of Judge #3, who had the lowest correlation with the group average as a member of both the faculty and student groups. The regrade correlations, by category, are given in Table 6.

The regrade correlations are higher than most of the inter-judge correlations, with most values falling within or slightly above the range for correlations of each judge with the average of the other judges' scores. In particular, the high consistency shown by Judge #3 indicates a stable grading procedure based on standards different than those used by the majority of the judges.

Inter-Category Correlations

There are two possible sources of inter-category dependency. First, training and experience may tend to improve all aspects of performance, so that the performer who sings with the proper intonation will be more likely to have an acceptable vibrato than one who sings with poor pitch control. Second, it may be tacitly assumed by a judge that the preceding is true, in which case his judgement of one category may bias his grade in another category (called a "halo" effect.)

The between-category correlations of the average scores are shown in Table 7 for the five categories scored by the judges. Correlations between the first four categories and the "Avg. of 1-4" category would contain spurious components, and are not shown. However, the correlation between the "Overall" and "Avg. of 1-4" categories was .96, indicating a very high predictability for the "Overall" grade on the basis of scores in the four other categories.

"Intonation" is the most distinct of the categories, having the lowest inter-category correlations. It is also unique in having as high a reliability as the "Overall" category.

Summary of Criterion Results

An analysis of the faculty and student judges' scores for the sixty-two performances has shown that the average grade for each performance is sufficiently reliable in each judging

Table 6

Regrade Correlations for Three Judges

<u>Judging Category</u>	<u>Judge #1</u>	<u>Judge #3</u>	<u>Judge #4</u>
1. Intonation	.67	.58	.58
2. Vibrato	.60	.66	.45
3. Rhythm	.50	.66	.26
4. Dynamics	.55	.57	.31
5. Overall	.59	.69	.43
6. Avg. of 1-4	.69	.74	.56

Table 7

Intercorrelations of Performance Judging Categories

<u>Judging Category:</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
1. Intonation	1.00	.66	.63	.68	.86
2. Vibrato		1.00	.77	.82	.84
3. Rhythm			1.00	.84	.84
4. Dynamics				1.00	.86
5. Overall					1.00

category to serve as a stable criterion for the multiple regression analysis. The most stable categories are "Intonation", "Overall" and "Avg. of 1-4" with predicted reliabilities for the faculty judges ranging from .83 to .84. The reliabilities of the "Vibrato", "Rhythm" and "Dynamics" categories are considerably lower, ranging from .69 to .73. Comparison of group averages for the faculty and student judges substantiates the reliability prediction, except for the "Vibrato" category. The two groups have a greater level of agreement on the "Vibrato" scores than would be predicted from the inter-judge correlations. The very low correlations of Judge #3 with the other judges in both groups is the primary source of this disagreement, so that the inter-group value for "Vibrato" (.86) is considered to be a valid measure.

ANALYSIS OF THE PREDICTOR VARIABLES

As indicated in "Procedures", the predictor variables were based on deviations in the performance features from a set of standard values. Two different standards were used: Predictor Set #1 was derived from the best performance ("Operational Standard"); Predictor Set #2 was based on a literal interpretation of the musical score ("Literal Standard"). The correlations between these predictors and the criterion scores are presented in Tables 8 and 9. The predictors are grouped according to category: duration, #1-5; pitch, #6-11; sound power, #12 and #13. All the predictors in Set #2 differ from those in Set #1, except for the duration ratio, since its definition is not relative to any standard.

MULTIPLE REGRESSION ANALYSIS

Simulation of the Judges Response

The results of the computer simulation of the human judgments are presented in Table 10, along with some associated statistics. In the first column, the reliability estimate for the average judges' score is given for each category. The multiple regression coefficients found using the thirteen predictors from Sets #1 and #2 on the sample of 62 performances are tabulated in column two. Except for the "Intonation" category for Predictor Set #2, all of these values are statistically significant beyond the 5% level, determined using the F-test. Two categories in Set #1, "Vibrato" and "Avg. of 1-4", are significant beyond the 1% level.

Column three, the "shrunken" multiple regression

Table 8
 Correlation of Predictors with Criterion Scores
 Predictor Set #1: Operational Standard

Predictors	1	2	3	4	5	6
1. Duration Ratio	-.03	-.26	-.28	-.24	-.11	-.22
2. Tonal Duration	.16	.28	.24	.23	.24	.25
3. Tonal Duration, Sq.	.06	.18	.16	.16	.15	.15
4. Non-tonal Duration	.03	.17	.09	.12	.07	.11
5. Non-tonal Duration, Sq.	.12	.28	.20	.23	.19	.23
6. Initial Pitch	.36	.29	.24	.36	.32	.35
7. Initial Pitch, Sq.	.23	.17	.12	.26	.20	.23
8. Middle Pitch	.38	.38	.31	.41	.41	.41
9. Middle Pitch, Sq.	.25	.25	.18	.30	.28	.27
10. Final Pitch	.39	.35	.30	.40	.41	.41
11. Final Pitch, Sq.	.26	.19	.15	.26	.25	.24
12. Sound Power	.18	.15	.29	.28	.25	.25
13. Sound Power, Sq.	.19	.16	.30	.30	.27	.26
1. Intonation						
2. Vibrato						
3. Rhythm						
4. Dynamics						
5. Overall						
6. Avg. of 1-4						

Table 9

Correlation of Predictors with Criterion Scores
 Predictor Set #2: Literal Standard

Predictors	Judging Category					
	1	2	3	4	5	6
1. Duration Ratio	-.03	-.26	-.28	-.24	-.11	-.22
2. Tonal Duration	-.05	.06	.10	.05	.04	.04
3. Tonal Duration, Sq.	.05	.15	.21	.16	.16	.16
4. Non-tonal Duration	-.22	-.11	-.05	-.10	-.20	-.14
5. Non-tonal Duration, Sq.	-.14	-.05	.07	-.02	-.10	-.05
6. Initial Pitch	.25	.06	.07	.07	.12	.13
7. Initial Pitch, Sq.	.20	-.01	.02	.02	.06	.07
8. Middle Pitch	.31	.24	.19	.24	.29	.23
9. Middle Pitch, Sq.	.21	.29	.17	.24	.26	.25
10. Final Pitch	.31	.15	.15	.16	.23	.22
11. Final Pitch, Sq.	.22	.10	.10	.11	.16	.15
12. Sound Power	.17	.00	.15	.17	.17	.13
13. Sound Power, Sq.	.21	.04	.20	.21	.22	.18

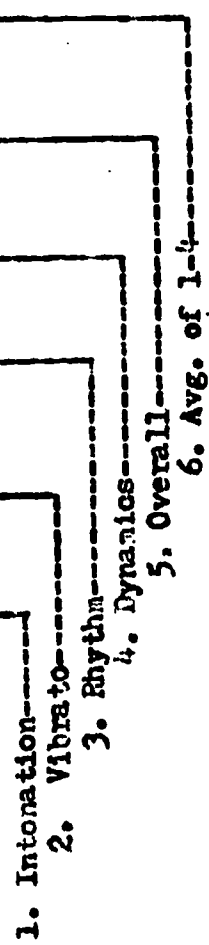


Table 10

Computer Simulation of Human Judgments
(62 Performances, 13 Predictors)

Predictor Set #1: Operational Standard

Judging Category	Alpha (k=7)	Mult-R	Shrunken Mult-R	Corrected for Attenuation
1. Intonation	.84	.62*	.47	.51
2. Vibrato	.73	.76**	.68	.79
3. Rhythm	.72	.67*	.55	.65
4. Dynamics	.69	.67*	.55	.66
5. Overall	.83	.67*	.55	.61
6. Avg. of 1-4	.83	.72**	.62	.67

Predictor Set #2: Literal Standard

Judging Category	Alpha (k=7)	Mult-R	Shrunken Mult-R	Corrected for Attenuation
1. Intonation	.84	.59	.41	.45
2. Vibrato	.73	.65*	.52	.60
3. Rhythm	.72	.68*	.57	.67
4. Dynamics	.69	.69*	.57	.69
5. Overall	.83	.62*	.47	.51
6. Avg. of 1-4	.83	.67*	.55	.61

Note: *indicates $p < .05$ (F-test)
**indicates $p < .01$

61

coefficient, is an estimate of the correlation between predicted and the averaged judges' scores which may be expected for a new sample of the same size from the same population, when the predicted scores are calculated using the predictor weightings (b-weights) found for the original sample. This validity estimate will always be lower than the value found for the first sample, since it is expected that some of the variance accounted for will not correlate in the new sample. For a fixed number of samples, the amount of variance accounted for by the multiple regression coefficient increases as the number of predictors increases. The equation used to calculate the expected shrinkage is called the Wherry Formula (Kelly, 1947, p. 474) and has the form:

$$R_s^2 = \frac{(N-1) R^2 - n}{N - n - 1},$$

where " R_s " is the shrunken multiple regression coefficient, " R " is the coefficient found for the sample, " N " is the number of subjects in the sample, and " n " is the number of predictors.

For comparison purposes, it is convenient to have a coefficient that has been normalized over varying criterion reliabilities, so that it represents the predictability relative to a perfectly reliable criterion. A multiple regression coefficient calculated from the shrunken coefficient and corrected for attenuation due to criterion unreliability is presented in column four. This value is calculated by dividing the shrunken multiple regression coefficient by the square root of the reliability of the criterion variable (Kelley, 1947, p. 412). From this value, an expected multiple regression coefficient can be calculated for a new sample, where criterion reliability is known, by multiplying by the square root of the new reliability.

In order to determine the uniformity of the sample under multiple regression analysis, the scores for each of the subsets used in the crossvalidation were calculated, using the b-weights figured for all 62 performances. These were then correlated with the criterion scores. The results are shown in Table 11. This was of particular interest because of the non-random subdivision of the performance selections (labeled "36" and "26"). A clear difference between predictor sets 1 and 2 can be seen here. Whereas for both sets of predictors "31B" is somewhat higher than "31A" for most categories, a reversal occurs for the "36" and "26" subsets when Set 2 predictors are used.

Cross Validation

Two different partitions have been used in the cross-

Table 11

Correlation of Simulated Scores with Criterion
for Performance Subsets

Predictor Set #1: 13 Predictors, b-wts. from 62 perf.

Judging Category	Performance Groupings (Number of Perf.)					
	<u>62</u>	<u>31A</u>	<u>31B</u>	<u>36</u>	<u>36</u>	<u>26</u>
1. Intonation	.62*	.52	.59	.51	.49	.62
2. Vibrato	.76**	.72	.79	.80	.62	.59
3. Rhythm	.67*	.67	.69	.71	.57	.57
4. Dynamics	.67*	.62	.73	.70	.68	.55
5. Overall	.67*	.53	.79	.68	.55	.59
6. Avg. of 1-4	.72**	.64	.78	.76		

Predictor Set #2: 13 Predictors, b-wts. from 62 perf.

Judging Category	Performance Groupings (Number of Perf.)					
	<u>62</u>	<u>31A</u>	<u>31B</u>	<u>36</u>	<u>36</u>	<u>26</u>
1. Intonation	.59	.65	.57	.50	.58	.57
2. Vibrato	.65*	.63	.70	.67	.57	.72
3. Rhythm	.68*	.72	.68	.61	.76	.66
4. Dynamics	.69*	.63	.76	.61	.50	.62
5. Overall	.62*	.57	.71	.50		
6. Avg. of 1-4	.67*	.66	.72			

Note: *indicates $p < .05$ (F-test)
**indicates $p < .01$



validation calculations. The first partition was randomly chosen to provide two subsets of 31 performances each. The second partition divided the sample into the 36 Handel selections and the 26 Mendelssohn selections. Tables 12 and 13 present the results of these partitions and the subsequent calculations. The two tables are identical in format.

In the first column of each three column block of Tables 12 and 13, the multiple regression coefficient for the subset is presented. This coefficient is calculated, using the six predictors listed in Table 2, by applying step-wise multiple regression to the subset indicated at the top of the column. The b-weights (regression coefficients) generated in this step are then applied to the remaining performances in the partition, and are used to predict the judges' scores on this subset, which was not involved in the multiple regression analysis. The results of correlating the predicted scores with the judges' averaged scores are presented in column three, for the subset indicated at the top of the column.

The significance levels for column one (multiple regression coefficients) were calculated using the F-test, with degrees of freedom appropriate for six predictors and $N=31$, 36 or 26 performances. Significance indications for column three were based on a one-tailed t-test, which places the levels of rejection for the null hypothesis lower than those used in column one, appropriate for a correlation between two independently generated sets of scores.

The second column, labeled "R", presents the shrunken coefficient calculated from the values given in the first column. This is an estimate, based on the Wherry formula, of the success of prediction in a new sample from the same population, and is used here as a guide in evaluating the cross-validation results. When column three equals or exceeds the value in column two, reasonable confidence can be placed in the generalizability of the coefficients of the regression equation. This occurs for Predictor Set One in all categories except "Intonation". There is not, however, complete reciprocity between the performance halves. The b-weights generated for "31B" do not generalize as well as those generated for "31A" when the predictors from Set 1 are used. The reverse is true for the Set 2 predictors. The "Intonation" predictions are uniformly intransigent, indicating that the information concerning the relevant intonation characteristics was lost in either the feature extraction or subsequent reduction.

The natural subsets, as shown in Table 13, provide useful levels of generalizability in only one case: the

Table 12

Cross-Validation
of Random Subsets

Predictor Set #1: Operational Standard, 6 Predictors

Judging Category	31A		b-wts Perf.	31B		31A
	31A	R _a		31B	R _a	
1. Intonation	.58	.41	.33*	.66	.54	.19
2. Vibrato	.75#	.68	.56**	.62	.48	.54**
3. Rhythm	.68	.58	.58**	.54	.33	.59**
4. Dynamics	.60	.45	.60**	.67	.56	.50**
5. Overall	.57	.40	.62**	.69	.59	.41*
6. Avg. of 1-4	.68	.57	.62**	.67	.56	.53**

Predictor Set #2: Literal Standard, 6 Predictors

Judging Category	31A		b-wts Perf.	31B		31A
	31A	R _a		31B	R _a	
1. Intonation	.66	.54	.12	.62	.48	.23
2. Vibrato	.66	.54	.43**	.71#	.66	.53**
3. Rhythm	.69	.59	.42**	.68	.57	.64**
4. Dynamics	.61	.47	.50**	.77#	.70	.54**
5. Overall	.57	.39	.43**	.74#	.67	.52**
6. Avg. of 1-4	.67	.56	.41*	.73#	.64	.62**

Note: #indicates $p < .05$ (F-test)
 *indicates $p < .05$ (t-test)
 **indicates $p < .01$ (t-test)

application of the b-weights generated from the Mendelssohn performances ("26") to the Handel performances ("36"). Again, intonation remains unpredictable in the new sample.

The lack of reciprocity in each of these cases is due solely to the accidental correlation, and the inability of the multiple regression analysis to distinguish between accidental and generalizable regularities in the data.

SUMMARY

The results reported in this section, based primarily on correlational analysis, have shown the level of inter-judge and inter-group agreement to be sufficiently stable to allow the averaged scores to be used as a reasonable criterion variable.

Thirteen predictor variables, based on the acoustic features extracted from the performance, separately correlate with the criterion scores with a maximum value of .41. Used in combination in an optimised multiple regression equation to simulate the judges' averaged response, the maximum correlation (occurring for the "Vibrato" category) is increased to .76.

Cross-validation checks indicate that a generalizable set of coefficients is possible for all categories but "Intonation" when these predictors are used. The best cross-validation results approximate the highest judges' correlations with the averages of other judges' scores.

Table 13
 Cross-Validation
 of Natural Subsets
 (Handel and Mendelssohn)

Predictor Set #1: Operational Standard, 6 Predictors

Judging Category	36			<u>b-wts</u> <u>Perf.</u>	26		
	<u>36</u>	<u>R_s</u>	<u>26</u>		<u>26</u>	<u>R_s</u>	<u>36</u>
1. Intonation	.53	.36	.15		.66	.51	.25
2. Vibrato	.75#	.69	.40*		.53	.24	.62**
3. Rhythm	.71#	.64	.15		.58	.35	.52**
4. Dynamics	.68#	.59	.27		.60	.40	.49**
5. Overall	.63	.53	.20		.62	.44	.47**
6. Avg. of 1-4	.72#	.65	.29		.64	.47	.53**

Predictor Set #2: Literal Standard, 6 Predictors

Judging Category	36			<u>b-wts</u> <u>Perf.</u>	26		
	<u>36</u>	<u>R_s</u>	<u>26</u>		<u>26</u>	<u>R_s</u>	<u>36</u>
1. Intonation	.50	.31	.27		.58	.36	.13
2. Vibrato	.70#	.62	.20		.64	.47	.15
3. Rhythm	.66#	.56	.23		.74#	.63	.41**
4. Dynamics	.62	.51	.46**		.77#	.68	.28*
5. Overall	.54	.38	.14		.69	.56	.19
6. Avg. of 1-4	.65	.55	.34*		.71	.58	.30*

Note: #indicates $p < .05$ (F-test)
 *indicates $p < .05$ (t-test)
 **indicates $p < .01$ (t-test)

DISCUSSION AND CONCLUSIONS

CRITERION SCORES

The performances chosen for this experiment were selected from a real adjudication situation. The performers were high-school girls who were considered by their teachers to be the most capable singers to send to the All-State competition. The range of the performance abilities represented in the sample was considered by the judges to be quite narrow. Under these circumstances, one might expect widely divergent responses from judging panels.

The reliabilities and inter-group correlations found (Table 5) show that, while reliabilities for single judges are low, a stable criterion can be established by averaging the scores of a number of judges. The highest seven judge reliabilities, in the range of 0.85, were found for the "Intonation" and "Overall" categories. Strong confirmation of the validity of these reliability values is provided by the inter-group correlations, which equal or exceed the predicted values in almost every case.

It should be noted that the judges did not discuss the meanings of the category, names or otherwise attempt to standardize their responses. Presumably, any such procedure would improve the inter-judge agreement. A wider range of performance abilities represented in the sample should also increase the homogeneity of the judges' responses, as well as providing for easier predictor selection and testing.

STANDARDS FOR EVALUATING COMPUTER PERFORMANCE

Three levels of agreement on performance scores are represented in the data for the criterion variable. At the lowest level are the inter-judge correlations. The middle values are found for the regrade correlations, and the correlations of each judge with the average of the other judges. The highest values are the reliabilities and inter-group correlations. Representative values for both high and low judging categories within each level are given in Table 14.

The three levels are presented here to provide bench marks in evaluating the simulation results. Previous studies have compared the computer results to the inter-judge correlations, which are at the low end of the agreement level scale. It seems quite reasonable, however, to expect simulation results to exceed the middle range values, and to begin to approach

Table 14

<u>Judging Category</u>	<u>Criterion Correlation Ranges</u>		
	<u>Low</u>	<u>Agreement Level Middle</u>	<u>High</u>
High*	.42	.63	.84
Low**	.26	.47	.71

* The high categories are Intonation, Overall, Avg. of 1-4.

** The low categories are Vibrato, Rhythm, Dynamics. In several cases, Vibrato moves into the high ranges.

Table 15

Homogeneity and Reliability for 16 Judges

<u>Judging Category</u>	<u>Homogeneity</u>	<u>Reliability</u>
1. Intonation	.41	.92
2. Vibrato	.30	.87
3. Rhythm	.28	.86
4. Dynamics	.24	.83
5. Overall	.40	.91
6. Avg. of 1-4	.42	.92

the values of the inter-group correlations. The computer, in effect, has information none of the human judges has: a representative set of average scores. If it is to be compared with the human judge, the simulation must not only correlate on the lowest level with each of the human judges, but must have a correlation with the judges' average scores which approximates the middle level.

For the purpose of assigning grades in a school situation, correlations of 0.80 or greater between the predicted and the judges' average scores would be desirable. Since prediction accuracy cannot be expected to exceed the criterion reliability, somewhat higher reliabilities are required. With a homogeneity value of 0.40, fourteen judges are required to exceed a reliability of 0.90. For example, with the combined scores of the student and faculty panels (16 judges), the homogeneity and the resulting reliability for the average score in each category is shown in Table 15.

PREDICTOR SIMULATION OF CRITERION SCORES

Multiple Regression Results

The linear regression approximation to the criterion has been shown to be moderately successful using the predictors based on the operational standard (Set 1). The shrunken multiple regression coefficients for these predictors are all within the middle range specified in Table 14. Two anomalies should be noted for these values, however. The value for "Intonation" is at the typical value for the lower group of categories, whereas, according to the reliability grouping, it should be at 0.63 or above. The vibrato value, at 0.68, exceeds its expectations as a member of the three categories with lower reliabilities. This is particularly notable since there are no predictors which are nominally "Vibrato" predictors, and there are three nominal "Intonation" predictors.

The literal standard (Set 2 Predictors) proved to be less effective, overall, than the operational standard. Losses in four categories far outweighed the very slight gain in the other two. This may be taken as an indication that whatever constitutes a musically acceptable performance, it is not the literal interpretation of the score, at least not for these vocal performances.

Cross Validation Results

Again with the exception of the "Intonation" category, the cross validation results are in the middle range specified in Table 14, for at least one set of predictor weights.

The results are not reciprocal: that is, the b-weights generated on subset "31A" predict well for subset "31B", but those generated on "31B" are in general less effective on "31A". The reverse is true for the Set 2 predictors.

The difference between the results of Set 1 and Set 2 predictors is less pronounced in the cross validation than it is in the simulation based on the full sample of 62 performances. This is probably due to the choice of predictors used in the reduction from 13 to six for the cross-validation.

The natural subsets (the Handel and Mendelssohn performances) show somewhat lowered cross validation results, and again reciprocity is lacking. However, except for "Intonation", the best results are in the mid-range indicated in Table 14.

GENERALIZABILITY

Because of the preliminary nature of the simulation attempted in this experiment, and the improvements in feature extraction expected in the near future, it does not appear to be useful to specify the particular, normalized regression weights found for this sample. In addition it is expected that the regression weights will be both instrument and context sensitive. Many more experiments will be needed to test the range of commonalities over various musical sources and forms.

The methodology, however, as it is used to extract features, compare them to a standard and then reduce them to predictors, is expected to generalize to most musical sources and forms. In particular, because this test was conducted with a random selection of performances from a real adjudication situation, and because of the high level of intergroup agreement on the criterion scores, it is believed that the levels of significance achieved here can be equalled and improved upon for other similar samples in many musical performance areas. However, considerably higher levels of prediction than those represented here will be required before practical applications can be pursued.

FUTURE RESEARCH

Many branches for investigation appeared during the course of this project which could not be included. Some of these will be pursued under a grant provided by the United States Department of Health, Education and Welfare. The following items were considered to be of particular importance:

1. In place of the multiple regression approach, the methods of the "adaptive set" using maximum likelihood approximations (Sebestyen, 1962) should be tried. This approach makes class assignments, rather than predicting a scale value, and can take into account disjoint sub-classes.

2. The performance standard used in this study was taken from within the sample set. An "ideal" performance standard should be tested. That is, a performance which is significantly better than any of those in the sample set in all respects.

3. An analysis of variance test for the uniqueness of the judging categories should be applied, and an attempt made to extract the "halo" effect.

4. Automatic feature extraction procedures must be developed, so that adequate samples may be analyzed in a reasonable amount of time. Because of the enormous amounts of information to be processed, hand processing at any point in the sequence is an intolerable bottleneck.

In addition to these modifications of the present experiment, a number of new, related experiments should be started:

A. Patterns and methods of structuring human judgments of musical performance need a great deal of investigation before models can be developed that have wide applicability. The patterns suggested by the student and faculty judgments in this study would have important implications if they were found to persist for other groups. Direct responses, such as "galvanic skin response" (skin resistivity), respiration rate, and pulse rate, should be recorded for test groups as well as the verbal response to sets of performances.

B. Synthesis, in the form of graded performances, based on the results of studies such as this one, should be used to determine what elements of the performance can be manipulated to produce predictable responses in a panel of judges. The goal of such synthesis should be to maintain as natural a setting as possible, while changes in only one element of the performance (such as pitch or loudness contour) are made.

SUMMARY

While this investigation has not established conclusively the practicality of simulating the pooled responses of human judges to short vocal performances, the results are

encouraging. The average judges' score can be accepted, on the bases of the results reported here, as a stable criterion for simulation. The simulation effort produces scores which correlate with the criterion (the average of the judges' scores) in the same range (0.47 to 0.63) as do the human judges. This is considerably higher than the inter-judge correlations (0.26 to 0.42) and appears particularly encouraging in view of the lack of precedent to provide guidance in selecting and calculating the acoustic features used for prediction.

The only exception to the general level of simulation and prediction was found in the "Intonation" category. Low correlations here are particularly puzzling because of the high criterion reliabilities for this category. In spite of this exception, the results found in this experiment show that operationally defined variables can be found which will predict, within limits, the subjective response of a group of musically adept listeners. This is a necessary (but not a sufficient) condition for the inverse process described in the "Statement of Purpose": the definition of acceptability limits for performance in terms of operationally defined variables.

Considerable work is still required before this approach can be considered a useful tool, either for direct application to adjudication or for the uncovering of basic correlational relationships between subjective and objective acoustic variables. The results of this experiment lend strength to the belief that this will be a productive approach, and that it deserves serious consideration in any research program directed toward understanding musical perception and performance in terms of measurable, rather than mystical, concepts.

REFERENCES

- Benson, M. "Spurious Correlation in Hydraulics and Hydrology." Journal of the Hydraulics Division. Proceedings of the American Society of Civil Engineering, July 1965, pp. 35-42.
- Kelley, T. L. Fundamental Statistics. Cambridge: Harvard University Press, 1947.
- Page, E. B., and Paulus, D. H. The Analysis of Essays by Computer. Final Report, HEW Project No. 6-1318. Storrs: University of Connecticut, 1968.
- Rozeboom, W. W. Foundations of the Theory of Prediction. Homewood, Ill.: Dorsey Press, 1966.
- Sebestyen, G. S. Decision-Making Processes in Pattern Recognition. New York: Macmillan, 1962.

APPENDIX A

INFORMATION GIVEN TO JUDGES

A. Instructions to Judges

Sixty-two student "performances" of fifteen to thirty seconds duration each, are presented on tape with a performance number preceding and a ten second pause following each selection. All the performers are sopranos or altos. Thirty-six sing the first few measures of "He shall feed His Flock: from the "Messiah". The remaining twenty-six performances consist of the first and second occurrences of the phrase beginning "If with all your hearts..." from "Elijah", sung by thirteen singers. The halves of these thirteen selections have each been given a performance number and are presented separately on the listening tape. While some voices may be distinctive enough to link the two performance halves, they are to be judged independently, as far as this is possible.

The judging form consists of a general evaluation of several categories, with ratings from "A", the best, to "E", the worst. Because of the number of judgements to be made, it may be necessary to listen to the tape more than once, or stops between the performances may be preferred, with an immediate replay when necessary. Any auditioning scheme which produces consistent results can be applied.

An attempt should be made to distribute the grades so that the categories are approximately equal, eg. 15 A's, 15 B's, etc. However, wide latitude can be taken with this distribution if the data demands it. More important to remember is that "A" is defined here to mean the best of five categories present in these performances, and "E", the worst in these performances.

The column headed "Overall" is not intended to be a summary of the other categories. It is recognized that aspects of performance other than those specifically mentioned may greatly modify the overall impression. This judgement should therefore be made independent of the ratings in the other categories, if at all possible.


In addition, please indicate how you would delineate the judging categories in order to produce the most useful form for grading or diagnostic purposes. Also include any critical comments on the mechanics of presentation, eg. selections too short, silent spaces too long, etc.

B. Sample from Judging Form


	Intonation A B C D E	Vibrato A B C D E	Rhythm A B C D E	Dynamics A B C D E	Overall A B C D E
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					

C. Excerpt from Handel's "Messiah"

Larghetto ($\text{♩} = 112$)



He - shall feed His flock like a shep - herd, and



He - shall gather the lambs with His arm, with - - His arm,

D. Excerpt from Mendelssohn's "Elijah"

Andante con moto ($\text{♩} = 72$)



If with all your hearts ye tru-ly seek me,



ye shall ever surely find me. Thus saith our



God. If with all your hearts ye tru-ly



seek me, ye shall ever sure-ly find me.



Thus saith our God, thus - - saith our God.

APPENDIX B

DATA NORMALIZATION AND REDUCTION

Six features were measured for each tone of each performance, and were stored in a data matrix for that performance. The values punched into data cards were the chart scale readings for pitch and amplitude, and chart lengths in inches for durations.

The following steps were needed to transform the data matrix, and to calculate the predictors from the normalized data. Table A-1 shows a sample data matrix with the column headings referred to in the description of the calculations.

1. Determination of Time Factor from Starting Pitch. The time factor (speed change) used to fit the performance range into the frequency "window" of the FM Discriminator (frequency meter) was determined by comparing the original starting frequency to the one indicated by the discriminator calibration:

$$\text{Adj. Start. Freq.} = C_1 (\text{Chart Value}) + C_2 ,$$

where C_1 and C_2 are calibration factors associated with the frequency meter. The time factor is simply the ratio of the actual and the adjusted starting frequencies:

$$\begin{aligned} \text{Time Factor} &= \text{Actual Start. Freq.} / \text{Adj. Start. Freq.} \\ \text{Effective Chart Speed} &= \text{Actual Chart Speed} * \text{Time Factor} \end{aligned}$$

2. Conversion of Duration Data to Seconds.

With the effective chart speed known, the duration data (columns 1 and 2) can be converted from chart distance to seconds:

$$\text{Duration} = \text{Chart Value} / \text{Effective Chart Speed}$$

3. Conversion of Frequency Data to Cents.

The frequencies (columns 3, 4 and 5) are calculated from chart values as in step one. Tape Speed change is accounted for by multiplying by "Time Factor". Conversion to cents normalizes the pitch values so that all performances of the same selection are comparable:

$$\text{Freq.} = (C_1 (\text{Chart Value}) + C_2) * \text{Time Factor}$$

$$\text{Pitch} = \text{Start. Pitch} + (1200 * \text{Log}_2 (F/F_0)).$$

F_0 is the frequency of the starting tone.

4. Conversion of Sound Power Data to Decibels.

The chart value representing output voltage (column 6)

is converted to a decibel scale referenced to the smallest peak value, V_0 . Since recording levels were not standardized, only changes within each performance can be compared:

$$\text{Sound Power (db)} = 20 * \text{Log}_{10} (V/V_0).$$

5. The first predictor: Duration Ratio.

The duration columns, #1 and #2 are summed to give the total time for each category. The ratio of these two values, called the "Duration Ratio" is the first predictor.

6. Conversion of Data Matrix to Deviation Scores.

The values in each cell of the data matrix for each performance were subtracted from the corresponding values of the performance standard. Two performance standards, the "Operational" and the "Literal" standard, were employed. In matrix notation, this can be expressed:

$$D - S - A,$$

Where "D" is the difference matrix, "S" is the matrix of standard scores, and "A" is the original matrix. The absolute values of the difference or deviation scores replace the original values in the matrix. By summing each column and dividing by the number of tones, a "mean deviation" score is found for each column. These values are the next six predictors.

If each matrix entry is squared and the columns again summed, and divided by the number of tones, a "mean squared deviation" from the standard score is produced. These six values bring the total predictors to thirteen.

Table A-1
Sample Data Matrix

Tone	Duration		Fitch			Sound Power
	Tonal	Non-tonal	Initial	Middle	Final	
1	*	*	*	*	*	*
2	*	*	*	*	*	*
.						
.						
.						
n	*	*	*	*	*	*

APPENDIX II

THE EFFECT OF THE ATTACK TRANSIENT ON AURAL RECOGNITION OF INSTRUMENTAL TIMBRES

Ralph C. Thayer, Jr.

Introduction

One characteristic of musical performance which must be dealt with by performer and teacher without a clearly defined set of standards, is tone quality. Tone quality can be discussed by teachers and students only in vague terms, such as "dark", "bright", "round", "thin", etc. These descriptions mean different things to different people, and very little, if anything, to the young student. The instructor can only suggest physical changes, such as position of the bow or adjustment of the embouchure, none of which are sure answers to a tone quality problem. The student does not establish a goal of good tone quality until many years of experience and exposure to what is assumed to be good tone quality have been accomplished. Even then the steps that must be taken to achieve this goal are not obvious. Indeed, the goal itself may be faulty, with no standard by which to judge it.

However, before standards of tone quality can be established, it must be determined what aspects of a musical tone are important in the aural perception of timbre. A musical tone consists of three basic parts: the attack, steady-state, and decay. Most research in the area of tone quality has dealt with the steady-state portion of the tone, assuming that this area is virtually unaffected by what precedes or follows it. "As has been shown by Helmholtz and others, the timbre of a given tone is determined by its harmonics, i.e., by the greater or lesser prominence of some of these harmonics over the others" (Apel, 1947). Recent investigation has indicated that this definition may no longer be an accurate description of timbre; that other factors, notably the attack of the given tone, may have some bearing on the perception of timbre.

Before any meaningful qualitative evaluation of timbre can be accomplished, it must be determined what constitutes timbre — what effect, if any, or to what degree do factors other than harmonic structure influence one's judgement of tone quality. If these factors do have an influence, it must be decided if they can be studied separately or must be considered in relation to each other in any study of timbre. Particularly since the advent of magnetic recording

tape it has been recognized that the attack transient seems to play an important role in the recognition of instrumental timbre. In synthesizing trumpet tones, the attack is found to be one of the subjectively important parameters (Risset and Mathews, 1969). If an attack is important to instrumental timbre recognition, is the quality of the attack also important? Should the attack be considered as merely a necessary embellishment or as an integral part of the characteristic qualities of a tone?

Problem

The purpose of this study is to determine what effect the attack transient has on the aural recognition of instrumental timbres. As we have no objective standards for good attacks or timbre, this study approaches the problem by mechanically replacing the attack of one instrument by the attack of another, to determine if, in this way, the listener is influenced in his attempt to recognize the timbre more by the attack or by the remainder of the tone. If the attack plays an important role, the listener will either be confused in identifying the instrumental timbre or will identify the timbre as that of the attacking instrument. If the nature of the attack significantly influences the listener at this level of discrimination, it would present the possibility that the attack would affect evaluation at a much finer level, such as determining what good tone quality is.

For the purposes of this study "attack" refers to the characteristic beginning of an instrumentally produced tone, and includes all of the tone up to the steady-state portion, or relatively periodic section. "Decay", or "release", refers to the ending of the tone, or that portion from the point that can no longer be defined as steady-state to the cessation of sound. For want of a better word, "steady-state" is meant to include all of the tone beyond the attack, including the decay. As the decay is always included in the tones presented in this study, this should not cause any confusion as to the meaning of the term. "Timbre" and "tone quality" are used synonymously to mean those qualities which differentiate the tone of one instrument from another.

Limitations

This study must be limited in several respects. Only three pitches are represented. A larger number would become extremely awkward to handle statistically, and would present a much longer test period to the subjects, causing fatigue.

The three pitches selected are fairly representative of the ranges of the instruments involved. For similar reasons the number of instruments is limited to four. The instruments can all be classified as soprano wind instruments. Therefore, any discussion of results must be assumed to apply only to this family of instruments. Similar studies using instruments of different families and different ranges might produce quite different results. Any generalizations to time and nature of attack must also be limited. The number of possible lengths of tones is obviously infinite, and no study could attempt to include them all specifically. The length of tones presented in this study is from one and a half to two seconds. Results obtained from tones of much shorter or much longer duration could be entirely different, and generalizations must be limited to tones of this approximate duration. The attacks utilized in this study may be classified as normal, that is, not accented and not legato.

Hypotheses

For statistical purposes, the null hypotheses are stated as follows:

1. There is no significant difference between instrumental timbre recognition scores of subjects presented with unaltered instrumental tones and instrumental timbre recognition scores of subjects presented with instrumental tones altered by substitution of the attack transients of other instruments.

2. There is no significant difference between instrumental timbre recognition scores of subjects presented with unaltered instrumental tones and instrumental timbre recognition scores of subjects presented with instrumental tones altered by elimination of the attack transient.

3. There is no significant difference between instrumental timbre recognition scores of subjects presented with instrumental tones altered by elimination of the attack transient and instrumental timbre recognition scores of subjects presented with instrumental tones altered by substitution of the attack transients of other instruments.

Procedures

A test, designed to measure the effect of attack on timbre recognition, was administered to three groups: group A--57 high school instrumentalists, group B--43 college

non-music majors, and group C--38 college level and above music majors.

The members of group A were asked to indicate their instrument on the answer sheet. This would allow for possible future analysis based upon instrumental background. The members of group B were not music majors, but were members of a music history class, possibly indicating some musical background. The members of group C were music majors, including several music faculty and professionally performing musicians.

Test

Four instruments were selected from the soprano wind family--the flute, the oboe, the clarinet, and the trumpet. From the common range of these instruments three pitches were selected--d', c'', gb''; that is, D below the bottom line of the treble staff, C on the third space of the treble staff, and Gb above the top line of the treble staff. These pitches were agreed upon by the performers as fairly representative of the common range of the four instruments. Each instrument was played by a professionally employed specialist on that instrument.

Each of the three pitches was recorded on tape by each of the four instruments. A Conn Strobotuner and a Hewlett-Packard 400E AC volt meter were placed so as to be easily visible to the performers. A standard of pitch and loudness was determined for each note. Each performer was able to check his performance against these standards by use of the tuner and volt-meter. A length of approximately one and one-half seconds was selected for each note. Each performer practiced each pitch with the measuring instruments several times until the standards were matched, including pitch, loudness, and time. Several recordings were made of each note in order that the most accurate could later be selected.

The recording was supervised by a qualified recording engineer. A Neumann model U67 condenser microphone, an Ampex PR-10 mixer, and an Ampex AG-440 tape recorder were the recording instruments. The tones were recorded full track, monophonic, at 15 inches per second, on 3M Scotch 202 recording tape (1.5 mils). A minus 2 db on the VU meter was used as constant peak level.

The resultant recording was analyzed and twelve tones were selected as most accurately meeting the standards, one tone of each of the three pitches performed by each of the four instruments. These twelve tones became the raw material from which the test tape was prepared.

Each tone was then prepared for the test tape in the following manner: unaltered, that is, exactly as recorded; with the normal attack replaced by the attack of each of the other three instruments on the same pitch; and with the attack portion deleted entirely. Each tone treated in this manner resulted in a total of sixty items. On the final test tape each item was repeated, making a total of 120 items.

The portion of the initial tone which could be considered "attack" had to be determined. The tones were played at slow speeds and the approximate length of the attack was determined aurally. All subjective decisions were made by a panel of professional musicians. Several experimental splices were made to ascertain the exact length of attack which would result in the most normal transition from attack to steady-state. Also, the effects of straight and diagonal splices were determined. These evaluations resulted in the use of diagonal splices and an attack length of one-twentieth of a second (three-quarters of an inch to the middle of the diagonal splice, at 15 inches per second). Using these criteria, the attacks of the tones to be altered were replaced with the attacks of the other instruments on the same pitch. The attacks of the tones to be presented with no attack were deleted with a square cut.

The 120 items were placed on the final tape in randomly selected order. They were placed in groups of five items each. A pause of four seconds was placed between each of the items in the first two groups, three seconds between the remaining items. Between each group of five, an eight second pause was placed. The items were grouped in this manner in order to correspond with the answer sheet, which was prepared with the items placed in groups of five. It was felt this would facilitate keeping one's place on the answer sheet without using audible numbers, which might tend to distract from the purposes of the test. Also, the longer spaces between the first ten items would give the subjects time to become familiar with the mechanics of the test before requiring more rapid responses.

"Print-through," or the phenomenon of the recorded sound appearing on the previous layer of tape, presented problems on the completed tape. In order to eliminate the print-through, each item was separated by an appropriate length of paper tape, which has no recording properties.

The subjects were presented, then, with a tape of twelve minutes, thirty-seven seconds duration, consisting of the 120 prepared items. They were asked to indicate which instrument each item suggested to them. The answer sheet was prepared with the four possible choices—flute,

oboe, clarinet, and trumpet. Each subject was asked to place a check in the appropriate box for each item. They were not told what alterations had been made to the tones, only that the tones were synthesized.

Scoring

In order to establish a basis for scoring, each response which did not identify the steady-state portion of the item was considered an error. Using this criterion, each paper was corrected and three scores were obtained: number of normal tones correct, number of altered tones-- or tones with attacks of other instruments--correct, and number of no-attack tones correct. As the test consisted of 24 normal tones, 24 no-attack tones, and 72 altered tones, the scores of correct normal tones and no-attack tones were multiplied by 3 in order to arrive at a possible score of 72 correct items for each category. These scores were used to calculate the means and to test the null hypotheses.

Each error was then charted to show what instrument the subject had indicated on the answer sheet. This chart provided detailed information as to the actual pitch, attack, and steady-state of the item incorrectly identified, as well as the response of the subject. This information was then compiled on a master chart for each group from which analyses were made.

In order to facilitate compiling the data a system of symbols has been devised. The normal tones are represented by two letters: the initial letter of the instrument (F-flute, O-oboe, C-clarinet, T-trumpet) plus the letter name of the pitch; hence TD indicates trumpet sounding d'. The initial letter of the attacking instrument is placed before this symbol to indicate the altered tones; hence OTD represents oboe attack and remainder of the tone trumpet sounding d'. The symbol for normal tones followed by (NA) indicates the tones with no attack; hence CG(NA) represents clarinet sounding g^b' with no attack. The flat symbol (b) is not used.

Results

The means and standard deviations (S.D.) for each group and each type of stimulus are presented in Table 1. Table 2 represents the conversion of these data to percentages.

The statistical comparisons for each group which tested

TABLE 1

MEANS AND STANDARD DEVIATIONS *

Group	n=	Normal		No Attack		Altered	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
A	57	53.32	9.59	43.74	9.50	38.32	6.93
B	43	52.81	12.04	45.35	11.46	38.93	9.62
C	38	59.63	10.26	50.61	9.37	45.13	8.86

* Highest possible score = 72

TABLE 2

MEANS CONVERTED TO PERCENTAGES

Group	Normal (%)	No Attack (%)	Altered (%)
A	74.01	60.75	53.22
B	73.35	62.99	54.07
C	82.82	70.29	61.85

Group A = High School Instrumentalists
 Group B = College Non-Music Majors
 Group C = College (and above) Music Majors

TABLE 3

RESULTS OF STATISTICAL TESTS OF SIGNIFICANCE

Group	Comparisons	Means	t (df=)	Probability of Null Hypothesis
A	(a) Normal vs No Attack	53.32-43.74	5.293	.001 (Null Hyp. 2)
	(b) No Attack vs Altered	43.74-38.32	3.452	.001 (Null Hyp. 3)
	(c) Normal vs Altered	53.32-38.32	• •	Rejected Null Hyp. 1 on basis of (a) and (b).
B	(a) Normal vs No Attack	52.81-45.35	2.903	.01 (Null Hyp. 2)
	(b) No Attack vs Altered	45.35-38.93	2.779	.01 (Null Hyp. 3)
	(c) Normal vs Altered	52.81-38.93	• •	Rejected Null Hyp. 1 on basis of (a) and (b).
C	(a) Normal vs No Attack	59.63-50.61	6.502	.001 (Null Hyp. 2)
	(b) No Attack vs Altered	50.61-45.13	2.585	.01 (Null Hyp. 3)
	(c) Normal vs Altered	59.63-45.13	• •	Rejected Null Hyp. 1 on basis of (a) and (b).

the three null hypotheses stated for the study, are presented in Table 3. The comparisons between normal versus altered tones (c) were not calculated since each of these mean differences were larger than the first two comparisons, and therefore were at least as significant as the (a) and (b) comparisons.

Further tests of significance determined no significant difference in any of the categories (Normal, No-Attack, Altered) between Groups A (high school) and B (college non-music majors). There were, however, significant differences found between Group C (college level and above music majors) and both Groups A and B in all categories (See Table 4).

Tables and graphs reporting detailed analyses of the test results appear at the end of this report. The results for Group C (Tables 5 and 6) have been extracted from these tables for discussion here, since this group (music majors) had the highest correct recognition for normal tones and therefore gives the clearest picture of the effects of recognition of instrument when the attack is altered or eliminated.

In Table 6, the capital letters in the left column refer to the attacking and steady-state instruments (CF = clarinet attack, flute steady-state; (NA) = no attack). The lower case letters refer to the pitches d', c'', and g^b'''. The numbers indicate the number of errors made by the group in each category, and what instrument was incorrectly identified.

The oboe tone was least often correctly identified among the normal tones. Table 6 shows that most of these errors, 55 (2+8+45), resulted from confusion with the clarinet, most of these occurring in the upper register, 45--g^b. A much smaller number, 12 (1+3+8), were incorrectly identified as trumpet; again, the number of errors increasing as the range increases. Although the oboe was most often confused with the clarinet, and next the trumpet, the reverse did not hold true. The number of incorrect clarinet identifications was quite small, 18 (4+6+8) incorrectly identified as oboe, 8 (5+3) as flute, and 2 (1+1) as trumpet. The trumpet tone was most often confused with the clarinet, 25 (5+20). Practically all of these occurred in the upper register, 20--g^b. A small number of errors in the lower register, 8, resulted from confusion with the oboe. Relatively few errors occurred in the identification of the flute tone. Those that did occur were confined to the low and middle register and were fairly evenly divided among the other three instruments.

TABLE 4
STATISTICAL TESTS BETWEEN GROUPS

GROUPS A AND B		GROUPS B AND C	
Category	t	Category	t
Normal0.226	Normal2.696
No Attack0.759	No Attack2.219
Altered0.404	Altered2.967
No significant difference in any category		Significance beyond the .05 level in all categories	

GROUPS A AND C

Category	t
Normal3.014
No Attack3.435
Altered4.078
Significance beyond the .01 level in all categories	

TABLE 5
CORRECT RESPONSES (IN PERCENTS) FOR GROUP C

Flute (Steady-state)	Oboe (Steady-state)	Clarinet (steady-state)	Trumpet (Steady-state)	Av.
% of Normal Tones Correctly Identified				
92.5	70.2	87.7	82.9	82.82
% of No-Attack Tones Correctly Identified				
89.0	66.6	82.0	46.1	70.29
% of Altered Tone Steady-States which Were Correctly Identified				
75.6	50.1	78.5	45.9	61.85
% of Altered Tone Steady-States which Were Identified as the Attacking Instrument				
11.3	26.0	12.0	22.4	17.90
% of Altered Tone Attacks which Were Identified as the Attacking Instrument				
6.3	15.8	27.8	21.8	17.90
% of Altered Tone Attacks which Were Identified as the Steady-State Instrument				
61.8	67.8	59.3	61.2	61.85

TABLE 6

RESPONSE ERRORS FOR GROUP C

FLUTE	Oboe	Clar	Trpt	OBOE	Flute	Clar	Trpt		
FF	d c g ^b	3 5 ..	4 2 ..	3	OO	d c g ^b 1	2 8 45	1 3 8
F (NA)	d c g ^b	8 3 1	1 3 7	1 1 ..	O(NA)	d c g ^b	.. 1 5	4 5 45	2 4 10
OF	d c g ^b	12 6 ..	10 12 7	1 1 ..	FO	d c g ^b 5	3 11 54	1 20 3
CF	d c g ^b	8 5 6	13 10 16	1	CO	d c g ^b	.. 1 2	13 10 59	1 11 8
TF	d c g ^b	7 4 2	4 13 9	6 11 3	TO	d c g ^b 3	3 7 36	13 49 28
CLARINET	Flute	Oboe	Trpt	TRUMPET	Flute	Oboe	Clar		
CC	d c g ^b 3	4 6 8	1 .. 1	TT	d c g ^b 2	8 2 2	.. 5 20
C (NA)	d c g ^b	1 10 6	5 5 12 2	T(NA)	d c g ^b 5	27 16 16	2 15 42
FC	d c g ^b	.. 4 10	5 12 4	1 1 ..	FT	d c g ^b	1 1 21	31 11 9	3 15 35
OC	d c g ^b	.. 3 6	6 8 15	2 .. 3	OT	d c g ^b 5	36 13 12	2 15 45

TABLE 6 (continued)

CLARINET		Flute	Oboe	Trpt	TRUMPET		Flute	Oboe	Clar
TC	d	1	8	1	d	..	26	5	
	c	5	7	6	c	..	10	17	
	g ^b	2	5	32	g ^b	5	16	42	

The tendencies that are displayed in the identification of the normal tones are the bases against which the results of the identifications of the no attack tones are compared: a strong tendency for oboe to be confused with clarinet, particularly in the upper register; a tendency for trumpet to be confused with clarinet, particularly in the upper register, less often with oboe in the low register; a slight tendency for clarinet to be confused with oboe over the entire range, increasing some as the range increases; and a tendency for flute to be fairly accurately identified.

The no-attack portion of the test resulted in a comparatively larger number of errors in trumpet recognition. While the scores of the other three instruments dropped from 3.7 to 5.7 percentage points, the trumpet identification scores dropped 36.8 percentage points, making the trumpet the least identifiable of the four instruments without attack. This occurred as a result of an amplification of the tendencies noted for trumpet in the normal tones: confusion with clarinet in the upper register (42), and, somewhat less, with oboe in the lower register (27). The scores of the other three instruments were primarily a result of a continuation of the trends already noted for normal tones.

The replacement of the normal attack with the attacks of the other instruments resulted in substantially lower scores in two categories, flute and oboe. Clarinet identifications were somewhat less accurate than the no-attack scores (3.5 percentage points), and trumpet very little less (0.2 percentage points).

Flute steady-states were incorrectly identified more often as clarinet, regardless of the attacking instrument-- 29 with oboe attack, 39 with clarinet attack, and 26 with trumpet attack. With oboe attack, 18 were identified as oboe, 2 as trumpet; with clarinet attack, 19 as oboe, 1 as trumpet; with trumpet attack, 13 as oboe, 20 as trumpet.

The effect of the clarinet and trumpet attacks on the incorrectly identified oboe steady-states was to cause the steady-states to be identified predominantly as the attacking instrument. Eighty-two oboe steady-states with clarinet attacks were identified as clarinet, mainly in the upper register (59), and 90 oboe steady-states with trumpet attack were identified as trumpet, mainly in the middle register (49). The flute attack resulted in 68 oboe steady-states being identified as clarinet, mainly in the upper register (54), 24 as trumpet, mainly in the middle register (20), and only 5 as flute, in the higher register.

The clarinet steady-state was most often incorrectly identified as trumpet (39) when preceded by the trumpet attack, predominantly in the upper register (32), and as oboe when preceded by the oboe (29) or flute attacks (21).

The trumpet steady-state was influenced by all three attacks in relatively equal numbers: 127 errors with flute attack, 128 with oboe, and 115 with clarinet. The flute attack resulted in 23 being identified as flute, 21 in the upper register; 51 as oboe, mainly in the lower register (31); and 53 as clarinet, mainly in the upper register (35). The oboe attack resulted in only 5 being identified as flute (upper register), 61 as oboe, and 62 as clarinet. As to register, the tendency remained the same--lower for oboe (36) and upper for clarinet (45). With clarinet attack, the trumpet steady-state was most often identified as clarinet (69), next as oboe (43), and least as flute (3). The range tendencies remained the same.

The percentage of attack transients which were identified correctly--that is, the subject chose the attacking instrument as the predominant timbre--was much smaller than the percentage of steady-states correctly identified. The clarinet attack was most often correctly identified (27.8%). The largest number occurred when it was attached to the oboe steady-state (82). When preceding the trumpet steady-state, 69 were correctly identified, and 39 when preceding the flute steady-state.

Of the trumpet attacks, 21.8 percent were correctly identified--90 when preceding oboe steady-state, 39 before clarinet, and 20 before flute. Oboe attacks were correctly identified 15.8 percent of the time: 61 with trumpet, 29 with clarinet, and 18 with flute steady-states. The least often correctly identified attack, flute, occurred only 6.3 percent of the time: 23 with trumpet, 14 with clarinet, and 5 with oboe steady-states.

Summary of Results

In general, the flute and clarinet steady-states were the most recognizable of timbres, under these conditions. The identification of the flute was slightly less accurate than that of the clarinet when preceded by other attacks, but was more accurate when preceded by its own or no attack. The attack portions of these two instruments are, however, in marked contrast to each other. The flute attack was least often correctly identified and the clarinet attack most often correctly identified of all of the instruments included in this study. This would indicate

that the flute provides very strong identification information in its steady-state, but very little in its attack. The clarinet provides strong identification information in its steady-state and in its attack. However, the combination of clarinet steady-state and attack does not result in any substantial reinforcement of this information.

The oboe was the least recognizable of the normal tones, followed, but at a relatively large interval, by the trumpet. With the removal of the attack portions, the trumpet dropped well below the oboe in degree of recognizability. With the addition of attacks of the other instruments, oboe identification scores dropped considerably, the trumpet scores only slightly, but the trumpet remained the least accurately identified timbre. The oboe provides, overall, the least identification information of any of the instruments included in this study. It resulted in its steady-state being most greatly influenced by the attacks of other instruments, and its attack having the least influence on the steady-states of other instruments.

The trumpet attack provided a great deal of identification information in combination with the trumpet steady-state. Without it, the steady-state was influenced by other attacks almost as much as the oboe steady-state. But, in combination with other steady-states the trumpet attack did not provide as much identification information as the clarinet attack. The trumpet attack, then, does reinforce the information provided by the trumpet steady-state in identification of this timbre. The steady-state by itself or in combination with other attacks is easily confused or identified as the attacking instrument, and the trumpet attack with other steady-states does not provide as much information as one might anticipate, in light of its effect on the trumpet steady-state.

It must be kept in mind that this discussion of results is made in relation to the overall effect reported in the results. This overall effect is that the identification of timbre becomes less accurate (statistically significant, $p < .01$) as the tone progresses from normal, to no-attack, to altered.

Conclusions

This study has attempted to determine what effect the attack transient has on the aural recognition of instrumental timbres. In general, the results show significantly less identification accuracy when the normal attack transient is removed. This difference is not equally displayed

by each of the instruments, however. The accurate identification of flute timbre was affected very little by the removal of the attack, in one case, Group B, actually increasing in accuracy. The identification of trumpet timbre was much less accurate when the attack was removed, with the degree of accuracy of identification of oboe and clarinet lying between these two extremes.

With the addition of foreign attacks the identification of the flute and oboe became considerably less accurate than in the no-attack state. The identification of clarinet and trumpet was only slightly less accurate.

Further analysis of the results show that the flute and clarinet steady-states were least apt to be identified as the attacking instrument; trumpet and oboe, the most apt to be identified as the attacking instrument. The clarinet attack, when combined with other steady-states, was most apt to be identified as the predominant timbre; the flute, least. When combined with the clarinet attack, other steady-states were least apt to be identified as the predominant timbre; with the oboe attack, most.

It is apparent that the identification of each of these instruments is affected differently by the alterations made in this study, and, from these differences certain characteristics can be attributed to each instrument.

The flute attack is relatively unimportant as an indicator of timbre. The flute timbre is quite readily identifiable either with or without the attack and the flute attack does not greatly influence the recognition of other steady-states. The flute steady state retains its identity quite well in combination with other attacks. The flute might be characterized as having a weak attack and a strong steady-state.

The oboe is the least accurately identified, even in its normal state, and the absence of attack affects this but little. The steady-state is noticeably influenced by other attacks, and the oboe attack influences other steady-states very little. The oboe might be characterized as consisting of both a weak attack and a weak steady-state, either in combination or separately.

The clarinet is only slightly less accurately identifiable without its attack than it is with it. The steady-state retains its identity the best of any of the instruments when combined with other attacks, and the clarinet attack substantially influences other steady-states. The

clarinet might be characterized as having, separately, a strong attack and a strong steady-state, but these strengths not necessarily reinforcing each other in the normal state.

The trumpet becomes the least accurately identifiable without its attack. The substitution of foreign attacks does not result in reducing this accuracy substantially, but the trumpet steady-state is relatively easily influenced by the attacking instrument. Although the trumpet attack plays an extremely important role in connection with the trumpet steady-state, this influence is not carried over to other steady-states as strongly as the influence of the clarinet attack. The trumpet might be characterized as having, separately, a weak steady-state and a not too strong attack. But, in combination with each other in the normal state, they reinforce one another substantially.

It is obvious that no one specific definition of timbre can be applied to all of these instruments. On the basis of this study, timbre may be defined only as those qualities that differentiate the tone of one instrument from another. Further studies in this area of all instruments, and more comprehensive studies of individual instruments, are indicated in order to arrive at a more accurate appraisal of the distinguishing characteristics of each instrument. When these characteristics have been determined and standardized, the basis for studies dealing with the standardization of tone quality (in this sense, the determination of good tone quality for each instrument) will be established.

REFERENCES

- Apel, Willi, Harvard Dictionary of Music. Cambridge: Harvard University Press, 1947.
- Resset, Jean-Claude, and Mathews, Max V. "Analysis of Musical Instrument Tones." Physics Today, 22 (February, 1969), 23-30.

TABLE 7: TABLES OF ERRORS

(The capital letters in the left column refer to the attacking and steady-state instruments, CF = clarinet attack, flute steady-state; NA = no attack. The lower case letters refer to the pitches d', c'', and g^b'. The numbers indicate the number of errors made by the group in each category, and what instrument was incorrectly identified.)

GROUP A

FLUTE	Oboe	Clar	Trpt	OBOE	Flute	Clar	Trpt
FF	d 30 c 7 g ^b 3	22 19 7	1 .. 1	OO	d 1 c 1 g ^b 5	3 10 54	3 2 5
F (NA)	d 31 c 9 g ^b 2	19 16 11	3 1 2	O (NA)	d 2 c 3 g ^b 9	4 24 59	7 3 12
OF	d 68 c 11 g ^b 5	22 26 8	.. 6 2	FO	d .. c 3 g ^b 12	2 32 72	7 5 3
CF	d 53 c 11 g ^b 11	28 30 22	3 3 6	CO	d .. c .. g ^b 9	14 34 66	3 7 21
TF	d 22 c 9 g ^b 2	10 21 11	46 39 2	TC	d 1 c 1 g ^b 9	8 19 59	10 33 12

CLARINET	Flute	Oboe	Trpt	TRUMPET	Flute	Oboe	Clar
CC	d 1 c 11 g ^b 16	12 6 29	.. 4	TT	d 2 c 2 g ^b 20	17 8 8	6 4 33
C (NA)	d 2 c 30 g ^b 19	14 12 28	.. 1 5	T (NA)	d .. c 3 g ^b 29	47 19 26	3 29 42

GROUP A--Continued

CLARINET	Flute	Oboe	Trpt	TRUMPET	Flute	Oboe	Clar		
FC	d c g ^b	2 24 29	15 4 7	1 .. 5	FT	d c g ^b	2 2 46	55 21 8	3 37 52
OC	d c g ^b	.. 7 22	15 12 24 6	OT	d c g ^b	1 2 18	61 8 29	10 42 52
TC	d c g ^b	4 15 14	11 9 10	1 12 43	CT	d c g ^b	.. 2 23	51 20 13	28 34 49
No response		FD(NA)-1	OG-1	CC-1	TG-1				
		FC(NA)-1	TOC-2	CG-1	OTD-1				
		TFD -1		FCD-1	OTG-1				
					CTC-2				

GROUP B

FLUTE	Oboe	Clar	Trpt	OBOE	Flute	Clar	Trpt		
FF	d c g ^b	11 3 ..	20 7 5	1 3 1	OO	d c g ^b	2 1 8	4 17 40	4 4 10
F (NA)	d c g ^b	15 6 1	12 7 8	.. 1 ..	O(NA)	d c g ^b	.. 3 14	10 15 51	4 5 6
OF	d c g ^b	27 10 1	27 16 9	3 3 1		d c g ^b	1 4 16	4 22 54	1 16 2
CF	d c g ^b	19 12 7	27 13 13	4 1 5	CO	d c g ^b	2 3 18	11 22 48	4 13 13
TF	d c g ^b	12 3 4	10 8 9	33 35 10	TO	d c g ^b	.. 1 12	9 15 44	17 46 20

GROUP B--Continued

CLARINET	Flute	Oboe	Trpt	TRUMPET	Flute	Oboe	Clar		
CC	d	..	15	..	d	..	7	1	
	c	9	10	1	TT	c	3	2	
	g ^b	10	16	8	g ^b	20	6	21	
C(NA)	d	..	22	2	T(NA)	d	..	27	6
	c	19	8	..	c	2	14	14	
	g ^b	12	12	3	g ^b	18	17	40	
FC	d	1	16	2	FT	d	..	34	6
	c	12	13	..	c	3	9	15	
	g ^b	19	6	1	g ^b	38	6	40	
OC	d	1	17	3	OT	d	..	36	6
	c	8	16	1	c	..	10	20	
	g ^b	15	11	..	g ^b	15	15	44	
TC	d	1	16	5	CT	d	..	27	14
	c	9	5	13	c	3	19	15	
	g ^b	12	4	38	g ^b	19	5	34	

No response OC(NA)-1 CG-1 TD(NA)-1
 COD -2 OCG-1 OTC -1
 TOD -1 CTD -1

Group C responses appear in the text of this report.

TABLE 8: TABLES OF PERCENTAGES

PERCENT OF NORMAL TONES
CORRECTLY IDENTIFIED

Group	Steady-State				Average
	Flute	Oboe	Clar	Trpt	
A	73.7	75.4	76.9	70.8	74.01
B	80.2	65.1	73.3	76.7	73.35
C	92.5	70.2	87.7	82.9	82.82
Average	82.1	70.2	79.3	76.8	76.73

PERCENT OF NO-ATTACK TONES
CORRECTLY IDENTIFIED

A	72.5	64.0	67.5	42.1	60.75
B	80.6	58.1	69.8	46.5	62.99
C	89.0	66.6	82.0	46.1	70.29
Average	80.7	62.9	73.1	44.9	64.68

PERCENT OF ALTERED TONE STEADY-STATES
CORRECTLY IDENTIFIED

A	52.6	55.0	71.5	34.8	53.22
B	58.4	46.0	68.3	44.1	54.07
C	75.6	50.1	78.5	45.9	61.85
Average	62.2	50.6	72.8	41.6	56.38

PERCENT OF ALTERED TONE STEADY-STATES
IDENTIFIED AS THE ATTACKING INSTRUMENT

Group	Steady-State				
	Flute	Oboe	Clar	Trpt	Average
A	25.1	18.7	15.8	25.2	21.22
B	21.8	23.9	17.1	21.3	21.03
C	11.3	26.0	12.0	22.4	17.90
Average	19.4	22.9	15.0	23.0	20.05

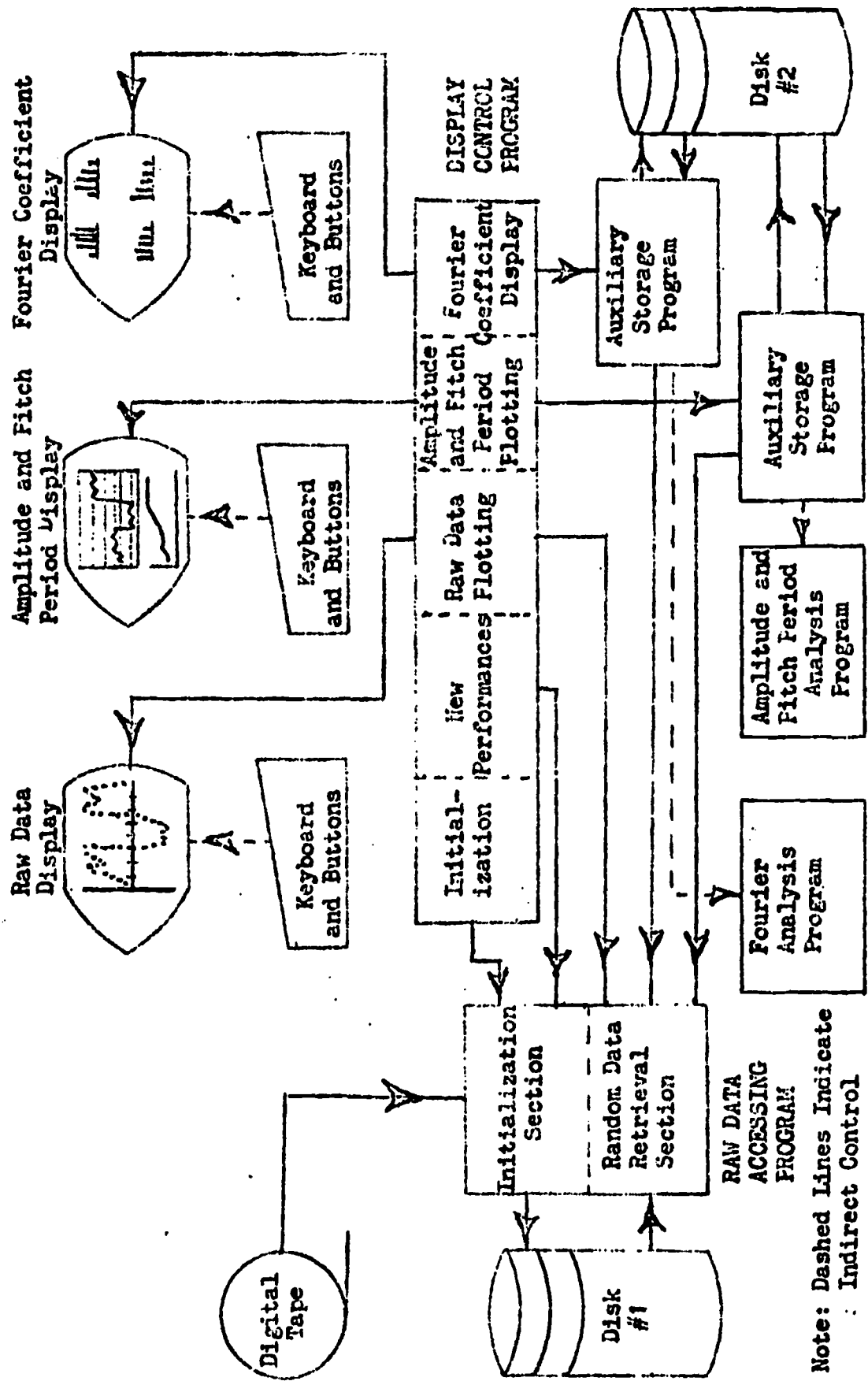
PERCENT OF ALTERED TONE ATTACKS
IDENTIFIED AS THE ATTACKING
INSTRUMENT

Group	Attack				
	Flute	Oboe	Clar	Trpt	Average
A	11.7	22.7	29.7	20.8	21.22
B	12.1	18.5	25.5	28.0	21.03
C	6.3	15.8	27.8	21.8	17.90
Average	10.0	19.0	27.7	23.5	20.05

PERCENT OF ALTERED TONE ATTACKS
IDENTIFIED AS THE STEADY-STATE
INSTRUMENT

A	56.2	55.5	47.3	55.8	53.22
B	55.9	59.3	52.1	49.5	54.07
C	61.8	67.8	59.3	61.2	61.85
Average	58.0	60.9	52.9	55.5	56.38

Figure 1. Program Configuration and Logic for use with the 2250 Display Console



Note: Dashed Lines Indicate Indirect Control

Appendix III: COMPUTER ANALYSIS SYSTEM SOFTWARE

Jack Owens

This section describes a library of computer programs used to perform certain basic mathematical analyses of the digitized musical performances. A package of programs performing optional supporting functions is also described.

The computational subroutines extract from the raw data quantitative information usually used to characterize acoustical signals. There are three subroutines, one to obtain the signal intensity, another the fundamental frequency, and the third the overtone spectrum. Associated with each is an optional I/O subroutine which may be used to store and retrieve randomly on a disk pack the output of the subroutines. This reduces the necessity to re-compute results when they must be used repetitively. There is likewise a program which translates the digitized information on the input tape into a computer format and then stores the results on a disk pack for subsequent random access. Finally a main-line program is provided which controls these subroutines and displays their output on a 2250 cathode ray display terminal as directed by the terminal operator. This enables the operator to specify features of interest in the data efficiently, so that they may be retained for subsequent processing. Visual access to the data in this way was deemed necessary because of the large volume involved. It would be unnecessarily expensive to process completely all of the data, or even to graph all of it on a computer print-out.

Figure 1 shows the relationship among these programs when they are used in conjunction with the 2250. It emphasizes the independence of the programs from each other. Each may be incorporated individually in another application. Thus the user may select to use only those programs performing functions needed for his particular problem.

Programming Considerations

The design philosophy of these programs is based upon several factors. In the first place, it is expected that they will be used by people not necessarily familiar with the details of the programming. This requires them to be easy to operate. Input arguments must be specified in a fashion relating in an obvious way to the application. Reliability is important. Faulty specification of arguments should lead to a default assumption or an error indication, rather than to ambiguous execution-time errors.

Secondly, the programs will probably receive heavy usage. This is because the functions they perform are basic to many types of analyses of musical performances. Thus they must be efficient and not use excessive central processor time. They must be able to operate independently, so that only programs performing those functions required for a particular application need be used. Provisions must be made so that their performance can be optimized easily in different applications.

Finally, allowances must be made for the possibility of future modifications in the logic of the programs. This applies especially to the computational subroutines and the display control program. This means that programming logic should be kept free of complexities and similar logic should be used to solve similar problems.

Implementation

In carrying out the programming choices must be made of the programming language, how the analysis is to be segmented, and how the subroutine arguments are to be specified. It was decided to use PL/I for the computational subroutines (called "procedures" in PL/I) and the main-line display control program. It is expected that the former will be used in main-line programs written in PL/I so they could not be written in FORTRAN. The use of assembly language would make future modifications of their logic extremely difficult, so PL/I was chosen. This requires the calling program, which in this case is the display control program, to be in PL/I. The loss of efficiency resulting from this choice was deemed insignificant. To minimize this loss, none of the more sophisticated, and consequently, less efficient, features of PL/I were used. It should therefore be easy to translate these programs into FORTRAN should it prove desirable.

In choosing a language for the I/O programs efficiency was the deciding factor. It turns out to be quite easy to access the data in assembly language, but very difficult in any higher-level language. This is because the organization of the data set does not conform to any of the standard IBM-supplied formats. A considerable savings in efficiency is effected by using assembly language. Fortunately the logic of the programs is probably not subject to alteration, so including parameters to modify table lengths and the like provides them with sufficient flexibility. These matters are discussed in more detail below.

It was decided to limit segmenting of the analysis as much as possible to the minimum amount dictated by the

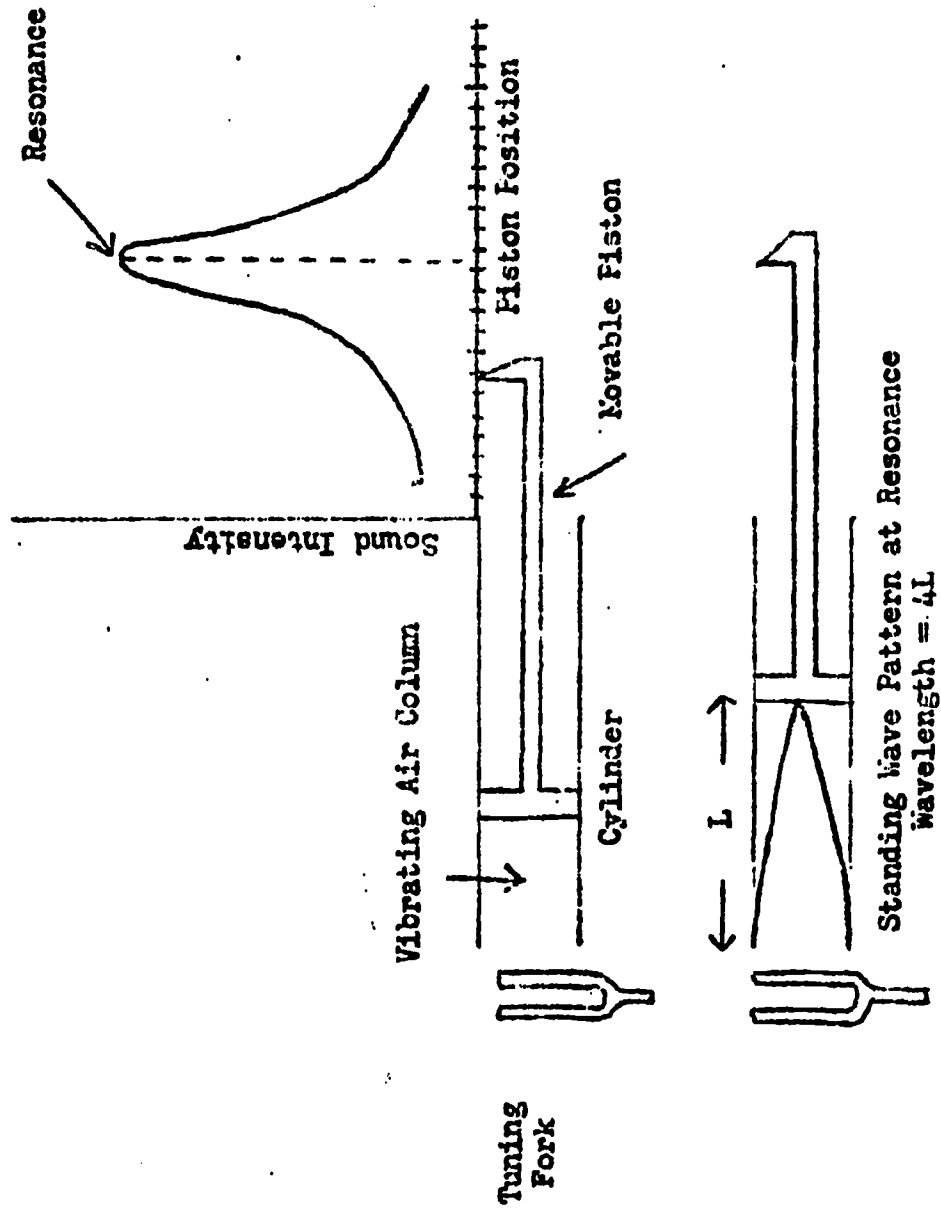
problem. This is because a large number of housekeeping chores are performed at the beginning and end of each PL/I subroutine making extensive use of subprograms quite costly. Thus the number of entries into a computational subroutine are limited by having each calculate a whole segment of data as specified by its arguments, rather than just one data point at a time. This produces a gain in efficiency and makes the programs easier to use. To some extent it also makes them more difficult to understand and therefore harder to modify.

The manner in which the subroutine arguments are specified was chosen to make them easy to operate from the standpoint of the user. For example, the starting and ending points of a segment of data are specified in terms of their times in seconds (and decimal fractions thereof) relative to the beginning of the performance in the actual time frame of the performance. Time intervals are likewise expressed in seconds. Other variables specify such things as the number of data elements per second and the number of data elements in a record of raw data.

As mentioned above, the computational subroutines determine the signal intensity, its fundamental frequency (or pitch period), and its overtone spectrum. The programs are controlled by specifying the points in time at which the information is required. The starting time, the time interval between points, and the number of points required suffice to determine this. By computing several points of data at one entry to the subroutine the total number of entries is minimized. The programs perform the analysis by operating on the raw data contained in an interval straddling the data point of length equal to the pitch period. Knowledge of the pitch period is of fundamental importance.

The technique for determining the pitch period is the mathematical analog of the following physical experiment. Suppose one holds a vibrating tuning fork near the open end of a hollow cylinder whose other end is closed off with a movable piston (see Figure 2). This causes the air column inside the cylinder to vibrate. Moving the piston changes the length of the air column. As this is done, a position will be found at which the tone produced by the tuning fork becomes considerably louder. The condition of maximum loudness is called "Resonance". If there is no node between the piston and the mouth of the tube then knowing the length of the air column gives the wavelength of the tone and hence its frequency. This may not be the fundamental frequency of the tuning fork because one of its overtones may be excited. This may be determined by searching for resonances of longer wavelength by lengthening the air column.

Figure 2. Obtaining the Wavelength from a Resonance



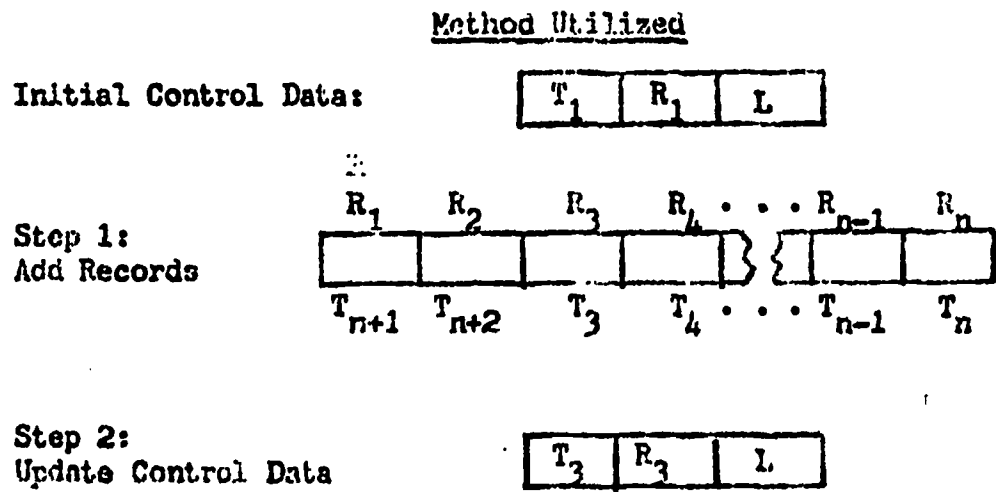
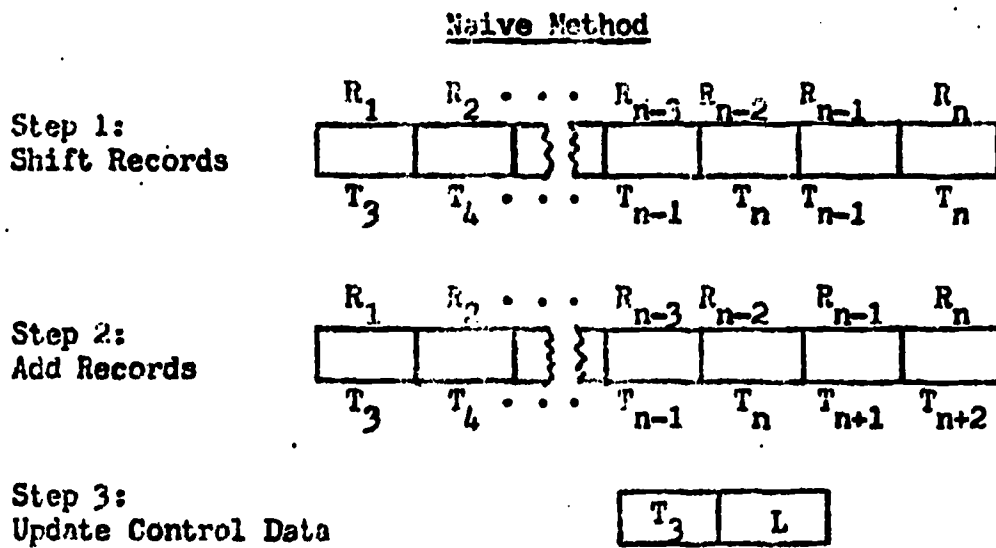
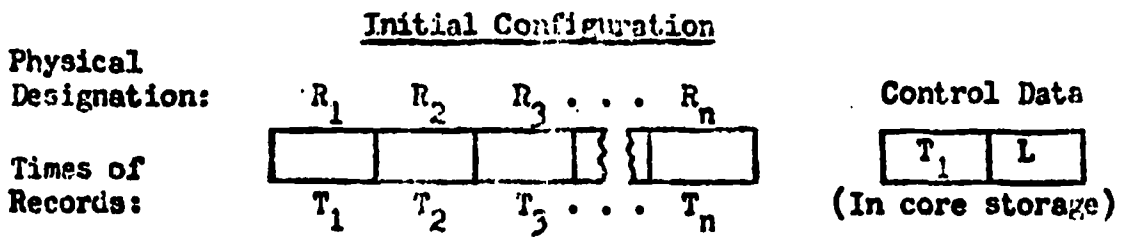
The mathematical analog of this used in this program performs a discrete Fourier analysis of the raw data over some interval about a point. The first Fourier coefficient is examined and the size of the interval is varied until it attains a maximum value. This corresponds in the above example to changing the length of the air column to obtain a resonance. The lowest resonance is sought by doubling the length of the interval and repeating the Fourier analysis. If the size of the first Fourier coefficient is sufficiently different from zero the process is repeated. Otherwise the length of the previous interval in seconds is taken to be the pitch period and its reciprocal the fundamental frequency. Since the data involved is primarily music, the search is first made using intervals of $\frac{1}{2}$ step on the tempered musical scale. Smaller intervals are then used to locate the maximum. The Cooley-Tukey fast Fourier transform algorithm is used for the Fourier analysis.

The intensity of a signal may be easily found once its fundamental frequency is known. It is equal to the square of the signal integrated over one period and divided by the length of the period. In this instance the trapezoidal rule was deemed accurate enough for doing the integration, so that the problem reduces to summing the squares of the data points over a period. The overtone spectrum is determined by carrying out the complete Fourier analysis over an interval of length equal to the pitch period.

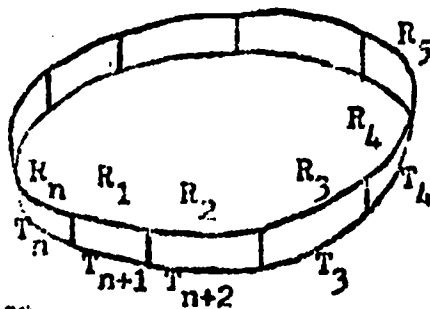
The I/O programs servicing the computational subroutines are similar to each other in function and design but quite distinct from the program which is used to access the raw data. The function of the former is to eliminate the necessity of re-computing results when a large volume must be processed repetitively. For example a person operating the 2250 may wish to scan back and forth through a large segment of data representing the pitch of the signal. Storing the calculations the first time they are needed and calling them back subsequently results in economical utilization of the central processor and core storage. The program handling the raw data allows an entire performance to be accessed randomly by converting all of the data into the computer format and storing it on the disk. This minimizes the number of cumbersome tape operations required and makes it necessary to translate the data only once.

Output data from the computational subroutines are stored on the disk in records of fixed length, representing a fixed interval in time. To simplify handling the data each record is defined to begin at a pre-determined point in time relative to the beginning of the performance, regardless of the time at which data is requested. These times are determined from the record length and the time interval between

Figure 3. Updating Disk Data Records



Conceptual Visualization of Disk Storage:



points, both of which are established by the user. From this information, given a request for a data element as a time expressed in seconds, the program can uniquely decompose the request into two integers, one giving the record number relative to the beginning of the performance and the other the location of the desired data element relative to the beginning of the record. If the time of the request does not correspond to one of the pre-assigned data points this procedure effectively truncates the time downward so that it does correspond to a valid data point.

With these conventions, accessing the data on the disk would be very easy if disk storage space could be reserved for all of the data that might be computed for a performance. The amount of space required for this is prohibitive, however, so space is allotted for only a fixed number of records, corresponding to a time interval within the performance of fixed length. This length is established by the user for optimum efficiency. The location in time of the interval relative to the beginning of the performance is allowed to vary as requests are made for data not represented by the current time interval.

Accessing the data under these circumstances is a little complicated. To understand the difficulties which arise, consider the example shown in Figure 3. Here it is assumed that space has been allotted for n records, stored contiguously in physical locations on the disk designated by $R_1, R_2, R_3, \dots, R_n$. The data stored in each record represent specific points in time relative to the beginning of the performance. The time of each record is given by specifying the time of the first data point in each record, say $T_1, T_2, T_3, \dots, T_n$ in chronological order with the earliest record of time T_1 stored in the first physical location R_1 and the latest record of time T_n stored in the last physical location R_n . If the record in storage with the earliest time is always to be stored in the first physical location, then there is no fixed relationship between the time of a given record and its location in storage. Control data such as the time of the earliest record and the length of the space allotted can be used to decide if a given record of a particular time is represented by the records which are currently stored.

Now suppose it is desired to add two new records to the data set corresponding to times T_{n+1} and T_{n+2} . Since all of the available space is taken up, the time interval represented by the data stored must be shifted upward by deleting the two earliest records, T_1 and T_2 . If the physical location R_1 is always to contain the record with the earliest time then all of the records must be shifted, so that

T_1 is replaced by T_3 , T_2 by T_4 , and so on. Then the two new records can be added to the end of the data set, so that the physical location R_n always contains the record with the latest time, which is now T_{n+2} . The control data is then updated to reflect the fact that the earliest record in disk storage now corresponds to time T_3 .

Although the logic involved in the above scheme is simple, it is clearly quite inefficient, as the entire data set has to be re-written every time the time-interval which the data represents is moved. The approach actually taken is to replace the records which are to be deleted by those which are added. In the foregoing example, T_1 and T_2 would be replaced by T_{n+1} and T_{n+2} respectively. The record corresponding to the earliest time stored is now not necessarily contained in the first physical record, R_1 , so that its physical location must be accounted for by an additional entry in the control data. The data set can be thought of as being organized in a ring, as shown in the bottom diagram of Figure 3. The record contained in R_1 may be considered as following that contained in R_n .

An important feature of this type of organization is that there is a correspondence between the time of a record and its physical location in the data set. This is obtained by computing the record number from its time, as discussed above, modulo the number of records allotted to the data set plus one; that is, modulo $n + 1$ in this example. When they are stored on the disk, records numbered 1 through n are stored sequentially in physical locations R_1 through R_n , as are records numbered $n + 1$ through $2n$, $2n + 1$ through $3n$, and so forth. This relationship may be calculated by expressing the quotient of the record number divided by $n + 1$ as an integer and a remainder. When the remainder is zero the record will be stored in R_1 ; if it is one, the record will be stored in R_2 , and so on. Because the data set organization is transparent to algorithms like this, accessing it is fairly efficient.

Information about each record is stored in a table which contains a series of entries for each record for which space is reserved. Rather than accessing the data directly, the program computes the location of the entries corresponding to the desired record. This tells the program if the data record has been computed and stored and if so where to find it. The actual device address of the record is stored so that it may be retrieved quickly. The table is kept updated to represent the current status of each record at all times.

If the requested data has not been computed the program

establishes the arguments required by the appropriate computational subroutine and returns control to the calling program with an indication of this. The latter can then call the computational routine and subsequently return control to the I/O program which writes the new data on the disk. This process continues until the original request for data is satisfied. Thus the user need not control directly the computational subroutines, but rather he can think in terms of the segments of data required.

The program handling the raw data performs three basic functions. It provides calibration information in the form of the number of data elements per second and the time interval in seconds between data points so that the exact time relative to the beginning of the performance of each data point is determined. It reads the entire performance off the digital tape, translates it into the computer format, and stores the results on the disk. Finally, it accesses the data off the disk at random as requested.

Calibration information is obtained from a calibration record at the beginning of the tape consisting of a wave such as a sine wave of known frequency. The program counts the number of pitch periods and data elements so that the calibration parameters can be computed from the frequency. A sufficiently long calibration record will provide a large number of complete pitch periods and data samples so that instabilities in the equipment can be averaged out. If the calibration record is recorded at the same speed as the performance all times will be in the actual time frame of the performance, so a stop watch can be used to determine the times of features of interest in the original tape.

The process of reading the tape, translating the data into computer format, and writing it out again on disk storage poses no problem. The translation is easily accomplished in assembly language. Two buffers are used and controlled by the program. The tape reading operation is the slowest of the three operations so translation and writing on the disk proceed concurrently with reading in of the next tape record. Each performance is preceded by a header record and the tape is terminated by a trailer record. The operation continues until one of these is encountered. Then it stops and control reverts to the calling program.

Providing for random access of the data from the disk posed some problems because the records produced by the analog - to-digital converter are not of uniform length. Thus it is not possible to compute directly the record in which a data element of a given time resides. This problem is circumvented by establishing a table of data with a series of entries for each record for which space is allotted on the disk.

When the program is initialized disk space is allotted for the maximum number of records of the maximum length to be encountered. These values are determined by the user. The device address of each record is stored in the table at this time. When the performance is read in, the number of data elements in each record and the starting time of each record is recorded in the table as the records are processed. The length of each record is determined from the maximum record length and data and is used by the control program for reading the tape, which is accessible to the assembly language program. The calibration factor, number of data elements in the previous record, and size of the tape record gap are used to calculate the starting time of each record. When all of the records of the current performance are read in the (variable) total number of records is recorded and an indication is placed after the entry in the table corresponding to the last record read.

When a request for a data element of a given time is made the record containing that data element is located by searching this table. The search is not sequential, but proceeds in a more efficient pattern. This is done to reduce the search time, although it leads to more complicated logic.

The display control program is a main-line PL/I program which enables the user to obtain a graphic display of the raw data and the outputs of the computational sub-routines on the 2250 cathode ray display console.

The console is equipped with a typewriter-like keyboard and a set of 32 push buttons. Using the keyboard the user can enter data into the buffer of the 2250 from which it is displayed and made available to the program. When a push button is pressed it can be sensed by the program enabling it to respond in a pre-determined fashion.

The display control program uses the keyboard to enable the operator to specify the type of display he wants, the time of the frame he wishes to be displayed, and the scale option he desires. There are three types of displays: one for the raw data, one showing the pitch period and amplitude, and one showing the Fourier coefficients. The scale options determine the time spanned by a single frame. There are two options for plotting the raw data and three for the other two displays.

The push buttons are used to manipulate the display and perform certain control functions. The display may be translated forward or backward in time by a full frame or one-

third of a frame. Its scale may be expanded or compressed. Control functions performed enable the operator to revert to functions performed on the keyboard such as requesting a different type of display or a new time. There is a button which causes the program to begin processing the next performance in sequence and another which terminates processing altogether.

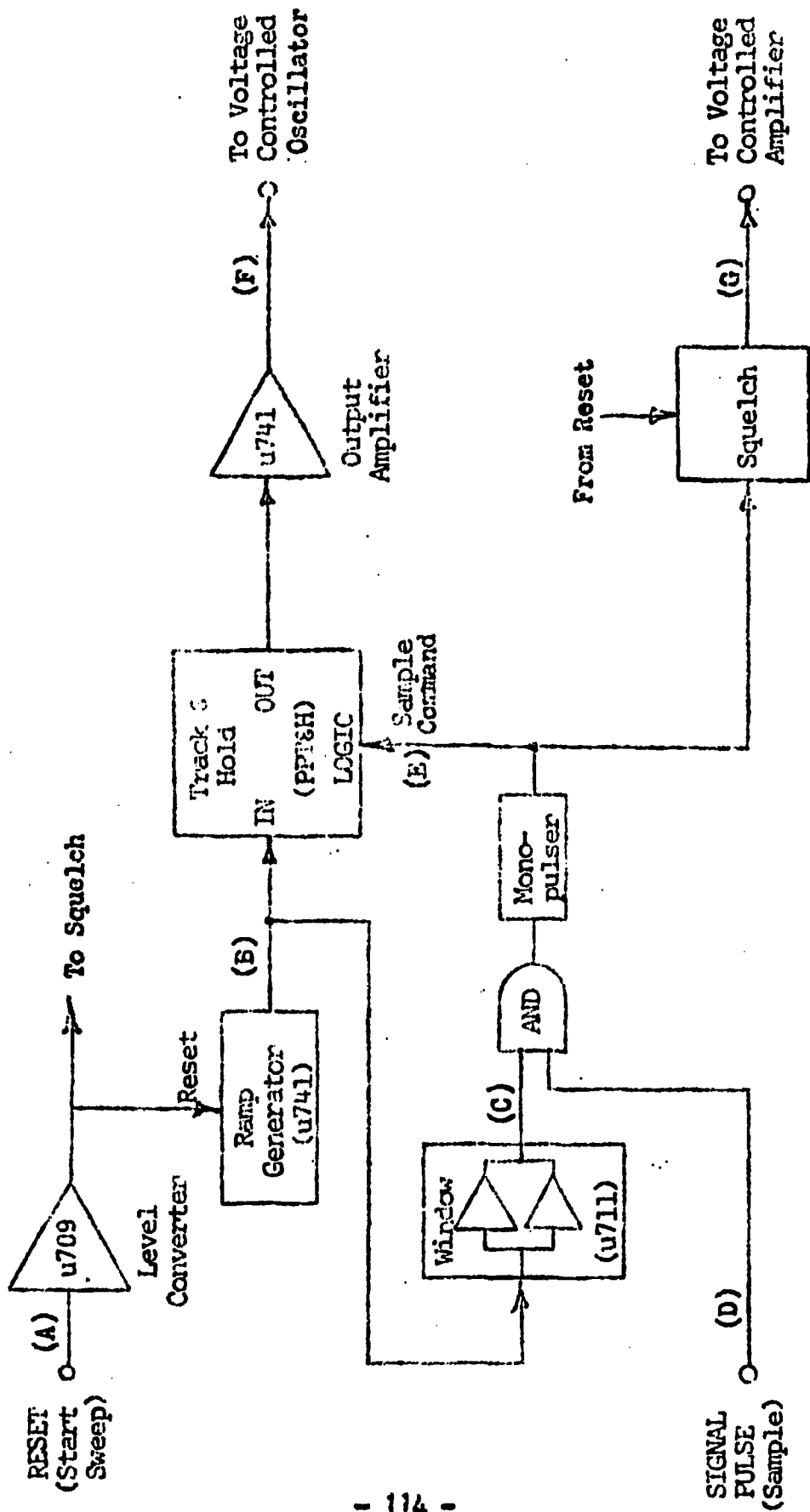
The program logic is diagrammed in Figure 1, where it is divided into several blocks according to function. When generation of the display begins the program is placed in the wait status until a button is pushed or the end key on the keyboard is depressed, signaling the end of data entry. The program senses exactly what has taken place and transfers control to the appropriate point in the program to handle the condition.

The display is generated and controlled using the Graphic Subroutine Package supplied by IBM. These generate the actual graphic orders and data. It is through them that the program enters the wait status and senses the condition that caused it to leave this status. They also provide the program with the data entered from the keyboard.

It is possible not to use these subroutines; subroutines could have been written in assembly language using the graphics access method especially for this application. It was decided not to do this because of the difficulty of such a task and the problem of making future modifications of the program. Using the graphic subroutine package makes it much easier to make modifications, although there is some loss in efficiency.

Figure 1.

SCANNER / SYNTHESIZER INTERFACE



Appendix IV: SCANNER/SYNTHESIZER INTERFACE

An electronic solid-state interface for the EMR optical scanner and a voltage controlled synthesizer. Original design - Ernest Guignon; design modifications - Philip Bogner.

The scanner/synthesizer interface was designed to convert the timing and signal pulses from the scanner to a suitable input for a Moog (or equivalent) voltage controlled oscillator (VCO) and/or a voltage controlled amplifier (VCA). The interface is represented schematically in Figure 1 and operationally in Figure 2. A second interface channel is available which is identical in function, but which utilizes the ramp generator output from channel one.

The two inputs provided by the optical scanner to the interface are identified as: A. Reset and D. Signal Pulse. At the beginning of each optical scan, a reset pulse (A) is generated. After level shifting it is applied to the ramp generator, and resets the ramp output to zero volts. The ramp generator employs an operational amplifier in an integrator configuration to produce a linear, sawtooth shaped output function, designated a "ramp" (Figure 2, line B). This waveform, which provides a voltage proportional to the position of the optical scan, is applied to the input of the track and hold module and to the window generator.

The window generator limits the active scan region for the channel. These limits can be selected by the operator, and correspond to boundaries on the paper fed to the scanner. This allows the paper to be divided lengthwise into two regions, each controlling a separate information channel. Whenever the ramp generator output voltage is within the preset limits, the output of the window generator (a dual comparator) is at a logical "1" voltage level. During the remainder of the sweep, the window is "closed" at a logical "0" level (line C). During normal operation the channel 2 window will be a logical complement to channel 1 (channel 1 is open when channel 2 is closed, and the converse).

A signal pulse from the optical scanner (line D) indicates that a mark appears on the paper at the concurrent scan position. The signal pulse is presented as input to a logical "and-module," along with the output of the window generator. The mono-pulser will be triggered only when the signal pulse occurs within the window limits set for the channel. In Figure 2, signal pulses 1 and 2 (line D) occur within the window, so that sample command pulses are generated (line E). Signal pulse 3 occurs outside the window, so that no sample command is generated.

A sample command pulse from the mono-pulser causes the track and hold module to sample the concurrent value from the ramp generator, and to transfer it to the output amplifier (line F). The output voltage is a linear function of the position of the line sensed by the optical scanner. The output amplifier provides adjustable output gain and zero offset voltage, and protects the track and hold module from a short circuit condition at the output.

The squelch circuit produces an "off" condition when no sample command has appeared for two or more consecutive optical scans (line G), as indicated by the reset pulse. The two level output of the squelch provides an on-off input to a voltage controlled amplifier.

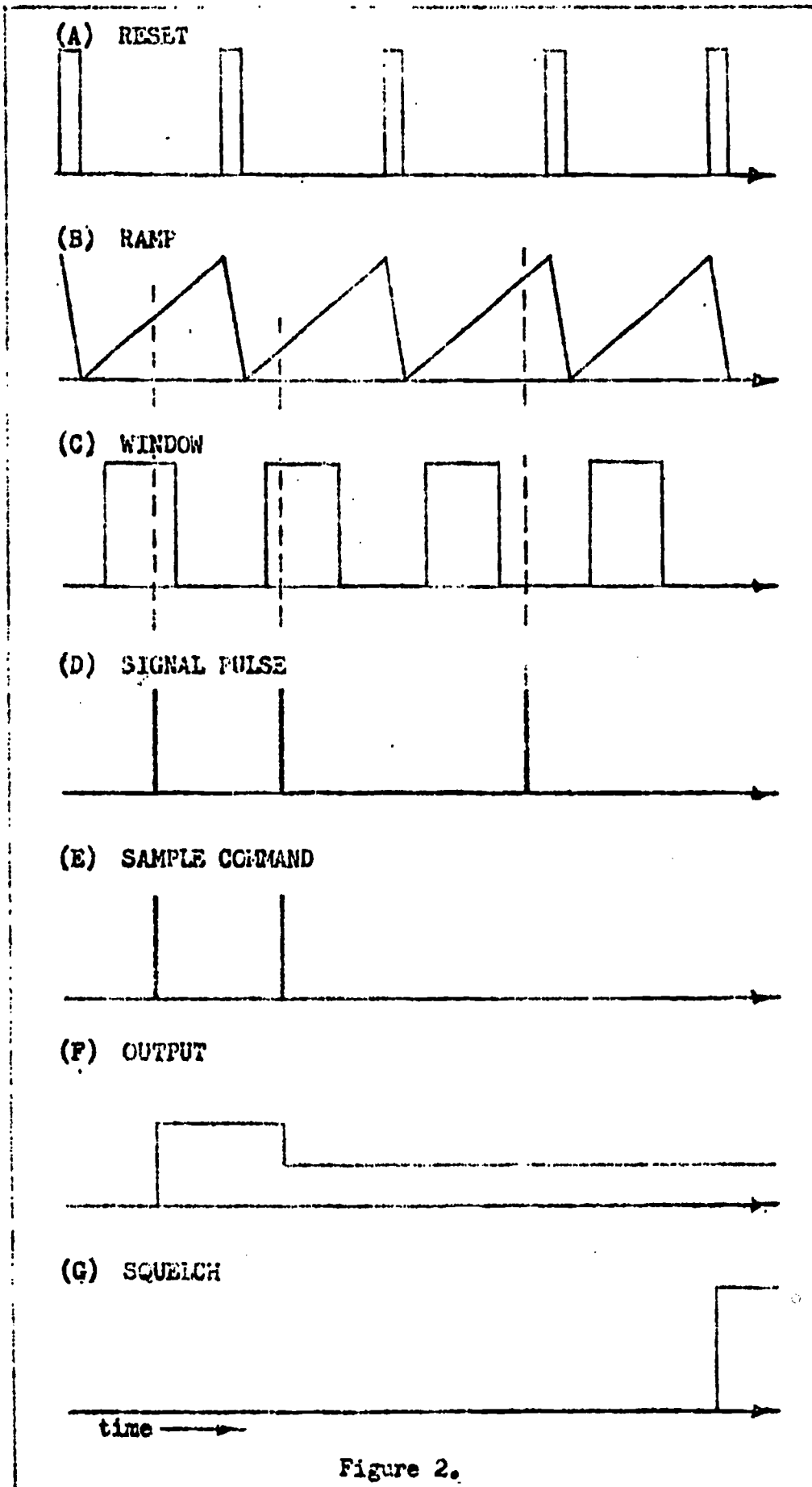


Figure 2.

J. Ricci

The harmonic synthesizer was originally conceived as a flexible instrument for on-line realization of complex musical events, and promises to be an extremely useful tool for experiments in auditory perception. What is desired is a device which will permit real-time synthesis of functions which are "small-scale periodic" - that is, functions whose Fourier component amplitudes are essentially invariant over an appreciable number of periods of the fundamental, although they may be varying dynamically with time. At any instant of time we would like to be able to specify the amplitudes of the first N harmonics of such a waveform by means of knob settings or, preferably, by means of control voltages. Pictorially, we would have the input-output relationship shown in Fig. 1, where input to the device are the control voltages:

$v_f(t)$: Proportional to fundamental frequency at time t.

$v_1(t), v_2(t) \dots v_N(t)$: Proportional to the amplitude at time t of the N harmonics.

And the device outputs an auditory waveform whose Fourier representation is:

$$v_0(t) = k_a \sum_{i=1}^N [v_i(t)] \cos [i k_f v_f(t)] t$$

Where $k_f v_f(t)$ is the radian frequency of the fundamental around time t.

An extra feature which we would certainly accept if we could get it would be some control over the phase relationships among the harmonics (Fig. 2). At $t=0$ we would initialize parameters ϕ_1 through ϕ_N so that the output function becomes:

$$v_0(t) = k_a \sum_{i=1}^N [v_i(t)] \cos \{ [i k_f v_f(t)] t - \phi_i \}$$

An early attempt at a design which has subsequently been abandoned involved extracting the harmonics of a spike train by means of voltage controlled bandpass filters and then scaling them individually using voltage controlled amplifiers. Useful as a conceptual tool, the approach would have been impractical for a number of reasons, to wit:

1. The VCF's would have been exorbitantly expensive to construct if, in fact, it proved possible to construct them at all.
2. Each such filter would require its own control voltage merely to position its center frequency at a given instant precisely where the desired harmonic is to be found.
3. Each VCF would introduce a phase shift and a considerable temperature-stability problem.

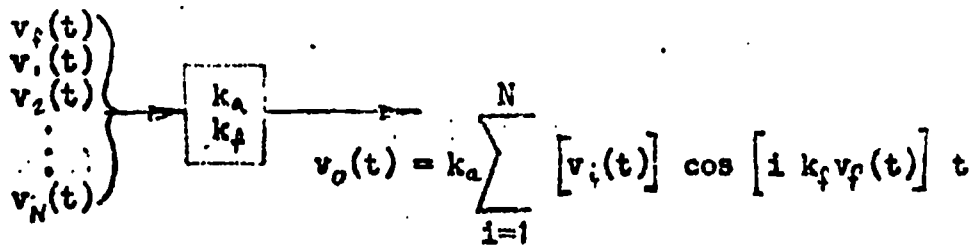


Figure 1.

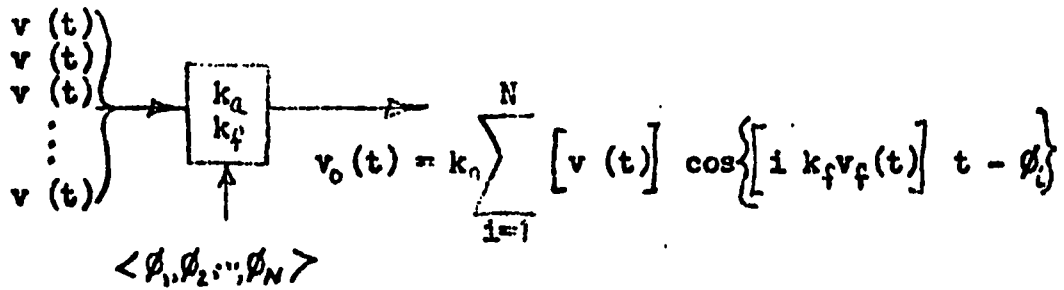


Figure 2.

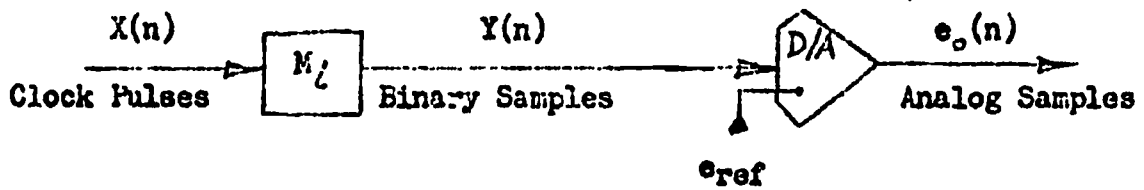


Figure 3.

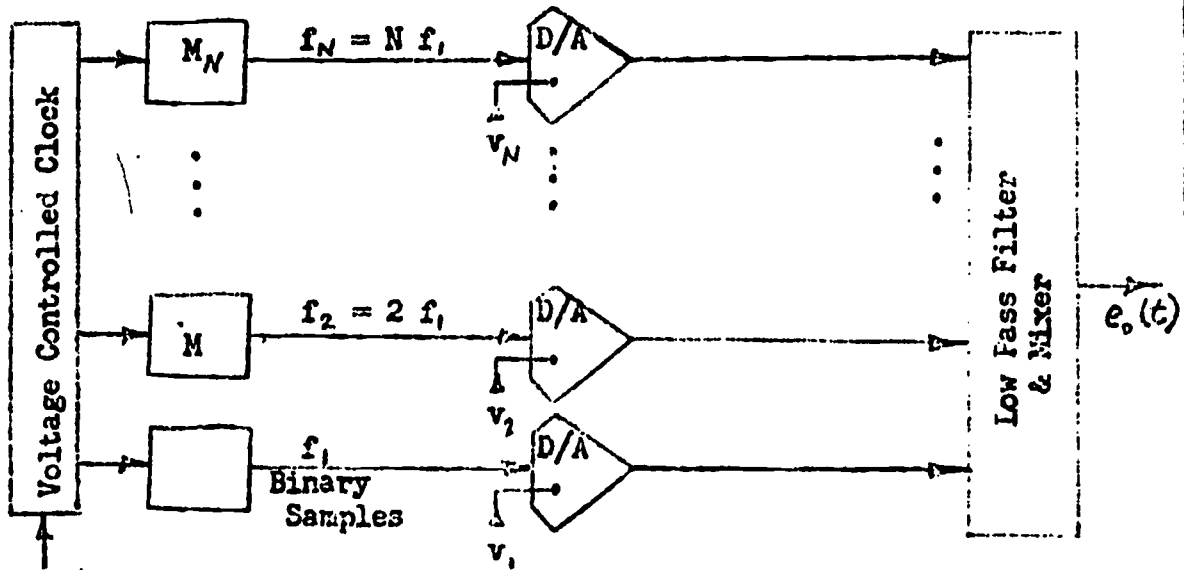


Figure 4. Digital Harmonic Synthesizer

Accordingly, some other approach is required. Suppose we had a sequential machine which would output discrete binary samples of an archetypal sinusoid (Fig. 3). Input, $X(n)$, to this machine is a train of clock pulses, and output consists of the binary samples:

$$Y(n) = \left[\cos \left(\frac{2\pi n}{T_n} \right) \right]^2$$

Where T_n is the number of sample points in one period. If this is then input to a D/A converter having reference voltage e_{ref} , output becomes:

$$e_0(n) = e_{ref} \cos \left(\frac{2\pi n}{T_n} \right)$$

Now, say we want this voltage to have frequency f . We pulse the machine at the rate $n/t = fT_n$ pulses/sec, and the output samples as a function of time become:

$$e_0(t) = e_{ref} \cos \left\{ 2\pi f s(t) \right\}$$

where $s(t) = \frac{i}{fT_n}$ for $\frac{i}{fT_n} < t < \frac{i+1}{fT_n}$ and $i = 0, 1, 2, \dots$

If this output is then low pass filtered at less than half the pulse rate, we have by the sampling theorem:

$$e_0(t) = e_{ref} \cos(2\pi ft)$$

Before considering further the difficulties posed by this kind of approach, I should like to make note of a few inherent advantages which make the attempt seem well worth pursuing:

1. The reference voltage, e_{ref} , need not be constant. In fact, the D/A converter is effecting a multiplication between e_{ref} and the sinusoidal samples. Thus, with a converter of the proper design, we should be able to replace e_{ref} by one of the $V_i(t)$ of Fig. 2 and thus obviate the need for voltage controlled amplifiers for amplitude scaling.
2. If we can synthesize each harmonic with one of these sections such that each section can be pulsed from the same voltage controlled clock, we eliminate all the control voltages which were required to center the VCF resonant frequencies.
3. There need be no phase difference among the harmonics unless we want them; and if special phase relationships are desired, we can specify them at will by advancing each individual section an arbitrary number of steps.

The form of synthesizer we are approaching is shown in Fig. 4. To make this a bit clearer, let us use some realistic numbers. Suppose we want a range of fundamental frequencies from 100 through 3000 HZ., and $N=20$ harmonics. Frequency response of the instrument is to be 20 KHZ., thus we will low pass filter at that frequency. To find the number of sample points, T_n , per cycle of the fundamental, we note that the sample frequency at worst-case condition is $100T_n$. The sampling theorem requires $100T_n > 2(20\text{KHZ.})$ which means we can use $T_n = 400$ samples/cycle and filter a shade below 20 KHZ. The machine for f_1 has one input

and (say) ten outputs (for 10 bit resolution); and it requires 400 states. Clock pulse rate is $400(100)=40\text{KHZ}$. For $f_1=100\text{HZ}$. And it becomes $400(3000)=1.2\text{MHZ}$. For $f_1=3000\text{HZ}$; well within the capabilities of modern logic components.

However, the device as postulated has one serious drawback - each of the 20 sequential machines has 400 states. The f_2 machine uses its 400 states to store samples of two cycles, the f_3 machine stores three cycles, etc. In order to make the device more economical, we would like to have each machine store only one cycle, resulting in the number of states getting small for the higher frequency machines. In fact, if we can manage this kind of trick, the number of states for M_i will shrink quite rapidly as i increases. Using the typical numbers above, consider that M_1 stores one cycle consisting of 400 samples. Let M_2 store one cycle of 200 samples (200 state machine). If we pulse both these machines at the same rate we will get two cycles out of M_2 in the time it takes to get one cycle from M_1 , which is precisely what we want. Similarly, M_4 has 100 samples, M_8 has 50, M_{16} has 25, etc. That is, the number of states for machine M_i is $400/i$. Naturally, the difficulty arises for those values of i which do not divide evenly into 400. In fact, if we wanted to design the device such that each harmonic were exactly generated by this scheme, the number of states of M_1 would have to be the product of all the prime numbers in the interval (1,20), and the pulse rates required would be unthinkable.

A solution to this dilemma currently under investigation is to approximate desired behavior in the following way. Machine M_i stores the integer part of $(400/i)$ samples. Clearly if we stopped here we would be in serious trouble, since inexact harmonics would be slightly higher in frequency than they should be, leading to successive interference effects from those pseudo-harmonics. However, the machines can be designed to inhibit the correct number of clock pulses so that at the start of each cycle of the fundamental all harmonics are in proper relation to each other. To make this clearer, let's consider M_3 . We store $\text{int}(400/3)=133$ samples in this machine. At the end of one cycle of the fundamental, the f_3 waveform would be $1/400$ ahead of where it should be. If, however, the 201st pulse is inhibited, then after 200 pulses f_3 will be $(1/800)$ ahead, after 201 pulses it will be $(1/800)$ behind, and after 400 pulses it will be coincident with the ideal f_3 . It is hoped that the amount of harmonic distortion introduced by such approximations will be negligible. If not negligible for a psychoacoustical laboratory instrument (in which case we can resort to equal-length machines), it might still be more than adequate for music and/or speech. Further, if necessary, the sampling theorem might be confoundable to some extent by low pass filtering the output mix at different frequencies as a function of f_1 . For example, for $f_1 < 500\text{HZ}$. We have $f_{20} < 10,000\text{HZ}$. And if we filter at 10,000 HZ. We only require 200 samples/cycle. For $500 < f_1 < 3000$, filtering at 20 KHZ, we would have a worst-case requirement of $500T_n > 40,000$ or $T_n > 80$. Thus with such a scheme, 200 samples would be adequate.

To reduce the number of states of the component machines, it may be possible to filter the outputs of the individual stages individually at

different cutoff frequencies, before mixing. Thus the number of states required by the sampling theorem may thereby be reduced. Of course, for an instrument of laboratory quality, much attention will have to be paid to the phase effects of such an approach.

Fig. 5 presents a set of waveforms generated directly via D/A conversion by computer simulation. For this case we have 160 samples for f_1 and are generating 11 harmonics. The "break" (pulse inhibition) positions for this set of photographs are arbitrary. In Fig. 6 we see the effects of the approximations on the filtered waveforms. Of course f_8 is a perfect sinusoid while f_7 and f_9 are not. The perceptual correlates of these distortions are under investigation. In particular, it will be interesting to discover whether apparent smoothness of a waveform effects perceived audio distortion. That is to say (Fig. 7) breaks which occur at points of zero derivative result in a waveform which appears to the naked eye to be a pure sinusoid, while breaks elsewhere are readily visible. In terms of Fourier spectrum the harmonic distortion is the same in both cases, but the visual difference is striking enough to make one wonder. In any event, we can apparently reduce the harmonic distortion as much as we are willing to pay for, and it seems that the attainment of an instrument of laboratory quality would be a relatively straightforward matter if adequate funding were available.

Figure 5. Data Samples
for f_1 through f_{11} .

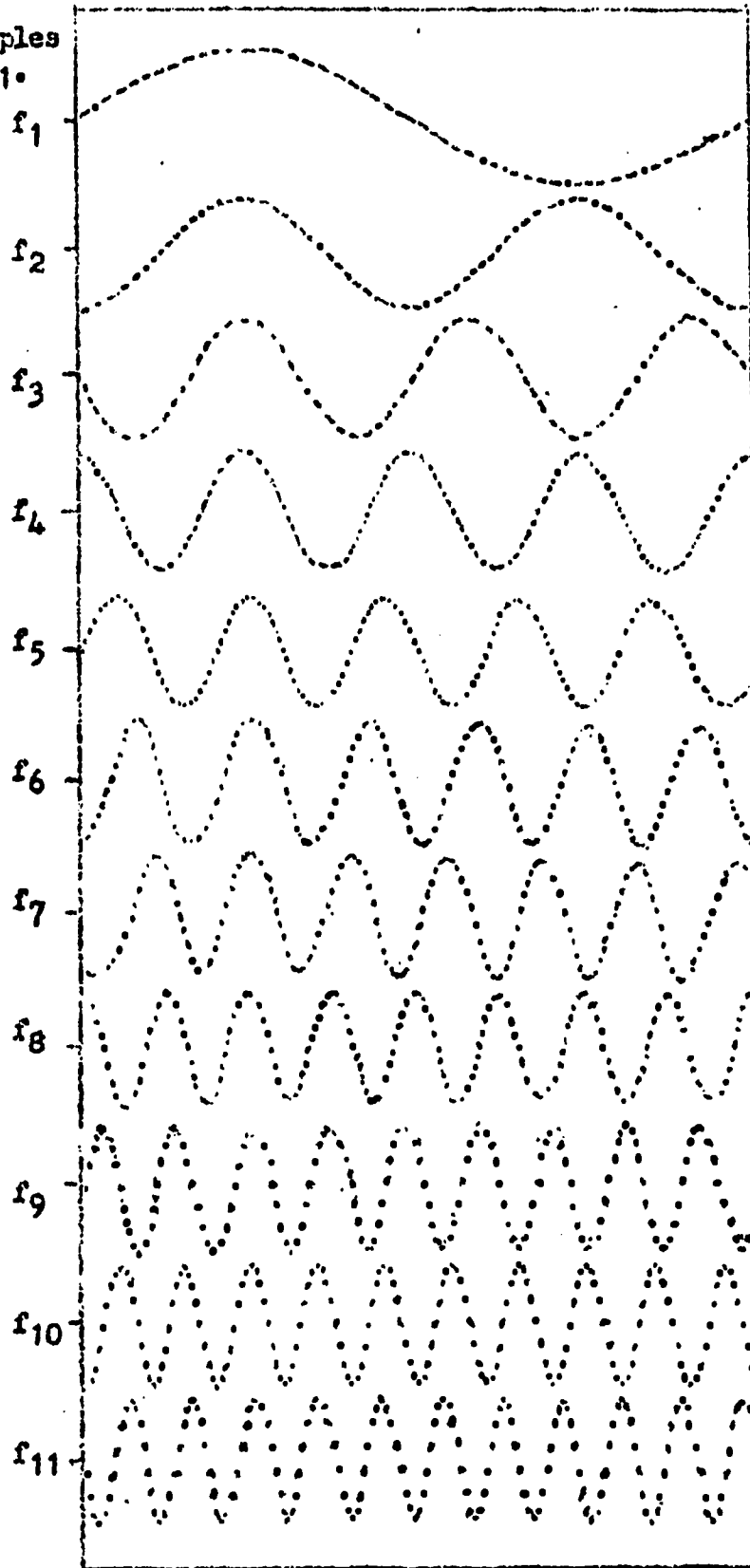


Figure 6.
Data Samples,
Before and After
Filtering

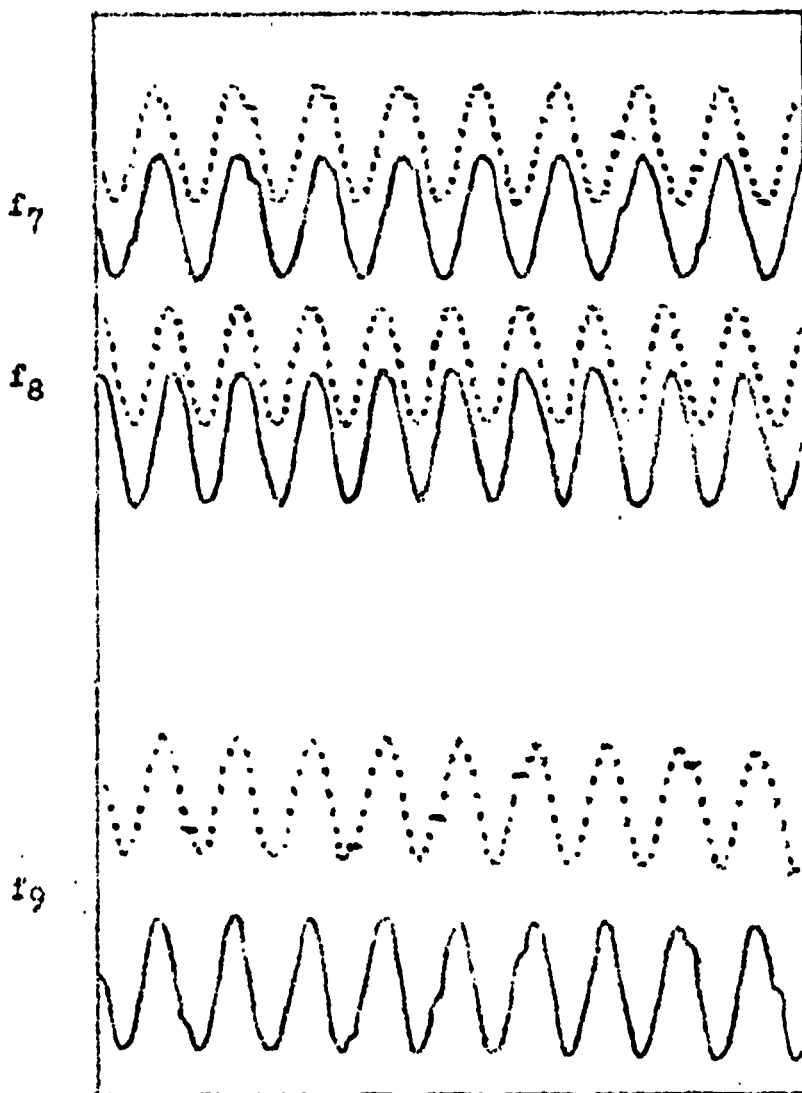


Figure 7.
Distortion at Breaks

