A Computer-Assisted Program in Timbral Ear Training — A Preliminary Study

by

René Quesnel
Department of Music Theory
McGill University, Montréal

September 1990.

A Thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of the requirements
for the degree of Master of Music

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr Wieslaw Woszczyk, who believed in the project from the start. His insights and advice were most helpful. I owe many thanks to Wayne Zelmer, the technician of the McGill Recording Studios, for his invaluable assistance in the assembling and maintenance of the system used for this thesis. I also want to thank my students in the Technical Ear Training course who patiently used early and unstable versions of the software. Finally, I would like to thank the Music Faculty at McGill University for providing the equipment used for this research.

J'aimerais addresser un remerciement très spécial à mon frère Roger qui, le premier, m'a donné la chance d'apprendre l'informatique, ainsi qu'à ma famille pour son support continu Et à Marie, pour m'avoir donné une raison de plus de mener à terme ce projet.

TABLE OF CONTENTS

Abstract

One of the main responsibilities of a sound engineer is to control the quality of the sound during the recording process. An important qualitative aspect of a recording, besides the musicality of the performance, is its timbral content. Proper level and spectral balance between the mixed elements of the recording and the absence of extraneous noises and distortion are key elements. Therefore, timbre perception acuity is an essential skill for sound engineers.

This thesis proposes a computer-assisted system as a training tool for developing and maintaining aural skills related to timbre perception. A set of criteria for the design of such a system based on current knowledge in timbre perception is presented and an exploratory implementation is described. Limits of the current system are discussed and areas that need further investigation are identified.

Résumé

Une des tâches principales de l'ingénieur du son est de contrôler la qualité du son tout au long d'un enregistrement. Mise à part la musicalité de l'interprétation, un des éléments qualitatifs importants d'un enregistrement sonore est son timbre. L'équilibre des niveaux sonores et la répartition spectrale entre les divers éléments d'un enregistrement, de même que l'absence de distortion et de bruits indésirables sont particulièrement importants. L'acuité de la perception du timbre est donc une habileté essentielle à l'ingénieur du son.

Un système assisté par ordinateur est proposé comme outil de formation pour le développement et la préservation des facultés auditives reliées à la perception du timbre. Un ensemble de critères pour l'élaboration d'un tel système est présenté, basé sur les connaissances actuelles sur le timbre. Un système préliminaire est décrit. Les lacunes de ce système sont examinées et les aspects nécessitant de plus amples recherches sont identifiés.

INTRODUCTION

Timbre perception acuity and a good timbral memory are essential tools that sound engineers must develop and maintain in order to achieve excellence in their work. At every stage of the recording process, from microphone placement to final mixing, sound engineers must rely on their ability to perceive fine timbral changes and on their knowledge of the physical parameters of sound responsible for a given sound quality, in order to successfully shape the timbre content of the recording.

A substantial body of research on timbre perception has developed over the past 30 years. Chapter two of this thesis will review the current knowledge on timbre. When one considers the great importance of timbre perception in the field of sound engineering, it is surprising to realize how little attention the topic has attracted in books and journals on sound recording. Chapter 3 will discuss aspects of sound recording that are related to timbre perception.

If timbre perception is a perceptual skill, it should be possible to develop that skill so that engineers hear better and make better recordings. Topics in perceptual learning and memory will be discussed in chapter 4. Chapter 5 will specifically address the subject of timbral ear training. Attempts to develop aural skills for sound engineers have already been reported in the literature. Some of these were successful; in particular, a course was developed at the Chopin Academy of Music in Warsaw (Letowski 1985). The results presented in that paper indicate that very high levels of performance can be achieved with proper training.

The present thesis proposes criteria for the building of a computer-controlled system for timbral ear training. Advantages of using computer technology for timbral ear training will be presented. A preliminary computer-assisted system was built and results from informal tests will be examined. Finally, the limits of the current system together with suggestions for further research are presented in the final chapter. To the knowledge of the author, no other computer-assisted system for timbral ear training has been developed elsewhere.

CHAPTER 1

TIMBRE PERCEPTION: CURRENT STATE OF KNOWLEDGE

1.0 Introduction

Of all the attributes of complex tones, timbre is probably the most poorly understood one. Its multidimensional character, its complex interactions with other perceptual attributes, and the high complexity of the signals encountered in musical contexts have contributed to the difficulties involved in that research area. The recent development of computer-based tools for sound analysis and synthesis allow a more accurate study and description of the phenomena associated with the perception of timbre. However, the physical and acoustical factors related to the perception of timbre are better understood than the perceptual dimensions or scales along which timbre can be evaluated and classified.

1.1 Definitions and concepts

It seems impossible, with the current state of knowledge, to define accurately and unambiguously what is timbre. It is nevertheless necessary at the onset of this discussion to at least delimit the possible meanings the word can take. It must be made clear that timbre is not a physical characteristic of sound: it is an attribute of perception[1]. Strictly speaking,

[1]"Hauteur et timbre ne sont pas des paramètres physiques: ce sont des attributs de la perception" (Risset 1978, 1).

timbre is not a given distribution of spectral energy; it is the resulting aural perception of that spectrum when made available to the hearing system. Timbre is generally considered to be, together with pitch, loudness, and perceptual duration, one of the main subjective attributes of complex tones. But as mentioned above, the concept of timbre is not well defined and researchers do not always agree on what the word "timbre" should include or know what it should mean.

Words like tone quality, sound quality, tone color, texture are often used interchangeably with timbre, the precise meaning varying to a certain degree among individuals Timbre may refer to the characteristic tone quality of a particular musical instrument or family of instruments as opposed to another instrument or family of instruments. Musicians will often use the word timbre to refer to the tone colors of the different registers of a given instrument (e.g., the tone quality of the chalumeau register vs. the tone quality of the clarion register of the clarinet). Timbre is also used to refer to the characteristic "sound" of one player compared to the sound of another player of the same instrument. Timbre refers also to the palette of instrumental colors available to the orchestrator, the composer, and the conductor.

Attempts have been made to standardize the definition of timbre but with limited success. In almost every published study on timbre perception, the definition provided by the American Standards Association (1960) is cited. It states that:

> Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same pitch and loudness are dissimilar ... Timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus.

This definition has been criticized by several authors (Bregman 1990; Strawn 1982, Risset and Wessel 1982) as being vague and ambiguous. It restricts the meaning and measurement of timbre to very specific experimental conditions while implying that any difference between two sounds besides pitch and loudness can be labeled as a difference in timbre. Seen with this angle, timbre can be related to a very large number of acoustical and physical phenomena, which makes its study very complex. Letowski (1989) made a useful distinction between timbre and sound quality. In his view, timbre is "the perceived sound spectrum" and timbre parameters should be evaluated on quantitative scales. Sound quality is a more global parameter which contains both timbre and spaciousness and implies a judgement of appreciation. Sound quality should be evaluated on preference scales. In the present thesis, timbre will indeed refer to evaluation and quantification processes and "sound quality" will be used to refer to listeners' preferences and qualitative assessments[2].

In conclusion, the question of standardizing the timbre definition precisely must be left open in the present state of knowledge. Until we can identify more accurately the perceptual dimensions of timbre, it will remain a somewhat intuitive notion with multiple, varying but not necessarily contradictory meanings.

1.2 Physical parameters of timbre

Although a meaningful definition of timbre seems difficult to formulate, certain acoustical parameters of sound generally recognized as having an effect on timbre —i.e., parameters whose variation produces changes in perceived timbre to various degrees—are well identified and can be precisely measured. However, their relative importance for timbre is still a matter of debate.

These parameters combine together to produce a potentially unlimited number of different timbres. As a consequence, one can much more easily identify acoustical parame-

[2]Others have proposed definitions of timbre: see for example Kitamura 1986 and Bregman 1990.

ters related to timbre perception than define its perceptual dimensions, i.e. scales along which timbre sensations can be ordered and evaluated. Unlike pitch and loudness which can both be evaluated along single scales from low to high and from soft to loud respectively, timbre is a multidimensional attribute of sound.

The structure of a sound event

When discussing timbre perception, it has become usual to model in time a single musical note as a sound event consisting of three sections. The attack or starting transient is the portion during which the sound starts and builds up to full amplitude. The next portion is called the steady-state during which the sound is relatively stable. Finally, the decay is the section between the end of the excitation and the return to silence (Meyer 1978, 21-23)[3]. Although much simplified, this model will be useful for the present thesis' purposes The acoustical parameters underlying each section are different and their relative importance to the overall timbre will vary depending on the musical instrument, the musical context, the acoustics of the room, etc.

1.2.1 Frequency spectrum

The distribution of energy along the frequency spectrum is a major determinant of timbre and has been recognized as such by virtually all researchers in the field. References to it are found from the pioneering work of Helmholtz in the nineteenth century to the latest research on timbre in musical contexts.

---

[3]This is a rather crude division and a much finer resolution is possible. For example, Luce (1963) further subdivides the attack transient of the string family into 3 "epoches".

Many researchers have limited their study of timbre perception to the steady-state portion of isolated sounds for reasons of simplicity and ease of manipulation (Slawson 1968; Plomp 1970; von Bismarck 1974, etc.) or because they lacked powerful enough tools to analyze the transient states of sound (Helmholtz 1954). All these studies identified the distribution of energy along the frequency spectrum as being responsible for the timbre of a sound. Helmholtz's classical theory proposed that timbre depends on the presence and relative strength of the harmonics of a sound. This results from the ability of the ear to decompose a complex sound into its spectral components through the frequency-selective action of the hair cells along the basilar membrane.

The resolution of the ear for the perception of the discrete partials of a complex sound was further investigated by Plomp (Plomp 1964; Plomp and Steeneken 1968). It was found that for both harmonic and inharmonic complex sounds, between five to seven partials can be heard separately under optimal experimental conditions. The results of these experiments were found to be in accordance with the concept of critical bands.

According to this concept, the basilar membrane can be viewed as a bank of band-pass filters the bandwidth of which varies with center frequency. The bandwidth corresponds typically to 10% to 15% of the center frequency (Moore 1989). For pure tones, each frequency will produce a vibration pattern with a single maximum whose place along the basilar membrane depends on the frequency. For complex sounds, individual components falling into separate critical bands can be heard separately. Partials within a single critical band will interact and their respective loudness will add to produce a single complex response on the basilar membrane. They will therefore combine into a single percept.

Formants and the source-filter model.

In the voice, the different vowel sounds are characterized by formants, distinctive peaks in their spectra that remain relatively fixed in frequency as the fundamental is changed. As a result, the vowel "a", for example, can be identified as such whether a bass or a soprano is producing it. This phenomenon has been explained by the source-filter model (Slawson 1981). In this model, sound production is viewed as a two-stage process. Vibrations in a resonating body are produced when acoustical or mechanical energy is applied to it. This is the source stage. Due to its physical characteristics, the resonating body will impart on the original source spectrum a characteristic envelope: amplitudes of certain frequency regions will be increased or lowered in a distinctive pattern (due to the physical layout of the body). The body or cavity thus acts as a filter on the source.

In the source-filter model, the source is independent from the filter: they are weakly coupled. Therefore, the spectral envelope theoretically remains unchanged when the frequency or the intensity of the source is varied[4]. The model has been successfully applied to speech and vowel sounds where the source is constituted by the vocal cords and the filter by the vocal tract, the mouth, and the nasal cavities.

Because musical instruments can be identified throughout their range although their spectra can change significantly, invariant factors in the sound were sought and the formant model was proposed (Fletcher 1934, Bartholomew 1942). According to Bartholomew, "the characteristic tone quality of an instrument is due to the relative strengthening of whatever partial lies within a fixed or relatively fixed region of the musical scale" (p. 17).

[4]The spectrum, hence the timbre, of the singing voice does change from the low to the high registers. But the difference in timbre is of a lower magnitude than the difference between two different vowels (Bloothooft and Plomp 1988).

Using vowel-like synthesized tones that were to be judged both on their vowel quality and their musical timbre quality, Slawson (1968) determined that maximum timbral similarity between two sounds an octave apart is achieved when the lower two formants of the higher tone are shifted up about 10 %. He also found strong similarities between invariances in musical timbre and vowel quality in terms of the underlying acoustical factors responsible for that invariance: "At least in musical sounds with fairly pronounced, broad spectral peaks, the complex of auditory attributes that make up what is known as musical color are identical with the auditory attributes of vowel color" (p. 100). He also mentions that "the sounds of most musical instruments have spectrum envelopes that are either fixed in frequency or that shift in proportion to shifts in the fundamental frequency" (p. 100). More restrictively, Plomp demonstrated that "timbre depends upon the frequency position of the spectral envelope and not upon its position relative to the frequency of the fundamental" (Plomp 1976, 108).

However, the source-filter model cannot be applied without reserve to all musical instruments. As Grey pointed out, "many musical instruments do not have physical formant regions fixed in frequency across a wide range of pitches but rather exhibit transposable patterns of harmonic amplitudes relative to the fundamental" (Grey 1975, 5). In the case of vowels, the source (the vocal cords) and the filter (the vocal tract and mouth/nasal cavities) are independent from each other: they are weakly coupled. Most musical instruments, on the other hand, have strongly coupled source (bow/buzzing reed) and filter (string/horn): on many musical instruments, to change the pitch, one must change the filter and not the source. The result is that the spectral envelope, instead of remaining fixed, will be displaced as the pitch is changed. On many instruments, the timbre will be markedly different from registers to registers. However, a number of musical instruments have additional fixed resonances that color the sound in a characteristic manner. Good examples are the resonances from the body and the back plate of the violin, or the resonance in the double-reed instruments originating from the tube connecting the reed to the body of the instrument (Slawson 1988).

Several researchers have analyzed musical instrument tones. Unfortunately, their results do not always agree. Analyzing sounds produced by a C trumpet in an anechoic chamber, Risse and Mathews (1969) found that for a given intensity, the shape of the spectral envelope remains relatively unchanged with a peak (formant) between 1000 and 1500 Hz. Meyer (1978) identified one main formant in the B-flat trumpet around 1200 Hz and three subsidiary higher formants. On the other hand, an analysis of the trumpet by Grey and Moorer didn't reveal any formant peaks in the spectrum (Strawn 1977) Meyer (1978) identified formants for the members of the string family, double-reed instruments (oboe, bassoon), brass (trumpet, trombone, French horn, tuba), and clarinet. The clarinet presents a more complex spectral structure and must be described by two spectral envelopes, one for the odd and one for the even harmonics: odd harmonics are stronger than even ones for lower harmonics while even harmonics have a peak around 3000 Hz with decreasing amplitude on either side (Strawn 1977). Luce (1963) studied 14 non-percussive instruments of the orchestra and found a number of invariant factors both in the steady state and in the attack transients of the acoustic waveforms, that allow a source (here a musical instrument) to be identified under a wide variety of conditions such as: "different pitch (subjective frequency), different loudness (subjective intensity), different style and/or intensity evoked by player, different players, different environments, different samples of the same instrument" (pp. 16-17)[5]. Among the possibly perceptually significant

---

[5]One must not forget that the sample rate (20.833 kHz) for A/D conversion and the 10-bit resolution used by Luce produced sounds that were not always very faithful to the originals. It would even seem that the D/A conversion was done at 27.777 kHz (p. 86). Still, the dissertation provides useful findings on resonances and invariants within the limits of the method's accuracy. (See Strawn 1985).

"regularities of the waveforms which represent invariants with respect to the instrument sounding the tone" (p. 97), Luce includes: phase changes in time, amplitude and waveform modulations, duration of both amplitude and structure transients, inharmonicity both during the attack transient and steady-state portions, invariance (or regular changes) of phase with respect to fundamental frequency that may reveal information on the characteristics of the excitation of the instrument, random characteristics in the waveform, short time fluctuations in the intensities and frequency characteristics, changes from note to note, presence of additive noise which can be a sign of poor execution and/or "an important and perhaps not unpleasant identification clue for the instrument" (p. 99). Luce did not test the aural significance of the regularities of the waveforms that he presented.

That not all musical instruments seem to have a formant structure may be partially explained by the fact that the source-filter model cannot be applied integrally: while the construction of musical instruments do produce characteristic resonances that filter the original vibrations, there is also a strong coupling between the excitation and the body vibrations/resonances in a feedback process that causes the resonating body to control the pitch and the loudness of the sound (Slawson 1981).

In summary, the source-filter model of sound production and the resulting formant structure cannot be generalized to all musical sounds. It has been found that the fixed-pitch theory or the modified fixed-pitch theory (formant frequencies remain fixed when fundamental is changed) assures a minimum change in timbre. Most musical instruments are strongly-coupled systems in which the spectral envelope is not fixed when the fundamental is changed. This is in agreement with Slawson and Plomp since the timbre of musical instruments does vary along their range. The identification of sounds of different timbres to the same instrument has been explained by learning processes (Slawson 1988)[6], —a listener would learn to associate a collection of sounds with different timbres as originating from the same instrument—, by the fact that some instruments exhibit additional, weakly coupled resonances that do not vary with pitch (violin, oboe, bassoon, etc.) and by the characteristic evolution in time of certain acoustical parameters.

Spectrum in musical contexts

Grey (1978) pointed out that in a musical context, a listener can compare the different spectra of a given instrument at different registers and that timbre perception and evaluation may be influenced by that "composite spectral map" of resonance patterns across the registers of the instrument. For his experiments, Grey used computer synthesized versions of three timbres that were presented within single-voice patterns, multi-voice patterns and in isolated form for comparison. All notes were equalized in loudness and presented uniform spectral and temporal shapes (within each timbral category).

From the results of his study, Grey concluded that single-voice musical patterns emphasize spectral differences between different versions of a timbre. Grey suggested that the extended stimuli as in the single-voice patterns help the listener to store and compare the different timbres. In addition, the poor timbre discrimination observed in the multivoice

---

6"In most cases, one cannot isolate a single physical invariant characteristic of the timbre of a musical instrument. Throughout the pitch and loudness range, the physical parameters of the sound of a given instrument vary considerably, to the extent that the perceptual invariance, the unity of timbre of an instrument like the clarinet seems to be a learned concept. However a property, a law of variation, rather than an invariant, often appears to play an important part in the characterization of an instrument (or a class of similar instruments)" (Risset 1977, 29-30). In addition, "with experience, we learn the kinds of transformations in spectral relations within a clarinet tone that can accompany register changes, loudness changes, and so on" (McAdams 1982, 293).

patterns may be caused by masking and by the increased complexity of the spectral energy distribution.

Strawn (1985) studied more closely the region between successive notes in a musical context. He defines such a region as including "the ending part of the decay of one note, the beginning and possibly all of the attack of the next note, and whatever connects the two notes" (p. 2). In general, these transitions may be characterized by a drop in amplitude, a roll-off of the high frequencies producing a low-pass filtered version of the spectrum at the end of the steady state, and the addition of noise components, especially when the second note is detached. Even with slurred notes, noise may be produced by the action of the fingers. No systematic differences seem to exist between ascending and descending intervals, although the articulation between some notes may reveal audible idiosyncratic difficulties. Strawn's experiments indicate that test subjects "consistently preferred the transitions with spectral changes to those without" (p. 179)[7].

1.2.2 Phase

In his definition of timbre, Helmholtz concluded that timbre depend "in no respect on the differences under which [the] partial tones enter into composition" (Helmholtz 1954, 127). Phase was an important issue in early research on timbre because it strongly affects the waveform and it was believed that the waveform of a sound was related to its timbre.

Plomp and Steeneken (1969), studying steady-state complex sounds of equal loudness and pitch, found that phase does affect timbre but that the effect is very small

---

[7]Strawn's study tries to identify the ingredients for a good synthesis of transitions. The importance of transitions for the perception of timbre is not clear. From his definition, the magnitude of the contribution depends on the perceptual importance of the attack and decay portions of sounds. Attack and decay transients in context should be different from those for isolated notes.

compared to the effect of the amplitude pattern of the harmonics. They determined that the effect of phase is greater for tones of lower fundamental and that its greatest possible effect on timbre can be quantitatively compared to the difference in timbre between the vowels [ø], [e], and [œ].

As will be seen later, phase plays a more important role in natural, reverberant listening environments. Woszczyk (1978) points out that "phase interference is largely responsible for the particular timbre of many instruments. It plays an important part in shaping the sounds of instruments in a room or next to a single reflective surface. In practice, phase interference is always contributing to the sound we perceive" (p. 92). As will be seen in chapter 3, the perceptual effect of phase on timbre of reproduced sounds is still a matter of debate.

1.2.3 Temporal factors

Dynamic features of a sound also play an important role in the perception of its timbre. Studies that investigated timbre using a single looped period or isolated steady-state portions of synthesized sounds do not provide a complete, faithful description of sounds as they are produced and perceived in a normal listening context. Helmholtz (1954) recognized that "many of [the] peculiarities of musical tones depend upon the way in which they begin and end" (p. 66). Similarly, other researchers investigating only the steady state (Fletcher 1945; Plomp 1970, 1976, 1987) agreed on the importance of the temporal evolution of sound for timbre.

Attack and Decay

The acoustic features of the attack portion of a sound differ substantially from those found during the steady-state. Because of the sudden character of the attack, all resonance modes of an instrument are brought into vibration. The very high number of partials present results in a strong noise component in the sound. Certain inharmonic components are only present during the attack and provide very important cues for the identification of different instruments or family of instruments (Meyer 1978, 30-31).

The early studies investigating the attack and decay portions of sound used magnetic tape splicing techniques. Clark et al. (1963) conducted a series of experiments on various portions of sounds from orchestral instruments. Their results suggested that the attack transient is more important than the steady state for instrument identification. Berger (1964) studied the effects on timbre of removing the attack/decay portion of a sound, of playing tones backwards, and also of filtering out all harmonics. Results suggested that the spectrum was the most critical factor in timbre recognition (subjects obtained a score of only 18 % in a musical instrument recognition task when all partials were removed) while the attack/decay portion was the second most important (subjects scored 35 % when the attack and decay were spliced out). One should note the extreme character of the sound processing for the filtered version. More subtle variations of the spectrum could have yielded different results. Saldanha and Corso (1964) investigated the relative importance of attack/decay transients, spectrum, and vibrato for timbre identification. They found that best identification was obtained with both attack transients and the steady-state portion of the sounds present. Both Clark (1963) and Saldanha and Corso (1964) concluded that the decay portion of a sound doesn't affect instrument recognition. Obviously, the decay portion of a sound is very important to determine the timbre of sounds which do not have any steady state such as percussion instruments, the piano, plucked strings.[8]

Data Reduction

Results from a computer analysis of the physical parameters of timbre can be tested by resynthesizing the tone using the data obtained from the analysis. If the original sound and the resynthesized one cannot be told apart, this is a good indication that all essential information for the timbral description of the sound was detected in the analysis.

This type of analysis produces a huge amount of data describing the amplitude and frequency functions for each harmonic as they evolve over time (Grey 1975, 17). Several attempts have been made to simplify the acoustical representation of timbre without altering the perceived quality and hence to determine the most perceptually important parameters of timbre (Risset 1966; Grey 1975; Charbonneau 1981; Chowning et al 1982).

Such a data reduction performed by Grey (1975) revealed that:

1º removal of low-amplitude, inharmonic components in the attack section of a note is most discriminable;

2º the replacement of the time-variant frequency functions by constant frequencies approximations is very easily detectable;

3º line-segment approximations of amplitude and frequency functions, whereby minute fluctuations were discarded, were very hard to detect, suggesting that they are not essential to the perception of timbre.

[8]While percussion instruments do not have a steady state, modifications of the spectrum will produce strong perceptual changes of the timbre. Spectral manipulations affect both the attack portion and the decay which carries most of the timbre.

The perceptual significance of the various types of data reduction mentioned above varies strongly between instruments. For example, the attack simplification was highly discriminable for the bass clarinet but not for the French horn; the soprano saxophone timbre was greatly affected by the constant frequencies approximations while the change in timbre of the English horn was very small.

Similar results were reported by Chowning et al. (1982). Patterns of onsets-offsets and the spectral envelope of the harmonics seem to be the most important features of the time-variant amplitude functions for timbre. In addition, it appears that the highly complex timing pattern of onsets/offsets can be simplified (producing differences in the order of 5 ms.) without being detectable. On the other hand, even slight changes in the amplitude levels of the harmonics are obvious[9].

Computer analysis of musical instrument tones provides us with a four-dimensional physical model for timbre consisting of amplitude, frequency, time, and phase. Since phase does not affect timbre significantly (Plomp 1969), we can reduce the model to three dimensions. Data reduction can further simplify the model by discarding physical features that are aurally insignificant. However, as Risset and Mathews point out, a more detailed physical description of sound may be necessary in order to confuse all listeners in A-B comparison tests[10].

---

[9]"The most important aspects of the temporal amplitude functions of the harmonics are their patterns of onset-offset and their spectral envelope, and not necessarily the fine details and differences of their individual shapes." The authors also achieved "a simplified representation of the timing pattern of onsets and offsets for the harmonics of a tone. ... The differences that were produced were generally on the order of 5 ms; hence the indiscriminability of the operation points out a temporal integration in hearing... The perceptual cues important in the temporal structure of an instrumental timbre in the attack and decay, then, may often be represented in a more simplified manner than actually occurs acoustically." (pp. 2-3).

[10]<u>See</u> the section on levels of timbre perception sophistication in chapter 4.

As with spectrum, the context in which sounds are heard can modify the perceptual significance of temporal factors. Grey (1978) found that isolated tones help the discrimination of temporal-related features of a sound. He further suggested that masking could cause greater difficulties in hearing fine temporal details within musical patterns. On the other hand, the isolation of single notes allow the listener to concentrate on the fine details of attack and decay of the sound. Campbell and Heller (1978) defined the 'legato transient' as "the transition between two notes in a legato passage played on a continuous tone instrument. It is initiated when the performer interrupts an existing standing wave and ends when a new standing wave has been established" (p. 1). The authors investigated the relative importance of attack and legato transients and steady-state portions of sounds for identification of different instruments. They concluded that both types of transients were more important than the steady-state alone and that possibly the legato transient would be a better cue than the attack transient with some reserve due to the possible contribution of steady-state elements during the legato transient[11].

Other time-variant features of a sound contribute to its timbre. Computer analysis of musical instrument tones (Risset and Mathews 1969; Grey 1975; Strawn 1977; see also Luce 1963) have shown that each partial of a complex sound has its own temporal envelope. Amplitude and frequency modulation associated with these envelopes are characteristic of the different instruments. For example, low harmonics in the trumpet have a shorter attack time and a longer decay than higher harmonics (Risset and Mathews 1969).

[11]See also Strawn 1985.

1.2.4 Additional factors

Musical instruments are imperfect mechanical systems that produce various noises and irregularities in their sound. Puffs of air in the attack of a flute, the sound produced by the impact of the hammers on the strings of a piano, the blips in intonation in brass instruments, bowing noise in the violin, all are highly idiosyncratic features of various sound sources[12]. The importance of these imperfections for the overall tone quality of musical instruments was recognized early in the pioneering work of Helmholtz (1954) who stated that "such accompanying noises and little inequalities in the motion of air, furnish much that is characteristic in the tones of musical instruments" (p. 68). In a musical context, inequalities intentionally or unintentionally introduced by the player will come up and influence the overall perceived tone quality. Playing styles, phrasing, and various articulation techniques all become part of the overall sound produced (Risset and Mathews 1969). And as Risset (1977) points out: "Very often idiosyncrasies of sound production result into tone particularities which are strong cues for instrument recognition: frequency glides between notes in the trombone, because of the slide; intonation troubles in the horn, because of the difficulty to hit the right mode; initial erratic vibration in string instruments, when the string is first set in motion by the bow; burst of tonguing noise at the beginning of recorder sounds" (p. 30). Balzano (1986) even suggested that "the kinds of things we are capable of hearing that are important for timbre perception are events like pounding, blowing, bowing, plucking, rolling, whistling, screaming, and all sorts of physical processes that words can only hint at but which are nonetheless specified in the underlying dynamics of the signal, and therefore just as potentially 'available' to the perceiver as a Fourier spectrum" (p. 13).

---

[12]The noise background results from irregularities in the excitation of the source that put into vibration all the natural modes of a particular instrument, hence its distinctive character and its importance for the sound quality and naturalness (Meyer 1978).

1.3 Spectral fusion, perceptual grouping, scene analysis

The complexity of timbre perception phenomena reaches a maximum in musical contexts such as a symphony or in pop/rock music, when different instruments playing different notes or chords (or the same ones) are mixed together. In that case, timbre is not determined only by the harmonics but rather on how they group together into distinct sound objects.

When we hear a dense, complex piece of music, some of the sound components reaching our ears will group sequentially while others will group simultaneously. In such complex sound settings, coherent sound images are the result of two complementary processes, namely fusion and parsing. Perceptual fusion of components will occur according to some rules: for example, sounds that start and stop at the same time, sounds that undergo synchronized frequency or amplitude modulation will tend to group together. Two pure tones that group together will result in a perceived richer tone. Therefore, timbre can be viewed as the result of the fusion or spectral grouping of several components (Bregman and Pinker 1978). McAdams (1982) identifies three cues influencing perceptual fusion of complex tones: harmonicity of the frequency content of a tone, coordinated modulation of the spectral components, and the relative familiarity of the spectral envelope (p.281).

Perceptual fusion increases as the degree of harmonicity increases. It would seem that the auditory system is biased toward fusing harmonic elements into single sources. It should be noted however, that many musical instrument are characterized by the inharmonicity of their partials, notably percussion instruments in general, but also the piano (Meyer 1978).

It seems that sounds with familiar spectral envelopes fuse more readily than others. Fusion then depends on "the shape and location in the frequency spectrum of resonance peaks characteristic of sounds we frequently encounter" (McAdams 1982, 285). This implies that the auditory system would be able to "evaluate the shape of the spectral envelope of a source independently of its evaluation, at any given moment, of the presence of the frequency components whose amplitudes are shaped by that spectral envelope" (p. 285).

A monophonic recording of a symphony orchestra will present the listener with a single wave front containing all discrete sound sources that were originally recorded. The listener will nevertheless be able to parse with great accuracy the various original sources into separate streams and will hear individual instruments and individual timbres.

Complex listening situations involve the concept of pattern recognition. As Moore (1982) explains, "Firstly, complex stimuli will contain more than one frequency component, so that the patterning of spectral energy as a function of frequency will be important. Secondly, auditory stimuli typically vary with time, and the temporal patterning can be of crucial importance to perception" (p. 186).

A multiplicity of auditory cues can be used in a typical listening situation. As pointed out by Grey (1978), Risset and Wessel (1982), and Strawn (1982), the nature of the context of presentation will strongly influence the listener's choice of individual parameters as being important cues in the elaboration of a timbre percept. As Bregman (Bregman 1984) further points out,

> Human perceptual decisions appear to make use of many sources of information, some of which are undoubtedly redundant. Their effects are made to compete and collaborate in shaping the final percept. Perhaps this is the source of the great robustness of human perception in the face of all sorts of distortions (p. 173).

1.4 Interactions between timbre and other attributes of sound

In studies on timbre, variations of the parameters of sound other than timbre are usually removed from the stimuli so that the factors determining timbre can be isolated and the experimental variables better controlled. However, in a musical context, attributes of sound interact with each other to produce the final percept.

Intensity

Clark and Milner (1964) found that the spectrum of a non-percussive instrument varies with intensity: as the intensity increases, the proportion of high-frequency components increases as well. Their results, however, indicated a very low correlation between playing intensity and perceived timbre. The method used casts some doubts on the validity of the results. Tones from musical instruments were recorded in an anechoic room with an omnidirectional microphone at different dynamic markings (*pp*, *mf*, *ff*) and then played back at an equalized level.This procedure may have introduced errors as the frequency-dependent sensitivity of the ear can affect the timbre perceived, independent of the original markings. It may be significant that for several instruments, recordings made at *pp* were judged as *ff* when played back at *ff*. Clark and Milner' s findings contradicts more recent research (Risset 1966, Meyer 1978, Hall 1980). Risset (1966) described this factor for the trumpet as being characteristic of the spectrum of the brass family. Meyer (1978) takes as an example three spectra of a high note played on a horn at three different levels. At *ff*, harmonics around 10 kHz have amplitudes of about 45 dB while in a *mf*, the sixth harmonic around 2.2 kHz will barely stand above the noise floor (pp. 33-34). The effect of intensity on timbre may of course vary with different instruments. The poor sensitivity of the ear at low and high frequencies is another factor contributing to the influence of intensity on timbre. Even though they may be present, low and high frequencies will be less (or not at all) perceptible at low levels. Loudness, which is largely determined by intensity, was found to correlate well with sound quality of reproduced sounds (Illényi and Korpássy

1981). As will be discussed in the next section, distortion products may increase as the listening level is increased.

Frequency and physical duration also interact with timbre but the nature of their influence is not well understood yet (Grey 1975, 21-22). The perceived quality of a sine wave changes from low frequencies to high frequencies even though, obviously, the spectrum stays the same (Risset 1977, 15)[13]. Wapnick and Freeman (1980) presented evidence of the interactions between timbre and perception of musical flatness and sharpness. The effect of duration on timbre would seem to depend on the spectral integration time of the ear (Wang and Gossick 1978).

Reverberation

Little attention has been paid to the effects of reverberation on the perceived timbre of sounds. At any point in a room, the perceived sound will be a mixture of direct and reflected sounds. The closer to the source is a listener, the greater the proportion of direct to reverberant sound will be heard. Reverberation modifies significantly the phase and frequency relationships of the sound spectrum. Differences in frequency response up to 20 dB have been reported (Risset and Wessel 1982, 29). In addition, these variations in the physical structure of the sound will be different at every location in the room.

Plomp and Steeneken (1973) studied the effect of reverberation on the timbre of steady-state complex tones. They found that for tones above 100 Hz, the perceptual effect of phase randomness is negligible. Reverberation impairs identification of sounds whose spectral differences are less than the variation introduced by the reverberation. The magnitude of the spectral differences caused by reverberation are close to the differences between

[13]The spectrum does not change but the sensitivity of the ear does.

the same vowel sound pronounced by different speakers (Plomp 1976, 101)[14]. This should obviously depend on the amount of reverberation added to the sounds. More recently, Toole and Olive (1988) studied the effect of reverberation on the audibility threshold of spectral resonances. It appears that the addition of reverberation lowers the threshold of nondelayed medium- and low-Q resonances in impulsive or transient sounds while low-Q resonances delayed by more than 1 ms are more easily detected without reverberation with differences of up to 10-14 dB (pp.130 and 138).

The importance of room reflections and reverberation on musical instruments' timbre depend on the source. For percussive instruments such as timpani in which the decay is most important, reverberation becomes the main carrier of the timbral information of the instrument. But "for percussion instruments with low damping (like bells, gongs, low piano tones) the instrument's decay usually dominates the room's reverberation. In such cases, the details of the decay have a strong bearing on the timbre" (Risset 1977, 31).

It has been recently pointed out that since musical instruments exhibit frequency-dependent radiation patterns[15], a complete spectral description of an instrument's timbre is only possible if information from the important directions of radiation is made available to the listener (Woszczyk 1978, Benade 1985)[16]. In a reverberant room, such information is provided by early reflections (Benade 1985). As Woszczyk points out, "the environment plays an important part as the integrator and carrier of (an) instrument's timbral information" (p. 78). Woszczyk conducted listening tests which indicated the listeners' preference

---

[14] See also Kates 1984.

[15] "... les instruments de musique ont un rayonnement anisotrope: lorsqu'on tourne autour de l'instrument, le spectre entendu varie considérablement, mais cela n'affecte guère la reconnaissance de l'instrument" (Risset 1978. 6).

[16] See also Clark et al. 1963, 46; Luce 1963; and Risset 1978.

for recordings using multiple microphones positioned to integrate an overall spectrum of the sources over single-microphone recordings providing only a truncated spectral view of the source. Benade (1985) obtained high identification scores on steady-state one-second samples of an oboe when both source and microphones used for the recording were moving so as to sample the instrument's sound at different locations. Unfortunately, he didn't include any supporting data. Benade further suggests that results from earlier studies indicating the predominance of attack transients as identification cues could be explained by the fact that the recording technique used (stationary source and microphone) provided only "a limited single-view auditory 'picture' of the instrument, from which recognition turns out to be very difficult until some additional piece of information is made available (such as onset behavior)" (p. 229). This could explain why identification scores for unmodified sounds as reported by Berger (1964) and Saldanha and Corso (1964) were so low.

1.5 Dimensions of timbre

Physical parameters of sound that affect timbre can be precisely measured by calibrated instruments. The measure of timbre perception is more complex. Because timbre cannot be evaluated on a single scale from low to high like pitch, it is said to be multidimensional. At present, we know more about the physical factors affecting timbre than about the corresponding perceptual dimensions. In other words, we still do not have a satisfactory psychophysical model that could be used to quantify differences in timbre. Because timbre is a multidimensional attribute, its study is much more complex than other attributes like loudness or pitch.

Several attempts have been made to determine and describe a manageable number of dimensions along which timbre can be "measured". Plomp (1970, 1976, 1987) suggested a model using the spectrum level in 1/3-octave bands for steady-state periodic complex sounds. He obtained a three-dimensional model that correlated well with the intensity levels in 18 1/3-octave frequency bands. It would seem however that these 1/3-octave band levels are not independent factors due to masking patterns (Grey 1975, 13) Using 35 steady-state timbres evaluated along 30 pre-defined bipolar semantic scales, Bismarck (1974a, 1974b) performed semantic differential scaling and factor analysis to obtain four dimensions, among which the scale dull-sharp seemed to be the most important. Bismarck identified sharpness as the main perceptual attribute of timbre. According to Bismarck, sharpness is defined by the relative proportion or balance between low and high frequencies. It increases as the amount of energy in the high frequency region increases compared to the lower one. Furthermore, the sharpness of a sound is independent from peaks or dips in the spectrum[17]. Although the reliability of Bismarck's results has been criticized (Grey 1975; Strawn 1982), the attribute of sharpness (dull-sharp, dull-bright) has emerged from other research as well (Lichte 1941; Grey 1975; Grey 1977; Wessel 1979).

A better approach to study timbre dimensions may be the use of multidimensional scaling where stimuli are presented in pairs to subjects who have to judge their similarity. The data is then fed to a computer program which produces a simplified geometrical map of n dimensions in which the distance between any pair is inversely proportional to the similarity between the stimuli. The investigator then links these psychological evaluations to possible physical correlates.

Grey (1975, 1977) used this technique to investigate the possible perceptual dimensions of timbre. He used 16 instrumental tones that were analyzed, simplified and resynthesized by a computer. All stimuli were equalized in pitch, loudness and duration in order

[17]There is some ambiguity in the litterature about the conflicting use of the words sharpness and brightness and the difference in their meanings is not always clear.

to focus on timbre dimensions. A three-dimensional spatial configuration was found to be the best solution to interpret the psychophysical relationships between tones. One of the dimensions was related to energy distribution along the frequency spectrum. At one end of the scale (dimension), the bandwidth is narrow with a low ratio of high/low frequencies. At the opposite end, the bandwidth is larger with a corresponding higher ratio of high/low frequencies. The physical description of this dimension closely resembles the sharpness attribute described by Bismarck (1974b) and the brightness reported by Wessel (1979).

The two other dimensions described by Grey referred to temporal characteristics of the tones. One seemed to be related to the degree of synchronism between upper harmonics while the other one involved the presence (or absence) of high-frequency inharmonic energy being present before the full attack. Another interpretation of the last two dimensions indicated a clustering of stimuli according to instrumental family. Subjective correlates for the last dimension were suggested (Grey 1978): the more there are inharmonic components before the attack, the more noise-like is the sound. In addition, the perceived "explosiveness" or "hardness" of the sound decreases as the amount of "precedent high-frequency energy in the attack" increases (Grey and Gordon 1978, 1499).

Using multidimensional analysis techniques, the investigator does not need to make any assumptions about the possible dimensions prior to the experiment. However, the identification of the obtained dimensions will depend on the experimenter' s interpretation of the results.

In the studies reported above, the experimenters looked into the parameters of the sound produced for factors determining perceptual dimensions. A number of researchers considered instead the factors pertaining to the methods of sound production, such as hardness of mallet for percussion instruments (Freed and Martens 1986). More generally, Balzano (1986) states: "Perceiving timbre is more a matter of perceiving underlying dynamics of physical processes [involved in the sound production] than perceiving the places of things in abstract ordered structures" (p. 16).

CHAPTER 2

TIMBRE PERCEPTION IN SOUND ENGINEERING

2.0 Introduction

Sound engineers must evaluate timbre of live sources (in the studio or the concert hall), and of reproduced sound (in the control room) through loudspeakers and headphones. The study of timbre of reproduced sound must take into account the additional physical factors introduced in the recording/reproducing chain which shape, favorably or unfavorably, the resulting sound quality of the program material. The sound engineer can control these parameters to obtain the desired timbre.

2.1 The recording/reproducing chain

The recording/reproducing chain or more simply the audio chain is the usual path that audio information will follow from the source (e.g.., a musical instrument) to the listener. Between the two end points of the transmission system are several stages which can each modify the timbre of the sound arriving at the listener's ears. Therefore, the sound engineer will evaluate the original sound source indirectly and the knowledge of these signal transformations is essential for effective control of sound quality.

Typically, a musical instrument produces a sound which will be more or less affected by the acoustical environment depending on the type and placement of the microphone(s) and the size of the room. The sound waves from the instrument will be converted to an electrical signal by the microphone, and then will go through amplification and various signal transformations (spectral filtering, level balance) in a mixing console, and then amplified again and fed to loudspeakers. The signal is then converted back to sound waves, transmitted through the listening room, and finally reaches the listener's ears. In an alternate transmission path, the signal goes from the mixing console to optional outboard signal processors to a storage device and then played back through the loudspeakers Another stage in the chain is the final mixing when multiple channels are mixed together, generally to two channels. Here, the delicate balance between all sources will shape the overall sound of the recording, allowing certain instruments or certain timbres to come out, and masking or modifying others at other times.

Each element in the chain can modify the original spectrum of the source by being frequency discriminating (microphone), by adding resonances of various kinds (equalization, signal processing, loudspeakers, listening room), by introducing distortion elements (basically every link). The significance of these spectral and temporal perturbations depend on the extent to which they are audible. Therefore, the perceptual thresholds of these phenomena must be taken into account.

Some of the links in the chain are variable and can be controlled by the engineer: microphone choice and placement, signal processing. Others are modifiable to a limited extent: acoustics of the recording room (through the use of acoustic baffles, for example). Others are basically fixed: monitoring system and listening environment (once a control room has been designed and is built, it is usually not modifyable to a large extent).

The elements of the audio chain and their respective "timbre modifiers" can be schematically listed as follows:

Source	musical instrument, performer, score.
Recording Room Acoustics:	reflections, phase cancellations/reinforcement, reverberation, time delays, resonances.
Microphone:	directional properties, resonances, frequency response, proximity effect, off-axis colorations.
Mixing console:	level balance, equalization, compression/limiting, other signal processing.
Tape recorder:	storage limitations (saturation), distortion, hiss (analog), dither (digital), wow/flutter (analog), drop outs.
Loudspeakers:	resonances, distortions (linear and nonlinear), clipping (amplifiers), off-axis irregular spectral response.
Listening Room acoustics:	room modes, reverberation, reflections, spectral aberrations.
Listener:	perception acuity, attention, biases, fatigue, hearing impairment[17].

The elements of the audio chain are intimately related to each other. For example, a particular microphone will strongly determine the nature of the signal transmitted to the loudspeaker which in turn will heavily bear upon the sound quality perceived by the engineer in the control room (Somerville 1952; Toole 1973; Børja 1977; Olive 1990).

As Somerville points out, the way a recording is mixed or a microphone is placed depends largely on the monitoring system on which the balance or microphone is evaluated.

[17]Olive (1990) provides an extensive list of all interactions of the elements of the audio chain that can modify the original spectrum of the source except for the listener. The reader may refer to this paper for more details.

When the frequency response of the loudspeakers or the room reveals serious irregularities (peaks/dips), equalization will be applied by the engineer to compensate for the bad speaker/room response rather than for artistic/creative enhancements of the recording (Børja 1977). The three last elements of the chain interact strongly with each other: "The loudspeakers, room and listener comprise a system within which the sounds and spatial illusions of stereo are decoded, and they must be considered together" (Toole 1990, 1).

Toole provides a list of factors in the loudspeakers/room/listener combination that can affect the perceived timbre: the acoustical coupling of loudspeakers and the room which depends on room modes[18], listeners seating position in respect to the loudspeakers, acoustical interferences (comb filtering) resulting from the combination at the listener's ears of the direct sound and one or more strong early reflections, the frequency-selective sound absorptive qualities of room boundaries which modify the integrated spectrum of the sound field, off-axis colorations of loudspeakers, various effects of reverberation (amplification/attenuation) on delayed and non-delayed resonances. Fortunately, room resonances often have medium to high Q's to which the hearing system is less sensitive (Toole and Olive 1988, Toole 1990a).

2.2 Timbre of reproduced sound

The timbre of sounds reproduced over loudspeakers and headphones is modified by variables specific to such transmission systems. Generally, a non-flat response of a system will result perceptually in a change of timbre:

---

[18]"The proportions of a room, length to width to height, determine the distribution of room modes in the frequency domain, whether there are clusters or gaps in the distribution. The dimensions themselves determine the frequencies at which the resonances occur" (Toole 1990, 4).

> ... response refers to the transmission characteristics (of a linear system) which determines the spectral composition of the output when the input is a signal of complex frequency composition. In musical terms, a non-flat response is recognized by a change in timbre, the amount of change depending on the source spectrum and the response of the sound system (Salmon 1950, 14).

Most of the research in the field of timbre of reproduced sound has been aimed at the evaluation of the elements (loudspeakers mainly) of the reproducing system. The focus of the engineer remains the sound itself, not the devices through which it is reproduced. One must therefore be cautious about the conclusions of the reported research as they may not be totally relevant or readily applicable to the listening strategies of the sound engineer.

2.2.1 Timbre modifications

Distortion

Whenever the output of a circuit is not the exact replica of the input signal, we can say that the original sound is distorted. When recording a sound source in a room, many variables can introduce error factors that will modify the original sound. Whether they are beneficial or detrimental to the desired sound quality, a knowledge of their nature and perceptual significance is important.

Linear distortion

Linear distortion modifies the spectral balance of a sound by either changing the amplitudes of some harmonics or by delaying some harmonics, or a combination of the two. The delay will produce cancellations and reinforcements at particular frequencies, hence a change in timbre. As Preis (1984) states, "in electrical or mechanical systems, linear distortion occurs because portions of the signal energy are selectively absorbed, or reinforced, or stored and released at a later time, or reflected, or propagated through materials at different relative speeds" (p. 4).

The perceptual significance of phase distortion is still a matter of debate. Findings from Plomp and Steeneken (1969) reported in chapter 2 were obtained from isolated artificial signals and are not representative of music recordings. The listening context in the control room introduces several variables that can affect (mask, enhance, or modify in some other way) the perceptual importance of phase relationships between components of a complex sound.

There is some evidence that phase shifts can produce audible effects on selected signals in specific listening conditions. Group delays, defined as "the difference between the time delay occurring towards the extremes of the audio frequency band and the delay that occurs at some suitable reference frequency in the middle of the band" (Moir 1976, 84), are often used as a measure of phase sensitivity. Sensitivity to group delay depends on frequency. The ear is most sensitive to delays between about 500 Hz and 5000 Hz where group delays of ± 0.5 msec can be detected, using broadband clicks or impulses on headphones. Below 500 Hz, the threshold is around ± 2.5 msec. Other experiments on group delay distortion produced by anti-alias filters indicated that delays up to 4 msec in a 3000 Hz bandwidth about 15 kHz are not audible (Preis 1984). Blauert and Laws (1978) point out that perceptual thresholds for group delays depend "on the shape of the group delay curve, the nature of the test signal, the conditions of stimulus presentation and the state of training of the subjects" (p. 1479). It was found out that the highest sensitivity occurred for short impulses under anechoic listening conditions. Using 25-µsec rectangular pulses with headphones, the authors obtained thresholds of 0.1 to 0.5 msec. Since group delays introduced by non-minimum phase loudspeakers and headphones are on the order of 400 μsec, the authors concluded that in practical situations (music programmes heard over loudspeakers in a room), group delays normally encountered are not perceptually significant.

Deer, Bloom, and Preis (1985) investigated the perceptual significance of group delay distortions introduced by second-order all-pass filters which allow the independent adjustment of peak value (in millisecond) and of bandwidth of group delay distortion (p. 2). A two-alternative forced-choice test procedure was used, with six subjects listening over headphones. The required response from the subjects was to determine if the two members of the pair were the "same" or "different". A perceptual threshold of around 2 ms at 2 kHz was obtained. Peak group delay ranged from 1 to 16 ms and broad, medium, and narrow delay bandwidths were used.

Moir (1976) reported an experiment in which simple square-wave type signals were additively constructed. The listeners themselves could modify the phase of the third and fifth harmonics in respect to the fundamental while monitoring visually on an oscilloscope the resulting changes in waveforms. Changes in sound quality could not be heard. The author reports another experiment in which the steepness of the wave front was seriously altered by shifting higher harmonics by 180 º. No audible difference could be heard. Moir reports the results of some experiments on the effects of group delay on speech. He notes that the effects of differential time delays are more apparent on speech than on music and that the audible effect sounds like a "metallic echo following each syllable or word" (p. 84). Thresholds of five to eight milliseconds for the frequency band between 5 and 8 kHz that was delayed behind the 1 to 3 kHz band were observed. At the low frequency end, delays up to 70-90 msec were inaudible. In these tests performed at Bell Laboratories, subjects had to increase the amount of delay until a change in quality was perceived in a process of direct comparisons.

It seems that the effect of phase shifts is very subtle if not negligible (Toole 1986). Large phase shifts cannot be reliably detected even by experienced listeners when listening to loudspeakers in a normal room. In such complex listening contexts, it is difficult to determine if an audible effect is due to a shift in phase or to a related spectral amplitude change. Toole points out that the situation could be different for rapid local phase changes associated with resonances but then, one may wonder if it is the phase effect which is audible or the associated resonance.

Suzuki et al. (1980) tested the audibility of phase distortion on transient artificial signals presented both over headphones and loudspeakers. Subjects were presented with four different pairs of sounds: A-B, B-A, A-A, B-B. Because the difference between the two sounds was very small, each sound of the pair was repeated five consecutive times. Again, the signals were short transients and the subjects needed the sound to be repeated in order to remember the sound quality. The task consisted in determining if the two signals sounded the same or if they were different.

Results indicated that although subjects were all sound engineers, the ability to perceive phase distortion was highly variable between individuals. Unfortunately, the authors calculated percentages of correct response. No threshold values were determined. The frequencies studied were at 300 Hz and 1 kHz. The authors concluded that although some people can readily detect phase distortion at low frequencies "when highly artificial signals are used" (p. 573), most people are not very sensitive. None of the subjects were able to detect a change in sound quality when popular music from commercial records was used as stimulus.

Lipshitz et al. (1982) reviewed a number of papers that tended to demonstrate that phase distortion is audible. For example, they duplicated experiments by Mathes and Miller (1947) and Craig and Jeffress (1962) using a two-component tone consisting of a fundamental and a second harmonic. They concluded that the timbre of the tone is a function of the relative phase between the two components. They went further and adjusted the phase difference to produce an asymmetrical waveform from positive to negative. They claim that a polarity reversal (simultaneously at the two channel) can be readily detected as a change in timbre. They also claim that this effect can be heard, although with much more difficulty, with speech and music signals. Lipshitz and Vanderkooy (1981) qualified the change on percussive signals as a change in "perceived depth, high-frequency detail, clarity and even level" (p. 489). The ability to hear polarity reversal on asymmetrical signals is related to the half-wave rectifier behavior of the inner ear. "Neural output from the acoustic waveform occurs predominantly during the rarefaction half of the acoustic waveform, for frequencies below approximateley 1 kHz. Therefore the ear is able to detect waveform asymmetries and hence polarity reversal of asymmetrical signals" (p. 383). Although many musical and speech signals are asymmetrical, the effect of polarity reversal is very subtle.

The authors dispute the argument that phase distortion cannot be audible because in a reverberant environment, severe phase aberrations are observed for very small displacement in space. They claim that phase nonlinearities can be detected for the direct portion of the sound, before the reflections disturb the phase further.

Although "to a greater or lesser extent all audio components contribute some phase distortion" (p. 585), loudspeakers appear to be the main source for midrange phase distortion (between 100 Hz and 3 kHz)[19].

One of the authors could hear phase distortion over stereo records. It is well known that phase effects on stereo programs are quite obvious (Moir 1976). Stereo material is

[19] See Lipshitz and Vanderkooy 1981 p. 585 for contributions of other parts of the audio chain to phase distortion.

probablv not a good choice for test stimuli. Results from one subject are not very conclusive even when 99% confidence results are reported.

We know that phase shifts can be heard in some cases when there are two sound sources. If the sources are separated in space as with a pair of loudspeakers, spectral cancellations/reinforcements can occur or the stereo image can be perturbed. When the two sources occupy the same horizontal position, there are claims that time delays that reduce the steepness of transient peaks can be heard (Moir 1976, Moore 1989). But beyond the obvious fact, the question of phase distortion audibility must be left open for now.

Non-linear distortion

Non-linear distortion occurs when additional components not present in the original input of a circuit or device are observed at the output. If the input is a sine wave, the output will produce new components which will be integer multiples of the fundamental. This is commonly called harmonic distortion. If the input is a complex wave, the output will consist (in addition to the harmonics of the input frequency) of the sums and differences of the original input frequencies, and all combinations of these input frequencies. This is called intermodulation distortion.

Cabot (1984) identifies five main factors that affect the perceptual significance of non-linear distortion products:

1. characteristics of the non-linearity
2. type and complexity of the distorted audio signal
3. characteristics of the listening environment
4. other distortion in the other links of the audio chain
5. the "ability, experience, concentration, etc." of the listener.

The nature of the nonlinearity has a strong effect on the spectral content of the distortion components and their dependence on signal characteristics. For example, the amplitude of harmonic and intermodulation components produced by crossover distortion increases as the signal amplitude decreases. Clipping distortion products only appear when input level increases beyond a certain threshold limit and the amplitude of distortion components increases as the input level increases.

The minimum audible distortion threshold is limited by the hearing threshold at any frequency, by signal-to-noise ratio in rooms, and by the masking properties of the ear. The upward spread of masking (Moore 1982, 76) implies that components lower in frequency than the main tone should be more noticeable than components higher than the main tone. Harmonic distortion can therefore be viewed as a special case of masked threshold perception (Cabot 1984, 6).

Bryan and Parbrook (1960) obtained perceptual thresholds of less than 0.05% for harmonics above the fourth and a listening level of 70 dB. The frequency of the fundamental was 357 Hz. Whyte (1977) suggested that the ear's sensitivity to interfering tones vary with frequency, according to the equal loudness contour curves.

The characteristics of the listening room have important effects on the audibility of distortion products. Ambient noise may mask the distortion components. In a control room, noisy equipment that tends to be masked by the output from the loudspeakers could possibly mask in turn distortion components present in the music when the output from the loudspeakers is low. Reverberation can also mask low level distortion whereas standing waves can lower the threshold if the fundamental of the signal corresponds to the center frequency of a dip in the room response. Whyte (1977) observed variations in threshold in the order of 6 dB above 100 Hz between two listening positions.

Higher thresholds have been observed for music signals compared to pure tones (Gabrielsson et al. 1972). In a first experiment, two types of stimuli were used: pure tones at 500, 1200, and 2000 Hz at a level of 60 dB SPL, and clarinet tones at 150, 470, and 1200 Hz at a level of 80 dB SPL. Both signals were gated with an onset time of 70 msec. 70 subjects were tested: 6 sound engineers, 4 musicians and 60 "normal" subjects. The duration of the stimulus was 4 sec and x^2 or x^3 distortion was introduced during either the first or the second half. Five levels of distortion were tested. Subjects were instructed to determine if during each presentation, the tone was changed in any way. Both loudspeakers and headphones were used. Among the results, it appeared that threshold values for pure tone stimuli were systematically lower than for the clarinet tones. Threshold values for headphones were lower or equal to the ones obtained with loudspeakers. In general, engineers and musicians obtained lower thresholds.

In a second experiment, 101 subjects were presented with pairs of tones and were asked for each one if they could hear a difference between the two tones. Clarinet and flute tones were used, with a duration of 650 msec. Listening was done with headphones. These experiments revealed that detection of x^2- and x^3-distortion depends on complex interactions between physical variables such as distortion type, spectrum, transient type, and frequency. For example, it was found that threshold values for x^2-distortion were higher than x^3- distortion for the flute whereas they were lower or about the same with the clarinet.

Because most distortion types vary with signal amplitude, it is rather difficult to determine a perceptual threshold for music programmes, which vary in level continuously (less so for pop music) (Moir 1981). To alleviate the problem, some authors provided statistical data. For example, Fryer (1975) determined that clipping distortion was audible if it occurred for 1.5 % of the time in a 5-minute presentation (in Cabot 1984, 8).

Wigan (1961) found that the ear can accept a certain level of distortion if it is conditionned to a higher level of distortion. It seems that as the reproduction bandwidth is reduced, the threshold of distortion increases. For example, Olson (1957) obtained a 0.7 % threshold for a 14 kHz bandwidth and a 1.4 % threshold for a 4 kHz bandwidth. This would indicate that distortion components of higher frequencies are more objectionable than lower ones.

Distortion products introduced by the ear itself may impose some limitations on perceptual thresholds when listening is done at high levels. According to Killion (1976), harmonic products exceed 1 % for sound levels above 80 dB and intermodulation products reach the same percentage above 90 dB. (Moore 1982, 57 and 132; Killion 1976).

Fielder (1982) reported that binaural cues can lower the threshold of low level signals. Tones 10 dB lower than the ambient (average) level could thus be detected. Cabot (1984) hypothesizes that this same phenomenon could play a role in the perception of distortion.

He points out that the method used to calculate thresholds has a strong impact on the meaning of the data obtained. For example, threshold values averaged over a large number of listeners with varying listening abilities will not be good indicators of what can be perceived by trained critical listeners. For the latter, one would need to average several results from the <u>same</u> subject. As the author words it, "to say that 1 % distortion is acceptable because the average person cannot perceive the difference is not to say that 0.1 % distortion is inaudible" (p. 12).

Because various amounts of nonlinear distortion are often present in real sounds, perception of non-linear distortion can often be equated to the perception of an increment in distortion from an initial value (Letowski 1975). The difference limen for nonlinear distortion (DLND) is defined as "the smallest perceptible increment of the product of the nonlinear distortion" (p. 106). This "just noticeable increment" depends on the initial value of distortion, the nature of the distortion, the fundamental frequency of the signal , the sensation level of the signal, and its spectrum.

Pure tones and electronic organ sounds were used. Sounds were presented in pairs, and the subjects (6 professional sound engineers) had to judge if the timbre of the two sounds was the same of different. The difference limen was calculated as the value producing 75 % "different" judgments. It appears that all physical variables tested interact to affect DLND's; the initial distortion level is the most important. Differences in x^3-distortion are generally easier to detect. The results on the effect of spectrum indicate that distortion of complex sounds is harder to detect than for simple sounds. While difference limen were markedly higher at 4 kHz for pure tones, it was much less so for musical sounds. This can be due to the different masking powers of sine and complex signals, and the upper frequency limit of the hearing system. DLND's are lower at 50 dB SL than at 80 dB SL for sine signals whereas they are independent of level in that range for complex sounds.

An interesting point made by Moir (1981) is that threshold depends on the duration of the distortion product. Tests were performed on sine waves to which clipping distortion was applied. For a 4 msec burst, the threshold value reached around 10 % and for a 20 msec pulse, the threshold was reduced to around 0.3 %. Therefore, "ten peaks each lasting 2 msec are less objectionable than one peak lasting 20 msec in the same listening period" (p. 33).

According to the author, threshold values between 1 % to 5 % should be expected in practical situations. Equipment with distortion levels around 0.01 % should not sound cleaner than another piece of equipment having distortion levels around 0.1 %.

Colorations and Resonances

Besides phase and non-linear distortions which can produce audible colorations, resonances in the audio chain are a major cause of such colorations. The particular timbres of musical instruments and voice are produced by resonance mechanisms. Several links in the audio chain can introduce additional resonances which may modify the quality or timbre of the desired sound. Colorations are often added deliberately but they can also be side effects of the signal processing.

The effect of these resonances on timbre depends on several factors among which Q, center frequency, and amplitude of the resonances were studied by Bücklein (1962). Audibility of peaks and dips in the spectra of music, speech, and white noise samples was measured for 10 listeners. The audibility of these irregularities was measured as a function of their form and depth, the type of stimuli, and the frequency region of the perturbation. Various types of music, from solo instrument and voice to orchestra were used. Listening was done with headphones. The subjects' task consisted in determining the order of presentation of pairs of stimuli, one of which was distorted. The frequencies tested ranged from 150 Hz to 11 kHz.

One-octave wide 10-dB valleys are easily heard except at f = 11 kHz. The threshold of audibility was estimated to be for 1-octave wide 5-dB dips. Antiresonant type dips of 20 dB for Q values of 1.8, 2.9, and 5 were also used. Threshold increases as the Q increases. Wide dips are better detected than narrow ones.

Peaks in the frequency response proved to be easier to hear. 6-dB peaks at 1 kHz (Q = 6.7) were detected by 60 % subjects. Listeners reported that valleys, even when detectable, didn't change the perceived timbre as much as corresponding peaks. Unfortunately, threshold values for peaks are not included in the Bücklein's paper.

The author also notes that spectral differences were harder to hear when solo instrument sounds were used. Tones had to be in the frequency region of the irregularity in order to reveal it.

As with musical sounds, peaks were much more easily detected than dips when using 6-sec white noise stimuli. Curve depths of 5, 10, and 15 kHz were tested. Only 15-dB dips at 3 and 5 kHz were heard. Interstimulus interval played a more important role for dips than for peaks. Most subjects could hear 5 dB peaks with a Q of 10 while only 60 % could hear the corresponding dips. A subject with perfect pitch could detect 90 % of the dips in white noise[20]. Fryer (1965) determined that the threshold for a resonance increases by approximately 3 dB per doubling of the Q value.

Toole (1986) suggested a few explanations for the greater audibility of low-Q resonances compared to large-Q ones:

1º "... from a purely statistical point of view, a broad resonance will be excited more often by sounds in music than a narrow resonance" (p. 232).

2º Low-Q resonances reach full amplitude more often and more quickly than high-Q resonances.

3º A low-Q resonance "can ring at a frequency significantly different from the one that initiated the response, thus imparting a monotonal coloration to a range of exciting frequencies" (p. 233).

Toole and Olive (1988) studied the audibility of resonances in the context of subjective evaluation of timbre in live performances and in reproduced sound. In addition to the factors mentionned above of Q, center frequency, and amplitude, the perceptual

---

[20]One might investigate whether absolute perception of pitch influences in any way the perception of frequency response irregularities.

significance of resonances depends also on the "temporal relationship between the resonance and the signal it modifies" (p. 123; see Barlow 1978).

When there is no time delay between the resonance and the signal, as is the case for resonances produced by spectral filtering, resonances inherent in the sound source (voice and musical instruments), and some resonances from loudspeaker and microphone construction, the addition of reverberation at the recording or reproduction stage can lower the threshold of low- and mid-Q resonances in "discontinuous, impulsive, or transient sounds" by up to 10 to 14 dB (p. 138). The authors conclude that subtle timbral shades in transient sounds can be more audible when listening through loudspeakers in a moderately reverberant room than when listening with headphones. It is also implied that added artificial reverberation can enhance the audibility of artificial formants in transient sounds but the effect is opposite when using continuous sounds such as vowels and organ.

Delayed resonances occur when delayed reflections with different spectral contents are present. For example, many musical instruments radiate sounds with different spectral components in different directions. Interaction of these with room boundaries can produce delayed resonances. With white or pink noise, threshold increases as the amount of delay is increased. In headphone listening, threshold can increase by 25 dB for delays from 0 to 60 msec (p. 139). With transient signals, the opposite occurs. Threshold can drop by 40 dB as the delay is increased from 0 to 20 msec in headphone listening (p. 139). The higher the frequency, the lower the threshold value.

When adding reverberation either at the recording stage or at the reproduction stage, variations in threshold as a function of time delay are reduced as threshold values are systematically raised (p. 139). The interaction between reverberation and the audibility of resonances implies that spectral equalization applied in a given environment will not necessarily produce the same perceptual results in another environment; the results will depend on the amount of reverberation in the recording and the amount of reverberation added by the listening environmnent.

2.2.2 Perceptual dimensions of timbre

In several experiments investigating perceptual thresholds of distortion products and colorations, subjects' task usually consists in detecting the presence of distortion by indicating if, basically, there is a change in timbre. No qualitative judgments are required. A good reason is that it is much more difficult to perform qualitative judgments than quantitative ones. Nevertheless, a number of authors have attempted to derive psychophysical relations for the timbre (usually denoted by the broader term of sound "quality") of reproduced sound (Eisler 1966; McDermott 1968; Nakayama/Mivra 1971; Staffeldt 1974; Gabrielsson, Rosenberg, and Sjögren 1974; Gabrielsson and Sjögren 1979; Gabrielsson and Lindström 1985; Toole 1985, 1986; Gabrielsson et al. 1988). All these investigations are concerned with the reproduction side of the audio chain and the aim of these experiments is to evaluate the reproducing equipment itself (mainly loudspeakers) whereas the main interest of this thesis is the evaluation of the sound as produced, modified and perceived along the whole audio chain. Thus the choice of certain dimensions or scales might apply more to the characteristics of the reproducing equipment than to the characteristics of the sound. Furthermore, the particular sound stimuli used in the various experiments influence the type of dimensions obtained. For example, studies using speech samples will tend to emphasize dimensions related to clarity and loudness whereas studies using music samples will produce dimensions related to spaciousness and aspects of timbre

(McDermott 1968; Gabrielsson et al. 1988). Perceptual dimensions can be used to organize and quantify, to scale comparisons between different timbres[21].

In experiments conducted by Eisler (1966), 4 sound engineers evaluated the overall quality of 10 different reproducing systems on 24 different sound examples. After factor analysis nine dimensions were found, seven of which were interpreted: sound level, purity of transients, environmental information, bass boost, full-treble reproduction, high-treble relative midrange, and disturbing directional effects.

In experiments on listeners' preferences for multichannel reproduction, Nakayama et al. (1971) obtained 3 dimensions: fullness, depth (of the image sources), and clearness. Their respective importance for overall sound quality is in the same order. From the results of Nakayama, there seems to be a close relation between fullness and the amount of early reflections from the side. That fullness is the most important attribute in respect to this particular experiment is in agreement with the findings of Woszczyk (1979) and Benade (1985).

Staffeldt (1974) found two dimensions, "emphasized bass" and "emphasized treble", both related to frequency and phase response of loudspeakers.

The most substantial body of research on timbre of reproduced sound came from the work of Gabrielsson and his coworkers, and from Toole (1985, 1986). Gabrielsson and Sjögren (1979) elaborated a set of 8 dimensions that were subsequently used in various listening tests (Gabrielsson and Lindström 1985; Gabrielsson et al. 1988; Toole 1985). The dimensions and the possible physical parameters underlying them are:

---

[21]"... il est utile d'explorer l'espace perceptif du timbre, pour dégager certaines dimensions suivant lesquelles s'articulent chez l'auditeur ressemblances ou relations de timbres" (Risset 1978, 9).

1. Clarity (clearness/distinctness): a sound reproduction will score high on this scale if it has a broad frequency range, a flat frequency response, and a low non-linear distortion level. Systems with a narrow frequency range, strong resonance peak(s), and high distortion levels will get a low rating. It is suggested that opposites for this dimension could be termed roughness/harshness. If the sensation of roughness increases as the number of components within a single critical band increases, then the more distortion products are present, the rougher becomes the sound. Clarity and definition of the sound enables the listener to hear and distinguish individual instruments, voices, notes, and attacks (Toole 1985).
2. Sharpness/hardness - softness: sharpness/hardness is related to a "steeply" increasing energy level at higher frequencies or to resonance peaks at higher frequencies, especially around 2-4 kHz. It is also characterized by a weak bass response and depends on distortion and sound level. Softness is related to the naturalness of the high-frequency sounds (Toole 1985).
3. Brightness - darkness/dullness: similarly to the dimension in 2, this dimension is related to the energy balance between low- and high-frequency components. This is an example of two dimensions being sometimes judged as independent, sometimes covarying. The authors explain this ambiguity by saying that they may be different dimensions depending on the same physical factors.
4. Fullness - thinness: a sound is full if it has a broad frequency range and some emphasis of the low-frequency region. It also varies with sound level.
5. Spaciousness (feeling of space): the psychoacoustical relations for this dimension are not conclusive. The stereo aspect was not examined in this experiment.

6. Nearness: this is related to intensity of the sound. There are also no clear conclusions about the relation with the frequency response.
7. Absence of extraneous sounds (disturbing sounds): noisy, crackling systems would be related to an "increased response at high frequencies" (Gabrielsson and Sjögren 1979, 1031). Obviously, a high-frequency boost may reveal noise problems caused by other factors.
8. Loudness: this would be more or less synonymous to sharp or hard.

Half of these dimensions (1-4) have clear correlates with spectrum energy distribution, and hence with timbre. Gabrielsson et al. (1988) conducted listening tests to determine the effects of different frequency responses on the evaluation of sound quality along the 8 scales: fullness, loudness, brightness, softness, nearness, spaciousness, clarity, and total impression. Monophonic sound examples included female and male speech samples, jazz and female solo voice. The various frequency responses were achieved by using 5 different filter circuits. The settings were:

1- flat;

2- boost of 6 dB/8ve between 1 and 4 kHz

3- 6dB/8ve cut below 1 kHz and 6dB/8ve boost between 1 and 4 kHz

4- 12dB/8ve cut below 1 kHz and 6 dB/8ve boost between 1 and 2 kHz

5- 12 dB/8ve cut below 1 kHz

Subjects listened through supra-aural headphones in sound-isolated rooms.

Settings 3 through 5 led to a decrease in fullness, softness, nearness, spaciousness, clarity and total impression, and an increase in brightness. Setting 2 features an increase in brightness, nearness, spaciousness, clarity and total impression and a decrease in softness relative to setting 1. Number 2 got the highest rating of all 5 on the total impression scale. For the normal hearing subjects, highest correlation with the total impression scale occurred for fullness, followed by clarity, nearness, spaciousness, softness, loudness, and brightness (p. 173).

The relative importance of the dimensions for good sound quality were ordered by subjects with normal hearing as follows:

	Music	Speech
spaciousness	1	7
fullness	2	6
clarity	3	1
brightness	4	5
loudness	5	2
softness	6	4
nearness	7	3

The usefulness of the perceptual dimensions developed by Gabrielsson and his colleagues was demonstrated in extensive listening tests conducted by Toole (1985) who obtained reliable and repeatable sound quality ratings for a large number of loudspeakers.

2.2.3 Descriptive labels for timbre

The use of labels to identify the different dimensions of timbre presents a number of problems as pointed out by several authors (von Bismarck 1974b; Toole 1985; Gabrielsson 1979; Meyer 1978; Letowski 1989). Words can have different meanings for different listeners. The meaning of descriptive terms may be determined by an individual's frame of reference. This frame of reference is determined by past listening experience, context in which the sound is listened to, and even cultural background (Solomon 1958). There are also translation problems. Shades of meaning may not be easily transposable from one language to another.

The ability to communicate timbre evaluations is nevertheless essential in sound engineering. Some words are more intuitively accepted as describing a particular aspect of timbre, e.g., sharp, dull, bright, etc. And yet the information they carry is not precise and is subject to many interpretations. A number of authors have attempted to relate verbal descriptors to specific physical characteristics of the sound (Gabrielsson 1979; Bartlett 1983). But until a standard vocabulary of timbre is developed (elaborated), using words to describe timbre will remain problematic.

Salmon (Salmon 1950a, 1950b) proposed a collection of words and expressions to describe timbre and overall quality of reproduced sounds. He identified four general categories of terms: 1- adjectives that describe the perceptual effects of noise, distortion, etc.; 2- terms characterizing specific frequency regions; 3- terms used to compare sensations produced by two frequency regions; and 4- terms used to describe more subtle sensations related to overall sound quality. Category 1 includes terms such as clean, dirty, harsh, rough, etc. Salmon subdivided the audible range of spectrum frequencies into slightly overlapping regions defined as:

extreme lows:	< 100 Hz
lows:	100 - 300
lower middles:	300 - 800
upper middles:	800 - 1500
lower highs:	1.5 - 4 kHz
highs:	4 K - 8 K
extreme highs:	> 8 kHz

An excess of lows will result in a "boomy" sound, a deficiency of lower middles will produce a "thin" sound. Increasing the energy in the lower highs will produce sounds from brilliant to bright, metallic, brassy and shrill. An emphasis of the high frequencies will result in a crisp, hard, or harsh sound. Terms used in frequency range comparisons include bump, peaked, shelving, etc. Terms in category 4 are more general in nature: "intimacy presence" can be obtained by a reduction of reverberation, audibility of certain noises (breathing of a singer, key clicks, bow scraping, etc). "Detail presence" relates to the ability of a listener to pick out individual sources in an ensemble. Other related terms are transparency and clarity. Emphasis in some frequency ranges such as the low middle frequencies may impair this quality. "Source size" seems to be determined be liveliness and spectral balance. Other terms are used to evoke spatial perception, source size, realism.

The vocabulary described by Salmon does not identify independent perceptual dimensions. The same words are used in different categories, causing some confusion in their meaning. His particular choice of words may seem at times questionable. His work nonetheless points to a few possibilities for the development of a vocabulary of timbre: for example, equating frequency-based timbre categories with selected descriptive words.

Vowel labels

Vowel labels can be used as timbre identifiers. The similarity between vowel timbre and musical timbre was pointed out in chapter two (Slawson 1968). Meyer (1978) proposed to relate the timbre of musical instruments to the formant regions attached to the vowel timbres, as an alternate method of "illustrating frequency regions in the sound of musical instruments" (p. 26). He used a set of 8 German vowels each identified by a frequency range and a peak frequency:

Vowel	**Formant Region**	**Peak of the curve**
u (boot)	200 - 400 Hz	349.23 Hz
o (beau)	400-600 Hz	493.88 Hz
å (fall)	600-800 Hz	740 Hz
ö (her)	1.2 - 1.8 kHz	1396.91 Hz
ü (rue)	1.2 - 1.8 kHz	1567.98 Hz
ä (air)	1.2 - 1.8 kHz	1760 Hz
e (née)	1.8 - 2.6 kHz	2217.48 Hz
i (geese)	2.6 - ≈ 4.5 kHz	3153.96 Hz

Meyer then associated verbal descriptors to formant characteristics. For example, "sonority" would be characterized by the presence of strong components in the "u" formant region (200-400 Hz) and especially the "o" formant (400-600 Hz). A powerful timbre has a distinct "a" formant; a pungent sound is similarly characterized with emphasis especially in the 1 - 1.2 kHz region; a nasal sound has emphasis in the 1200-1800 Hz region with a weak fundamental and a lack of energy at higher frequencies; clarity and brilliance are characterized by strong "e" and "i" formants.

Opolko (1982) proposed a system in which a given (musical) sound is spectrally analyzed in order to determine peaks that can then be equated with formants F1, F2, and F3. The ratios between these peak frequencies are then compared to those characterizing eight pre-defined vowel types[22]. After a vowel identity is attached to the sound, a descriptive term can be used to more fully characterize the sound quality. The eight vowel-type categories are: heed, hid, head, had, hawed, hood, who'd. The descriptive terms used to supplement the vowel-type categories for string instruments are: full/solid, bright/trebly, open/clear, sharp/harsh, present, nasal, thin/constricted, bassy/soft/dull.

A number of questions are raised by the use of vowel labels to describe timbre. Using formants of vowel sounds as anchors for timbre identification limits the spectrum to frequencies below 5 kHz. Such a system implies that timbre is strongly categorized, what happens when a particular timbre doesn't correspond to any of the pre - defined vowel types? In polar scales such as those used by Toole and Gabrielsson, "in-between" values can be easily expressed; could there be transitional values for the vowel-category system? There are also advantages. The method eliminates the ambiguity inherent in the use of descriptive adjectives. Vowel formants provide an already universally accepted set of frequency regions. Of course, there are differences between the different languages. One should perhaps choose the one with greatest resolution or combine vowel sounds from different languages to obtain a more complete scale, as in Meyer (1978). Additionally, such a system can be quite precise in terms of the objective frequency correlates.

The findings presented in chapter 2 and 3 suggest a clarification of the meaning of timbre. In musical contexts, timbre is essentially the perception of spectral variations over

---

[22]The use of formant ratios is arguable. Vowel formants are characterized by absolute frequencies and not ratios. This is an attempt to apply the formant model to instruments which do not have a clear formant structure. In reproduced sound, the vowel formant analogy can be applied to artificial formants produced by room resonances or microphone placement without the need to use ratios.

time. These variations can be triggered by acting directly on the spectrum itself — modifying the distribution of energy among the harmonics and partials — and indirectly by varying other attributes of sound — intensity, duration, reverberation, etc. The manipulations can be done globally (mixing) and locally (equalization of a single instrument, modification of the attack rate of a note or of a single partial).

CHAPTER 3

PERCEPTUAL JUDGMENTS, LEARNING, AND MEMORY

This chapter covers topics related to hearing and listening performance. First, the notion of levels of timbre perception sensitivity is presented. Factors determining perceptual judgments accuracy are enumerated as well as general criteria for the evaluation of a listener's hearing abilities. Then, studies on auditory perceptual learning are reviewed and finally the subject of timbre memory and learning is discussed.

3.1 Levels of timbre perception acuity

As mentioned earlier, timbre has various meanings. Often, timbre refers to the "tone quality" of a musical instrument, to the "texture" of a sound. In several studies on timbre, the term is rather associated with the identification and discrimination mechanisms between **different** instruments (the timbre of a clarinet as opposed to the timbre of a bassoon). Timbre is also used to talk about the subtle shades of color that the sound of the **same** instrument can have. Grey (1975, 104-105), Strawn (1982, 7), and others have suggested that the perception of tone quality (or timbral quality) and the identification of musical instruments are two different processes. One strong argument for the distinction between the perception of tone quality and musical instrument identification (source identification, in general) is that the tone quality of a given instrument can be significantly altered by filters or other signal processing methods and yet be readily identified as a clarinet, a piano, etc.[23] I

[23]Another explanation for that is the redundancy of timbral cues: See Bregman 1984

would like to further suggest that this difference in the meaning of timbre may also be considered as a difference in level of perception sophistication or "refinement".

One can distinguish four basic levels of sophistication in the perception of a sound's timbre:

1. detection: this is the most basic level of sound perception related to the intensity of the sound (Roederer 1979, 138). (Is there a sound present? or Are there additional harmonics present?)
2. discrimination: this could be described as the ability to notice a <u>difference</u>, if any, between two sounds. (Does sound A have the same quality as sound B? Or is it different?)
3. identification: this is the ability to <u>recognize</u> a given sound, to associate a label of some sort with a particular sound; e.g., the ability to determine which instrument is playing (Is it an oboe or a bassoon?).
4. evaluation: this may be the highest level of sophistication in timbre perception. It implies the ability to <u>describe</u> the tone quality of a sound (Is the sound dull or bright? How bright is it? Which region(s) of the spectrum is (are) emphasized or lowered?).

The first three levels correspond to basic perceptual tasks to be described in the next section. In the fourth one, the ear is used more or less as a measuring instrument.

3.2 Listener performance

Hearing acuity and listening abilities are critical issues in sound engineering.The resulting quality of a recording or broadcast depends for a major part on the engineer's perception of the sound. Many factors can alter a listener's perceptual performance: physiological factors related to the functioning of the hearing system, judgment ability, listening experience. These will be reviewed in this section.

3.2.1 Criteria to evaluate a listener's performance

A person's perceptual performance is determined by sensitivity (the ability to detect an uncertain stimulus), resolving power (the ability to discriminate), and by channel capacity and storage (the ability to identify an uncertain stimulus). If all three are demonstrated, then we can say that this person has the ability required to process a given stimulus, or a given set of stimuli (Watson 1981). Detecting, discriminating, and identifying are thus basic perceptual tasks on which subjects can be evaluated. As will be seen, these tasks vary in difficulty, memory requirements, and learning time.

3.2.2 Detection

The ability to detect a sound or a change in a sound depends on the sensitivity of the ear. Sensitivity can be evaluated by measuring absolute and differential thresholds. Threshold values for various types of stimuli were given in chapter 3. Sensitivity can be decreased by hearing impairment, listening fatigue, lack of attention

3.2.3 Discrimination

Discrimination can be defined as the process of delayed or paired comparison by which two objects, in our case two sounds, are judged as being the same or different. A delayed comparison consists in presenting a standard stimulus and, after a time interval, a comparative stimulus to be judged relative to the standard. It has been found that the variability of judgment increases with the interstimulus interval. Using a single standard repeatedly in a rehearsal process helps "to reduce trace variance for the standard and also for criterion variability" (Sandusky 1975, 72). Bindra et al. (1965) found that it consistently takes more time to judge that two sounds are the same than to judge that they are different. Also, it is more frequent to judge two sounds to be different when there are not than it is to judge two sounds to be the same when they are different. It takes more time when the task is to decide wether or not 2 sounds are the same than when the task is to say if the 2 sounds are different. So, it seems easier to compare two different sounds than it is to compare two identical sounds.

3.2.4 Identification

In identification tasks, no immediate comparison standard is available. "The observed stimulus values are assumed to be compared to internal criteria which serve as standards and result in classification" (Sandusky 1975, 75). The accuracy of the identification is limited by the "channel capacity" of the subject (Miller 1956): it decreases as the number of presented stimuli increases. The correlation between the stimulus input and its perception increases at first and then levels off at some asymptotic value. This asymptotic value is the channel capacity of the subject: it is the greatest amount of information that the observer can extract from the stimulus through an absolute judgment. The amount of input information (variance or the size of the stimulus catalogue) can be increased by increasing the number of alternative stimuli that must be distinguished (identified). Confusion will appear near the point of maximum "channel capacity" (p. 83). The human capacity for absolute judgments of unidimensional attributes seems to be finite, the upper limit being around 5 to 9[24]. There is an obvious discrepancy between this number and our judgment capacity for multidimensional stimuli such as complex sounds[25]. It seems that "the addition of independently variable attributes to the stimulus increases the channel capacity" (p. 88) but at the same time, the accuracy for individual variables is decreased.

3.2.5 Hearing losses

Hearing losses can significantly reduce hearing acuity. Sound engineers are a target population for hearing losses introduced by prolonged exposure to high-level sound intensities. Hearing loss is caused by physiological, biochemical, and anatomic changes of the hearing system. In particular, damage to the hair cells in the cochlea —"where pressure waves are transduced into electrical impulses that the nervous system interprets as sound" (Martinez and Gilman 1982, 686) — impairs the system's ability to transduce an acoustic signal. "Destroyed hair cells are not replaced by other hair cells, nor do they regenerate" (p 686). Perceptual effects of hearing loss include:

---

[24]This range is not absolute and could reach around 14 (Norwich 1981).

[25]We can easily identify individual sounds among a large collection of instrumental timbres. It would seem that the number of different sounds that can be identified increases as the perceptual distance between the various sounds increases. One must also consider that identifying a stimulus is a more difficult task than discriminating among a set of stimuli: "our ability to discriminate is more acute than our ability to make absolute judgments" (Norwich 1981).

- reduced ability to discriminate between frequencies;
- reduced ability to encode rapidly changing frequency/intensity information in a signal;
- need for greater intensity for the detection of an acoustic signal (p. 686).

The initial frequency region initially affected is between 3000-6000 Hz. Hearing loss is also caused by the aging process and initially affects frequencies above 4000 Hz, the ear acting as a low-pass filter. Both effects can combine.

It appears that a strong indicator of listener's judgment variability is the sensitivity (hearing level) below 1 kHz (Toole 1985). Particularly important is the range of levels below 20 dB, a range considered as normal in standard audiometric tests. Obviously, hearing requirements for critical listening are different than for speech communication. Impaired listeners "hear less, and less well" (p. 27). The perception of several aspects of sound can be affected by reduced sensitivity: "temporal integration of short-duration sounds, ability to localize sounds". The first one implies that there could be problems in the detection and loudness perception of short sounds, causing "signal-dependent dynamic-range distortion" (p. 27) The second one could reduce "the ability to differentiate sounds in space", therefore diminishing the possibility of using binaural cues to discriminate unwanted sounds and reverberation. Also, the "loudness curves" could be significantly modified.

One must make a clear distinction between "how well people typically do perform from how well they can perform in highly controlled and nearly optimal situations" (Trahiotis 1983, 63). Most studies done in the laboratory investigate the highest levels of auditory behavior in ideal environments, striving to eliminate all possible interferences outside of the one aspect under scrutiny. In practice, a sound engineer's hearing system is not always "operating with maximum efficiency" (Toole 1985, 27). Results have been obtained indicating that "25-minute exposures to 100-dBA popular music produced measurable temporary threshold shifts and related changes in spectral balance in recordings made before and after the exposures (Toole 1985, 28)[26]. More recently, Laroche and Hétu (1988) suggested that the decrease of temporal integration and the improvement of difference limen in intensity might be better, more sensitive factors than temporary threshold shifts to measure the effects of long exposures at low sound pressure levels.

3.2.6 Listening attention

A listener's attention to various aspects of a sound can be very selective. A study from Greenberg and Larkin (1968) and reported in Trahiotis (1983) reveals that "subjects who were highly practiced at detecting a 1000-Hz tone in the presence of broadband noise were completely unable to detect infrequently presented tones which were 100 Hz or so above or below the 1000-Hz target" (Trahiotis 1983, 65). Apparently, listeners were attending to a narrow frequency band around the target, thus discarding tones sufficiently remote from it. These results agree with the critical band concept. Trahiotis concludes that "listeners may miss physically present and significant information simply because they are attending to some other aspect of the input" (p. 65), depending on which auditory cues they are using. In that respect, training of auditory acuity can include the learning of new clues and anchors that help detecting and discriminating signals in a complex sound presentation.

---

[26]A threshold shift is a temporary elevation of the sensitivity floor of the ear after exposure to a stimuli of sufficient magnitude and duration. See Moore (1982), 60-66

3.3 Auditory perceptual learning

Listeners considered to have normal hearing audiometrically do not necessarily have the same auditory capabilities (Johnson et al. 1987). This is especially true of untrained subjects. The observation that subjects' performance in aural tasks improves with practice has been reported by several authors of psychoacoustical studies. A recent paper (Watson 1981) presented a review of current knowledge on auditory perceptual learning. Proper training enables subjects to learn to hear selected properties of sound and hence to focus their attention on parameters of importance. The time required for learning depends on the complexity of the task and on the type of stimuli. In general, time course will increase from detection to discrimination to identification.

For simple sounds, e.g., single musical instrument notes, "asymptotic detection performance is approached in 1-2 hours, discrimination in 2-6 hours, and identification in 10-100 hours" (p. 101), the required time depending on the stimulus uncertainty[27]. When more complex stimuli are used, learning time increases significantly: detection can take several hours, discrimination 4-12 hours, and identification from 20 to several hundreds of hours. Here, the required time depends both on the complexity of the stimuli and on stimulus uncertainty. The required time depends also on the particular task and the way it is presented. The studies reviewed by Watson didn't specifically address timbre perception learning. Detection experiments were concerned with either sinusoids in quiet or in background noise, the latter taking more time to learn than the former. Discrimination experiments were concerned with frequency, duration, or intensity for sounds presented in pairs

---

[27]Watson defines stimulus uncertainty as "the predictability of the pattern to be heard on the next trial, typically scaled by the number of patterns in a catalog from which a sample is drawn on each trial" (Watson 1981, 101).

or within tonal patterns. The tasks in the auditory identification experiments included Morse code reception (the most demanding one), musical pitch identification, and temporal order identification.

Various studies of reproduced sound included sections on perceptual learning. Blauert and Laws (1978) investigated the effects of training on group delay perception. Delay values from 0.25 to 1 msec were used. There were 15 training sessions of 30 minutes each. Feedback was provided after each response and in case of error, the subject had to reevaluate the judgment. Threshold values dropped from 0.86 to 0.54 msec after one training session and reached an asymptotic value of 0.4 msec the second day. These remarkable results suggest that the effects of group delay under the specified conditions are quite easy to learn. However, these training tests were performed only for one subject. There is no reason to believe that this subject is representative of a larger population, even if we narrow it down to a sample of critical listeners. Second, no attempts were done to qualify the perceived differences in term of sound quality. A triadic forced-choice procedure was used and the task was basically one of detection.

Audibility of distortion is limited by listener "education" and characteristics of the hearing system. Familiarity with the sound of a given instrument may help to detect departures from it: "an individual trained in the sound of a particular instrument will be able to recognize slight departures in the reproduction of this instrument at the highest frequencies of the overtones of the instrument provided they are non-harmonically related tones" (Jacobs and Wittman 1964).

Measuring crossover distortion thresholds, Moir (1981) estimated the improvement factor in sensitivity to be at least 10 times for a 10-15 minute practice session. Apparently, the listener would learn after only a few comparisons to recognize the nature of timbral changes associated with a particular form of distortion[28].

3.4 Timbre memory and learning

Empirical evidence presented in the previous section indicates that learning does occur for various auditory tasks. However, the memory processes involved and the nature of what is actually learned is still under investigation.

Tonal information is memorized in a specialized storage area separate from speech (Deutsch 1975). Deutsch suggested that timbre memory could be a subdivision of a tonal memory system (p. 122). Roederer (1979) describes the process of timbre identification as consisting in the storage of information about the auditory signal in memory with an associated label of identification and the subsequent comparison of a new signal with the stored one (p. 139). He further mentions that timbre memory probably consists more in the storage of information-processing instructions than in the storage of the actual timbre data.

Profile analysis

A new body of research has been recently developed (Green et al. 1983; Green and Kidd 1983; Green 1983; Green et al. 1984; Kidd et al. 1986; Kidd and Mason 1988), focusing on auditory intensity discrimination, the perception of spectral shape modification,

[28]The retention period of this newly acquired skill is obviously of particular interest. It would be interesting to verify how long a particular performance level is maintained after a training session and how this retention period varies with experience. Also, the amount of practice time necessary to achieve asymptotic performance, the magnitude of this minimum value, and the frequency of practice sessions necessary to maintain the acquired sensitivity would be most useful.

and the related memory processes. The authors describe a theory called profile analysis which states that when faced with the task of detecting a change in the intensity of a single component of a complex sound (a change in spectrum), the hearing system will perform a simultaneous comparison of spectrum level at different frequency regions (across critical bands) in order to determine if there is a bump in the spectral envelope, rather than performing successive measurements of levels (within one critical band).

Whenever a listener has to discriminate between two sounds occurring at different instants in time, a comparison between the two must be performed. For a comparison to occur, the first tone must somehow be remembered, therefore it must be stored in memory. For complex sounds, it has been suggested that sounds are stored in memory in a "highly quantized categorization form" (Green et al. 1983, 641). Thus, a complex spectral description of a sound would be stored in long-term memory as "a few bits of information indicating that the sound is signal-like or non-signal like" (p. 641).This memory model seems in contradiction with the high levels of timbre perception acuity that can be attained The authors reply: "The relative gross categories assumed to represent the memory stores do not imply that the observer is insensitive to fine changes in the spectrum Logically, the accuracy of the categorization process is independent of the number of categories created" (p. 641).

Kidd and Mason (1988) found evidence that "the detection of a difference in spectral shape between a pair of sounds does not require the conversion of the sensory trace to a more abstract, nondecaying memory representation" (p. 147) and that a "sensory-trace mode" is also used in profile analysis in which "the listener maintains through rehearsal an image or trace of a sound stimulus in short-term memory" (p 144). In this mode, threshold increases with inter-stimulus interval This suggests that listeners use short-term memory for paired comparisons and long-term memory for absolute judgments

Kidd et al. (1986) investigated the learning processes involved in the discrimination of sounds that differ in spectral shape. Signals were 100 ms long and were presented binaurally through headphones. A two-alternative forced-choice procedure was used. The signal to be detected was an intensity increment to the 948-Hz component of the complex sound.The "boost" was obtained by the addition in phase of a second component of 948 Hz to the one already in the complex sound. Threshold then was defined as the difference in level between the two 948-Hz tones. Feedback was provided after each trial. Sounds were low-pass filtered at 3200 Hz.

The signal increment threshold was determined as a function of the number of trials for naive listeners. Between 150 and 3000 trials, threshold decreased by about 2 dB for each doubling of the number of trials. After the 3000 trials, threshold had dropped by 11 dB. This is not to say that no improvement should occur for more trials: "the most rapid improvement in threshold occurred during the first 500-750 trials with continued gradual decrease in threshold through at least 2500-3000 trials" (p. 1046). As to how the reference spectrum is stored in memory and subsequently compared with the test spectrum, the author suggests that:

> ... the critical band levels representing the spectrum of a sound are compared with a similar set of stored values reflecting the relative critical band levels of the reference stimulus, with the decision variable depending on detecting a difference between the two sets of values. The learning process is likely to involve the development of an accurate set of values of the reference stimulus in memory. Experimented listeners appear to be able to generalize this process of encoding to novel spectra or to new manipulations of a previously learned spectrum more quickly than naive listeners (p. 1051).

Papcun et al. (1989) suggested a theory of prototypes to explain how subjects remember and forget unfamiliar voices. A prototype is defined as a representative member of a category. "What listeners remember is a characterization of the voice they heard in terms of a prototype and deviations therefrom. As times passes, listeners lose information about the manner in which the voice they heard deviates from the prototype". Listening strategies for familiar and unfamiliar sounds seem to be quite different: "whereas unfamiliar voices are recognized in terms of the prototypes plus deviations, familiar voices are recognized by deviations alone. In other words, when listeners become familiar with a voice, they learn its idiosyncracies and no longer perceive it with respect to a prototype" (pp. 923-924). Learning, then, would consist in developing an ability to attend and remember the deviations that differentiate a sound from a known prototype without making further references to the prototype. The memory trace would hence be more impervious to interferences accumulating over time[29].

---

[29]See also Watson et al. 1976, Brandsford 1979, Spiegel and Watson 1981, Jones 1976, Zwislocki et al. 1958, and Siegel 1972 for other material on memory.

CHAPTER 4

TIMBRAL EAR TRAINING

4.0 Introduction

The discussion presented in chapter 3 provided substantial evidence that auditory perceptual acuity and memory can be significantly improved by controlled practice and training. In addition, perceptual judgment consistency, which is vital in sound engineering, also increases with experience.

The accurate control of sound quality at every stage of the audio chain requires highly specialized listening skills. Sound engineering practice involves the detection, discrimination, identification, and evaluation of timbre variations both in terms of the underlying physical parameters of the sound and the resulting percepts. Perception acuity and memory demands for these tasks are high. However, surprisingly little attention has been paid in the audio field to the importance of training hearing abilities of sound engineers and few reports can be found in the literature on the subject.

4.1 Review of past research

The need for trained listeners has been pointed out by several researchers conducting listening tests on loudspeakers (Toole 1985; Gabrielsson 1987; Beck 1987). In listening tests on perceived sound quality of reproducing systems using three groups of `isteners (naive, musicians, and hi-fi listeners), Gabrielsson et al. (1974) found that "naive" listeners were less reliable than musicians and hi-fi listeners. A tendency was also noted for naive listeners to give higher ratings to the poorer loudspeakers and to the reproductions with decreased treble than the other two groups. 'Hi-fi' listeners were characterized by having substantial experience in listening to high-quality reproducing equipment and in attending live concerts. Their performance was higher than listeners in the music group. Gabrielsson (1987) also reported that "experienced listeners may differ in ratings among themselves due to more of less familiarity with the type of music programs used in the test" (p. 58).

Beck (1987) defined "acoustic memory" as "the process that enables subjects to remember certain characteristics of an acoustic stimulus" (p 25). Unfortunately, he didn't specified what these characteristics could be. He found that a period of 24 hours was ideal between listening sessions in order to maintain a high reliability of the subjects He also found that having more than 2 sessions within 24 hours introduces fatigue and decreases a listener reliability. When sessions were separated by more than 24 hours, subjects gave ratings that tended to gather around the 'mean' loudspeaker[30]. Beck compared two basic test procedures in listening tests on loudspeakers: the paired comparisons and the single judgment method. In paired comparisons, loudspeakers are presented in groups of two and they are evaluated in respect with each other. In the single rating method, loudspeakers are evaluated separately. Beck found that subjects were more reliable and consistent in their judgements when using paired comparisons than when using single ratings. He concluded that for listening tests on loudspeakers, the paired comparison was a better approach.

[30]This phenomenon could perhaps be linked to the theory of prototypes proposed by Papcun, Kreiman, and Davis (1989) for unfamiliar voices: with time, listeners tend to forget the distinctive characters of sounds that differentiate them from some prototype and their perceptual judgements tend to cluster around that prototype.

Beck (1990) conducted experiments to investigate how training would affect subjects' performance (their reliability or "usability") in listening tests on loudspeakers. Ten subjects evaluated six pairs of loudspeakers on four different programme selections. There were six 38-minute sessions, the first two being given the same day, the remaining four being given one per day. A procedure of paired comparisons was used and each sound presentation was approximately one-minute long. Results indicated that the main effects of the training period were "a decrease in level of error variance and an increase in the perceived differences between loudspeakers" (p. 21)[31].

Hansen (1987) described the establishment and training of a panel of listeners for listening tests on sound reproducing equipment (loudspeakers, pickup cartridges, compact disk players, crossover networks, power amplifiers, etc.). The purpose of the training program was to allow panel members to develop their "ability to express their impressions at a listening test reliably and clearly" (p. 91). Weekly sessions of one hour and a half were scheduled. In addition to these sessions, four to five live concert attendances a year were planned in order to "help members in building up an absolute reference and standard of judgment" (p. 91).

A set of standard terms was developed to allow "more precise and accurate communication" between panel members. Terms such as full, bright, hard, and sharp were used to characterize various spectral balances in four pre-defined frequency bands (bass, low midrange, high midrange, and treble). "Training in terminology resulted in members using the same words to describe identical aural experiences, ensuring that descriptions were fully comprehensible" (p. 94).

---

[31]Beck defines error variance as "an estimate of the variance due to the errors made by subjects when making repeated ratings of the same stimulus" (p. 7).

Familiarity with the sound of specific musical instruments was found to be essential since "a precise knowledge of the sound of acoustic instruments is needed to accurately quantify certain parameters, for example, definition" (p. 96). Both recorded and live sounds were used for that purpose.

Both the paired comparisons and single stimulus rating modes of presentation were used. The latter produced a significant increase in the variance of listeners' responses at the beginning, but eventually became the procedure preferred by the listeners after about a year of training.

While listeners were reported to improve, no supporting data are provided. One should note that the training program described by Hansen was to help listeners evaluate the performance of audio equipment in terms of sound quality, and not specifically to train their timbre perception acuity — which obviously improved in the process. The design of the program reflects this objective.

Everest (1982) presented a course designed to develop acuity of listening skills for various aspects of sound quality. The author suggests that "the usual random learning process [of listening abilities] can be accelerated by subjecting the trainee to a wide variety of listening experiences in a carefully structured manner" (p. 2). Using cassettes and an accompanying manual, the author covers topics such as perception of frequency response irregularities, distortion, frequency band restrictions, etc. General theoretical concepts are presented along the sound examples.The limited scope of the program and the fixed delivery mode of instruction make this program inadequate for extensive training of hearing abilities.

More recently, Letowski (1985) reported on the development of a course called "timbre solfeggio" at the Chopin Academy of Music in Warsaw, Poland. The purposes of the course are to develop timbre perception skills – increase timbre sensitivity and improve timbre memory – and to elaborate a vocabulary "for exchanging information regarding timbre impressions" (p. 241). A wide range of signal types (complex tones, noises, isolated phonemes, samples of connected speech, electronic music signals, solo instruments, duets, music groups and orchestras with music genres ranging from pop to classical) are used to demonstrate parameters of sound quality that affect timbre: linear and nonlinear distortions, music dynamics, sound decay and attack times, reverberation times, and spectrum.

Two main types of exercises are used. In the active exercises, the sound to be evaluated is transmitted through two identical but independent channels which contain the same sound processing device. By listening alternately to both channels, students must duplicate in channel B the modifications introduced by the instructor in channel A. The processes involved in this exercise are described as follows by Letowski:

> To match both timbres, a student needs first to analyze, compare, and identify the elements that differentiate the two signals, and next to adjust the controls to neutralize the difference. This requires analytical listening since, due to the extremely large number of possible changes, guessing is practically useless. Comparative analysis of timbres is an essential requirement in searching for the most appropriate sound by changing microphone positions or by selecting a certain equalization circuit" (Letowski 1985, 241).

This type of exercise helps to establish relationships between timbre changes and the related physical parameters that the sound engineer can manipulate in the recording studio.

Passive exercises are used to develop both relative and absolute timbre perception. In the former, the students have to detect, identify, and verbalize timbral differences between two sounds. In the latter, identification and description of the sound quality must be performed without comparison with a reference.

The multidimensional nature of timbre and the practically unlimited number of possible shades of color called for the establishment, at the beginning of the course, of a limited set of timbral categories or standards that are stored in long-term memory. In the solfeggio course, nine categories identified by nine center frequencies an octave apart are used. They are : 63, 125, 250, 500, 1000, 2000, 4000, 8000, and 16000 Hz. This initial set is eventually extended to 27 1/3-octave frequencies. Such a model of frequency-based timbre perception thus uses well-defined, easily reproducible objective timbre categories against which sounds can be compared. In addition, descriptive adjectives are used as well as a system of vowel-like timbre categories.

The program described in Letowski's paper lasts three years (6 semesters). During the first year (the basic course), the 27 standards are learned. Very pronounced at the beginning, the timbre changes become more and more subtle as perception acuity is refined. Complex harmonic multitones and wide band noises are used to develop the nine basic categories. Sound analysis is not limited to the center frequencies, but also include "the width of the formants and the lowest and highest components in the sound" (p. 243).

The intermediate (second year) and the advanced (third year) courses deal also with "the identification of the natural formants of various musical instruments (frequency location, relative strength, approximate formant width), the audibility of various distortions in laboratory and commercial sound recordings" (p. 243), as well as the analysis of the sound stage (the number and the name of instruments and their respective location, etc.).

Results of this training program are reported to be excellent: in some cases, identification of center frequency, bandwidth and relative amplitude of up to 4 artificial formants was achieved by the students.

4.2 Program design criteria

It is proposed in this thesis that a computer-assisted program for timbral ear training would provide an excellent tool for the improvement and maintenance of the aural skills that are essential to sound engineers. The following section introduces the training system.

Advantages of the use of a computer-controlled configuration are discussed. A list of design criteria is presented along with hardware and software requirements. The last section presents preliminary data obtained from informal tests that were conducted with an alpha version of the system.

4.2.1 Rationale for a computer-controlled system

Research on the use of computers to assist in the development of aural skills has been going on for the last 20 to 30 years. Although early studies such as Swets et al. (1962) lacked the technological means to test viable systems (poor user interface, limited sound capabilities), essential questions were already addressed such as the use of feedback, self-pacing, task difficulty depending on previous performance.

In musical ear training, the use of computer-assisted instruction has proven to be very effective (Eddins 1981; Hofstetter 1981; Gross 1984; Killam 1984). Advantages of computer-assisted instruction include the individualization of the learning process, the possibility for the student to progress at her/his own pace, and the interactive environment made available for each student (Hofstetter 1981; Witlich et al. 1986; Bork 1986, 1987). In addition, the use of a computer allows the gathering of a wealth of information about a student's performance that can be invaluable in research areas such as auditory perception, learning strategies and memory.

The advantages of the use of a computer-controlled system for timbral ear training are manifold:

1) Individualized instruction:

Aural abilities vary greatly among individuals (Johnson et al. 1987), as well as learning speed (Watson 1981). The computer system allows the student to spend less time on mastered material and invest more on harder tasks. The trainee can thus take the time necessary to build her/his competence away from the frequently stressful context of a recording session.

2) Regular practice:

Training and practice on a regular basis are important to develop and maintain aural acuity (Letowski 1985). A computer system allows regularly scheduled training sessions and the possibility of extra time for individuals who need it.

3) Avoidance of group pressure:

Perceptual judgments are very sensitive to others' influence during early stages when confidence and skills are in development (Asch 1958) Individual training should remove such stress and improve the rate of learning. The student should thus be better equipped when he/she faces the task of a recording session.

4) Interactive learning:

Timbral ear training is particularly well suited for an interactive environment since the manipulation/modification of sound parameters is a key aspect. Traditionally, sound engineers improve their hearing acuity while working on recording sessions. However, a recording session doesn't necessarily provide an optimum learning environment and time for experimentation is scarce. Theoretical knowledge of perception is useful but knowledge of how the ear perceives sounds does not make one's hearing acuity better. practical knowledge is necessary. Interacting directly with the sound parameters and being able to hear the perceptual result is essential for thorough knowledge:

> To know and object is to act on it. To know is to modify, to transform the object, and to understand the process of this transformation, and as a consequence to understand the way the object is constructed (Piaget 1964, 176).

5. Constant, objective verification

The monitoring of the student's answers and manipulations by the computer makes it possible to objectively and systematically verify the performance and progress of the user.It also allows the collection of data on various strategies used by the student to accomplish a given task.

4.2.2 Delivery modes

There are various possible modes of interaction between the user and the computer. They vary in complexity and serve different purposes. Probably the most common is the drill-and-practice mode "in which the student is repeatedly exercised by being placed in a loop involving presentation, response, and feedback" (Witlich et al. 1986, 75). According to the authors, this method can be very effective provided that the feedback — the computer's response to the user's response — "is varied and appropriate to the task ... and if lessons permit flexible as well as fixed sequencing of the instruction" (p. 75). Keller (1987) points out that drill-and-practice is mostly useful to reinforce a skill that has already been learned. But it cannot be used to teach effectively new notions. A second mode called simulation is defined as "a replication of the behavior of a phenomenon of one's universe designed to substitute for the phenomenon" (Witlich et al. 1986, 76). While this mode has not found many applications in music, it could be used for example to simulate the directional properties of musical instruments. However, simulators are " best used in contexts where performing the necessary work would be either time consuming, dangerous, or exper ,ive (Keller 1987, 137) Finally, a tutoring system could be described as "a simulation of the interaction between the ultimate instance of an expert teacher and an arbitrary listener" (Witlich et al. 1986, 76). Another possibility is the 'heuristic' mode in which the student learn by exploring and discovery (Hofstetter 1981). Dolson et al. (1987) reported on a system for exploration and learning about sound and sound processing. The authors promote the idea that the best use of computers in education is not as in the drill-and-practice mode but rather as an "open-ended vehicle for exploration and active learning" (p. 314). In the system described by Dolson and his coworkers, students can use the computer to conduct their own experiments and to discover by themselves principles of acoustics and psychoacoustics like Fourier series, timbral brightness, insensitivity to phase, residue pitch , etc. This is accomplished by direct manipulation and modification of the signals with both graphic (visual) and aural feedback. The authors found that "the coordinated presentation of graphic and aural information can be an enormously powerful aid to learning" (p. 342)

Although it is the simplest to implement, the drill-and-practice mode with its repetitive and somewhat mechanical nature should not be the sole mode of presentation in timbral ear training. A combination of drill-and-practice and elements of tutoring where the computer could suggest appropriate listening strategies and demonstrate relationships between physical parameters of sound and the perceived timbre should be more appropriate and effective In the training program, there should be a mode in which the computer is in control: specific exercises have to be accomplished. In addition, there should be a mode in which the user is free to explore various timbre manipulations. The computer would still respond to the user actions, either upon request or at a predetermined moment, but would not impose any specific actions.

4.2.3 Hardware requirements

Killam (1984) lists five basic needs in an environment for aural skills development: sound, real-time interaction, individualization, student records, and research (p. 2). A computer-controlled system for timbral ear training imposes even higher and very specific demands on the hardware. Among them are:

1) fast and accurate access to high quality playback of a wide variety of sound examples
2) real-time sound processing
3) immediate feedback (when needed)
4) graphic interface allowing a viable metaphor of sound processing devices used to manipulate timbre in the recording studio; the graphics must also allow fast and accurate modifications of sound parameters under study
5) graphic representation of the changes in the physical parameters of sound or in the processing parameters
6) storage of student records and production of reports of progress
7) ease of use: most people are still unfamiliar with computers. The time required to learn the system should be reduced to a minimum. Energy should be spent improving hearing acuity, not computer proficiency.

Random access to sound material can be implemented in two ways. Hard disks can be used to store digital sounds. Using a sampling rate of 44.1 kHz, storage capacity is demanding and expensive but access is fast. The other possibility is to use a computer-controlled audio recorder (DASH or R-DAT) and/or CD-player. The disadvantage of the tape recorders is that access is sequential and hence adds delays in the presentation of aural material. Real-time signal processing must be implemented with dedicated hardware. There are two possibilities: a) The use of a DSP chip such as the Motorola 56000. This offers the most flexibility in the actual type of sound processing needed. b) The control of a signal-processing device by a computer. Here the use of MIDI links offers interesting possibilities. MIDI (Musical Instrument Digital Interface) is a standard protocol for exchanging control data between synthesizers, sound processing devices and computers[32]. Programming is reduced to sending the proper MIDI byte streams to specify parameter values to the device; the actual processing is done by the machine itself. It is less flexible though: the type of processing is limited to what the device can do – it cannot usually be expanded by the programmer.

Listening environment

Room, loudspeakers, and listening position choices

Several authors have stressed the importance of an appropriate monitoring system (Toole 1985; Somerville 1954; Børja 1977; Letowski 1985). Some of the effects on timbre of the interactions between the loudspeaker, the listening room, and the listener, were pointed out in chapter two. A thorough explanation and description of the problem cannot find its place in the present thesis. The International Electrotechnical Commission has published specifications for a standard listening room[33]. Since timbral ear training aims at developing perception acuity of timbre, it logically follows that a room with the least possible interference with the sound to be evaluated should be used together with high

[32]MIDI is mostly used for music performance, either live or in the recording studio. But the use of 'System Exclusive' messages in which only the first and last byte are included in the specification allows the transfer of data of arbitrary length and content.

[33]See Toole 1990 for more detailed description.

quality loudspeakers. Therefore, when listening is done to critically and technically evaluate sound, a room with a short reverberation time, the use of directional loudspeakers, and a near- to free-field listening position should be preferred. If headphones are used for monitoring, problems caused by a room with deficient acoustics can be avoided. However, other problems are associated with headphone listening. As we saw previously (Toole and Olive 1987), some aspects of sound quality are more easily perceived on headphones and others are most easily perceived on loudspeakers. This implies that the perceived sound quality is different depending whether headphones or loudspeakers are used.

Toole (1984) discussed important aspects of headphone reproduction that may affect sound quality The author states that "there are good reasons to be cautious about the choice and use of headphones in critical listening situations...Headphones are in essence small loudspeakers suspended close to the ears" (p. 1). An important problem in headphone listening is the coupling between the ears and the headphones: all ears are different. "Just as the listening room is the unpredictable link between a loudspeaker and a listener, the external ear is the uncertain acoustical coupling between a headphone and the eardrum. [So] the acoustical performance of a headphone [depends] upon the individual listener" (p. 2). The external ear acts as a sound collector for frequencies above 2 kHz:

> In normal listening sounds arriving from different angles are 'encoded' with distinctive frequency responses. Some of the resonances in the external ear are excited by sounds arriving only from rather narrow angular windows. ... Because of the substantial acoustical activity in the external ear at high frequencies a uniform sound field outside the ear can be elevated or reduced by up to 20 dB at the eardrum at some frequencies (p. 3).

The type of acoustical coupling between the ears and the sound source is different: "A headphone placed on the ear does not result in the same kind of acoustical interactions with the external ear as the complex sequences of more-or-less plane sound waves that are incident in normal listening" (p. 3). As a consequence, the sound quality in headphone listening is distinctive . Supraaural headphones tend to cause "large deficiencies in bass due to excessive air leaks" (p. 4). In that respect, circumaural headphones are better, but not ideal.

Letowski (1985) suggested that timbral ear training should be done on various types of loudspeakers and headphones to provide an array of timbres available to the listener. As pointed out by Kirk (1957), there is a tendency for listeners to prefer the sound of equipment they are used to even though it may be worst. It is well known that the sound quality can vary significantly from one pair of loudspeakers to another, from one listening room to another. Sound engineers often strive to make recordings that have a good sound quality both in the recording studio and in other listening environments. Listening to a variety of monitors might help to build a more stable set of timbral references across different listening conditions.

4.2.4 Software requirements

General Requirements

A number of points must be kept in mind when developing instructional software (also called 'courseware') in order to insure the most effective learning. As Gross points out, "the courseware designer faces two main problems: presenting material and helping the student learn" (Gross 1984, 40-41). Also stressed is the importance of keeping student records: "To help in the learning process, it is desirable to program diagnostics, and to keep records of the students' recent history. Records are essential in distinguishing between guessing and misunderstanding, a distinction that can only be made on the basis of a repeated record of responses" (p. 41).

Software for timbral ear training

Timbral ear training aims to improve the aural performance and reliability of the last element of the audio chain: the listener. Since there is still a lot to be investigated and learned about timbre perception and the listening strategies used by sound engineers, software for timbral ear training should meet two main requirements: teaching and research.

Teaching requirements

Timbral ear training should be concerned with "low-level" perception of timbre; emphasis should be on the development of perceptual skills, not of aesthetic judgment aptitude. The program should accommodate the development of both quantitative and qualitative evaluation of timbre. Quantitative evaluation involves the use of the ear as a measuring instrument. Qualitative or categorical evaluation involves the use of descriptive words and analogies with vowel colors. Exercises should be provided to:

1º develop in long-term memory a set of timbral category references against which sounds to be analyzed can be compared.

2º provide the student with a knowledge of the relationships between perceived timbral changes and the physical parameters of sounds responsible for these changes and under her/his control.

3º develop perception acuity of fine timbre differences (paired comparisons); this is the ability to compare a given timbre with an absolute reference.

4º develop the ability to describe the timbre of a given sound (absolute judgment); this is the ability to compare a given timbre with a personal internal reference.

Research requirements

Student records should be kept to evaluate their performance. Information should include data defining both the questions and the answers, response time, session duration, etc. Ideally, every user's actions (fader manipulations, button and menu selection, and pauses) should be recorded in a time-ordered sequence so that all the steps the user took to arrive at a particular answer are known. Doing so would provide the computer the information necessaary to interact effectively with the user.

User's evaluation

There are a number of problems involved in the evaluation process For the case where an exercise requires the identification of the center frequency of a resonance, the evaluation is more or less straightforward. A tolerance must be introduced to account for small deviations within critical bands: depending of the sounds used, a collection of answers around the target frequency will sound the same and hence should be considered as equivalent answers.

A formula to calculate the critical bandwidth[34] at any frequency was provided by Zwicker and Terhardt (1980):

$$CB_c = 25 + 75\,[1 + (f\,/\,0.85\ \mathrm{kHz})^{2}]^{0.69}$$

where CB_c = critical band in Hz and f = the center frequency in kHz.

[34]The critical bandwidth is defined as the width of the frequency band within witch a modification of different frequencies produces the same perceptual result.

Values obtained with this formula should be used with caution: audibility of a given spectral modification can vary depending on the listening conditions, the sound stimulus, the listener, etc. Recent findings by Moore et al. (1990) and Shailer et al. (1990) indicate that values usually calculated for the critical bandwidth function should perhaps be revised. The equivalent rectangular bandwidth (ERB)[35] of the auditory filters obtained in these two studies were narrower than the corresponding critical bandwidths usually cited in the literature. The values were obtained using the formula:

$$ERB = 24.7\ (4.37F + 1)$$

where ERB = the equivalent rectangular bandwidth in Hz and F = the center frequency in kHz.

For exercises where the task consists in replicating multiple modifications applied to the original sound on several parameters (frequency, Q, gain), evaluation of the student's answer by the computer becomes quite complex. The same perceptual result can be attained by different combinations of settings. Theoretically, a "good" answer is one that produces the same perceived timbre. The computer could provide the feedback in the form of a display of the original settings of the problem and the student would then compare it at will with her/his answer. The instructor, later, when collecting data, could test aurally the answers.

---

[35]"The ERB of a filter is the bandwidth of a rectangular filter which has the same peak transmission as that filter and which passes the same total power for a white noise input" (Moore 1989, 334).

4.2.5 The sound library

The basic teaching material used in timbral ear training consists of sound. The students manipulates, compare, and evaluate the timbre of sound examples. These must be selected very carefully in relation to the exercises they are used for. The relationship between the sound material and the exercises is most critical: modifications (equalization, filtering) will be audible only if the material allows them to be audible. For example, consider an exercise in which the student must discriminate between spectral peaks with center frequencies ranging between 100 and 500 Hz If the sound example presented to the student is a short excerpt played by a piccolo in the upper range of the instrument, it will be very difficult to find the right answer. The sound examples should illustrate clearly the problem presented in an exercise. It should be chosen so that it helps the student focus her/his attention on the proper parameters of sound.

A collection of sound samples should be assembled with the following general criteria in mind:

1. The sound example needs to contain proper spectra for the exercise. The sound could be analyzed spectrally to verify objectively its suitability (as long as the analysis reflects the perceived timbre).

2. The sound example needs to be of sufficient length so that proper evaluation of timbral aspects can be performed without constantly having to go back to the beginning of the selection. Since the time needed to evaluate timbre can vary with the difficulty of the task, an alternative is to provide facilities to loop examples so that they can be continuously repeated.

3. Spectral content and the loudness of the sound example should be as homogeneous as possible to avoid perceived timbre fluctuations during its presentation.

4. Because the student must often listen repeatedly to a sound example in order to provide the proper answer to a given problem, sounds should not be monotonous or irritating[36].

5. The type of sound should be carefully selected according to the particular task at hand, its difficulty, etc. For example, simple stimuli (white/pink noise, single musical instruments) should be preferred at early stages of the training and more complex sounds (ensembles, large orchestras) should be used at later stages.

6. In listening tests on loudspeakers, mono recordings are generally preferred when the purpose of the test is the evaluation of timbre or sound quality. One of the main reason is that loudspeakers differ widely in their spatial reproduction characteristics which in turn affect the reproduced timbre of the sound. Toole reported that "variations in listeners' opinions, as a function of hearing threshold level, increased much more rapidly in stereo than they did in mono test" (Toole 1990, 6)[37]. Poor loudspeakers judged as such in mono reproduction are sometimes given a better rating when listened to in stereo. As Toole points out, "stereophonic listening presents a listener with much more information, and therefore a much more difficult task" (p. 6). Stereophonic reproduction can help masking distortion elements in the sound and thus prevent proper response to exercises that focus of perception of spectral modifications due to distortion products. In technical ear training, the purpose is not to evaluate the relative merits of different pairs of loudspeakers; it is to develop and improve the ability to perceive changes in timbre and to build a stable set of timbre standards or references in long-term memory. Interferences caused by the listening

---

[36]Stimuli such as white and pink noise are an exception. They are very useful, especially at early stages, because they contain all frequencies. However, they shouldn't be used for extensive periods of time.

[37]See also Toole (1985).

environment and the mode of presentation can preclude the building of objective timbre standards. For that reason, monophonic sound examples should be used when the main focus of an exercise is timbre evaluation. The use of monophonic sound material however imposes a restriction on the choice of sources. Only sounds recorded with one microphone or stereo sounds which are 100 % compatible with mono should be used in order to avoid spectral cancellations.

4.3 Preliminary implementation

An alpha version of a computer-assisted system for timbral ear training was built and informal tests were run with 9 subjects during the 6-month period of development. The following section describes the hardware and software for that system. Then, exercises are presented together with preliminary data obtained from these early trials.

4.3.1 Hardware

The most direct and simplest way to manipulate the timbre of a sound is to modify its spectral balance. Frequency equalizers and filters are thus well suited for timbral ear training. Parametric equalizers are necessary so that center frequency, gain, and bandwidth of the modification can be directly controlled. A system was assembled using an Apple Macintosh IIx computer with 2 megabytes of RAM, an Opcode Studio Plus Two MIDI interface, and two Roland E660 digital parametric equalizers. Operation of the digital equalizers was entirely controlled through the computer using MIDI. A Sony CD player (model CDP-C100) with multiple disks capacity was used to present audio material. Sounds (white and pink noise, male speech, solo piano, solo cello, harpsichord, and orchestra) were selected from the Denon Audio Technical CD no 38C39-7147. Listening was done through AKG K240 Monitor headphones. Listening level and balance between channels A and B, as well as channel selection for listening was managed by a MIDI-controlled Yamaha DMP-7 digital audio mixer (see diagrams below).

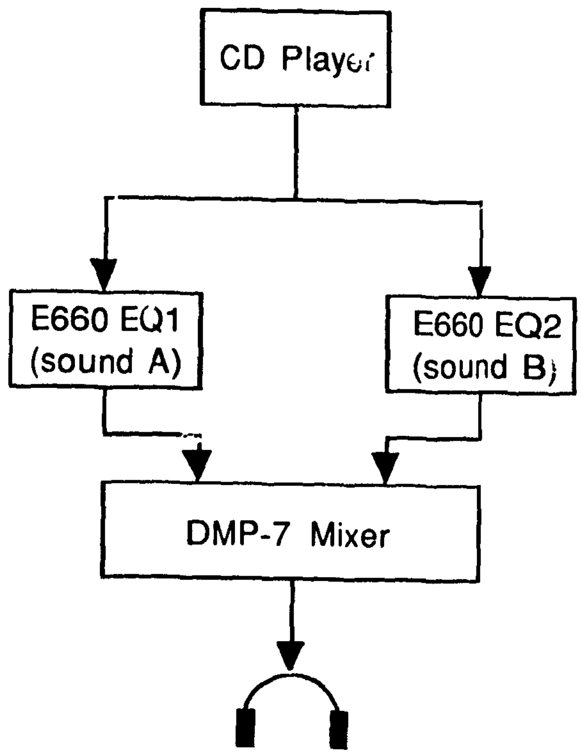

Fig. 1. Audio section of the technical ear training system

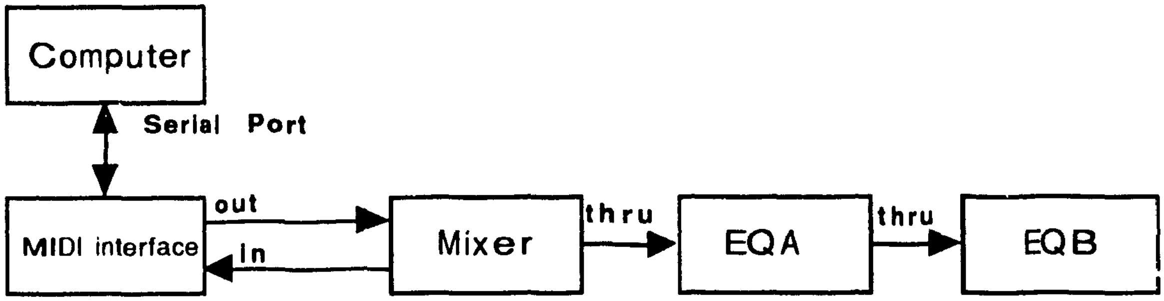

Fig. 2. MIDI section of the technical ear training system

4.3.2 Software

Custom software in C and Assembly was written by the author to control the presentation of exercises to the subjects and for the evaluation of the answers. The feedback provided by the computer was simple, indicating if the answer was right or wrong and allowing the subjects to try again. Elements of the user interface will be presented as the exercises are described. Following is an algorithm describing the basic steps involved in a typical Technical Ear Training session, from the user point of view:

```
repeat
{
     choose an exercise;
     choose a sound;
     get the first problem;
     repeat
     {
          compare sounds A and B;
          give an answer;
          get next problem;
     } until exercise is completed;
} until done;
```

4.3.3 System performance

It seems clear that the transmission speed of MIDI together with the use of a high-performance personal computer allow real-time sound processing of sound material. Some distortion elements were audible when the faders on the computer screen were moved very fast (and thus sending large quantities of MIDI data to the equalizers). However this seems to be more a hardware design problem than a software problem. The equalizers are not capable of processing the data quickly enough when the amount is too large.

4.3.4 Student records

Extensive records were kept on disk during the development period in order to collect data on subject's performance. For each work session at the computer, recorded information included date, starting and ending time, and the number of exercises completed during the session. For each exercise, the information included the amount of time spent on the exercise, and the number of problems that were done for that exercise. Finally, for each problem within a given exercise, stored data included: the maximum number of trials allowed, the number of trials necessary to achieve a right answer, response time (the time necessary to achieve a right answer), and the answer data for each trial (the data describing the right answer).

4.3.5 Exercises

Two main types of exercises were used: comparative listening exercises (CLE) and absolute identification exercises (AIE). They are each related to a test procedure method used in listening tests on loudspeakers. CLE uses the paired comparisons procedure and AIE uses the single judgement procedure. Both methods reflect situations frequently encountered in the recording studio. In CLE, the student is presented with two sounds A and B. Sound A has been spectrally modified by the computer but the modifications are unknown to the student. The task is to duplicate the timbre of sound A by adjusting on the computer screen the settings of the equalizer. The student can compare at will the reference sound (A), the sound he/she has to modify (B), and a flat, non-modified version of the sound. Three parameters of the sound can be manipulated with the equalizer for each of the frequency bands: center frequency, bandwidth of the modification (Q), and the amount of gain or cut applied to a frequency region. The difficulty of the exercise can be modified by varying the size of the catalog of stimuli, the number of parameters to be identified in one single exercise, the magnitude of the modifications, the frequency regions to which they are applied, etc.

In a variant of CLE, a spectrally modified sound A is presented, but this time, sound B is a duplicate of sound A. A flat version of the sound is also available. The task consists in modifying sound B parameters so as to remove the modifications that were applied to it. The possibility to listen to the flat version of the sound may or may not be available.

The frequency range was divided into 4 overlapping bands (30-960 Hz, 800-4,000 Hz, 4,000-12,000 Hz, and 10,000-20,000 Hz). For each band, 3 potentiometers could be used to modified center frequency, Q, and gain between -12 and +12 dB. Q values ranged between 0.3 and 9.9. The student could only see and manipulate the settings of sound B (the sound to be modified). When sound A was selected, the potentiometers became inactive. One digital equalizer was used for each sound (A and B). All controls relating to the equalizer section were gathered in the main window. A second window was used for monitoring the progress of the student during an exercise and for the selection of sounds to be used with each exercise. Total number of exercises, number done/to do, number of right/wrong answers as well as percentage, were displayed. The monitor window was used to allow the subjects to know where they are within an exercise at any given moment. Constant indication of performance level proved to be useful in order to provide incentive for the student to strive for best performance (see figure 3).

Comparative listening is constantly performed by the sound engineer in the recording studio. Typical situations include comparing timbre produced by two microphones, or two positions of the same microphone type, comparing two equalization settings applied to the same sound, or comparing a newly recorded sound with one already on the tape, possibly from an earlier recording session, and trying to match the timbre of both sounds, etc. In other occasions, undesirable resonances and antiresonances are present in a recording and the engineer must remove them by applying a boost or cut at the resonance frequency. The second variant of CLE is a replicate of this task. In the tasks described above, the engineer must be able to hear the resonances and timbre differences and also know the physical characteristics of such timbral irregularities.

For AIE (see figure 4), the student was presented with a spectrally modified sound. The task was to identify the center frequency to which the modification was applied. For that exercise, the sound could not be manipulated. Instead, the student had to select the answer from a menu of predetermined frequencies. The difficulty of the exercise could be varied by modifying the size of the catalog of stimuli and by providing or not the possibility to listen to a flat version of the sound for comparison. Absolute identification of timbre is a much more difficult task than comparative listening. The listener must compare the stimulus with an internal reference stored in long-term memory. Absolute identification of resonances allows a sound engineer to pinpoint with accuracy and efficiency spectral regions that need attention.

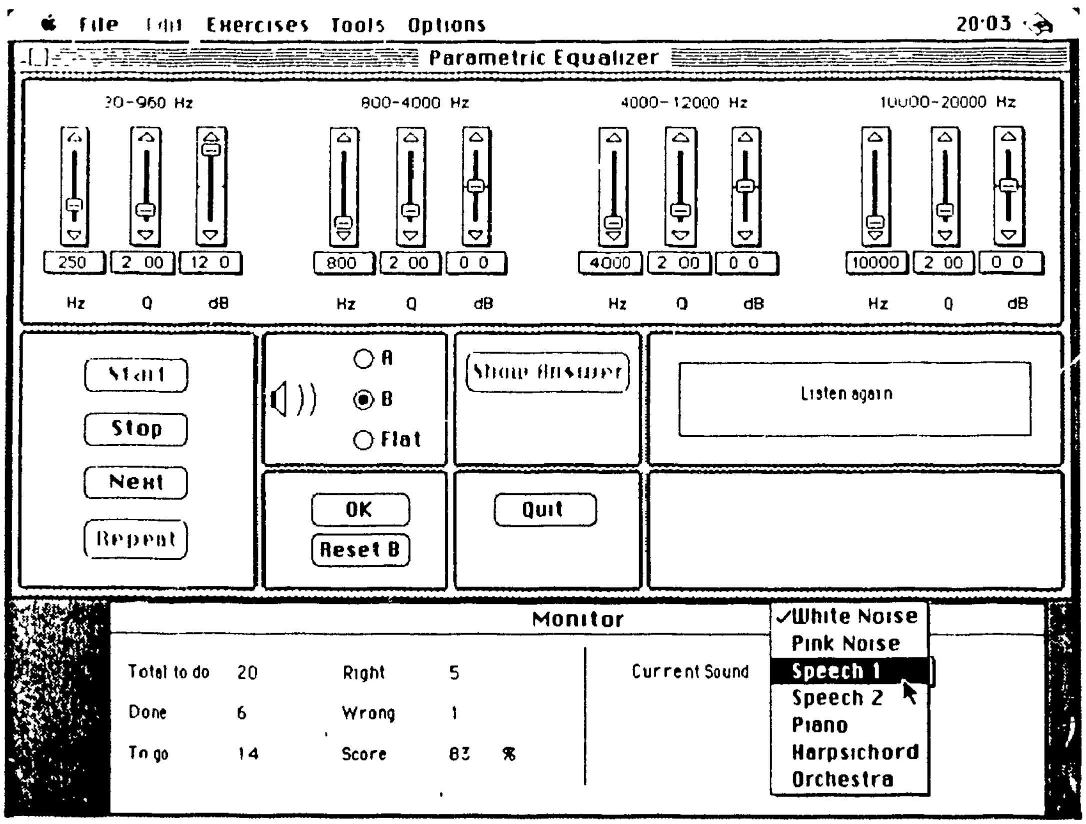

Fig. 3. User interface for comparative listening exercises

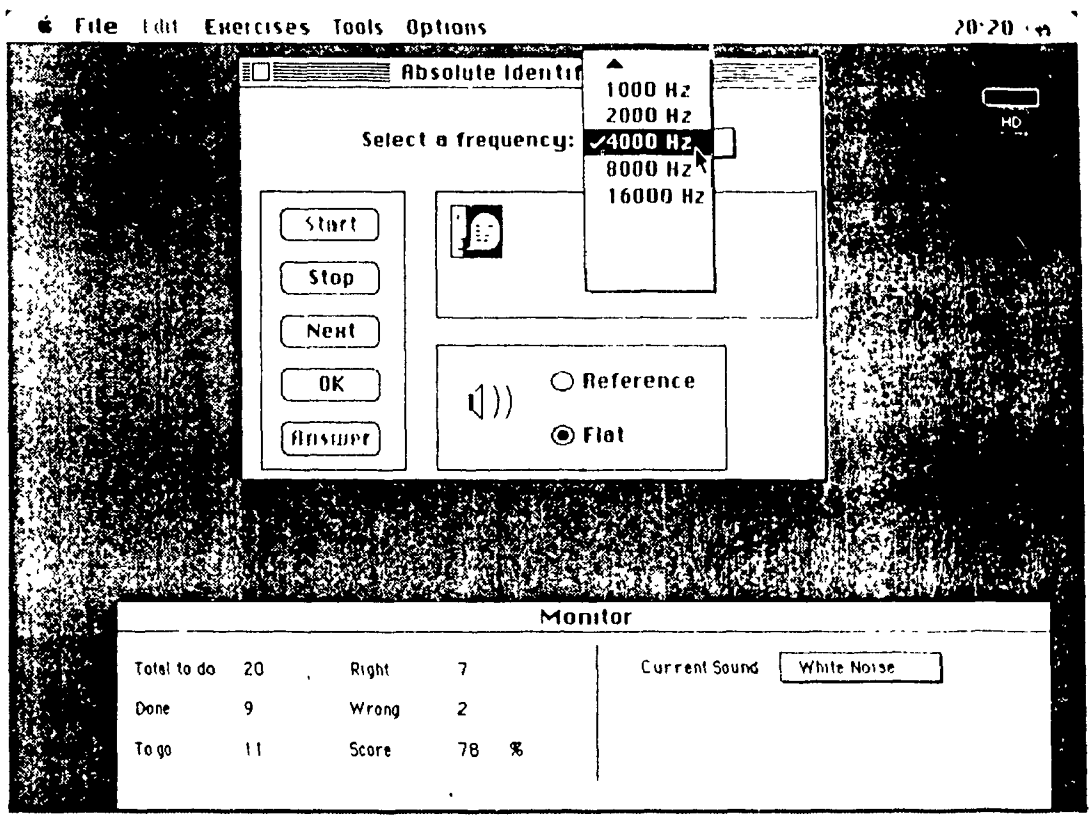

Fig. 4. User interface for absolute identification exercises

For both types of exercises, two modes of interaction could be used. In the "free mode", problems were presented at random and their number was not fixed. The student could explore at will the sounds with the tool at hand (computer) and practice her/his hearing skills. In the "fixed mode", exercises containing a fixed number of problems had to be completed before proceeding to the next exercise. Within each exercise, the problems were presented at random[38]. All exercises were pre-determined by the author: no computer-generated problems were used.

4.3.6 Preliminary data

As mentionned above, data on subjects performance were collected during program development. Some of these data will be presented here as a first indication of the performance and usefulness of the program. However, they should not be considered as final proofs of the validity of the system. The preliminary nature of the system prevented the collection and analysis of all data necessary for a fully thorough evaluation.

4.3.6.1 Procedure

Nine subjects used the program for a period of 6 months as it was being developed. They were all students in the Graduate program in Sound Recording at McGill University. Four of the subjects were in the second year of their program of study, and five in the first year. The testing stage was done in the context of the Technical Ear Training course for which all students received credits. Two periods of 45-60 minutes per week of practice time

---

[38]Here the term exercise is used to denote a collection of related problems.

were recommended for a total of 24 weeks. Factors such as schedule constraints resulted in a more irregular time allocation: session duration ranged between a few minutes to 2 hours. A first examination of the data indicates that most sessions for most subjects lasted between 30 and 60 minutes. All sessions were done individually without supervision by the instructor. The following section compares results obtained at the beginning and at the end of the 6-month period for one exercise in Comparative Listening and one exercise in Absolute Identification. Only the frequency parameter was varied. Q and gain were fixed at 2.0 and ±12 respectively. Modifications could be applied at 9 pre-determined center frequencies: 63, 125, 250, 500, 1000, 2000, 4000, 8000, and 16,000 Hz. A total of 20 problems per exercise were presented and within each exercise, two problems were repeated twice. The order of presentation within a given exercise was random. The sound used was a looped section of white noise.

4.3.6.2 Results and discussion

In figure 5, scores obtained by the nine students in the comparative listening exercise are plotted. The lower curve (black squares) represents the scores obtained at the beginning of the six-month period. The upper curve (empty squares) shows the scores for the same exercise at the end of the period. The values on the ordinate range between the lowest and highest scores obtained by the students to emphasize differences between the two sets of values. Two main observations can be made about the graph. First, performance (score value) improved for all students. In some cases (students 3, 5, 7, and 8), the difference is quite significant. Second, individual ability within the group varied substantially at the beginning whereas it was much more homogenous at the end, all students but one obtaining perfect scores. Students 1, 4, and 9 had already developed great ability for that particular aural task when they started the training so their progress was smaller in comparison with those students who started at a lower level.

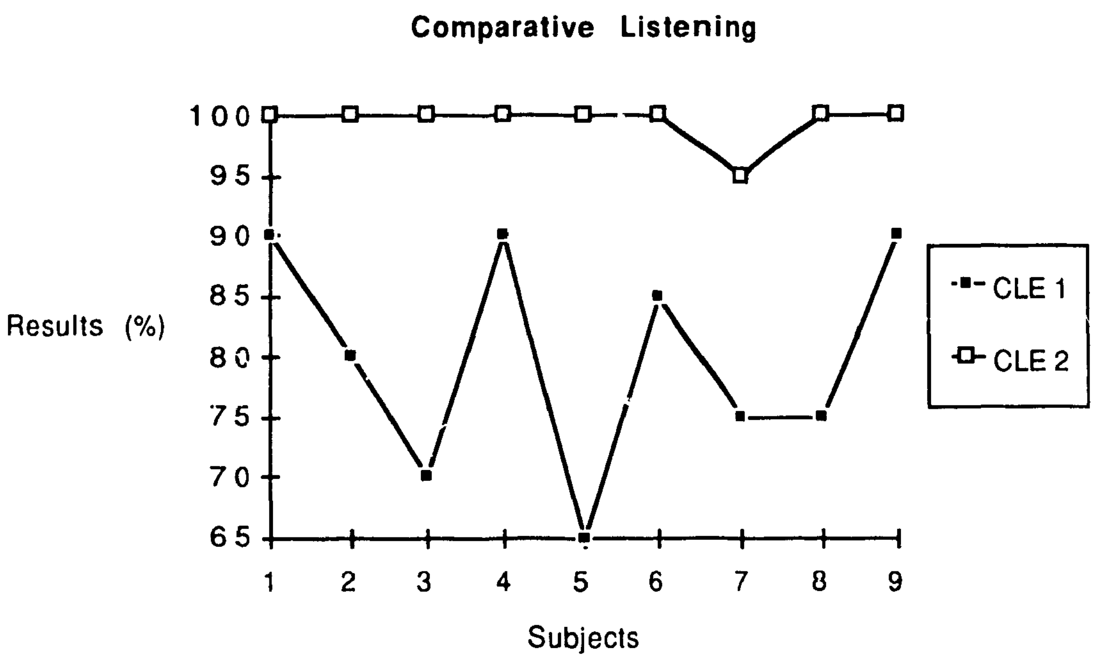

Fig. 5 Scores for CLE

highest scores obtained by the students to emphasize differences between the two sets of values. Two main observations can be made about the graph. First, performance (score value) improved for all students. In some cases (students 3, 5, 7, and 8), the difference is quite significant Second, individual ability within the group varied substantially at the beginning whereas it was much more homogenous at the end, all students but one obtaining perfect scores. Students 1, 4, and 9 had already developed great ability for that particular aural task when they started the training so their progress was smaller in comparison with those students who started at a lower level.

A significant aspect of performance for sound engineers is the time necessary to achieve the desired result. The time necessary to obtain the desired sound quality depends on the speed with which the engineer can modify appropriately the sound, how quickly decisions are made in the choice of center frequencies to alter, on the choice and placement of microphones, etc., which all depend on the speed with which the engineer can analyze and evaluate the sound and determine the modifications that need to be applied in order to obtain the desired timbre. This aspect was further examined for the comparative listening exercise. For each exercise, average response time for all problems was calculated. Results are presented in figure 6.

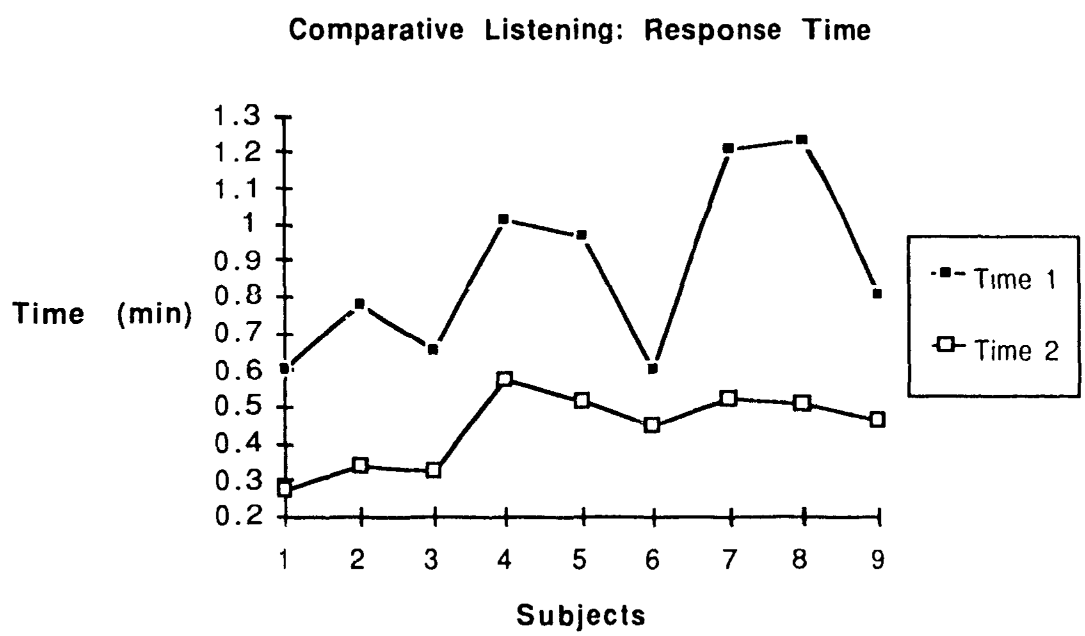

Fig. 6. Average response time

The upper curve (black squares) represents average response time at the beginning of the training and the lower curve (empty squares) is the response time achieved at the end of the training. A few observations can be made. First, all students responded more quickly at the end, i.e. the average time spent on each problem within the exercise was shorter at the end. Second, the difference was very significant. For most students, response time decreased by a factor greater than 2. Third, the homogenization effect of practice that was observed in the score data is not as apparent in the time data. Students who were faster at the beginning were still faster than others at the end. Although inter-individual differences were less marked at the end, the general shape of the curve is basically the same[39].

The data above must be used with caution. Even though students were already familiar with the system when they first did the exercise, it is possible that additional experience in the use of the computer and the program contributed to a certain extent to the quicker performance at the end. In that respect, one would think that there should be some relation between practice time and performance. Figure 7 shows a graph of total time spent by each student using the program during the 6-month period. One must realize that the total number of hours for most students is low, with an average of only 27.4 hours. There is no clear relation between practice time (experience) and performance. Perhaps practice time was too short to reveal any significant effects. The comparative listening exercise described here was a simple task and all students but one obtained a perfect score at the end. Data obtained from a more difficult task might reveal more inter-individual differences.

---

[39]The only exception is subject 8 who was slower than 7 at the beginning and then faster at the end.

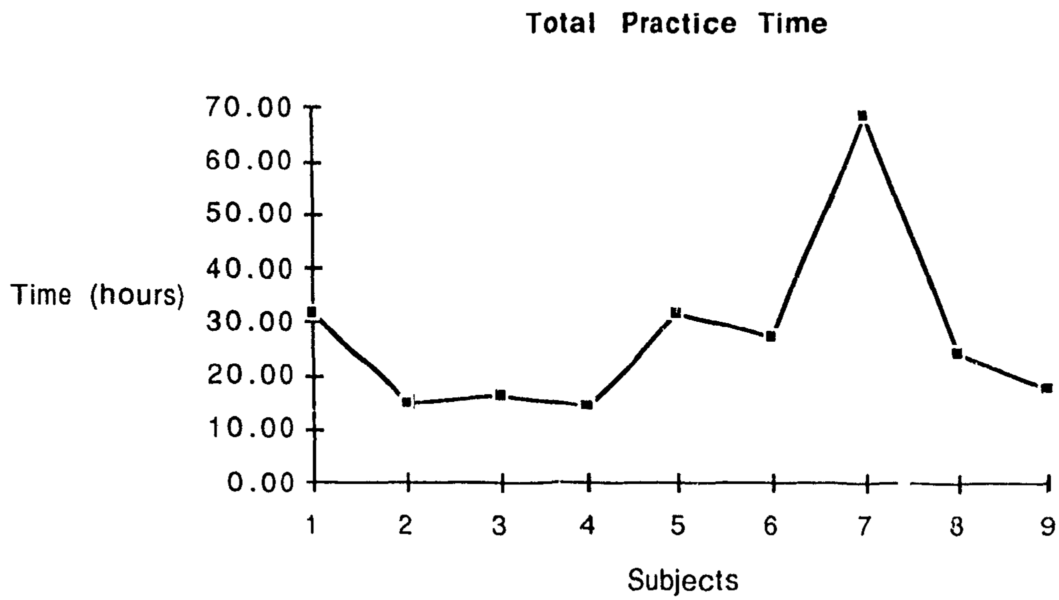

Fig. 7. Total practice time for each student for the 6-month period

Data obtained from the Absolute Identification exercise is presented in figure 8. This was a more difficult exercise. It calls on the use of stored internal references of timbres. Students couldn't modify the sound and explore different equalization settings. The students could however compare the equalized sound with a flat version of the same sound. Seven students only are represented since two members of the original group didn't do the exercise at the beginning of the training period, thus preventing comparison of performance. The graph indicates that students' abilities for that task remained unequal after the training period. Generally, the rank of students remained the same: students who were better than others at the beginning remained in the same relative position at the end. It seems that practice time was not long enough to attain asymptotic performance. It is also reasonable to say that asymptotic levels and the time necessary to attain them may vary between individuals[40]. This is not surprising since the aural task is quite difficult. All students improved although the magnitude of the improvement for students 3, 4, and 7 is small. There seems to be a relation between total practice time (on all types of exercises) and performance levels obtained for AIE. Students 3 and 4 practiced the least and they improved very little for AIE. Student 7 practiced the most and improved very little as well. The latter student had numerous sessions that were very long (around 2 hours). Listening fatigue might have hindered improvement. Other factors may also be responsible for the low performance of student 7. Errors in the use of the program, lower ability of the student, etc.. Some students might have lower perceptual abilitiy limits than others. Students 1, 5, and 6 all practiced around 30 hours for an average of 1.3 hours a week and improved best. The data suggest that there might be an optimum practice time value in order to maximize improvement. More experiments are needed to determine the time necessary to obtain asymptotic performance.

---

[40]See Jonhson et. al 1987 and Watson 1981.

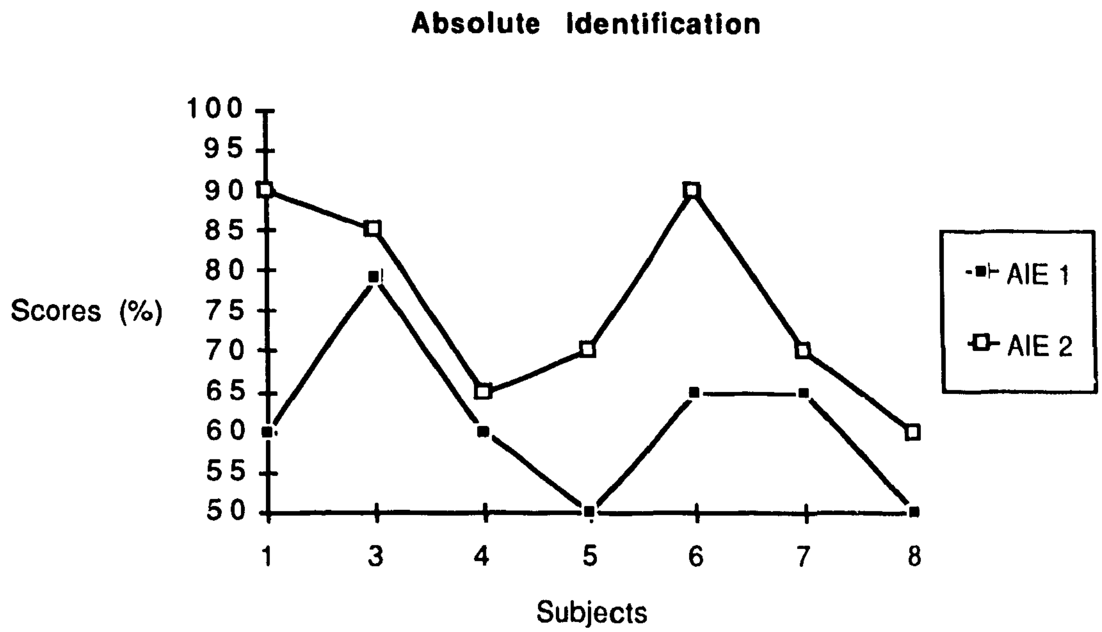

Fig. 8. Scores for AIE

4.3.6.3 Conclusion

This section presented a simple computer-controlled system for timbral ear training. The data presented here do not allow any final conclusions to be drawn about the validity of the system in its current state as a training/teaching tool for timbre perception. It does however give indications that such a system could be a useful tool for sound engineers to improve the perceptual abilities they need in their profession. The next chapter will discuss the limits of the current system and will present suggestions for further research.

CHAPTER 5

THE NEXT STEP

5.1 General conclusion

A lot of energy and time have been spent over the years in the audio industry to improve the performance of components of the recording/reproducing chain, from microphones to loudspeakers. Comparatively little attention has been paid to the last element of the chain: the listener. Although relatively few papers addressed the subject of timbre in sound engineering[41], it is quite clear that timbre perception acuity is an ability that sound engineers need in order to be successful in their profession.

Literature on timbre perception was reviewed. We know reasonably well what are the physical parameters of sound that affect the perception of timbre. What is less clearly understood is the relative perceptual significance of these factors in the complex listening situations that are typical in sound engineering. As a consequence, we still lack a complete, standardized, and acknowledged set of scales or dimensions along which timbre could completely be measured and evaluated. Even the word "timbre" itself still needs to be clearly defined.

Although much research remains to be done to fully understand what is timbre and the nature of all complex interactions that determine it, numerous studies in psychoacoustics and papers that investigated issues related to listening tests on loudspeakers have shown that similarly to the other elements of the audio chain, the listener's performance and

---

[41] See recent papers by Toole, Olive, Letowski (1985).

reliability can be improved. Previous work has indicated that individuals can attain very high levels of timbral perception acuity and memory when they are systematically trained (Letowski 1985). However, we still don't understand very well the mechanisms governing this learning process and how timbre is stored in long-term memory. Findings have started to emerge, mainly from research on profile analysis, but many aspects are still hidden in the dark. It is not certain that the results obtained in studies on profile analysis would be the same if the stimuli used were isolated notes and sequences of notes played by single and multiple musical instruments instead of synthesized multitone sounds. Other factors such as the time necessary to attain asymptotic levels of performance in spectral shape learning and the practice time necessary to maintain such levels when musical sounds are used need to be investigated. Factors impairing and aiding learning processes need to be identified.

This thesis proposed that a computer-assisted training system could help improve timbre perception acuity and memory of sound engineers. Criteria for the design of such a system were proposed. Custom software was developed and a simple system was assembled. The preliminary data presented in this thesis tend to support the hypothesis that a computer-assisted program for timbral ear training should help improve users' performance for particular aural tasks. Generally, students performed better at the end of the training period than at the beginning for specific exercises. The improvement was more significant for easier exercises than for more difficult ones. The total average practice time was short (around 27 hours) and it is not known if levels of performance would have continued to increase with time. Tests were informal and the results should be considered as encouraging indicators of the potential of such a system as a teaching tool for timbre perception skills.

However, a true measure of the validity of such a system is achieved only if improvements in the exercises can also be measured in the recording studio. Such a direct comparative measure can be difficult to obtain since experimental control would probably be impractical during recording sessions. However, if aural tasks can be successfully modeled at the computer so that the decisions to arrive at a solution to a given problem are the same than those used in the studio, then perhaps we can assume that improvements observed in timbral ear training will result in comparable improvements in the recording studio.

5.2 Limitations of the system and suggestions for further research

The use of the exploratory system presented in this thesis revealed a number of limitations and problems that will need to be addressed in order to design an effective system for timbral ear training. These limitations will be enumerated in this section and a number of solutions will be proposed as additional design criteria.

5.2.1 Exercises

The exercises were limited to the direct manipulation of the spectrum through the use of parametric equalizers. The gain and bandwidth were fixed and known to the students and only single center frequencies were manipulated. Exercises in which multiple modifications (multiple frequency bands and combinations of parameters) are applied to the sound should be built. Other exercises related to spectrum should be added: for example, low-, high-, and band-pass filtering as well as shelf-type spectrum equalization should be used in exercises where the task would be to identify cut-off frequencies and bandwidths.

Our perception of timbre is determined by other factors than direct manipulation of the spectrum. The set of exercises should be expanded to include these other factors: recognition of loudness, characteristics of spaciousness (reverberation, panning), various types of distortions, and evaluation of spectral modifications introduced by sound processing devices commonly used in the recording studio: compressors and limiters, artificial reverberation devices, flangers, etc.

5.2.2 Sound material

The choice of sound examples was limited to only a few choices. Some of the exercises were not always accompanied with appropriate sounds in terms of spectral content. As mentioned in chapter four, it is critical that exercises be accompanied with sounds specifically chosen for them.

5.2.3 Procedure

In the present system, the computer didn't control the presentation of the sound material. The required sounds were specified but their selection on the compact disk player was made by the student. The student had to specify in the program the sound used for a particular exercise so that the selected sound was included in the stored record. Presumably, the sound in the data file and the actual sound used for the exercise were the same but the lack of control might have introduced errors in some cases. Because the instructor is not present when the students are practicing, the whole system must be under computer control. The computer should also be able to detect and notify the student when the sound used doesn't correspond to the prescribed one.

For the comparative listening exercises, a fixed tolerance of 10 % was applied to the students' answers (value of center frequency). Because the critical bandwidth varies with center frequency, a more accurate tolerance function should be used. The equation provided by Zwislocki and Terhardt (1980) for the critical bandwidth or the more recent function used by Moore et al. (1990) for the equivalent rectangular bandwidth should be tested.

5.2.4 Courseware design

Several critical issues must be addressed in order to improve the system. What are the most effective ways of presenting the material to be learned? How should differences in timbre/sound quality be expressed and demonstrated using a mixture of sound, graphics, and text? How should students' answers be evaluated? What type of and how much control should the student have on the material and or the learning sequence? In the system used for this thesis, feedback provided by the computer was limited to messages displayed on the screen indicating if the answer was right or wrong. The interactions between the computer and the student must be expanded. The computer should be able to guide the student towards the resolution of a given problem, and different strategies should be suggested. What is needed is a model of an expert faced with the same problem: How does a professional engineer tackle a given task such as removing multiple resonances? It would be useful to define a collection of strategies adapted to different tasks and different contexts.

BIBLIOGRAPHY

Abeles, H. F. 1982. The effect of verbal descriptions on clarinet tone quality. In Proceedings of the Research Symposium on the Psychology and Acoustics of Music 1981, edited by P. E. Sink, 66-76. Lawrence, Kansas: University of Kansas.

American Standard Acoustical Terminology S1.1-1960. New York: American Standards Association, Inc., 1960.

Asch, S. E. 1958. Effects of group pressure upon the modification and distortion of judgments. In Readings in Social Psychology, edited by E. E. Maccoby, 174-183. T.M. Newcomb, and E. L. Hartley. New York: Bolt.

Atal, B. S. and M. R. Schroeder. 1962. Perception of coloration in filtered gaussian noise - Short-term spectral analysis by the ear. In Proceedings of the 4th International Congress on Acoustics. Copenhagen, Paper H31.

Bales, K. W. 1986. Computer-based instruction and music technology in education. Journal of Computer-Based Instruction 13 (1): 2-5.

Balzano, G. J. 1986. What are musical pitch and timbre?. Music Perception 3: 297-314.

Barlow, D. A. 1978. Loudspeaker coloration. Wireless World 84: 34-37.

Bartholomew, W.T. 1942. Acoustics of music. Englewood Cliffs: Prentice-Hall.

Bartlett, B. 1981. Tonal effects of close microphone placement. Presented at the 69th Convention of the Audio Engineering Society, Los Angeles, Preprint 1782.

Bartlett, B. 1983. Description of audio quality: A glossary. [Photocopied].

Bartlett, B. 1987. Introduction to professional recording techniques. Indianapolis: H.W. Sams.

Bauer, B. B. 1974. Audibility of phase distortion. Wireless World 80: 27-28.

Baxandall, P. J. 1977. Audible amplifier distortion is not a mystery. Wireless World 83: 63-66.

Beck, Søren. 1987. Listening tests on loudspeakers: An experimental study of nuisance variables and correlation between results of objective and subjective tests. The Acoustics Laboratory, Technical University of Denmark, Report No 43

Beck, Søren 1990. Listening tests on loudspeakers: A discussion of experimental procedures and evaluation of the response data. Paper presented at the 8th International AES Conference, The Sound of Audio, Washington D.C.

Benade, A. H. 1985. From instrument to ear in a room: Direct or via recording. Journal of the Audio Engineering Society 33: 218-233.

Berger, K. W. 1964. Some factors in the recognition of timbre. Journal of the Acoustical Society of America 36: 1888-1891.

Bernstein, L. R. and D. M. Green. 1987. The profile-analysis bandwidth. Journal of the Acoustical Society of America 81 (6): 1888-1895.

Bernstein, L. R. and D. M. Green. 1987. Detection of simple and complex changes of spectral shape. Journal of the Acoustical Society of America. 82 (5): 1587-1592.

Bindra, D., J. A. Williams, and J. S. Wise. 1965. Judgments of sameness and difference: Experiments on decision time . Science 150: 1625-1627.

Blauert, J. and P. Laws. 1978. Group delay distortions in electroacoustical systems. Journal of the Acoustical Society of America 63 (5): 1478-1483.

Blauert, J. 1987. Some fundamentals of auditory spaciousness. In A Symposium on Perception of Reproduced Sound, 33-40. Gammel Avernæs.

Bloothooft, G. and R. Plomp. 1988. The timbre of sung vowels. Journal of the Acoustical Society of America 84 (3): 847-869.

Børja, S. E. 1977. How to fool the ear and make bad recordings. Journal of the Audio Engineering Society 25: 482-490.

Bork, A. 1986. Learning with personal computers. New York : Harper & Row.

Bork, A. 1987. The potential for interactive technology. BYTE 12(2): 201-202, 204, 206.

Borwick, J. 1976. Sound recording practice. Oxford: Oxford University Press.

Borwick, J. 1980. If we can hear it, we can measure it. dB (July): 49-52.

Bregman, A. S. 1978. Auditory streaming: Competition among alternative organizations. Perception and Psychophysics 23: 391-398.

Bregman, A. S. and S. Pinker. 1978. Auditory streaming and the building of timbre. Canadian Journal of Psychology 22: 10-31.

Bregman, A. S. 1984. Auditory scene analysis. In Proceedings of the 7th International Conference on Pattern Recognition, 168-175. Silverspring, Maryland: IEEE Computer Society Press.

Bregman, A. S. 1990. Auditory scene analysis: The perceptual organization of sound. Cambridge and London: The MIT Press, Bradford Books.

Bryan M. E. and H. D. Parbrook. 1960. Just audible thresholds for harmonic distortion. Acustica 10: 87-91.

Bücklein, R. 1981. The audibility of frequency response irregularities. Journal of the Audio Engineering Society 29: 126-131.

Cabot, R. C. 1984. Perception of nonlinear distortion. Presented at the 2nd International Conference of the Audio Engineering Society. Anaheim, Preprint C1004

Cabot, Richard C. 1990. Audible effects vs. objective measurements in the electrical signal path. Paper presented at the 8th International AES Conference, Washington D.C

Campbell, R. A. and A. M. Small. 1963. Effect of practice and feedback on frequency discrimination. Journal of the Acoustical Society of America 35. 1511-1514.

Campbell, W. C. et al. 1978. The contribution of the legato transient to instrument identification. In Proceedings of the Research Symposium on the Psychology and Acoustics of Music 1978, 30-44 Edited by E. P. Asmus. Lawrence, Kansas: University of Kansas.

Charbonneau, Gérard R. 1981. Timbre and the perceptual effects of three types of data reduction. Computer Music Journal 5: 10-19.

Chopp, J. M. 1986. The computer: Integrating technology with education. Music Educators Journal (December): 22-25.

Chowning, J. M., J. M. Grey, J. A. Moorer, and L. Rush. 1982. Instrumental timbre and related acoustical phenomena in the perception of music. Stanford University, Department of Music Report No. STAN-M-11.

Clark, M. et al. 1963. Preliminary experiments on the aural significance of parts of tones of orchestral instruments and on choral tones. Journal of the Audio Engineering Society 11: 45-54.

Clark, M. and P. Milner. 1964. Dependency of timbre on the tonal loudness produced by musical instruments. Journal of the Audio Engineering Society 12: 28-31.

Clark, M., P. Robertson, and D. Luce. 1964.A preliminary experiment on the perceptual basis for musical instrument families. Journal of the Audio Engineering Society 12: 199-203.

Cleveland, T. F. 1977. Acoustic properties of voice timbre types and their influence on voice classifications. Journal of the Acoustical Society of America 61: 1622-1629.

Craig, J. H. and L. A. Jeffress. 1962. The effect of phase on the quality of the two-component tone. Journal of the Acoustical Society of America 34: 1752-1760.

de Bruijn, A. 1978. Timbre-classification of complex tones. Acustica 40: 108-114.

Deutsch, D. 1975. The organization of short-term memory for a single acoustic attribute. In Short-Term Memory, edited by D. Deutsch and J. A. Deutsch, 107-151. New York: Academic Press.

Dolson, M., A. Singer, and D. Rivas. 1987. Personal-computer microworlds for learning about signals and sound. ICMC Proceedings, 341-348.

Dworak, P. E. et al. 1978. The contribution of phase changes to differences in timbre. In Proceedings of the Research Symposium on the Psychology and Acoustics of Music 1978, edited by E. P. Asmus, 14-29. Lawrence, Kansas: University of Kansas.

Eddins, J. 1978. Random access audio in computer-assisted music instruction. Journal of Computer-Based Instruction 5 (1-2): 22-29.

Eddins, J. 1981. A brief history of computer-assisted instruction in music. College Music Symposium 21 (2):7-14.

Eisler, H. 1966. Measurement of perceived acoustic quality of sound-reproducing systems by means of factor analysis. Journal of the Acoustical Society of America 39: 484-492.

Everest, F. A. 1987. Instruction in critical listening. Presented at the 83rd Convention of the Audio Engineering Society, New York, Preprint 1890.

Farrar C. L. et al. 1987. Spectral-shape discrimination I: Results from normal-hearing listeners for stationary broadband noises. Journal of the Acoustical Society of America 81 (4): 1085-1092.

Fielder, L. 1982. Dynamic-range requirements for subjectively noise-free reproduction of music. Journal of the Audio Engineering Society 30 (7/8): 504-511.

Fletcher. H. 1934. Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure. Journal of the Acoustical Society of America 6: 59-69.

Fletcher, H. 1958. Hearing, the determining factor for high-fidelity transmission. Audio (July): 24, 26, 49; (August): 45-52, 63; (September): 34, 36, 53.

Fletcher, H. 1962. Quality of piano tones. Journal of the Acoustical Society of America 34: 749-761.

Fletcher, H. 1969. The ear as a measuring instrument. Journal of the Audio Engineering Society 17: 532-534.

Fletcher, H. 1978. Some experiments with the bass drum. Journal of the Acoustical Society of America 64 (6): 1570-1576.

Freed, D. J. and W. L. Martens. 1982. Deriving psychophysical relations for timbre. ICMC Proceedings, 393-415.

Fryer, D. A. 1975. Intermodulation distortion listening tests. Presented at the 50th Convention of the Audio Engineering Society, Collected Papers. Cited in R C. Cabot. Perception of nonlinear distortion. Presented at the 2nd International Conference of the Audio Engineering Society. Anaheim, Preprint C1004.

Gabrielsson, A., U. Rosenberg, and H. Sjögren. 1974. Judgments and dimension analyses of perceived sound quality of sound-reproducing systems. Journal of the Acoustical Society of America 55: 854-861.

Gabrielsson A. and H. Sjögren. 1979. Perceived sound quality of sound reproducing systems. Journal of the Acoustical Society of America 65: 1019-1033.

Gabrielsson, A. and B. Lindström 1985. Perceived sound quality of high-fidelity loudspeakers Journal of the Audio Engineering Society 33. 33-53.

Gabrielsson Alf, Björn Lindström , and Ove Till. 1986. Loudspeaker frequency response and perceived sound quality: measurements in listening room. Stockholm: Karolinska Institute, Technical Audiology Report No TA114.

Gabrielsson Alf, Björn Lindström , and Ove Till. 1987. Loudspeaker frequency response and perceived sound quality: Comparison between measurements in listening room, anechoic room and reverberation room. Stockholm: Karolinska Institute, Technical Audiology Report No TA115

Gabrielsson, A. 1987. Planning of listening tests. Listener and experimental variables In A Symposium on Perception of Reproduced Sound, 51-60 Gammel Avernæs.

Gabrielsson, A., B N. Schenkman, and B. Hagerman. 1988. The effects of different frequency responses on sound quality judgments and speech intelligibility Journal of Speech and Hearing Research 31 (June): 166-177.

Gabrielsson Alf et al. 1988. Perceived sound quality of reproductions with different frequency responses and sound levels. Stockholm: Karolinska Institute, Technical Audiology Report No TA117.

Gagné, Jean-Paul, and P. M. Zurck. 1988. Resonance-frequency discrimination. Journal of the Acoustical Society of America 83(6): 2293-2299.

Green, D. M. 1983. Profile analysis: A different view of auditory intensity discrimination American Psychologist 38: 133-42

Green, D. M. et al. 1983. Successive versus simultaneous comparison in auditory intensity discrimination. Journal of the Acoustical Society of America 73: 639-643

Green, D. M. and G. Kidd. 1983. Further studies of auditory profile analysis. Journal of the Acoustical Society of America 73· 1260-1265

Green, D. M. et al. 1984. Profile analysis: critical bands and duration. Journal of the Acoustical Society of America 75: 1163-1167.

Green, D. M. and C. R. Mason. 1985. Auditory profile analysis Frequency, phase, and Weber's law. Journal of the Acoustical Society of America 77: 1155-1161.

Green, D. M. et al. 1986. Auditory profile analysis of irregular sound spectra. Journal of the Acoustical Society of America 79: 1045-1053.

Green, D. M. et al. 1987. Frequency effects in profile analysis and detecting complex spectral changes. Journal of the Acoustical Society of America 81: 692-699.

Green, D. M. 1988. Profile analysis: Auditory intensity discrimination. New York: Oxford University Press.

Grey, J. M. 1975. An exploration of musical timbre. Ph.D. Thesis, Dept. of Music Report No. STAN-M-2, Stanford University.

Grey, J. M. 1977. Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America 61: 1270-1277

Grey, J. M. 1977. Perceptual evaluation of synthesized musical instrument tones. Journal of the Acoustical Society of America 62: 454-462.

Grey, J. M. 1978. Timbre discrimination in musical patterns Journal of the Acoustical Society of America 64: 467-472.

Grey, J. M. and J W. Gordon. 1978a. Perception of spectral modifications on orchestral instrument tones Computer Music Journal 2(1): 24-31

Grey, J. M. and J. W. Gordon. 1978b. Perceptual effects of spectral modifications on musical timbres. Journal of the Acoustical Society of America 63: 1493-1500.

Gross, D. and R. E. Foltz. 1981. Ideas on implementation and evaluation of a music CAI project. College Music Symposium 21 (2): 22-26.

Gross, D. 1984. Computer applications to music theory: A retrospective. Computer Music Journal 8 (4). 35-42.

Hall, D.E. 1980 Musical acoustics: An Introduction. Belmont: Wadsworth.

Handel Stephen 1989. Listening: An introduction to the perception of auditory events. Cambridge and London: The MIT Press, Bradford Books.

Hansen, V. 1987 Establishing a panel of listeners at Bang & Olufsen: A report. In A Symposium on Perception of Reproduced Sound, 89-98. Gammel Avernæs.

Hartmann, W. M. et al. 1986. Discrimination of spectral densities. Journal of the Acoustical Society of America 79 (6): 1915-1925.

Harwood, H D 1976 Audibility of phase effects in loudspeakers. Wireless World 82: 30-32

Hofstetter, F. 1981. Applications of the GUIDO system to aural skills research, 1975-1980 College Music Symposium 21 (2): 46-53.

Howard, J. H., and E. B. Silverman. 1976. A Multidimensional scaling analysis of 16 complex sounds. Perception and Psychophysics 19: 193-200.

Illényi, A. and P. Korpássy. 1981. Correlation between loudness and quality of stereophonic loudspeakers. Acustica 49: 334-336.

International MIDI Association. 1983. MIDI Musical Instrument Digital Interface Specification 1.0. North Holliwood: International MIDI Association.

Jacobs, J. E. and P. Wittman. 1964. Psychoacoustics, the determining factor in stereo Disk distortion. Journal of the Audio Engineering Society 12: 115-123.

Jones, B. L. 1987. Subjective quality assessment methods: The old international standards are changing. Presented at the 83rd Convention of the Audio Engineering Society, New York, Preprint 2551.

Johnson, D. M., C. S. Watson, and J. K. Jensen. 1987. Individual differences in auditory capabilities. Journal of the Acoustical Society of America 81 (2): 427-438.

Karjalainen M. A. 1986. Studies on the perception of distortion in the light of an auditory model. In Proceedings of the 12th International Congress on Acoustics, Paper B2-3 Toronto: ICA.

Kates, J. M. 1985. A Central spectrum model for the perception of coloration in filtered gaussian noise. Journal of the Acoustical Society of America 77 (4): 1529-1534

Keller, Arnold. 1987. When machines teach: Designing computer courseware New York: Harper & Row.

Kidd G. et al. 1986. Auditory profile analysis of irregular sound spectra Journal of the Acoustical Society of America 79: 1045-1053.

Kidd, G. and C. R. Mason. 1988. Evidence for sensory-trace comparisons in spectral shape discrimination. Journal of the Acoustical Society of America 84 (1): 144-149.

Killam, R. et al. 1981. Research applications in music CAI. College Music Symposium 21 (2):37-45.

Killam, R. 1984. An effective computer-assisted learning environment for aural skill development. Music Theory Spectrum 6: 52-62.

Killion Mead C. 1976. Noise of ears and microphones. Journal of the Acoustical Society of America 59 (2): 424-433.

Kirk, R. E. 1957. Learning, a major factor influencing preferences for high-fidelity reproducing systems. Journal of the Audio Engineering Society 5 238-241.

Kitamura, O. 1986. On the definition of tone color. In Proceedings of the 12th International Congress on Acoustics, Paper B2-6. Toronto: ICA.

Klein, W., R. Plomp, and M. C. W. Pols. 1970. Vowel spectra, vowel spaces, and vowel identification. Journal of the Acoustical Society of America 48: 999-1009.

Laroche Chantal and R. Hétu. 1988. Effets réversibles du bruit sur l'audition: recherche d'une mesure fidèle et plus sensible que le DTS (Décalage Temporaire des Seuils). Acustica 66 (September): 179-189.

Larson, R. J. and A. J. Adducci. 1961. Transient distortion in loudspeakers. IRE Transactions on Audio 9 (May/June): 79-85.

Lebel, C. J. 1947. Psycho-acoustical aspects of listener preference tests. Audio Engineering (August): 9-12, 44-48.

Leipp, E. 1976. Acoustique et musique. Paris: Masson.

Letowski, T. 1975. Difference limen for nonlinear distortion in sine signals and musical sounds. Acustica 34: 107-110.

Letowski, T. 1985. Development of technical skills: Timbre solfeggio. Journal of the Audio Engineering Society: 240-243.

Letowski, T. 1989. Sound quality assessment: Concepts and criteria. Presented at the 87th Convention of the Audio Engineering Society, New York, Preprint 2825.

Lichte, W. H. 1941. Attributes of complex tones. Journal of Experimental Psychology 28: 455-480.

Lipshitz, S. P. and J. Vanderkooy. 1981. The great debate: Subjective evaluation. Journal of the Audio Engineering Society 29: 482-490.

Lipshitz, S. P., M. Pollock , and J. Vanderkooy. 1982. On the audibility of midrange phase distortion in audio systems. Journal of the Audio Engineering Society 30: 580-595.

Luce, D. 1963. Physical correlates of nonpercussive musical instrument tones. Ph.D. Thesis, Massachusetts Institute of Technology.

Luce, D. and M. Clark. 1965. Durations of attack transients of nonpercussive orchestral instruments. Journal of the Audio Engineering Society 13: 194-199.

Luce, D. and M. Clark. 1967. Physical correlates of brass instrument tones. Journal of the Acoustical Society of America 42: 1232-1243.

Luce, R. D. 1972. What sort of measurement is psychophysical measurement ? American Psychologist (February): 96-106.

Lutfi Robert A. 1989. Informational processing of complex sound. I: Intensity discrimination. Journal of the Acoustical Society of America 86 (September): 934-944.

Martinez, C. and S. Gilman. 1988. Results of the 1986 AES audiometric survey. Journal of the Audio Engineering Society 36: 686-691.

Mathes, P. J. and R. L. Miller. 1947. Phase effects in monaural perception. Journal of the Audio Engineering Society 19:780-797.

McAdams, S. 1982. Spectral fusion and the creation of auditory images. In Music, Mind, and Brain. The Neuropsychology of Music, edited by Manfred Clynes, 279-298. New York: Plenum Press.

McAdams, S. 1988. Spectral fusion, spectral parsing and the formation of auditory images. Ph.D. Thesis, Stanford University, 1984. (Ann Arbor, Michigan: University Microfilms International).

McDermott, B. J. 1969. Multidimensional analyses of circuit quality judgments. Journal of the Acoustical Society of America 45: 774-781.

Meyer, J. 1978. Acoustics and the Performance of Music. Frankfurt/Main: Verlag Das Musikinstrument .

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63: 81-97.

Moir, J. 1976. Phase and sound quality. Wireless World 82: 32-38.

Moir, J. 1981. Just detectable distortion levels: Attempts to arrive at a practical criterion for assessing audio equipment. Wireless World 87: 32-38.

Moore, Brian C. J. 1989. An introduction to the psychology of hearing. London: Academic Press Inc.

Moore Brian C. J., Simon R. Oldfield, and Gary J. Dooley. 1989. Detection and discrimination of spectral peaks and notches at 1 and 8 kHz. Journal of the Acoustical Society of America 85 (February): 820-836.

Moore Brian C. J. et al. 1990. Auditory filter shapes at low center frequencies. Journal of the Acoustical Society of America 88 (1): 132-140.

Moore, F. R. 1982. The computer audio research laboratory at UCSD. Computer Music Journal 6 (1): 18-29.

Moorer, J. A., J. M. Grey, and J. Strawn. 1977a. Lexicon of analyzed tones. Part 1: A violin tone. Computer Music Journal 1 (2): 39-45.

Moorer, J. A., J. M. Grey, and J. Strawn. 1977b. Lexicon of analyzed tones. Part 2: Clarinet and oboe tones. Computer Music Journal 1: 12-29.

Moorer, J. A., J. M. Grey, and J. Strawn. 1978. Lexicon of analyzed tones. Part 3: The trumpet. Computer Music Journal 2: 23-31.

Moylan, W. 1987. A Systematic method for the aural analysis of sound sources in audio reproduction/reinforcement, communications, and musical contexts. Presented at the 83rd Convention of the Audio Engineering Society, New York, Preprint 2514.

Nakayama, T. et al. 1971. Subjective assessment of multichannel reproduction. Journal of the Audio Engineering Society 19: 744-751.

Norwich, Kenneth H. 1981. The magical number seven: Makin a "bit" of "sense". Perception and Psychophysics 29 (5): 409-422.

Olive, Sean E. and Floyd E. Toole. 1989. The detection of reflections in typical rooms. Journal of the Audio Engineering Society 37 (July/August): 539-553.

Olive, Sean E. 1990. The preservation of timbre: Microphones, loudspeakers, sound sources and acoustical spaces. Paper presented at the 8th International AES Conference, Washington D.C.

Olson, H. F. 1957. Stereophonic sound reproduction in the home. Presented at the 9th Convention of the Audio Engineering Society, Preprint no 26.

Opolko, F. J. 1982. A comparative method of evaluating recorded timbres. In The Combinative Application of Contact and Air Transducers on Selectcd Acoustical Instruments for Multi-Channel Recording. M.Mus. Thesis, McGill University.

Osgood, D. 1987. The difference in higher education. BYTE 12 (2): 165-178.

Padgham, C. 1986. The scaling of timbre of the pipe organ. Acustica 60: 189-204.

Papcun George, Jody Kreiman, and Anthony Davis. 1989. Long-term memory for unfamiliar voices. Journal of the Acoustical Society of America 85 (February): 913-925.

Patterson, J. H. and D. M. Green. 1970. Discrimination of transient signals having identical energy spectra. Journal of the Acoustical Society of America 48 (4): 894-905.

Peters, G. and J. Eddins. 1978. Applications of computers to music pedagogy, analysis, and research: A selected bibliography. Journal of Computer-Based Instruction 5 (1): 41-44.

Peters, G. D. 1981. Hardware development for computer-based instruction. College Music Symposium 21 (2): 15-21.

Piaget, J. 1964. Development and learning. Journal of Research in Science Teaching 2: 176-186.

Plomp, R. and A. M. Mimpen. 1968. The Ear as a frequency analyzer II. Journal of the Acoustical Society of America 43: 764-767.

Plomp, R. and H. J. Steeneken. 1969. Effect of phase on the timbre of complex tones. Journal of the Acoustical Society of America 46: 409-421.

Plomp, R. 1970. Timbre as a multidimensional attribute of complex tones. In Frequency Analysis and Periodicity Detection in Hearing, edited by R. Plomp and G. F. Smoorenburg. Leiden : A.W. Sijthoff.

Plomp, R. and H. J. M. Steeneken. 1973. Place dependence of timbre in reverberant sound fields. Acustica 28: 50-59.

Plomp, R. 1976. Aspects of tone sensation: A psychophysical study. London : Academic Press.

Plomp, R. 1987. A review of basic research on timbre. In A Symposium on Perception of Reproduced Sound, 13-22. Gammel Avernæs.

Pollack, Irwin. 1963. Discrimination of sharp spectral changes within broad-band noises. Journal of Auditory Research 3: 165-168.

Pollard H. F. and Jansson E. V. 1982a. A tristimulus method for the specification of musical timbre. Acustica 51: 162-171.

Pollard H. F. and Jansson E. V. 1982b. Analysis and assessment of musical starting transients. Acustica 51: 249-262.

Pols, L. C. W. 1970. Perceptual space of vowel-like sounds and its correlation with frequency spectrum. In Frequency Analysis and Periodicity Detection In Hearing, edited by R. Plomp and G .F. Smoorenburg. Leiden: H.W. Sijthoff.

Pols, L. C. W., L. J. T. van der Kamp, and R. Plomp. 1969. Perceptual and physical space of vowel sounds. Journal of the Acoustical Society of America 46: 458-467.

Poulsen, T. 1987. Application of psychoacoustic methods. In A Symposium on Perception of Reproduced Sound, 3-12. Gammel Avernæs.

Powell, R. 1986. The challenge of music software. BYTE (June) : 145-150.

Pramanik, S. K. 1987. Inadvertent bias in listening tests. In A Symposium on Perception of Reproduced Sound, 99-102. Gammel Avernæs.

Pratt, R. L. and P. E. Doak. 1976. A subjective rating scale for timbre. Journal of Sound and Vibration 45: 317-328.

Preis, D. 1984. Linear distortions, measurement methods and audible effects: A survey of existing knowledge. Presented at the 2nd International Conference of the Audio Engineering Society, Anaheim, Preprint C1005.

Prével, M. 1980. Low-cost, computer-assisted ear training. Journal of Computer-Based Instruction 6 (3): 77-78.

Raney Jill J. et al. 1989. Signal uncertainty and psychometric functions in profile analysis. Journal of the Acoustical Society of America 86 (September): 954-960.

Rasch, R. A. and R. Plomp. 1982. The perception of musical tones. In The psychology of music, edited by D. Deutsch. London: Academic Press.

Risset, J.-C. and M. V. Mathews. 1969. Analysis of musical instrument tones. Physics Today 22: 23-30.

Risset, J.-C. 1977. Musical acoustics. I.R.C.A.M.

Risset, J.-C. 1978. Hauteur et timbre des sons. I.R.C.A.M.

Risset J.-C. and D. L. Wessel. 1982. Exploration of timbre by analysis and synthesis. In The psychology of music, edited by D. Deutsch. London: Academic Press.

Roads, C. 1980. Artificial intelligence in music. Computer Music Journal 4 (2): 13-25.

Roads, C. 1983. A report on SPIRE: An interactive audio processing environment. Computer Music Journal 7 (2): 70-74.

Roederer, J. G. 1979. Introduction to the physics and psychophysics of music. New York: Springer-Verlag.

Saldanha, E. L. and J. F. Corso. 1964. Timbre cues and the identification of musical instruments. Journal of the Acoustical Society of America 36: 2021-2026.

Salmon, V. 1950a. Imagery for describing reproduced sound. Audio Engineering 36 (8): 14-15, 28.

Salmon, V. 1950b. Imagery for describing reproduced sound II". Audio Engineering 36 (9): 14, 29-35.

Sandusky, A. 1974. Memory processes and judgment. In Handbook of perception. Vol. 2, Psychophysical judgment and measurement. Edited by E. C. Carterette and M.P. Friedman. New York: Academic Press.

Schouten, J. F. 1968. The perception of timbre. In Proceedings of the 6th International Congress on Acoustics, paper GP-6-2, 35-44. Tokyo: ICA.

Seashore, C. E. 1938. Psychology of music. New York: McGraw-Hill.

Shailer M. J. et al. 1990. Auditory filter shapes at 8 and 10 kHz. Journal of the Acoustical Society of America 88 (1): 141-148.

Shrader, D. L. 1981. Microcomputer-based teaching. College Music Symposium 21 (2): 27-36.

Siegel William. 1972. Memory effects in the method of absolute judgment. Journal of Experimental Psychology 94 (2): 121-131.

Slawson, A. W. 1968a. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. Journal of the Acoustical Society of America 43: 87-101.

Slawson, A. W. 1968b. Phonetic notation for timbre in computer synthesized music. In Proceedings of the 6th International Congress on Acoustics, 29-32, Paper N-2-8. Tokyo: ICA.

Slawson, W. 1981. The color of sound: A theoretical study in musical timbre. Music Theory Spectrum 3: 132-141.

Slawson, W. 1985. Sound color. Berkely: University of California Press.

Smurzynski, J. 1986. New approach to the rise time differential sensitivity. In Proceedings of the 12th International Congress on Acoustics, paper B1-7. Toronto: ICA.

Sneegas, J. E. 1987. Aural acuity and the meaning of sound quality: A cultural approach. Presented at the 83rd Convention of the Audio Engineering Society, New York, Preprint 2489.

Solor..on, L. N. 1958. Semantic approach to the perception of complex sounds. Journal of the Acoustical Society of America 30: 421-425.

Somerville, T. 1954. The establishment of quality standards by subjective assessment. Acustica 4: 48-50.

Staffeldt, H. 1974. Correlation between subjective and objective data for quality loudspeakers. Journal of the Audio Engineering Society 22: 402-415.

Staffeldt, H. and E. Rasmussen. 1982. The subjectively-perceived frequency response in small and medium-sized rooms. SMPTE Journal, (July): 638-643.

Staffeldt, Henrik. 1984. Measurement and prediction of the timbre of sound reproduction. Journal of the Audio Engineering Society, 32 (6): 410-414.

Staffeldt, H. 1987. Evaluation and scaling of timbre in listening tests on loudspeakers. In A Symposium on Perception of Reproduced Sound, 23-32. Gammel Avernæs.

Strawn, John. 1982. Research on timbre and musical contexts at CCRMA. In ICMC Proceedings.

Strawn John M. 1985. Modeling musical transitions. Ph.D. Thesis, Center for Computer Research in Music and Acoustics, Department of Music Report No STAN-M-26, Stanford University.

Strawn, John. 1986. Orchestral instruments: Analysis of performed transitions. Journal of the Audio Engineering Society 34: 867-879.

Strawn, John. 1987a. Analysis and synthesis of musical transitions using the discrete short-time Fourier transform. Journal of the Audio Engineering Society 35: 3-13.

Strawn, John. 1987b. Editing time-varying spectra. Journal of the Audio Engineering Society 35: 337-352.

Strong, W. and M. Clark. 1967a. Synthesis of wind instrument tones. Journal of the Acoustical Society of America 41: 39-52.

Strong, W. and M. Clark. 1967b. Perturbations of synthetic orchestral wind instrument tones. Journal of the Acoustical Society of America 41: 277-285.

Summerfield Q. et al. 1987. Auditory enhancement of changes in spectral amplitude. Journal of the Acoustical Society of America 81 (3): 700-708.

Suzuki, H., S. Morita, and T. Shindo. 1980. On the perception of phase distortion. Journal of the Audio Engineering Society 28: 570-574.

Swets, J. A. et al. 1962. Learning to identify nonverbal sounds: An application of a computer as a teaching machine. Journal of the Acoustical Society of America 34: 928-935.

Terhardt, E. 1978. Psychoacoustic evaluation of musical sounds. Perception and Psychophysics 23: 483-492.

Toole, F. E. 1973. Loudness – Applications and implications to audio, part 2". dB (June): 25-28.

Toole, F. E. 1979. High fidelity in the control room – Why not ? dB (February): 30-33.

Toole, F. E. 1982. Listening tests: Turning opinion into fact. Journal of the Audio Engineering Society 30: 431-442.

Toole F. E. 1984. The acoustics and psychoacoustics of headphones. Presented at the 2nd AES International Conference, Anaheim, Preprint C1006.

Toole, F. E. 1985. Subjective measurements of loudspeaker sound quality and listener performance. Journal of the Audio Engineering Society 33: 20-32.

Toole, F. E. 1986. Loudspeaker measurements and their relationship to listener preferences. Journal of the Audio Engineering Society 34: 227-235, 323-348.

Toole, F. E. and S. E. Olive. 1986. The perception of sound coloration due to resonances in loudspeakers and other audio components. Presented at the 81st Convention of the Audio Engineering Society, Los Angeles, Preprint 2406.

Toole, F. E. 1987a. Planning of the listening test – Technical and environmental variables. In A Symposium on Perception of Reproduced Sound, 41-50. Gammel Avernæs.

Toole, F. E. 1987b. Correlation between the results of objective and subjective tests. In A Symposium on Perception of Reproduced Sound, 119-124. Gammel Avernæs.

Toole, F. E. and S. E. Olive. 1988. The modification of timbre by resonances: Perception and measurements. Journal of the Audio Engineering Society 36: 122-142.

Toole, Floyd E. 1990a. Loudspeakers and rooms for stereophonic reproduction. Paper presented at the 8th International AES Conference, Washington D.C.

Toole, Floyd E. 1990b. Identifying and controlling the variables. Paper presented at the 8th International AES Conference, Washington D.C.

Trahiotis, C. 1983. Progress and pitfalls associated with scientific measures of auditory acuity. In Collected Papers from the AES Premiere Conference, Rye, New York, June 1982, 63-66 . New York: Audio Engineering Society.

Ueda, Kazuo and Akagi Masato. 1990. Sharpness and amplitude envelopes of broadband noise. Journal of the Acoustical Society of America 87 (February): 814-819.

Vangellow, A. 1987. MIDI controlled effects devices. Electronic Musician (January): 33-34.

Vaughan, D. 1982. The Perception and appreciation of richness and timbre in rooms for music. Applied Acoustics 15: 289-306.

Viemeister, Neal F. 1990. An overview of psychoacoustics and auditory perception. Paper presented at the 8th International AES Conference, Washington D.C.

von Bismarck, G. 1974a. Timbre of steady sounds: A factorial investigation of its verbal attributes. Acustica 30: 146-159.

von Bismarck, G. 1974b. Sharpness as an attribute of the timbre of steady sounds. Acustica 30: 159-172.

von Helmholtz, H. 1954. On the sensations of tone as a physiological basis for the theory of music. Second English edition; Translated by A. J. Ellis, 1885; reprint: New York: Dover (page references are to reprint edition).

Wagenaars, W. M. et al. 1986. Subjective evaluation of dynamic compression in music. Journal of the Audio Engineering Society 34: 10-17.

Walker, D. and R. Hess. 1984. Instructional software: principles and perspectives for design and use. Belmont: Wadsworth.

Wang, C. C. and B. Gossick. 1978. Relation between timbre perception and tonal duration. In Proceedings of the Research Symposium on the Psychology and Acoustics of Music 1978, edited by E.P. Asmus, 1-13. Lawrence, Kansas: University of Kansas.

Wapnick, J. and P. Freeman. 1980. Effects of dark-bright timbral variation on the perception of flatness and sharpness. Journal of Research in Music Education 28 (3): 176-184.

Watson, Charles S. 1980. Time course of auditory perceptual learning. Ann. Otol. Rhinol. Laryngol 89: Suppl.74, 96-102.

Watson, C. S. and D. C. Foyle. 1985. Central factors in the discrimination of complex sounds. Journal of the Acoustical Society of America 78: 375-380.

Webster, J. C. 1970. Perceptual constancy in complex sound identification. British Journal of Psychology 61: 481-489.

Wessel, D. V. 1979. Timbre space as a musical control structure. Computer Music Journal 3 (2): 45-52.

Williams, M. 1987. Choice of programme material for critical listening to loudspeakers. In A Symposium on Perception of Reproduced Sound 81-87. Gammel Avernæs.

Wilson, Geoffrey L. 1978. Trends in audio education – A symposium. Presented at the 60th Convention, of the Audio Engineering Society, Los Angeles, Preprint 1379.

Winckel, F. 1967. Music, sound, and sensation. New York: Dover.

Wittlich Gary E., John W. Schaffer, and Larry R. Babb. 1986. Microcomputers and music. Englewood Cliffs: Prentice-Hall.

Wood, R. W. and P. J. Clements. 1986. Systematic evaluation strategies for computer-based music instruction systems. Journal of Computer-Based Instruction 13 (1): 17-24.

Woszczyk, Wieslaw. 1978. Improved instrument timbre through microphone placement. Recording Engineer/Producer 10 (5): 78-95.

Woszczyk, Wieslaw. 1979. Multimicrophone pickup of solitary acoustical instruments for single-channel transmission. Presented at the 63rd Convention of the Audio Engineering Society, Los Angeles, Preprint 1491.

Woszczyk, W. R. and F. E. Toole. 1983. A subjective comparison of five analog and digital tape recorders. Presented at the 74th Convention of the Audio Engineering Society, New York, Preprint 2033.

Whyte, D. 1977. Perceptible levels of audio-frequency tones in the presence of programme. BBC Research Department Report No RD 1977/31. Cited in R.C. Cabot. Perception of nonlinear distortion. Presented at the 2nd International Conference of the Audio Engineering Society. Anaheim, Preprint C1004.

Wigan, E. R. 1961. New distortion criterion. Electronic Technology 38 (4): 128-137; (5): 163-174.

Yankovskii, B. A. 1976. Methods for the appraisal of violin tone quality. In Musical Acoustics, Part 2, Violin family functions, 231-245. Edited by C.M. Hutchins. Stroudsburg: Dowder, Hutchinson, & Ross, Inc.

Yavelow, C. 1986. MIDI and the Apple Macintosh. Computer Music Journal 11 (4): 11-47.

Yavelow, C. 1987. Personal computers and music: The state of the art. Journal of the Audio Engineering Society 35: 160-193.

Zwicker, E. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). Journal of the Acoustical Society of America 33 (2): 248.

Zwicker E. and E. Terhardt. 1980. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. Journal of the Acoustical Society of America 68 (5): 1523-1525.

Zwislocki, J. et al. 1958. On the effect of practice and motivation on the threshold of audibility. Journal of the Acoustical Society of America 30: 254-262.