8409413

Pierce, Linda K.

PERCEPTION OF COLORATION IN DIOTIC REVERBERANT NOISE

City University of New York

PH.D. 1984

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

PERCEPTION OF COLORATION IN DIOTIC REVERBERANT NOISE

by

Linda K. Pierce

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1984

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

January 23, 1984
Date

Chairman of Examining Committee

January 23, 1984
Date

Executive Officer (acting)

David Berkley

Arthur Boothroyd

James M. Pickett

The City University of New York

Abstract

PERCEPTION OF COLORATION IN DIOTIC, REVERBERANT NOISE

by

Linda Pierce

Adviser: Harry Levitt

These experiments explored the perception, termed coloration, of the frequency domain characterization of reverberant sound. The beginning premise was that the standard deviation, σ, of the room frequency response, H(f), was a good measure of coloration. Using simulated rooms, reverberant noise stimuli were created which differed only in their perceived frequency spectra and not in their perceived temporal characteristics.

Four experiments were run. In two experiments subjects heard pairs of reverberant noises representing different rooms and made difference judgments; data from these experiments were analyzed using multidimensional scaling techniques. It was determined that coloration has both quantity and quality. Both quantity and quality of coloration are derived primarily from the early reflections, i.e., the first 10 msec, of a room's impulse response, h(t).

Quantity of coloration is related to the standard deviation of H(f), but is best described by a model which uses a critical band-like filter to smooth H(f) before calculating the standard deviation, σ_{CB}. Quantity of coloration can be manipulated with little variation in coloration quality by varying the reflectivity, β, of the surfaces in a given room.

Quality of coloration is a complex pitch-like quality associated with the particular reflections in a room. Differences in quality between rooms were described accurately by calculating the standard deviation of the difference, σ_{CBDIF}, between two smoothed room spectra.

In the third experiment, subjects judged quantity of coloration in two tasks, a paired comparison task and an absolute judgment task, and σ_{CB} was substantiated as a measure of quantity of coloration.

In the last experiment, a Thurstone paired comparison task and analysis was used to determine that the range of the coloration quantity continuum is about 5 1/2 jnd's.

ACKNOWLEDGEMENTS

Many people helped me while I was doing this work, and I would like to acknowledge their efforts and thank them for it.

This research was carried out in the Acoustics Research Department at the Murray Hill AT&T Bell Laboratories. The individuals in that department provided a stimulating and supportive atmosphere in which to work. My supervisor at Bell Laboratories, David Berkley, and my advisor at CUNY, Harry Levitt, contributed both substance and encouragement. In particular, David Berkley suggested the stimulus generation method used throughout these experiments, and provided day-to-day guidance in using the extensive computer laboratory facilities in the department as well. In addition to his advisory functions, Harry Levitt suggested the idea for Experiment 4.

Two other individuals gave considerable time and help. I had many valuable discussions of the data with Michael Pavel; he suggested the filter used in the σ_{CB} measure and helped with programming it. Barbara McDermott gave me extensive instruction and advice with respect to the use of the multidimensional scaling programs and regression analyses that were needed in both Experiments 1 and 2; she also acted as a daily source of encouragement.

I had useful and interesting discussions with Jont Allen and Tom Landauer. Also, Kathy O'Connor's instructions concerning the facilities for running subjects made that process much easier. Finally, I would like to thank Arthur Boothroyd for a meticulous reading of this dissertation which provided corrections to both form and content.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Reverberation is reflected sound; it can be described physically in terms of the time or frequency domain. Temporally, reverberation in a room is described by the room's impulse response, i.e., the delay and level of the echoes produced by an impulse in the room. The room frequency response, H(f), is calculable from the impulse response, h(t), and describes the effect of the delayed impulses, the echoes, on the flat wideband spectrum of an impulse. Aspects of both the temporal and frequency domain characteristics of reverberant sound are perceptible, and have been studied in a variety of contexts for two main purposes: (a) to pursue an understanding of fundamental auditory phenomena and (b) to describe how reverberation interferes with the perception of speech.

With respect to the first of these purposes, the perception associated with the frequency domain description of reverberant sound has increased the understanding of the auditory filter (or critical band). Reverberation expressed in the time domain, as echoes, has been informative with respect to echo suppression and the ear's ability to integrate sound over short periods, i.e., less than about 50 msec. In combination with interaural phenomena such as masking level differences and localization, these perceptions associated with echoes represent the ear's ability to use briefly delayed sounds to enhance perception.

The second reason for studying reverberant sound concerns the ear's inability to integrate sound over longer time periods. Because of this, reverberant energy may act to interfere with the perception of the original sounds. This is especially true of speech stimuli which have relatively short steady state portions. It is also especially true for hearing impaired listeners and the reasons for this are not entirely clear.

Studies concerned with the first purpose, the understanding of fundamental phenomena, have most often used well defined experimental manipulations, e.g., single echoes; noise or sinusoids, rather than speech, have been the typical base stimuli. Studies in the second area, the effect of reverberation on speech perception, have most frequently used complex experimental conditions, i.e., real rooms, and complex stimuli, i.e., speech. As frequently happens, real world conditions are gained at the cost of experimental definition and control.

There has been a gap between these two major research directions which is only beginning to be filled by the ability of high speed computers to simulate real room conditions without giving up the ability to define exactly the experimental conditions. The first step in this direction was a computer program which simulates room acoustics (Allen and Berkley, 1979). The availability of this tool has made it possible to begin to study the perception of complex reverberant stimuli in a well-controlled environment. The studies to be reported here are a part of this research effort.

2. REVIEW OF THE LITERATURE

Let us begin with a review of the two main research directions and a brief review, as well, of other areas which will be needed to carry us through an understanding of the experiments to be discussed. These experiments will be concerned with coloration, the perception of the frequency domain characterization of reverberation, and the first step is to understand the critical band concept and relate it to coloration.

2.1 AUDITORY PHENOMENA OF DELAYED SOUND

2.1.1 Frequency Selectivity, Critical Bands

An echo perturbs the frequency response of the signal by producing a ripple in the frequency response inversely proportional to the delay, e.g., $f = 1/t$ where f is the spacing of the peaks in the ripple and t the echo delay. Thus an echo with 30 msec delay produces a ripple with peaks spaced at intervals of 33 Hz.

The sensitivity of the ear to variations in the frequency spectrum depends on its peripheral processing of the incoming waveform. The output of this filtering process by the peripheral ear represents the signal available for central processing. Both speech and music perception as well as the perception of the coloration of the waveform produced by reverberation are the results of this filtering process.

The degree to which the ear is capable of frequency analysis has been widely studied psychologically through studies of the auditory filter, termed the critical band.

Fletcher's classic work (1940) introduced the critical band concept and provided several demonstrations of it. He showed that the threshold for a tone in broadband noise increases with frequency. Also, increasing the bandwidth of noise about a tone will produce an increase in tone threshold up to some bandwidth (i.e., a critical bandwidth) and thereafter the threshold does not increase. He also demonstrated that the critical band increases with frequency. This concept has been fundamental to both psychological and physiological acoustics since that time, and evidence concerning the critical band's width, shape, and how it relates to other auditory phenomena have steadily accumulated.

2.1.2 Critical Bands and Coloration Perception

Let us be clear about the implications of critical band phenomena for the perception of echoes, and, in particular, the perception of the coloration of reverberant sound. The idea is that fluctuations in the frequency spectra that are within a single critical band will be smoothed by the auditory filter and, therefore, less perceptible than fluctuations that range over more than a critical band. Thus short delays, which produce only a few ripples, widely spaced, in the frequency response, should yield lower thresholds than long delays which produce many ripples.

Atal, et. al., (1962) looked at this directly: the threshold for the perception of ripple in white noise was determined for rippled noise created from impulse responses with single echo delays of up to 160 msec. Sensitivity to the ripple was most sensitive for delays of about 5-10 msec with the threshold increasing up to a delay of about 60-70 msec. The ripple of the frequency response was not detected for delays greater than 70 msec, which translates to a ripple frequency of about 15 Hz. This is consistent with what should be expected based on critical band data.

This result has since been substantiated by Zurek (1976) and Koenig (1979) in studies of echo perception and by Bilsen and Ritsma (1970) studying repetition pitch. The Zurek and Koenig studies provide important binaural processing information about echo detection. They, independently of each other, determined ripple thresholds for a dichotic condition in which the ripple at one ear was displaced 180° from the ripple at the other. The combined frequency response, displayed diotically, will yield a flat, unrippled frequency spectrum in contrast to the sinusoidal ripple in the frequency spectrum of their diotic listening condition. Both sets of results demonstrated that the dichotic condition required ripples with larger peak-to-valley ratios for threshold than the diotic condition. The inability to listen optimally, i.e., monaurally, in the dichotic condition, demonstrated a previously unknown constraint for binaural listening.

2.1.3 Physiological Frequency Selectivity

Some context is now needed in which to understand more broadly the ear's capability to perform frequency analysis. While it is the psychological phenomena associated with frequency selectivity that are of interest in this paper, considerable physiological work has sought the underlying mechanism (or mechanisms). From von Békésy's early studies with human cadavers to studies with increasingly more refined techniques (Johnstone and Boyle, 1967, Rhode, 1971, and Russell and Sellick, 1977) it has been clear that the frequency response of the basilar membrane is one of the determiners of the ear's frequency selectivity, and whether it suffices or whether there is further mechanical sharpening of the frequency response between the membrane and the hair cells is not yet known. Russell and Sellick's recordings within the hair cells found tuning curves as sharp as neural tuning curves (Kiang, 1965), thus effectively moving the sharpening site into the mechanical or chemical realm preceding the firing of the first neural layer.

The auditory filter is an area of audition in which both psychological and physiological observations have been obtained. Difficulties in both areas arise due to method: measures of frequency selectivity vary greatly depending on how they are measured and how the results are interpreted. It has been tempting for experimenters to compare measures of frequency selectivity obtained in the two areas, such as the half power frequency range for basilar membrane displacement and for a critical bandwidth. It is reasonable to expect that such bandwidths are correlated, i.e., if the basilar membrane response widens with frequency, then the critical band should also increase with frequency. However, it requires an intuitive leap for which there is no good empirical evidence to expect them to be directly comparable. As we shall see, studies within the psychological realm yield variance enough.

2.1.4 Critical Band Width

Masking studies have dominated the psychological investigations of frequency analysis. Fletcher's (1940) paper hypothesized that the noise power needed to mask a tone is equal to the power of the sinusoid. From thresholds in wideband noise, he calculated the rectangular bandwidths needed to mask sinusoids at different frequencies: 51 Hz at 250 Hz, 63 Hz at 1000, and 204 Hz at 4000. Hawkins and Stevens (1950) also determined threshold measurements in wideband noise and used them to calculate the same critical band as Fletcher, and their results were in good agreement with his. Table 2.1 includes these and other critical band estimates.

Although Fletcher and Hawkins and Stevens refer to critical bands, a different nomenclature has been more recently suggested and generally adopted (Scharf, 1970). The bandwidths determined from thresholds in wideband noise are termed critical ratios (rows 1 and 3 in Table 2.1) while empirically determined measures (rows 4-9 in Table 2.1) are termed critical bands. Empirically obtained critical bands are about 2 1/2 times as large as the calculated ones obtained by Fletcher (1940) and Hawkins and Stevens (1950). This is seen as a simple empirical fact: the noise power needed to mask a tone is not equal to, but about 4 dB greater, than the signal power (Zwicker, et. al., 1957). Psychological measurements, of either the critical band or critical ratio, have produced much smaller estimates than the same bandwidths which Green (1965) cited from von Békésy's physiological measurements of basilar membrane displacement: 190, 760, and 3000 Hz at 250, 1000, and 4000 Hz, respectively (row 2 of Table 2.1).

Table 2.1 Critical Bandwidth Estimates

		250	500	800	1000	3000	4000
1.	Fletcher	51			63		204
2.	von Békésy	190			760		3600
3.	Hawkins & Stevens	36			63		320
4.	Schafer, et. al.	65		65		280	
5.	Hamilton			140			
6.	Zwicker, et. al.		160		180		
7.	Swets, et. al.				95 41		
8.	Greenwood		109		185	420	
9.	Zwicker	100	115		160	470	700

In the 1950's and 1960's the question that predominated the study of the auditory filter was the width of that filter, or more specifically, the width of the filter as a function of frequency. Besides determining auditory filter information from wideband masking studies such as Fletcher (1940) and Hawkins and Stevens (1950), a host of experimenters (Schafer, et. al., 1950, Hamilton, 1951, Webster, et. al., 1952, Green, McKey and Licklider, 1959, Greenwood, 1961, Swets, et. al., 1962, and Green, 1965) used other masking approaches to evaluate the critical band.

Schafer, et. al. (1950) used narrow bands of noise (as did Greenwood, 1961), centered at 200, 800, and 3200 Hz, created from groups of tones to obtain sharp masker edges, and determined the thresholds for sinusoids in the area of each noise band. Their observed critical bands were 65, 65, and 280 Hz for the noise bands at 200, 800, and 3200 Hz, respectively. Since their data showed no sharp break in the masking function such as a rectangular filter would predict, they suggested a resonance curve for a single tuned circuit might better approximate the ear's frequency selectivity, and they presented their data along with predicted curves based on their and Fletcher's assumptions. Their predictions appeared to better fit the data. Studies such as this led other experimenters to question the relation between the critical band and assumptions made about the shape of the auditory filter.

From another point of view, Egan and Hake (1950) showed the asymmetric upward spread of masking, and Patterson (1974) demonstrated that much of this effect could be predicted from the widening of the filter with frequency. Egan and Hake also showed that the upward spread of masking increased dramatically with masker level. While the bandwidth for the 3 dB down point does not vary much with masker level, the tails of the masking function do widen considerably with level. The implication here is that the filter broadens at high noise levels. Patterson (1974) concurred with this; he found that the filter shape was relatively constant for noise levels up to 40 dB spectrum level and that it widened with higher noise levels.

2.1.5 <u>Filter Shape and Filter Width</u>

This particular issue has been an important one and Swets, et. al. (1962) explored the effect of filter shape assumptions on the critical band. Using a four alternative forced choice signal detection paradigm, for a tone they determined d' as a function of the width of a narrow band of noise and also d' as a function of the level of broadband noise. They assumed that the d' function for the narrowband data represented performance when an external filter determined the data, and that the d' function for the broadband data represented the internal filter. Changes in d' as a function of energy permitted through the external filter were then assumed to be equal to changes in d' due to energy permitted through the internal filter. The latter d' difference could be referred to a difference in noise power and used in combination with assumptions about the filter shape to calculate the critical bandwidth. They assumed a single-tuned, rectangular, or Gaussian filter, and the critical band at 1000 Hz was calculated to be 41, 95, and 80 Hz, respectively, which indicates the sizable effect of assumed filter characteristics.

Noise masking paradigms are not the only situations in which critical band phenomena are observed. One of the early fundamental studies in this area, performed by Zwicker, et. al. (1957), shows loudness summation to be affected by the range of tones in a complex. A four tone complex was matched in loudness to a single tone at the center frequency (or vice versa). The level of the comparison remained unchanged until a critical band was reached at which width the loudness of the complex increased.

In a second study, the SPL of a narrow band of noise was held constant while varying the bandwidth; the subjects performed loudness matches with a tone at the center of the noise bandwidth. The same result occurred; the loudness of the noise band remained constant with increasing bandwidth until, at a critical bandwidth, the loudness of the noise band began to increase.

2.1.6 Summary of Critical Band Data

In general, the critical band is relatively constant for frequencies up to 500 Hz and is nearly a linear function of log frequency thereafter. (Very recently, Moore and Glasberg, 1983, have suggested that the width of the critical band continues to decrease with frequency below 500 Hz.) Table 2.1 includes Zwicker's (1961) summary of critical bandwidths. A comparison of these numbers with a sampling from the rest of the literature (Table 2.1) indicates considerable variability across experiments. In spite of this, the functions relating critical band and critical ratio to frequency are similar, and they are also similar to the jnd scale for frequency and the mel scale, and, probably, the position of stimulation on the basilar membrane. It is clear that Zwicker's summary at this time suggested a fundamental auditory phenomenon at the physiological level accounting for diverse psychological effects, and this general conclusion still holds today.

The Zwicker summary has been used as a standard for frequency selectivity, and Zwicker has used it as well as a basis for his widely accepted measure of loudness (Zwicker and Scharf, 1965). This work is entirely empirically based, and subsequent work in this area, much of it by Patterson and his colleagues (Patterson, 1974, Patterson, 1976, Patterson and Henning, 1977, Weber, 1977, and Patterson and Nimmo-Smith, 1980) has aimed at developing a structure, a model with psychological or physiological overtones, from which to predict critical band phenomena.

2.1.7 Approach to a Critical Band Model

The Patterson effort to determine the auditory filter shape has the same basis as Fletcher's (1940) work: the power of the signal at threshold is proportional to the masker power within the auditory filter, i.e., $P_s = k\int N(f)|H(f)|^2 \partial f$ where P_s is the signal power, k the constant of proportionality, N(f) the noise power spectrum, and H(f) the auditory filter shape (from Patterson, 1974). Where Fletcher assumed k to be unity, Patterson has explored it as a fitting parameter (Patterson, 1974) and as an index of processing efficiency (Patterson, et. al., 1982).

The Patterson work has been characterized by the collection of extensive data in order to fit multi-parameter functions for H(f). An important influence determining his empirical approach has been the potential for off-frequency listening. Consider a sinusoid and a nearby band of noise to one side (in the frequency domain) in a tone detection task. If a subject listens at, i.e., centers his auditory filter on, the tone frequency, some proportion of the noise power will be under the filter and will act as a masker. By listening off the tone frequency, i.e., centering the filter to the side of the tone away from the noise, a considerably lesser proportion of the noise power may be under the filter without an equal decrement in the signal, thus improving the signal to noise ratio available for detection. If listeners can do this, then studies which have moved a signal through a narrowband masker, Schafer, et. al. (1950) and Greenwood (1961), for example, have probably yielded estimates of the critical band which were too narrow. On the other hand, data from other sources, such as wideband noise data, (Fletcher, 1940, Hawkins and Stevens, 1950, Green, et. al., 1959) or loudness summation data (Zwicker, et. al., 1957) should not require reinterpretation.

2.1.8 Off-Frequency Listening and Asymmetric Filters

Leshowitz and Wightman (1971) demonstrated off-frequency listening in a variety of tone detection paradigms. For example, when the detection task is to detect a brief pulsed tone increment of a masking tone at the same frequency (tone-in-tone detection), the listener may listen on-frequency to detect the increment in the masker or he may listen off-frequency to the energy in the sidebands produced by brief signals. The authors used a 1300 Hz masker and signal with on times of the signal of 10 or 100 msec, and determined signal thresholds in an unfiltered condition and in a condition with a very narrow passband at 1300 Hz. The signal levels required for threshold in the filtered conditions were about 20 dB higher than in the unfiltered conditions; the authors attribute this difference to off-frequency listening to side-band energy in the unfiltered condition.

In an effort to gain qualitative information about the shape of the filter, they also obtained thresholds for conditions where the masker and stimulus were high-pass and low-pass filtered. The results indicated the threshold for the low pass condition was only about 5 dB greater than for the unfiltered condition while the high pass condition produced a 25 dB higher threshold. The high pass condition had the high frequency sideband present; since this did not appear to be used as much as the low frequency sideband by the listeners, it suggests an asymmetrical auditory filter with a steeper high frequency than low frequency slope. The masker level was set at 80 dB SPL in this study, and these data are consistent with others (Egan and Hake, 1950) which show considerable upward spread of masking at high stimulus levels.

The size of the off-frequency listening effect has led to empirical work designed to control or evaluate its effect by using noise bands placed both above and below (notched noise) the tone to be detected. In such a case, moving the filter above or below the signal frequency helps relatively little because the energy under the filter skirts in one direction increases as it decreases in the other; the extent to which this is true depends, of course, on the symmetry of the filter. In an important paper, Patterson and Nimmo-Smith (1980) point out that most of the work concerned with the shape of the filter (Patterson, 1974, Margolis and Small, 1975, Houtgast, 1977, Weber, 1977, Patterson and Henning, 1977, and Small and Tyler, 1978) has either assumed the filter to be symmetric or has assumed on-frequency listening. They show the power masking equation expressed in terms of a notched noise masker:

$$P_s = kN_o \int_{-\infty}^{f_l} |H(f)|^2 \partial f + kN_o \int_{f_u}^{\infty} |H(f)|^2 \partial f$$

where f_l is the lower frequency cutoff and f_u the upper. If there is on-frequency listening, a single noise band can be used, one of the terms on the right hand side of the equation disappears, and moving the edge of the noise band (f_l or f_u) moves the amount of noise under the filter's edge to provide a measurement of the filter area overlapped by the noise. Off-frequency listening, however, means the filter as well as the noise edge moves and the threshold measure is no longer providing a measure of the area of the filter overlapped by the noise.

If, on the other hand, the filter can be assumed to be symmetric in linear frequency, and if f_l and f_u are equidistant from the signal, the two terms on the right side of the equation can be added, and threshold again is a measure of filter area. While data such as Egan and Hake's (1950) suggest that a symmetric filter assumption may not be bad at low masking levels, their data and others (especially Leshowitz and Wightman, 1971, and also Weber, 1977) show this is not appropriate for high intensities. Based on the literature, Patterson and Nimmo-Smith conclude, however, that the symmetric filter assumption at moderate levels is to be preferred to an assumption of on-frequency listening, and prior to their work, one assumption or the other was necessary to make the theoretical equations solvable.

2.1.9 Critical Band Model

Patterson and Nimmo-Smith suggest that the best simple model for filter skirts is a pair of negative exponential functions with a rounded top. In particular, they point out (1) the signal threshold in dB plotted against linear frequency yields a straight line (data from Patterson, 1976, and Weber, 1977) and (2) a straight line is also obtained when the threshold is plotted against linear frequency for two tone masking data (Patterson and Henning, 1977). However, for two tone masking data, Patterson and Henning showed the power model with the signal in the middle of two masking tones reduced to

$$P_s = 2kM_of_o|H(\Delta f/f_o)|$$

where M_o is the level of each masking tone, f_o the signal frequency, and Δf the distance in frequency between the masker and the signal. In this case the threshold data is a direct measure of the filter shape rather than of its area, and exponential functions have the same form as their derivatives ($de^x/dx = e^x$). Thus the fact that these functions are straight lines and that they represent measures of both the filter area (noise maskers) and shape (tone maskers) are arguments for an exponential function in an auditory filter model.

The authors used an experimental paradigm which encouraged off-frequency listening. They fixed the first noise band at a distance from the signal frequency (2000 Hz) and then varied the distance from the signal of a noise band on the other side of the signal. The distance of the first noise band from the signal was varied systematically as was whether it was above or below the signal frequency, resulting in a family of empirical functions in which the fit of the model to on-frequency listening/off-frequency listening assumptions and symmetric filter/asymmetric filter assumptions could be tested. The function they used to represent the filter was $H(g)^2 = p(\lambda g)e^{-\lambda g}$ where $g = (|f_0 - f|/f_0)$, and λ was a cubic polynomial. Thus three parameters and k were used to fit the data.

Their data indicated the presence of off-frequency listening and an asymmetrical filter, with the off-frequency listening being by far the larger factor. The resulting filters had steep slopes, about 100 dB/octave, near the center of the filter, and the filter widths at the 3 dB down point were 216 and 196 Hz for the two subjects. The slopes broaden to about a 40 dB/octave slope after the first 30 dB drop at 100 dB/octave. While the steep slopes are reminiscent of tuning curves, the symmetry of the filter and the broadening slope on the high frequency side are not like neural data. This is true even if the data are plotted in log frequency; this steepens the high frequency side but not enough to match tuning curves. The authors do not attempt to interpret their model in the physiological domain and it is, indeed, not clear at all how to relate them. It is, of course, not necessary to relate them in order for this body of work to be significant in pulling together and interpreting data in the psychological realm.

2.1.10 Application of Model

Patterson, et. al. (1982) used both the notched noise experimental method and a negative exponential model for the auditory filter to study frequency selectivity as a function of age. They suggest that an increase in threshold in broadband noise for tones or speech can be attributed either to a broadened auditory filter or to poorer processing efficiency following the filter. For a normal ear, the threshold drops about 35 dB at 1000 Hz from a noise with a $\Delta f = 0.0$ notch width (broadband noise) to a noise with a notch width of 0.4 (800 Hz at 1000 Hz). In a normal ear with impaired processing efficiency, this same function would be expected except with a higher threshold of some fixed dB throughout. In terms of the model explored in their paper, k would be larger for an individual with impaired processing efficiency. In an ear with impaired frequency selectivity, the decrease in threshold from no notch to an 800 Hz notch should be less than the 35 dB for a normal ear; that is, the function would be flatter than the normal ear function. Notched noise thresholds were determined for subjects between the ages of 23 and 75 and it was found that filter width increased systematically with age from ages 23-65. For subjects over 70, however, filter width was very variable across subjects. The constant, k, as a measure of processing efficiency did not increase with age and, across frequencies, did not remain constant for a subject. While their data suggest that notched noise thresholds may be a sensitive measure of auditory filter width, the parameters of the model used in this experiment do not mesh particularly well with parameters concerning auditory acuity. Future work may show how best to fold auditory filter measurements and models into the diagnostic battery available for assessing auditory impairments.

2.1.11 Monaurally Perceived Echo Delays

When a primary sound is followed by reflections, these successive reflections appear to be suppressed. Green (1976) gives intuitive, anecdotal evidence for this when he points out that we rarely hear echoes in normal sized rooms, although perhaps most of the sound energy is reflected. Haas (1949, English publication, 1972) studied the suppression of a single echo by seating his subjects on the rooftop facing two loudspeakers 90° apart, and varying the delay between the loudspeaker outputs. Using speech stimuli, he found that the delayed speech had to be 10 dB more intense with delays of 5-30 msec in order for the loudspeakers to be perceived as equally loud. If the loudspeakers were set at equal intensities with a 10 msec delay, the second loudspeaker was not perceived at all. At equal intensities and at a delay of 40 msec, the second loudspeaker was "recognized as an additional sound source" and over 50 msec delay finally yielded a "separate echo." This is evidence for a strong echo suppression effect.

Although Haas' subjects listened binaurally, Green (1976) points out that echo suppression is a monaural effect. It is the fact that the echoes occur sequentially, in a short time period, that produces the suppression. Green also points out that if the second, delayed, "unperceived" sound source is silenced, it is noticeable. This argues for a merging or integration of sounds within some time period, rather than complete suppression.

Lochner and Burger (1958) further explored the suppression of a single echo in terms of the ear's ability to integrate the energy of the source and a delayed signal. Their subjects listened in an anechoic environment to two loudspeakers, 90° apart, to speech material consisting of Harvard PB 50 articulation test lists. Using low speech levels to produce articulation scores of less than 100% in quiet, they determined that a non-delayed stimulus emitted from both loudspeakers gave a 3 dB articulation advantage over a stimulus from a single loudspeaker. They then increased the delay between the loudspeakers to determine how long the 3 dB advantage would endure. They found the full 3 dB advantage was maintained with up to 30 msec delay and that there was some partial advantage until the delay reached 80 msec. In good agreement with Haas' finding that a delayed source must be louder that the primary to be equally loud, with a 10 msec delay they found an additional 12 dB was needed at the second loudspeaker to create a perception of equal loudness.

Nábělek and Robinette (1978) studied delays from 5-160 msec for binaural and monaural listening and for both normal and hearing impaired listeners. For both groups of binaural listeners, they found no decrement in speech perception with delays of up to 20 msec. Their binaural and monaural data are based on different subjects which makes it difficult to compare data for the hearing impaired listeners, but it is possible to compare the binaural and monaural data in the case of the normal hearing subjects. The binaural condition for the normal hearing subjects produces higher scores than the monaural condition for delays as high as 80 msec, in agreement with Lochner and Berger. Nábělek and Robinette point out their results can be discussed in terms of effects in the time domain, i.e., temporal integration of sounds close in time and masking of more separate sounds, or in the frequency domain, i.e., the ripple in the frequency spectrum produced by delay. They and the other authors concerned with this topic find their results most plausibly discussed in terms of a temporal parameter, reverberation time.

2.1.12 Binaural Delay Effects

Reverberant sound produces monaural delays in the reflections that follow the source. The monaural phenomenon of echo suppression deals with a considerably longer delay interval, up to 80 msec, than the binaural delay of less than a millisecond associated with localization and MLD's. A short interaural delay from a single sound is produced by the head itself. A sound source located out of the medial plane will produce both interaural time and intensity differences. Interaural time differences yield perceptible phase differences at low frequencies (below 1000 Hz), and the head shadow produces large interaural intensity differences at high frequencies (over 4000 Hz). Stevens and Newman (1936) demonstrated a listener's sensitivity to both sets of these differences by locating a subject in a semi-anechoic environment, a chair elevated above the flat roof of a university building, and testing the listener's ability to localize tones of different frequencies as well as other stimuli. Performance was good at low and high frequencies while errors peaked in the mid-range, around 3000 Hz.

An extension of the capacity for processing interaural differences is represented by masking level differences (MLD's). MLD's demonstrate improved signal detection in the presence of noise by capitalizing in some way on interaural phase differences in the signal or noise. First demonstrated by Licklider (1948) and Hirsh (1948), an MLD of 12-15 dB can be achieved over a base condition (in which both signal and noise are diotic) by reversing the phase of the signal or the noise. Maximum MLD's of 12-15 dB are achieved at low frequencies where interaural time differences represent perceptible phase differences. This phenomenon probably produces (or is partially responsible for) the "cocktail-party" effect (Koenig, 1950), the ability to attend to a particular speaker in a noisy environment. Studies of speech perception in a noisy environment (Nabelek and Pickett, 1974, for example) also demonstrate MLD's in improved performance for binaural vs. monaural (near ear) perception. Also notable in this work is that increased reverberation reduces the magnitude of the MLD phenomenon.

Evident in both binaural and monaural settings is the precedence effect; this is the considerably greater importance of a primary sound in relation to its echoes. Wallach, Newman, and Rosenzweig (1949) demonstrated the importance of the primary sound for lateralization. They presented a pair of clicks binaurally within 3 msec of each other. The second pair of clicks was offset some amount temporally in one direction and they varied the amount the first pair was offset in the other direction to determine when a centered image was obtained. They found the temporal difference between the second pair of clicks needed to be about 6 times as great as the opposite temporal difference of the first pair for a centered image.

2.2 REVERBERATION AS INTERFERENCE

Let us now turn to our second major area of concern: reverberant sound as interference. We have seen that interaural delays are used as cues for localization and for signal selection and detection, and that the effect of echoes produces the precedence effect and can be used to study the ear's frequency selectivity. Binaural delays created by the head are significant when they are somewhat less than a millisecond and monaural delays are most significant for echo suppression when on the order of 5-30 msec and have some echo suppression effects out to about 80 msec.

2.2.1 Reverberation Time

In most listening environments, reflected sound is produced with delay times considerably longer than one or even 80 msec. The descriptive statistic used for reflected sound is reverberation time, the time for sound to decay 60 dB from the SPL of the source measured at the position of the receiver. This interval will be referred to as T_{60} in the following sections. It is noteworthy that this statistic is a temporal measure, although we have seen the effect of echoes in both time and frequency domains. This single statistic is used to describe a complex phenomenon, and, with respect to the literature concerning speech perception performance in a reverberant environment, it appears to be a useful descriptor.

Nábělek and Pickett (1974a) refer to 300-600 msec reverberation times as typical for medium-sized classrooms and Knudsen (1929) cites cases of auditoriums with 2-8 sec reverberation times. Echo delays that are longer than the ear's ability to integrate, and most rooms produce them, interfere with, rather than enhance, speech perception. Both normal hearing and hearing-impaired subjects show impaired performance as reverberation time increases. This is probably the most general finding in the reverberation perception literature, and it suggests that in most rooms the enhancing effects of the early echoes are more than offset by the later ones.

2.2.2 Variables for Reverberation Perception

The primary variable in this area has been reverberation time. In some experiments subjects have listened in real rooms with changeable surfaces to permit manipulation of reverberation time (Nábělek and Pickett, 1974a and b, and Finitzo-Hieber and Tillman, 1978). In other cases a recording has been made in the reverberant room, sometimes using KEMAR (Moncur and Dirks, 1967, Bloom, 1980, and Newman, 1982), and sometimes not (Gelfand and Silman, 1979). Subjects have then listened to the test material through headphones. The first set of conditions is most realistic and the second introduces an additional element of control over the stimuli. Gelfand and Hochberg (1979) introduced more control with artificial reverberation from an unusual technical set-up: using a continuous tape loop, they added delay and attenuation between the output of a tape recorder and its input, which thereby delayed and attenuated each echo by a fixed, manipulable amount. In another study, Nábělek and Robinette (1978) used a single echo and thereby moved from the realm of reverberation time to a variable described in terms of echo delay.

The range of reverberation times obtained from the experimental rooms in these studies has ranged from .3 (Nábělek and Pickett, 1974a and b) to 2.3 sec (Moncur and

Dirks, 1967), with an additional longer time obtained from the Gelfand and Hochberg study of 3 sec. The range of T_{60} most often evaluated is from about .4 to 1.3 sec. Nábělek and Pickett (1974a) suggest that .3-.6 sec is typical for classrooms, with larger rooms having the potential for a larger range of reverberation times. Many studies (Moncur and Dirks, 1967, Finitzo-Hieber and Tillman, 1978, and Newman, 1982) have also included a non-reverberant condition for a performance baseline.

Other main independent variables associated with speech perception in a reverberant environment are (1) whether the listener has normal hearing or is hearing-impaired (Nábělek and Pickett, 1974b, Gelfand and Hochberg, 1976, and Bloom, 1980), (2) whether listening is done binaurally or monaurally (Moncur and Dirks, 1967, Nábělek and Pickett, 1974a and b, Gelfand and Hochberg, 1976, Nábělek and Robinette, 1978, Nábělek and Mason, 1981, and Newman, 1982), (3) whether listening is with or without hearing aids (Nábělek and Pickett, 1974a, and Finitzo-Hieber and Tillman, 1978), and (4) whether there is added noise in addition to reverberation (Moncur and Dirks, 1967, Nábělek and Pickett, 1974a, and Nábělek and Mason, 1981).

The main dependent variable has been some measure of speech perception, most frequently the Modified Rhyme Test (Nábělek and Pickett, 1974a and b, Gelfand and Hochberg, 1976, Nábělek and Robinette, 1978, Gelfand and Silman, 1979, and Nábělek and Mason, 1981). Some studies have also attempted to analyze the effect of reverberation on speech sounds such as consonant type or the location of the consonant in the test word (Nábělek and Pickett, 1974a and b, Gelfand and Silman, 1979, and Newman, 1982). In addition, one study (Allen, et. al., 1979) has used a non-performance measure, preference, to evaluate reverberation perception; this study will be described in section 2.3 which discusses simulated reverberation.

2.2.3 Reverberation and the Binaural Advantage

The binaural advantage is the ability to take advantage of the different inputs to the two ears to decode a signal from background noise. It is the capability displayed in MLD's. A binaural advantage is consistently displayed in reverberant conditions as well as in other noisy environments. Moncur and Dirks (1967) gathered data on normal-hearing subjects who listened binaurally or monaurally with the near or far ear. They tested both quiet and noise (speech babble) conditions for reverberation times from 0 to 2.3 sec. Except for the no reverberation-quiet condition, a small binaural advantage over the near-ear condition was consistently found as well as a constant large effect of performance based on near-ear over far-ear listening. Nábělek and Pickett (1974a and b) also found a binaural advantage of 3-5 dB in all their listening conditions with normal subjects and a smaller binaural advantage of about 1.5-2 dB for most of the conditions with hearing-impaired subjects. This difference in binaural advantage between normal hearing and hearing-impaired listeners holds true when the impairment is noise or babble as well as reverberation (Licklider, 1948, Jerger and Dirks, 1961, Jerger, et. al., 1961, and Dirks and Wilson, 1969).

The binaural advantage interacts with background noise level and the presence/absence of a hearing aid as well as with the subject's hearing ability. As noise or reverberation increases to the point where the input to the far ear deteriorates, the binaural advantage lessens. Moncur and Dirks' data show the binaural advantage in their noise conditions; the difference between binaural and monaural near ear listening dropped from 14.2% to 5.3% as reverberation time went from 0.0 to 2.3 sec.

Nábělek and Pickett (1974a and b) found a binaural advantage in all their listening conditions, but found a smaller binaural advantage for their longer reverberation time (.6 sec) than for their shorter (.3 sec) for their normal hearing subjects. Also, when Nábělek and Pickett (1974a) put a hearing aid on their normal hearing subjects, their binaural advantage dropped from 4-5 dB to 3 dB. They suggest the aid may be suppressing some of the low frequency contribution to binaural gain, and may be providing some temporal distortion. The latter factor is supported by the Finitzo-Hieber and Tillman study which used only monaural listening conditions. They tested hearing-impaired listeners when 'aided' by a loudspeaker or a hearing-aid and found the loudspeaker condition provided a small, but consistent, improvement over the hearing aid.

If noise is added to a reverberant environment, speech perception is impaired additionally by the noise. It is difficult to determine whether there is interaction between T_{60} and noise because of the ceiling and floor effects associated with the response measure, percent intelligibility. When the interaction occurs, its effect is to reduce intelligibility more than the sum of the individual effects of noise and T_{60} would predict (Moncur and Dirks, 1967, Finitzo-Hieber and Tillman, 1978, and Nábělek and Mason, 1981).

The papers by Nábělek and Pickett (1974a and b) are probably among the most significant in this body of literature; they provide as near as is available to a factorial experiment with the major reverberation-relevant variables. They contrast normal and hearing-impaired listeners, the binaural and monaural near ear, and include levels of reverberation and types of background noise. Nonetheless, the range of reverberation levels is limited by the classroom in which they conducted their experiment. This is a general limitation of most studies which have examined reverberation perception.

In their discussion, Nábělek and Pickett regret the lack of data concerning the ability of the impaired ear to integrate temporally since an impaired ear may be integrating less efficiently than a normal ear; this is the temporal analog to the wider auditory filter concept that Patterson, et. al. (1982) were concerned with in their study of frequency selectivity in the ear as a functon of age and is a foreshadow of the need for a theoretical basis for an understanding of perception in a reverberant environment.

2.2.4 Speech Features and Reverberation Effects

To further analyze the effect of reverberation, experimenters have mostly analyzed the structure of the responses to the input speech material. They have found that consonant location is important; initial consonants are perceived correctly more often than final consonants (Nábělek and Pickett, 1974a and b, Gelfand and Silman, 1978). Since initial consonants are preceded by the pause between the carrier phrase and the test word and finals are preceded by and are perhaps part of a vowel, this is a predicted finding. Nábělek and Pickett find and account for a variety of consonant feature effects. Place is most sensitive to impairment by reverberation, and more in the final than initial position. Place perception depends on brief formant transitions, especially easily masked at the end of a word. With increasing reverberation time they found voicing better perceived than manner of articulation (whether the consonant was a stop, fricative, glide or nasal consonant) and manner better perceived than place. They found more binaural gain for manner and voicing than place and attribute these effects to the higher proportion of low frequency energy available for MLD-like processing in manner and voicing.

These results are similar to those of Gelfand and Silman (1978); they found initial consonant better than final consonant perception and found reverberation most affected place, stop and frication. Newman (1982) found place and manner errors were most common and she found few nasalization and voicing errors. These results have a common theme based on reverberant energy as noise; reverberation of vowel sounds is loud and will mask nearby, especially following, low level consonant signals. A second theme is that low frequency energy is used to produce binaural gain, and the upward spread of masking (Egan and Hake, 1950) would also support better perception for consonants with low frequency energy. These studies have analyzed the speech signals to see how reverberation time affects them. A second approach to understanding the effects of reverberation is to analyze the reverberation to determine its perceptual components. We have seen that echoes with short delays appear to be integrated with the signal and enhance it, and that echoes with long delays impair the signal. Additionally, echoes introduce a ripple into the frequency spectrum of a signal and listeners are sensitive to wide ranges of ripple densities.

Almost the only measure of reverberation that has been used in speech perception experiments is reverberation time which amounts to an implicit assumption that it is the temporal effects of reverberation that are important for perception. It isn't that workers in this area are unaware of the complexity of reverberant sound. For example, Nábělek and Pickett (1974a) discuss the importance of early reflections and the integration of delayed speech, and Gelfand and Silman (1978) are concerned about the density of reflections in the tiny, highly reverberant room in which they tested compared to single echo studies such as the one performed by Nábělek and Robinette (1978). Rather, the analytical tools to pursue such an investigation have not been available to easily manipulate the physical variables determining the reverberant sound field.

2.3 SIMULATED REVERBERATION

In 1979, Allen and Berkley developed an image method of calculating the impulse response of a rectangular room from a set of physical room parameters. They used 15 parameters: the room dimensions (3), a wall reflection coefficient, β, for each wall (6), the location of the source (talker) in the room (3), and the location of the receiver (listener) in the room (3). From the room impulse response, $h(t)$, its equivalent, $H(f)$, is calculable, and so also is the room reverberation time, T_{60}. Convolving the impulse response with a signal (speech, noise) yields a simulation of room reverberation in the simulated room on the signal. This program is the basis for stimulus generation for all the experiments to be reported in section 3. The program, named ROOM, will be described more fully in section 3.1. Another room reverberation measure calculated by Allen and Berkley is spectral variability (Jetzt, 1977), measured in dB as the standard deviation, $\sigma_{L(f)}$, of the room frequency response; $\sigma_{L(f)}$ will be termed σ in this paper for easy reference:

$$L(f) = \text{SPL of } H(f) = 20\log_{10}[H(f)]$$

$$\sigma_{L(f)} = \left[\frac{\sum_f [L(f) - \bar{L}(f)]^2}{f_{max} - f_{min}}\right]^{1/2}$$

This gives us a frequency domain measure of reverberant energy, and, while T_{60} and σ both increase with increasing reverberant energy in a room, i.e., with β, they can be manipulated independently using other room parameters in order to explore their importance for perception. In particular, the distance between the source and the receiver, Δ, will not affect reverberation time, the rate of decay of reverberant energy, but it will affect the value of σ. Increasing Δ will lower the level of the direct sound relative to the reverberant sound and thereby increase σ. Also of potential perceptual importance is the size of the room; evidence cited earlier (Knudsen, 1929) shows reverberation time increasing with room size. Since the earliest echoes will be the largest, short distances between the talker and a wall may create special perceptual effects.

As can be seen this approach already enriches the consideration of reverberation. From a single measurement, T_{60}, we move to a manipulable set of physical and derived room statistics with potential perceptual relevance.

Allen, McDermott, and Berkley (1979, see also Berkley, 1978) examined some of these room variables. They sought to evaluate the perceptual effect of σ and T_{60} and to determine what physical parameters are perceptually important for reverberant speech. Holding room size constant at 12.5 x 15.0 x 16.25 feet, they varied the physical room parameters, β and Δ, to yield rooms in which σ and T_{60} were not perfectly correlated. Table 2.2 presents their experimental

Table 2.2 Experimental Design of Allen, et. al. Study

	β T_{60} (ms)	.38 75	.54 124	.70 175	.79 263	.86 525
	.63	.32 (1)		.77 (6)	1.05 (10)	1.46 (13)
Δ	1.25					3.55 (14)
	2.50	1.27 (2)		3.74 (7)	4.75 (11)	5.24 (15)
(feet)	5.00	2.64 (3)		5.57 (8)	5.78 (12)	5.59 (16)
	10.00	4.25 (4)	5.08 (5)	5.59 (9)		

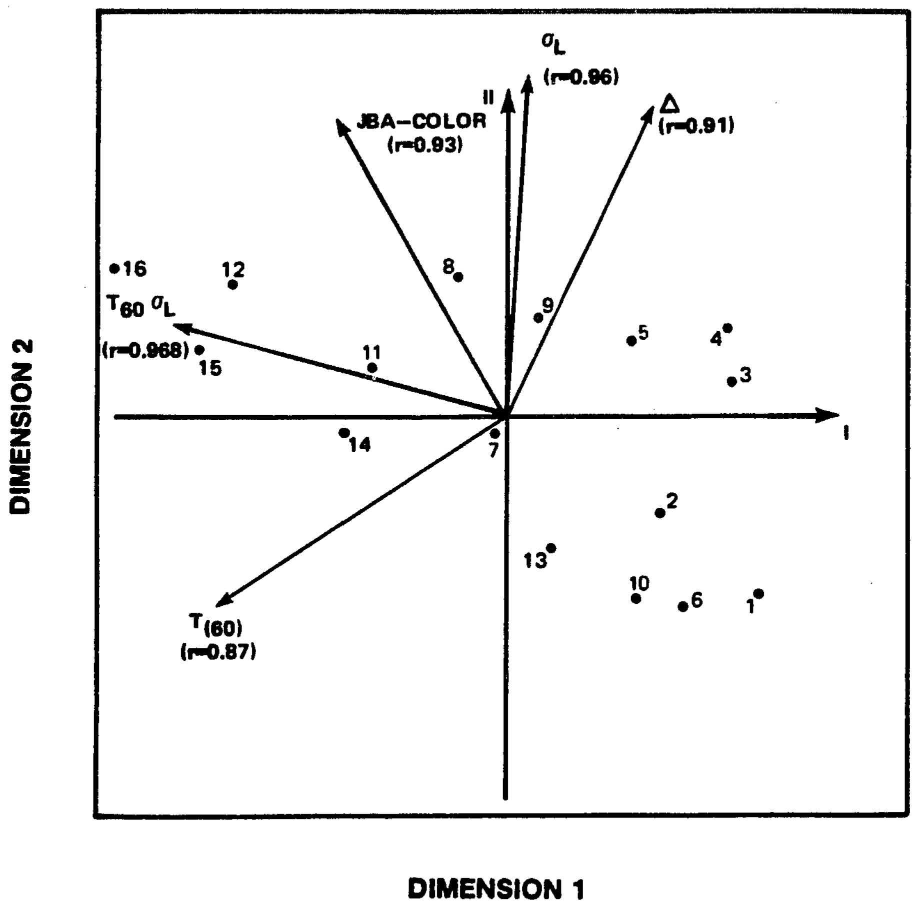
design. The column headings are the input values of β and the resulting values of T_{60}, and the row headings are the values of Δ, from .63 to 10.0 feet. The entries in the body of the table are the values of σ for that room, and, under the σ value, the stimulus number. This was not a factorial experiment; only the rooms for which there is a value of σ in Table 2.2 were actually used, a total of 16 rooms. Ten Phonetically Balanced sentences convolved with h(t) in the reverberant room program were the reverberant speech stimuli. Their subjects listened to all pairs of the room conditions and judged, on a 10-point scale, how different each pair was.

Results across subjects were averaged and a multidimensional scaling analysis was performed on them. A section on multidimensional scaling is included later in this review; for now let it be said that the difference judgments are used to create a "space" of pre-specified dimensions in which the stimuli, i. e., the room conditions, are located so that distances between the stimuli reflect the size of the difference judgments.

The results supported a two-dimensional solution as representing a practical compromise between stress (an error measure) and parsimony. Their two-dimensional scaling solution is presented in Figure 2.1. Based on these data, this solution represents the perceptual range of reverberant speech. A two-dimensional scaling solution suggests that two variables are needed to account for the data, and they were sought by regressing various reverberant room variables into the multidimensional solution, i.e., by doing a multiple regression analysis of each room variable with the coordinates of the stimulus points in the two-dimensional space. Figure 2.1 also includes the best fitting regression in the form of a vector, of some of the room variables explored. σ and σT_{60} both regressed well into the space ($r = .96$ and and .97, respectively) and were nearly orthogonal.

FIGURE 2.1

TWO-DIMENSIONAL SOLUTION FOR ALLEN, ET. AL. REVERBERANT SPEECH EXPERIMENT



The authors found these physical qualities were related to subjective qualities by an experienced listener (one of the authors, JBA) who rated the individual room conditions for "echo" and "coloration". The coloration judgments were most highly correlated with σ and the echo judgments created a vector which fell between the σT_{60} and the T_{60} vector. The coloration and echo judgments were nearly orthogonal when regressed into the multidimensional space, thereby making the two of them, in combination, a good predictor of the space. The results gave rise to the hypothesis that the perception of reverberant speech could be defined by the perceptual variables, coloration and echo, and that σ and T_{60} (or σT_{60}) are variables derived from physical room parameters which can be used to characterize the perceptual dimensions.

A second experiment was done in which subjects were asked for their listening preference on all pairs of rooms. These preference judgments resulted in a one-dimensional solution; the authors found that preference increased as both the perception of coloration and echo decreased. With low values of both coloration and echo, the preference function appeared to flatten out; whether this is a ceiling effect on the response scale or whether small amounts of coloration and echo are actually preferred by listeners is not known at this time. In any case, this study was a diotic listening experiment, and the preference judgments may not be entirely valid for dichotic listening. In general these preference data are consistent with speech perception experiments: the less reverberation, the better.

The experiments just reported are the immediate basis for the studies to be reported in this paper. These studies will seek, first, to validate the physical variables underlying coloration perception and, second, to determine the range of the subjective functions. The first of these tasks is best done using multidimensional scaling techniques, and the second with psychophysical scaling methods. A multidimensional scaling analysis may show us important variables for perception, but it does not fully determine the psychophysical scale associated with the variable or inform as to the range of the scale. Once a perceptually important measure is determined, we must turn to the more traditional methods of psychophysical scaling to help us put our measures into context. The studies to be reported will include both multidimensional scaling and psychophysical scaling analyses. These areas will be very briefly reviewed.

2.4 PSYCHOPHYSICAL SCALING

The psychophysical methods are much older, especially in the area of audition, and we will start there. The problem is an old one: to determine the function relating the growth of subjective magnitude to an increase in stimulus magnitude. The two most striking auditory stimulus continua are sound pressure and frequency, and, for a tone, to a great extent, sound pressure level determines its loudness and frequency its pitch. (Stevens, 1934, showed loudness decreasing slightly as frequency increases and pitch increasing slightly with increasing intensity, but these are very much second order effects.)

The first auditory psychophysical scale was stated by Fechner in 1877 for loudness; he integrated Weber's law, which states that the subjective size of a stimulus jud is constant ($\Delta x/x = k$), to yield a logarithmic law for the growth of loudness. Following Luce and Edwards, 1958, Weber's law may be rewritten as $\Delta u/\Delta x = A/x$ where Δu is a sensation jnd; Fechner rewrote $\Delta u/\Delta x$ as du/dx and integrated the resulting differential equation: $du/dx = \int A/x$, and integration produces $A\log x + b$. Fechner's law has been attacked on various grounds, e.g., (1) Weber's law isn't true, and it isn't, quite, (Jesteadt, et. al., 1977, and Luce, 1977). The Weber fraction has been shown to decrease slowly with frequency, from about .4 at low frequencies, to about .1 at high frequencies. (2) Fechner's math was wrong, and it was wrong, (Luce and Edwards, 1958) although Krantz (1971) reformulated the problem and established Fechner's law from Weber's law using considerably more elaborate mathematics than did Fechner.

The most controversial dispute concerning Fechner's law, however, has been its assumption that a unit of discrimination, of variability, is the appropriate unit for psychological magnitude. This issue has affected the methodology preferred by experimenters as well as arguments concerning the nature of the psychophysical function. Workers who accept variability as an appropriate unit (Garner, 1954, and Thurstone, 1927a and b, and Torgerson, 1958, on Thurstone) have worked with methods such as categorical scaling, jnd scaling, and paired comparison methods which are heavily influenced by or based on subjects' abilities to discriminate between the stimuli.

In opposition to this approach are those who prefer direct methods such as magnitude and ratio estimation and production. Stevens (1957, and 1971, and Stevens and

Galanter, 1957) has been a strong proponent of this school of thinking which claims that using a measure of discriminability as a unit of subjective magnitude is simply irrational, and that to learn subjective magnitude, one should ask subjects about it directly.

In 1955, Stevens summarized a large body of experiments based on direct methods and offered a power law with a .3 exponent as the psychological scale for loudness, $L = kI^{.3}$ where L is loudness and I is stimulus intensity. In addition to summarizing a large body of data, this law has a simplicity which equals that of Fechner's, and the practical advantage that adding 10 dB doubles loudness. The representation of loudness as a power function of intensity has become a component of an international standard (ISO R532) for a simple method of calculating the loudnes of a steady state sound.

In support of the direct methods, Stevens (1957) presented a summary of power law exponents for a variety of psychological continua. The exponents for continua such as loudness and brightness were around .3, the exponent for visual length and duration were cited at about 1.0, and the exponent for auditory flutter and visual velocity were about 1.7. A generalization that can be drawn from such exponents is that continua which cover many orders of magnitude have low exponents and, as the range of a physical continuum decreases, the exponent associated with the psychophysical continuum increases (Poulton, 1968). This suggests that the power law for a subjective continuum depends on how people use numbers, i.e., that a certain range of numbers will be used to describe the psychophysical range of a physical continuum. This is an interesting generalization which links together many perceptual continua, albeit with a non-perceptual link. For loudness, Zwislocki (1978) has suggested an even stronger generalization, that the link between numbers and the subjective continuum is "absolute."

Neither class of methods, direct or categorical, is free from methodological bias. Categorical scaling methods are very susceptible to context effects such as stimulus presentation order and the number of scaling categories (Stevens, 1957). Recent work (Green, Luce, and Smith, 1980) shows the direct methods to be susceptible to context effects also. In particular, the smaller the difference between stimuli, the smaller the response variability. Also, the direct methods have always been noted for the inter-subject variability displayed. The stability of the results is based on the use of the median which is less susceptible than the mean to data extremities.

A bias-free method for evaluating loudness does not appear to exist. Nevertheless, direct methods yield power laws for a large number of psychological continua and have proved useful as well as convenient in application. This has not eliminated discrimination-based scales from use. If the fineness of discrimination on a continuum is not a good measure of magnitude, it is still fundamental to the measurement of threshold and an indicator of the range of a continuum. Both types of methods are still widely used in psychophysical studies.

We have discussed the psychophysical function for loudness and have so far ignored that other outstanding auditory dimension, pitch. Stevens (1957) distinguishes between two types of stimulus continua; prothetic continua, such as loudness, are based on magnitude or quantity at the physiological level, and the growth of subjective magnitude on a prothetic continuum is best characterized by a power function. Other examples of prothetic continua are brightness, heaviness, saturation of hue (grayness), and duration. On the other hand, metathetic continua, such as pitch, are represented by place at the physiological level and different "levels" of stimulation produce a substitution rather than an increase in excitation. Other examples of metathetic continua are visual place and inclination and, probably, hue and different kinds (rather than intensities) of taste and smell (Stevens and Galanter, 1957). Viewing pitch as physiological substitution suggests that it doesn't make sense to apply a direct method to pitch scaling; "produce a pitch twice as high as the standard" does not produce twice as much pitch. Stevens and Volkmann (1940) had subjects divide the freqency range into intervals of equal pitch to develop a psychophysical function for pitch, the mel scale. As discussed earlier, Zwicker, et. al. (1957) show it in good agreement with critical band and pitch discrimination data.

Nonetheless, the mel scale has not resulted in widespread useful applications, probably due to its divergence from musical pitch. Ward (1970) points out pitch scaling results in within subject differences which are large, and much larger than judgments about pitch related to musical pitch. Stevens and Volkmann cited a 9.9% within subjects error for equal interval judgments while Ward mentions a 0.6% error for determining an "octave above 1000 Hz."

The preceding discussion of auditory psychophysical scaling has emphasized loudness over pitch because most of the important scaling issues have revolved about loudness. Let us step back and put pitch (and frequency selectivity) and loudness into some relative perspective. At a rough estimate there are about 10 times as many jnd's for pitch (around 1200-1500 based on gross extrapolations from Harris, 1952 and Stevens and Volkmann, 1940) as there are for loudness (around 120 based on an extrapolation from Green's 1976 summary of intensity discrimination). For all languages frequency selectivity and duration are the major bases for speech production and perception. In terms of auditory processing, frequency selectivity is fundamental and loudness is mostly important for getting the frequencies to the internal processor.

2.5 MULTIDIMENSIONAL SCALING

Psychophysical scales such as sones for loudness and mels for pitch relate a single physical and psychological continua. If both frequency and sound pressure level are varied, a single psychological continuum is not sufficient to describe a listener's response; the response would be multidimensional, consisting of the perception of both pitch and loudness. If a set of sinusoids varying in pitch and loudness were presented to listeners in pairs, and the listeners asked to judge the size of differences between them, we would expect the difference to reflect both the difference in pitch and in loudness, and would expect to be able to account for the response using the independent variables of frequency and sound pressure level. This would not be an especially interesting experiment to do in the case where we believe we already know the underlying physical parameters which would determine the responses. In a case where it isn't known what underlying variables are important, this kind of experiment may provide important insights.

Multidimensional scaling provides an analytical structure for creating a multidimensional space in which the distances between points (stimuli) in the space agree as well as possible with the difference (or similarity) measurements which are the input data. The output of a multidimensional scaling solution with n dimensions is a coordinate for each stimulus on each of the n dimensions. Each stimulus is thereby represented as a point in the multidimensional space. Distances between the points can be calculated and when the stimulus coordinates are plotted on rectangular coordinates, a graphical representation of the multidimensional scaling solution is displayed.

An interpretation of the dimensions of the space is sought in terms of underlying physical or psychological measures. In general, a successful multidimensional scaling analysis results in a scaling solution which provides a good fit to the difference judgments. The multidimensional scaling analysis itself does not provide an interpretation of the dimensions of the space; these must be determined, usually through the use of regression techniques, by the experimenter. The goals of multidimensional scaling are quite different from the goal of psychophysical scaling; the latter goal is to determine a (the) function relating a known physical and perceptual continuum. Rather, it structures the data in a way that makes it easy to explore how different variables may account for the data, and how these variables are related to one another.

Multidimensional scaling tells us nothing about the range of a perceptual continuum and does not usually provide a well defined psychophysical function. It may be based on the assumption that the distance judgments are ordinal (Kruskal, 1964a and b) or interval (Carroll and Chang, 1970, and Carroll and Wish, 1974). The most commonly used multidimensional scaling methods are based on only ordinal information which provides no firm theoretical ground for determining a psychophysical function. In practice, a large set of ordinal judgments will supply considerable interval information (Coombs, 1964) just as the product-moment correlation coefficient will agree well with the rank order correlation coefficient for most large datasets.

The use of only ordinal information is both a strength and a weakness. Since categorical and direct scaling methods result in scales that are monotonically related, the use of only rank order information obviates objections concerning the categorical nature of difference judgments. On the other hand, the resulting scaling solutions are ordinal solutions. Distances in the space, although they may be (and will be here) used for convenience in exploring the solution and potential underlying variables, are not interval scale distances. They are commonly treated as though they were interval (and will be in this paper), but further substantiation with other methodology is needed for a psychological scale based on a non-metric multidimensional scaling solution. This also introduces constraints with respect to describing the variability in the solution.

The non-metric approach to multidimensional scaling requires only weak assumptions about the data and, in consequence, permits only relatively weak statements about the solution. This is a standard trade-off for data analysis methods; payment for analytical power must be made with increasingly astringent assumptions about the data. Other multidimensional scaling methods (INDSCAL) will be briefly discussed which make much stronger assumptions about the difference data. An experimenter's choice of method will reflect both the nature of the data and the goal of the experiment.

Let us look in more detail at the concepts on which multidimensional scaling is based. A major contribution in this area was made by Roger Shepard (1962a and b); these papers introduced the idea of a monotone relationship between the judgments of differences and the distances in a multidimensional space. Also spelled out is the need for a solution with minimum necessary dimensionality. Shepard and others have used an iterative procedure which obtains successively better approximations to monotonicity as stimulus points in the multidimensional space are moved from some random starting configuration. Shepard demonstrated the validity of his procedures by taking distances for fixed configurations and showing that he could recover them. While it was apparent that the quality of the multidimensional solution depended on how close a monotone function could be approximated, it was Kruskal (1964a and b) who applied the concept of goodness-of-fit to the monotone multidimensional scaling function and developed a measure, termed stress, to measure it.

Kruskal calculates a least squares monotonic regression function between the difference judgments and the distances in a scaling solution, and the deviation of the distances from this function is the basis for stress determination. Using the representation of Kruskal and Wish (1978), the stress equation is:

$$S = \left[\frac{\sum_{ij} [f(\delta_{ij}) - d_{ij}]^2}{\sum_{ij} d_{ij}^2} \right]^{1/2}$$

where d_{ij} is the distance between stimuli i and j in the scaling solution, δ_{ij} is the difference judgment for stimuli i and j, and $f(\delta_{ij})$ is the least squares monotonic regression function for δ_{ij}. Thus stress is small when the deviations of the distances from the monotone regression function are small, and is large when they are large. Stress is calculated for each iteration of a scaling analysis and the main criterion for determining when a solution has been found is when stress changes only a little, or not at all, from one iteration to the next. The determination of whether the final value for stress represents a good or bad scaling solution is left to the judgment of the experimenter, for there are no significance tests to apply. Kruskal (1964a) suggests that S = .10 is "Fair", S = .05 is "Good", and S = .025 is

"Excellent", and these terms are clearly as ordinal as the scaling assumptions.

The next question that must be considered is dimensionality; a scaling solution that provides a perfect monotone relationship can always be provided by a space with one fewer dimensions than the number of stimuli. A parsimonius solution, however, would provide a solution with as many dimensions as there are independent underlying variables influencing the subjects' difference judgments. In general multidimensional scaling techniques fix the dimensionality of the scaling solution and then solve for minimum stress; experimenters will usually determine a scaling solution for one to five or six dimensions and seek to determine the "true" dimensionality from the stress function and from outside factors such as dimension interpretability.

The stress function is stress plotted as a function of dimensionality. As dimensionality increases, stress will decrease. An elbow in the stress function is expected at the true dimensionality. The idea is that a dimension based on a variable underlying the responses will reduce stress by considerably more than a dimension based on data error reduction. Thus stress should decrease sharply as a function of dimensionality as long as the dimensions represent true response continua and stress should decrease much more slowly as only error variance is accounted for--thereby producing an elbow in the stress function. Kruskal (1964a) and Kruskal and Wish (1978) discuss this issue and cite data, some of which support the stress elbow and some of which are much more ambiguous. Dimensionality remains, finally, a matter of judgment.

A final consideration, the method for solving each iteration, has generally been based on established numerical procedures. Kruskal's work (1964a and b) has served as a basis for a computer program, termed KYST-2 (for the initials of Kruskal, Young, Shepard and Torgerson, founders of the concepts and techniques incorporated) (1973), which has been developed to analyze scaling data to provide multidimensional solutions. (This is the program used to analyze the data to be reported in this paper).

This is not the only important direction multidimensional scaling has taken; workers who have assumed difference judgments provide interval data (Carroll and Chang, 1970, Carroll and Wish, 1974) have created methods which result in interval scaling solutions with weights for subjects which indicate the relative degree to which each dimension determines their judgments. This technique is termed INDSCAL; and it also has received wide application. It assumes the dimensions of the solution space are the true dimensions in order to provide a rationale for the subject weighting procedure. This is in contrast to scaling methods which don't assign weights to subjects. For scaling methods other than INDSCAL, it is common to rotate the axes of a scaling solution to align them with variables which seem to represent the underlying dimensions.

Once a scaling solution has been obtained, the underlying variables which account for the solution are yet to be determined. Linear regression techniques are the most prevalent way of approaching this problem. These are used by performing a multiple linear regression betwen the coordinates of the stimuli in the multidimensional space and the values associated with the stimuli on a continuum. In our earlier example with difference judgments made on sinusoids varying in frequency and sound pressure level, we would correlate the coordinates of the stimuli in the solution space (probably two-dimensional) with the frequency and with the sound pressure level of each stimulus.

The correlation coefficient indicates how well a variable accounts for a dimension of the solution, i.e., regresses into the multidimensional space. This regression function may be projected into the multidimensional space to permit seeking a graphic interpretation of the variable in the solution. As would be expected, two underlying variables which are 90° apart in their projections in the space are independent and account for different dimensions of the solution. (In Figure 2.1 of section 2, the projections of σT_{60} and σ into the two dimensional solution space are nearly at right angles to each other and thus these vectors represent different dimensions of the solution. They are close to, but not an exact match with, Dimensions 1 and 2 of the space, respectively.)

Ideally for a scaling solution of n dimensions, n independent underlying variables would be determined to account for the scaling solution, and each would achieve a correlation of 1.00 within the space. In practice, of course, both the level of correlation and independence of the vectors for the variables are less than ideal. Nonetheless, such regression techniques have proved to be by far the strongest tool for interpreting the underlying dimensionality of multidimensional scaling solutions.

2.6 RATIONALE FOR EXPERIMENTS

The goal of the experiments to be reported here is to explore the frequency domain aspects of reverberant sound. On the basis of the Allen, McDermott and Berkley (1979) work, a perceptual variable of reverberant rooms, coloration, appeared to be defined by a measure of spectral variability. Validating the link between spectral variability and a perceptual attribute termed coloration, and establishing a psychophysical scale of coloration, are some of the next steps in understanding the perception of reverberation.

The beginning step was to separate the perception of coloration and echo so that coloration could be studied alone. Berkley suggested this might be done using reverberant noise stimuli created using the Allen-Berkley room simulation program. Such stimuli would have temporal continuity with echo perceptible only at the beginning and end of each stimulus. By then appropriately editing the leading and trailing edges of the noise, even these echo components could be eliminated, and the only remaining perception should be that of coloration. These assumptions about reverberation perception rest on coloration being derived from frequency domain variations. The hypothesis was then that a replication of the Allen, et. al. experiment using these noise stimuli should now yield a single dimensional solution and that this dimension should correlate very highly with σ. With this as a start, more standard psychophysical methods could be used in succeeding experiments to try to relate σ and coloration.

To anticipate a little, the results of this study were considerably more complex than had been expected. The results led to another multidimensional study in which a variety of measures of spectral variability were explored. In addition, several physical parameters of rooms were investigated, among them room size, the movement of the source and receiver in a room while keeping Δ fixed, and the distance of the source and receiver from a wall.

A third, categorical scaling experiment supported the σ variant as a correlate of coloration. Finally, the range of coloration jnd's for a given room, varying only in reflectivity, was determined experimentally for two simulated rooms in order to lend perspective to coloration as a psychophysical phenomenon. The final experiment permitted the determination of the number of jnd's on a continuum representing quantity of coloration.

3. EXPERIMENTS

3.1 GENERAL METHODOLGY

Four experiments were run. The first two studies were multidimensional experiments in which subjects made difference judgments on all pairs of the stimulus matrix. A fairly large number of naive subjects were used to obtain stable data. The third and fourth studies used more traditional psychophysical methods which required repeated observations by relatively few subjects as their means to stable data. Data from the third study consisted of absolute judgments and paired comparisons, and they were analyzed using only regression techniques. The fourth study was a more extensive paired comparison experiment; it required more elaborate analytical procedures: Thurstonian scaling using a maximum likelihood scaling solution followed by curve fitting with second order polynomials.

In spite of these analytical differences, there were many similarities among the experiments. The stimuli in all the experiments were noise stimuli produced in the same way: white noise bursts one second long were convolved with simulated rooms created by the Allen-Berkley ROOM program. Also, the experiments were all run in an IAC sound isolation room and there were many commonalities in the experimental set-ups and procedures.

These commonalities between experiments will be described here.

3.1.1 Stimulus Production

The procedures followed for creating the stimuli and the analog tapes used to present them were similar for all the experiments. A Data General Eclipse S/200 series computer with system and user disks was used in each step of the process. Peripheral units included a terminal with a CRT display and horizontal and vertical cursors, a magnetic tape unit, and an A/D and D/A converter. Figure 3.1 presents a schematic of part of the computer laboratory.

Figure 3.2 outlines the main steps of the procedure used to produce the stimuli. A frozen sample of white noise was the basis for all the reverberant noise samples in each experiment; the purpose of this was to insure that the only difference between stimuli would be due to the effects of the simulated room. The noise sample was input to the A/D converter from a General Radio 1382 noise generator. The A/D output was sampled at a rate of 8000 samples/sec and stored on the user disk. This sampling

FIGURE 3.1

SCHEMATIC OF COMPUTER LABORATORY

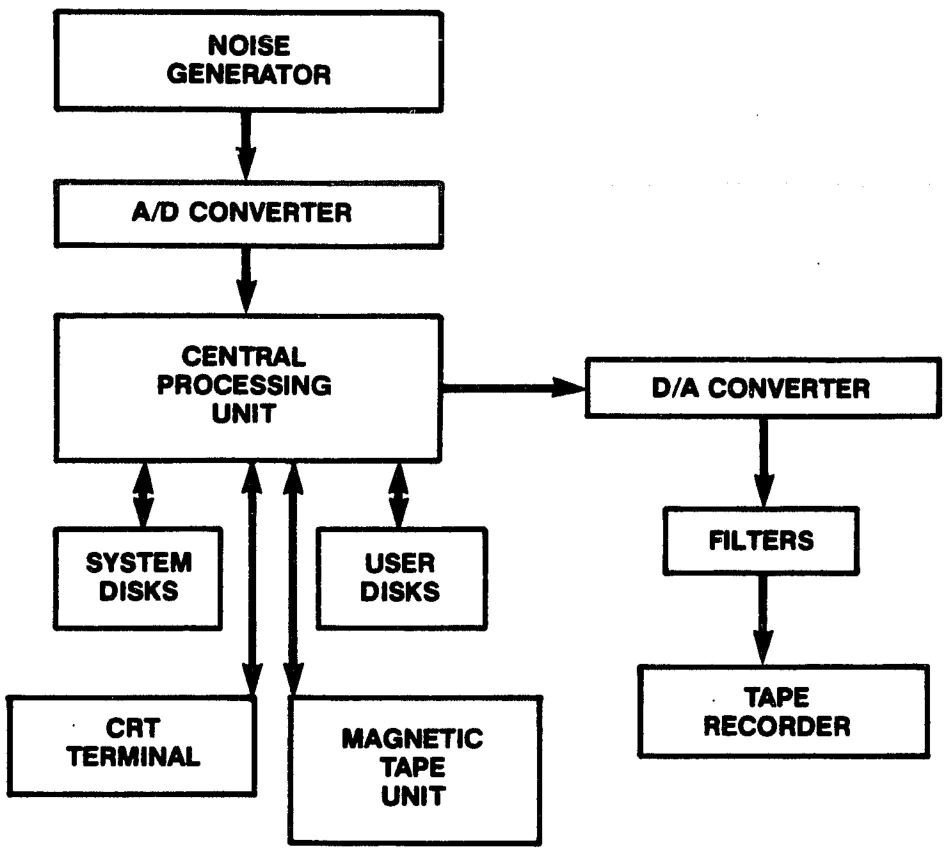
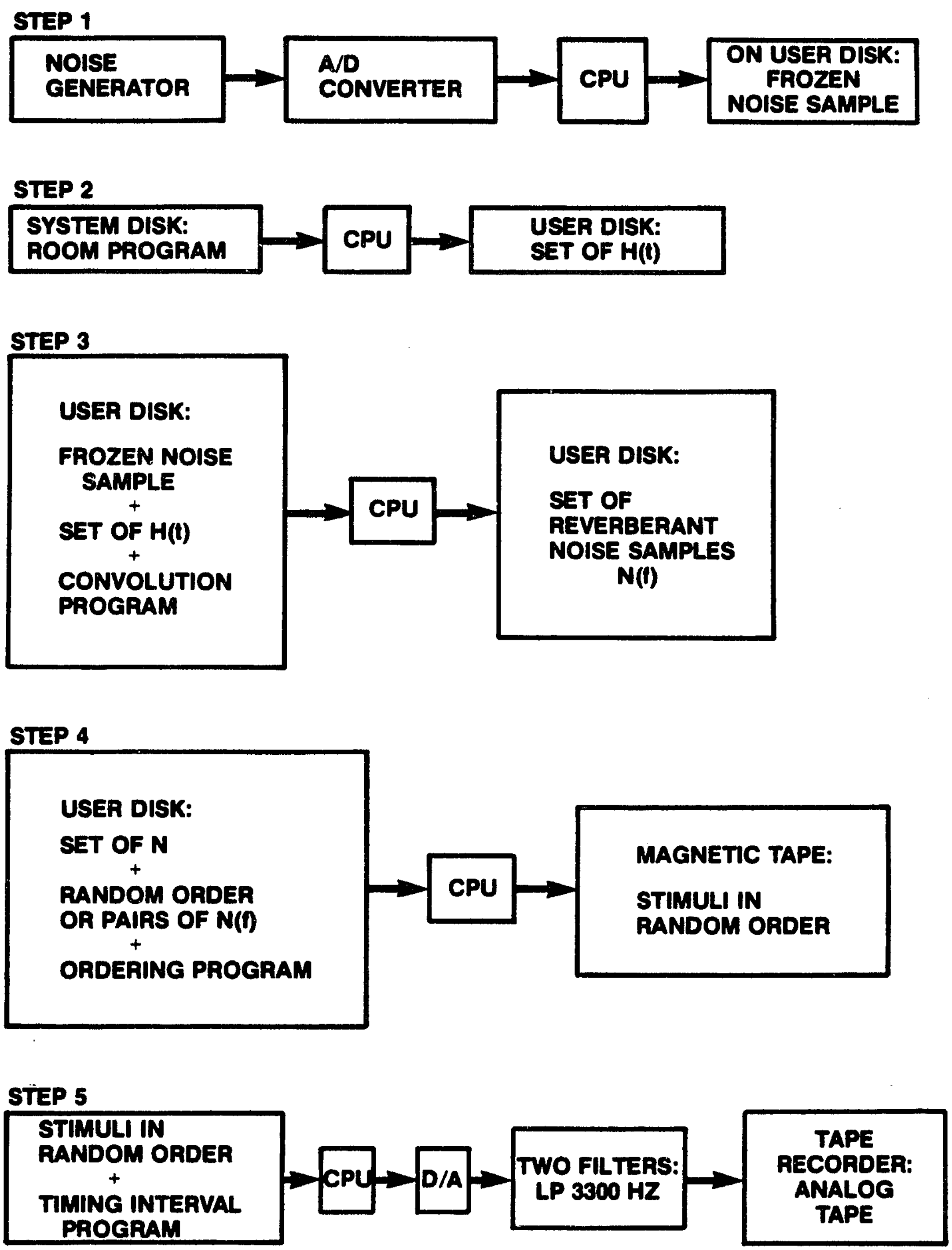


FIGURE 3.2

STIMULUS PRODUCTION PROCESS



rate limits the spectral range to frequencies under 4000 Hz. The frequency range of concern in all of these studies is primarily that most used in telephony, and its upper limit is about 3300 Hz. An anti-aliasing filter was not used on the input noise, because the aliased frequency bands are also Gaussian and adding the reflected Gaussian bands into the below 4000 Hz range will maintain the Gaussian quality of the noise. The output stimuli were filtered.

The room impulse responses were simulated using the Allen-Berkley (1979) computer program, ROOM, rather than using real room responses. An overview of the program's features was given in Section 2.3.

ROOM calculates the impulse response, a time domain calculation, using an image method. The time and amplitude of each reflection in a fixed time period (or of each image in a sphere centered about the receiver) is calculated from a set of 15 parameters which describe a rectangular room. The fixed time period used for the calculations was 512 msec; i.e., all reflections within 512 msec after the source impulse are calculated.

The room parameters are the room dimensions, the location of the talker and listener in the room, and a wall reflectivity coefficient, β, for each wall. In all the experiments reported here, a single value for β was used for a given room as a simplifying step. In the program, the assumptions are made that β is independent of frequency, and angle of incidence, and that there is no scattering of energy associated with reflection. These assumptions violate physical reality, although not greatly in the frequency bandwidth used here, and were selected by Allen and Berkley as useful constraints to keep computational time reasonably low. Each impulse response used in these studies took about 8 minutes to compute. (Computational time depends on room size; small rooms will yield more reflections in a fixed time and will therefore require a longer computation time.)

After the ROOM program calculates the room impulse response, h(t), it high pass filters at 80 Hz (to remove a dc anomaly due to the assumed closed room with no pressure release), performs a fast Fourier transform and calculates the room frequency response, H(f). ROOM also permits the calculation of the reverberation time using a display created from the backward integration of h(t) (Schroeder, 1965) and the terminal's cursor features.

To create the stimuli, H(f) was calculated from h(t) for each h(t), and then used as a filter for a frozen noise sample. An overlap-add method of convolution was used to create the reverberant noise samples (Allen and Berkley, 1979). The build-up and decay of the reverberant energy in the samples were cut from the stimuli and uniform 32 msec ramps were used to smooth their onset and offset. The resulting stimuli were each 1 sec in duration. These stimuli, the output of Step 3 in Figure 3.2, vary in their frequency response, and our goal is to determine if a frequency domain measure, σ, is a sufficient descriptor of the perceptual response to these stimuli. A final step was to equalize the loudness for the stimuli by normalizing them to have equal SPL's to the nearest tenth of a dB.

Experiments 1 and 2 and part of Experiment 3 presented pairs of reverberant noise samples on each trial and got a response from the subjects indicating the degree of the difference between them. In these experiments the stimulus set consisted of either the whole set or a subset of all possible pairs of noise samples. In the one experiment which used single stimuli, the stimulus set simply consisted of all the reverberant noise samples. In all cases randomized orderings of the stimulus pairs were used, and, where more than one randomization was used, a complete stimulus set was presented before another was started. Both orders of a stimulus pair were always included in the stimulus set.

The set of reverberant noise samples and the desired number of random orders of the stimulus set were used as input to a program which put the stimulus orders onto a magnetic tape. The final step in the preparation of the analog tape was a program which inserted the appropriate interstimulus and intertrial intervals into the stimulus orders and put this sequence out through the D/A converter. The D/A output went through two Rockland model 1024F filters, each with 48 dB/octave slopes, both low pass filtered at 3300 Hz, into an Ampex MR 70 tape recorder. After the first experiment, modifications were made to the programs which permitted combining steps 4 and 5 (see Figure 3.2) and creating analog tapes directly instead of using magnetic tapes as an intermediary stage. The primary advantage afforded by the magnetic tapes was to avoid interruptions on the analog tapes produced by time-sharing the computer system with other users. This same advantage was gained by making the analog tapes in time periods when no other computer use occurred.

3.1.2 Experimental Set-up and Procedures

All the experiments were run in an IAC sound isolation room in which 6 subjects could be seated and run simultaneously. The analog stimulus tapes were played on an Ampex tape recorder whose output was amplified by a Crown D-40 amplifier. Six Yamaha YH2 headsets were connected in parallel across the amplifier output. The stimulus sound pressure level out of the headsets was 82 dB SPL (about 50 dB spectrum level); this calibration was performed using a Bruel and Kjaer sound level meter with a flat plate 6 cc coupler designed at Bell Laboratories for measuring sound output from the Yamaha headsets. Daily calibration was performed at the amplifier output using a dBmO meter. The background noise level was 47 dB below the stimulus level.

As stated previously, stimulus duration was one second; when two stimuli were presented on a trial, the interstimulus interval was .5 sec. The time between trials, in which the subjects responded by writing a number on an answer sheet, was 3 sec. In order to help subjects keep their place on the answer sheets with respect to the stimulus sequence, there was an additional 5 sec interval in the response interval every 10 trials. Every 80 trials (8 minutes), there was a short break of 1-3 minutes. Most of the test sessions were 1-1 1/2 hours in duration, including 2 or more longer breaks. The purpose of moderately loud stimuli and frequent rest periods was the reduction of subject fatigue and inattention.

An experimental session went as follows: a group of up to 6 subjects read a set of intructions and then went through a set of practice trials until they reported feeling comfortable with the experiment. (The practice procedure was more formalized in Experiment 4; this will be discussed in that section.) Both practice and test trials consisted of listening to a pair (except for part of Experiment 3 which used single stimuli) of reverberant noise samples.

3.2 EXPERIMENT 1

The experimental hypothesis for Experiment 1 is that if the perceptual effects of echo (associated with T_{60} or σT_{60}) are removed from the stimuli, then only the effect of coloration will be perceived. A second, stronger hypothesis is that the multidimensional solution space will consist of one dimension, that of coloration.

The key to testing these hypotheses is the elimination of the perception of echo, the second perceptual dimension of reverberant speech. The method of echo elimination was described briefly in section 2.2; since this is an important point, let us consider it in a little more detail. The assumption is that echo can be eliminated by using reverberant stimuli without the usual leading and trailing edges in which reverberant energy rises and falls away. At D. A. Berkley's suggestion, such stimuli were created by, first, using noise stimuli for temporal continuity, and, second, editing digitally the noise stimuli so that the leading and trailing edges were removed and replaced with uniform ramps. The resulting stimuli will be noise stimuli of equal SPL's and durations which will differ only in their frequency response. Thus the standard deviation of the frequency response, σ, the physical variable used previously (Allen, et. al., 1979) to describe frequency response differences and account for their perceptual effect, would be expected to do so again.

3.2.1 Experimental Design.

Experiment 1 was a near replica of the Allen, et. al. experiment, except that the stimuli were composed of colored noise rather than reverberant speech. Table 3.1, on the pattern of Table 2.2, shows the characteristics of the stimuli in Experiment 1. Subjects listened to all pairs of the stimuli and judged the size of the difference within each pair on a 10-point difference scale (0-9).

Table 3.1 Experimental Design of Experiment 1. For each Δ, the First Row Contains the σ of the Stimulus, and the Second Row Contains the Stimulus Number.

	β T_{60} (ms)	.38 75	.54 124	.70 175	.79 263	.86 525
	.63	.32 (1)		.77 (6)	1.05 (10)	1.46 (13)
Δ	1.25		1.21 (17)		2.81 (18)	3.55 (14)
	2.50	1.27 (2)		3.74 (7)	4.75 (11)	5.24 (15)
(feet)	5.00	2.64 (3)		5.57 (8)	5.78 (12)	5.59 (16)
	10.00	4.25 (4)	5.08 (5)	5.59 (9)		

The same room parameters used by Allen, et. al. were used to simulate the stimulus "rooms." The earlier study used 16 rooms chosen from a 5x5 matrix of stimuli varying in β, or T_{60}, and in talker-listener distance (Δ). The room dimensions were the same for all the stimuli. (See Table 2.2 for their experimental design.) This study used the same 16 rooms, plus two more from the same matrix (Stimuli 17 and 18), in order to have at least two rooms for each β and Δ.

3.2.2 Experiment 1 Methods

Procedures. For these 18 rooms, h(t) and H(f) were calculated using the Allen-Berkley ROOM program. Two white noise samples were recorded and each was filtered by the 18 H(f)'s, and the build-up and decay of reverberant energy was removed and replaced with uniform 32 msec ramps. The set of stimuli associated with each noise sample was kept separate and comparisons were made only within each set. With the exception of comparing each stimulus with itself, all pairs of stimuli were presented to the subjects, including both orders of presentation for each pair. This resulted in 306 pairs for each noise sample.

Subjects were run in two sessions on two consecutive days. Prior to each session the subjects listened to a brief practice tape once or twice which consisted of stimulus pairs, and which included at least one sample of each stimulus. During the practice session, the subjects were told they could respond with practice difference judgments, but these responses were not required and were not used in the data analyses.

Subjects. These subjects were drawn from the Murray Hill Bell Laboratories subject pool, composed mostly of people related to or acquainted with BTL personnel. Subjects were paid for their time. There were 34 subjects in this experiment, all of whom were female. Audiometric evaluations were not made of the subjects; subjects were questioned with respect to whether their hearing was normal, and no subject reported any hearing difficulty. Subjects were run in groups of up to 6 in an IAC booth, as described earlier. Only data from subjects who used at least half the response scale were used in the analysis. This criterion caused one subject's data to be excluded.

This procedure represents a compromise between a concern for perceptual acuity and the recognition that subjects use numbers differently. When subjects have roughly the same acuity and use the number scale differently, we wish to normalize their responses in order for subjects to contribute equally to the solution. On the other hand, when a subject does not perceive the differences between stimuli which are perceived by others, it is better not to use those data as an input to an analysis of that perception. A criterion of use of half the response scale was selected.

3.2.3 Experiment 1 Results

Stimulus Processing. Difference judgments on all pairs of stimuli were collected in order to do a multidimensional scaling analysis of the data. A few processing steps were necessary before running that analysis. These steps were performed in the order listed.

1. Normalization within each stimulus set. The data for each subject was normalized. If a subject did not use the full range of difference judgments available, i.e. 0-9, that subject's data were linearly transformed so that her responses did cover that range. The data from one subject were omitted because she did not meet the response criterion. The particular subject omitted used only numbers between 0 and 3.

2. Averaging within subjects. For each stimulus set, two responses, one for each order of presentation, were obtained from each subject; the mean of these responses was used as that subject's response to a stimulus pair. In the few cases where there were missing responses, the single difference judgment was used.

3. Averaging across subjects. For each stimulus set, the mean of the subjects' responses to each pair of stimuli was calculated and used as input to the multidimensional scaling program. A separate analysis was performed for each stimulus. (A later multidimensional analysis, INDSCAL, Carroll and Chang, 1970, was used to confirm that averaging over subjects was appropriate. INDSCAL produces subject weights for each dimension; the weight on a dimension for a subject indicates its relative importance in determining the subject's responses. The INDSCAL analysis indicated that, although subject weights varied across individuals, most of the subjects were responding to most of the dimensions, and that groups of subjects did not cluster into different subject types. Thus the general multidimensional scaling solution is believed to be representative of a typical subject.)

4. Averaging across stimulus sets. The mean for the two values, one from each stimulus set, was calculated for each stimulus pair and used as input to the multidimensional analysis program.

Data Analysis. The multidimensional scaling analysis used was KYST-2 (Kruskal, et. al., 1973), a non-metric method which accepts as input one value for each stimulus pair, except for the comparison of each stimulus with itself which is not accepted or analyzed. The resulting partial matrix is termed a lower half matrix. Table 3.2 presents the difference judgment data, averaged over noise samples, subjects, and order of presentation.

A difference between data differing only in the white noise sample used would not be expected, and separate analyses for each noise set provide an indication of the reliability of the multidimensional scaling results. The first two dimensions of the solution for each noise set are presented in Figures 3.3 and 3.4. While there is not perfect agreement between the location of the points in the two spaces, the main features of the solutions are very similar. To quantify this agreement, the correlations between corresponding dimensions were calculated for the three-dimensional solution; r = .94, .93, and .76 for Dimensions 1, 2, and 3, respectively.

Table 3.2 Data from Experiment 1. Averaged Difference Judgments Used as Input to Multidimensional Scaling Analysis.

Stim	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	1.8																
3	2.4	2.3															
5	5.3	4.5	5.1	1.7													
6	1.7	1.5	1.9	3.7	5.5												
7	2.2	1.7	2.4	3.2	4.2	2.2											
8	3.9	4.4	2.5	3.8	4.5	4.4	4.6										
9	6.1	5.5	5.0	2.9	1.5	5.3	4.8	4.3									
10	1.5	1.7	2.5	3.5	4.9	1.6	2.4	4.5	4.6								
11	1.9	2.0	2.8	3.1	5.0	2.4	1.5	4.1	4.4	2.5							
12	4.3	5.2	3.1	4.4	4.1	4.8	4.7	1.2	3.8	5.0	4.8						
13	1.4	1.6	1.9	3.8	4.5	1.4	2.2	4.1	5.7	2.0	2.4	5.2					
14	1.9	2.3	2.4	4.8	5.3	1.8	2.4	3.6	5.7	1.8	2.6	4.7	1.9				
15	2.7	2.1	3.3	3.7	4.0	2.3	1.9	3.5	4.7	2.9	1.5	3.5	2.5	3.2			
16	5.5	5.6	3.8	5.0	4.3	5.4	5.6	1.6	4.0	5.5	5.0	1.5	5.7	4.8	4.7		
17	1.4	1.8	1.8	4.1	5.2	1.4	2.4	3.8	5.9	1.6	2.7	4.8	1.6	1.8	2.7	4.8	
18	1.9	2.1	2.4	4.1	5.1	1.7	2.6	3.6	5.6	1.9	2.6	4.9	2.0	1.6	2.7	5.9	1.5

FIGURE 3.3

EXPERIMENT 1: TWO-DIMENSIONAL SOLUTION FOR DIFFERENCE JUDGMENTS BASED ON FIRST FROZEN NOISE SAMPLE.

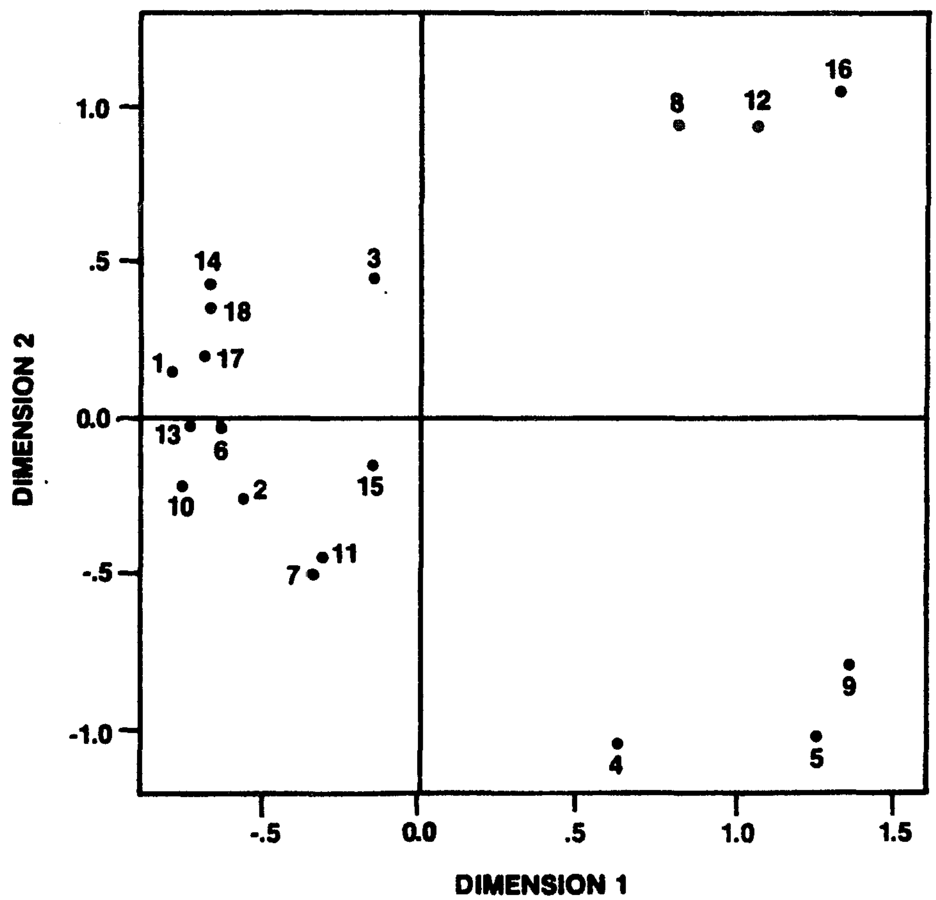
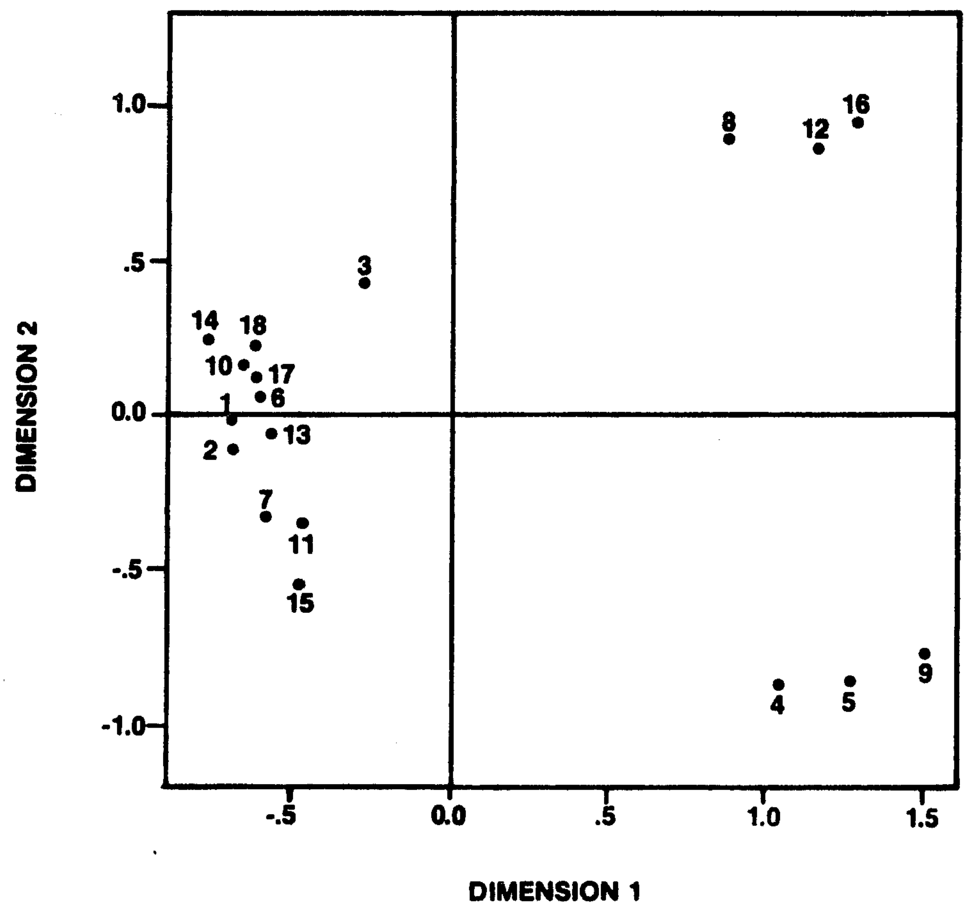


FIGURE 3.4

EXPERIMENT 1: TWO-DIMENSIONAL SOLUTION FOR DIFFERENCE JUDGMENTS BASED ON SECOND FROZEN NOISE SAMPLE.



With 18 comparisons, $r \geq .47$ is needed for significance at the .05 level. Agreement between the solutions for the two noise samples appears to be quite good for Dimensions 1 and 2, and there is significant agreement for Dimension 3 as well. These features will be discussed in terms of the general solution which combined both noise sets.

Figures 3.5 and 3.6 present Dimensions 1 and 2 and Dimensions 2 and 3, respectively, of the solution with the combined data, i. e., all the data gathered in Experiment 1.

The stress function (which provides a measure of goodness-of-fit) for solutions from 1 to 5 dimensions is presented in Figure 3.7. It is clear that a solution for these data required more than one dimension. At least two and perhaps three dimensions appear to be needed in order to account for the data adequately. The stress function does not present a sharp enough elbow to make the dimensionality clearcut.

In an effort to understand these data, several physical properties associated with aspects of reverberation were regressed into the three dimensional solution. Some of these variables and the correlation associated with their multiple linear regression into the

FIGURE 3.5

EXPERIMENT 1: DIMENSIONS 1 AND 2 OF THREE-DIMENSIONAL SOLUTION FOR WHOLE DATA SET.

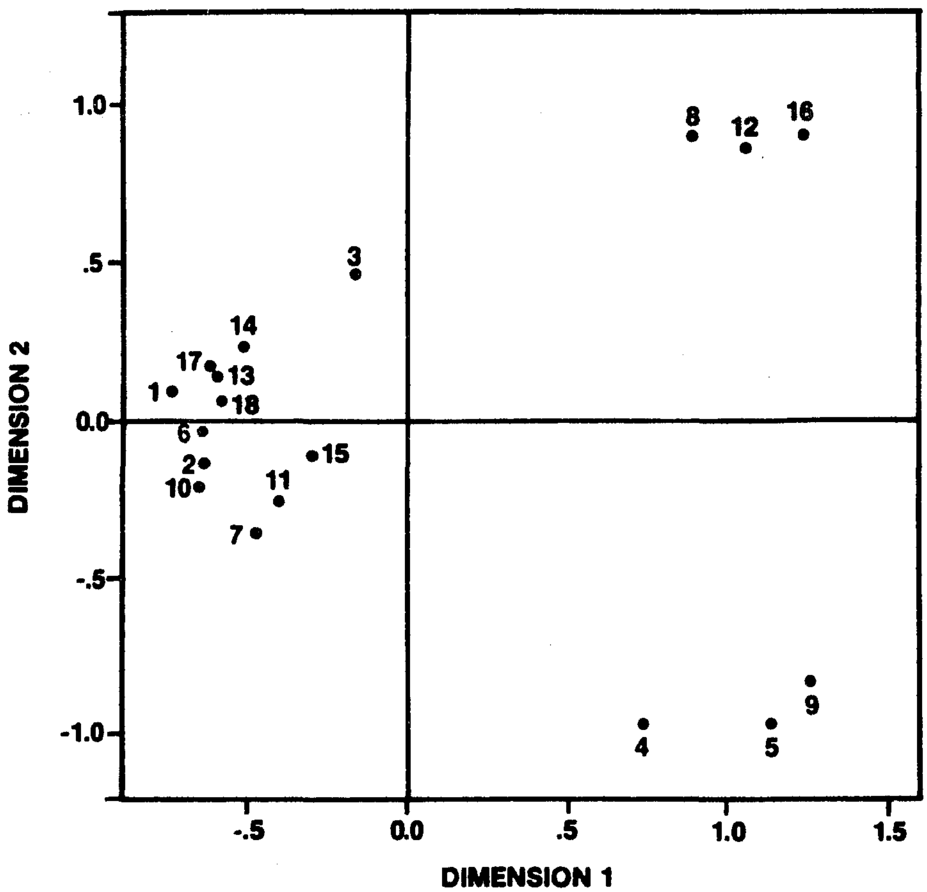


FIGURE 3.6

EXPERIMENT 1: DIMENSIONS 2 AND 3 OF THREE-DIMENSIONAL SOLUTION FOR WHOLE DATA SET.

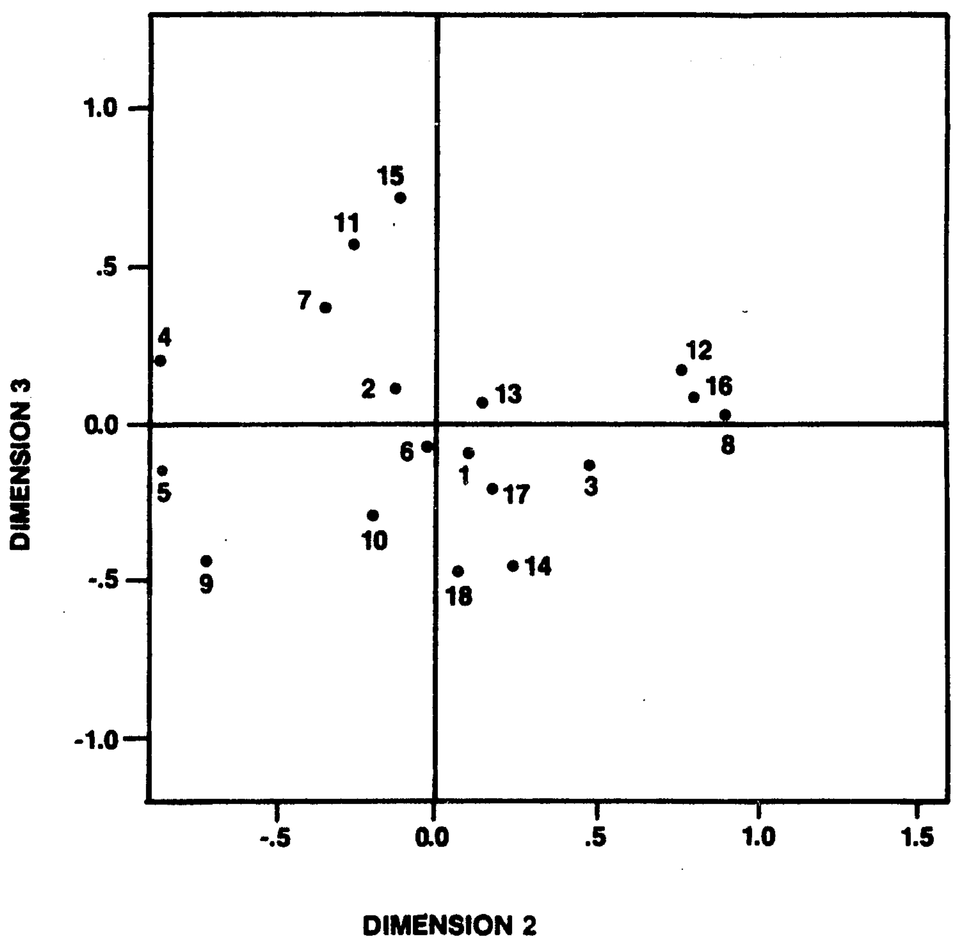


FIGURE 3.7

EXPERIMENT 1: STRESS FUNCTIONS FROM MULTIDIMENSIONAL SOLUTIONS.

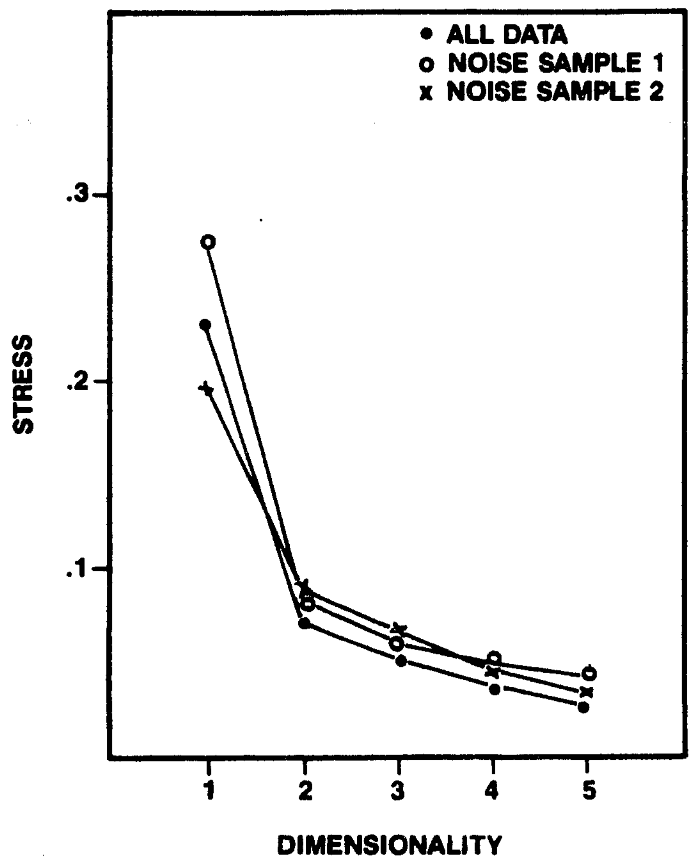


Table 3.3 Correlations of Variables with Multidimensional Scaling Solution

	Variable	Correlation
1.	Reverberation Time, T_{60}	.37
2.	β (Reflectivity)	.30
3.	Talker-microphone distance, Δ	.95
4.	σ	.86
5.	σ_{10}	.99
6.	σ_{20}	.96
7.	σ (Rectangular 47 Hz filter)	.96
8.	σ (Rectangular 140 Hz filter)	.98
9.	σ_{CB}, exponential filter	.97
10.	σ[Hamming window on noise stimuli]	.96

solution space are listed in Table 3.3.

These effects will be discussed in the following pages. Variables 1-4 in Table 3.3 have already been defined, but variables 5-10 have not and a brief definition will be given of each before they are discussed in more detail. Each is a measure of the standard deviation of H(f) after H(f) has been smoothed in some way.

1. σ_{10}, variable 5, is the standard deviation of an H(f) calculated from an h(t) where h(t) has been truncated to include only the first 10 msec of the room impulse response, h(t).

2. σ_{20}, variable 6, is the same as variable 5, except that h(t) has been truncated at 20 msec.

3. σ(Rectangular 47 Hz filter), variable 7, is the standard deviation of an H(f) that has been convolved with a 47 Hz wide rectangular filter.

4. σ(Rectangular 140 Hz filter), variable 8, is the same as variable 7 except that the rectangular filter is 140 Hz wide.

5. σ_{CB}, exponential filter, variable 9, is the standard deviation of H(f) after convolution with a filter with exponential slopes.

6. σ[16 msec wide Hamming window], variable 10, is the standard deviation of an H(f) produced by a short time spectral analysis of the noise stimuli.

Figures 3.8 and 3.9 display the projections of these variables into the three dimensional solution space. The length of the vector in each pair of displayed dimensions is determined by its projection onto that pair of dimensions. For example, in Figure 3.8, three vectors, σ, σ_{20}, and σ_{CB}, are all nearly parallel to Dimension 1. These measures will therefore be correlated with the Dimension 1 stimulus coordinates, and will be nearly independent ($r \simeq 0$) of the stimulus coordinates for Dimension 2.

Although linear regression has been used for convenient comparisons, it should be noted that most of the σ measures (those with $r < .97$) exhibit a little curvilinearity, being slightly concave upward, as is typical of rating scale data plotted as a function of dB rather than stimulus intensity (Stevens and Galanter, 1957). This represents a limit on the usefulness of linear regression for comparisons among these measures, but, as can be seen, the size of the correlations represent an even stronger limit.

Neither reverberation time nor its perfect ordinal correlate, β, regress well into this space. This was part of the initial hypothesis: the use of noise stimuli with controlled on and off ramps have removed the perception

FIGURE 3.8

EXPERIMENT 1: DIMENSIONS 1 AND 2 OF THREE-DIMENSIONAL SOLUTION WITH PROJECTIONS OF ROOM VARIABLES INTO THE SPACE.

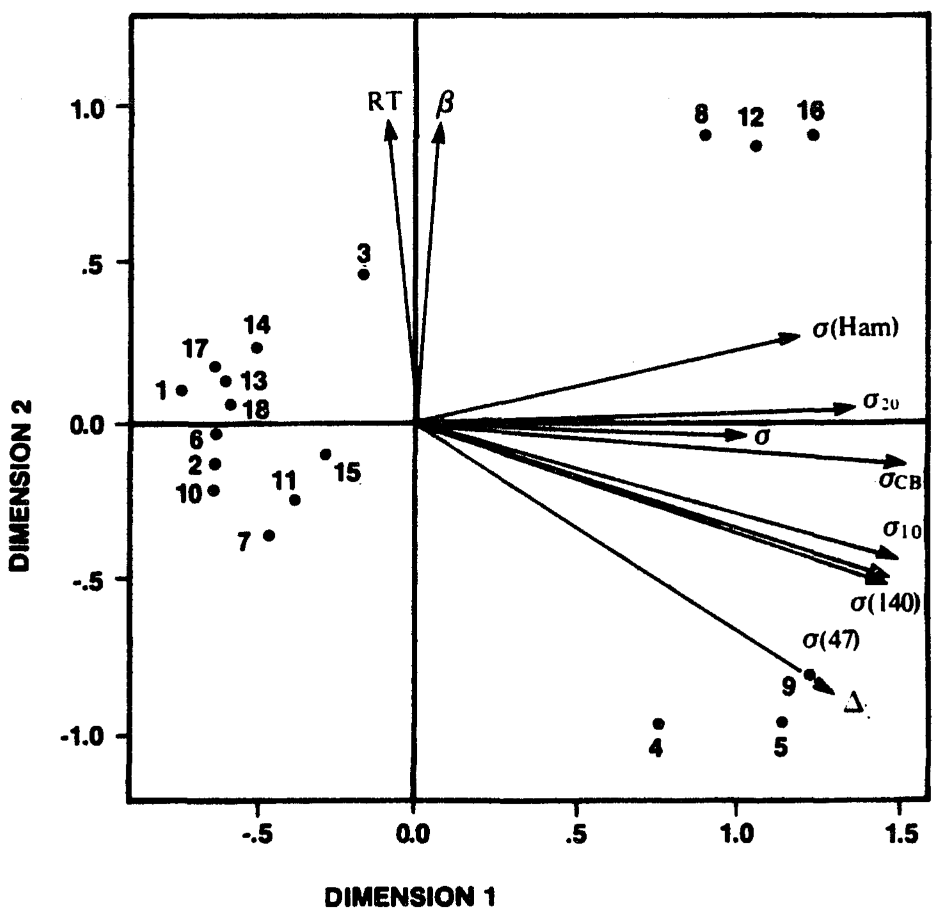
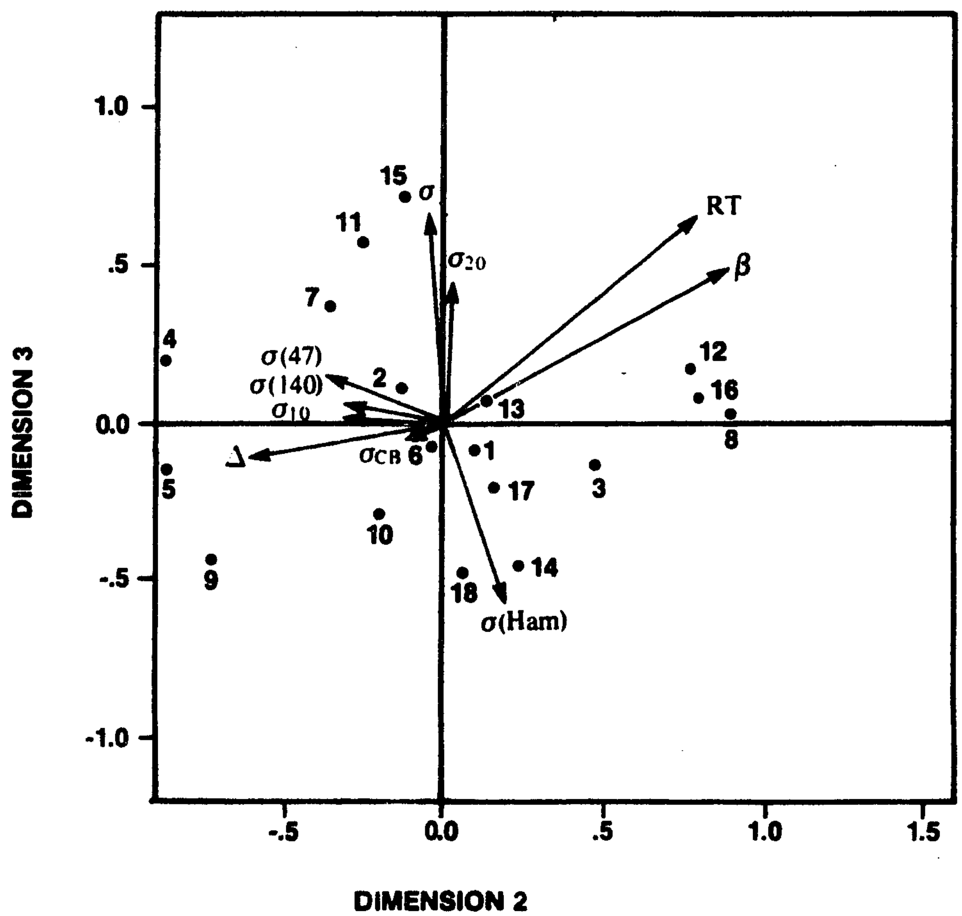


FIGURE 3.9

EXPERIMENT 1: DIMENSIONS 2 AND 3 OF THREE-DIMENSIONAL SOLUTION WITH PROJECTIONS OF ROOM VARIABLES INTO THE SPACE.



associated with reverberation time.

The talker-microphone distance, Δ, correlates highly with a dimension of the solution space, and its projection is not too different from the family of σ-type measures to be described next. In general, σ would be expected to increase with Δ because increasing Δ will reduce the energy in the original impulse compared to the energy in its reflections. For the most part, stimuli with a given Δ are clustered fairly closely together in this space. Another issue, which can't be examined within these data, is whether, for a fixed Δ, with the talker and/or microphone moved about in a room, or in different rooms, this correlation with a perceptual dimension would be maintained. This high correlation, .95, with a simple physical measure warrants further study.

The physical measure of the most _a priori_ interest is the standard deviation of the frequency response, σ. Its correlation within this space is .86; while this is not a low correlation coefficient, it is not as high as its correlation with the earlier speech experiments (Allen, et. al., 1979) would lead us to expect. Since this experiment was designed to show what σ could do, a rather detailed look was taken at the discrepancy between the σ of the stimuli and the projection of the stimuli on the σ

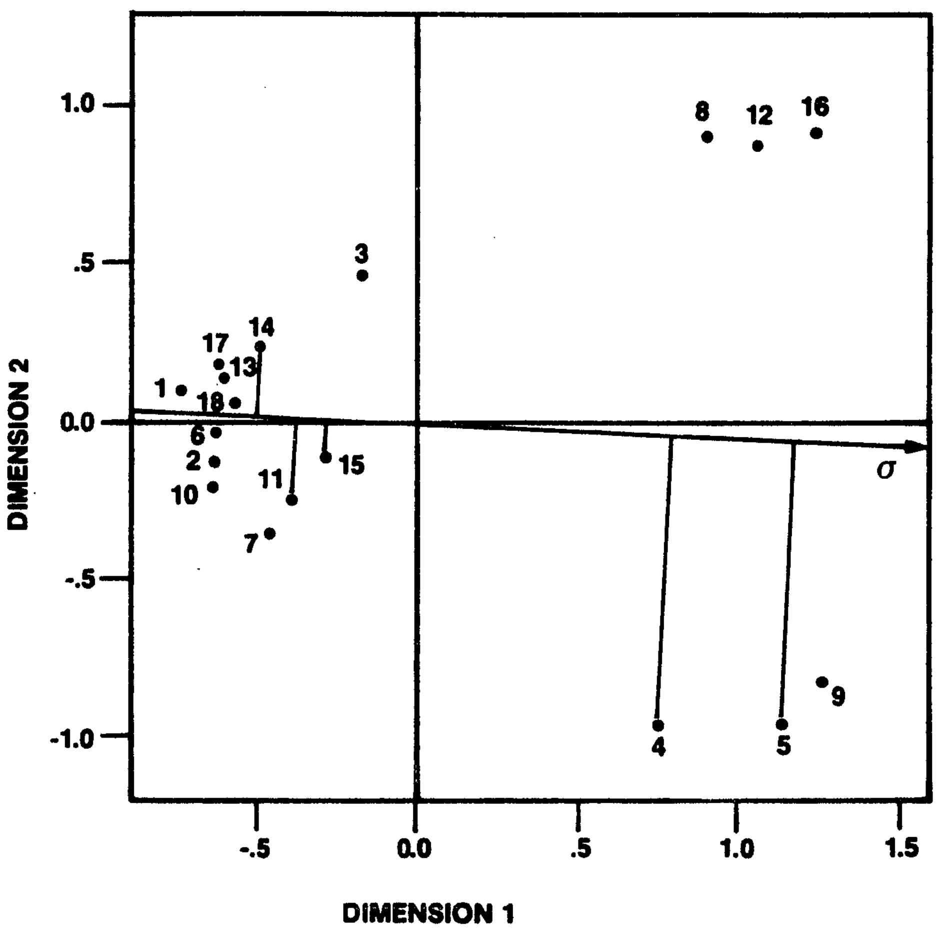
vector. Figure 3.10 presents the one- and two-dimensional space with only the σ vector plotted in it. Projections from some of the points to the vector have been dropped to show the values of σ predicted by the solution space.

The difference is particularly striking between Stimulus 5 and 15; both stimuli have a σ of about 5 with Stimulus 15 having a slightly greater σ. Yet Stimulus 15 is much closer to the low σ stimuli than would be expected. The same is true for Stimuli 11 and 14. When these three stimuli are examined in more detail, it is seen that they all have fairly low Δ with high β. This means they have relatively low energy in their early reflections with their large σ's deriving from a lot of energy in their late reflections. The opposite is true for Stimuli 4 and 5 which have high Δ and low β; these stimuli have more of their energy in their early reflections.

This effect of early reflections makes sense in terms of the discussion in the introduction of echo perception and, more generally, monaural peripheral processing. As was noted in the literature review, short echo delays produce perceptible ripples (inter-critical band ripples) in the frequency response; long echo delays do not (Atal, et. al., 1961, Zurek, 1976, Koenig, 1979). Accordingly, a

FIGURE 3.10

EXPERIMENT 1: σ VECTOR IN DIMENSIONS 1 AND 2 WITH PROJECTIONS OF POINTS ONTO VECTOR.



variety of measures which, to some extent, mimic this processing, by means of temporal or spectral filtering, were evaluated as underlying variables for describing the three-dimensional space. These are variables 5-10 in Table 3.3. These variables have been briefly defined above: variables 5-9 are σ's based on a modified frequency response while variable 10 is a σ based on the frequency spectrum of the stimuli.

As stated previously, variables 5 and 6, σ_{10} and σ_{20}, are based on an H(f) calculated from an h(t) where h(t) has been truncated to include only the first 10 or 20 msec of the impulse response. When these variables, σ_{10} and σ_{20}, are projected into the solution space, correlations of .99 and .96, respectively, are obtained.

Variables 7, 8, and 9 are also σ calculations based on smoothed versions of H(f). In each case a filter is convolved with H(f) to give the smoothed frequency response; these filters represent approximations to the critical band processing of the input frequency response. The two rectangular filters, variables 7 and 8, have a single parameter, bandwidth. The 47 and 140 Hz filters result in correlation coefficients of .96 and .98, respectively. (Other bandwidths within this range result in similarly high correlations.)

The third filter, variable 9, is more complex and a more detailed description is needed. It is somewhat closer to the steep sided Patterson model (1974, 1980) reviewed earlier. (This filter was suggested by Michael Pavel.) Rather than a single-parameter rectangular filter, this two-parameter filter is exponential with a lower slope described by $(f/f_c)^{\alpha}$ and upper slope described by $(f_c/f)^{\gamma}$ where f_c is the point in H(f) being filtered or is the characteristic frequency of the "tuning curve" represented by the filter. This filter results in a filter whose width increases with frequency. Critical band data indicate this is true for frequencies above about 500 Hz. For frequencies, (f_c), below 500 Hz, both f_c and f are increased by the difference between f_c and 500 Hz before exponentiation; this keeps the filter at a constant width below 500 Hz. The increasing width of the filter acts to produce smoothing at high frequencies and upward masking. Trial and error runs varying α and γ resulted in a best fit, i.e. highest correlation, of .97 for $\alpha = 8$ and $\gamma = 16$.

Another approach (suggested by J. B. Allen) involved short-time power spectral analysis of the noise stimuli rather than operations on h(t) or H(f). (This approach is similar to that used by Atal, et. al., 1961). A Hamming window is a cosine shaped time window. A window 16 msec in width was used as the short-time weighting function on each noise stimulus; a fast Fourier transform analysis was performed on each sample and, because the stimuli were stationary, the average of these short-time frequency responses was taken, and the standard deviation of this frequency response calculated for each of the 18 stimuli. The resulting correlation, when this property is projected into the solution space, is .96. (Other window widths were tried, up to about 32 msec, and the 16 msec window produced the highest correlation.)

3.2.4 Experiment 1 Discussion

All of these correlations between the solution and the σ-type variables are very high. Furthermore, by examining Figures 3.8 and 3.9, it is seen that there is fairly good agreement among them. Correlations between these measures are mostly greater than .9. By design, they are all measuring roughly the same thing, and their psychoacoustic correlate, coloration, appears to be represented by the first dimension of the solution. Although these properties agree well among themselves and with the data, and too well to permit choice among them, this may be partly or largely due to the clustering of points within the data. Viewing Figure 3.5 again, it appears there are three separate clusters of points rather than 18 stimulus points well dispersed in the space. Thus there is a lack of resolving power in the data. The meaning of the data clusters is linked to the interpretation of Dimensions 2 and 3 of the solution.

Dimension 2 is determined by the perceived difference in two sets of rooms, those with a Δ of 5 and 10, and dimension 3 is mostly determined by the difference between rooms with a Δ of 2.5 and other rooms. The extremes of Dimension 2 are two of the stimulus clusters; the $\Delta = 5$ stimuli create one stimulus cluster, and the $\Delta = 10$ stimuli create another. The rest of the stimuli are clustered fairly closely together in Dimensions 1 and 2. In order to try to understand the underlying variables for Dimensions 2 and 3, some additional exploration of the stimuli was done.

Listening informally to these stimuli, the sets of rooms with the same Δ sound "similar" with respect to something pitchlike, and increasing β within a set of rooms with fixed Δ produces mostly a stronger perception of that particular "pitchiness." The particular complex pitch perceived presumably depends on the peaks and valleys of the stimulus after monaural processing has done its job.

Some informal tests of musically experienced listeners were tried to verify this perception of pitch quality in rooms. Rooms with only one highly reflective surface were simulated; the subjects listened to the same type of noise stimuli as before. Such rooms yield an approximation to stimuli with a single echo. Listeners were asked to compare these rooms, with respect to pitch quality, to other rooms, some of which were the same except that reflectivity was high for the other surfaces as well. Usually the listeners could accurately state whether the pitch quality of the single echo room was present in the room with multiple echoes. When these listeners were asked to adjust a tone to match a ripple in a single or multi-echo room, they were generally unable to do so. The potential perceptual complexity of coloration quality is very great if the human capability for pitch discrimination can be applied to reverberant sound.

From this, coloration appears to have quality (pitchiness) as well as quantity associated with it. Quantity of coloration is roughly represented by dimension 1, is measured by some σ-like quantity, and is produced by increasing β while holding other room parameters constant. Perceptually, the effect is of an increasing amount of pitchiness. Quality of coloration is perceived as a pitchlike quality and is produced in this experiment by varying Δ.

In addition to these issues, there are variables which have not been manipulated in simulated room experiments and which suggest interesting questions, e.g., the effect of room size on coloration and whether coloration changes as a fixed Δ is moved about in a room.

To summarize the results from Experiment 1:

1. Coloration is a multidimensional perceptual phenomenon; amount of coloration is single dimensional.

2. A possible classification of coloration into quantity and quality attributes is suggested with quantity related to the various σ measures and quality to Δ.

3. σ is a reasonable descriptor of the quantity of coloration, but it does not appear to be the best measure of it. Other σ-type measures which model monaural processing in various ways better represent amount of coloration.

4. Δ correlates very highly within the solution space and fairly well with measures of coloration quantity. However, differences in Δ appear also to produce differences in coloration quality.

5. With all else held constant, quantity of coloration appears to increase with β. This is because increasing β increases the energy in the early reflections. (T_{60} also increases with β, but this is because increasing β increases the energy in and lengthens the reverberant tail of h(t).)

Questions for the next experiment concern the validity of this interpretation of coloration as having both quantity and quality. The dimensionality of the space is still in question as well as the variables underlying the dimensions. There is a set of σ-type variables, all highly intercorrelated, which account for the coloration quantity dimension of the space very well, but no measures which account for the second and third dimensions. It is desirable, if possible, to distinguish between the measures of coloration (variables 5-10, Table 3.3) to determine which provides the best measure.

3.3 EXPERIMENT 2

A first goal of this experiment was to determine the validity of the interpretation of Experiment 1, i.e. that coloration has both quantity and quality. Coloration quantity appeared to be manipulable by varying β, the reflectivity of the room's surfaces. In this experiment sets of rooms which varied only in β were used, and it was expected that the solution space would demonstrate increasing coloration with increasing β for each room set. Room sets which sounded different in quality were chosen, and it was expected that the multidimensional solution would reflect quality differences in some way.

Second, a selection of a measure of coloration quantity is needed as well as some grasp of the variables determining coloration quality. A more complex stimulus space may permit choice among the potential coloration quantity measures which were used in Experiment 1.

Third, the introduction of additional variables into the experiment, i.e., room size, source and receiver position in the room, and distance from a wall, may contribute insights into coloration. The smaller the room, the more reflections will occur within the 512 msec interval in which the impulse response is calculated. The effect of this, or some other perceptual effect of room size, is not known since neither Experiment 1 nor the earlier speech study manipulated this variable.

Another point deserves evaluation: σ_{10}, the standard deviation of H(f) based on the first 10 msec of h(t), correlated .99 with the Experiment 1 stimulus coordinates. This simplicity of this formulation as a measure of coloration makes it very attractive, and it can be tested directly. A source and receiver were located in rooms so that no reflections reached the receiver in the first 10 msec. If σ_{10} is a good measure of coloration, such rooms should produce little or no perception of coloration, no matter how large β is.

3.3.1 Experimental Design.

A larger set of stimuli was used than in Experiment 1 and greater variation in the physical room parameters was introduced. This greater complexity was an attempt to produce a richer stimulus space with the potential for (1) discriminating between σ-type measures and (2) a large number of dimensions if perception produces them. This complexity was introduced systematically so that particular questions could be addressed as well.

Table 3.4 contains a description of the 26-stimulus set.

Table 3.4 Room Characteristics of Experiment 2 Stimuli.

	Room	Volume (ft^3)	Δ (ft)	β	σ
Set 1	1	12.5x15.0x16.3	5.0	.01	.07
	2	12.5x15.0x16.3	5.0	.16	1.01
	3	12.5x15.0x16.3	5.0	.42	3.30
	4	12.5x15.0x16.3	5.0	.58	4.18
	5	12.5x15.0x16.3	5.0	.80	5.57
Set 2	6	12.5x15.0x16.3	10.0	.21	2.17
	7	12.5x15.0x16.3	10.0	.29	3.28
	8	12.5x15.0x16.3	10.0	.54	5.14
Set 3	9	12.5x15.0x16.3	5.0	.30	1.74
	10	12.5x15.0x16.3	5.0	.50	3.38
	11	12.5x15.0x16.3	5.0	.65	5.06
Set 4	12	12.5x15.0x16.3	7.2	.22	1.47
	13	12.5x15.0x16.3	7.2	.50	4.06
	14	12.5x15.0x16.3	7.2	.80	5.58
Set 5	15	9.0x12.0x10.0	5.0	.18	1.21
	16	9.0x12.0x10.0	5.0	.35	2.56
	17	9.0x12.0x10.0	5.0	.46	3.86
	18	9.0x12.0x10.0	5.0	.70	5.29
Set 6	19	9.0x12.0x10.0	3.6	.35	1.93
	20	9.0x12.0x10.0	3.6	.54	3.51
	21	9.0x12.0x10.0	3.6	.70	5.43
Set 7	22	20.0x30.0x19.0	5.0	.40	2.15
	23	20.0x30.0x19.0	5.0	.58	3.38
	24	20.0x30.0x19.0	5.0	.85	5.02
Set 8	25	20.0x30.0x19.0	7.2	.38	1.93
	26	20.0x30.0x19.0	7.2	.70	4.22

Figures 3.11 to 3.22 display the room impulse responses h(t)'s, and the room frequency responses for the least colored room, Stimulus 1 in Set 1, and for the most highly

colored rooms in some of the room sets (5, 8, 14, 18, and 26). While useful as examples, these figures show the difficulty of developing intuition about coloration phenomena from displays of complex derived room variables.

The rooms are divided into 8 sets; within each set all physical room parameters are constant except for β, the wall reflectivity. If quantity of coloration increases with β, this should determine the ordering of each stimulus set along a quantity of coloration dimension. There are 2 to 5 rooms in each set. Stimulus 1, though part of Set 1 with respect to the physical parameters which produced it, represents a room without coloration. For this room β is .01; the walls reflect almost no energy, and an almost flat H(f) should be produced. That is what happens; σ = .07 for this room.

Stimulus Sets 1-4 have the same room volume as the room in the first experiment. Stimulus Sets 1 and 2 are the room sets represented by Δ = 5 and Δ = 10, respectively, in the first experiment in that the talker and microphone are situated in the same place in the same room dimensions. Values of β differ somewhat.

FIGURE 3.11

EXPERIMENT 2 ROOM IMPULSE RESPONSE ROOM 1
DATA IN FILE HTRESP
PARAMETERS(Ft): LX= 13. LY= 15. LZ= 16. VOL=3047.
TRANSMITTER : X0= 5.0 Y0= 7.5 Z0= 7.5
RECEIVER : X= 5.0 Y= 2.5 Z= 7.5
BETA(X,Y,Z): .01 .01 .01 .01 .01 .01 SAMP RATE(KHz)= 8.0
ALPHA(X,Y,Z): 1.00 1.00 1.00 1.00 1.00 1.00 AVERAGE ALPHA= 1.00
VOLUME=3047. , AREA=1269. , T-R DIST= 5.0

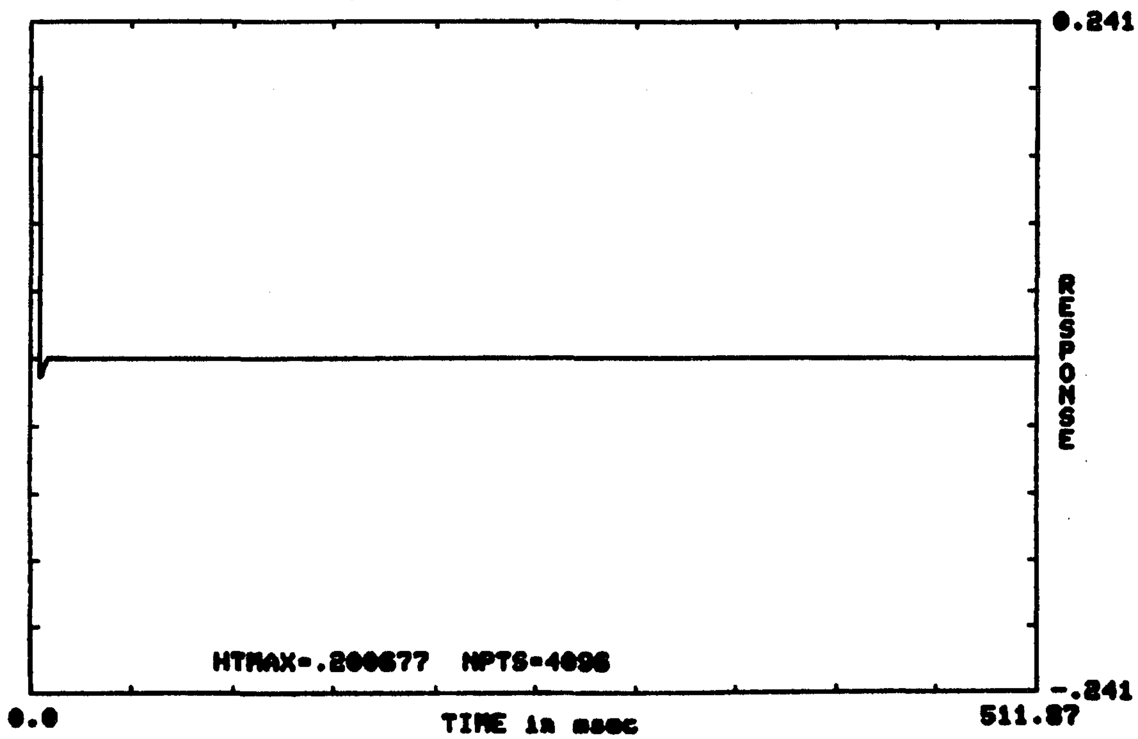


FIGURE 3.12

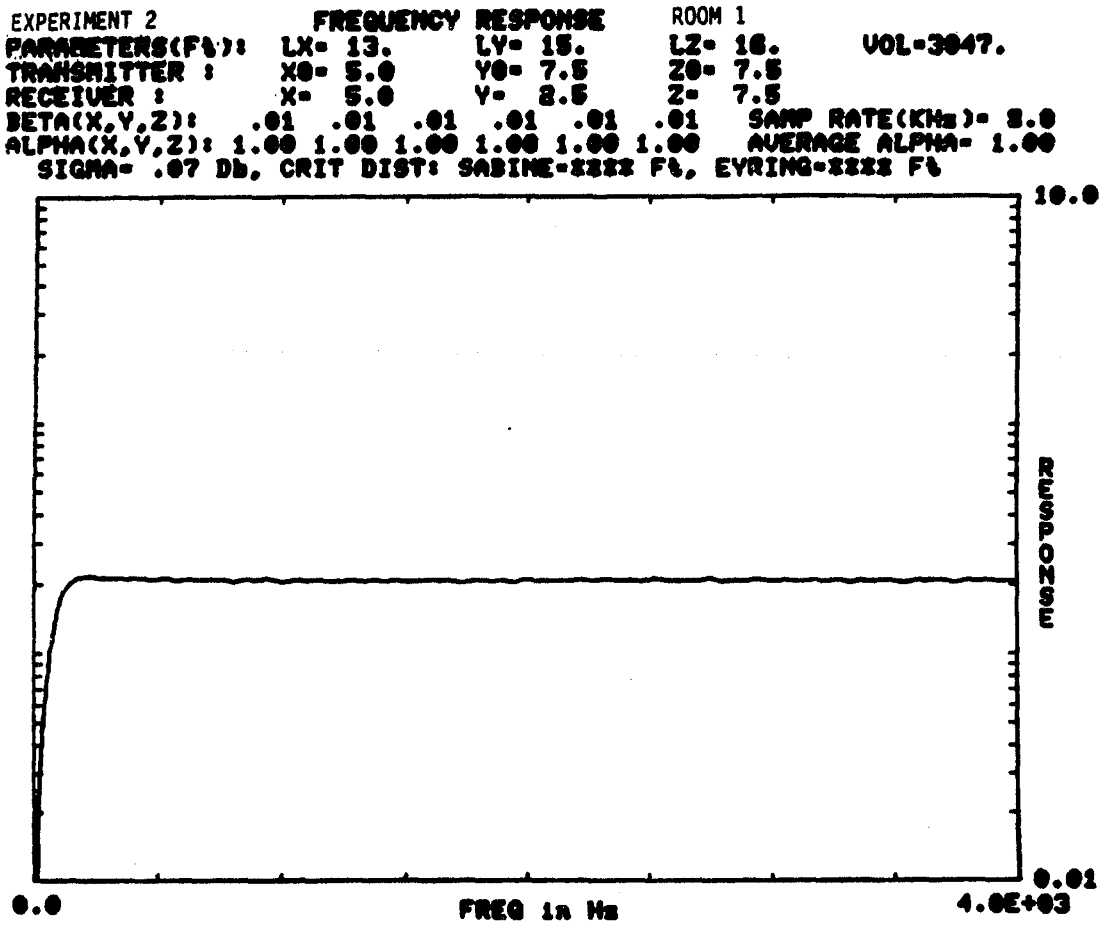

EXPERIMENT 2 FREQUENCY RESPONSE ROOM 1
PARAMETERS(Ft): LX= 13. LY= 15. LZ= 16. VOL=3047.
TRANSMITTER : X0= 5.0 Y0= 7.5 Z0= 7.5
RECEIVER : X= 5.0 Y= 8.5 Z= 7.5
BETA(X,Y,Z): .01 .01 .01 .01 .01 .01 SAMP RATE(KHz)= 8.0
ALPHA(X,Y,Z): 1.00 1.00 1.00 1.00 1.00 1.00 AVERAGE ALPHA= 1.00
SIGMA= .07 Db, CRIT DIST: SABINE=**** Ft, EYRING=**** Ft
10.0
RESPONSE
0.01
0.0
FREQ in Hz
4.0E+03

FIGURE 3.13

EXPERIMENT 2 ROOM IMPULSE RESPONSE ROOM 5
DATA IN FILE HTRESP

ROOM PARAMETERS(Ft):	LX= 13.		LY= 16.		LZ= 16.		VOL=3047.
TRANSMITTER LOCATION:	X0= 5.0		Y0= 7.5		Z0= 7.5		
RECEIVER LOCATION:	X= 5.0		Y= 2.5		Z= 7.5		
BETA(X,Y,Z): .80	.80	.80	.80	.80	.80	SAMP RATE(KHz)=8.00	
ALPHA(X,Y,Z): .36	.36	.36	.36	.36	.36	AVERAGE ALPHA= .36	

VOLUME=3047. , AREA=1269. , T-R DIST= 5.0

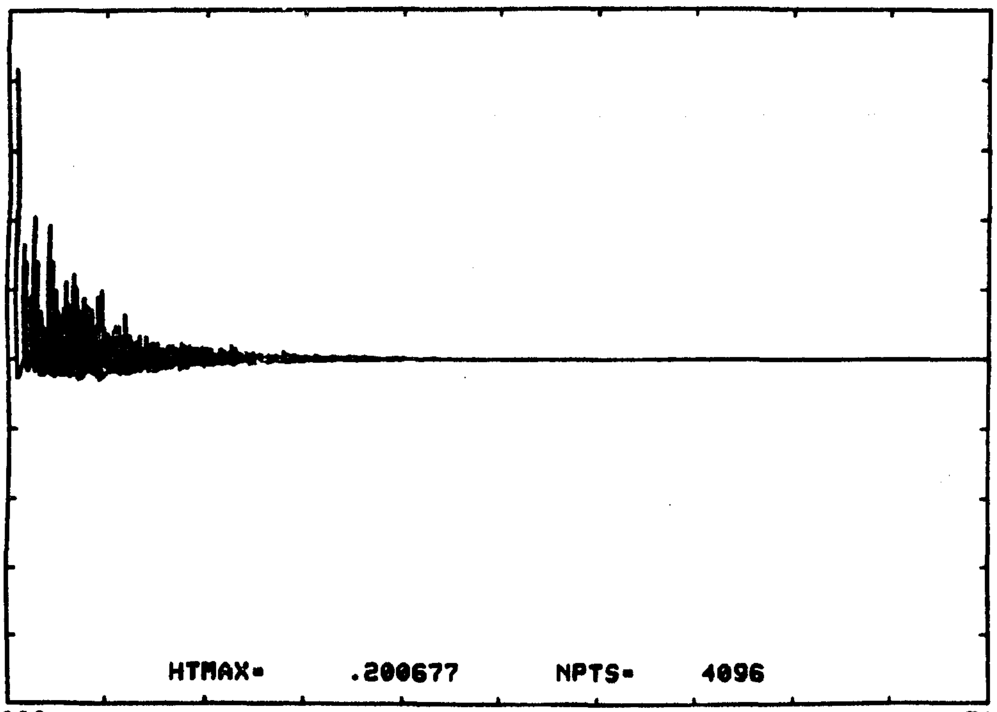


FIGURE 3.14

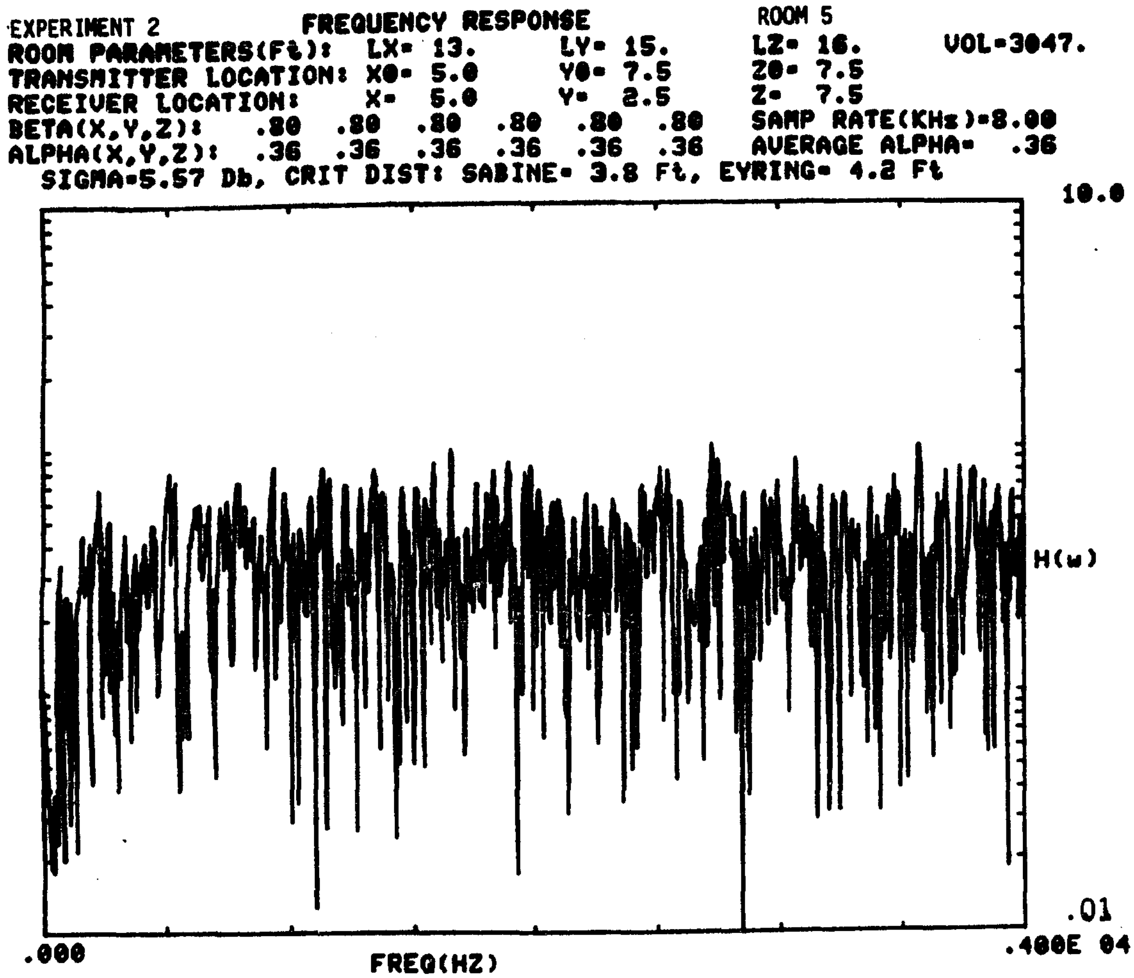

EXPERIMENT 2
FREQUENCY RESPONSE
ROOM 5
ROOM PARAMETERS(Ft): LX= 13. LY= 15. LZ= 16. VOL=3047.
TRANSMITTER LOCATION: X0= 5.0 Y0= 7.5 Z0= 7.5
RECEIVER LOCATION: X= 5.0 Y= 2.5 Z= 7.5
BETA(X,Y,Z): .80 .80 .80 .80 .80 .80 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .36 .36 .36 .36 .36 .36 AVERAGE ALPHA= .36
SIGMA=5.57 Db, CRIT DIST: SABINE= 3.8 Ft, EYRING= 4.2 Ft
10.0
H(w)
.01
.000
FREQ(HZ)
.400E 04

FIGURE 3.15

EXPERIMENT 2 ROOM IMPULSE RESPONSE ROOM 8
DATA IN FILE HTRESP
ROOM PARAMETERS(Ft): LX= 13. LY= 15. LZ= 16. VOL=3047.
TRANSMITTER LOCATION: X0= 7.5 Y0= 11. Z0= 7.5
RECEIVER LOCATION: X= 2.5 Y= 2.5 Z= 7.5
BETA(X,Y,Z): .54 .54 .54 .54 .54 .54 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .71 .71 .71 .71 .71 .71 AVERAGE ALPHA= .71
VOLUME=3047. , AREA=1269. , T-R DIST= 10.

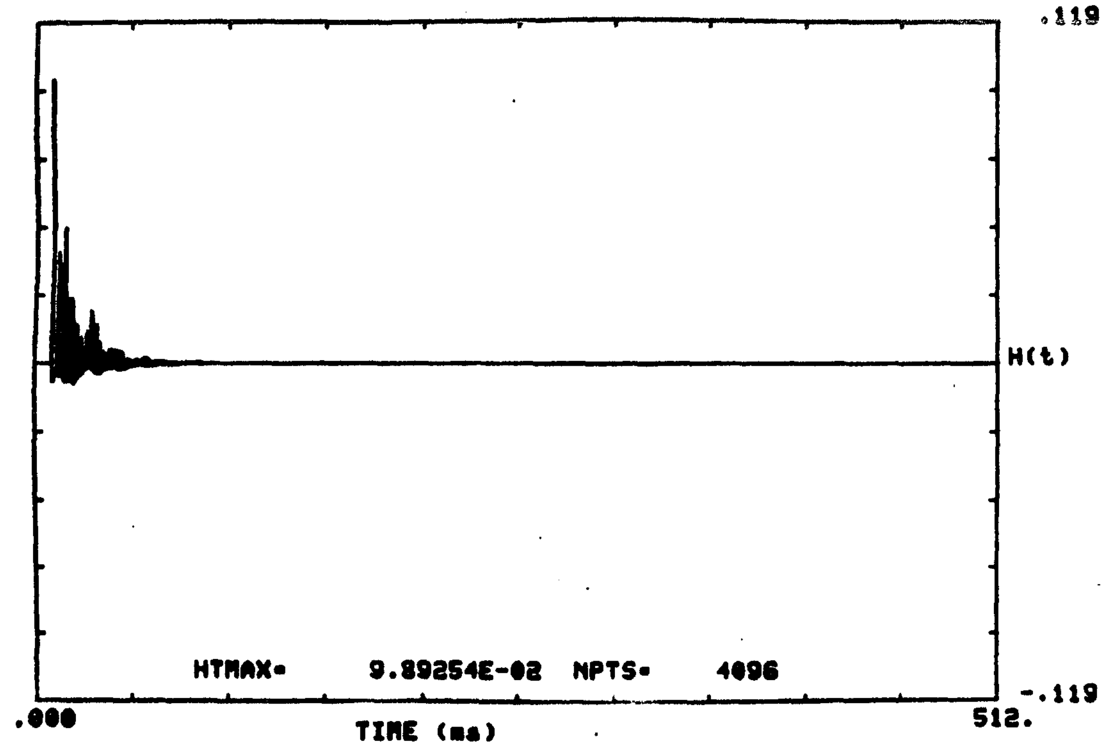


FIGURE 3.16

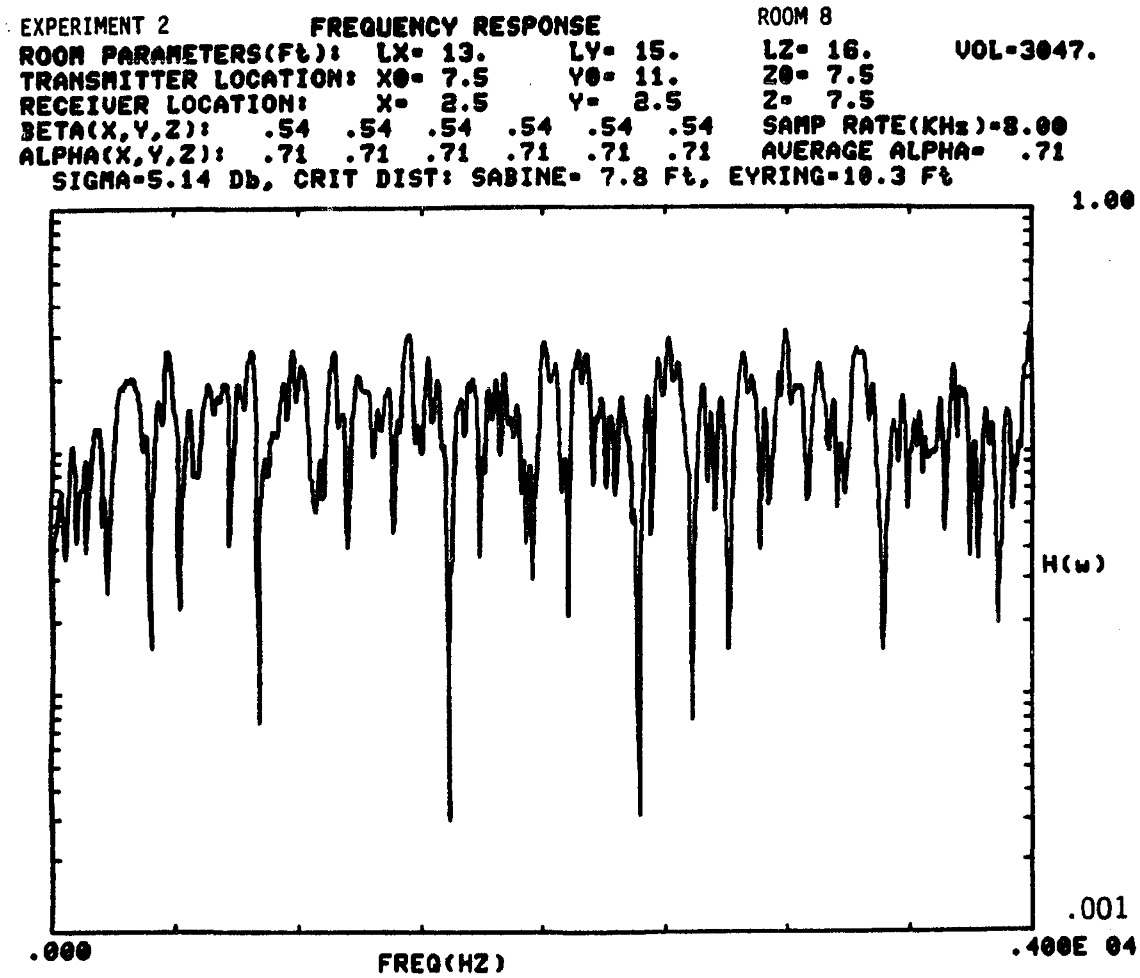

EXPERIMENT 2
FREQUENCY RESPONSE
ROOM 8
ROOM PARAMETERS(Ft): LX= 13. LY= 15. LZ= 16. VOL=3047.
TRANSMITTER LOCATION: X0= 7.5 Y0= 11. Z0= 7.5
RECEIVER LOCATION: X= 2.5 Y= 2.5 Z= 7.5
BETA(X,Y,Z): .54 .54 .54 .54 .54 .54 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .71 .71 .71 .71 .71 .71 AVERAGE ALPHA= .71
SIGMA=5.14 Db, CRIT DIST: SABINE= 7.8 Ft, EYRING=10.3 Ft
1.00
H(w)
.001
.000
FREQ(HZ)
.400E 04

FIGURE 3.17

EXPERIMENT 2 ROOM IMPULSE RESPONSE ROOM 14

DATA IN FILE HTRESP

ROOM PARAMETERS(Ft): LX= 13. LY= 15. LZ= 16. VOL=3047.
TRANSMITTER LOCATION: X0= 5.0 Y0= 4.5 Z0= 7.5
RECEIVER LOCATION: X= 6.5 Y= 12. Z= 7.5
BETA(X,Y,Z): .80 .80 .80 .80 .80 .80 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .36 .36 .36 .36 .36 .36 AVERAGE ALPHA= .36
VOLUME=3047. , AREA=1269. , T-R DIST= 7.2

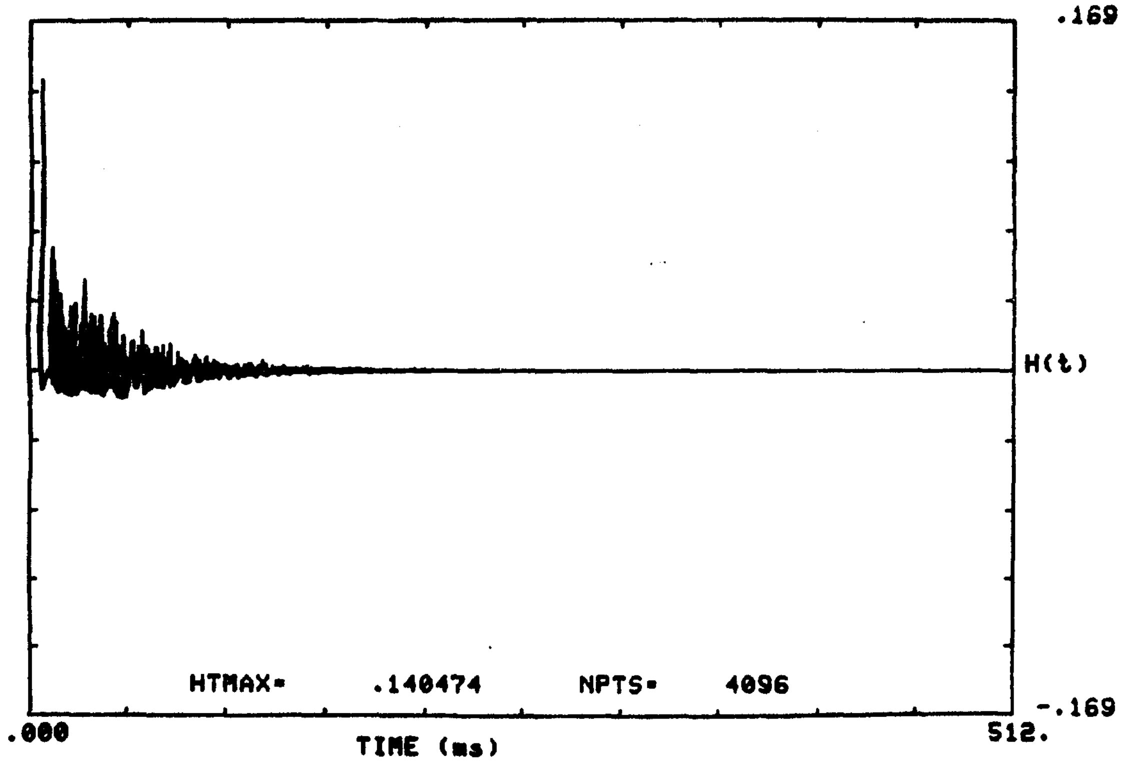


FIGURE 3.18

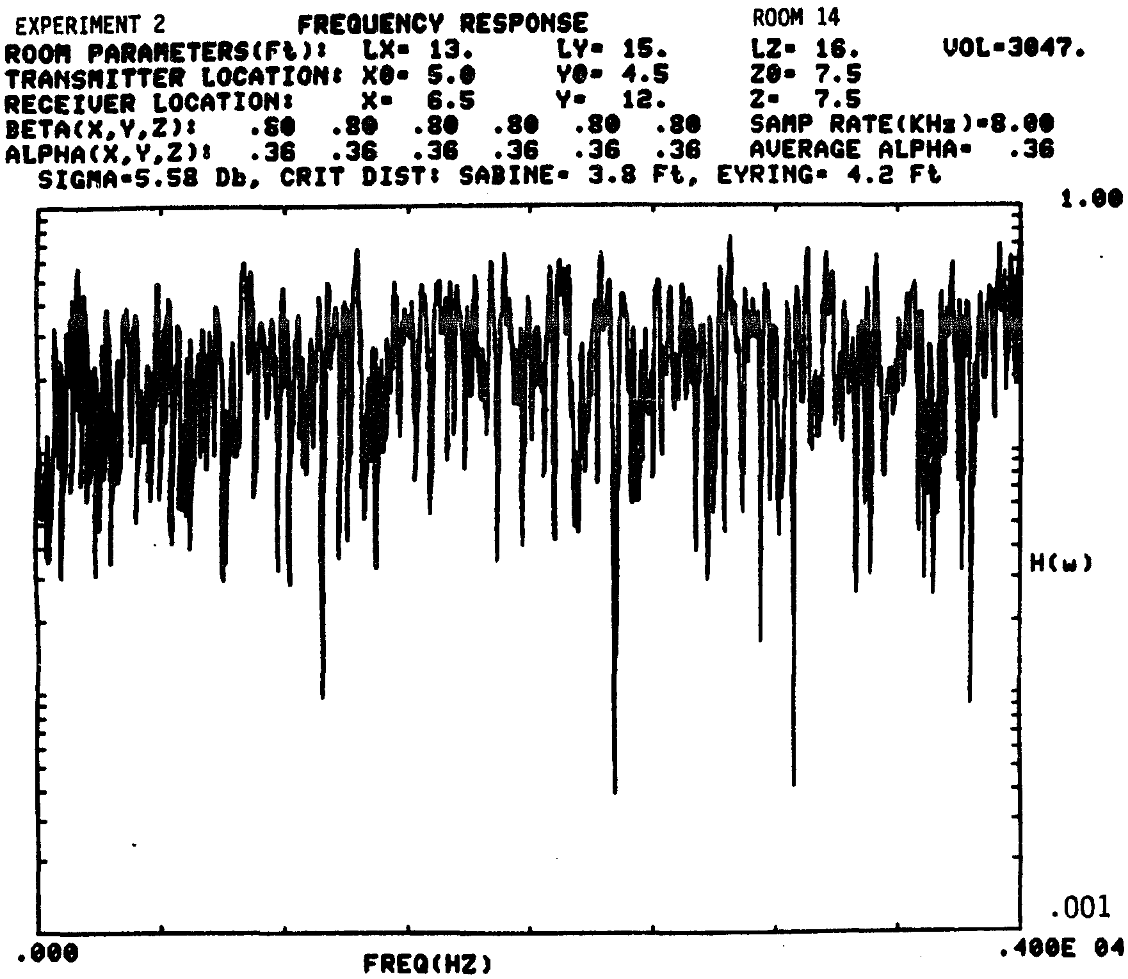

EXPERIMENT 2
FREQUENCY RESPONSE
ROOM 14
ROOM PARAMETERS(Ft): LX= 13. LY= 15. LZ= 16. VOL=3047.
TRANSMITTER LOCATION: X0= 5.0 Y0= 4.5 Z0= 7.5
RECEIVER LOCATION: X= 6.5 Y= 12. Z= 7.5
BETA(X,Y,Z): .80 .80 .80 .80 .80 .80 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .36 .36 .36 .36 .36 .36 AVERAGE ALPHA= .36
SIGMA=5.58 Db, CRIT DIST: SABINE= 3.8 Ft, EYRING= 4.2 Ft
1.00
H(w)
.001
.000
FREQ(HZ)
.400E 04

FIGURE 3.19

EXPERIMENT 2 ROOM IMPULSE RESPONSE ROOM 18
DATA IN FILE HTRESP
ROOM PARAMETERS(Ft): LX= 9.0 LY= 12. LZ=10.0 VOL=1080.
TRANSMITTER LOCATION: X0= 2.5 Y0= 3.0 Z0= 4.8
RECEIVER LOCATION: X= 5.5 Y= 7.0 Z= 4.5
BETA(X,Y,Z): .70 .70 .70 .70 .70 .70 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .51 .51 .51 .51 .51 .51 AVERAGE ALPHA= .51
VOLUME=1080. , AREA=636.0 , T-R DIST= 5.0

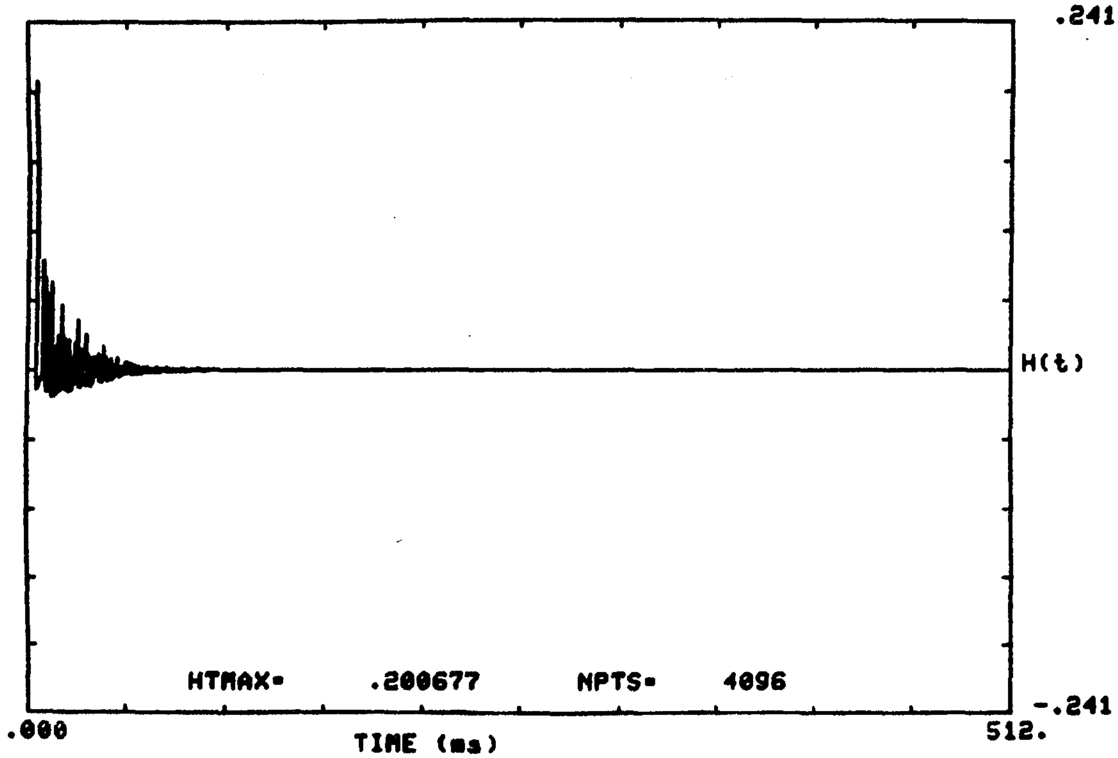


FIGURE 3.20

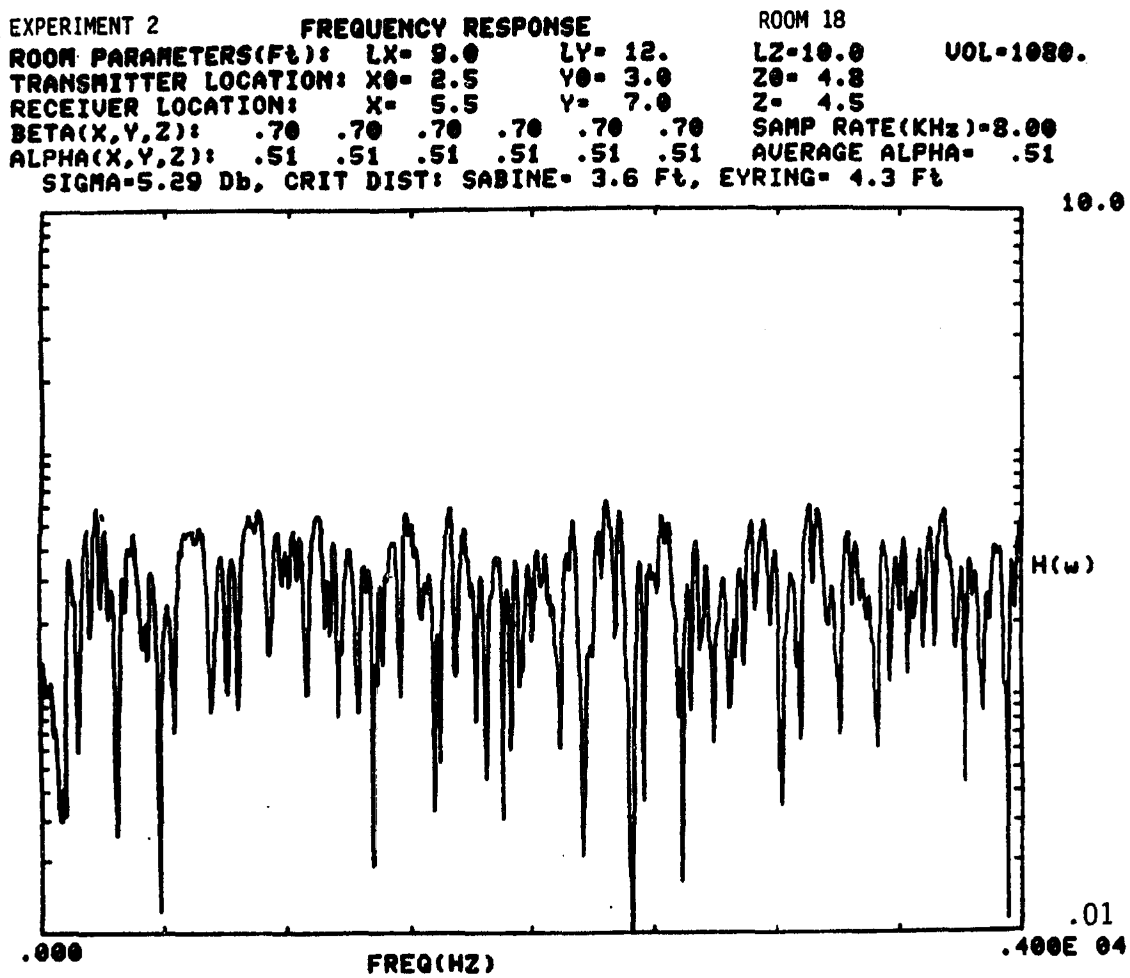

EXPERIMENT 2
FREQUENCY RESPONSE
ROOM 18
ROOM PARAMETERS(Ft): LX= 9.0 LY= 12. LZ=10.0 VOL=1080.
TRANSMITTER LOCATION: X0= 2.5 Y0= 3.0 Z0= 4.8
RECEIVER LOCATION: X= 5.5 Y= 7.0 Z= 4.5
BETA(X,Y,Z): .70 .70 .70 .70 .70 .70 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .51 .51 .51 .51 .51 .51 AVERAGE ALPHA= .51
SIGMA=5.29 Db, CRIT DIST: SABINE= 3.6 Ft, EYRING= 4.3 Ft
10.0
H(w)
.01
.000
FREQ(HZ)
.400E 04

FIGURE 3.21

EXPERIMENT 2 ROOM IMPULSE RESPONSE ROOM 26
DATA IN FILE HTRESP
ROOM PARAMETERS(Ft): LX= 20. LY= 30. LZ= 19. VOL=1.140E 04
TRANSMITTER LOCATION: X0= 8.0 Y0=10.0 Z0= 9.0
RECEIVER LOCATION: X= 9.5 Y= 17. Z= 9.0
BETA(X,Y,Z): .70 .70 .70 .70 .70 .70 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .51 .51 .51 .51 .51 .51 AVERAGE ALPHA= .51
VOLUME=1.140E 04, AREA=3100. , T-R DIST= 7.2

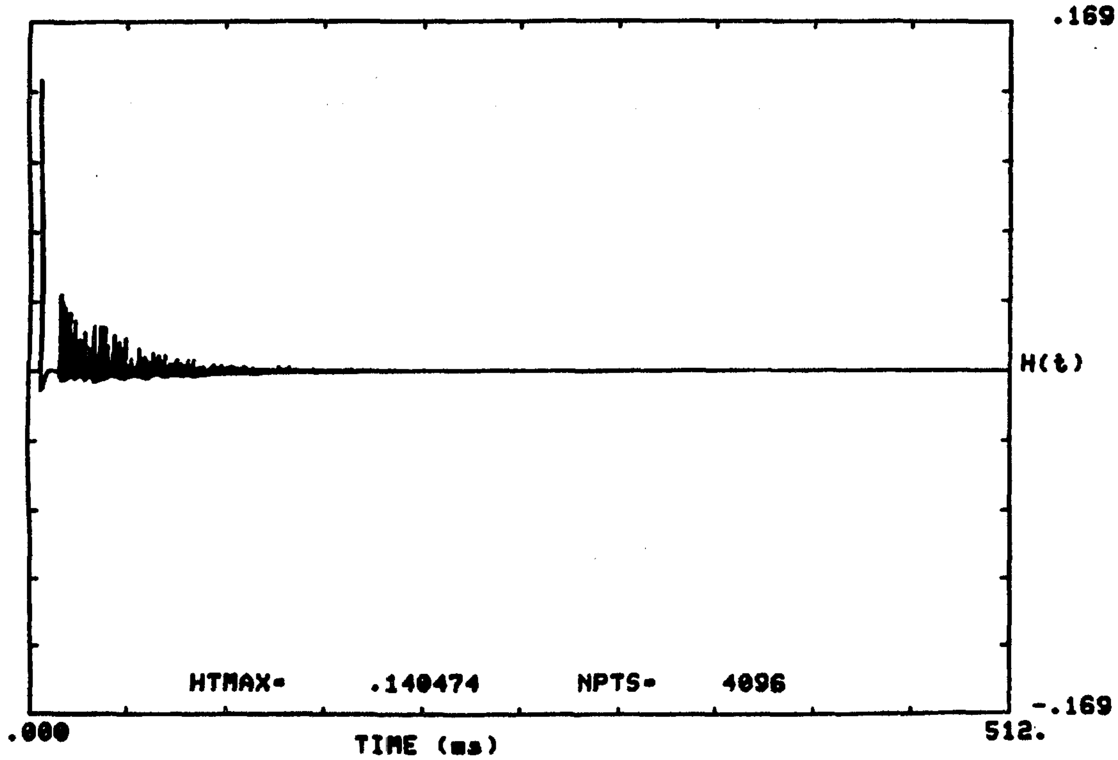
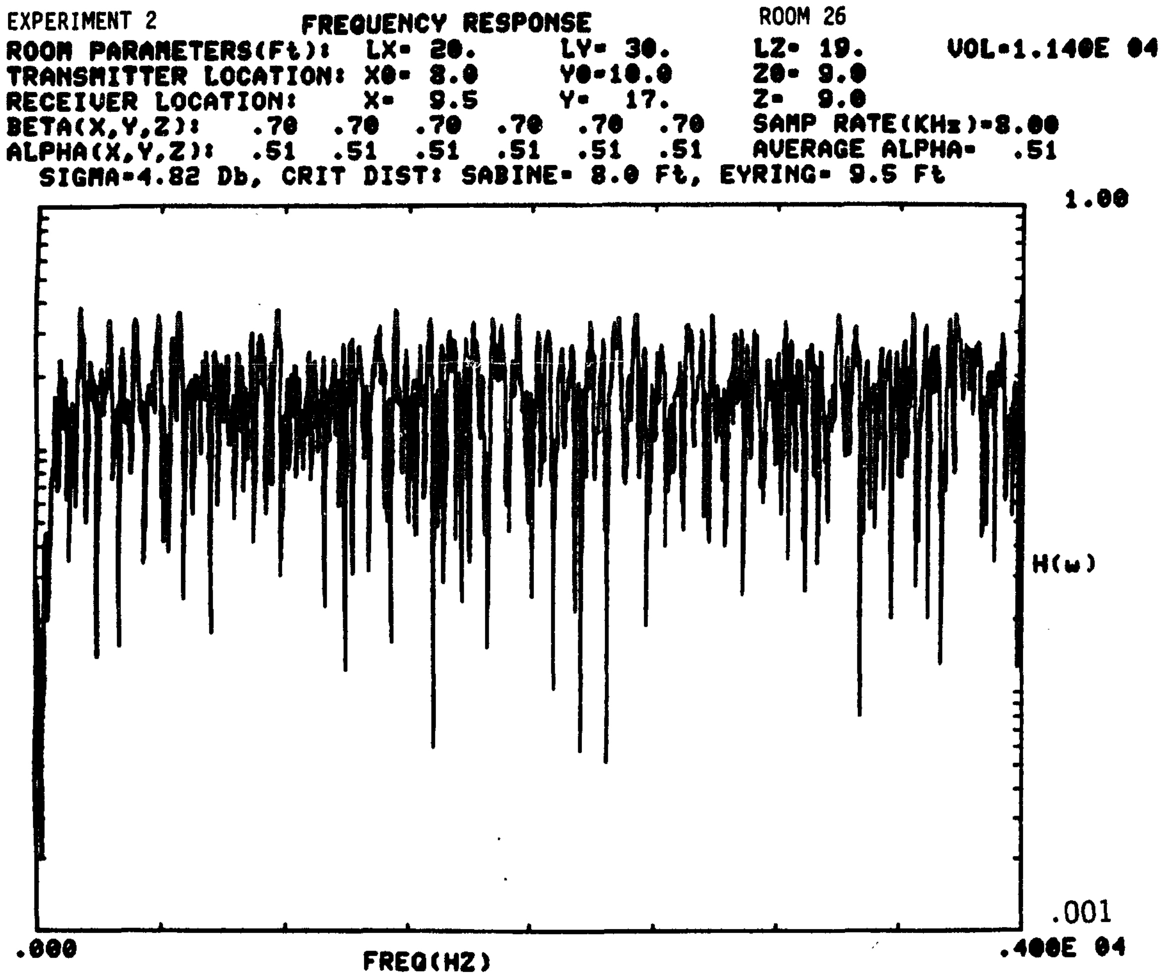


FIGURE 3.22


EXPERIMENT 2
FREQUENCY RESPONSE
ROOM 26
ROOM PARAMETERS(Ft): LX= 20. LY= 30. LZ= 19. VOL=1.140E 04
TRANSMITTER LOCATION: X0= 8.0 Y0=10.0 Z0= 9.0
RECEIVER LOCATION: X= 9.5 Y= 17. Z= 9.0
BETA(X,Y,Z): .70 .70 .70 .70 .70 .70 SAMP RATE(KHz)=8.00
ALPHA(X,Y,Z): .51 .51 .51 .51 .51 .51 AVERAGE ALPHA= .51
SIGMA=4.82 Db, CRIT DIST: SABINE= 8.0 Ft, EYRING= 9.5 Ft
1.00
H(w)
.001
.000
FREQ(HZ)
.400E 04

Several of the stimulus sets have the same talker-microphone distance. Sets 1, 3, 5, and 7 all have $\Delta = 5$ with Sets 1 and 3 using the same room volume and differing only with respect to where the source and microphone are placed in the room. Sets 1 and 3 have the same parameters in Table 3.4, but the source and receiver are located differently in the two stimulus sets. Also, Stimulus Sets 4 and 8 both have $\Delta = 7.2$ feet. Rooms with the same Δ should cluster together, or show some other similarity if the results of the first experiment hold.

Three room sizes were used, one smaller, one the same, and one larger than in the previous study. In increasing order of size, the three rooms' dimensions are 9 by 12 by 10, 12.5 by 15 by 16.25, and 20 by 30 by 19 feet.

In Stimulus Set 8, one of the sets from a large room, the source and receiver are located sufficiently far from all six surfaces so that no reflections reach the receiver in the first 10 ms. This is a direct test of σ_{10} since these stimuli should not be perceived differently from Stimulus 1 in Set 1 or from each other if coloration perception is produced only by the first 10 msec of reflections. Note that Stimuli 25 and 26 (Set 8) have $\sigma = 1.93$ and 4.22, respectively, well above the $\sigma = .07$ for

Stimulus 1.

Subjective criteria entered into stimulus selection. The stimulus with the lowest β in each stimulus set (and Stimulus 2 in Set 1) were all perceived by an experienced listener as different from Stimulus 1, (including Set 8, which is the test of σ_{10}). Also, for each set of stimuli, consecutive rooms, e.g., Stimuli 2 and 3, 3 and 4, and 4 and 5, were perceptibly different. And, finally, the stimuli with the highest β in each set were all chosen to be different from each other. If the results of Experiment 1 were interpreted correctly, this should create a perceptually non-redundant stimulus space.

3.3.2 Experiment 2 Methods

This experiment used the same methodology and method of analysis as Experiment 1. Subjects heard all pairs of a matrix of colored noise stimuli, judged how different the two members of a pair were, and the normalized, averaged results were submitted to the KYST-2 non-metric multidimensional scaling program. The methodology for Experiment 2 was the same as that of the first experiment except for the following differences.

Subjects. Subjects who had been in the first experiment were not used in this experiment. Forty-five subjects were run in 8 groups of 4-6. Data from one of these subjects was not used because, as before, the range of judgments was below the specified criterion.

Noise Stimuli. A single noise sample was used, and it was one of those used in the first study. All pairs of the 26 stimuli were used in this study; that is, both orders of stimulus presentation and the comparison of each stimulus with itself. A single random order of the 676 stimuli was recorded on 4 analog tapes, 169 pairs on each tape. The order of presentation of the 4 tapes was counterbalanced across each 4 groups of subjects. Each subject heard the whole stimulus set of 676 noise pairs only once.

Data Analysis. The data were normalized and averaged within and then across subjects as before. As a reliability check, a aplit-half technique was used in which the data for the first four groups (N = 23) and the last four groups (N = 21) of subjects were analyzed separately. The two solutions were found to be similar to each other and to the final solution obtained by averaging.

3.3.3 Experiment 2 Results

Table 3.5 contains the averaged difference judgments for all the subjects, stimuli, and stimulus orders in the experiment. Both of the split-halves solutions, and the solution for the whole data set, have configurations in which the different room sets seem to emanate from a central area somewhat like spokes of a wheel. (See Figures 3.23 and 3.24 for the two-dimensional split halves solutions. Figure 3.25 to 3.28 display the two- and six-dimensional solutions for the whole data set, and Figure 3.29 displays the stress functions for the solutions for all these data sets.)

In all of these solutions, a central point is represented by the least colored stimulus (1); stimuli from each set are on a single spoke, and as the distance from the center increases for each spoke, the β of the stimulus also increases. This solution is consistent with the view of coloration proposed from the last experiment, but in this solution quantity of coloration (distance from stimulus 1) does not create a dimension that obviously matches one of the solution dimensions. Nonetheless, each stimulus set displays it: distance along a spoke from the center appears to represent quantity of coloration, while the different spokes for the different rooms represent different coloration qualities.

Table 3.5 Data from Experiment 2. Averaged Difference Judgments Used as Input to Multidimensional Scaling Analysis.

Stim	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	26
2	2.4												
3	2.1	2.1											
4	3.4	1.9	2.2										
5	5.2	3.9	2.9	2.2									
6	2.4	2.7	3.0	3.3	4.6								
7	2.8	3.8	3.9	3.9	4.1	1.6							
8	6.1	6.0	5.8	5.4	4.1	3.4	2.5						
9	2.5	1.7	2.5	2.9	4.0	2.7	3.1	5.6					
10	3.4	2.9	2.7	3.5	4.9	2.7	3.5	5.2	1.8				
11	3.8	3.6	2.8	3.4	4.4	3.5	3.0	6.3	1.8	1.4			
12	1.9	1.9	2.8	3.5	5.2	2.5	3.3	6.0	2.0	2.4	3.4		
13	3.3	2.7	3.5	4.2	5.2	4.0	4.2	6.7	3.1	3.7	3.7	1.7	
14	4.4	4.7	4.5	4.2	4.6	4.2	4.9	6.4	5.2	5.1	5.2	3.9	2.0
15	2.1	2.1	2.7	3.3	4.4	2.5	3.8	5.7	3.0	3.0	3.4	2.5	3.7
	5.4												
16	3.0	2.5	3.7	3.7	5.4	3.9	4.0	6.0	3.4	4.7	3.8	3.5	4.2
	5.1	1.9											
17	3.2	4.0	3.4	5.0	5.0	3.6	4.5	6.0	3.3	4.0	4.3	3.7	4.7
	6.0	2.1	1.7										
18	3.3	4.1	4.4	4.6	5.7	4.1	4.5	6.6	4.2	4.2	4.3	3.8	4.9
	5.7	3.0	2.6	2.3									
19	2.0	2.3	2.5	2.9	4.4	2.3	3.3	5.9	2.7	3.0	3.3	2.2	2.6
	4.3	2.5	2.9	3.6	4.1								
20	2.6	2.5	2.3	3.4	4.5	2.7	3.4	6.0	2.5	3.0	3.4	2.0	3.4
	4.4	2.6	4.1	4.0	4.4	2.0							
21	3.6	3.1	2.7	3.7	4.0	3.4	3.7	5.4	3.4	3.9	3.3	2.6	3.4
	4.2	3.4	4.0	4.8	3.7	3.0	1.7						
22	2.3	1.8	2.1	2.7	4.5	2.4	2.4	5.7	2.8	2.5	4.0	2.7	3.5
	4.5	2.2	4.0	3.5	4.3	2.4	2.7	2.9					
23	2.9	2.7	2.8	2.6	4.4	2.1	3.0	5.1	2.3	4.3	4.4	2.9	4.0
	4.6	3.2	3.7	3.8	4.3	2.7	3.1	3.4	1.6				
24	3.6	3.3	3.4	3.9	4.0	3.3	3.2	4.6	3.9	4.2	4.4	4.6	4.9
	5.4	4.1	4.5	4.4	5.4	4.0	3.7	4.4	2.4	1.8			
25	2.0	2.0	2.5	3.3	5.1	2.5	3.6	5.6	2.4	2.9	3.4	1.9	3.2
	4.6	2.3	3.3	2.8	3.4	2.4	3.0	3.3	3.2	2.9	4.5		
26	2.4	2.9	2.9	3.7	5.7	2.9	3.8	5.3	3.3	3.5	4.0	2.6	3.4
	4.5	3.0	3.2	3.2	3.9	3.2	3.7	3.9	3.6	3.5	4.4	2.1	

FIGURE 3.23

EXPERIMENT 2: TWO-DIMENSIONAL SOLUTION
FOR THE FIRST HALF OF THE SUBJECTS (23 S'S).

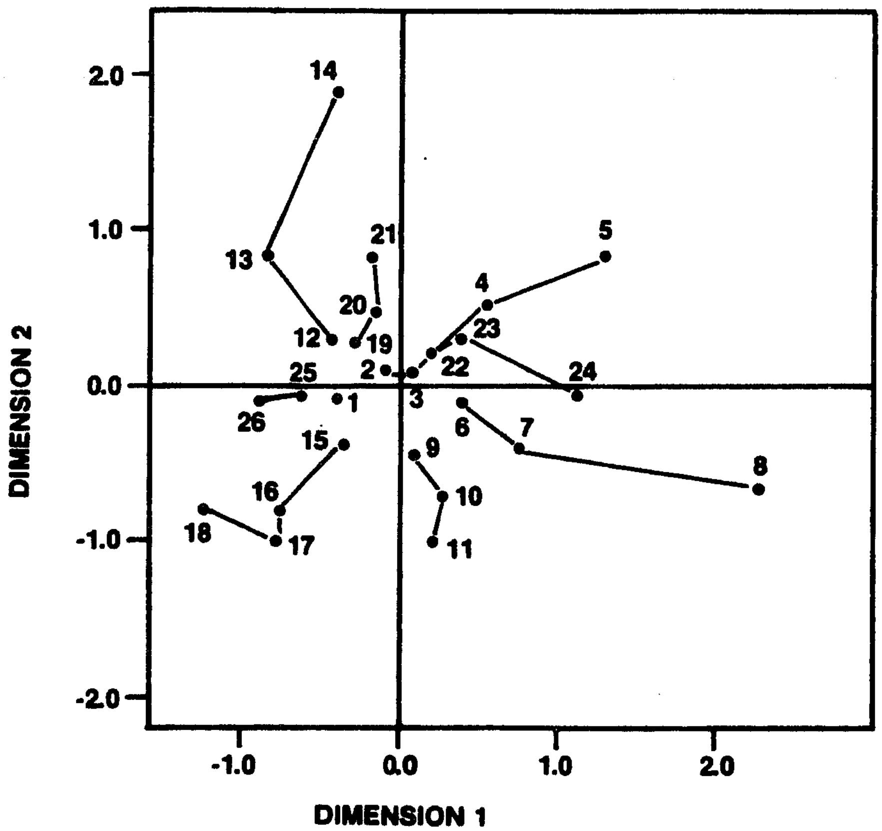


FIGURE 3.24

EXPERIMENT 2: TWO-DIMENSIONAL SOLUTION FOR THE SECOND HALF OF THE SUBJECTS (21 S'S).

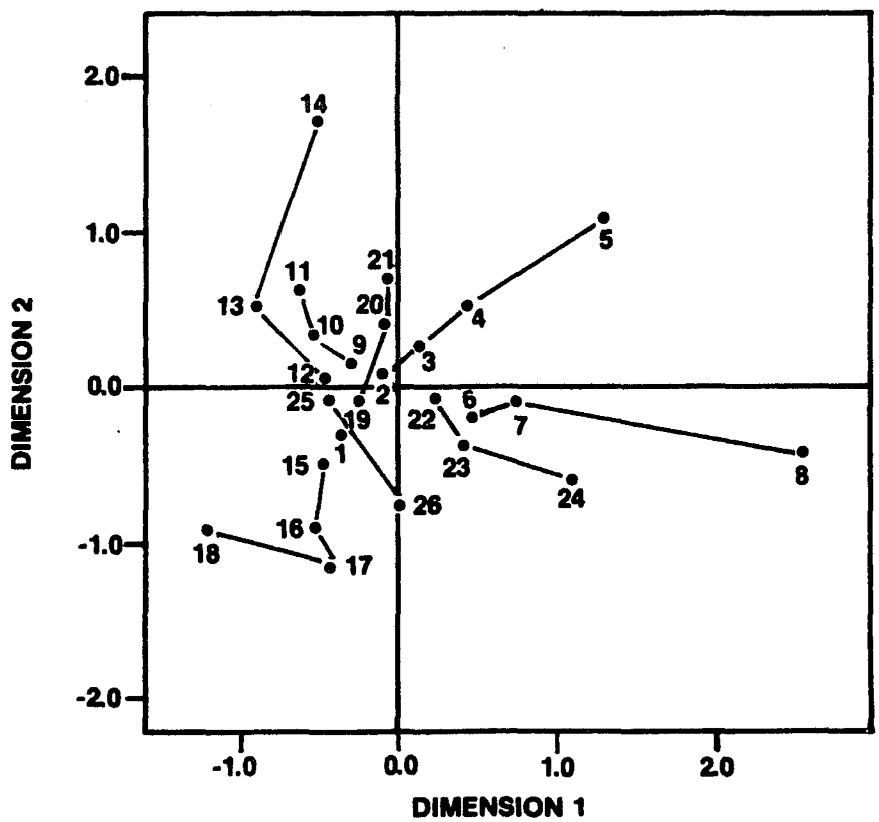


FIGURE 3.25

EXPERIMENT 2: TWO-DIMENSIONAL SOLUTION FOR WHOLE DATA SET.

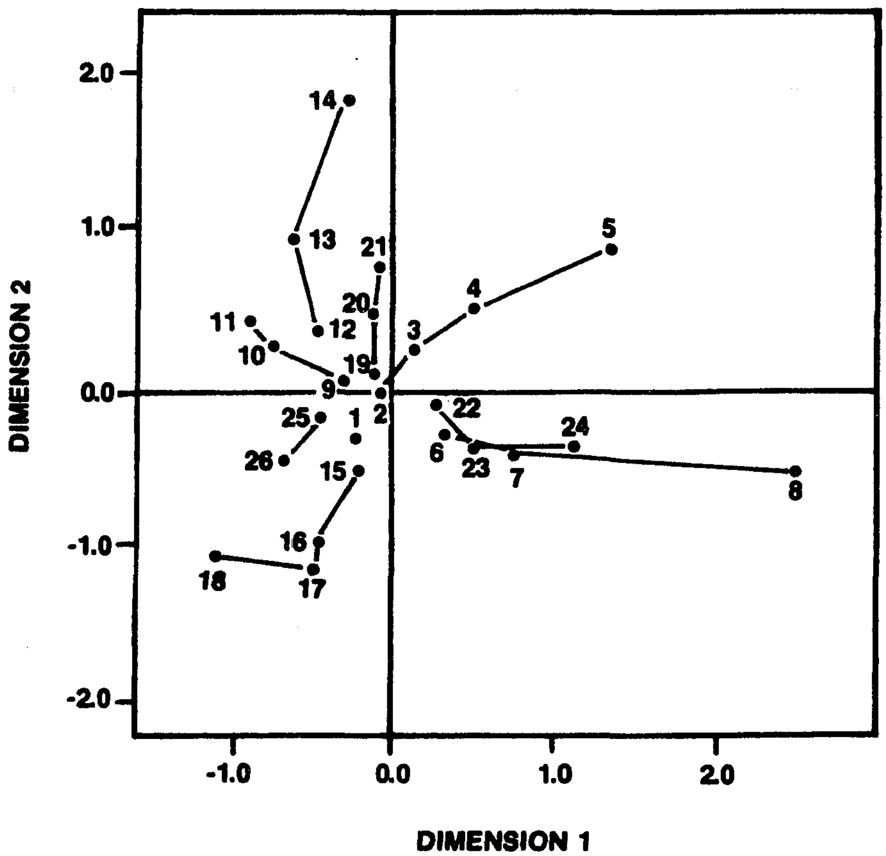


FIGURE 3.26

EXPERIMENT 2: DIMENSIONS 1 AND 2 OF SIX-DIMENSIONAL SOLUTION FOR WHOLE DATA SET.

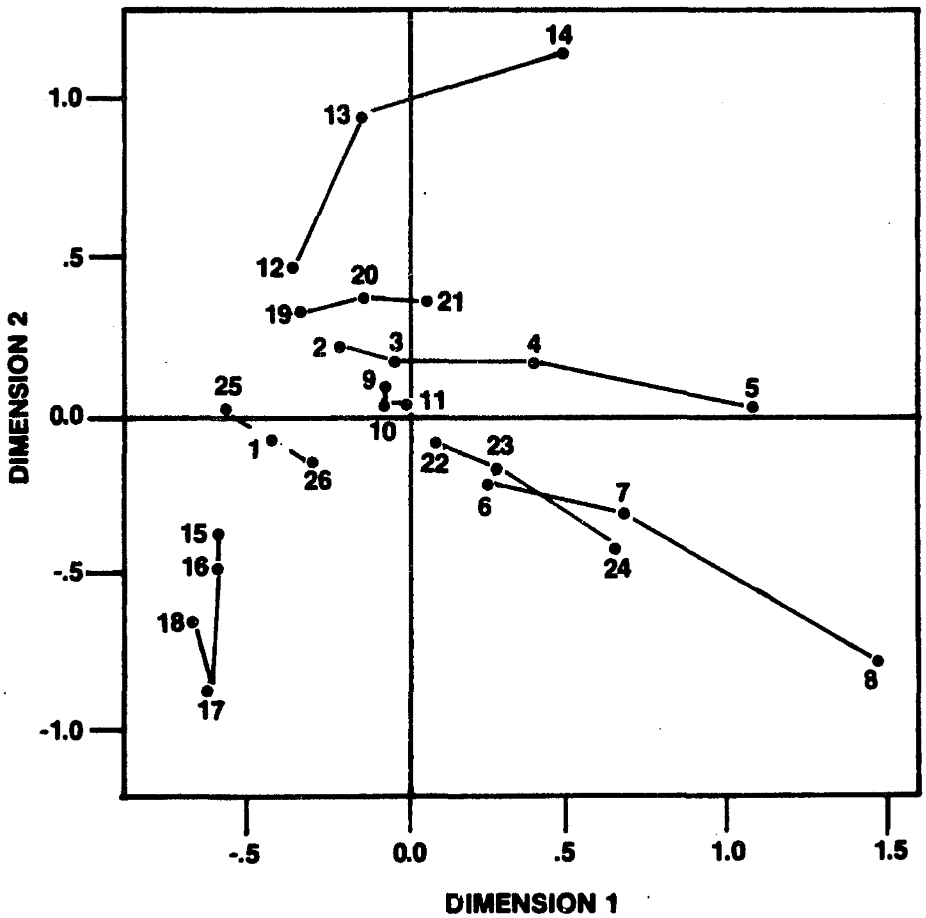


FIGURE 3.27

EXPERIMENT 2: DIMENSIONS 3 AND 4 OF SIX-DIMENSIONAL SOLUTION FOR WHOLE DATA SET.

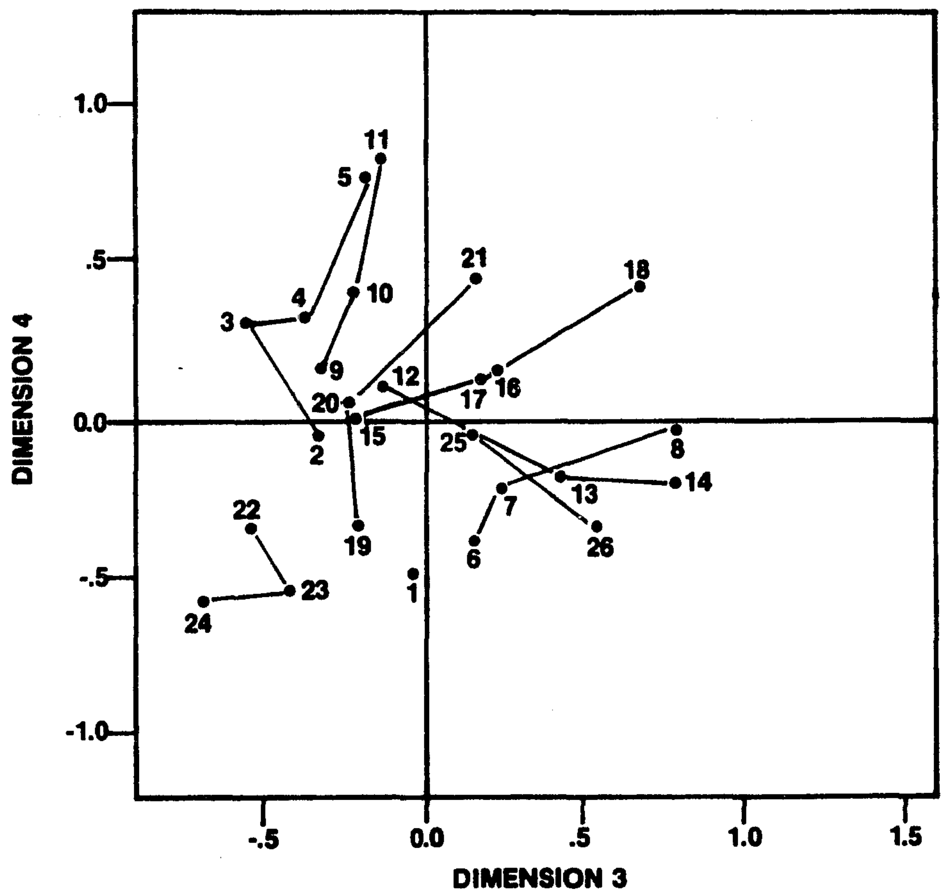


FIGURE 3.28

EYPERIMENT 2: DIMENSIONS 5 AND 6 OF SIX-DIMENSIONAL SOLUTION FOR WHOLE DATA SET.

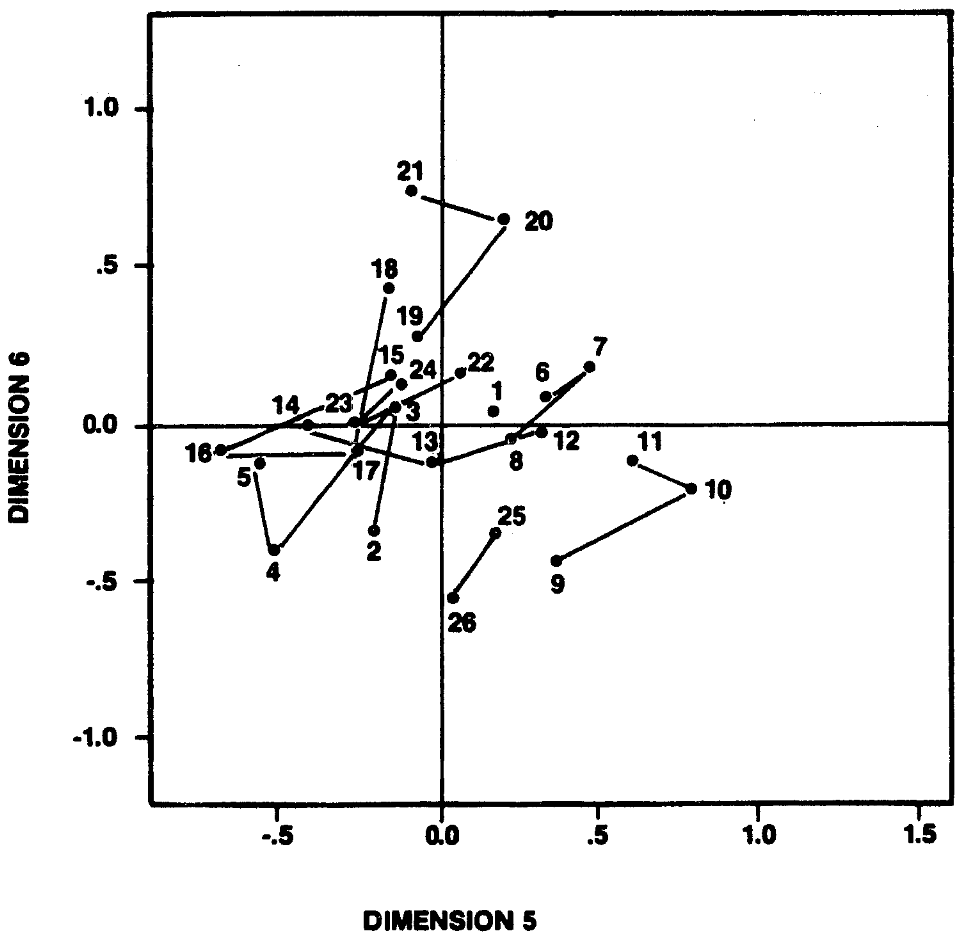
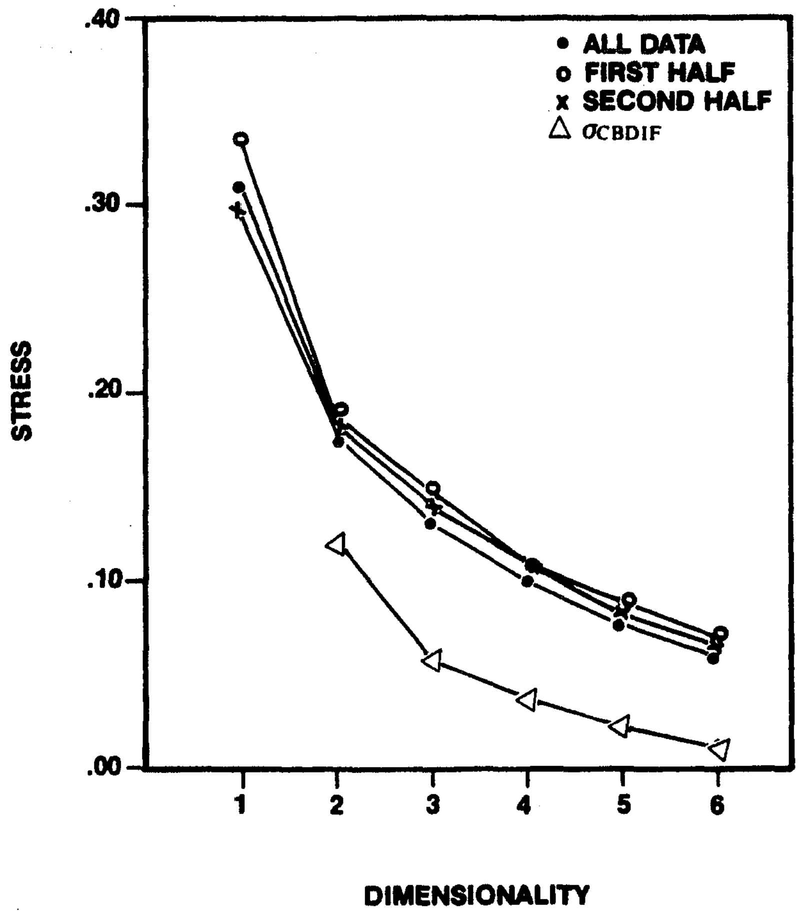


FIGURE 3.29

EXPERIMENT 2: STRESS FUNCTIONS FROM MULTI-DIMENSIONAL SOLUTIONS.



Corresponding dimensions in the split-halves solutions were correlated, as in Experiment 1. With 26 pairs of points, $r \geq .39$ is needed for correlation significantly greater than zero at the .05 level. For the two dimensional split-half solutions, $r = .91$ and .72 for Dimensions 1 and 2, respectively. Information about the dimensionality of the solution may be available from the extent to which reliability decreases with dimensions. Accordingly, the correlations for the corresponding dimensions of the six-dimensional solution were determined: $r = .89$, .76, .74, .61, .58, and .75, for Dimensions 1-6, respectively. These correlations are all high enough to suggest that each dimension represents reliably some proportion of the data variability, and low enough to suggest that error variance is high. They suggest that an effort must be made to account for a space of high dimensionality.

Visual inspection of the split-halves analyses in Figures 3.23 and 3.24 suggests commonalities in the solutions. Room Sets 1 (Stimuli 2, 3, 4, and 5), 2 (Stimuli 6, 7, and 8), 4 (Stimuli 12, 13, and 14), and 5 (Stimuli 15, 16, 17, and 18) have the lengthiest spokes and they are in about the same configuration in the two solutions. The spoke for Stimulus Set 3 (Stimuli 9, 10 and 11) differs in location in the two solutions. The difference in location involves the direction of the spoke representing these stimuli. It is supportive of the spoke interpretation while indicating variation in how subjects perceive this stimulus set relative to the others. Stimulus Set 6 (Stimuli 19, 20, and 21) is in the same relative position in the two solutions. Sets 2 (Stimuli 6, 7, and 8) and 7 (Stimuli 22, 23, and 24) reverse position relative to each other on Dimension 2 while remaining in mostly the same configuration relative to the other stimuli. Set 8 (Stimuli 25 and 26) with $\sigma_{10} \approx 0$ is fairly close to Stimulus 1 in both cases, suggesting low, if not zero, coloration.

The stress function for each of the split-half solutions and for the whole solution are plotted in Figure 3.29 for solutions from 1 to 6 dimensions. The stress functions drop rapidly at first as the dimensionality of the solution is increased, but they do not flatten out, nor are there sharp elbows in the higher dimensional solutions. It is not clear from the stress functions whether there are two dimensions to the solution with error variance being accounted for as dimensions are added or whether each added dimension accounts for some additional quantum of perception. It is the correlations between corresponding dimensions, as we have seen, that indicate each of the six dimensions account for more than error variance. Stress for the six-dimensional solution is .Ø58, a level termed "good" by Kruskal (1964). The stress for the two-dimensional solution is .176 and, by Kruskal's standards, this does not account for a respectably high amount of the variability in the data. These findings, in agreement with the reliability measures, suggest a space of high dimensionality.

The six-dimensional solution, as summarized in Figures 3.26, 3.27, and 3.28, plots Dimensions 1 and 2, 3 and 4, and 5 and 6, respectively. With a few exceptions, the spoke interpretation holds up well for Dimensions 1, 2, 3, and 4. Stimulus Set 3 (Stimuli 9, 1Ø, and 11) and 8 (Stimuli 25 and 26) are not spoke-like in Dimensions 1 and 2, but Stimulus 1 and Stimuli 9, 1Ø, and 11 run across most of the range of Dimension 4. Stimulus Set 8 is in the expected order in Dimensions 4 and 6, but these stimuli are always fairly close to the uncolored stimulus, Stimulus 1, an indication in six dimensions that what is perceived as coloration is produced by early reflections, mostly within the first 1Ø msec.

Aside from the spokes as an indicator of quantity of coloration, what else is to be learned from the solution? Consider some of the variables that were introduced into the design of this experiment: Room Sets 1, 3, 5, and 7 all have the same source-microphone distance, Δ = 5. Stimulus Sets 4 and 8 have also the same Δ, 7.2 feet. There are no obvious similarities or groupings based on common Δ's within the solution space. While Δ will be further examined in the following regression analyses, this result suggests that the effect of Δ in the earlier experiment may have been an artifact based on the small number of rooms sampled.

Three room sizes were used. Stimulus Sets 5 and 6 are from the smallest room, Sets 1, 2, 3, and 4 are from the middle room in size, and Sets 7 and 8 are from the largest room. No obvious commonalities based on size appear.

This lack of consistent effects due to the physical room variables, Δ and room size, are not surprising given an interpretation of coloration based on early reflections. Any room size and any Δ could produce coloration if the source and/or microphone is close enough to a wall.

These findings do not account for the solution space. Let us turn to the analytical methods. Two approaches will be explored. The multiple linear regression analysis used in the first experiment can be applied here also. While Dimension 1 of this solution doesn't seem to agree with coloration quantity the way it did in the first experiment, there may be a non-obvious coloration quantity dimension. A second approach is to use the distance of the other stimuli from Stimulus 1 more directly; these distances can be calculated and correlated with the previously developed coloration measures.

Table 3.6 presents the correlations between various room parameters and coloration measures and their multiple linear regression into the three and six dimensional solutions, respectively. The correlations are low for the low dimensional solution. Even in six dimensions only one of the coloration measures, σ_{CB}, based on a sharply peaked critical band-like filter, projects into the space with a high correlation (r = .95). Figures 3.30-3.32 show the projection of a few of the variables into the six-dimensional space. Again the coloration measures show considerable similarity in their projections although σ_{CB} regresses best into the space. The intercorrelations among all the σ-type measures are generally .9 or better. This occurs despite a successful effort to create a less clustered stimulus space, and is an indication that these measures are very similar. Reverberation time and β regress poorly, as in Experiment 1 and as expected.

Table 3.6. Correlations between Room Variables and Experiment Solution Spaces of 3 and 6 Dimensions

	Variable	3 Dim	6 Dim
1.	T_{60}	.35	.53
2.	Δ	.71	.86
3.	β	.36	.60
4.	σ	.34	.75
5.	σ_{10}	.48	.87
6.	σ_{20}	.39	.79
7.	σ(Rect. 47 Hz filter)	.29	.82
8.	σ(Rect. 140 Hz filter)	.36	.86
9.	σ_{CB}, exponential filter	.54	.95
10.	σ[Hamming window]	.30	.86

FIGURE 3.30

EXPERIMENT 2: PROJECTIONS OF SOME ROOM VARIABLE VECTORS INTO DIMENSIONS 1 AND 2 THE SIX-DIMENSIONAL SPACE.

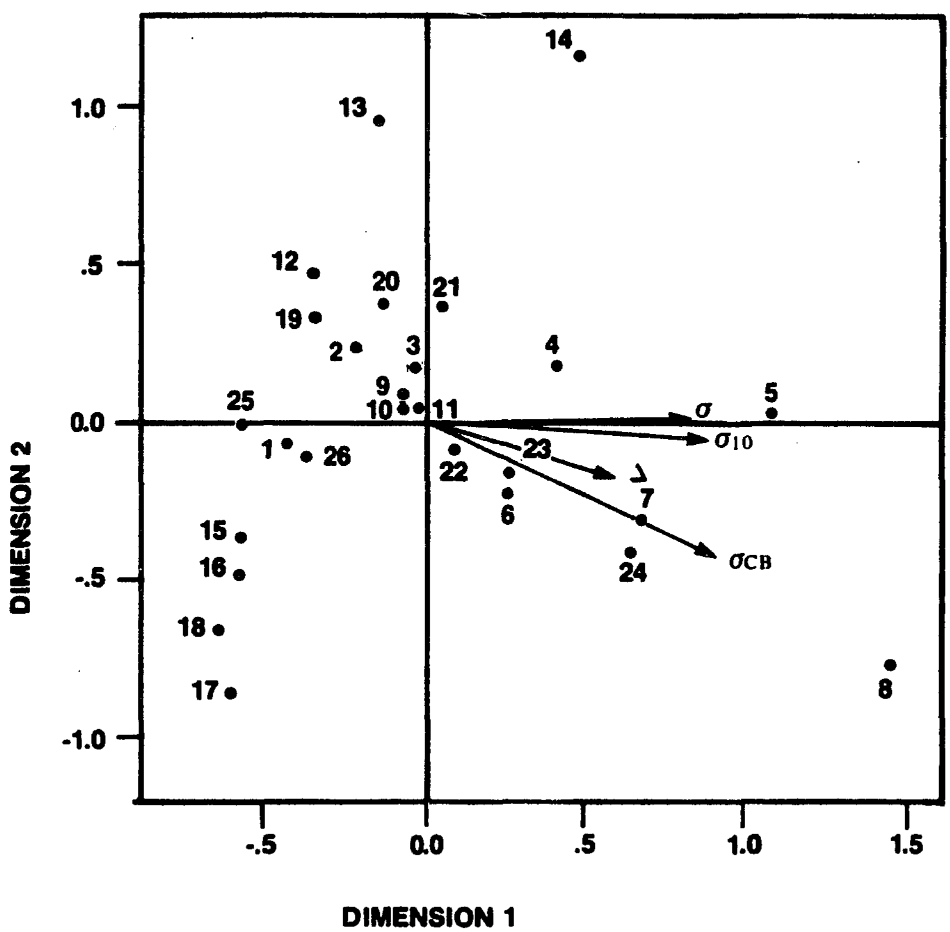


FIGURE 3.31

EXPERIMENT 2: PROJECTIONS OF SOME ROOM VARIABLE VECTORS INTO DIMENSIONS 3 AND 4 OF THE SIX-DIMENSIONAL SPACE.

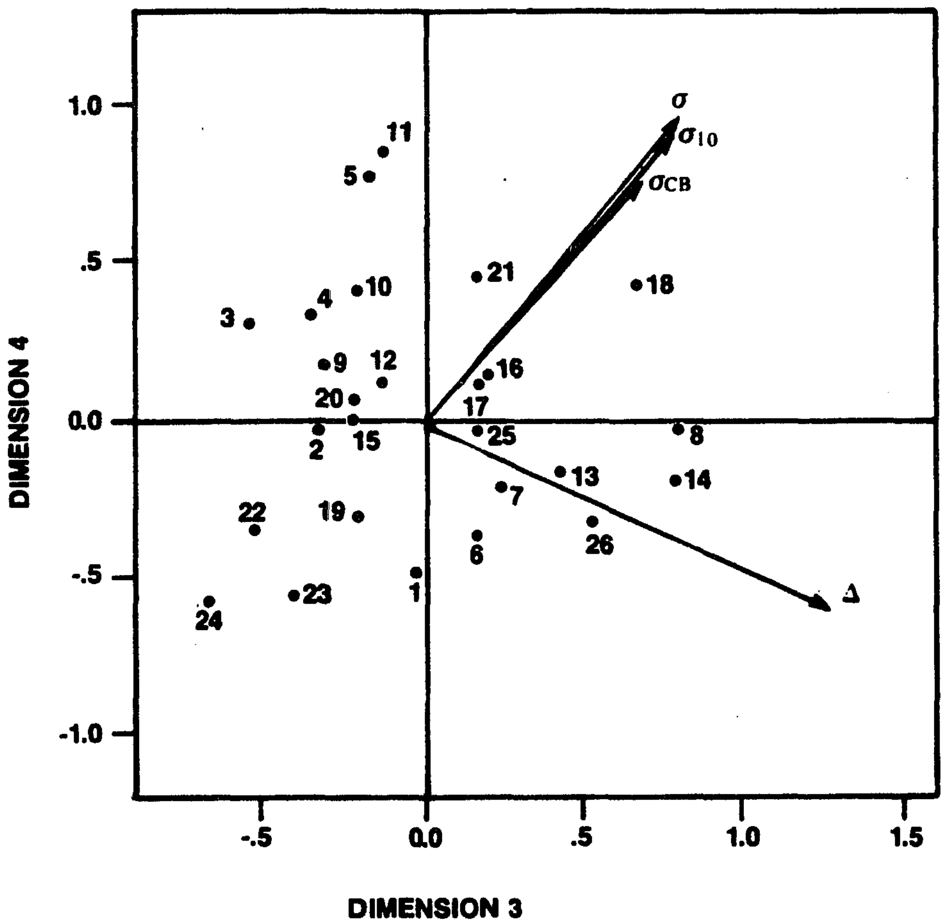
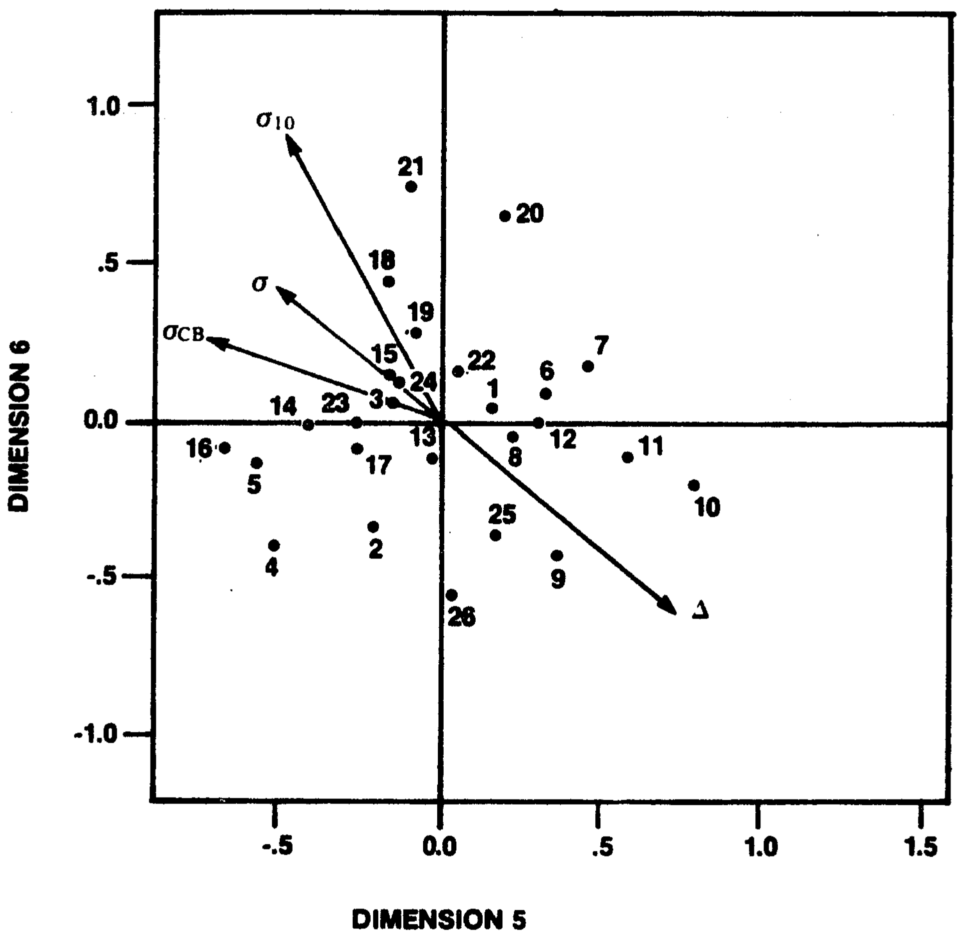


FIGURE 3.32

EXPERIMENT 2: PROJECTIONS OF SOME ROOM VARIABLE VECTORS INTO DIMENSIONS 5 AND 6 OF THE SIX-DIMENSIONAL SPACE.



Δ regresses best of all the variables in the three dimensional space although this does not hold up in higher dimensions. Also, the Δ vector does not agree with the coloration vectors as it did in the previous experiment. With only four possible values for Δ, inspection of the solutions suggests that the spoke-effect may permit a spuriously high correlation due to Δ by grouping stimuli on a vector when that vector cuts a spoke perpendicularly. This appears to happen for Stimuli 6, 7, and 8 in Figures 3.31 and 3.32 and for Stimuli 25 and 26 in Figure 3.32.

These results suggest there may be a coloration quantity dimension within the six dimensional solution, and that σ_{CB} may be the best of these measures. Given the spoke interpretation for quantity of coloration reached by first looking at the data, this suggests that there is a dimension in the space in which all the spokes go in the same direction. Since the spokes also radiate in many directions from a point with no coloration, this suggests a space in which spokes are structured like a multidimensional umbrella. The σ_{CB} vector in Figures 3.30-32 represents the best estimate of the coloration quantity dimension, but it is difficult to be satisfied with this interpretation when viewing these data.

The next analytical approach is to calculate the Euclidean distance from Stimulus 1 for the other stimuli in some of the solution spaces, and then determine the correlation of these distances with the various room and coloration measures. The results are presented in Table 3.7.

Table 3.7. Correlations between Room Variables and the Distance of the Stimuli from Stimulus 1 in the 3- and 6-Dimensional Solution Spaces (Experiment 2).

Variable	3 Dim	6 Dim
1. Δ	.58	.61
2. β	.69	.73
3. σ	.80	.84
4. σ_{10}	.91	.89
5. σ_{20}	.79	.82
6. σ(Rect. 47 Hz filter)	.79	.81
7. σ (Rect. 140 Hz filter)	.84	.86
8. σ_{CB}, exponential filter	.90	.92
9. σ[Hamming window]	.84	.85

The spoke interpretation is supported by these data: the coloration measures correlate fairly highly with distance from Stimulus 1, and this correlation is roughly constant for each measure regardless of the dimensionality of the solution. This is strong evidence for the importance of distance from the uncolored stimulus as a measure of quantity of coloration. The two measures providing the highest correlations are σ_{10} and σ_{CB}. These correlations are about .9.

It is hard to know how to compare the two sets of correlations, i.e., the ones from the projections into the space and the others based on the distance from Stimulus 1. For the latter, note that any error in the location of Stimulus 1 will introduce error into the whole set of distances. The relative orderings of the coloration measures in each set of correlations agree; σ_{CB} and σ_{10} are best in both, and the two analyses yield very similar results for such disparate approaches.

For the distance data, measures such as β and Δ correlate poorly although β correlates better with the distance measures than does Δ, a reversal from the projection data. This is because β is monotonic with coloration for each room set, while Δ will give the same value to all the stimuli on a spoke.

So far, the data from Experiments 1 and 2 do not permit selection of the best measure of coloration; both σ_{CB} and σ_{10} account adequately for a dimension of the data. It is also not clear to what extent the two interpretations of the data, i.e., amount of coloration as spokes or as a projected vector, are the same.

Still larger questions remain. Coloration is multidimensional and coloration quantity accounts for only one of the dimensions of the solution. Do the other dimensions represent quality of coloration? What accounts for the configuration of the stimuli in the multidimensional space? In order to answer these questions let us assume that the output of the peripheral hearing process is, in fact, one of the smoothed spectra represented by the coloration measures. The subjects were asked to rate how different each pair of stimuli were; let's assume they did just that. That is, they compare the two smoothed noise spectra, point by point, over the frequency continuum, and then take some sort of average of this set of differences in order to create their difference judgment. A statistic used for just this sort of evaluation is the standard deviation of the differences,

$$\sigma_{DIF} = \left[\frac{\sum_i (X_{DIF} - \bar{X}_{DIF})^2}{N} \right]^{1/2}$$

where the X_{DIF} represent the difference in levels of the smoothed frequency spectra at each point. This statistic is independent of the mean of the differences between the two continua. For these data, the 4000 Hz frequency range is sampled 1024 times with 795 of these points between 100 and 3200 Hz, the range of calculation used for the smoothed spectra; thus N = 795. The differences are the differences in amplitude in dB at each of the 795 frequency samples.

Thus σ_{DIF} represents a comparison of two frequency spectra and is the calculation of a measure of the difference between them. Using the σ_{CB} filter, this statistic, termed σ_{CBDIF}, was calculated for all pairs of stimuli, i.e. frequency responses, in the experiment. There is also a direct measure of the perceptual difference between each pair of stimuli, the 325 mean difference ratings that are the input to the multidimensional scaling analysis. The correlation between the 325 difference data points and the 325 σ_{CBDIF} measures is .88.

The set of σ_{CBDIF}'s were next analyzed the same way the difference data were, i.e., by doing a multidimensional scaling analysis of them. The two-dimensional solution for σ_{CBDIF} is presented in Figure 3.33, and the six-dimensional solution is presented in

Figures 3.34-3.36. Corresponding figures for the subjective data are 3.26-3.28, respectively. First, in Figure 3.29, note that the stress function for σ_{CBDIF} has considerably lower values than the stress function associated with the subjects' data; this is to be expected since σ_{CBDIF} represents theoretical values without the presence of subject error. The σ_{CBDIF} stress function, like the others, declines gradually, without a clear indication of when the "true" dimensionality is reached. The spoke-like phenomenon is even more apparent for this analysis than it is for the subjective data. It dominates every dimension of even the six-dimensional solution.

Now, let us compare the solutions for the model and the data. These can be compared in the same way reliability was evaluated: corresponding dimensions of the model and data solutions can be correlated. Again, for 26 pairs of points, $r \geq .39$ is needed for significance at the .05 level. For the two-dimensional solutions, $r = .86$ and .90, respectively. For the six-dimensional solution, r = .95, .92, .83, .81, .85, and .81, respectively. These are high correlations, especially in light of the reliability statistics, and suggest that σ_{CBDIF} accounts very well for the part of the data that is not error variance.

FIGURE 3.33

EXPERIMENT 2: TWO-DIMENSIONAL SOLUTION FOR σ_{CBDIF}. BLOW-UP ENLARGES CENTER OF GRAPH.

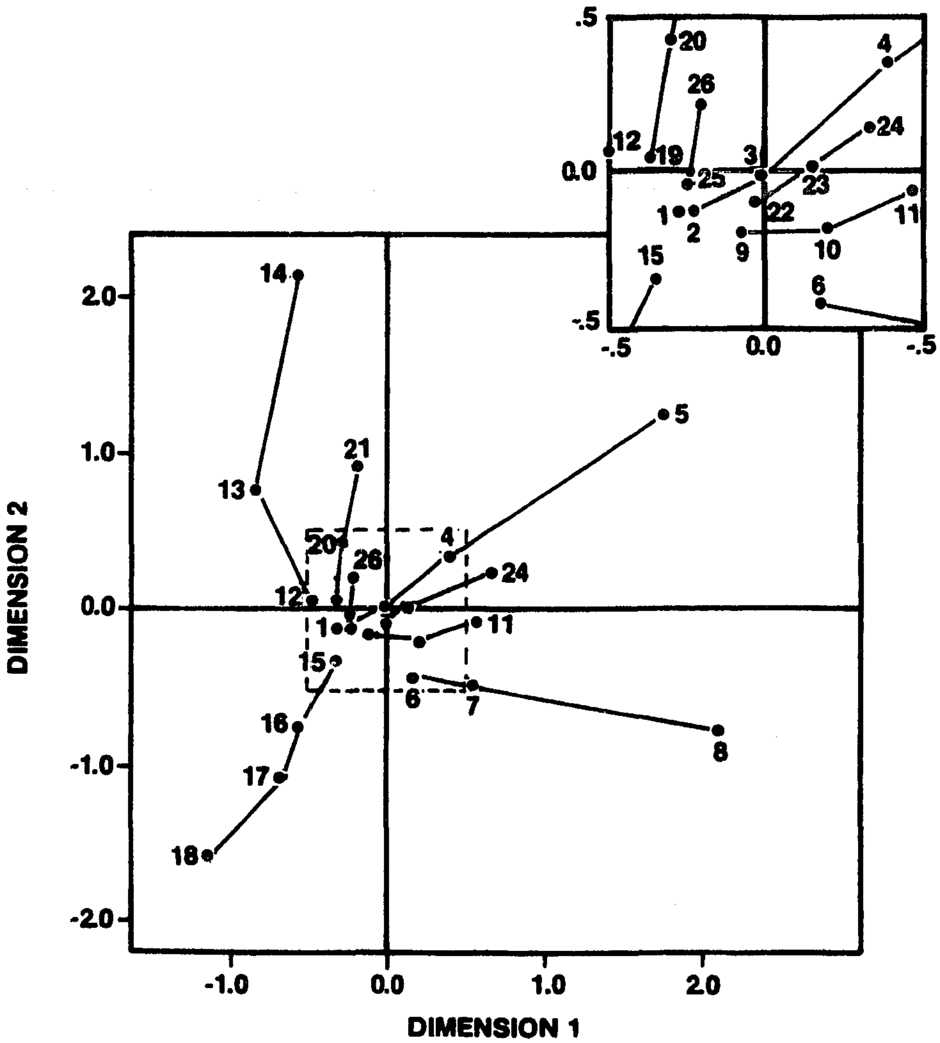


FIGURE 3.34

EXPERIMENT 2: DIMENSIONS 1 AND 2 OF SIX-DIMENSIONAL SOLUTION FOR σ_{CBDIF} .

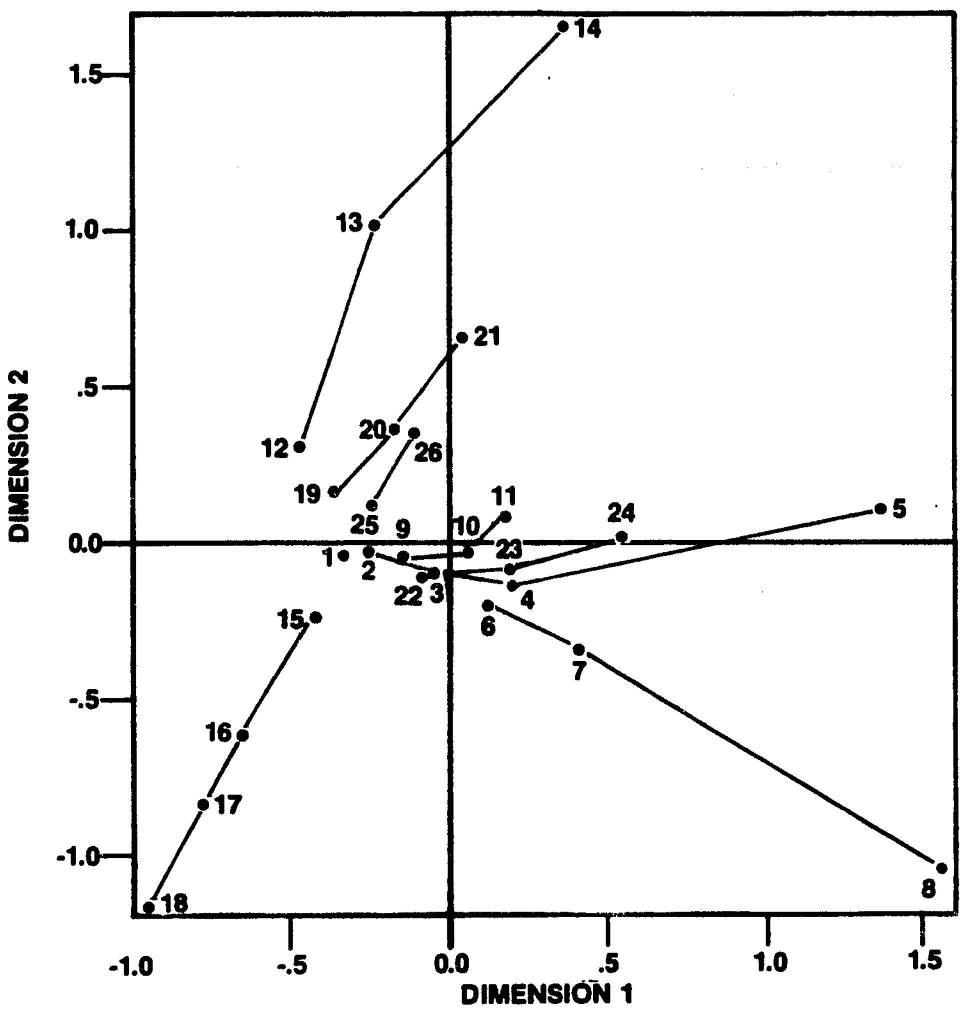


FIGURE 3.35

EXPERIMENT 2: DIMENSIONS 3 AND 4 OF SIX-DIMENSIONAL SOLUTION FOR σ_{CBDIF} .

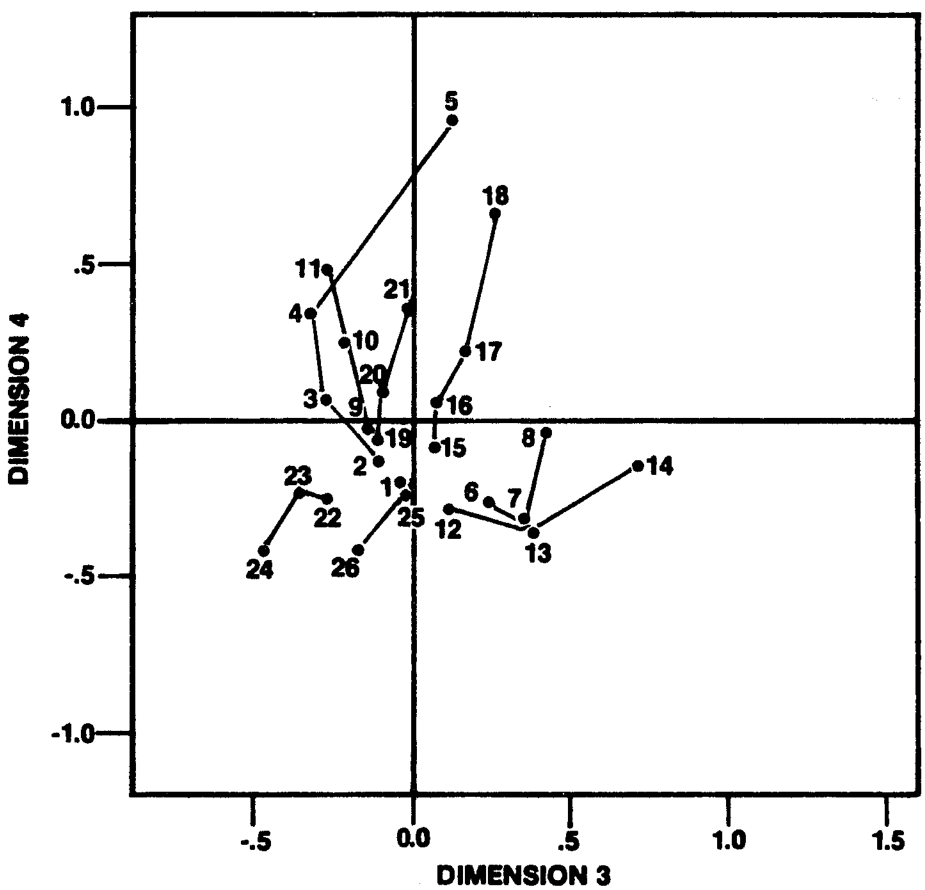
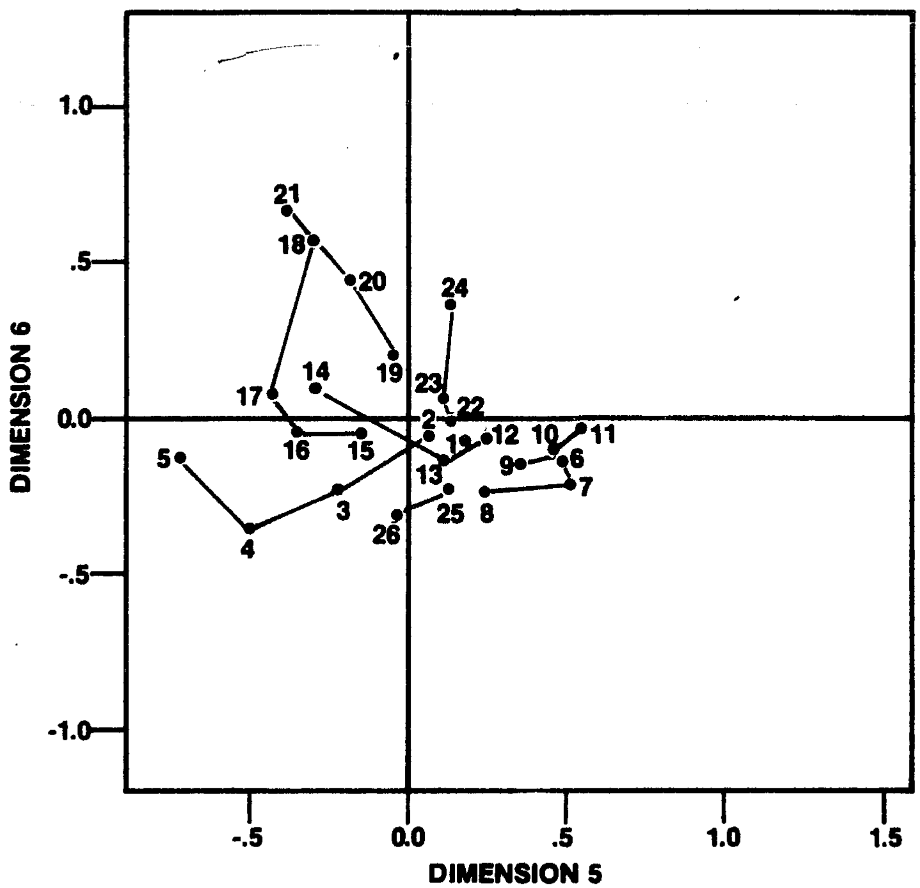


FIGURE 3.36

EXPERIMENT 2: DIMENSIONS 5 AND 6 OF SIX-DIMENSIONAL SOLUTION FOR σ_{CBDIF} .



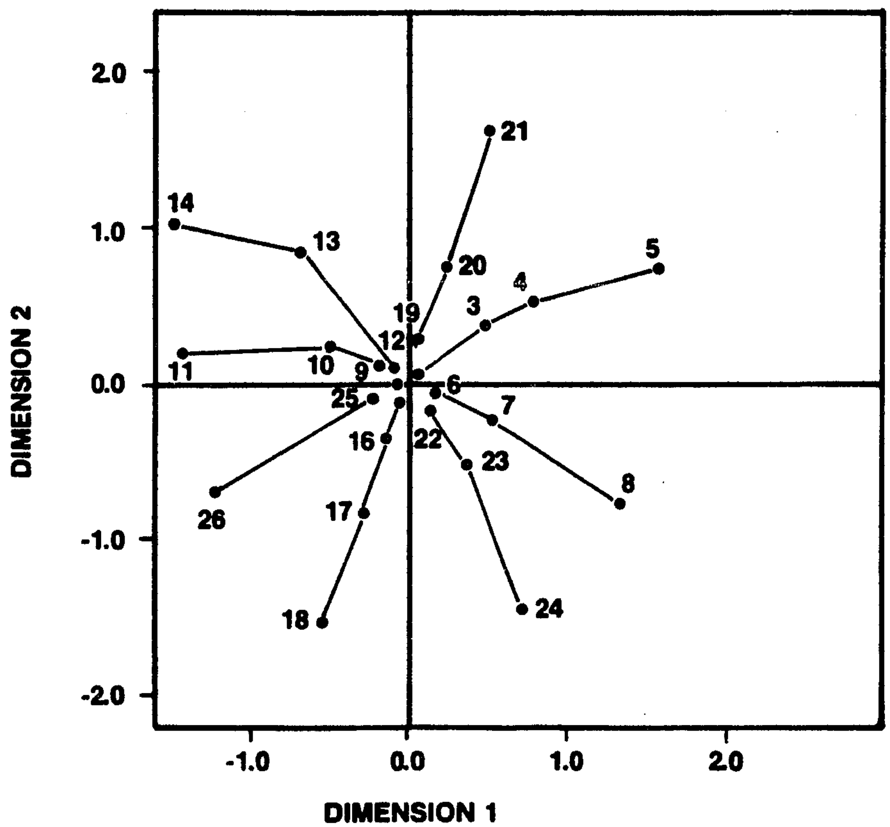
For both the two-dimensional and six dimensional solutions, there is striking qualitative similarity between the data and model solutions. This is especially true for the two-dimensional solution and the first two dimensions of the six-dimensional solution. The same four long spokes are the main determiners of both solutions, and there is fairly good agreement among the shorter spokes as well. As we move into the comparison of the higher dimensions of the solution, the agreement between model and data diminishes but does not entirely disappear. In higher dimensions, the subject's data are much less spoke-like, but the configurations of the room sets, relative to each other, are still similar in the plots of Dimension 3 by Dimension 4 and Dimension 5 by Dimension 6. It seems very plausible that what we see in the perceptual scaling solution is increasing error variance laid on top of fundamentally the same solution as displayed by the σ_{CBDIF}'s solution. It is the discrepancies in the orientation of the spokes in the model and data (e. g., Stimulus Set 7, with Stimuli 22, 23, and 24) rather than the lack of linearity of the spokes in higher dimensions that probably address most clearly the discrepancies between the model and perception.

The sharply peaked filter from which σ_{CB} and σ_{CBDIF} are derived is only one of several possible filters. Another measure which is a good predictor is σ_{10}. A difference measure, σ_{10DIF}, was also calculated from $H_{10}(f)$, the spectra used as the base for σ_{10}. Figure 3.37 shows the two-dimensional solution for σ_{10}. While the spoke-like character of the solution is present here also, the qualitative similarity between this solution and the perceptual two-dimensional solution (Figure 3.25) is strikingly less than for σ_{CBDIF}. The quantitative similarity is also less. Correlations with the data dimensions are .82 and .83, compared to .86 and .90 for σ_{CBDIF}. This measure clearly captures less of the comparative perceptual process.

A σ_{CBDIF} analysis was also performed with the data from Experiment 1. The correlation between the σ_{CBDIF} measures and the subjects' mean differences resulted in r = .93. Figure 3.38 shows a plot of the first two dimensions of the σ_{CBDIF} solution and a comparison with Figure 3.5 again shows a strong qualitative similarity. The data-model corresponding correlations for the three-dimensional solution are .94, .95, and .88 for Dimensions 1-3, respectively. The coloration quantity/quality distinction and the spoke description of coloration are supported in these data also. The coloration spokes for

FIGURE 3.37

EXPERIMENT 2: TWO-DIMENSIONAL SOLUTION FOR σ_{10} .



the two room sets with the largest Δ, 5 and 10 feet, determine Dimension 2, while the two room sets with smaller Δ, 1.25 and 2.5 feet, appear to determine Dimension 3 (Figure 3.6). The set of rooms with Δ = .63 remains clustered together and appears to have little or no coloration.

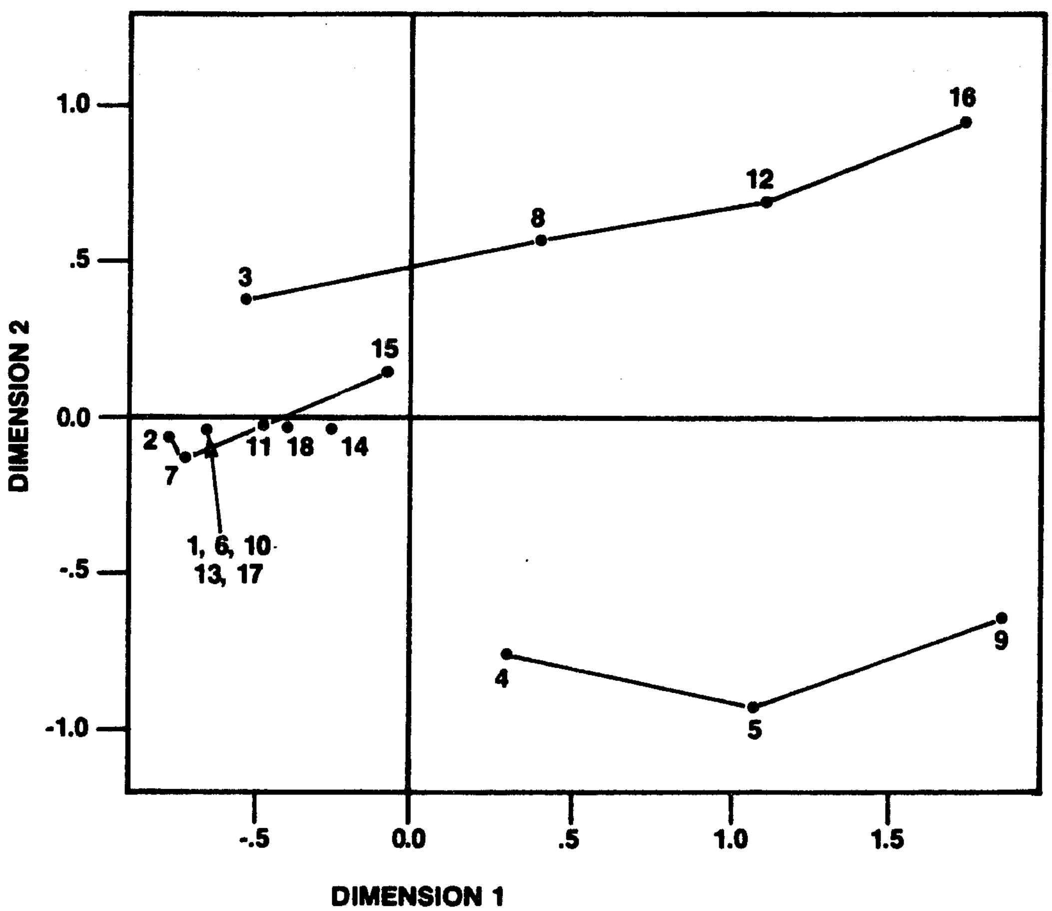
3.3.4 Experiment 2 Discussion

In spite of having multidimensional solutions for Experiments 1 and 2, these solutions can be accounted for in terms of one variable, a sharp, critical-band like filter, on the room frequency response. The solution spaces are difference spaces, and the underlying subjective dimensions are unknown and may not need to exist. Multidimensional spaces obtained in different experiments with different rooms need have nothing in common (such as dimensions) except some representation of coloration quantity.

The distance measure, σ_{CBDIF}, is a composite created from the comparison of two functions of frequency. A sufficiently large difference between the continua at any set of frequencies may be perceptible and thus would be expected to produce a change in the difference judgment. In the context of these experiments, a perceptible change in ripple density would be expected to

FIGURE 3.38

EXPERIMENT 1: DIMENSIONS 1 AND 2 OF THE THREE-DIMENSIONAL SOLUTION FOR σ_{CBDIF}.



corresponding change in difference judgments. One interpretation is that each just noticeable change in a ripple density or in a group of ripple densities may create a new dimension, to yield a multidimensional space of many hundreds of dimensions. Both parsimony and intuition about the integrated nature of the reverberant noise stimuli are violated by this interpretation.

An alternative outcome of repeated experiments is envisionable: a space with relatively few dimensions with characterizable quality dimensions, although just what those unifying dimensions might be is not clear at this time.

What has been modeled is first, coloration quantity for a single static signal, and second, a comparative process to describe the size of perceptual differences. The particular solutions achieved are a function of the stimulus sets chosen for investigation. While perceptual dissimilarity among the stimuli was a criterion of stimulus selection, no thought was given to a systematic sampling of the σ_{CBDIF}'s. (Indeed, the formulation of σ_{CBDIF} came after Experiment 4 was run, although the critical band-like filter and σ_{CB} resulted from Experiment 1.)

Informal listening to the stimulus sets, both singly and in pairs, again confirmed the impression of quantitative and qualitative aspects of coloration. The stimulus sets in Experiment 2, i.e. the spokes of the solution, sound different with respect to some complex pitch-like perception, and the difference in this perception between rooms is captured by σ_{CBDIF}.

3.4 EXPERIMENT 3

A third brief, experiment was done to substantiate σ_{CB}, the quantity of coloration measure, outside the framework of a multidimensional scaling experiment.

The stimuli were the same stimulus set used in the previous experiment, i.e., 26 reverberant noise samples representing white noise heard in 26 different rooms. Subjects who were sophisticated musically or with respect to reverberant sound were used, and they were asked to rate quantity of coloration in, first, a paired-compairson (PC) task and then in an absolute judgment (AJ) task.

3.4.1 Experimental Design

PC Task. Subjects heard all the 26 stimuli paired with Stimulus 1, the low coloration stimulus used as a reference in the previous experiment. Stimulus 1 was always presented first, and subjects were asked to rate, as before, how different the two noises sounded on a scale from 0-9. Subjects heard and rated 5 random orders of the 26 stimulus pairs.

AJ Task. After performing the first task, the subjects were presented the 26 stimuli individually and asked to make absolute judgments of how much coloration each stimulus contained, again on a 0-9 category scale. This task was performed in the same experimental session as the first task. Absolute judgments are expected to produce considerably more variability and it was thought that if subjects could do it at all, it would be while they were already well practiced with the stimulus set. The subjects were told they were hearing the same stimulus set as in the PC task. They heard and rated 10 random orders of the 26 stimuli.

3.4.2 Experiment 3 Methods

Subjects. Ten subjects were used. Four of them had done work concerning reverberation and/or coloration previously. The other six subjects all had many years of musical training and/or experience. Thus all the subjects were to some extent accustomed to listen closely to non-speech sound stimulation. The subjects were also all technical employees at the Murray Hill Bell Laboratories.

Experimental procedures. The stimuli were presented using the same set-up as previously, i. e., subjects sat in an IAC room and listened to taped stimuli over headsets and wrote their responses on answer sheets. The timing of the stimuli in the PC task was the same as in the first two experiments (1 sec stimulus on-time, 1/2 sec interstimulus interval, 1 sec stimulus on-time, 3 sec response interval). For the AJ task, a stimulus on-time of 1 sec was followed by a 3 sec response interval.

3.4.3 Experiment 3 Results

Data analysis was correlational, both for comparing subject groups within and across experiments as well as for evaluating the agreement of coloration measures with the data. The correlations are based on 26 pairs of statistics. The data have been averaged over subjects within type of subject and type of task. All the subjects in both tasks used the full range of the response scale, and normalization of the data was unnecessary. Correlations that refer to previous experiments use the normalized data.

The two types of subjects agree fairly well, r = .95 in the PC data, and r = .93 in the AJ data. Also, the agreement between the data from the two types of tasks, PC and AJ, averaged over subjects is quite good, r = .95. This agreement between the two task types is a little higher, r = .96, for the subjects with experience with coloration than for the subjects with extensive musical experience, r = .91. Difference testing based on the Fisher r to z transform does not result in significant differences between types of subjects or types of tasks.

A subset of the Experiment 2 data, all the comparisons of stimuli with Stimulus 1, can be compared with these PC data. Experiment 2 data included comparisons with Stimulus 1 first (1-i) and comparisons with Stimulus 1 last (i-1). The combined normalized data yields $r = .90$ with the PC data. For 1-i only, $r = .92$, with the PC data while the i-1 data yields only $r = .70$. Since the PC data for this experiment always presented Stimulus 1 first, a somewhat higher correlation is to be expected when the presentation order is identical. This difference is reliable ($z = 2.45$, $p < .05$). It is somewhat surprising that the correlation of i-1 is so low.

The above correlations are measures of consistency with and across similar data sets and, in general, they are fairly high. They represent measures of the data's reliability since they are measures of different subject groups' judgments on the same task. Since the degree to which subjects' responses are reliable will place an upper bound on how well the responses reflect an underlying continuum, the correlations also represent an upper limit on values that should be expected when the data is correlated with coloration measures. Both σ_{CB} and σ_{10} have been correlated with some of these data. See Table 3.8 below. Although individual comparisons are not statistically significant, σ_{CB} is consistently a slightly

Table 3.8 Correlations Between Dataset and Measures of Coloration

Dataset	σ_{CB}	σ_{10}
PC, All Ss	.92	.85
PC, Coloration Ss	.90	.83
PC, Musical Ss	.91	.85
AJ, All Ss	.82	.78
AJ, Coloration Ss	.86	.79
AJ, Musical Ss	.77	.75

better predictor within these data than is σ_{10}.

The correlation between σ_{CB} and the PC data is .92, and the internal correlation (between the two types of subjects) in those data is .95. The correlation of σ_{CB} with the AJ data is .82, and that internal correlation is .93. For the PC data, the correlation is probably as high as could reasonably be expected, while the AJ correlation seems a little low. Figures 3.39 and 3.40 present scatterplots of σ_{CB} plotted against the PC data and the AJ data. The range of judgments appears to be reduced at both the top and bottom of the distribution for the AJ data, attributable to the lack of a standard for comparison. It is not surprising that the correlation for the PC data is higher than the AJ data. The PC task is traditionally the easier, and is also the type of data from which σ_{CB} was derived.

FIGURE 3.39

EXPERIMENT 3: PAIRED COMPARISON JUDGMENTS AS A FUNCTION OF σ_{CB}.

r = .92 AND
PC′ = 6.4 σ_{CB} - .29

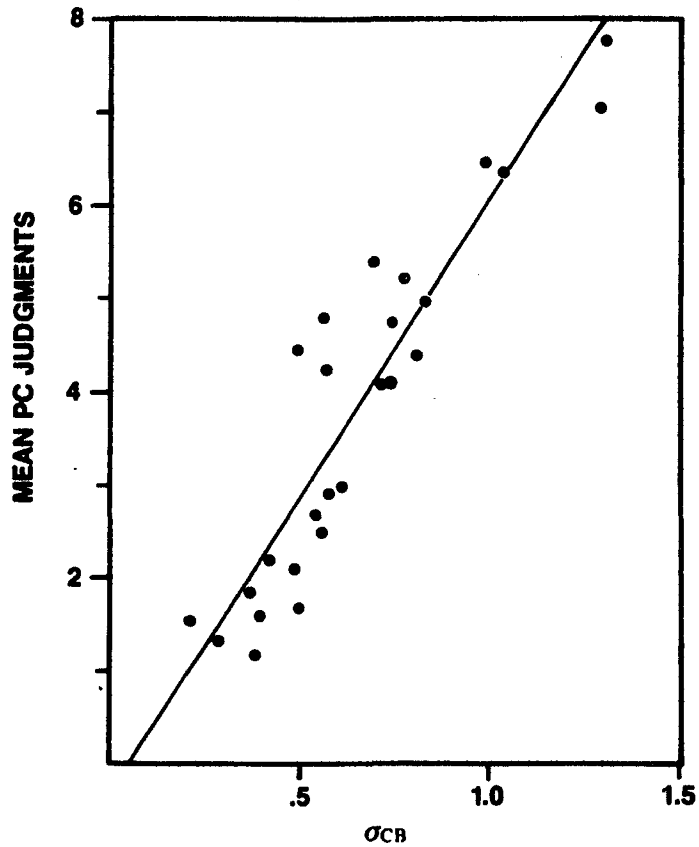
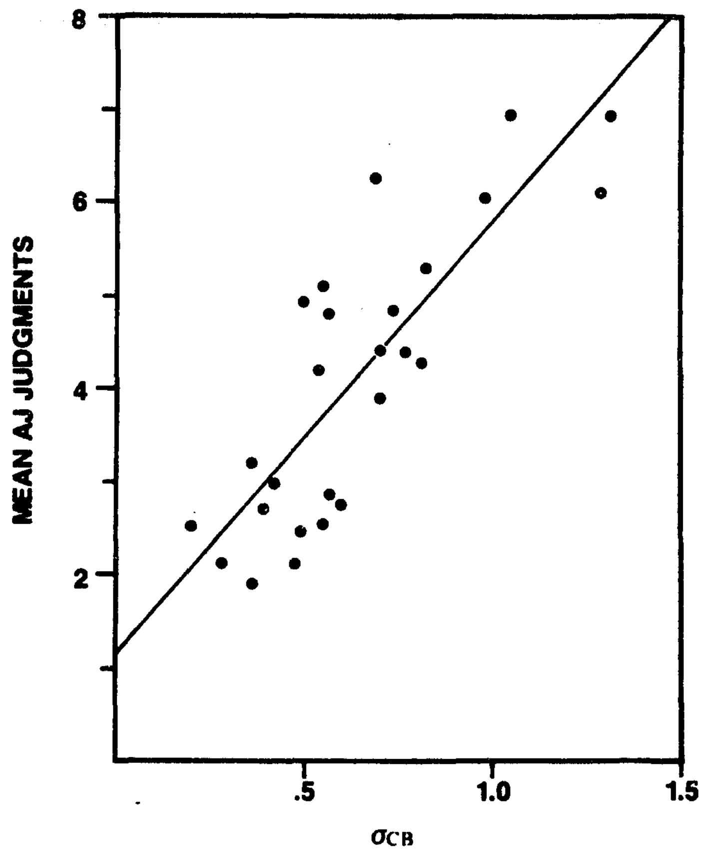


FIGURE 3.40

EXPERIMENT 3: ABSOLUTE JUDGMENTS AS A FUNCTION OF σ_{CB} .

r = .82 AND

AJ' = 4.67 σ_{CB} + 1.14



The difference in the correlations between the PC data and the l-i and i-l data from Experiment 2 (r = .92 and .70, respectively) suggested that it would be interesting to compare the two Experiment 2 data sets, i-l, and l-i, with σ_{CB} separately. These correlations are .90 and .77, respectively. Scatterplots for these data are presented in Figures 3.41 and 3.42, respectively. These correlations are in keeping with how well the Experiment 2 data agree with the PC data. The scatterplot for the i-l data shows a drop in range for the ratings but the most striking effect is among the stimuli with moderate values for σ_{CB}. Most of the middle stimuli are clustered fairly closely together.

3.4.4 Discussion

The data in both Experiment 2 and in Experiment 3 demonstrate a large effect of order of presentation. It is probably not due to interactions within the peripheral auditory system. The interstimulus interval is 500 msec throughout these studies, which is well beyond the 200-300 msec usually found in temporal masking paradigms to represent the limits of masking. Since the goal of these studies is oriented towards a general account of coloration, averaged over subjects, methods, and order of presentation, this finding is noted to be of interest but will not be further examined.

FIGURE 3.41

EXPERIMENT 2: MEAN DIFFERENCE JUDGMENTS FOR THE 1-/ STIMULUS PAIRS AS A FUNCTION OF σ_{CB}.

r =.90 AND

$Y' = 4.42\,\sigma_{CB} + .41$

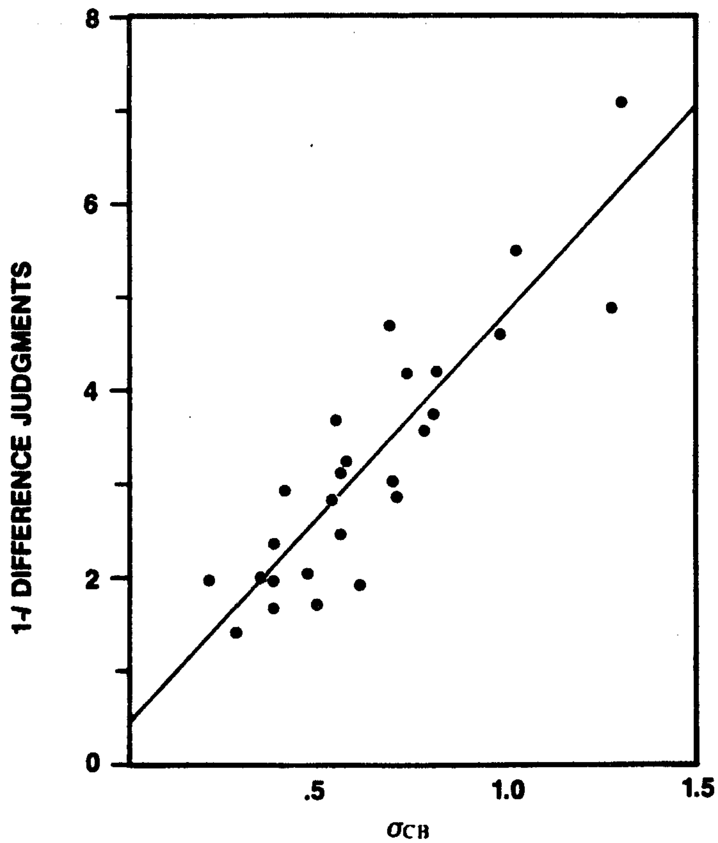
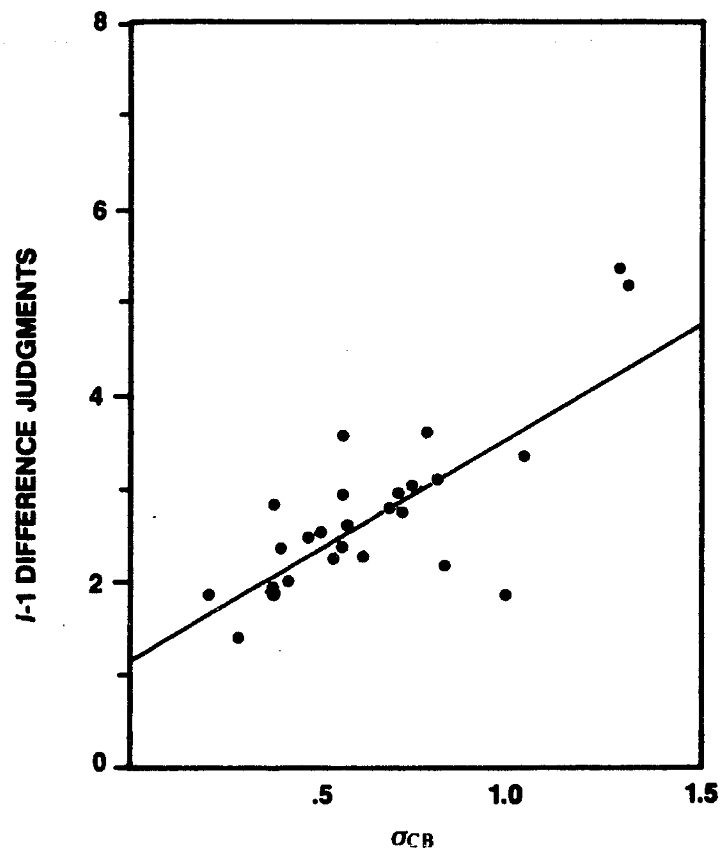


FIGURE 3.42

EXPERIMENT 2: MEAN DIFFERENCE JUDGMENTS FOR THE *I*-1 STIMULUS PAIRS AS A FUNCTION OF σ_{CB} .

r =.77 AND
$Y' = 2.42\,\sigma_{CB} + 1.19$



In this experiment, using slightly different experimental methods than were used in Experiments 1 and 2, σ_{CB} was further substantiated as a measure of coloration. There is sufficient evidence at this point, with the weight of these three experiments supporting the original speech work of Allen, et. al., 1979, that a measure for the non-time varying, frequency domain aspects of reverberation perception has been achieved. As in Experiment 2, σ_{CB} was compared with σ_{10} with respect to response predictability. In both experiments, the correlations associated with σ_{CB} are consistently slightly higher than those associated with σ_{10}.

3.5 EXPERIMENT 4

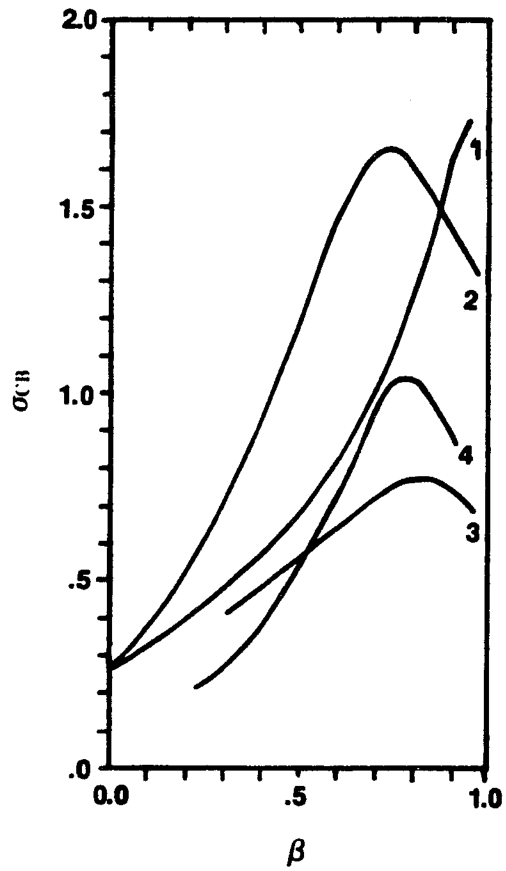
The previous studies describe how coloration is perceived by listeners, and they provide a scale of coloration, σ_{CB}, and a measure of the difference in coloration between rooms, σ_{CBDIF}. This last study is an effort to determine the perceptual range of the coloration scale. Using σ_{CB} as the coloration measure, an estimate of the number of jnd on the coloration scale is determined by doing a Thurstonian scaling analysis of paired-comparison data. This is an attempt to put coloration, and the coloration measure, σ_{CB}, into perspective. Discrimination on the two main auditory dimensions, pitch and loudness, is extensive. An estimate of the number of jnd on the coloration scale will help to give some idea of the magnitude of coloration effects relative to major auditory phenomena.

The stimulus set needed for this experiment should consist of stimuli which vary only in amount of coloration, not in quality of coloration. Both of the first two experiments demonstrate the multidimensionality of coloration, but they also demonstrate that varying β, while holding all other room parameters constant, results in both theoretical and empirical spokes of coloration. Neither these theoretical or empirical spokes are straight lines, however, and this suggests that though these spokes may approach unidimensionality, they do not attain it. Increasing β will increase proportionately all of the reflections that compose h(t) and thus all of the fluctuations in H(f). The effect may not be proportionate perceptually. For example, when β is low many of the low frequency fluctuations in H(f) may be below threshold and these fluctuations may become audible as β increases. This amounts to adding ripples to the stimulus spectrum and would seem to be a violation of unidimensionality with respect to pitch. However, there is no obvious way to vary coloration without violating unidimensionality in some way. Another approach would be to fix β as well as the other room parameters, and then add increasing amounts of white noise. This would fill in the deepest valleys of H(f) before affecting the shallow ones, and thereby also violate the potential for perceptual unidimensionality.

Increasing β has been chosen as the means of increasing coloration since this represents a real parameter of rooms. Although increasing β will be the means of increasing coloration in the ROOM program, it will be equal interval steps of σ_{CB} that will determine the stimulus set. The greatest possible range of σ_{CB} is desirable, but the maximum possible value that can plausibly be produced by a room isn't known. Since considerable trial and error listening went into the selection of the Experiment 2 rooms in order to select rooms with different and noticeable coloration, it was decided to explore those room sets for the stimulus selection. Two of the room sets, Sets 1 and 2 from Experiment 2, produce a greater range of σ_{CB} than any of the others. Figure 3.43 plots σ_{CB} as a function of β for some of the room sets. For most of them, as β becomes large, σ_{CB} peaks and then decreases. This effect hasn't been explored, but it seems to be produced by fluctuations due to late reflections becoming great enough and dense enough to begin to smooth out the fluctuations in H(f) which produce large σ_{CB}. For Sets 1 and 2, σ_{CB} reaches values over 1.5 while it is considerably less for all of the other Experiment 2 sets of rooms. Those sets not included in the figure were explored, but only sparsely, and functions can not be drawn for them. Accordingly, Room Sets 1 and 2 from Experiment 2 were selected as the basis for the stimuli in this experiment. This means two sets of stimuli were used in this experiment. One of the stimulus sets consists of rooms with the same parameters as those for Room Set 1 in Experiment 2 (with the exception of values of β) and the other stimulus set is drawn from Room Set 2 in Experiment 2 in the same way. For each room, a set of values for β was selected to result in equal interval steps for σ_{CB}.

FIGURE 3.43

EXPERIMENT 4: σ_{CB} AS A FUNCTION OF β FOR ROOM SETS 1-4 OF EXPERIMENT 2.



The assumption is that coloration is some monotonically increasing function of σ_{CB}. Because the purpose of the study is to evaluate the subjective range of coloration, it is important that consecutive stimuli not be easily discriminated. Also, σ_{CB} values will be nearly identical for the two rooms sets, and this will facilitate a comparison of their jnd scales.

An additional fact should be noted; this experiment was designed and carried out before the σ_{CBDIF} measure was developed. A perhaps better way to determine equal interval coloration steps would be to use equal σ_{CBDIF} steps, since varying β may not produce unidimensional coloration variation. Appendix A contains the σ_{CBDIF} values for the whole stimulus matrix for Stimulus Sets 1 and 2. As will be seen (Table 3.10) not all of these stimulus pairs were heard by the subjects during the test sessions. The data analyses will compare scales based on σ_{CB} and σ_{CBDIF}.

3.5.1 Experimental Design

Table 3.9 presents the stimuli used in this experiment. There were two sets of room stimuli. Stimulus Set 1 is drawn from Room Set 1 in Experiment 2 and Stimulus Set 2 is drawn from Room Set 2 in Experiment 2. Each stimulus set contains 13 rooms and the value of

Table 3.9 Stimulus Sets 1 and 2, Experiment 4.

Set 1 Stimuli	β	σ_{CB}	σ_{CBDIF}
1-1	.010	.27	
1-2	.200	.40	.164
1-3	.320	.50	.120
1-4	.420	.60	.121
1-5	.510	.70	.137
1-6	.580	.80	.133
1-7	.640	.90	.144
1-8	.695	1.00	.169
1-9	.737	1.10	.161
1-10	.773	1.20	.163
1-11	.807	1.30	.174
1-12	.840	1.40	.185
1-13	.875	1.50	.207
			Σ = 1.878

Set 2 Stimuli	β	σ_{CB}	σ_{CBDIF}
2-1	.010	.28	
2-2	.110	.38	.175
2-3	.170	.48	.112
2-4	.225	.58	.112
2-5	.275	.68	.108
2-6	.325	.78	.116
2-7	.370	.88	.115
2-8	.415	.98	.127
2-9	.455	1.08	.124
2-10	.495	1.18	.138
2-11	.535	1.28	.138
2-12	.570	1.38	.122
2-13	.615	1.48	.152
			Σ = 1.537

σ_{CB} ranges from slightly under .3 to about 1.5 for both stimulus sets. The values of σ_{CBDIF} refer to the difference between pairs of successive stimuli. The σ_{CBDIF} are not constant; they are larger at the two ends of the scale than in the middle. The ratio of the largest to the smallest difference is about 1.7 for both sets of stimuli.

All the subjects were run on the first set of rooms before testing was done on the second set. Two new frozen noise samples were used, one was convolved with the first set of rooms and the other with the second set. The practice stimuli for each room set used the same room set and noise sample as the test stimuli.

Each stimulus set consisted of 13 noise samples representing progressively more highly colored rooms (Table 3.9). All pairs of 13 rooms were not used in the experimental sessions. Pairs of rooms widely separated in coloration were not presented for comparison in the test sessions. Table 3.10 presents the pairs of stimuli which were used in the test sessions. Each stimulus was paired with the three stimuli closest to it in coloration, both below and above it, and both orders of each pair were used. The least, median, and greatest stimulus in each set was paired with itself. This made a total of 69 stimulus pairs with 33 AB and 33 BA comparisons and 3 pairs taken from the diagonal of the stimulus matrix. Ten different random orders of each set of 69 stimulus pairs were created and analog stimulus tapes were made which divided the 10 sets into two tapes with five orders each.

Table 3.10 Stimulus Pairs Used in Experiment 4 Test Sessions.

First Stimulus	Second Stimulus 1	2	3	4	5	6	7	8	9	10	11	12	13
1	x	x	x	x									
2	x		x	x	x								
3	x	x		x	x	x							
4	x	x	x		x	x	x						
5		x	x	x		x	x	x					
6			x	x	x		x	x	x				
7				x	x	x	x	x	x	x			
8					x	x	x		x	x	x		
9						x	x	x		x	x	x	
10							x	x	x		x	x	x
11								x	x	x		x	x
12									x	x	x		x
13										x	x	x	x

For each set of 13 room stimuli the subjects heard 25 replications of the 69 stimulus pair set; they heard one of the tapes twice and the other three times. Thus for each stimulus pair each subject responded to 50 presentations of each pair, 25 in each order of presentation. In all, each subject heard 1725 stimulus pairs for each of the two sets of room stimuli. Each daily test session consisted of one of these sets of 5 orders of the 69 stimulus pairs, and testing lasted 10 days, 5 days on each set of rooms. (This is in contrast to two days of test sessions for Experiments 1 and 2 and a session on one day for Experiment 3.)

3.5.2 Experiment 4 Methods

The experimental task consisted of subjects listening to a pair of colored noise stimuli and then judging which of them had the greatest coloration. This was a forced choice paradigm; subjects were told to guess when they were uncertain. Enough data were collected to determine individual psychophysical functions as well as a group function. Naive subjects were used and it was necessary to train them with respect to coloration by using practice trials with feedback before the experimental sessions. During an experimental session no feedback was given.

As in the earlier experiments, each stimulus was 1 sec in duration, there was a .5 sec interval between the stimuli in a pair and a 3 sec interval after each pair in which subjects wrote their response on an answer sheet. These responses consisted of a 1 if the first stimulus was perceived as having more coloration and a 2 if the reverse was true. The subjects were instructed, prior to any practice, that coloration was perceptible as an increased pitchiness, or roughness, in the stimulus.

Prior to the test session, subjects received practice trials until most of them were performing above chance and they reported feeling comfortable with the task. During these trials, after the 3 sec response interval one of two lights lit up to indicate the correct answer. Subjects scored themselves at that time. There were 4 short practice tapes. The first tape presented the lowest coloration stimulus paired with each of the others in order of increasing coloration, i.e., 1-2,1-3,1-4, etc., to permit the subjects to hear the coloration continuum. The next three tapes presented small subsets of the stimulus matrix, each progressively more difficult. In particular, stimuli were at least 3 consecutive stimuli apart on tape 2 (stimulus comparisons included 1-4, 1-7, 1-10, 2-5, 2-8, etc.). They were at least 2 consecutive stimuli apart on tape 3 (1-3, 1-5, 1-7, 2-4, 2-6, etc.). Tape 4 was representative of the test tapes and included consecutive stimuli (see Table A.2). There were two sets of practice tapes for tapes 2-4 with different random orders of the stimulus conditions on each.

In the early days of testing, it was common to go through both sets of the practice tapes at all difficulty levels, but as subjects became more experienced, the need for this was reduced. However, subjects always heard at least one of each of the 4 practice tapes before testing began each day. It should also be noted that the subjects were never told how many stimuli were in the study, or that the practice tapes included all the stimuli.

Subjects. Data were obtained for five subjects; a sixth subject chose to drop out before testing was completed. The subjects were high school students. They came in for testing every weekday morning for two weeks.

3.5.3 Experiment 4 Results

The daily performance for each subject for Stimulus Set 1 and 2 is presented in Figures 3.44 and 3.45. Each day's test session consisted of 345 trials, which creates a 95% confidence interval of .053 about the chance performance level of .50.

Figure 3.46a and b presents the mean performance data averaged over the five subjects for the two sets of rooms. It is desirable to use stable, asymptotic discrimination data in developing psychophysical scales, and analyses of variance were performed on both sets of data to determine the days on which performance was stable. For Stimulus Set 1 an analysis of variance indicated that performance on the five days on Set 1 were significantly different ($F_{4,20} = 12.7$, $p<.05$), while performance on the last 4 days did not yield a significant F ratio. An analysis of variance on Stimulus Set 2 data did not indicate a significant difference in performance across the five days of testing. As a result, the psychophysical scaling analyses were performed on the last four days of testing

FIGURE 3.44

EXPERIMENT 4; ROOM SET 1: PROPORTION CORRECT FOR EACH SUBJECT ON EACH DAY.

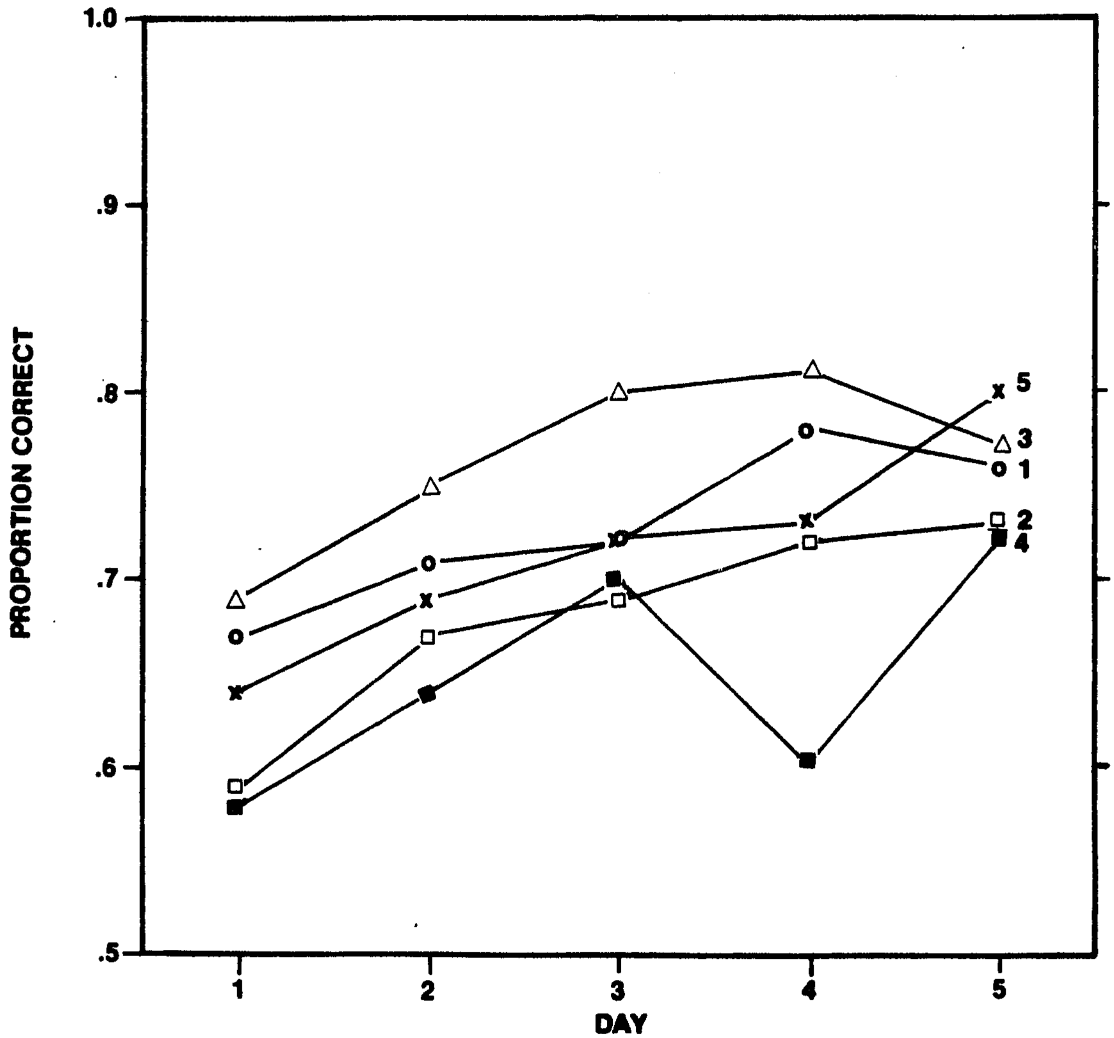


FIGURE 3.45

EXPERIMENT 4; ROOM SET 2: PROPORTION CORRECT FOR EACH SUBJECT ON EACH DAY.

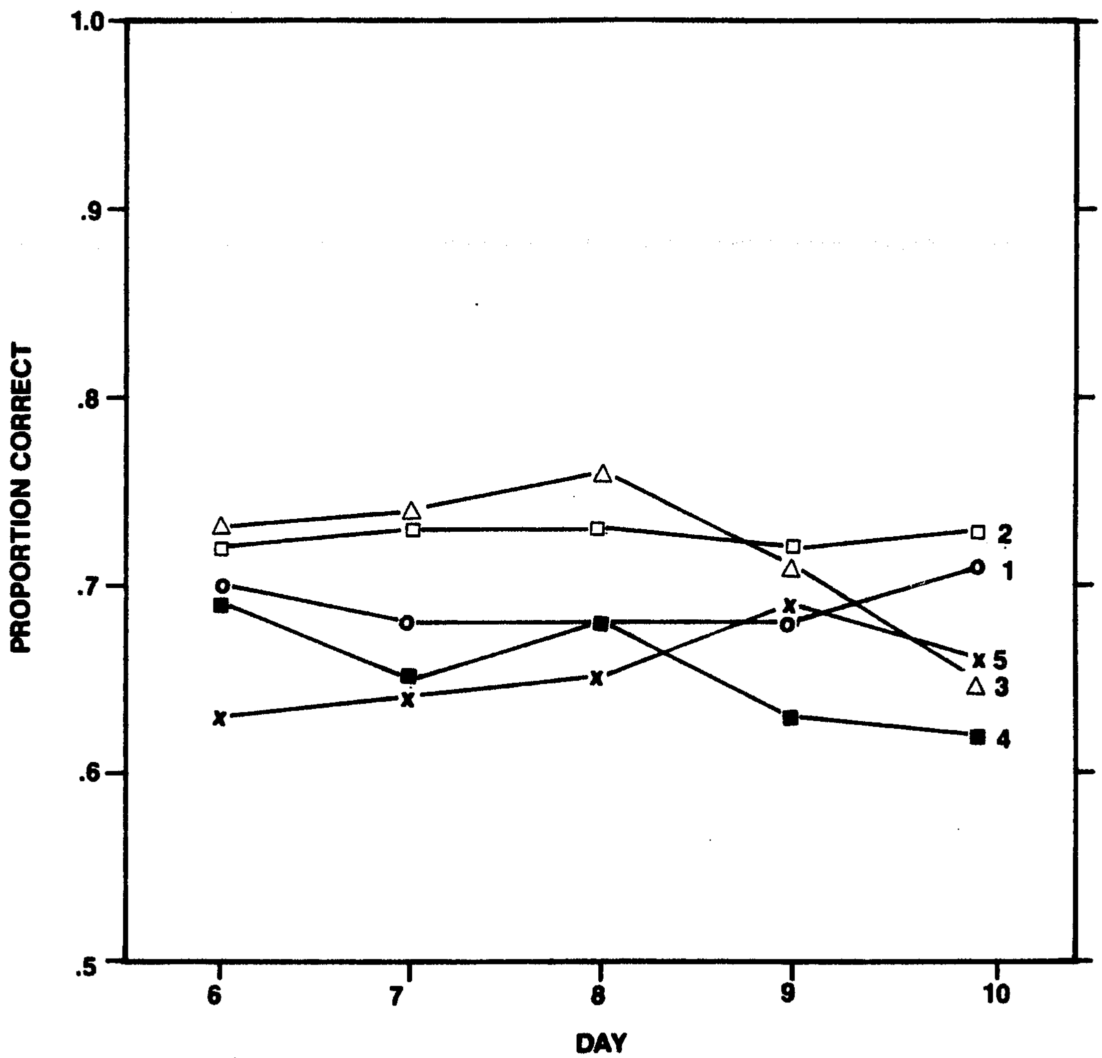
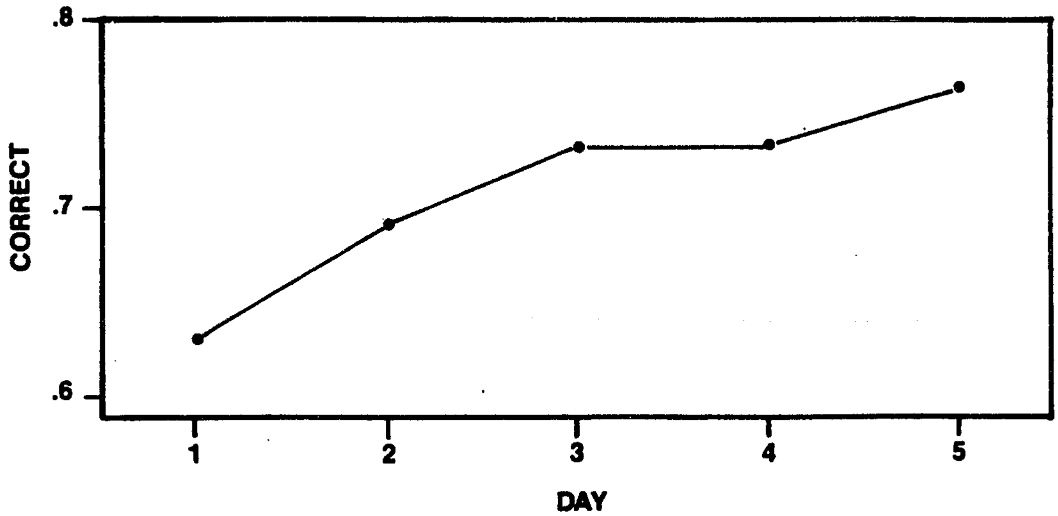


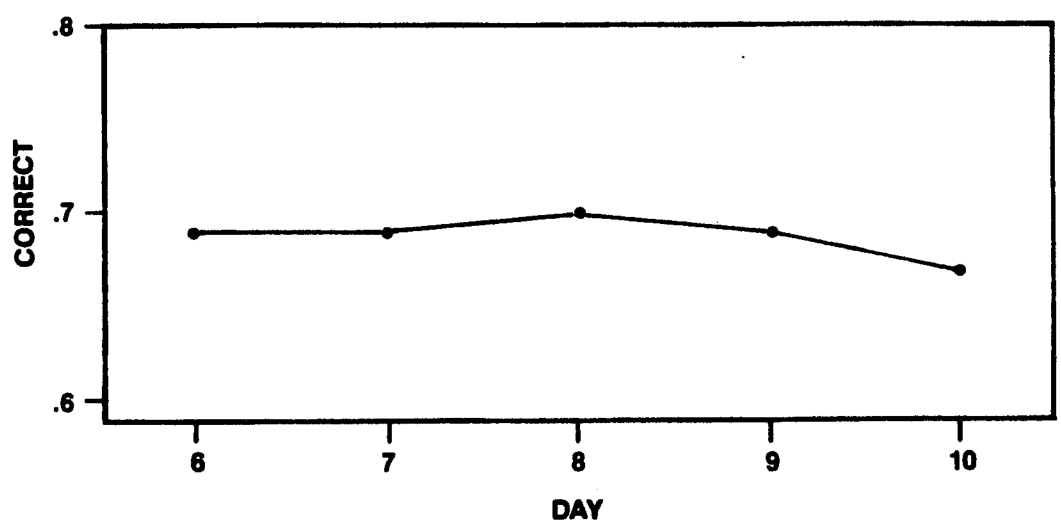
FIGURE 3.46

EXPERIMENT 4: MEAN PROPORTION CORRECT

A. ROOM SET 1



B. ROOM SET 2



for Stimulus Set 1 and all five days of testing for Stimulus Set 2.

Performance was about 73% correct for the first stimulus set and about 69% for the second stimulus set. Thus it appears that the first stimulus set provided easier discrimination. The stimulus sets seem to include about the right difficulty levels since the average proportion correct is in the middle of the scale.

The tabulated data for this study consist of the proportions of times each stimulus was judged to have more coloration than the other stimuli with which it was compared. Tables 3.11 and 3.12 present these proportions,

Table 3.11 Percent of Observations for Which the Row Stimulus Was Judged to Have Greater Coloration Than the Column Stimulus; Stimulus Set 1, Experiment 4.

	1	2	3	4	5	6	7	8	9	10	11	12
2	56											
3	61	55										
4	65	59	55									
5		64	60	56								
6			65	61	55							
7				66	61	56						
8					67	62	57					
9						68	63	56				
10							68	63	56			
11								69	63	57		
12									70	64	58	
13										71	65	58

Table 3.12 Percent of Observations for Which the Row Stimulus Was Judged to Have Greater Coloration Than the Column Stimulus; Stimulus Set 2, Experiment 4.

	1	2	3	4	5	6	7	8	9	10	11	12
2	57											
3	61	54										
4	65	59	54									
5		63	59	54								
6			63	59	55							
7				63	59	55						
8					64	59	55					
9						64	60	55				
10							65	60	55			
11								65	61	56		
12									65	60	55	
13										66	61	56

averaged over subjects, for Stimulus Sets 1 and 2 respectively. These data served as input to a Thurstonian paired comparison scaling analysis which converts these proportions to the corresponding normal deviate and determines a psychophysical scale from them (Thurstone 1927a and b, and Torgerson, 1958). The 33 comparisons in the tabulated data produce 33 degrees of freedom which were were used to predict the scale. To smooth the data further, second order polynomials were fit to the Thurstonian scales.

<u>Data Analysis Methods</u>. Thurstonian scaling consists of a model to use discrimination data to achieve an underlying psychological scale. Thurstonian scaling (Thurstone, 1927a and b, and Torgerson, 1958) assumes that

the degree of similarity of stimuli on an underlying continuum is reflected by subjects' ability to discriminate between the stimuli.

Briefly, Thurstone assumed that the effect of a stimulus is to give rise to a perception along the stimulus continuum. Over successive trials, slightly different perceptions occur with the same physical stimulus. The population of these perceptions along the stimulus continuum is assumed to be normally distributed with mean $\bar{x}_i$ and variance σ_i^2 for stimulus i. For different stimuli, the mean of the differences between two stimuli is equal to the difference of their means $\bar{x}_j - \bar{x}_i$ and the standard deviation of the difference distribution,

$$\sigma_{x_j - x_i} = (\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j)^{1/2}$$

From the assumption of normal distributions, it follows that the proportion of times a paired comparison of stimuli i and j results in j being judged greater can be used to determine the normal deviate z_{ij}:

$$\bar{x}_j - \bar{x}_i = z_{ij}\sigma_{x_j - x_i}$$

is the general form of Thurstone's law of comparative judgment. The inputs to the equations for a set of stimuli are the z_{ij} values which are determined from the proportions in the data. Various cases of Thurstone's law of comparative judgment introduce simplifying assumptions in order to reduce the number of unknown quantities to no more than the number of equations. Thurstone's Case V, which includes the most simplifying assumptions, is used here. This case assumes that on each trial, the observations x_j and x_i are uncorrelated and that the variances associated with the stimuli, σ_j^2, σ_i^2, etc. are constant. The resulting equation is

$$\bar{x}_j - \bar{x}_i = cz_{ij}$$

where c is a constant. For these experiments, Case V has 33 proportions from which to estimate 12 scale values; there are 13 scale values, but one of them is fixed to set a scale origin.

Another Thurstone case was also explored with these data, Case III, in which the variances are not assumed equal, but are assumed to be uncorrelated. The resulting scaling solutions must then solve for both the scale value and its variance for each stimulus, which doubles the number of parameters to be estimated. For these data, this would mean the 33 observed proportions would be used to calculate 12 scale values and 12 variances associated with the scale values (one of the 13 variances is fixed to determine the scale unit).

The psychophysical scales resulting from the scaling solutions can be used to determine the number of jnd on the scale. It is assumed that two stimuli, i and j, are one jnd apart in coloration when stimulus j is reliably and correctly perceived to have more coloration than stimulus i 75% of the time. In terms of the difference distribution which characterizes the coloration difference between i and j, this jnd difference is the normal deviate associated with 75% of the normally distributed difference distribution. Thus,

$$jnd = .67\sigma_{x_j - x_i}$$

For the Case V analysis, we have assumed the standard deviations of the stimuli to be fixed and equal ($\sigma_i = .2$). From this, $\sigma_{x_j - x_i} = (1.414)\sigma_i$, and a jnd, in the same scale units as the Thurstone scale is

$$jnd = .67\sigma_{x_j - x_i} = .67(1.414)\sigma_i$$

The Thurstonian Case V scaling of the paired comparison data was accomplished by means of an iterative maximum log likelihood solution developed and programmed by C. A. King of Bell Laboratories. His approach uses the first two terms of the Taylor expansion of the differentiation of the log likelihood function with respect to the parameters to be estimated, i.e., the means (and variances for Case III) of the stimuli on the psychological continuum:

$$\frac{\partial L}{\partial \Theta_k} = \sum_{i,j} \frac{n_{ij} \partial P_{ij}}{N_{ij} \partial \Theta_k} = 0$$

The above is the maximization condition, where L = the log likelihood function, Θ_k = the parameters to be estimated, n_{ij} = the number of times stimulus i is chosen over stimulus j, N_{ij} = the number of comparisons of i and j (independent of order of presentation), and

$$P_{ij} = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{z_{ij}} e^{\frac{-z_{ij}^2}{2}} \partial z$$

The first order Taylor expansion of $\partial L/\partial \Theta$ yields equations which may be iteratively solved for Θ which yield P_{ij} near n_{ij}/N_{ij} and σp_{ij} near the sample standard deviation.

This procedure does not require a complete upper or lower half matrix as input. King's procedure also computes the x^2 associated with the predicted probabilities of choice for each stimulus pair and the obtained data, and this is a measure of goodness of fit for the solution. The degrees of freedom associated with the data are the number of data entries in the upper half matrix which serves as input to the program, the n_{ij}/N_{ij}, and there are 33 such entries in this case. The number of parameters to be estimated must be subtracted from this to determine the x^2 degrees of freedom: 33 - 12 = 21 degrees of freedom. After determining Thurstonian scales for both individual and grouped data, second order polynomials were fit to the scales. These polynomials smooth the scales and give an indication of how well the scales are fit with three parameters.

Data Analysis Results. The data for each subject were combined over orders of presentation to determine the number of times stimulus i was preferred to stimulus j out of n comparisons. For Stimulus Set 1 the data for each subject was based on the last 4 days of testing to give a total of 40 observations for each ij comparison; Stimulus Set 2 data were based on 5 days of testing for a total of 50 observations for each ij comparison. The subjects' data were also combined within a stimulus set, and a scale based on the total dataset was developed for each stimulus set.

The scales determined were based on the Case V model discussed in the previous section. The King program was run using both Case III and Case V assumptions. The Case III assumptions led to results which were heavily dependent on the starting configuration and which gave very highly variable estimates for the standard deviations. The orderliness of the data (Tables 3.11 and 3.12) suggest no reason for this, and it seems probable that this was due to trying to estimate too many parameters (24) with too few data points (33). As a result, the scales to be discussed are based on Case V analyses only. For these data, the standard deviation of the scale value was set equal to .2, and the midpoint of each of the scales was fixed at 1.0 to give each a scale a point in common thereby facilitating inter-subject comparisons. For example, a subject who responds more accurately than another subject will produce a scale with a larger range because the fixed standard deviation unit of the scale will produce distances between points dependent on the number of σ units between the points.

Tables 3.13 and 3.14 present the Thurstone scales for Stimulus Sets 1 and 2, respectively.

Table 3.13 Thurstonian Scales for Group Data and Individual Subjects for Stimulus Set 1, Experiment 4.

Stimuli	σ_{CB}	σ_{CBDIF}	S1	S2	S3	S4	S5	Group
1-1	.27	.00	.27	.54	.04	.48	.50	.39
1-2	.40	.16	.44	.65	.18	.66	.53	.52
1-3	.50	.28	.59	.69	.31	.78	.63	.62
1-4	.60	.39	.68	.76	.47	.82	.71	.70
1-5	.70	.53	.86	.84	.65	.95	.84	.84
1-6	.80	.66	.91	.94	.80	1.02	.94	.93
1-7	.90	.81	1.00	1.00	1.00	1.00	1.00	1.00
1-8	1.00	.97	1.13	1.13	1.09	1.07	1.11	1.10
1-9	1.10	1.13	1.21	1.24	1.20	1.13	1.15	1.18
1-10	1.20	1.29	1.37	1.29	1.31	1.15	1.25	1.27
1-11	1.30	1.46	1.41	1.36	1.41	1.21	1.36	1.34
1-12	1.40	1.66	1.47	1.37	1.45	1.23	1.42	1.38
1-13	1.50	1.85	1.55	1.43	1.50	1.32	1.54	1.46
Scale Range			1.28	.89	1.46	.84	1.04	1.07
χ^2			27.1	22.1	18.9	22.4	21.7	28.0
Number of jnd			6.8	4.7	7.7	4.4	5.5	5.7

In each table, the second column contains the σ_{CB} value for the stimulus and the third column presents the cumulative σ_{CBDIF} value for the stimulus.

Appendix A (Tables A.1 and A.2) shows how these cumulative σ_{CBDIF} are derived. In Tables 3.13 and 3.14, note, first, equal intervals of σ_{CB} do not produce equal σ_{CBDIF} intervals, and note, second, that the range of σ_{CBDIF} is larger for Stimulus Set 1 (1.85) than for

Table 3.14 Thurstonian Scales for Group Data and Individual Subjects for Stimulus Set 2, Experiment 4.

Stimuli	σ_{CB}	σ_{CBDIF}	S1	S2	S3	S4	S5	Group
2-1	.28	.00	.47	.28	.45	.33	.63	.45
2-2	.38	.17	.57	.40	.56	.51	.68	.55
2-3	.48	.28	.62	.46	.63	.60	.75	.62
2-4	.58	.40	.69	.60	.70	.74	.77	.71
2-5	.68	.50	.77	.74	.79	.81	.88	.80
2-6	.78	.62	.91	.90	.94	.94	.90	.92
2-7	.88	.74	1.00	1.00	1.00	1.00	1.00	1.00
2-8	.98	.86	1.10	1.08	1.11	1.03	1.01	1.06
2-9	1.08	.99	1.18	1.20	1.17	1.04	1.07	1.14
2-10	1.18	1.12	1.22	1.27	1.26	1.08	1.24	1.21
2-11	1.28	1.26	1.26	1.28	1.36	1.10	1.26	1.25
2-12	1.38	1.38	1.32	1.34	1.42	1.14	1.29	1.30
2-13	1.48	1.54	1.35	1.41	1.50	1.12	1.39	1.35
Scale Range			.88	1.13	1.05	.79	.76	.90
χ^2			10.4	21.5	11.6	16.4	13.5	9.7
Number of jnd			4.7	6.0	5.6	4.2	4.0	4.8

Stimulus Set 2 (1.54).

Two other measures are also presented at the bottom of these tables.

The χ^2 values calculated for each scale have 21 df. A χ^2 of 32.7 is needed for significance at the .05 level and none of these χ^2's approach significance. This means that the size of the differences between the observed and predicted proportions did not reach statistical significance.

The second set of statistics is the number of jnd in each scale. With fixed $\sigma_i = .2$, and jnd = $(.67)(1.414)\sigma_i = .189$, the number of jnd in a scale is simply the range of the scale divided by .189. Thus for the group data for Stimulus Set 1, the scale range is (1.46 - .39) = 1.07 and 1.07/.189 = 5.7 jnd's on the coloration scale. For Stimulus Set 1, there are about 5 1/2 jnd's in the coloration scale, and for Stimulus Set 2, there are about 4 1/2 jnd's. This difference is in keeping with the range of each scale as indicated by the range of σ_{CBDIF}. Indeed, the ratio of the scale range to the σ_{CBDIF} range is a constant for the group data for both stimulus sets (1.07/1.85 = .90/1.54 = .58), although this is not true for the individual data.) For Stimulus Set 1, the cumulative σ_{CBDIF} range is 1.85 while it is 1.54 for Stimulus Set 2. Based on these ranges, the number of jnd should be greater for Set 1 than for Set 2. Four out of the 5 subjects do display better discrimination on Stimulus Set 1; Subject 2 is the exception. This is additional evidence that the σ_{CBDIF} measures represent accurately the comparison of spectral variation. The near agreement between σ_{CB} and σ_{CBDIF} suggests that varying β provides a nearly unidimensional variation in coloration. The data from the two stimulus sets is sufficiently similar to conclude that quantity of coloration results in a psychophysical scale with a range of about 5 1/2 jnd's for the perceptually most extensive room set evaluated.

Figures 3.47 and 3.48 preseent the individual Thurstone scales for Stimulus Sets 1 and 2, respectively. The abscissa is σ_{CB} and the ordinate is the Thurstone scale value. For Stimulus Set 1, Subject 3 is conspicuous for his markedly better performance on the lower part of the σ_{CB} scale. For Stimulus Set 2, Subject 5 is somewhat worse than the others on the lower part of the scale. In both stimulus sets, Subject 4 performs most poorly on the upper half of the σ_{CB} scale. Also, Subject 4 is the only subject whose scales display a non-monotonicity, i.e., between .7 and .9 σ_{CB} for Stimulus Set 1 and at the highest σ_{CB} value for Stimulus Set 2.

Each subject's scale and the group scale was smoothed using a second order polynomial, i.e., 3 parameters, fit to the Thurstonian scale values. These polynomials are presented as a function of σ_{CB} and σ_{CBDIF} for each individual subject and for the group in Appendix A (Tables A.3 and A.4). Figures 3.49 and 3.50 present the polynomials for the group data for Stimulus Sets 1 and 2, respectively. Clearly a second order polynomial is needed, and is sufficient, to account for most of the variability in the data. Some examples of individual data are presented in Figures 3.51, 3.52, and 3.53, recall that

FIGURE 3.47

EXPERIMENT 4: STIMULUS SET 1: THURSTONE SCALE AS A FUNCTION OF σ_{CB} FOR INDIVIDUAL SUBJECTS.

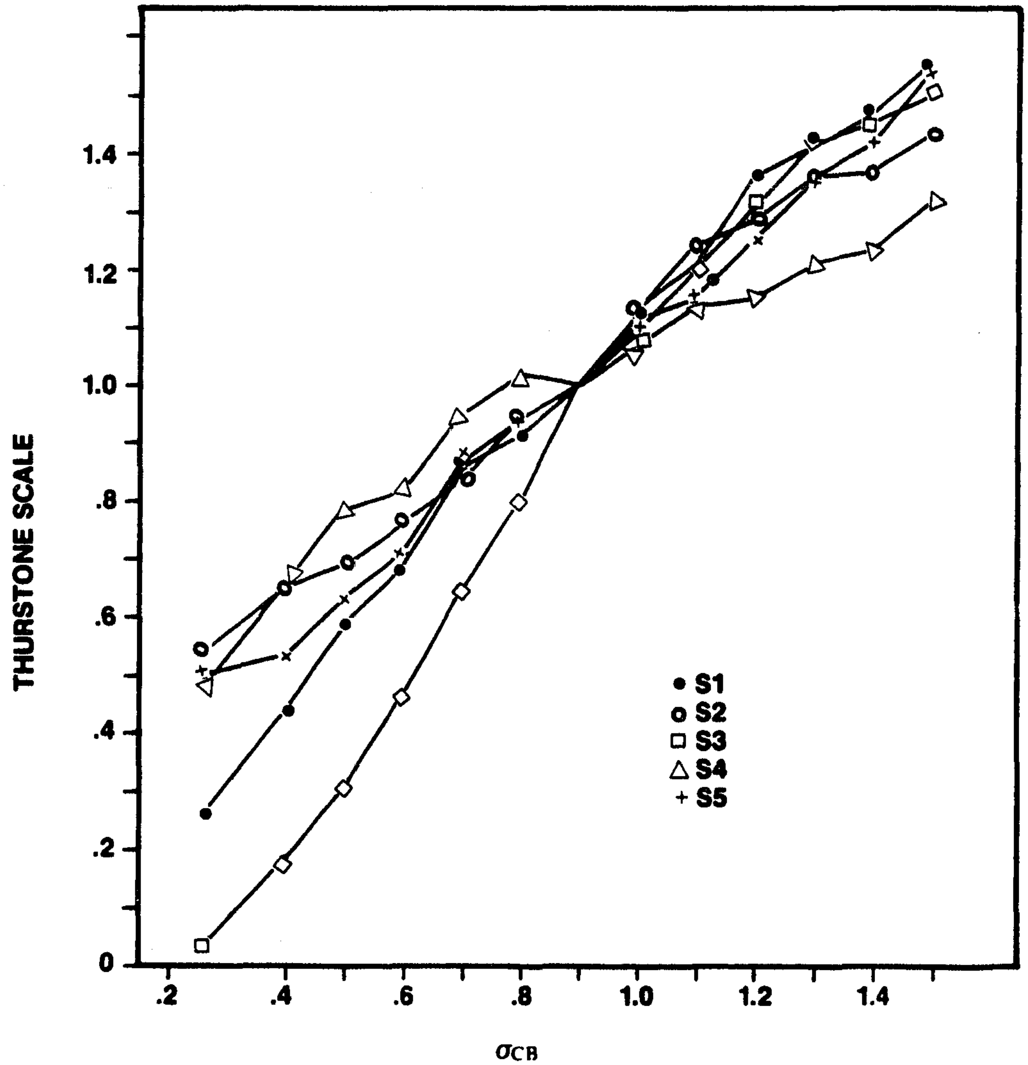
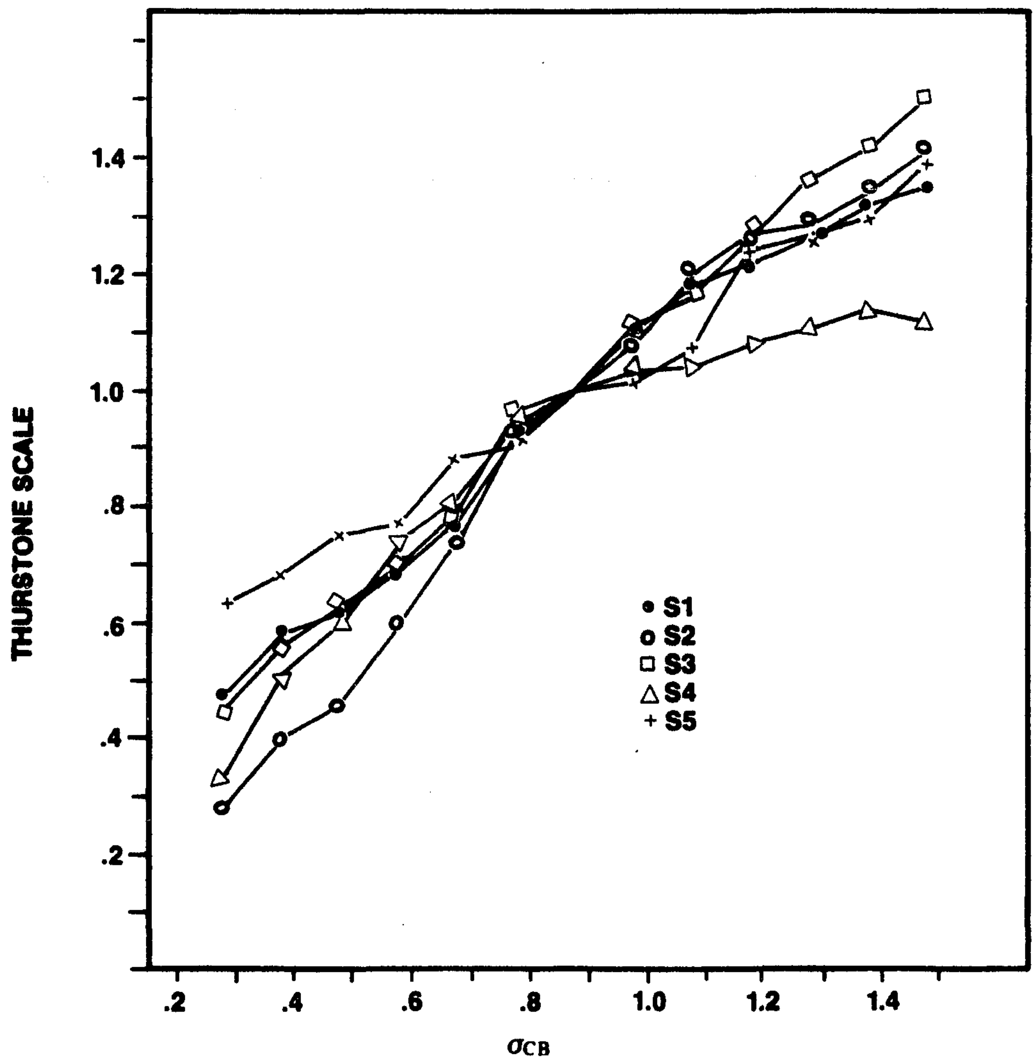


FIGURE 3.48

EXPERIMENT 4: STIMULUS SET 2: THURSTONE SCALE AS A FUNCTION OF σ_{CB} FOR INDIVIDUAL SUBJECTS.



Subject 3 tended to perform better than most and Subject 4 worst than most of the other subjects while Subject 1 gave roughly median performance data. Individual data are also well fit by second order polynomials.

Figures 3.49 and 3.50 permit comparison of the polynomials for the two stimulus sets. These polynomials are in good agreement. Creating polynomials based on the σ_{CBDIF} scale makes this good agreement even better, as Figure 3.54 shows.

3.5.4 Experiment 4 Discussion

These data are a strong demonstration of subjects' ability to judge amount of coloration. With the two small exceptions in Subject 4's data, these scales are monotonically increasing with small steps of σ_{CB}. Since we have just determined that the 13 stimuli cover a range of about 5 jnd, each of these steps is less than half a jnd. For the most part the scales are concave downward, which is consistent with obtaining paired-comparison data for a prothetic scale, due to smaller jnd's at the low end of a scale (Stevens, 1957 and 1971).

Although the range of coloration quantity determined from these data is between 5 and 6 jnd's, since the range of σ_{CB} or σ_{CBDIF} isn't known, it can't be concluded that

FIGURE 3.49

EXPERIMENT 4: THURSTONE SCALE AND SMOOTHED THURSTONE SCALE AS A FUNCTION OF σ_{CB} FOR STIMULUS SET 1.

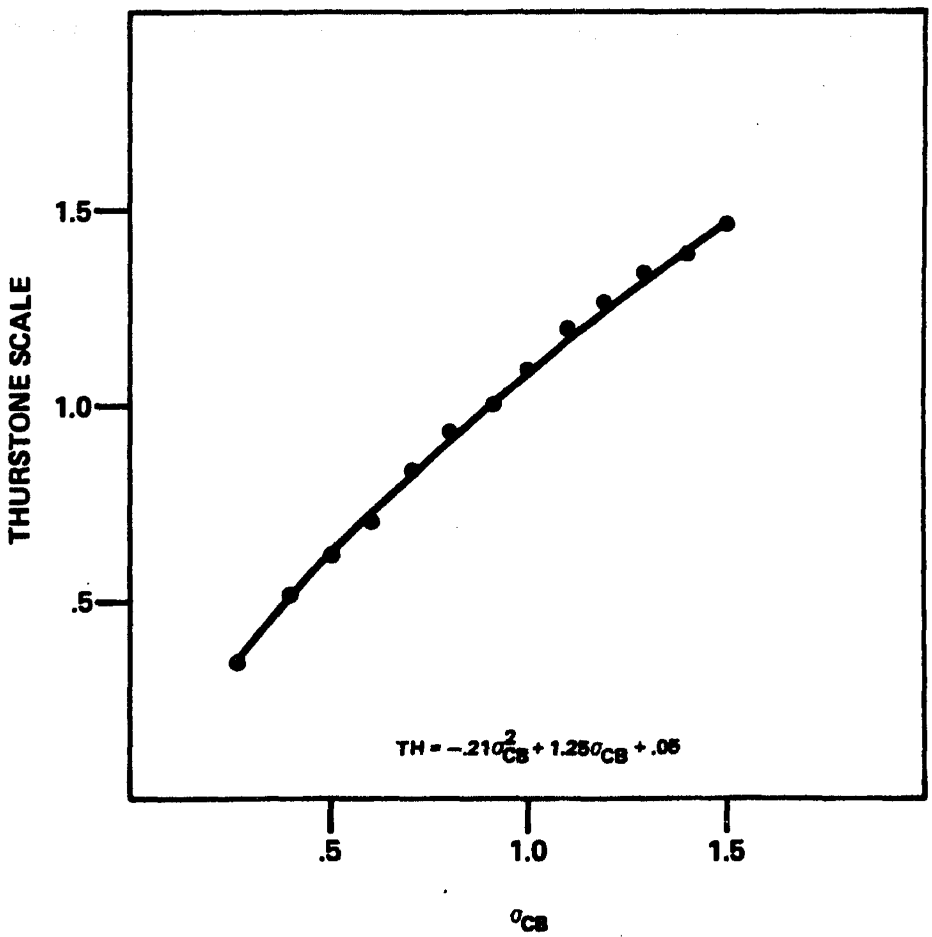


FIGURE 3.50

EXPERIMENT 4: THURSTONE SCALE AND SMOOTHED THURSTONE SCALE AS A FUNCTION OF σ_{CB} FOR STIMULUS SET 2.

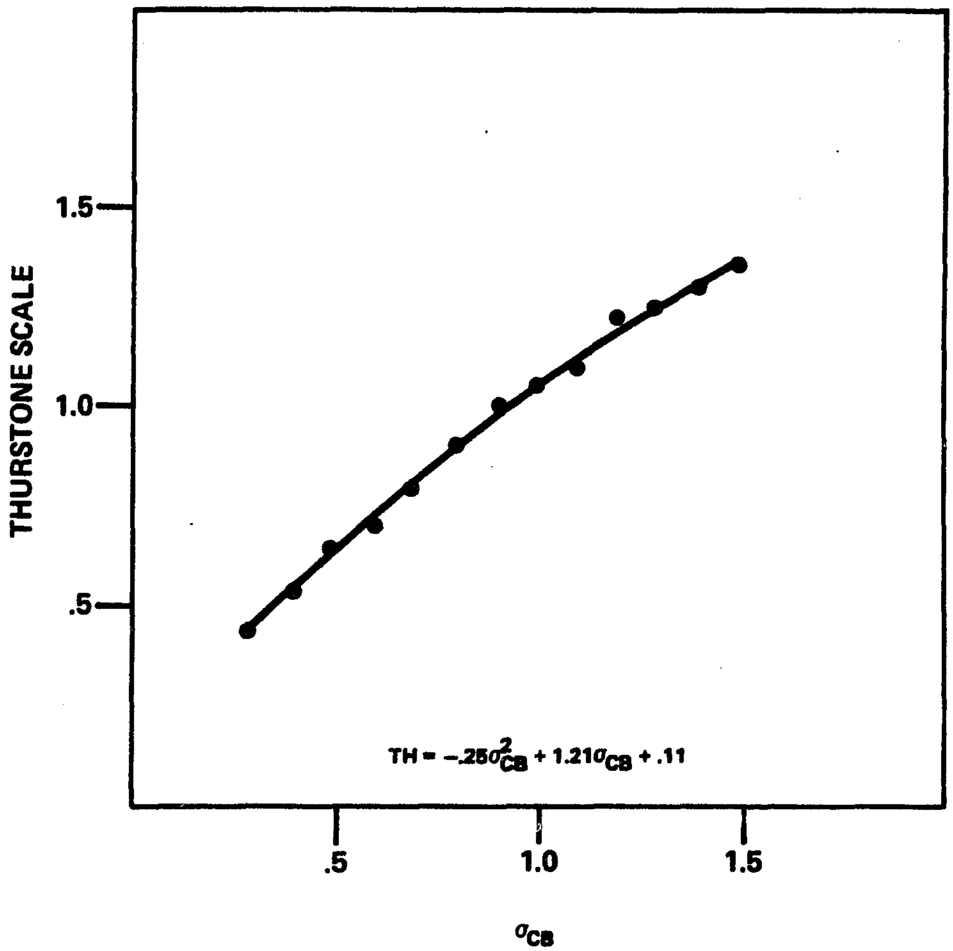


FIGURE 3.51

EXPERIMENT 4: THURSTONE SCALE AND SMOOTHED THURSTONE SCALE AS A FUNCTION OF σ_{CB} FOR SUBJECT 1 ON STIMULUS SET 1.

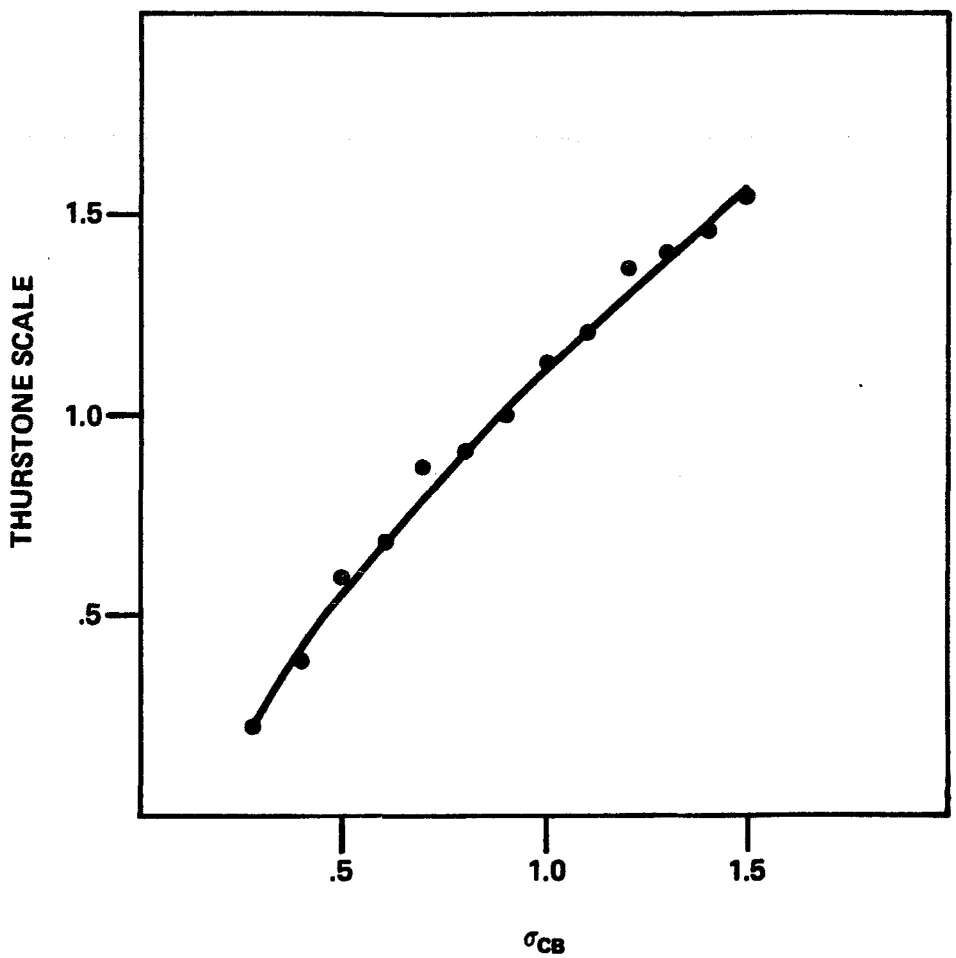


FIGURE 3.52

EXPERIMENT 4: THURSTONE SCALE AND SMOOTHED THURSTONE SCALE AS A FUNCTION OF σ_{CB} FOR SUBJECT 3 ON STIMULUS SET 1.

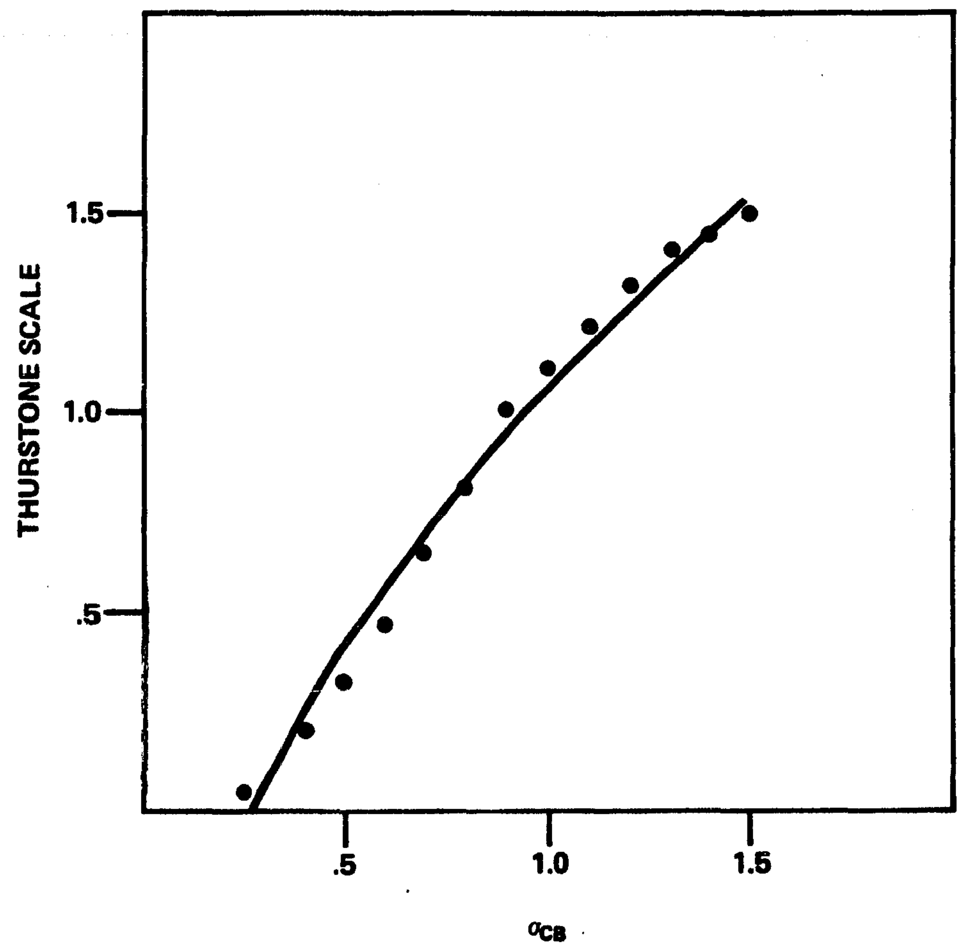


FIGURE 3.53

EXPERIMENT 4: THURSTONE SCALE AND SMOOTHED THURSTONE SCALE AS A FUNCTION OF σ_{CB} FOR SUBJECT 4 ON STIMULUS SET 1.

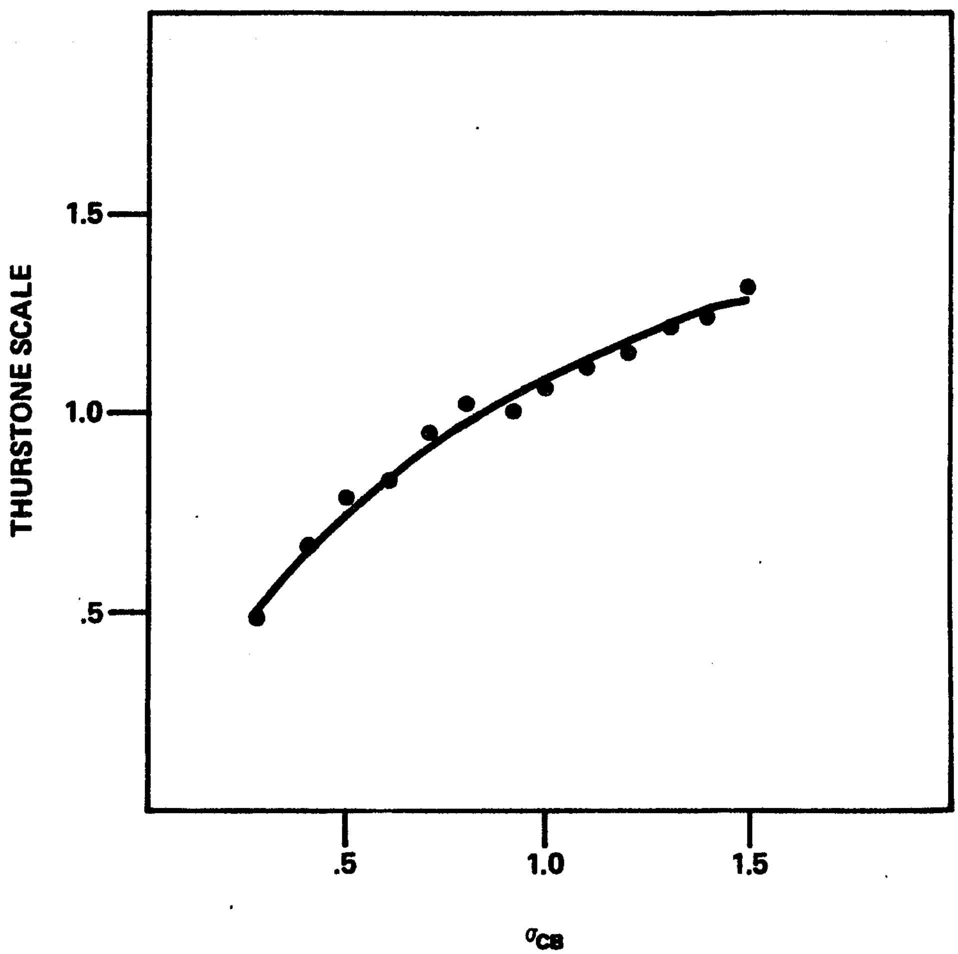
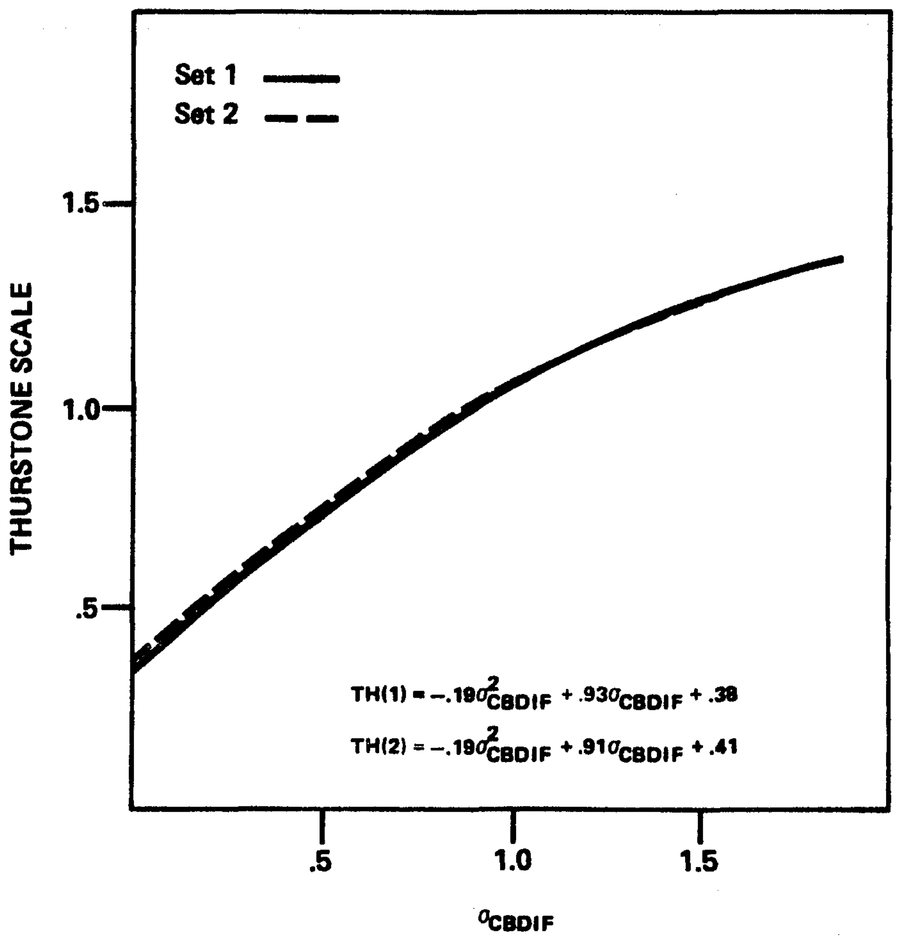


FIGURE 3.54

EXPERIMENT 4: SMOOTHED THURSTONE SCALES AS A FUNCTION OF σ_{CBDIF} FOR STIMULUS SETS 1 AND 2.



this is the limit of the range of amount of coloration. What is clear is that this range is very much smaller indeed than the range for the major auditory continua of pitch and loudness,

These data do not address the range of a scale or space which involves variation in coloration quality, and the most interesting questions concerning coloration may lie in that direction.

4. DISCUSSION

4.1 SUMMARY

A brief recapitulation is in order. The beginning premise was that the frequency domain characterization of reverberant sound is perceived as coloration. These effects consist of ripples in the room frequency response produced by the reflections of sound waves from the room's surfaces. Evidence from earlier work (Allen, et. al., 1979, and Berkley, 1978) suggested the standard deviation of a room's frequency response was a good measure of coloration. These studies have aimed at validating that observation and further exploring the perception of the frequency domain characterization of reverberant sound.

Using diotic reverberant noise stimuli, it was determined that coloration is not a single dimensional quantity, but rather a phenomenon with both quantity and quality. Quantity of coloration is related to the standard deviation of H(f), but is best described by a model which uses a critical band-like filter to smooth H(f) before calculating the standard deviation. This finding is very much in keeping with the literature concerning critical band effects (Atal, et. al., 1962, Zurek, 1976, and Koenig, 1979) and models (Patterson, 1974, and 1976, and Patterson and Nimmo-Smith, 1980); in retrospect these results are plausible and satisfying.

Quantity of coloration can be manipulated with little variation in coloration quality by varying the reflectivity, β, of the surfaces in a given room. Performing this manipulation results in a coloration quantity scale with a range of about 5 1/2 jnd's. Nonetheless, it should be understood that β is not a measure of coloration quantity when other room parameters are varied. It is the size of the reflections compared to the direct sound that determine the depth of the ripples in H(f) and the talker-microphone distance, Δ, is another important determiner of this.

Quality of coloration is the complex pitch-like quality associated with a particular room's H(f); by particular room is meant a room with fixed length, width and height parameters, and a fixed location of both the talker and microphone. Quality of coloration depends on the number and times of occurrence of the early reflections in a room and both the size of the room and the location of the talker and microphone relative to the room's surfaces will determine this complex pattern.

In the experiments reported in this paper subjects were asked to judge the size of the difference between two colored noise samples and the nature of this comparative process was revealed. Subjects responded as though they stored the smoothed room frequency responses, calculated the difference between them as a function of frequency, and then determined a weighted average difference with higher weights for larger differences. A summary statistic which does these things, the standard deviation of a difference distribution, was used as a measure of the difference between smoothed room spectra, σ_{CBDIF}, and it produced results which agreed well with the perceptual data. This difference measure describes the perceived frequency domain differences between reverberant rooms. From earlier work, this perceptual phenomenon is termed coloration and differences in quality and quantity of coloration have been distinguished and described quantitatively with this difference measure.

Nevertheless, there is a difference between coloration quantity and quality with respect to what has been learned about (1) how to manipulate them and (2) the underlying scales. To manipulate coloration quantity independently (nearly) of quality in a given room, β was manipulated in Experiment 4 to determine quantity jnd's. No comparable procedure has emerged for coloration quality. A unidimensional scale for coloration quantity, σ_{CB}, has been developed which predicts well quantity of coloration in a room. Coloration quality is multidimensional and it is not clear that analogous quality scales exist. This leaves the coloration quality phenomena with many more unanswered questions than coloration quantity.

4.2 COLORATION, REVERBERANT SPEECH AND HEARING IMPAIRMENTS

Next in order is a consideration of how and where these data fit into the literature on audition. First, what does this work have to do with the perception of reverberant speech? Reverberant speech, because of its non-stationary nature, includes perceivable effects in both the time and frequency domains. An effort to integrate time domain effects, i.e., the perception of echo, into the work done here is needed if this work is to be applied to reverberant speech. An obvious next step is to create a function based on the σ_{CB} and/or the σ_{CBDIF} measures which also includes a time domain measure, such as reverberation time, and use this to account for the earlier speech work. Very sketchy first efforts in this direction suggest that this modeling is not trivial, and it will not be attempted in this paper.

Even without a complete model for reverberant speech, these experiments have implications for some of the concerns discussed in the literature review. In Experiment 4 the perceptual range of coloration quantity was determined to be about 5 1/2 jnd's; this is very small compared to the number of jnd earlier cited for the perceptual scale of pitch (about 1200 jnd) or loudness (about 120 jnd). When the reduced sensitivity of hearing impaired listeners is taken into account, these data suggest they may be insensitive to coloration. However, because of the exponential slopes of its filter, the coloration measure reflects mostly low frequency perturbations, and this is the frequency range to which sensory-impaired individuals are most likely to be sensitive. Thus, the extent to which these individuals perceive coloration is an unanswered empirical question. These data support the prediction that it is listeners with good residual low frequency hearing (under 1000 Hz) that will be most sensitive to coloration.

One qualification needs to be made about the above prediction. The stimulus level in these studies was 82 dB SPL with a spectrum level of about 50 dB. Both the work of Patterson (1974) and Egan and Hake (1950) support the widening of the critical band at levels above 40 dB spectrum level, and it would be expected that lower stimulus levels would produce more sensitivity at high frequencies. In most cases stimulus levels used with the hearing-impaired will be at relatively high levels. This consideration of stimulus SPL also implies that the exponents for σ_{CB} are unlikely to be constant over SPL's. At levels under 40 dB spectrum level, it would be expected that the exponents would be more nearly equal. An expectation of greater sensitivity to fast spectral variations suggests that the σ_{CB} exponents should both be high (about 16) rather than low (about 8) at lower SPL's.

4.3 MULTIDIMENSIONAL SCALING AND σ_{CBDIF}

The efficacy of σ_{CBDIF} has implications for the interpretations of multidimensional scaling experiments. In general these experiments are interpreted in terms of underlying dimensions which it is believed are meaningful to the judges. These experiments with σ_{CBDIF} suggest this need not be true. A six-dimensional solution for Experiment 2's data matched (though more and more poorly for higher dimensions) a six-dimensional theoretical solution based on σ_{CBDIF}, and σ_{CBDIF} is a summary of a continuous function of frequency. This single difference measure accounts for a multidimensional solution. As was discussed in Experiment 2, the potential for discriminating among rooms is as great as the ear's capability for frequency analysis; the solution space for a set of rooms may simply be a difference space based on σ_{CBDIF}. While it is possible to seek perceptual dimensions within σ_{CBDIF}, it is a violation of parsimony to try to use many underlying variables when one will do. From this point of view, the only commonality expected among solution spaces for different room sets is some indication of coloration quantity in the solution. It is an unanswered question whether coloration quality must be interpreted in this way or whether there is a small number of dimensions which can be used to account for it.

The approach considered above, that a solution space is a difference space without necessary underlying dimensions, leads to the conclusion that the psychological reality of underlying dimensions is an interesting empirical finding when it occurs rather than a necessity for a multidimensional scaling solution.

The comparison of the subjective and σ_{CBDIF} solutions for Experiments 1 and 2 suggest another observation about multidimensional scaling results. A common intuition about such data is that the low dimensions reflect the true configuration of points while error variance is displayed in the higher dimensions of a solution. Such an interpretation is in keeping with the elbow in the stress function discussed earlier. What a comparison of the subjective and theoretical solutions in the present experiments suggests is that there is significant subject variability displayed in every dimension, and only in the lower dimensions of the solution are the perceptual effects large enough to stand out. Such an interpretation makes a gradual stress function the expected result in cases where the size of the experimental effect (about 5 jnd for a coloration spoke) is not a great deal larger than subject variability.

4.4 CRITICAL BANDS AND σ_{CBDIF}

The difference measure σ_{CBDIF} has been discussed in terms of its implications for multidimensional scaling, but it should also be considered with respect to the particular auditory mission it fulfilled: it demonstrated a strong qualitative superiority of σ_{CB} over σ_{10}. In Experiments 2 and 3, σ_{CB} regressed slightly better or correlated slightly more highly than σ_{10} with the data, but the statistics were never different enough to make a compelling case for σ_{CB}. The σ_{CBDIF} space, however, clearly captured features of the data solution space better than σ_{10DIF}.

One further comment about σ_{CBDIF} is in order. In spite of the fact that it is this measure that provides the most compelling evidence for a choice between σ measures, σ_{CBDIF} is not a measure of the capability of the ear. It rather describes the decision process that occurs after peripheral auditory processing is completed.

These experiments yield a satisfactory measure of coloration by developing a model of the critical band. The model used has two steep-sided slopes in common with Patterson's modeling work. It also has a steeper higher than low frequency slope, in keeping with neural tuning curve data. The Moore and Glasberg (1983) summary of critical band width's at low frequencies showed the critical band continuing to decrease with frequency for frequencies below 500 Hz. This would be an interesting refinement to apply to the model developed in this paper. Other interesting directions are to (1) use other models (e.g. Patterson's) to predict these data and (2) use the model developed here to predict critical band data obtained in other contexts.

APPENDIX A

EXPERIMENT 4 TABLES

The tables in Appendix A contain support information for Experiment 4. Tables A.1 and A.1 contain the σ_{CBDIF} matrix for Stimulus Set 1 and 2, respectively. These tables demonstrate that σ_{CBDIF} is not an interval scale. For example, for Stimulus Set 1 the range of σ_{CBDIF} if successive steps are summed is 1.88 while the σ_{CBDIF} between Stimulus 1 and 13 in Set 1 is 1.41. The cumulative Thurstone scales in Tables 3.13 and 3.14 were obtained by averaging values for σ_{CBDIF} for the i $\pm$ 1, $\pm$ 2, $\pm$ 3 stimulus pairs.

Tables A.3 and A.4 contain the second order polynomial functions for each subject and for the group data as a function of σ_{CB} and σ_{CBDIF}, respectively.

Table A.1 σ_{CBDIF} for all Pairs of Stimulus

Set 1, Experiment 4

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.0												
2	.164	0.0											
3	.278	.120	0.0										
4	.386	.238	.121	0.0									
5	.502	.366	.255	.137	0.0								
6	.611	.487	.382	.268	.133	0.0							
7	.722	.610	.512	.404	.274	.144	0.0						
8	.845	.745	.655	.555	.434	.309	.169	0.0					
9	.957	.868	.785	.693	.581	.463	.317	.161	0.0				
10	1.068	.988	.913	.829	.726	.615	.484	.322	.163	0.0			
11	1.183	1.113	1.045	.961	.875	.772	.648	.491	.335	.174	0.0		
12	1.297	1.236	1.177	1.109	1.024	.929	.813	.664	.514	.356	.185	0.0	
13	1.407	1.355	1.305	1.246	1.172	1.087	.981	.843	.701	.552	.387	.207	0.0

Table A.2 σ_{CBDIF} for all Pairs of Stimulus Set 2, Experiment 4

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.0												
2	.175	0.0											
3	.288	.115	0.0										
4	.395	.226	.112	0.0									
5	.495	.331	.219	.108	0.0								
6	.597	.440	.332	.223	.116	0.0							
7	.692	.542	.439	.334	.229	.115	0.0						
8	.792	.652	.554	.454	.353	.241	.127	0.0					
9	.888	.757	.665	.570	.472	.362	.250	.124	0.0				
10	.989	.869	.783	.692	.599	.492	.381	.256	.133	0.0			
11	1.096	.986	.906	.820	.730	.626	.517	.393	.270	.138	0.0		
12	1.190	1.088	1.013	.932	.845	.743	.636	.513	.392	.260	.122	0.0	
13	1.304	1.212	1.143	1.066	.983	.885	.780	.659	.539	.410	.273	.152	0.0

Table A.3 Polynomials for Stimulus Sets 1 and 2, Experiment 4, as a Function of the σ_{CB} scale

Stimulus Set 1

$$Y(1,1) = -.28\sigma^2_{CB} + 1.54\sigma_{CB} - .12$$

$$Y(1,2) = -.81\sigma^2_{CB} + .92\sigma_{CB} + .27$$

$$Y(1,3) = -.46\sigma^2_{CB} + 2.10\sigma_{CB} -.57$$

$$Y(1,4) = -.35\sigma^2_{CB} + 1.22\sigma_{CB} + .22$$

$$Y(1,5) = .03\sigma^2_{CB} + .81\sigma_{CB} + .24$$

$$Y(1,Group) = -.21\sigma^2_{CB} + 1.25\sigma_{CB} + .05$$

Stimulus Set 2

$$Y(2,1) = -.23\sigma^2_{CB} + 1.18\sigma_{CB} + 1.27$$

$$Y(2,2) = -.43\sigma^2_{CB} + 1.75\sigma_{CB} -.21$$

$$Y(2,3) = -.07\sigma^2_{CB} + 1.01\sigma_{CB} + .17$$

$$Y(2,4) = -.69\sigma^2_{CB} + 1.84\sigma_{CB} -.11$$

$$Y(2,5) = .09\ \sigma^2_{CB} + .48\sigma_{CB} + .49$$

$$Y(2,Group) = -.25\sigma^2_{CB} + 1.21\sigma_{CB} + .11$$

Table A.4 Polynomials for Stimulus Sets 1 and 2, Experiment 4, as a Function of the σ_{CBDIF} scale

Stimulus Set 1

$Y(1,1) = -.23\sigma^2_{CBDIF} + 1.11\sigma_{CBDIF} + .28$

$Y(1,2) = -.13\sigma^2_{CBDIF} + .76\sigma_{CBDIF} + .51$

$Y(1,3) = -.36\sigma^2_{CBDIF} + 1.49\sigma_{CBDIF} - .03$

$Y(1,4) = -.19\sigma^2_{CBDIF} + .93\sigma_{CBDIF} + .38$

$Y(1,5) = -.08\sigma^2_{CBDIF} + .73\sigma_{CBDIF} + .46$

$Y(1,Group) = -.19\sigma^2_{CBDIF} + .93\sigma_{CBDIF} + .38$

Stimulus Set 2

$Y(2,1) = -.18\sigma^2_{CBDIF} + .90\sigma_{CBDIF} + .42$

$Y(2,2) = -.31\sigma^2_{CBDIF} + 1.27\sigma_{CBDIF} + .20$

$Y(2,3) = -.08\sigma^2_{CBDIF} + .84\sigma_{CBDIF} + .42$

$Y(2,4) = -.45\sigma^2_{CBDIF} + 1.20\sigma_{CBDIF} + .33$

$Y(2,5) = -.25\sigma^2_{CBDIF} + .47\sigma_{CBDIF} + .61$

$Y(2,Group) = -.19\sigma^2_{CBDIF} + .91\sigma_{CBDIF} + .41$

REFERENCES

1. Allen, J. B. and Berkley, D. A. "Image Method for Efficiently Simulating Small-Room Acoustics," J. Acoust. Soc. Amer., 65, 943-950, 1979.

2. Allen, J. B., McDermott, B., and Berkley, D. A., "A Method for Measuring Subjective Perception and Preference of Small Room Reverberation," unpublished manuscript, 1978.

3. Atal, B. S., Schroeder, M. R., and Kuttruff, H., "Perception of Coloration in Filtered Gaussian Noise---Short Time Spectral Analysis by the Ear," Proceedings of the Fourth International Congress on Acoustics, 21-28, August, 1962.

4. Berkley, D. A. "Normal Listeners in Typical Rooms: Reverberation Perception, Simulation and Reduction," Acoustical Factors Affecting Hearing and Performance, ed. by G. A. Studebaker and I Hochberg, 1978.

5. Bilsen, F. A. and Ritsma, R. J., "Some Parameters Influencing the Perceptibility of Pitch," J. Acoust. Soc. Amer., 47, 469-475, 1970.

6. Bloom, P. J. "Evaluation of a Dereverberation Process by Normal and Impaired Listeners," IEEE Transactions, 500-503, 1980.

7. Carroll, J. D. and Chang, J. J. "Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of Eckart-Young Decomposition," Psychometrika, 35, 283-319,1970.

8. Carroll, J. D. and Wish, M. "Models and Methods for Three-Way Multidimensional Scaling" in Contemporary Developments in Mathematical Psychology, Vol.2, ed. by D. H. Krantz, R. C. Atkinson, R. D. Luce and P. Suppes, 57-105, 1974.

9. Coombs, C. H. A Theory of Data, John Wiley and Sons, 1964.

10. Dirks, D. D. and Wilson, R. H. "The Effect of Spatially Separate Sound Sources on Speech Intelligibility," J. Spch. Hrg. Res., 12, 5-38, 1969.

11. Egan, J. P. and Hake, H. W. "On the Masking Pattern of a Simple Auditory Stimulus," J. Acoust. Soc. Amer, 22, 622-630, 1950.

12. Finitzo-Hieber, T. and Tillman, T. W. "Room Acoustics Effects on Monosyllabic Word Discrimination Ability for Normal and Hearing-Impaired Children," J. Spch. Hrg. Res., 21, 440-458, 1978.

13. Fletcher, H. "Auditory Patterns," Review of Modern Physics, 12, 47-65, 1940.

14. Garner, W. R. "A Technique andd a Scale for Loudness Measurement," J. Acoust. Soc. Amer., 26, 73-88, 1954.

15. Gelfand. S. A. and Hochberg, I. "Binaural and Monaural Speech Discrimination under Reverberation," Audiology, 15, 72-84, 1976.

16. Gelfand, S. A. and Silman, S. "Effects of Small Room Reverberation upon the Recognition of Some Consonant Features," J. Acoust. Soc. Amer., 66, 22-29, 1979.

17. Green, D. M. "Masking with Two Tones," J. Acoust. Soc. Amer., 37, 802-813, 1965.

18. Green, D. M. *An Introduction to Hearing*, Lawrence Erlbaum Associates, 1976.

19. Green, D. M., McKey, M. J., and Licklider, J. C. R. "Detection of a Pulsed Sinusoid in Noise as a Function of Frequency," *J. Acoust. Soc. Amer.*, *31*, 1446-1452, 1959.

20. Green, D. M., Luce, R. D., and Smith, A. F. "Individual Magnitude Estimates for Various Distributions of Signal Intensity," *Perception and Psychophysics*, *27*, 483-488, 1980.

21. Greenwood, D. D. "Auditory Masking and the Critical Band," *J. Acoust. Soc. Amer.*, *33*, 484-502, 1961.

22. Haas, H. "Influence of a Single Echo on the Audibility of Speech," *J. Audio Engineering Society*, *20*, 146-159, 1972.

23. Hamilton, P. M. "Noise Masked Thresholds as a Function of Tonal Duration and Masking Noise Band Width," *J. Acoust. Soc. Amer.*, *29*, 506-511, 1951.

24. Harris, J. D. "Pitch Discrimination," J. Acoust. Soc. Amer., 24, 750-755, 1952.

25. Hawkins, J. E., Jr., and Stevens, S. S. "The Masking of Pure Tones and of Speech by White Noise,", J. Acoust. Soc. Amer., 22, 1950, 6-13.

26. Hirsh, I. J. "The Influence of Interaural Phase on Interaural Summation and Inhibition,", J. Acoust. Soc. Amer., 20, 536-544, 1948.

27. Houtgast, T. "Auditory Filter Characteristics Derived from Direct-Maskign Data and Pulsation Threshold Data with a Rippled Noise Masker," J. Acoust. Soc. Amer, 62, 409-415, 1977.

28. ISO R532, Method A, 1956, and Method B, 1958, International Standards Organization.

29. Jerger, J. and Dirks, D. D. "Binaural Hearing Aids: An Enigma," J. Acoust. Soc. Amer., 33, 537-538, 1961.

30. Jerger, J., Carhart, R., and Dirks, D. D. "Binaural Hearing Aids and Speech Intelligibility," J. Spch. Hrg. Res., 4, 137-148, 1961.

31. Jesteadt, W., Weir, C. C., and Green, D. M. "Intensity Discrimination as a Function of Frequency and Sensation Level," J. Acoust. Soc. Amer., 61, 169-177, 1977.

32. Jetzt, J. J. "Critical Distance Measurement of Rooms from the Sound Energy Spectral Response," J. Acoust. Soc. Amer., 65, 1204-1211, 1979.

33. Johnstone, B. M., and Boyle, A. J. "Basilar Membrane Vibration Examined with the Moessbauer Technique," Science, 158, 389-390, 1967.

34. Kiang, N. Y-S. "Discharge of Single Fibers in the Cat's Auditory Nerve," Research Monograph, MIT Press, 35, 1965.

35. Knudsen, V. O. "The Hearing of Speech in Auditoria," J. Acoust. Soc. Amer., 1, 56-82, 1929.

36. Koenig, A. H. "Diotic and Dichotic Perception of Single Echoes," City University of New York doctoral dissertation, 1979.

37. Koenig, W. "Subjective Effects in Binaural Hearing," J. Acoust. Soc. Amer., 22, 61-62, 1950.

38. Krantz, D. H. "Integration of Just-Noticeable Differences," J. Math. Psychol., 8, {591-599, 1971.

39. Kruskal, J. B. "Multidimensioal Scaling by Optimizing Goodness-of-fit to a Non-Metric Hypothesis," Psychometrika, 29, 1-27, 1964a.

40. Kruskal, J. B. "Non-metric Multidimensional Scaling: A Numerical Method," Psychometrika, 29, 115-129, 1964b.

41. Kruskal, J. B. and Wish, M. Multidimensional Scaling, Paper 11, Sage University Press, 1978.

42. Kruskal, J. B., Young, F. W. and Seery, J. B. "How to Use KYST-2, A very Flexible Program to do Multidimensional Scaling and Unfolding," Bell Laboratories unpublished manuscript, 1973.

43. Leshowitz, B. and Wightman, F. L. "On Frequency Masking with Continuous Sinusoids," J. Acoust. Soc. Amer., 49, 1180-1190, 1971.

44. Licklider, J. C. R. "The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise," J. Acoust. Soc. Amer., 20, 150-159, 1948.

45. Lochner, J. P. A. and Burger, J. F. "The Subjective Masking of Short Time Delayed Echoes by their Primary Sounds and their Contribution to the Intelligibility of Speech," Acustica, 8 1-10, 1958.

46. Luce, R. D. "Thurstone's Discriminal Processes Fifty Years Later," Psychometrika, 42, 461-489, 1977.

47. Luce, R. D. and Edwards, W. "The Derivation of Subjective Scales from Just Noticeable Differences," Psych. Rev., 65, 222-237, 1958.

48. Margolis, R. H. and Small, A. M. "The Measurement of Critical Masking Bands," J. Spch. Hrg. Res., 18, 571-587, 1975.

49. Moncur, J. P. and Dirks, D. D. "Binaural and Monaural Speech Intelligibility in Reverberation," J. Spch. Hrg. Res., 10, 186-195, 1967.

50. Moore, B. C. J. and Glasberg, B. R. "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns," J. Acoust. Soc. Amer., 74, 750-753, 1983.

51. Náběլek, A. K. and Mason, David "Effect of Noise and Reverberation on Binaural and Monaural Word Identification by Subjects with Various Audiograms," J. Spch. Hrg. Res., 24, 375-383, 1981.

52. Náběլek, A. K. and Pickett, J. M. "Reception of Consonants in a Classroom as Affected by Monaural and Binaural Listening, Noise, Reverberation and Hearing Aids," J. Acoust. Soc. Amer., 56, 628-639, 1974a.

53. Náběլek, A. K. and Pickett, J. M. "Monaural and Binaural Speech Perception through Hearing Aids under Noise and Reverberation with Normal and Hearing-Impaired Listeners," J. Spch. Hrg. Res., 17, 7274-739, 1974b.

54. Náběլek, A. K. and Robinette, L. "Influence of the Precedence Effect on Word Identification by Normally-hearing and Hearing-Impaired Subjects," J. Acoust. Soc. Amer., 63, 187-194, 1978.

55. Newman, A. "Effect of Reverberation on Phoneme Discrimination of Children - A Developmental Study," City University of New York doctoral dissertation, 1982.

56. Patterson, R. D. "Auditory Filter Shape," J. Acoust. Soc. Amer., 55, 802-809, 1974.

57. Patterson, R. D. "Auditory Filter Shapes Derived with Noise Stimuli," J. Acoust. Soc. Amer., 59, 640-654, 1976.

58. Patterson, R. D. and Henning, G. B. "Stimulus Variability and Auditory Filter Shape," J. Acoust. Soc. Amer., 62, 649-654, 1977.

59. Patterson, R. D. and Nimmo-Smith, I. "Off-frequency Listening and Auditory Filter Asymmetry," J. Acoust. Soc. Amer, 67, 229-245, 1980.

60. Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. "The Deterioration of Hearing with Age: Frequency Selectivity, the Critical Ratio, the Audiogram, and the Speech Threshold," J. Acoust. Soc. Amer., 72, 1788-1803, 1982.

61. Poulton, E. C. "The New Psychophysics: Six Models for Magnitude Estimation" Psychol. Bulletin, 69, 1-19, 1968.

62. Rhode, W. S. "Observations of the Vibration of the Basilar Membrane on Squirrel Monkeys Using the Moessbauer Technique," J. Acoust. Soc. Amer., 49, 1218-1231, 1971.

63. Russell, I. J. and Sellick, P. M. "Tuning Properties of Cochlear Hair Cells," Nature, 267, 857-860, 1977.

64. Schafer, T. H., Gales, R. S., Shewmaker, C. A. and Thompson, P. O. "The Frequency Selectivity of the Ear as Determined by Masking Experiments," J. Acoust. Soc. Amer., 22, 490-496,1950.

65. Scharf, B. "Critical Bands," in Foundations of Modern Auditory Theory, Vol. I, ed. by J. V. Tobias, Academic Press, 159-202, 1970.

66. Shepard, R. N. "Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function, I," Psychometrika, 27, 125-140, 1962a.

67. Shepard, R. N. "Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function, II," Psychometrika, 27, 219-246, 1962b.

68. Schroeder, M. R. "New Method of Measuring Reverberation Time," J. Acoust. Soc. Amer., 37, 409-412, 1965.

69. Small, A. M. and Tyler, R. S. "Additive Masking Effects of Noise Bands at Different Cycles," J. Acoust. Soc. Amer., 63, 894-904, 1978.

70. Stevens, S. S. "The Attributes of Tones," Proc. Natn. Acad. Sci. U. S. A., 20, 457-459, 1934.

71. Stevens, S. S., "The Measure of Loudness," J. Acoust. Soc. Amer., 27, 815-829, 1955.

72. Stevens, S. S. "Calculation of the Loudness of a Complex Noise," J. Acoust. Soc. Amer., 28, 807-832, 1956.

73. Stevens, S. S. "On the Psychophysical Law," Psych. Rev., 64, 153-181, 1957.

74. Stevens, S. S. "Issues in Psychphysical Measurement," Psych. Rev., 78, 426-450. 1971.

75. Stevens, S. S. and Galanter, E. H. "Ratio Scales and Category Scales for a Dozen Perceptual Continua," J. Exp. Psychol., 54, 377-411, 1957.

76. Stevens, S. S. and Newman, E. B. "The Localization of Actual Sources of Sound," Amer. J. Psychol., 48, 297-306, 1936.

77. Stevens, S. S. and Volkmann, J. "The Relation of Pitch to Frequency: A Revised Scale," Amer. J. Psychol., 53, 329-353, 1940.

78. Swets, J. A., Green, D. M., and Tanner, Jr., W. P. "On the Width of Critical Bands," J. Acoust. Soc. Amer., 34, 108-113, 1962.

79. Thurstone, L. L. "A Law of Comparative Judgment," Psych. Rev., 34, 273-286, 1927a.

80. Thurstone, L. L. "Three Psychological Laws," Psych. Rev., 34, 424-432, 1927b.

81. Torgerson, W. S. _Theory and Methods of Scaling_, John Wiley and Sons, 1958.

82. Wallach, H., Newman, E. B., and Rosenzweig, M. R. "The Precedence Effect in Sound Localization," _Amer. J. of Psychol._, _62_, 315-336, 1949.

83. Ward, W. D. "Musical Perception," _Foundations of Modern Auditory Theory_, _Vol. I_, ed. by J. V. Tobias, Academic Press, 405-447, 1970.

84. Weber, D. L. "Growth of Masking and the Auditory Filter," _J. Acoust. Soc. Amer._, _62_, 424-429, 1977.

85. Webster, J. C., Miller, P. H., Thompson, P. O. and Davenport, E. W. "The Frequency Selectivity of the Ear as Determined by Masking Experiments," _J. Acoust. Soc. Amer._, _24_, 147-152, 1952.

86. Zurek, P. "An Investigation of the Binaural Perception of Echoed Sound," Arizona State University doctoral dissertation, 1976.

87. Zwicker, E. "Subdividion of the Audible Frequency Range into Critical Bands," _J. Acoust. Soc. Amer._, _33_, 248, 1961.

88. Zwicker, E., Flottorp, G., and Stevens, S. S. "Critical Band Width in Loudness Summation," J. Acoust. Soc. Amer., 29, 548-557, 1957.

89. Zwicker, E. and Scharf, B. "A Model of Loudness Summation," Psych. Rev., 72, 3-26, 1965.

90. Zwislocki, J. J. "Absolute Scaling," J. Acoust. Soc. Amer., 63, S316, 1978.