Order Number 9213551

Concurrent timbres in orchestration: A perceptual study of factors determining "blend"

Sandell, Gregory John, Ph.D.

Northwestern University, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

NORTHWESTERN UNIVERSITY

Concurrent Timbres in Orchestration: A Perceptual Study of Factors Determining "Blend"

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Music Theory

By

Gregory John Sandell , G. J., Ph. D.

EVANSTON, ILLINOIS

December 1991

Abstract

Concurrent Timbres in Orchestration: A Perceptual Study of Factors Determining "Blend"

Gregory John Sandell

Orchestration often involves selecting instruments for concurrent presentation, as in melodic doubling or chords. One evaluation of the aural outcome of such choices is along the continuum of "blend": whether the instruments fuse into a single composite timbre, segregate into distinct timbral entities, or fall somewhere in between the two extremes. This study investigates, through perceptual experimentation, the acoustical correlates of blend for 15 natural-sounding orchestral instruments presented in concurrently-sounding pairs (e.g. flute-cello, trumpet-oboe, etc.).

Ratings of blend showed primary effects for centroid (the location of the midpoint of the spectral energy distribution) and duration of the onset for the tones. Lower average values of both centroid and onset duration for a pair of tones led to increased blends, as did closeness in value for the two factors. Blend decreased (instruments segregated) with higher average values or increased difference in value for the two factors. The musical interval of presentation slightly affected the relative importance of these two mechanisms, with unison intervals determined more by lower average centroid, and minor thirds determined more by closeness in centroid. The contribution of onset in general was slightly more pronounced in the unison conditions than in the minor third condition. Additional factors contributing to blend were correlation of amplitude and centroid envelopes (blend increased as temporal patterns rose and fell in synchrony) and similarity in the overall amount of fundamental frequency perturbation (decreased blend with increasing jitter from both tones).

To confirm the importance of centroid as an independent factor determining blend, pairs of tones including instruments with artificially changed centroids were rated for blend. Judgments for several versions of the same instrument pair showed that blend decreased as the altered instrument increased in centroid, corroborating the earlier experiments. Other factors manipulated were amplitude level and the degree of inharmonicity.

A survey of orchestration manuals showed many illustrations of "blending" combinations of instruments that were consistent with the results of these experiments. This study's acoustically-based guidelines for blend augment instance-based methods of traditional orchestration teaching, providing underlying abstractions helpful for evaluating the blend of arbitrary combinations of instruments.

Acknowledgements

I would like express my appreciation and thanks to a number of people who have contributed in one way or another towards this project.

To my principle adviser, Gary Kendall, who suggested this line of research, patiently read numerous versions of the manuscript, and offered invaluable guidance, suggestions and ideas throughout its duration. To my other advisors, Alan Stout and Tom Carrell, who followed the project with interest, provided many helpful suggestions, and lent their valuable perspectives to the work.

To other teachers I have learned from during my odyssey through various schools and universities: at the Eastman School of Music, Robert Morris, who introduced me to new ways at looking at music theory, and Edwin Hantz, whose graduate seminar in music cognition in 1984 stimulated my interest in the subject; to Carol Krumhansl at Cornell, for broadening my understanding of research methodology and the work of other researchers in music perception; to Lois Elliott at Northwestern, for her valuable classes in psychoacoustics.

To various colleagues for the suggestions for research, ideas, and assistance with technical problems: to Bill Martens for advice in experimentation and graphic visualization of data; to Hans Pelleboer, for many interesting conversations about timbre and comradeship during long hours in the studio; and to David Wessel, for helping me acquire the stimuli used in this study, and for many helpful comments and suggestions.

For financial assistance during my years at Northwestern: the Graduate School, for a University Fellowship and a Dissertation Year Grant; and to Susan Lee and the Committee on Interdisciplinary Research in the Arts (CIRA) for a University Fellowship and a CIRA Research Fellowship.

To Northwestern's Institute for the Learning Sciences, whose understanding of my situation as self-supporting graduate student was more generous than one usually expects from an employer. Besides having the opportunity to do very interesting work for the last two years, I benefitted significantly from their computing facilities, for which I am most grateful. Thanks especially to Roger Schank, Chris Riesbeck, Alex Kass, and Eli Blevis.

To people who provided, supported, or maintained the various facilities that were necessary for this research: to Gary Kendall, for making Northwestern Computer Music a place where this kind of work can take place; to Amnon Wolman who also made generous efforts to insure the maintenance of the equipment I depended on, often solely for my benefit, and for the valuable use of his office and computer in the late stages of writing; to Frank Horowitz for repairing my Macintosh; to Dirk Ruiz and Peter Roll who helped with getting graphics software updated; and to various other people who helped keep large computing environments working: Rich Karstens, Jim McCoy, Christopher Chen, and Marty Wilde.

To my wife Elena, without whose support it is hard to imagine having ever completed this project. With her nurturing love, constant affirmation of what I was doing, and faith in my ability to finish, she contributed immeasurably to this work. Thanks also to the love and support of my dear friends Milton Stern and Jonathan Rothe.

To my family: my brother, and my Mother and Father. Sadly, while I was at work on the very latest stages of writing, Dad passed away unexpectedly. Among the flood of thoughts surrounding that event I recalled the memory of a lecture he took me to in the mid-1970's that was instrumental in developing the interests that led to this project: a presentation at Cal Tech on the digital recording and analysis-by-synthesis of a violin. In memory of this and the many other influences that helped shape my interests and character, this thesis is dedicated to Martin L. Sandell.

Table of Contents

List of Figures

List of Tables

List of Formulae

Introduction

> By musical theory I mean nothing very profound and nothing at all new: I refer to the familiar disciplines of harmony, counterpoint, analysis, and orchestration. (Cone, 1989, p. 29)

> While orchestration books abound, even the best of them serve up little more than a few tricks-of-the-trade illustrated by examples. There are no general statements that are nearly as persuasive as the simplest generalizations about counterpoint or chord progression found in the most elementary texts. (Slawson, 1978, p. 106)

The inclusion of orchestration within music theory, while an attractive prospect for the future, is a dubious proposition at present. As Slawson notes, few abstractions, paradigms, or generalizations govern orchestration's teaching or its practice; strategies for balancing chords and reinforcing melodies, for example, are learned largely in terms of specific instances, and seldom by rule. Furthermore, orchestration teaching involves nothing so generative as triadic chord construction or harmonic inversion; there are no standard resources such as major and minor scales; no pedagogically-graded exercises such as species counterpoint; and no orchestration-specific procedures equivalent to invention, fugue, or sonata form, to facilitate the comparison of student exercises to masterpieces.

This poses two problems to musicians. First, orchestration is hard to learn. Since its knowledge is acquired by imitation of examples, without access to a sight-reading orchestra or an apprenticeship to a "master" orchestrator, the prospect of developing one's own instinct for orchestral combinations is remote. Second, lacking robust generalizations and principles in its pedagogy, orchestration provides few techniques through which an analyst can assess the orchestration of a work and communicate it to others--which in turn further hinders the advancement of orchestration pedagogy.

What orchestration manuals tend to offer in place of generalizations and paradigms is a prescriptive, instance-based approach to learning. The passage below on melodic doubling illustrates advice typical of this approach:

> In *piano,* octave doublings are of good effect between trumpet and flute, horn and flute, oboe or clarinet. . . . Horns make effective unisons with clarinets and bassoons. Woodwind unisons with trombones are not often useful. The tuba combines quite well with bass clarinet and contrabassoon. (Piston, 1955, p. 427)

The first problem with such an approach is its short-sightedness: at best, prescriptions can only teach a student to repeat a solution for a particular texture. Schillinger (1941), remarking on such prescriptions, observed "If the composer is fortunate enough to apply them to just such a texture, he may meet with success. In all other cases he is bound to be a failure" (p. 1603). The prescriptions would be far more valuable, however, if accompanied by underlying abstractions from which to draw similar conclusions about other instrument pairings, rather than standing alone as rules to be memorized by rote. Second, the conditions under which the prescriptions hold true are under-specified. The registers of the instruments, and the articulations and durations of the notes being played can have a major impact on the sound of any particular pairing. The reference to the *piano* dynamic is helpful, but even this observation would be more valuable if described in more general terms; for example, the *effect* yielded by the *piano* dynamic should be clarified. Finally, and the most severe problem of all, the criteria "effective," "useful," and "combines well" are undefined here, and are poorly defined at best in any orchestration manual. Such vagueness no doubt contributes to the common perception of the Art of Orchestration as an arcane craft--which in turn explains why few musicians gain orchestration skill.

Recently, the psychologist Albert Bregman echoed some of these complaints, and suggested a solution:

> Although some parts of the crafts of counterpoint, orchestration, and construction of melodies are often codified into explicit rules, others are not, and the composer must learn through listening to music and through trial and error. A scientist would like to see all the principles written down. Many musicians assert that any principles that could be adduced, such as rules of counterpoint, would apply only to a particular form or style of music. As a scientist, I want to believe that this is not so, and that while many such rules will be specific to a style, if we dig deep enough, we will find universal principles of perception that can be converted to knowledge about orchestration. . . . Such a description would not be prescriptive, telling the composer what to do in the manner of traditional academic schools, but would

> simply say, "If you want this effect, do that, and here is the reason that it works." (Bregman, 1990, p. 458)

This study proposes that the best way to address the dilemma of theory and practice in orchestration is through rule-based approaches to evaluating instrumental combinations along the lines that Bregman suggests. To make this feasible--indeed, many musicians doubt that any "rule of orchestration" could predict how they would hear a given combination, or aid them in actualizing a timbral conception--the scope must be carefully defined. The plan here will be to (a) identify a criterion of timbral judgment common to a variety of orchestral applications, (b) define a continuum on which that judgment is made, (c) identify the general conditions that lead to various judgments along this continuum, and (d) identify which instruments or combinations of instruments meet those conditions. Such a "metric of orchestration" would go beyond the traditional prescriptions and provide a resource for solving problems in a variety of situations.

Orchestration manuals can be useful in providing hints for candidate topics of study, and clues to criteria for evaluating orchestral choices; however, one must look beyond their methods to satisfy the goals of this study. Acoustical analysis of musical instruments and empirical study of timbre perception will be necessary to identify the mechanisms underlying the criteria for evaluation and the relevant parameters of timbral variation necessary for generalization across different instruments. Although information from acoustical and timbre perception studies (e.g., Grey, 1975; Wessel, 1979; Ehresman & Wessel, 1978; Grey, 1978; McAdams, 1984) is relevant to orchestration and information from such studies has been used as source material for analytical or compositional theories of timbre (Erickson, 1975; Cogan, 1976; Cogan, 1984; Slawson, 1985; Lerdahl, 1987), no large-scale study has directly applied perceptual experiments to questions specific to orchestration. Central to this study, therefore, will be a series of perceptual experiments investigating some aspect of orchestration. Because of the absence of precedents, and in the interest of thoroughness, the study will not be able to address a wide range of aspects of orchestration. By necessity, it will be restricted to investigating one particular orchestral phenomenon. A survey of orchestration manuals and treatises (Chapter 1) will weigh various candidates for their suitability as topics for perceptual investigation. It will provide evidence that "blend," a quality obtained between concurrently-sounding musical instruments, emerges as a topic that is well-defined and widely applicable to a variety of orchestral activities.

There are two possible benefits to the activities of music theory and composition to be gained by this study. First, endowing orchestration pedagogy with paradigms and methods brings it closer to the other subjects of music theory: if there are metrics for producing orchestration, those same metrics can be used to analyze compositions as well. Conceivably, then, orchestration and orchestration analysis could be encompassed under one subject, each to the other's advantage. The second benefit relates more to the current musical scene. Multi-timbral synthesizers and micro-processor based controllers, widely available and at relatively low cost, have broadened the opportunities for composers to write "orchestrally" without the expense or limitations of conventional orchestras. The freedom to compose with multiple layers of timbral sound, and the opportunity to evaluate an unlimited number of combinations of synthesizer patches on one underlying compositional grid could trigger a renewal of wide interest in the art of orchestration. If so, the "orchestration manual" would need to be redefined for this new context: as electronic musicians are increasingly informed about synthesis algorithms and the psychoacoustics on which their instruments are based, they will require a system of orchestration founded on principles of auditory perception.

The remainder of the study will proceed as follows. Chapter 1 will examine traditional orchestration manuals and point out how the various questions they raise help to identify an appropriate candidate topic for modelling with a "metric." Chapter 2 will assess the contributions in the literature of auditory perception and psychoacoustics for material pertinent to the selected topic. Chapter 3 will propose an agenda for empirical study and introduce the working methods and stimulus materials for experimental investigation. Chapter 4 will report on the experiments themselves. Chapter 5 will summarize the overall findings of the experiments, evaluate their significance with respect to the art of orchestration, and consider future possibilities suggested by the outcome of this work.

Chapter 1:

Orchestration Manuals

For most musicians, the primary source of information on orchestration is its textbooks, manuals, handbooks and treatises. Literature reviews of orchestration manuals, encompassing the variety of topics of concerns and methods of instruction found in them, are rare or non-existent.[1] As part of the search for a suitable aspect of orchestration for perceptual investigation, a selective review of the contents of these sources is given · here.[2] Although several potential topics for investigation are encountered here, the topic which ultimately emerges as the strongest candidate is *blend*: the tendency for concurrently-sounding timbres to fuse into a single timbre.

---

[1] Carse (1964) and Becker (1969) offer surveys of the history of orchestration that provide brief allusions to orchestration manuals from various periods. Carse (1941) briefly considers treatises predating Berlioz (1856). Strawn (1985) offers a restricted survey of orchestration manuals pertaining only to discussions addressing matters of articulation and note-transitions.

[2] The survey is limited to sources that available in English. Not all sources surveyed were "orchestration manuals;" a few sources of more limited scope, or which address orchestration from non-pedagogical viewpoints, are included as well as well. Ott (1969), Schoenberg (1931), Scherchen (1933), Carse (1964), Becker (1969), Fowler (1980), Belkin (1988) are examples of these sources.

As an overview, consider the typical instructional layout of an orchestration manual:

1. Description of instrumental families and individual instruments; focus on practical limitations, qualities of sound, and strengths of registers.
2. Melodic writing: choosing combinations of instruments for doubling melodies at the unison, octave, or in thirds or sixths.
3. Harmonic writing: voicing and instrument selection to obtain homogeneous and balanced chords
4. Techniques for orchestral accompaniment: textures, polyphonic patterns
5. Orchestral transcription of piano pieces or chamber music

Topic (1) indeed plays a major role in orchestration teaching, usually occupying an initial, major portion of manuals. Nearly all manuals stress the necessity of a thorough understanding of the character, capabilities and limitations of every instrument as a precursor to all other orchestration activity. These sections develop a vocabulary for describing *quality of timbre* and *registral strength* that tends to permeate the instructional nature of the remainder of the book. That is, effective instrument choices, whether in solo or simultaneous presentation, involve applying information of timbral quality and strength wisely. Hence each of these two topics will be surveyed separately. Topics (2) and (3), however, can be considered together because both concern strategies for selecting simultaneous timbres, and authors tend to apply the same principles whether the issue is melodic doublings or chord voicings. Topics (4) and (5) involve high-level application of all three of these subjects; the latter, additionally, involves numerous stylistic questions. Hence the subjects of accompaniment and transcription are omitted from consideration *a priori* as topics too complex to consider for perceptual investigation. Therefore, the material to be reviewed here will be organized according to only three categories: (a) Semantic description of instrumental timbre, (b) characterizing strengths of instruments, and (c) concurrent timbre.

The goal is to find among these three categories some topic suitable for perceptual investigation. To guide the direction of this survey, topics will be evaluated according to the degree to which they satisfy the various criteria mentioned in the Introduction. These criteria are broken down, for clarity, into the following five categories:

Definability. A clearly defined sonic objective.

Explainability. General acoustical criteria for meeting that objective are suggested.

Demonstrability. Examples of successful and non-successful instances (in terms of specific instrument choices) of the sonic effect.

Continuity: The effect can be experienced to various degrees (i.e., along a continuum), not merely as an "either-or" phenomenon.

Relevancy. Relevance as an orchestral technique for application in diverse contexts.

Semantic Descriptors of Instrumental Timbre

One possibility as a point of departure for investigating orchestration may be to consider the types of descriptions musicians employ when writing about timbre. The terms they employ constitute a significant component in orchestration teaching, appearing throughout the definitions of various orchestral techniques and the descriptions of strengths and weaknesses of various instruments. Examining this aspect in detail, then, will provide valuable information about orchestration instruction. Furthermore, techniques which are shown to depend heavily on such descriptions may suggest candidates for perceptual investigation.

Adler (1989) stressed the importance of learning the various characters and capabilities of every instrument in the orchestra as follows:

> Of course, awareness of the range and limitations of each instrument is essential . . . . The timbre, strength, and texture of every segment of each instrument's range becomes crucial when you are creating color combinations that will characterize the music to be played. (p. 459)

Similarly, Richard Strauss's advice to the student of orchestration was that he should "above all ask instrumentalists of all kinds to familiarize him with the exact technique of their instruments and with the timbre of their registers" (Berlioz-Strauss, 1945, p. I).[3]

With this priority in mind, initial chapters of orchestration manuals are devoted to description of each instrumental family (strings, woodwinds, brass, percussion) broken down into discussions of individual family members. These chapters broadly discuss, for example, "clarinet timbre," or in other cases, give detailed descriptions of the changing qualities of an instrument's various registers. For woodwind and brass, the patterns of fingering and overblowing tend to determine the registral segments; for the string family, the changes in color tend to be determined by the individual strings (G, D, A, E, etc.). The effects of different playing methods (bowings, toungings), use of mutes, and dynamic on each register are also considered.

To characterize these qualities, authors of orchestration manuals do not use acoustical description; rather they use one-word semantic descriptors or phrases intended to suggest the aural impression of the instrument with evocative metaphors, simile, onomatopoeia, and comparisons to other sensual modalities. Words that are typically found in orchestration manuals are:

> dark, light, brilliant, dull, clear, veiled, transparent, and pale (visual modalities)
> fuzzy, velvety, smooth, soft, silky, coarse, rough, hard, grainy (tactile modalities)

---

[3] All references to Berlioz-Strauss (1945) refer solely to Strauss's editorial additions to Berlioz's treatise.

rich, mellow, bland, pungent, tangy (gustatory modalities)
thick, heavy, thin, rounded, full, hollow, metallic, reedy, brittle, woody, delicate, liquid, glassy (geometry, volumes, physical matter)
warm, windy, dry, cold, cool (environmental conditions)
calm, introspective, expressive, somber, poignant, melancholy (moods and emotions)

Orchestration manuals provide numerous examples of instruments that yield such qualities, and use semantic description as the basic building blocks in the definition of certain orchestral techniques, as shall be shown here. A primary concern is whether such terms are explainable (i.e., whether or not the terms correspond to acoustic regularities), and given such a basis, whether such terms provide a reliable method for instruction in orchestral technique. These issues are considered in the following sections.

Semantics as Guide to Similarity

Timbral similarity is an important relationship that plays a role in a number of orchestral techniques. The method used by orchestration manuals to describe similarities among instruments is rooted in semantic descriptors: a term such as "rich" acts as a category by which any two instruments so described can be assumed to be similar, for example.

Examples of applications where knowledge of timbral similarity is valuable are (a) "echo" passages, where melodic imitation is reinforced by a timbral imitation as well (see Adler, 1989, p. 217; Rogers, 1951, p. 105-106; Piston, 1955, p. 257; Rimsky-Korsakov, 1912, pp. 110-111); (b) "tone-color melody" techniques (*Klangfarbenmelodie*), where the similarity relationships among a succession of timbres define functional relationships (see Slawson, 1985, and Lerdahl, 1987); or (c) concurrent instrumental arrangements (i.e., melodic doubling or chord-construction).

This approach of using semantic categories to motivate choices for simultaneous presentations of instruments is recommended by many different authors (see Russo, 1968, p. 562; Ott, 1969, p. 53; Stiller, 1985, p. 8; Belkin, 1988, p. 50; Adler, 1989, p. 472; Schoenberg, 1931, p. 335). A frequently suggested strategy is to organize instruments according to semantic categories, and combine instruments from like categories (see Schillinger, 1941, p. 1326; Rogers, 1951, p. 5; Blatter, 1980, pp. 375-376, Rimsky-Korsakov, 1912, p. 14). Blatter's scheme, for example, has eleven categories:

dark and smooth
dark and mellow
dark and reedy
dark and full
neutral and full
bright and smooth
bright and clear
bright and full
nasal and bright
nasal and dark
percussive

Instruments included in Blatter's "nasal and bright" category, for example, are: oboes, all saxophones, English horn, bassoons, chalumeau register of the clarinets, harmon-muted brasses, and stopped horns. Blatter indicates that the quality "nasal and bright" will be yielded by any single instrument, or any combination of instruments from this list (Blatter, 1980, pp. 375-376).

Both Schillinger (1941) and Ott (1969) propose categorizations of instruments according to their similarity to *vowel sounds,* and consequently suggest them a guide for combining like instruments. In Schillinger's "Five Degrees of Tone Quality," each category is named by a semantic term (open, single-reed, stopped, double-reed/nasal, and closed, p. 1326) as well as a corresponding vowel ([u], [o], [ɑ], [ɛ] and [i], respectively, p. 1573).[4] Ott (1969) proposes the intriguing suggestion that the student undertake a project of imitating every instrument in the orchestra with his or her own voice, creating a personal categorization according to the vowels chosen (p. 54).[5]

Acoustic Correlates to Semantic Descriptors

On the surface, semantic terms in orchestration manuals appear to be loosely used, with their exact acoustical attributes left open to question. However, by examining other sources it can be shown that this "orchestral vernacular" has some basis in auditory reality. As this section will show, a number of terms are accorded some validity by either acoustical analyses of instruments, or empirical studies investigating semantic attributes of timbre.

Certain frequently-encountered terms appear to have meanings in orchestration manuals that are shared by researchers in timbre perception.

---

[4] The symbols used here and throughout the study for vowels are those of the IPA (International Phonetic Assocication). A pronunciation guide to IPA symbols may be found in Stiller (1985), pp. 99-102.

[5] The fact that students will be constrained to imitating only those instruments which fall within the limits of their own vocal ranges is something that Ott fails to note, however.

Two terms that are nearly ubiquitous in orchestration manuals are "dark" and "bright." They enjoy a widely-accepted interpretation by timbre researchers as well: they refer to "the location on the frequency continuum of the midpoint of the energy distribution" (Lichte, p. 472) or "the frequency position of the overall energy concentration of the spectrum" (von Bismarck, 1974, pp. 156-157). Dark refers to a low frequency position, and bright to a high position; many researchers employ the term *spectral centroid* to refer to values along a continuum from dark to bright (Gordon and Grey, 1978; Wessel, 1978; Beauchamp, 1982).[6] "Rough," and a similar term, "coarse," are sometimes used to describe instruments of especially low pitch compass: for example, low bassoon (Piston, 1955, p. 193, Rimsky-Korsakov, 1912, p. 16), low contrabassoon (Rimsky-Korsakov, 1912, p. 16), and low tuba (Rimsky-Korsakov, 1912, p. 24). Their meaning has a possible acoustic parallel in the measure of "roughness" based on the beats of closely-tuned sine tones within auditory critical bands (Plomp and Levelt, 1965; Terhardt, 1974) "Hard" is often employed in orchestration manuals to describe instruments of especially short attack time, such as high string pizzicato (Rimsky-Korsakov, 1912, p. 31), high oboe (Rimsky-Korsakov, 1912, p. 16, Berlioz, 1844, p. 81), or high English horn (Rimsky-Korsakov, 1912, p. 16); timbre researchers have found this descriptor useful as well (Grey & Gordon, 1978, p. 1499, attributed to Wessel).

---

[6] A quantitative measure of centroid will be introduced in Chapter 3.

The comparison of instrumental timbre to the sounds of the human voice is a universal phenomenon, and not surprisingly, some terms primarily associated with the voice have come into usage for instrumental timbre. The categorization systems of Schillinger (1941) and Ott (1969), based on similarities between instruments and vowels, were mentioned earlier. There is indeed some evidence for an acoustical basis of such comparisons. Meyer (1978), for example, identifies formants in certain instruments that correspond to first formants of rounded vowels;[7] indeed, some instruments are described as sounding "rounded" in orchestration manuals (English horn: Blatter, 1980, p. 94, Riddle, 1985, p. 132; cornet: Riddle, 1985, p. 62, Rogers, 1951, p. 61; tuba: Adler, 1989, p. 321, Piston, 1955, p. 287). Meyer lists examples such as [u] in the tuba and French horn (formant at 225 Hz), [o] in the tenor trombone (formant at 550 Hz), and a quality between those two vowels in the bass trombone (formant at 400 Hz; Meyer, 1978, pp. 42, 37). The term thin, often used to describe the oboe tone (Blatter, 1980, p. 93, Piston, 1955, p. 152, Rogers, 1951, p. 35, Berlioz, 1844, p. 81, Berlioz-Strauss, 1945, p. 176), invites comparison to those vowels that are produced with widely spread lips, such as [i]. Indeed, the oboe is known to possess a high formant in the neighborhood of the second formant of [i], near 3000 Hz (Strong & Clark, 1967, p. 44; Olson, 1967, p. 224; Moorer, Grey, and Strawn, 1977, p. 22; Meyer, 1978, p. 51).

---

[7] "Rounded" is a phonological term for vowels produced with rounded lips.

Consonants can also be used to characterize timbres. Stiller (1985) uses various consonants to characterize the kinds of sounds that brass mutes offer. With the plunger mute, for example,

> the sounds of the vowels [ɑ], [ɔ], [o], and [u] can be very closely approximated. . . . The consonants [m], [w], and [β] can also be produced, and such word-like combinations as [βɑ͡ʊmɔwo] can be easily done. (p. 83)

More specifically, Meyer (1978) uses "nasal" to describe formant-like concentrations of spectral energy in the area of 1.2 to 1.8 kHz (p. 26). According to Meyer, an increase in the level of these components can detract from the brightness, clarity, sonority and openness of an instrument's tone and lead to a dull sound (implied, pp. 42, 45, 57 and 58). He identifies significant nasal components in straight-muted trumpet (p. 44), high-register bassoon (p. 57), and the viola (p. 65), an instrument which orchestration manuals are virtually unanimous in describing as nasal (specifically with respect to the A-string). The oboe and other double reed instruments are also called nasal by nearly all orchestration manuals.[8]

Terms in orchestration manuals sometimes have other plausible associations with specific acoustical parameters. The description "hollow" for low clarinet notes (Forsyth, 1914, p. 236; Rogers, 1951, p. 40) is an apt description of its spectrum, with its characteristically weak energy at even-

---

[8] The practice of comparing wind instruments to nasal speech has had a long history. Note the clown's question to a group of performers of woodwind instruments in Shakespeare's Othello (Shakespeare, 1605): "Why, masters, have your instruments been in Naples, that they speak i' the nose thus?" (III.i)

numbered partials. The terms "full" and sometimes "heavy," often are applied to those instruments with relatively strong fundamentals in the lower range, such as contrabassoon (Rimsky-Korsakov, 1912, p. 16), bass clarinet (Rimsky-Korsakov, 1912, p. 16), bass trombone (Adler, 1989, p. 314), tuba (Stiller, 1985, p. 91; Blatter, 1980, p. 376), and saxophone (Blatter, 1980, p. 119; Berlioz, 1844, p. 233). The word "rich" is often used to refer to sounds with a large number of audible harmonics, as in the ubiquitous phrase "richly harmonic." Examples of instruments described as "rich" are the lower strings on all members of the string family (Blatter, 1980, pp. 45-56; Piston, 1955, p. 67), low alto flute (Blatter, 1980, p. 86, Adler, 1989, p. 179, Piston, 1955, p. 145) and chalumeau-register clarinet (Blatter, 1980, p. 101, Adler, 1989, p. 189, Piston, 1955, p. 168). An observation by Meyer (1978) corroborates this usage of the word with regard to the trumpet, "the instrument in the orchestra richest in overtones. . . . [it has] components right up to the limit of hearing in *ff*" (p. 43). The opposite condition, few harmonics, is sometimes characterized with the words "thin," "dull," "pale," or "pinched." One example of the latter, by Forsyth (1914), is particularly clever for its coincidence of poetic imagery and informed acoustical description: "the absence of the characteristic harmonics of the lower octaves becomes increasingly noticeable. The result is a somewhat pinched tone-quality as of someone complaining about his poverty" (p. 236).

Advantages and Disadvantages of Semantic Description

The value of semantic descriptors is a matter of some debate. The exceptions notwithstanding, their usage in orchestration, and even in some perceptually-oriented research in timbre, is frequently imprecise and the meanings unspecified.

However, a semantic term often concerns important *qualitative* aspects of timbre which may have a bearing on the factors motivating instrument choices in orchestration. In that respect, the use of semantic terms may be defended on cognitive grounds. It is well known from psychological studies of memory that any item is more easily encoded and retrieved from memory when one can represent it according to simple, but meaningful categories (Rosch, 1975). Regardless of whether semantic terms have consistent acoustic correlates, making associations between sounds and words aids in the learning and memorizing of qualities of sounds. As Erickson (1975) points out, the human tendency to attach labels to the sounds we hear is natural, subconscious and pervasive. Humans constantly make such classifications of everyday sounds based on the information--whether scanty or elaborate--we receive from our sensory inputs. We instinctively and quickly decide that this or that sound is a "squeak" or a "crunch" as best the prevailing conditions for our ear to analyze allows (Erickson, 1975, p. 11). Much of the activity of evaluating timbre (and hence, apprehending a musical structure which is projected through timbre) is no doubt based on spontaneous categorizations such as these. So understanding them could have significance for the creation of the rule-based approach to timbre and orchestration which is desired here.

Despite the fact that studies have provided empirical evidence of regularities between semantic descriptors and specific acoustical properties (Lichte, 1941; Solomon, 1958; Abeles, 1969; von Bismarck, 1974; Gridley, 1987), the usefulness of semantic descriptors for characterizing musical timbres has been questioned. Grey (1975) offered the following criticisms: (a) they do not "reveal factors which [are] *uniquely* related to independent properties of the stimuli," (b) "a single word may be associated with a number of independent stimulus dimensions," (c) they measure "complex aesthetic reactions to stimuli" rather than "direct information about many of the perceptual processes," and (d) some words do not exist or aren't available to describe certain perceived differences (pp. 7-8). Consequently, we may be unable to generalize, for example, from Adler's description of both the violin E-string and high-register trumpet as "brilliant" (Adler, 1989, pp. 59, 302): there may be numerous acoustical parameters responsible for the sensation, of which only a few are actually shared by both instruments. Consequently, without a precise definition there is no way to judge whether an arbitrary attribution of "brilliant" to an instrument is valid or invalid.

In addition the above criticisms, the following problems in the use of semantic terms are common in orchestration books:

1. The use of the term is with respect to an unnamed comparison: for example, when a tenor trombone is described as "bright," the

implicit meaning is often "the brighter member of the trombone family." In a larger context, for example including oboe and clarinet, the tenor trombone is relatively dark.

2. Terms are applied to instruments without specifying under what conditions (which dynamics, which registers) the attribution applies. A description of clarinet as "shrill" seems to be a reference to a particular stereotype of clarinet playing (high register, loud dynamic), but this is often not made explicit.

3. Terms often embody a variety of meanings beyond pure sonic characterization, often describing aspects exterior to actual timbre. Examples include words that convey the instrument's visual appearance ("silvery" for a flute), the character of the music written for such instruments ("regal" for trumpet), the cultural origins of the instrument ("exotic" for the oboe), and so on.[9]

Evaluation

Semantic description appears to play an important role as a tool for making basic qualitative decisions about timbre, especially in determining similarity between instruments. The existence of orchestration methods based on similarity helps satisfy the definability criterion. There is some evidence of acoustical regularities in the usage of semantic terms in

---

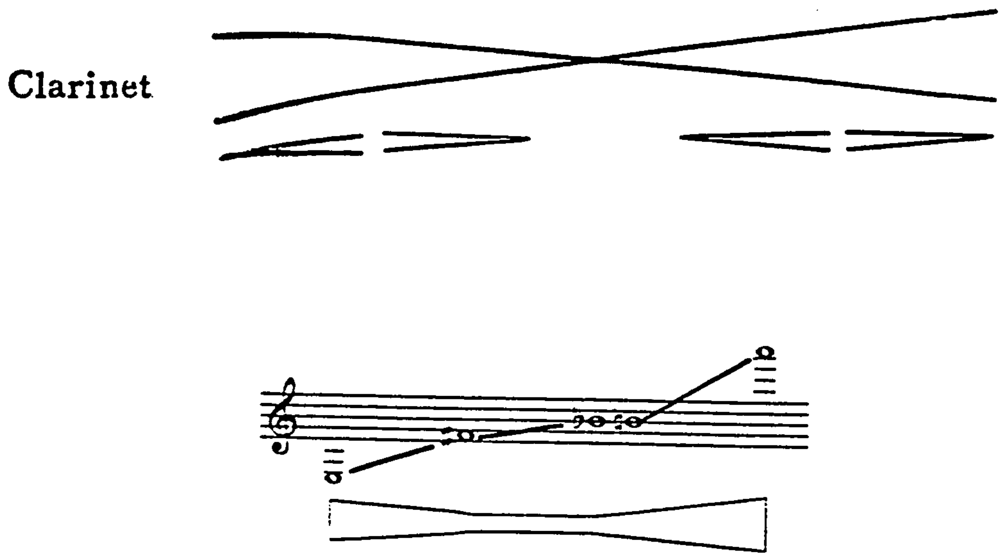
[9] This cross-modal confusion is especially characteristic of Rogers (1951).

orchestration, and there is no shortage of examples of instruments for each term: this helps satisfy the explainability and demonstrability criteria. Also, different degrees of similarity (e.g., somewhat similar, very similar, nearly identical), can be experienced, satisfying the continuity requirement. However, the basis for similarity is indirect and highly subjective: instruments are considered similar if they can be described with the same semantic labels. Orchestration manuals provide few solid acoustic cues to the mechanism of similarity itself, such as what makes a viola sound like a saxophone (Blatter, 1980, p. 320). Admittedly, it might be argued that since the perceptual attributes of timbral similarity has been extensively studied by other researchers (Plomp, 1970, 1976; Wedin & Goude, 1972; Miller & Carterette, 1975; Grey, 1975; Ehresman & Wessel, 1978), it may be possible to explicate this mechanism and extend its role in orchestration. Still, although orchestration manuals encourage using similarity relations as a strategy, they do not address the sonic implications of similarity in any detail; that is, rather than suggesting the different *ways* sounds may be considered similar, they simply offer similarity as a way to obtain "good combinations." This leaves the definability criterion poorly satisfied.

Characterizing Strengths of Instruments

Another category of description of timbre used frequently in orchestration manuals pertains to the strength or power of the instrument in its various ranges. This topic is also an important component of orchestration manuals and offers a possible point of departure for investigating orchestration. However, since this information is largely useful for attaining other goals rather than describing a specific sonic condition, it will be shown that the definability condition is poorly met.

The strength of tones within a given register of an instrument are typically a function of (a) its ability to play a wide range of dynamics (from soft to loud), (b) its ability to project its characteristic timbre, and (c) the ease of performance and freedom from idiosyncratic mechanical difficulties in the instrument. Orchestration manuals identify the registers of instruments which suffer due to one or more of these factors and recommend their cautious use in both melodic and harmonic writing. Typical examples are low register flute (cannot play loudly), low oboe (tone is raspy and loud), high bassoon (tone is thin and soft), and low brass or woodwinds (notes "speak" more slowly). Some manuals provide characterizations of the dynamic potentials of all notes across the instrument's pitch compass (Schillinger, 1941; Blatter, 1980; Stiller, 1985); see Figure 1 for examples.



EXAMPLE 128 Dynamic curve of clarinets.

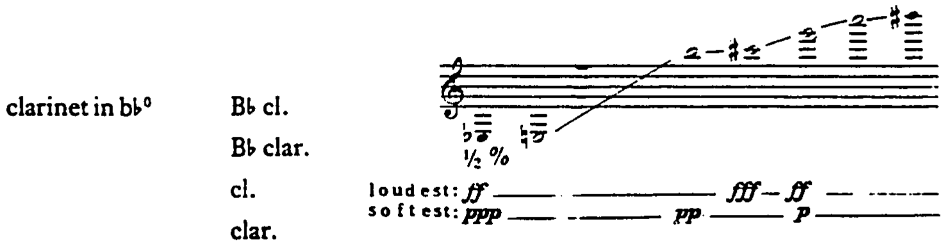


Figure 1. Characterizations of dynamic ranges of the Bb Clarinet from three orchestration manuals. (Sources, from top to bottom: Schillinger, 1941; Blatter, 1980; Stiller, 1985).

It is often not easy to draw a strong distinction between words that describe "strength" and words that describe other qualities of timbre. Many words apparently are simultaneously a characterization of the loudness and the timbral quality, such as "delicate," "light," "veiled," and "distant," or "penetrating," "incisive," "strident," and "piercing." Indeed, Fowler (1980) suggests that terms such as these have come to take on functional meaning among orchestrators:

> . . . they do more than define color characteristics; they also connote relative carrying power and relative masking capacity. . . . qualities like "dark," "gross," "thick," or "opaque" suggest strong masking capacity, while qualities like "dry," "thin," "lifeless," or "cold" do not. (pp. 68-69)

With melodic instrumental writing, the orchestration student is advised to guard against the accompaniment overshadowing a weak register, or choosing unfeasible instruments for the intended expression (e.g. calling upon the oboe to play a tender and lyrical melody in the lowest part of its pitch compass). For doubling melodies, Rimsky-Korsakov (1912) offers the rule of thumb that "The best and most natural combinations are between instruments whose registers correspond the nearest" (p. 58), echoing the paradigm that instruments with like qualities tend to combine well. As far as choosing *successions* of solo melodic timbres, orchestration manuals are mostly silent; one rare exception is Forsyth's piece of advice, warning against

> the danger of allotting any subject of a serious or poetical nature to the Bassoon if it involves repeated, detached, or *staccato* notes. Such a subject, especially if first heard on

> the Strings, will appear ludicrously commonplace when heard on the Bassoon. (Forsyth, 1914, p. 236)

For harmony, Piston's recommendation for matching strengths of instruments is typical of most orchestration manuals.

> Judged by the relative tone-weight, or carrying power, of each note, assuming that the intent is to create a vertical plane of sound in which no tone emerges or protrudes perceptibly[,] . . . contributing factors in balance are the number of instruments playing each note, and the relative dynamic power of the individual instruments in their particular registers. (Piston, 1955, pp. 397-398)

One compositional technique requiring an understanding of the *range* of strengths and weakness among instruments is the *orchestral crescendo,* involving the steady addition of instruments to a small ensemble until it reaches *tutti.* Rossini frequently used this technique in conjunction with repeated motivic material to achieve orchestral climaxes in his operatic overtures. Rogers (1951) offers the advice that "instruments must be added skillfully to the ensemble---often a beat or so before they are needed in the harmony; while the strongest instruments are the last to enter" (p. 105). Alternatively, the orchestral descrescendo (as in Haydn's *Farewell* Symphony) is possible.

The *echo* effect, cited earlier as a orchestral technique dependent upon similarity relationships, also involves exploiting relative differences in strength among instruments. In orchestral music, the repetition of material in call-response or subject-answer fashion is often intended to suggest a spatial idea, of the same instrument being heard first in close proximity, then from a distance. Creating this illusion depends a great deal on effecting a "decrease in volume of tone . . . the second instrument should be weaker than the first" (Rimsky-Korsakov, 1912, p. 110). Instrument successions that will be successful at achieving this are muted trumpets answering open trumpets (Piston, 1955, p. 257; Rimsky-Korsakov, 1912, p. 110; Rogers, 1951), or "muted brass answering woods or strings" (Rogers, 1951, p. 106), or strings answering brass. In the case of phrases echoed an octave above or below the original, Rimsky-Korsakov (1912) warns that such passages "are subject to the law of register. When a phrase is imitated in the upper register it should be given to an instrument of higher range and *vice versa*. If this rule is ignored an unnatural effect will be produced, as when the clarinet in its upper range replies to the oboe in the lower compass etc." (110-111). In other words, the composer must take care that the ranges the two instruments are playing in are not drastically different, or else the sense of imitation will not be apprehended.

Since the strength of an instrument is a matter of multiple factors (i.e. dynamic range, timbre and mechanical characteristics), a simple recipe for properly matching instruments to one another is seldom suggested in orchestration manuals. However, they do offer recommendations about how to handle specific instances. Some examples of these recommendations, paraphrased from orchestration manuals, are given below:

1. When the dynamic is forte or louder, two of any woodwind are needed to match the whole of one department of strings. In piano, just one woodwind is needed (Rimsky-Korsakov, 1912, pp. 33, 40).

2. It takes two woodwinds to match each single French Horn in power (Rimsky-Korsakov, 1912, p. 33). For a trumpet, four woodwind are needed (Rimsky-Korsakov, 1912, p. 57).
3. A weak woodwind line can be reinforced by adding strings (Rimsky-Korsakov, 1912, p. 50).
3. When the dynamic is forte or louder, the horns are 1/2 as strong as any other member of the brass. Thus two horns are needed to balance one other brass (Rimsky-Korsakov, 1912, p. 33, Adler, 1989, p. 328).
4. Clarinet is particularly good at matching the loudness of whatever it is paired with (Scherchen, 1933, p. 69-70).

Evaluation

Like semantic description, *strength description* defines no particular sonic condition in any detail, raising questions about definability. It appears that matching instruments in terms of strengths (for simultaneous presentation) is a basic preliminary condition to making various orchestral effects carry off well, but this activity itself does not seem to be the subject of orchestral inventiveness. There are few orchestral considerations described in orchestration manuals that depend on matching instrument strengths. The technique of *echo effects,* while it exploits strength differences, is a novelty which has little applicability to other practices: thus the relevancy criterion is poorly met. The primary value of strength matching, obtaining balanced chords, does not itself suggest interesting or fertile territory for experimental investigation.

Concurrent Timbre

Concurrent timbre, to paraphrase a term from speech research,[10] is here used to refer to combinations of timbres in simultaneous presentation, or, the perception of them (depending on the context in which it is used). This is a subject that attracts a great deal of attention from orchestration manuals, which describe the sound of certain combinations, suggest certain procedures that generate interesting combinations, and offer many rules of thumb for making such combinations blend. As this section will ultimately show, this latter subject (blend) satisfies the five criteria for topic selection quite well, and suggests the most promising avenue for perceptual investigation.

The qualitative outcome of combined timbres is a subject of central concern to orchestration and the object of much discussion in orchestration manuals. Musicians in general find the phenomenon of how various instruments sounding together yield a sound that is seemingly more than the sum of its parts to be somewhat mysterious and fascinating. Indeed, in one of Sir Francis Bacon's essays on natural history, over 300 years ago, one finds a proposal for a systematic inquiry into concurrent timbre that predates even the earliest manuals of orchestration by over a century:[11]

---

[10] Summerfield and Assmann (1989), "Auditory enhancement and the perception of concurrent vowels."

[11] Becker (1969) credits Valentin Roeser as having published the first orchestration manual in 1764 (Paris).

> All concords and discords of music are (no doubt) sympathies and antipathies of sounds. And so likewise in that music which we call broken music, or consort music, some consorts of instruments are sweeter than others (a thing not sufficiently yet observed): as the Irish harp and base viol agree well; the recorder and stringed music agree well; organs and the voice agree well, &c; but the virginals and the lute, or the Welsh harp and Irish harp, or the voice and pipes alone, agree not so well. But for the melioration of music there is yet much left (in this point of exquisite consorts) to try and inquire (Bacon, 1627, Century III, sect. 278).[12]

Description of concurrent timbres, for some orchestration manuals, can be as simple as combining the semantic terms of their constituent instruments. For example, "dark and rich" might be used as a description of combined low flute and chalumeau-register clarinet. Stiller (1985) offers a rule of thumb along these lines for anticipating the sound of concurrent timbres:

> The sound produced by a unison of nonidentical instruments is exactly intermediate between the instruments involved, in a very obvious way. One soon becomes able to imagine the sound of such combinations even when one has not heard them before (p. 9).

There is indeed some incidental evidence from a study by Grey (1975, Study D, experiment 2, pp. 82-85) which suggests that two instruments sounding together can sound like an "interpolation" between them; listeners who heard computer-synthesized interpolations between, for example, French

---

[12] Bacon continues from here to introduce the phenomenon of sympathetic resonance.

horn and oboe reported the sensation of having heard both French horn and oboe playing together (p. 101).

However, Stiller's generalization is too simple; the "interpolation" theory amounts to saying that concurrent timbre is merely the sum of its parts. It does not account for the myriad of textures, modulating qualities, and subtle colorations that are implied in such descriptions as "vibrant and exotic" (English horn and viola: Rogers, 1951, p. 37), "extreme pungency" (three oboes and English horn: Piston, 1955, p. 155), and "incisive" (electric guitar and straight-muted trumpet: Riddle, 1985, p. 136). Clearly there are qualities of concurrent timbre which strike the ear as more interesting than merely the sum or interpolation of different qualities.

The main way in which orchestration manuals characterize the timbral quality of concurrent timbres is by combining their corresponding semantic descriptors, as mentioned above. Because of the subjectivity of semantic descriptors, as discussed earlier, this does not suggest a promising avenue of investigation. It may be more instructive to examine the *sonic objectives* that orchestration manuals describe as the motivation for selecting concurrent timbres. Several techniques are identified throughout a sample of these sources; some of these are: (a) augmenting existing timbres, (b) softening timbres, (c) inventing timbres, (d) imitating timbres (e) blending timbres. Various techniques, procedures and evaluative criteria for each of these effects are considered below.

Augmenting Existing Timbres

In some concurrent timbres, the overall sound may be dominated by the timbre of one instrument, while added instruments augment it with the timbral qualities of other instruments. This is the most common and simple concurrent timbre arrangement; Brant (1971) refers to it as "functional doubling" (p. 541). References to one process "shading," "thickening," or "adding highlights" to another instrument are typical. Russo (1969), for example, cites trumpet and oboe as an example of a case where

> One of the two instruments . . . dominates the other . . . we hear principally the sound of the trumpet, somewhat modified by the oboe. The sound is not precisely a trumpet sound, but neither is it a new sound. (p. 562)

In another example, Russo describes a clarinet which is subordinate to, and only "slightly colors" the sound of a Harmon-muted trumpet (p. 562). An example by Rimsky-Korsakov (1912) is "The Eng. [*sic*] horn is absorbed in the musical texture, the principle colour being that of the 'cellos" (p. 40). Often the goal is to intensify some native quality of the original instrument by an instrument that possesses that same quality in greater measure. Adler (1989), for example, describes the doubling of cello and bassoon by remarking that "the cello tone predominates, but the doubling bassoon or bassoons give added body to the sound even as they are absorbed in the whole" (p. 210).

Descriptions of the qualities of such combinations is often through the additive use of semantic descriptors: the semantic attribute of the augmenting instrument is "added" to quality of the dominant instrument. For example, clarinet in low range adds "warmth or body" to whatever it combines; in high range, "added brilliance or focus" (Blatter, 1980, p. 102). The oboe adds "bite" or "cutting edge" to a clarinet (Adler, 1989, pp. 220, 185) or fullness to a flute (Rimsky-Korsakov, 1912, pp. 47-48). One author says the flute thickens the clarinet (Adler, 1989, p. 221), while another says it dulls the clarinet (Rimsky-Korsakov, 1912, pp. 47-48). Cello pizzicato is "excellent...for brightening outlines" (Rogers, 1951, p. 27). Note also that merely duplicating an instrument (adding an extra player on the same instrument) in an upper or lower octave can also achieve the effect of "brightening" or "darkening" the first instrument; several suggestions of this sort are found in Rimsky-Korsakov's discussions of octave doublings in his chapter on "Melody" (Rimsky-Korsakov, 1912, pp. 36-61).

Another way in which timbres are augmented is where the added instrument "does not add a new melodic voice so much as a harmonic *thickening* or underlying of the single voice" (Piston, 1955, p. 362, emphasis added). It appears that the phenomenon of "thickening" is due to disagreements in fundamental frequency and harmonicity between the constituent tones of a combination (Stiller, 1985, p. 8; Adler, 1989, p. 233, 250; Russo, 1969, p. 558), creating a "unique turbulent quality" (Stiller, 1985, p. 8); jazz orchestrators like to call this a "fat sound." For example, Adler (1989) comments on the addition of flute to clarinet: "In this pairing, the clarinet would be the most poignant factor, while the flute would contribute little more than a thickening of the resulting tone" (p. 221).

Russo (1969) asserts that when doubling a melody with duplicate instruments, *three* instruments are better than *two*. With only two instruments,

> discrepancies in intonation are made apparent, as are discrepancies in interpretation; and the actual quality of sound is dull and unrewarding. . . . The use of *more than two* identical instruments in unison is a different cup of tea. The discrepancies of intonation which disturb us in unisons of two trumpets constitute a decided advantage in unisons of three or more trumpets, for it is these discrepancies that give larger groups of instruments their body. (p. 558)

Adler (1989) also notes the fine line between effective and ineffective doublings, warning against unisons that "detract from the clarity of a line and thicken the sound by muddying the overtones of both instruments" (p. 250).

Softening Timbres

Orchestration manuals speak of some alterations of timbre where an added instrument does somewhat more than add a "shading," but modifies the quality of the first instrument as well. References to "taking the edge off" (Blatter, 1980, p. 321, Rogers, 1951, p. 105), "tempering the hardness" (Piston, 1955, p. 365, Berlioz-Strauss, 1945, p. 97) or reducing the brilliance, incisiveness or penetrating quality of a sound (Rogers, 1951, p. 105, Blatter, 1980, p. 321, Piston, 1955, p. 427) are frequent. The word most commonly used to describe this modification is "softening." For example, "the quality of the strings softens that of the wood-wind" (Rimsky-Korsakov, 1912, p. 58).

Furthermore, the most frequent agent for softening other instruments is the flute. Piston (1955) describes it as softening the oboe and clarinet (p. 365); similarly, Rimsky-Korsakov (1912) describes flute and clarinet joining together to soften the oboe (p. 78). Blatter (1980) comments on a case of "low-register flutes where the extra mass seems to . . . 'take the edge' off of the sound [and] the result is actually *less* penetrating than a single flute in the same range" (p. 321). Blatter's reference to "extra mass" is reminiscent of the process of "thickening," and a comment by Rogers reinforces the suggestion of a relationship between the two processes: "The high brass loses brilliance when doubled in unison by woodwinds. Its tone becomes thicker but less incisive. Some of the flashing edge is lost" (Rogers, 1951, p. 105).

Inventing Timbres

When instruments are combined not with the purpose of merely shading an existing timbre, but creating a new one, the goal is to obtain a quality quite unlike any of its constituent instruments, a practice Brant (1971) calls "expressive doublings" (p. 541). Four examples of references to this phenomenon in orchestration are given below:

> The best and most colorisitically valuable pairing is obtained between two dissimilar instruments. . . . the two instruments will blend into a quite new color unlike the timbre of either instrument by itself...[they] absorb each other, becoming a new color. (Russo, 1969, p. 562)
>
> The most striking mixed unisons are those that involve unlike timbres, for when the two parent timbres are very

> distinct from each other their combined sound will also be distinctive. (Stiller, 1985, p. 9)
>
> Doubling different instruments in unison blends their separate colors into a new hue of somewhat strengthened carrying power, but somewhat reduced clarity. (Fowler, 1980, p. 69)

Riddle (1985) suggests, as a way to obtain "fresh and unusual colors," combinations that include the less familiar auxiliary members of instrumental families or special playing methods; examples he includes are English horn with alto flute, alto flute with cup-muted trombone, oboe with straight-muted trumpet, and four trombones with baritone saxophone (pp. 132, 137). Some invented timbres have come to be widely used. For example, Carse (1964) cites the combination of two clarinets, two bassoons and celli as attributable to Mendelssohn, used as a "rich tenor unison melody" in the *Ruy Blas* overture (p. 261). One such combination--bassoon with flute and oboe--has the singular honor of having its own name, "Viennese unison" (Becker, 1969, p. 24).

A more complex kind of invented timbre can be called "artificial timbre." This is defined by Stiller (1985) as the creation of a single timbre by having a number instruments act as individual partials of a single complex tone (p. 9). That is, the specified pitches are chosen to reflect natural harmonic multiples of a fundamental (although inharmonic complexes are possible as well), resulting in a kind of "additive synthesis" with instruments. The celebrated passage in Ravel's *Bolero* (1927), in which several instruments duplicate a melody in parallel at the octave, fifth, second octave and third (i.e.

second through fifth harmonics) to imitate a rich organ registration, is an example of this. Stiller (1985) offers advice on how the student might go about writing such passages:

> The following conditions should generally be observed: first, that lower partials should be louder than higher ones, with the fundamental and/or second partial being loudest of all; second, that the fundamental should have a strong, bright timbre of its own, while the upper partials should be comparatively smooth and pure. (p. 9)

Stiller includes ordinary octave duplication of melodies as a primitive form of "artificial timbre" as well (p. 9); as Blatter (1980) explains,

> when the lower line is scored to be more prominent (louder) than the upper line(s). . . . This causes the upper line(s) to be perceived as partials of a new timbre which has the lower line as its fundamental. (p. 294)

A related practice recommended in some orchestration manuals is to enrich brass chords with woodwinds in such a way that they "enhance---and sound like---vivid upper partials of the brass" (Rogers, 1951, p. 105). Similarly, Piston (1955) describes the woodwinds acting "as reinforcement of upper partials of the brass" where they "add brilliance" (pp. 427, 449).

Finally, temporally-variable timbral effects can be obtained by strategic choices of dynamic markings and taking advantage of instruments' natural temporal qualities. For example, "a *fortepiano* attack played by a muted trumpet can be significantly altered by adding unison string pizzicati to each attack" (Blatter, 1980, p. 300). Rimsky-Korsakov (1912) calls such effects "*Sforzando-piano* and *piano-sforzando* chords" (p. 111). Another effect is the crossfade, where a "sound *mutates* or changes its identity" as a result of different instruments playing *crescendo* and *decrescendo* simultaneously (Belkin, 1988, p. 49). A related technique is suggested by Blatter (1980): "During a held oboe tone, the clarinet could enter much, much softer, crescendo up to match or exceed the oboe's loudness, and then decrescendo to nothing while the oboe remains unchanged" (p. 300).

Timbral Imitation

Occasionally composers write for combinations of instruments using a carefully selected choice of musical material, dynamics, register, and articulation to mimic the sound of another instrument or instruments. This can be accomplished by selecting instruments that in different ways bear resemblances to this or that quality of the target instrument; when played together, the different features sum together and evoke the image of the target. The effect of deception or masquerade is pursued as an expressive end in itself. For example, Rogers characterizes the opening motive played by low flute and oboe in Strauss's *Don Quixote* as an imitation of a trumpet fanfare; this imitation are obtains its charm because it "deliberately plays upon the listener's sensibility, [and] stirs the imagination by the subtle choice of an off-shade instead of an everyday coloring" (Rogers, 1951, pp. 105-106). In other cases this practice is applied towards perfectly practical ends: for example, where the effect of four French horns is desired, but only two are available, two bassoons can masquerade as the missing horns (Riddle, 1985, p. 53; Rogers, 1951, p. 106).

Blatter (1980) provides a remarkable list of no fewer than sixty possible "instrument substitutions," some of which include ways of simulating a solo instrument (e.g., low flute, high soprano saxophone), with the combination of two or more instruments of a different type. Among the possibilities he lists are: clarinet, bassoon and muted horn as an imitation of low register English horn; trombone and contrabass as an imitation of a contrabassoon; and French horn and viola as an imitation of low register alto saxophone (Blatter, 1980, p. 320).

Blending Concurrent Timbres

In all discussions of concurrent timbre, including the techniques described above, often the sole and ultimate criterion for their use is whether or not the combination "combines well." As mentioned earlier, one of the shortcomings of orchestration pedagogy is that this ultimate criterion is seldom defined. There is an exception, however, with the attribute of "blend." It is apparent in the writings of many authors that blend is an important metric for evaluating whether concurrent timbres combine well: furthermore, it is applied in many contexts of orchestration, the same basic definition appears to be used by a wide variety of authors, and similar methods are offered for achieving it. Thus, this section will give extensive consideration to how blend is defined, how it is obtained, and how it is used compositionally.

Defining Blend

Blend is defined in Webster's dictionary as "to combine or associate so that the separate constituents or the line of demarcation cannot be distinguished." Interpreted in musical terms, blended combinations would be those in which the distinctiveness or individuality of the constituent instruments is subordinated to obtaining an overall, uniform timbral quality. It is easy to see that this is an important requirement for the success of many of the effects cited earlier. If, for example, a bassoon is combined with cello for the purpose of adding a slightly grainy quality to the cello (augmented timbres), or if a horn and viola are to masquerade as an alto saxophone (timbre imitations), it is important that they form a composite sound, to suggest that the hybrid originated from a single source. Choosing instruments, ranges, dynamics and articulations to obtain blended combinations is one of the significant challenges of orchestration.

Orchestration manuals are mostly in agreement with one another on the use of this definition of blend. In Piston's description of the combination of English horn and cello, the meaning is fairly clear: "The two instruments blend as one, and neither predominates at any time" (Piston, 1955, p. 160). Rimsky-Korsakov (1912) provides several descriptions of blend:

> The strings do not blend so well with the brass, and when the two groups are placed side by side, each is heard too distinctly. . . . Muted strings do not combine so well with wood-wind, as the two tone qualities remain distinct and separate. . . . Uniting plucked strings and percussion with bowed instruments does not produce such a satisfactory blend, both qualities being heard independently. (p. 34)

Erickson's *fused ensemble timbre* includes the concept of blend as one of its meanings. Among the definitions of fused ensemble timbre is "a blend of the contributing elements in which timbral particularity is submerged in the more general sound of the whole" (Erickson, 1975, p. 165). When a chord obtains this quality, "the pitches comprising the chord are difficult to hear out. Individual instrument timbres seem to lose their identifiability" (p. 21). Similarly, Belkin (1988) speaks of combinations where the listener is "unable to distinguish which instrument is playing which note, and consequently hears blend" (p. 50).

The term "homogeneous" is often used to describe the quality of blended combinations: Webster's definition of the word includes the meaning "of uniform structure or composition throughout." Brant (1971), moreover, uses the term "homogeneous orchestration" to describe orchestration that emphasizes the goal of obtaining blends (p. 541). Stiller (1985), describing the problems writing for woodwind ensemble compares the heterogeneity of the woodwinds to the homogeneity of the strings:

> . . . the composer must constantly fight the tendency of this diverse ensemble to fly apart into its component parts

> in an unmediated five-voice wrangle. Contrast this with the string quartet (two violins, viola, cello) in which, because of the homogeneity of the ensemble, the blend is virtually perfect at all pitches. (p. 8)

Blend as a Continuum

One question about the use of blend in orchestration is whether it is utilized in an either-or, bi-polar fashion, or if there is a range of states between blended and segregated. The answer to this question depends in part on the musical style and compositional framework of the composer. Additionally, the question may be addressed from a historical perspective.

Orchestration styles have changed throughout history. Styles emphasizing homogeneous, blended combinations seem to alternate cyclically throughout history with styles emphasizing clarity of parts. Renaissance ensembles were heterogeneous collections of double reeds, plucked and bowed strings, and percussion instruments, but the Baroque period brought on a preference for string tone; this and the demise of loud outdoor instruments such as the shawm the led to ensemble music that emphasized "a highly finished blending technique" (Becker, 1969, p. 15). As new instruments were introduced into the orchestral world (oboe, clarinet, flute) in the eighteenth century and with increasing specializations of instrument families (e.g., English horn, bass clarinet), the trend turned towards heterogeny in the early nineteenth century (Koury, 1985, pp. 85-86). In the later nineteenth century, improved instrument construction resulted in instruments having more evenness of tone, and instrument families having greater tonal consistency across their aggregate range (Koury, 1985, p. 93). This once again promoted homogeneous orchestration, leading to a trend that culminated in the works of Wagner with their gigantic, rich sonorities. Early in our own century, both the "serialists" and the "neo-classicists" turned against this lush, romantic, Wagnerian style of orchestration and emphasized, in its place, transparency and clarity of line. Schoenberg in particular believed that the role of different timbres in the orchestra was "to underline the clarity of the parts, by making it easier for them to stand out from one another" (Schoenberg, 1931, pp. 333-334). His orchestration and that of other serialists who followed him avoided the doubling of instruments, and mixing of colors, "thus achieving maximum individuality of timbre" (Brindle, 1966, p. 127).

This historical synopsis suggests a polar approach to blend: composers either emphasized total homogeneity or heterogeneity. There have also been preferences for blend and segregation on a national basis: German-built woodwinds instruments in the 19th Century are said to have stressed homogeneity of tone and capacity to blend more than did their French- and English-built counterparts (Koury, 1985, pp. 87-88, 91). However, in our present, more pluralistic compositional scene, the desire is not to adhere to one extreme or the other, but to exploit the full range of possibilities. Composers today do not attach a value judgement to either end of the scale, as Ott (1969) observes:

> perhaps we do not even desire a blend between the two instruments. . . . Whether or not homogeneity of tonal qualities is to be sought after in actual orchestration is of

> little significance. In establishing homogeneous blends as a norm, we give ourselves a point of departure from which all other variations of blends are but lessening degrees of homogeneous blend. (p. 53)

In this context, Erickson (1975) observed an exploitation of this continuum in the music of Varèse. A passage characteristic of Varèse is the "explosive chord," such as the one shown in Figure 2. As is typical for such chords, it begins with staggered entrances of notes by several different instruments, which serves to create a highly segregated effect. The instruments each sustain their notes, and gradually a fused sound emerges, creating a single giant sonority. "Varèse is working not only with fused ensemble timbres, but with the whole range between separation and fusion, and movement between these states is fundamental to his art," Erickson states (p. 52). It is interesting to point out here that Varèse also spoke of seeking a "sensation of non-blending" in some of his music (Varèse, 1936).

In summary, the quality of blend appears to be exploitable along a continuum rather than merely in a bi-polar manner. Although certain style periods preferred particular degrees of blend to be operative in orchestrational style in general, music in the late 20th Century frequently exploits the entire continuum.

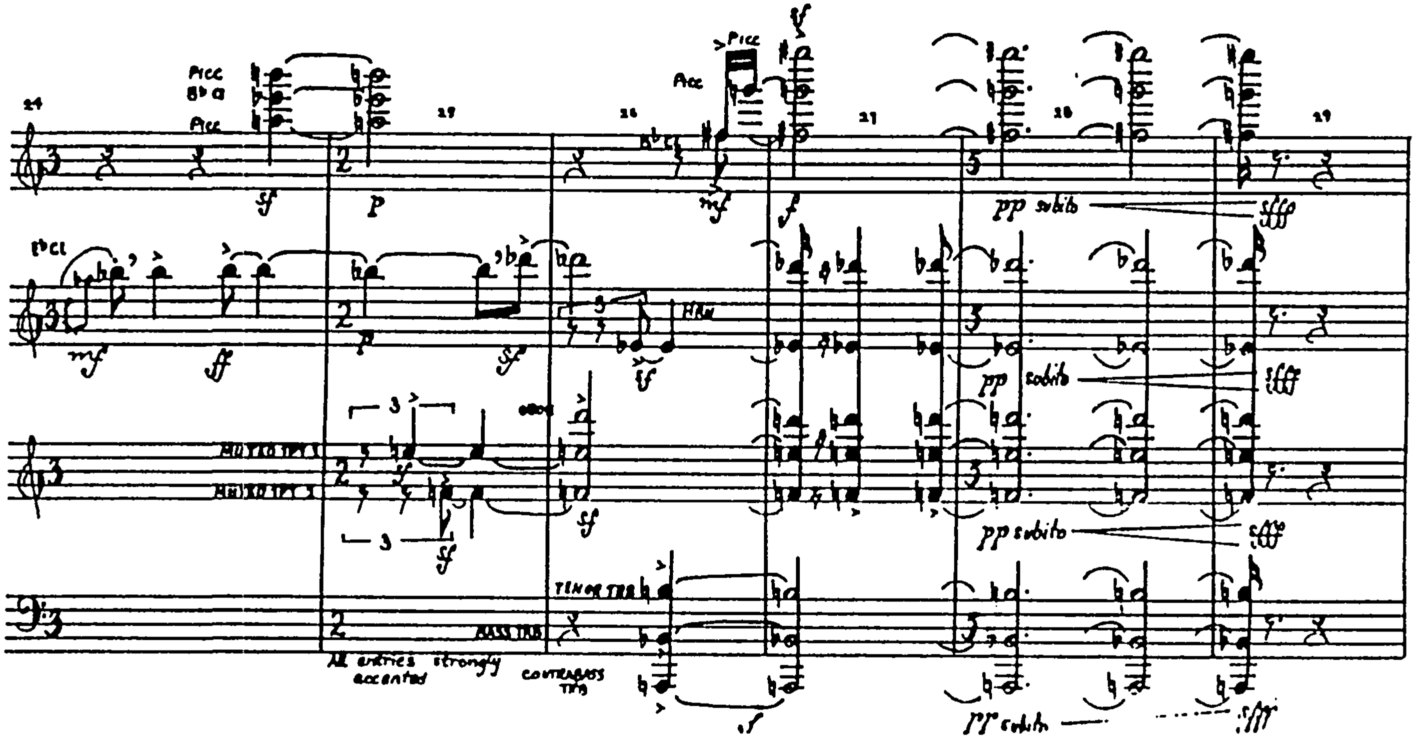


Figure 2. *Intégrales*, Edgard Varèse (1925), measures 24-29 (example reproduced from Erickson, 1975).

"Natural blenders"

The terms "homogeneous" and "heterogeneous" are not only used to describe resultant concurrent timbres (as in Brant's "homogeneous orchestration"), but the general blending capacity of instruments as well.

Orchestration manuals endorse particular groups of instruments--the strings and brasses most frequently--as "naturally homogeneous" (Blatter, 1980, p. 323), and encourage the student to count on them as "good blenders." Factors contributing to this innate capacity appear to be (a) evenness of tone across the pitch compass of the instruments, and (b) the fact that the various members of instruments in each family are sized in such a way to offer a timbral continuum across a wide pitch span.

For example, the reader will recall Stiller's earlier characterization of woodwinds in general as poor blenders (Stiller, 1985, p. 8), which is largely duplicated in observations by Blatter (1980, p. 320) and the following, by Adler (1989):

> The woodwind is perhaps the most quarrelsome of all the families within the orchestra since it is composed of largely heterogeneous instruments. It is even difficult for wind instruments to tune with one another and it takes the finest players to accomplish any kind of balance or blend between these most colorful and diverse orchestra members. (p. 152)

Two instruments which are universally regarded as "good blenders" are French horn and bassoon. Piston (1955) observes that "Horn tone combines well with that of all instruments" (p. 244), as does Riddle (1985, p. 51). Adler (1989) observes that the bassoon, "when accompanied . . . has a tendency to get swallowed up for it blends incredibly well with other instruments, especially in its higher registers" (p. 210). Blatter (1980) observes that bassoon pitches in the range g#2 to c#4 "blend well with almost any instrument or combination of instruments" (p. 112). As for the viola, Rogers (1951) warns that "Its propensity for blending with almost all colors often leads to monotony" (p. 26).

Blending by "Bridging Timbres"

The capacity for certain instruments to act as a sort of "glue" between different groups of instruments is frequently mentioned in orchestration manuals. For example, a group of strings and a group of woodwinds joining together on a single chord, yet occupying different pitch regions (e.g., the strings below and the woodwinds above), will resist blending. However, if certain instruments are carefully placed in the pitch region just between the two groups, they form a "connection," "link," "bridge," or "point of contact" between them, and draw the formerly heterogeneous elements together into a blend. What makes an instrument a particularly good "bridge timbre" appears to be a matter of its similarity to both groups of instruments, and its capacity as a "good blender" in general.

The French horn by far is the most frequently cited bridge instrument (Piston, 1955, pp. 244, 365; Rogers, 1951, p. 104; Rimsky-Korsakov, 1912, pp. 24, 35). Rimsky-Korsakov (1912) praises its ability to link brass and woodwind because of its strong similarity to the bassoon (p. 24). Rogers (1951) points out that the horns "provide a fine inner binding" (p. 104), and Piston (1955) offers:

> For many years, from the mid-nineteenth century on, it was a common habit of composers to use the four-part harmony of horns as a foundation for orchestral writing. This "blanket of horns," . . . could be so unobtrusive as to be unnoticed by the lay listener, and it offered a sure means of obtaining continuity and fullness of sound. (p. 244)

Other instruments described as providing bridge timbres are mentioned throughout orchestration manuals. Some of these are:

1. Bassoon bridging woodwind and brass (Rimsky-Korsakov, 1912, p. 35)
2. String harmonics bridging strings and woodwinds, because of their similarity with the flute (Rimsky-Korsakov, 1912, p. 10)
3. Viola and bassoon bridging strings and woodwind, since both are similar to clarinet ((Rimsky-Korsakov, 1912, p. 34)
4. Saxophone bridging winds and strings (Rogers, 1951, p. 125) or bridging French horns and woodwinds (Forsyth, 1914, p. 169)

Factors Affecting Blend

Orchestration manuals are unusually rich in information regarding particular features that play roles in determining blend. Some of these factors are: attack synchrony, spectrum, dynamic level, pitch height, pitch proximity, chord voicing, number of timbres, spatial position, and performers' effects. Such information provides clues as to the acoustical principles underlaying blend, the discovery of which is important for satisfying the explainability. Each of these factors are considered individually below.

Synchronicity of Attack

The fact that instruments must begin and end together in order to obtain a blend practically goes without saying. Erickson (1975) observes that "when individual instruments begin and end together fusion is more likely" (p. 46), as does Belkin (1988, p. 50). Schillinger (1941) stresses the importance of combining instruments of similar articulatory possibilities for obtaining blends (pp. 1598-1599). Fusion of chords containing instruments with a diverse set of attack and articulatory types can be overcome, however, by making attacks softer, as Erickson (1975) points out in a remark on Webern's *Five Pieces for Chamber Orchestra,* op. 10: "The dynamic marking of *PPP* [*sic*] certainly enhances the mixing of the sounds by minimizing the effects of the attack transients---they tend to mask one another." (p. 166)

There are two reasons why attack properties affect blend so significantly. First of all, human temporal acuity for detecting differences in onsets between events is as small as 2 ms (see Green, 1976),[13] an ability that has important implications for musical audition (Rasch, 1978, 1979). Unless instruments attack with great onset synchrony, cues that there are multiple

---

[13] The abbreviation *ms* stands for *milliseconds.*

sources present will become salient, and limit the likelihood that the ensemble will be perceived as fused. Second, since attack onsets play an important role in conveying the identity of an instrument (Stumpf, 1890; Clark, Luce, Abrams, Schlossberg and Rowe, 1963; Saldanha & Corso, 1964; Berger, 1964; Wedin & Goude, 1972; Elliott, 1975), asynchronous attacks will provide listeners significant cues as to the makeup of the ensemble, further increasing their awareness of the multiplicity of sources.

Spectrum

It will be noted that many of the instruments that are most frequently cited as "good blenders"-- French horn, bassoon, cello, bass clarinet--are all relatively "dark" in spectrum. That is, they have low centroids. There is also the implication that instruments that don't blend well tend to be the "brighter" varieties: woodwinds, many of which are very bright (clarinets and oboes especially), have been characterized by several authors as being poor blenders in general. It might be concluded, on this basis, that there is a tendency for darker timbres to blend well. Evidence which supports this conclusion is found in Riddle's remarks on the use of mutes in string sections: "one of the wonderful aspects of mutes---the blend, whether it be a small section or a large one, improves magically with their use" (Riddle, 1985, p. 124). Mutes on string instruments to a certain degree do cause an emphasis on the lower resonances of the instrument (Meyer, 1978, p. 64).

Dynamic Level

Soft dynamics in general can lead to greater blends: recommendations that certain instruments be counted upon to obtain blends are often qualified with the requirement that they be marked *mezzo-forte* or softer (Blatter, 1980, p. 321; Piston, 1955, p. 427; Adler, 1989, p. 328; Rimsky-Korsakov, 1912, p. 24; Rogers, 1951, p. 4). Obviously, a softer instrument tends to be absorbed by the sounds of others, and thus be less noticeable, while a loud instrument tends to protrude from the texture. Piston (1955) criticizes the ability of the saxophone to blend on this basis, claiming it to be intrinsically louder than other instruments (p. 206). It is sometimes suggested that the level of all instruments in a combination be marked low to encourage blend (see Erickson's remark on onsets above). Possibly this is because the spectral makeup of instruments tends to become darker as they are played more softly (Clark & Milner, 1964; Meyer, 1978, pp. 30, 34), and darker timbres may blend better than brighter ones (see above).

Pitch Height and Proximity

Instruments playing notes that are far apart in pitch will not blend as well as those that are close (Berlioz, 1844, p. 34; Ott, 1969, p. 53). As Blatter (1980) points out, "The more octaves between the instruments, the more the result will sound like separate instruments and not like one integrated tone." (p. 295) Another recommendation relating to pitch choice in combined timbres is to avoid extreme registers. Rogers (1951) observes that the choice of any extreme register leads to poorer blends (p. 50), while Stiller (1980) emphasizes that instruments tend to sound most distinctive in their low register, so combinations including these notes may not blend well (p. 8).

Chord Voicings

Possibly the most celebrated piece of "orchestration lore" is rule of thumb to follow the *natural order of register* (i.e. the order of instrument placement in the score) when constructing chords. According to this rule, woodwind chords should be scored, from low to high pitches, bassoon, clarinet, oboe, and flute, while strings are scored in the order contrabass, cello, viola, and violin. And furthermore, strings should more or less lay below the brass, which in turn should lay below the woodwinds. This method is endorsed as a way of promoting the blend of harmonies (Rimsky-Korsakov, 1912, pp. 45, 47, 52, 56; Piston, 1955, p. 423). Piston (1955), however, urges that the student not follow this advice dogmatically (p. 423); Strauss, specifically, recommends reversing the position of oboe and clarinet (Berlioz-Strauss, 1945, p. 213). In any case, the heuristic is somewhat suspect since the ordering of instruments in a score largely reflects their historical arrival as members of the orchestra (Forsyth, 1914, p. 263), rather than a convention for selecting instruments.

An equally well-known rule, borrowed from harmonic theory, is that chord voicings should observe *harmonic series spacing,* that is, wider intervals at the bottom than the top. This too will promote their blend: orchestration manuals observe that chords with large gaps between notes in the upper register tend to blend poorly (Stiller, 1985, p. 8; Rimsky-Korsakov, 1912, p. 68), and the effect is even worse at louder dynamics (Rimsky-Korsakov, 1912, p. 78). Similarly, notes in close voicing in the lower register will sound muddy, possibly due to the increased tonal "roughness" in that register (Plomp & Levelt, 1965).

A more specific voicing which is known to promote blend between different instruments is the "interlocking" arrangement. This usually refers to the use of two sets of same-instrument pairs, such as oboe-oboe and clarinet-clarinet, arranged in a four-part chord such that no instrument is adjacent in pitch to its duplicate, as Figure 3 illustrates. The blend of instruments that are otherwise very heterogeneous can be improved a great deal by this voicing (Riddle, 1985, pp. 81, 137; Rogers, 1951, pp. 52, 101, 104; Forsyth, 1914, p. 262). Belkin (1988) offers an interesting explanation of the perceptual process underlying the improvement. He suggests that in such cases, the ear confuses pitch with timbre: "the ear, while perceiving a certain fullness of color, is unable to distinguish which instrument is playing which

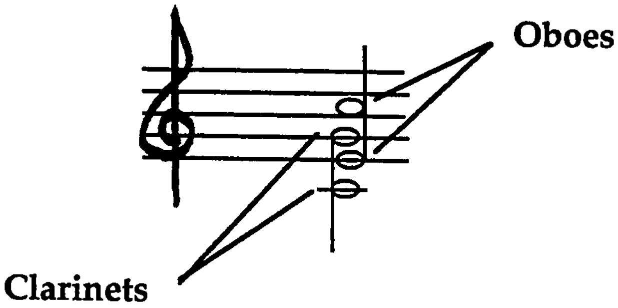


Figure 3. Example of an interlocking voicing of a chord.

note, and consequently hears blend" (p. 50). One might call this mechanism "blend by misattribution of pitch to timbre."[14]

Number of timbres

Yet another obvious deterrent to blend is the number of different instruments (and therefore timbres) involved in the combination. Although "bridge timbre" techniques are good at uniting otherwise heterogeneous groups, they will not apply when the timbres in a chord are many and different. As Adler (1989) remarks, chords "where each note has a different timbre . . . are difficult to balance" (p. 241). Robert Schumann's orchestration is said to be flawed in part due to this problem (Walsh, 1972). Similarly, Blatter (1980) states,

> If one is, for example, attempting to create a chord with an overtone series voicing, but assigned assertive or especially colorful voices to a unique assortment of pitches, then those pitches will be more audible and could thereby unexpectedly alter the final result. (p. 300)

Spatial Position

The closeness in space from which two or more sounds arise can play an important role in the degree to which they blend. In this regard, blend can be promoted or limited depending upon orchestral seating arrangements and/or the size of the ensemble. The capacity for a flute to blend with a

---

[14] Studies addressing the perceptual interactions between pitch and timbre are Plomp and Steeneken (1971), Singh (1987), Melara and Marks (1990), and Krumhansl & Iverson (in press).

contrabass, for example, will differ widely between the orchestral case, where the two instruments are separated by several meters, and the chamber music case, where the two performers may be seated very closely. Composers in the Twentieth Century have increasingly included seating specifications in their published scores in order to eliminate the randomness of such matters. The seating indications for Bartok's *Music for Strings, Percussion and Celeste,* seems to be chosen to emphasize the polychoral (and hence, non-fused) relationship between the two bodies of strings. Meyer (1978, pp. 137-141), additionally, points out the differences in sound between orchestral performances using 19th Century European seating (Violins I in front and on the conductor's left, and Violins II in front and on the conductor's right) and modern American seating (Violins II moved to the conductor's left and behind the Violins I).

Brant (1967) gives an example of how the perception of a timbral combination can be altered by its spatial presentation. In the passage below he suggests how a unison between different timbres that results in a "poor and confusing tone quality . . . [which is] disturbing to the overall harmonic effect" can be improved by a change in spatial location; reasoning by analogue, this example be applied to the subject of blend as well.

> This impression will be strongly felt if the sound comes from the same source and direction, as when all the performers are placed close together on one stage. If these same textures are now disentangled by distributing their respective performing groups into widely separated positions in the hall, the unisons occurring between the contrasted textures are no longer perceived, because the groups at this distance can no longer make harmonic

> contact between the tones that they simultaneously sound, and their respective tone qualities are now so diffused that no connection between them can impress itself on the listener. (p. 224)

Performers' Contributions

Finally, the importance of the role of the performers and the conductor in obtaining blends (or any other effect) can never be discounted. Effective orchestration of any type is of course not simply a matter of effective instrument selection or wise dynamics; there must be an

> awareness on the part of the players, a sensitivity as to phrasing and intonation, and a real desire to help the mixture "come off." They must listen keenly to what is going on around them and be able to fit the sound of their instruments into the general blend. (Riddle, 1985, p. 71)

The conductor Hermann Scherchen (1933) illustrated a situation in which performers might promote blend by adjusting to one another and moderating their tone: "All four woodwinds can colour a harmonic passage to blend with the tone-colour of the flute, the oboe, the clarinet, or the bassoon" (p. 117). Scherchen cited Schoenberg's *Farben* as a composition where this may be useful (p. 93).

Alternative Definitions of Blend

The definition of blend that has been employed throughout this discussion can be summarized as "fused, not separated." However, it should be noted that this definition is not entirely universal, as other authors offer alternative definitions of blend. Some of these will be considered briefly here.

One alternative definition is suggested by Belkin (1988):

> An important reason for blend is richness: it is often aurally more interesting to perceive two separate sounds intertwined than one alone. . . . Blend implies the mixing of distinct entities; if the sound lose all distinctness, they will no longer be heard as blended, but rather will be perceived as a unit. (p. 50)

Apparently, Belkin's "blend" is one which the individual identities of tones are not suppressed, but one in which they are still joined together in some interesting interaction. The important requirement seems that they be "interesting" in some undefined way is reminiscent of one of Erickson's descriptions of *fused ensemble timbre*: "an unexpected, or striking or otherwise memorable fused sound . . . in the perceptual foreground" (Erickson, 1975, p. 47).

There are also some indications for yet another interpretation of "blend," in the writings of jazz orchestrators. A few writings suggest an interpretation of blend as "separated, not fused"--the exact opposite of the usual definition. One might think that the jazz saxophone's sonic signature, with its strongly "chiffed" attack (see Grey & Gordon, 1978, p. 1499) and its characteristic pitch and intonation inflections would fairly insure its tendency to not blend with instruments of other families. This view is held only by "classical" orchestration manuals, however (e.g. Piston, 1955, pp. 186, 206); jazz orchestrators hold the opposite view, considering them to be "excellent mixer[s]" which blend "well with all instruments" (Russo, 1969, pp. 574 and 579, respectively). Similarly, "classical" orchestration teaching, which values timbral similarity as a guide to good (i.e., blended) combinations, is clearly on a collision course with Russo's recommendation that "The best and most colorisitically valuable pairing is obtained between two dissimilar instruments, at least one of which is highly colored by itself (the oboe) or used in a highly colored form (trumpet in Harmon mute)" (Russo, 1969, p. 562).

Evaluation

Concurrent timbre is clearly a topic of great interest to musicians, and a topic that occupies a central position in the teaching of orchestration. Much effort is devoted to trying to characterize and describe the sounds of concurrent timbres, and various prescriptions are suggested to obtain particular effects. This makes the motivation for studying concurrent timbre much stronger than the previously discussed topics (semantic attributes and strength), since the relevancy criterion is so well satisfied.

Several specific techniques or sonic objectives pertaining to concurrent timbre were surveyed: augmenting, softening, inventing, imitating and blending timbres. The first four are all techniques relevant to orchestration and applicable to a variety of different situations (revelancy criterion). Each defines a sonic effect fairly clearly (definability criterion); however, "augmenting timbres" depends mostly on semantic descriptors (e.g., the cello adds "richness" to the clarinet), which are only vaguely accountable for by acoustical properties (demonstrability criterion), a problem encountered earlier.

The phenomenon of "softening," however, deserves a closer look. This effect, described as a change in the perception of timbral quality due to the addition of a new element, strongly suggests principles of auditory masking (see Jeffress, 1970), and thereby invites the use of psychophysical methods for investigating tone-on-tone masking. Other masking literature, concerning the detection of tones in noise, might be applicable to the cases where flute acts as a "softener," since relatively high breath-noise levels accompany flute tones (Meyer, 1978, p. 49). Similar kinds of timbral modification relating to more advanced aspects of masking, such as auditory suppression (Shannon, 1976) and comodulated masking release (Hall, 1987), might be found to underlay the perception of concurrent timbres as well.

A significant drawback, however, is that there are few precedents for studying the masking of one temporally evolving complex sound by another: existing methodologies in masking pertain to masking of sine tones by noise, tone-on-tone masking, or at best, masking of speech by noise. But even in these contexts masking research involves detailed, low-level investigations. A research project addressing tone-on-tone masking of time-variant timbres would have to invest a great deal of time-consuming methodological groundwork, which would necessarily limit the investigation at first to only the lowest level phenomena. The results of such a study, even if desirable, would only remotely be applicable to any practical aspect of orchestration. Such a research project is outside the scope of this study.

A glance at the survey shows a large amount of evidence in support of the perceptual investigation of blend. The sonic objective is clearly defined (that combined instruments fuse into a composite sound), and it comes to bear on a variety of concurrent timbre situations: that is, in the techniques of augmenting timbres, inventing timbres, bridging timbres, and so on, the success of the effect depends a great deal on blend. Orchestration manuals offer generous information on the way in which the specific attributes of pitch, brightness, dynamics, and attack affect blend. In the following chapter, a great deal of information from the literature of auditory perception offers clues as to what makes instruments blend. The various instrumental combinations that orchestration manuals describe as blending "well," "not so well," and so on, suggest that blend can be evaluated along a continuum rather than on a mere accept-reject basis. All of these observations show that the criteria of definability, explainability, demonstrability and continuity are well-satisfied. These criteria provide some insurance that blend is an attribute that can be generalized, rather than described solely in terms of specific instances. This in turn suggests that investigating blend with perceptual experimentation should be fruitful.

Musical Examples

To show that the relevancy criterion is satisfactorily met, a few musical examples in which blend is central attribute will be examined briefly to underscore both the interestingness and relevance of the subject of blend.

The first shows an orchestral unison melody found in Wagner's *Parsifal* Prelude (Figure 4). The composer has ingeniously brought instruments in and out at various times to take advantage of their properties in various registers, with the objective of obtaining a smooth timbral quality throughout the arching line. The passage is dominated largely by cello tone, due to the position of the melody up on the A-string throughout; the violins, on their lower, less penetrating strings merely add support to this tone. The bassoon is removed as the pitch range approaches its "pinched" upper register, and then resumes once again. A similar strategy explains the entrance of the English horn, whose "honking" lower register is avoided. At the topmost arch of the melody, on the neighboring figure B-C-B, two interesting things occur: the Bb clarinet crosses over into its clarion register, lending a brighter quality on those notes, and the bright color of the oboe enters to add further emphasis. Another way of looking at the role of the oboe is as a continuation of the color of the English horn: Wagner has maintained the oboe-type color throughout, but has carefully allowed it to be represented by instruments playing in their best registers for the pitches in

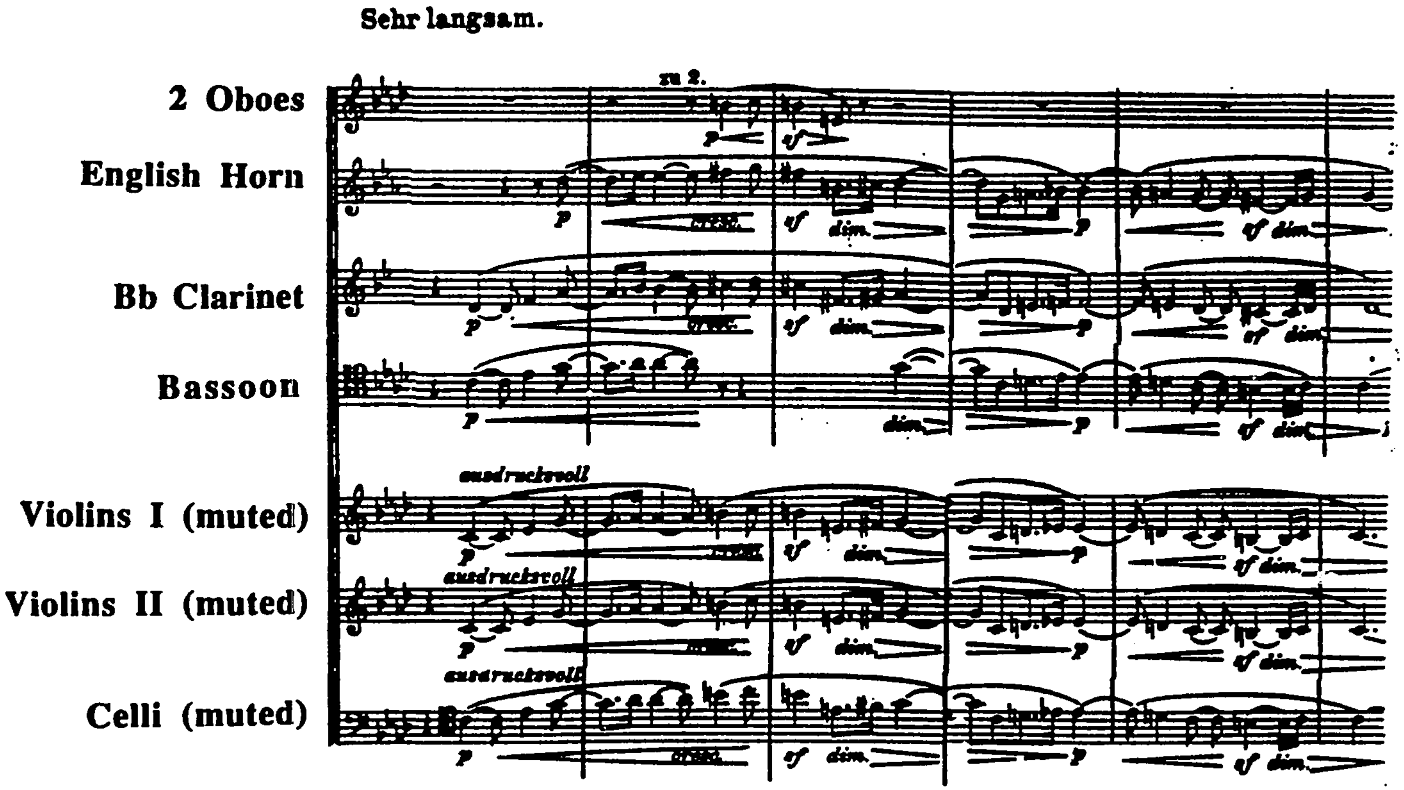



Figure 4. *Parsifal,* Richard Wagner (1892), measures 20-25.

question.[15] When performed competently, the effect yielded by the passage is not one of multiple colors--in fact, the listener will barely be aware of the presence of any of the wind instruments--but rather, a quality that is held constant despite the melody's changes in register and dynamics.

A passage from Schoenberg's *Five Pieces for Orchestra, Op. 16* (Figure 5) also projects a unison melodic line (shown in notation below the score), but here with a highly expressive, changing coloration. Embedded in the texture is a sequence of pitches orchestrated with various concurrent timbre pairs (cello-trombone, cello-flute, cello-oboe-flute, and trombone-English horn-flute). The use of the oboe to shade the expressive semitone neighbor A-Bb-A is strikingly reminiscent of the *Parsifal* example.

The example by Knussen (Figure 6) shows a portion of a work in which blend is applied on a far larger scale: the first forty measures consist entirely of a unison whose orchestration is in constant flux. In these three measures alone, the string parts carry the line continuously, but various brass and woodwind instruments are added sporadically to add timbral accentuation. The composer's own annotation at the bottom of the score[16] suggests that these accentuations should fuse with, rather than protrude from the overall tone of the strings.

---

[15] Maintaining timbral quality in a melodic line with instrumental successions such as these may be likened to a "relay race." Piston (1955) suggests using this approach as well for melodic lines that fall out the comfortable playing range of a single instrument (p. 422).

[16] "N.B. As far as possible, string timbre should dominate the unison passage m. 1-40."

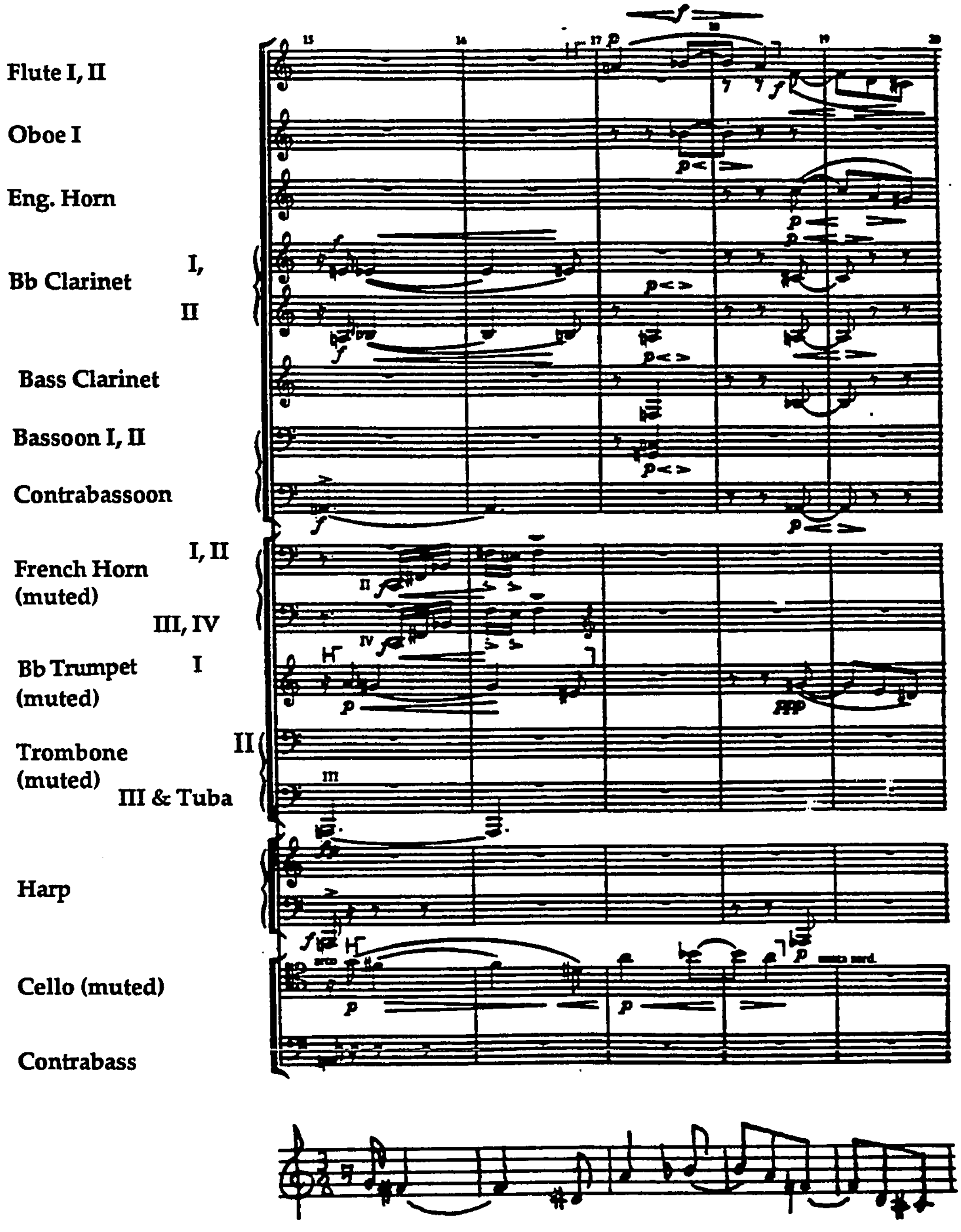


Figure 5. *Five Pieces for Orchestra, Op. 16,* Arnold Schoenberg (1949 version), no. 1, measures 15-19.

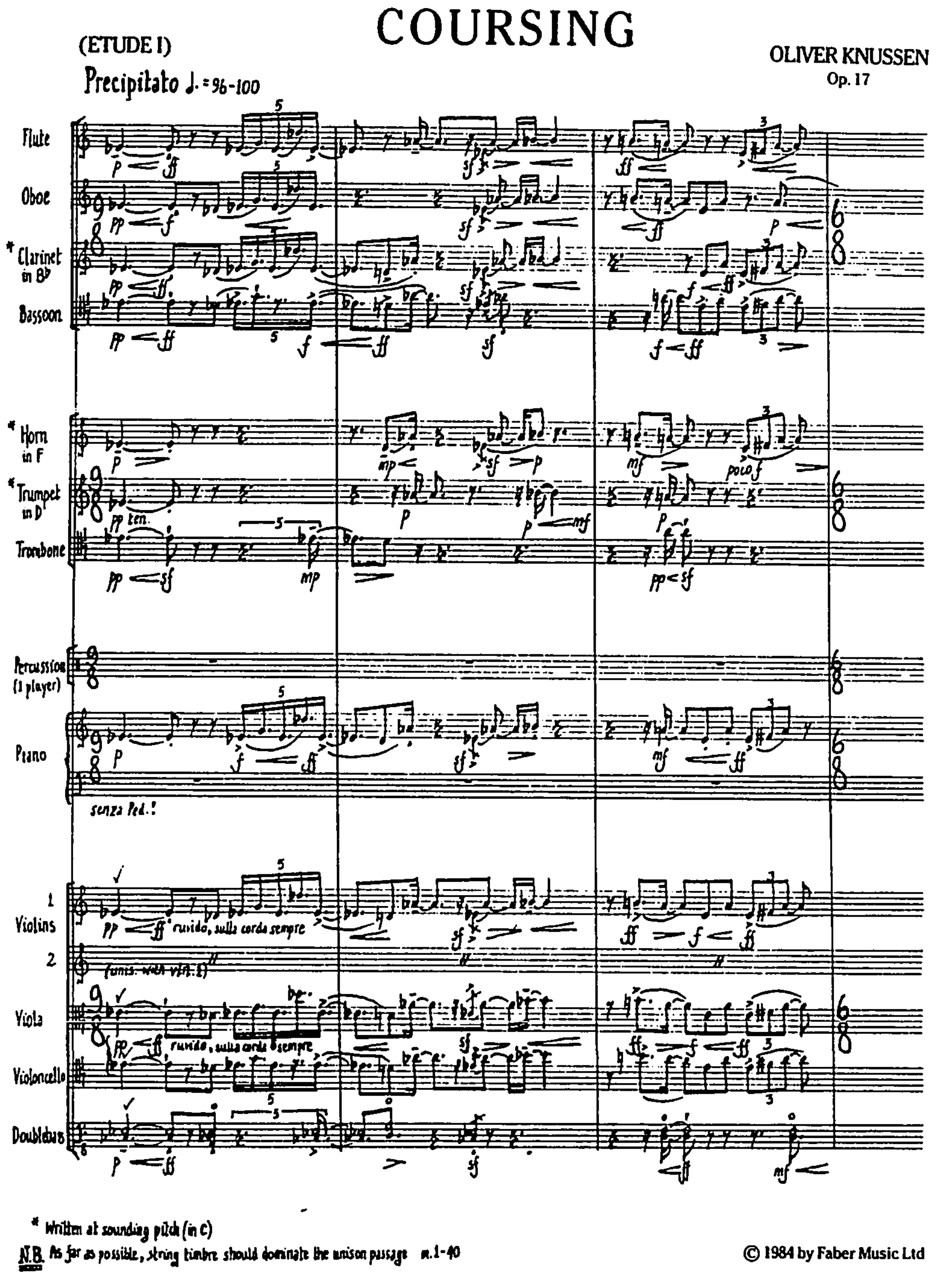



Figure 6. *Coursing, Op. 17,* Oliver Knussen (1979), measures 1-3.

Conclusion

Blatter (1980) endorses blend as a primary topic of orchestration when he writes, "One of the chief goals of the orchestrator is to mix, blend, match, and contrast the instrumental and vocal colors at his disposal" (p. 292). Similarly, Albert Bregman speculated recently on what a "theory of orchestration" might provide. Interestingly, he identified blend-related issues as the initial concern for such an enterprise:

> I can imagine a description of orchestration that addresses issues such as the following: How can you make sounds fuse to create a single timbre? Which qualities of the individual sounds will contribute which aspects of this global timbre? How can multiple, simultaneous timbres be kept perceptually distinct from one another? What are the acoustic factors that are responsible for all these effects? (Bregman, 1990, p. 458)

Indeed, "Blend" has been demonstrated to be the best candidate for the focus of this study. The study will proceed by examining the literature of psychoacoustics and auditory perception for cues as to how blend may profitably be studied through experimentation.

Chapter 2:

Perceptual Research and "Blend"

The previous chapter reviewed the observations on blend offered in orchestration manuals. This chapter will now examine the literature of auditory perceptual research pertaining to blend. No previously existing literature directly addresses concerns of musical orchestration, so by necessity a diverse set of sources must be sampled. Information acquired here will offer some perspectives for explaining how blend is attained and help determine appropriate ways of investigating the phenomenon experimentally.

Auditory Scene Analysis

A large body of literature known variously as "auditory stream segregation" or "auditory fusion," and largely associated with the work of Albert Bregman and his colleagues, is relevant to orchestral blend. Following the title of Bregman's recent book (Bregman, 1990) we may refer to this work in general as "Auditory Scene Analysis." The general features of Auditory

Scene Analysis that are of interest to the present study are defined in the following passage from McAdams (1984):

> The auditory system participates in the forming of images evoked by acoustic phenomena around us. An important aspect of the imaging process is the distinguishing of different sound sources. In order to be able to form images of sounds in the environment the auditory system must be able to decide which sound elements belong together, or come from the same source, and which elements come from different sources. (p. 4)

Many different levels of auditory perception are encompassed by this paradigm. One level of application concerns the conditions which lead to the perception of a single source: for example, how the mass of harmonic partials comprising an oboe--partials that may be dynamically changing in frequency and amplitude--becomes fused in the listener's ear as the object "oboe" rather than as a group of individually-heard sine tones. For the most part, however, this aspect of Auditory Scene Analysis falls outside the concern of the present study, since orchestral instruments in isolation involve little uncertainty as to whether they fuse as sonic objects.

On a higher level, however, Auditory Scene Analysis is concerned with how the auditory system shapes and organizes the perception of multiple sounding objects to which the listeners can then attend. A well-known example is the "cocktail party" effect: despite the fact that the acoustic signal reaching the ears includes spectral and temporal information pertaining to perhaps a dozen different voices, the auditory system can successfully attribute each voice to a different source and make out the message contained in any one of them. This aspect is far more relevant to the present study.

The auditory system, by default, looks for ways to segregate events that can plausibly be attributed to different sources so that the listener can attend to each individually. In some special situations, however, sounds which come from multiple sources fuse and are apprehended as arising from one source. For example, in music, chords will often blend into what Bregman (1990) calls "vertical grouping . . . in which the chord, rather than the individual tones, becomes the acoustic object that we hear" (p. 496). Ordinarily, the auditory system regards multiple pitches as strong evidence for the existence of multiple sources, so the phenomenon of pitches fusing into chords (i.e., sounding as a single object) can be thought of as an auditory illusion. Bregman (1990) provides an insightful characterization of this illusion as a "chimeric grouping" of sound:

> The Chimaera was a beast in Greek mythology with the head of a lion, the body of a goat, and the tail of a serpent. We use the word chimera metaphorically to refer to an image derived as a composition of other images. . . . Natural hearing tries to avoid chimeric percepts, but music often tries to create them. It may want the listener to accept the simultaneous roll of the drum, clash of the cymbal, and brief pulse of noise from the woodwinds as a single coherent event with its own striking emergent properties. The sound is chimeric in the sense that it does not belong to any single environmental object.
>
> To avoid chimeras the auditory system utilizes the correlations that normally hold between acoustic components that derive from a single source and the independence that usually exists between the sensory

> effects of different sources. Frequently orchestration is called upon to oppose these tendencies and force the auditory system to create chimeras. A composer may want the listener to group the sounds from different instruments and hear this grouping as a sound with its own emergent properties. (pp. 459-460)

It is this auditory illusion that underlies "blend." Although the majority of Auditory Scene Analysis research is applied to problems on a level just below the question of "chimeras"--namely, how single objects are fused from collections of harmonics--a few studies may be identified that address blend phenomena directly or indirectly. Studies pertaining to these paradigms will be discussed below, organized according to three topics: factors relating to pitch perception; onset synchrony; and temporal similarity.[17]

Factors Relating to Pitch Perception

One of the most robust mechanisms that shapes the segregation of sources is pitch perception. There are several theoretical models of the mechanism of pitch perception (see Bregman, 1990, pp. 235-237, for a review); the "pattern recognition" framework in particular is useful for addressing problems of Auditory Scene Analysis. In this approach the auditory system compares what it is hearing to various internal "templates" that correspond to all possible complex tones having harmonic relationships to one another. The auditory system accepts or rejects various collections of harmonics depending on the strength of the match to a template, and those that are

---

[17] For the remainder of this study, the reader should assume that the term "blended" is synonymous with "fused," and that the term "separated" is synonymous with "segregated."

accepted cohere into a global pitch. As a result of experience with sounds in the environment, the listener is likely to attribute the sounds to a single object, or source. The auditory system is strongly biased to make such attributions when collections of harmonically-related partials are present.

Pitch Separation

When the auditory system encounters simultaneous complex tones, it is very good at recognizing it as separate sources, and deterring the perception of a fused single source, or "chord." It is assumed that the system performs a pitch analysis on each (Houtsma & Beerends, 1986), and when the evidence is clear, is generally be successful at attributing it to separate sources. However, various factors can weaken this mechanism.

The system seems to require some separation in fundamental frequency for the mechanism to work effectively. One way this has been investigated is by asking listeners to identify individual objects in a mixture of several sounds at various pitch levels. The assumption of this paradigm is that the harder it is to identify individual objects in a mixture of sounds, the better the blend; conversely, the recognition of any one sound in a combination is taken to mean that it has segregated from the rest of the items in the mixture. This paradigm might be called "segregation by identification."

Scheffers (1983) explored recognition of concurrently-sounding vowel identities (two different Dutch vowels sounding simultaneously) and found that listeners could identify them easily with one to three semitones of pitch separation (described in McAdams, 1989, p. 2157). Similarly, Halikia (1985) found that improvement for concurrent vowel recognition with interval size reached asymptote with as little separation as 0.5 semitones (described in Bregman, 1990, p. 562). Stern (1972) ran a similar study using synthesized musical instruments. He presented them in concurrently-sounding pairs to listeners who were asked to identify them by name. His results showed that identifying the constituent instruments making up a unison interval was difficult, but instruments at other intervals (perfect fourth, major third, minor third, and minor second) could be recognized easily. The combination of a trumpet and clarinet at the unison, for example, "sounded like neither trumpet nor clarinet, nor like a predictable mixture of their qualities" (p. 209); however, at non-unison intervals the individual qualities of both trumpet and clarinet were apparent.

Apparently, then, the auditory system has difficulty separating sources having the same fundamental, but not those at the musical interval of a half-step or larger. This suggests that, in terms of pitch separation alone, only the unison has an innate tendency to fuse or blend.

Inharmonicity

Departures from purely harmonic relationships between partials can weaken the pitch detecting mechanism and create uncertainty in the source-segregation process (Bregman, 1990, pp. 243-244). If a pattern of partials is slightly inharmonic, the auditory system finds no single template to be the best match, but several templates making inexact matches. With more than one candidate pitch, the effect yielded presents a less clear picture of how many sources are present (McAdams, 1984, pp. 41-42), and the ear may regard combinations including such sounds as blended. The assumption behind this hypothesis is that if the system does not find sharply-defined separate elements in a sound, then the sound is blended. This paradigm might be termed "blend by indistinct numerosity."

Harmonic Coincidence

Harmonic coincidence refers to the number of instances of common partials between two complex sounds. The number of common partials may be calculated by the ratio of the fundamental frequencies of the complexes. For example, in a pair of tones in the ratio of 2:1 (an octave), every second partial of the lower tone will be contained in the upper tone; in 3:2 (a perfect fifth), every third partial in the lower tone will be found as every second partial in the upper tone. The larger the amount of coincidence, the less the system's pitch processing mechanisms finds evidence for multiple, independent tones, thus the pitches fuse into a single object (McAdams, 1984, p. 44; Bregman, 1990, pp. 246-247). This explains why octaves and fifths, in certain contexts (and depending on the tuning system) can easily be mistaken for single notes, because of their large coincidence of partials. Of course, this challenges the view that the unison is the only interval that tends to fuse intrinsically.

Stumpf (1890) investigated the fusion of various intervals by playing listeners concurrent tones at the octave, fifth, fourth, major third, tritone or major second, asking them to rate whether they heard one or two tones. The percentage of listeners who (wrongly) judged the intervals as being only one tone was directly related to the tones' harmonic coincidence: greater harmonic coincidence led to more instances of false perception of single tones. The percentages (given in Apel, 1972, p. 202) were:

8ve (2:1)	*5th* (3:2)	*4th* (4:3)	*3rd* (5:4)	*tritone* (7:5)	*2nd* (8:7)
75%	50%	33%	25%	20%	10%

Dewitt and Crowder (1987) re-ran Stumpf's experiment to investigate the fusion of various intervals. In addition to obtaining percentage correct judgments of one vs. two tones, they measured the reaction times of the responses as well. They also made the task slightly more complex: listeners had to judge if one or two tones were *added* to an already sounding tone. Their results showed that the judgments for intervals with simpler ratios (octave and perfect fifth) confirmed Stumpf's findings: they were judged as being one tone, and produced increased response latencies. However, the results for the remaining intervals they investigated (major seventh, tritone, perfect fourth and minor second) did not closely correlate to the complexity of the ratios as they did in Stumpf's data. For example, in their data the major seventh (11:6) was incorrectly judged a single tone more often than was the perfect fourth (4:3). It would appear then, that harmonic coincidence as a predictor of fusion is reliable only for intervals with large amounts of coincidence.

Stumpf (1890) regarded consonance as the equivalent to fusion: the degree to which two tones fused into a single image was a direct measure of their consonance. Conversely, two tones which did not fuse was regarded as a measure of their dissonance. One possible explanation for this is that the auditory system is accustomed to hearing interference patterns (beats) when two tones are sounding together; consequently, interference patterns are a familiar cue for "twoness," which in turn creates the sense of segregation. The greater the dissonance, then, the greater the sense of twoness, or segregation.

Recently Bregman (1990) questioned this assumption that dissonance and segregation were co-occurring phenomena. *Roughness,* the acoustical beats that arise from interference patterns between closely-tuned sine tones is sometimes used as a cue for dissonance, and sometimes not, Bregman hypothesizes (pp. 508-509). In other words, in certain situations listeners will hear beats arising from partials but discard that information in forming their impression of the quality of a combination; namely, when the two elements in question are in separate streams (for example, if they arise from different spatial locations; see Brant, 1967). In such cases the listener will not register the roughness cues and regard the combination, for example, as an accident of counterpoint rather than an intended simultaneity.[18] Only if the two events are in the same stream will the combination be perceived as dissonant. Bregman's opinion, then, in marked contrast to Stumpf's, is that dissonance is a percept *arising* out of the fusion of tones, not a cue that the tones are segregated.

---

[18] Bregman (1990) later presents evidence that this dichotomy underlies the rules behind Renaissance counterpoint as well (pp. 511-514; see also Wright and Bregman, 1987).

Onset Synchrony

One of the most powerful cues for either promoting or preventing blend of concurrent sounds is synchronicity of onset. Synchronizing the onsets for a group of sounds can often override other cues promoting segregation (e.g. pitch separation) and bring about fusion (Bregman, 1990, p. 263). It will be recalled that authors of orchestration manuals demonstrated a great awareness of the necessity for synchronized onsets for obtaining blend as well.

One important study in this area is by Rasch (1978), who investigated how sounds segregate from one another as a result of their onset synchrony. His experimental stimuli consisted of a two-voiced musical example, lasting two beats, with a low voice playing the same pitch on both beats and an upper voice leaping either up or down a fifth; the listener reported whether he heard the interval jump up or down. The amplitude level of the upper voice was low relative to the lower voice, so that the task of recognizing the direction of the jump was impeded by the masking of the upper tones by the lower ones. When both voices were synchronous, the listeners could maintain at least 75% accuracy at detecting interval direction with the upper voice being about 12.5 dB down in level from the lower voice.[19] When onsets were made asynchronous, the listeners were able to maintain 75% accuracy

---

[19] The use of 75% accuracy as a baseline is a convention of auditory masking studies. If a listener shows less than 75% accuracy in correctly detecting the presence of a tone in noise, then that tone is considered "masked" by the noise.

with larger level differences: -32 dB, -50 dB, and -64 dB for disparities of 10 ms, 20 ms and 30 ms, respectively. In other words, greater onset differences made it easier to distinguish the voices, and therefore decreased their degree of blend. Interestingly, the listeners did *not* seem aware that onset disparities were what were responsible for making the direction of the leap more recognizable, which suggests that the *cues* for blend may not actually enter into the conscious mind of the listener; rather, the auditory system only registers whether or not the tones blend.

Stern (1972) also showed that recognition of concurrently-presented pairs of tones could be improved with onset asynchronies. He used synthesized versions of musical instruments. As was discussed above, Stern's listeners easily recognized the constituent instruments at non-unison interval presentations, but had difficulty with unison intervals. Recognition for one of the two instruments in unison presentations could be improved if it was begun slightly earlier than the other, however.[20]

"Common Fate" Mechanisms

A number of the basic paradigms in Auditory Scene Analysis research have their roots in the grouping principles of the Gestalt psychologists. The principle which comes to bear on fusion and blend is "common fate," which

---

[20] Stern does not say what the duration of the disparity was, although he implies that it was something less than 100 ms.

says simply that sounds that change in similar ways are likely to have originated from the same source (Bregman, 1990, pp. 248-292).

Bregman (1990) offers a compelling example of common fate in a vision perception context:

> Let us imagine that we had a photograph taken of the sky. It shows a large number of birds in flight. Because they are all facing the same direction and seem to be at the same distance from the camera, we think that they are a single flock. Later we are shown a motion picture taken of that same scene. On looking at this view, we see that there were two distinct flocks rather than only one. This conclusion becomes evident when we look at the paths of motion. One group of birds seems to be moving in a set of parallel paths describing a curve in the sky. Another group is also moving in a set of parallel paths but this path is different from that of the first group. The common motion within each subgroup binds that group together perceptually and, at the same time, segregates it from the other group. The common motion within each group is an example of the Gestalt principle of common fate.

A great deal of work has been devoted to observing how the auditory system uses common fate to group frequency components from a complex auditory scene and form auditory images of single objects. By analogy it can be seen that the mechanism can pertain to the question of instrumental blend as well, regarding the individual instruments analogously to partials. To obtain a blend among several instruments, then, the logical strategy would be to have them all change according to one global temporal pattern.

There are two levels on which such global patterns can occur: through common micromodulation, or common temporal envelope. Having all the players in an ensemble adopt the same vibrato (frequency modulation) rate would promote their blend; this would be an example of micromodulation. Having the instruments share the same or similar pitch trajectories, the same attack-sustain-decay patterns in their amplitudes, or the same changes in brightness over time, would similarly promote their blend; these are examples of common temporal envelopes. In all of these cases, the synchronous temporal changes among multiple sources mimics the behavior of low-level elements (i.e., partials) that the auditory system customarily attributes to single sources, and the listener registers the combination as fused and arising from a single source.

Micromodulation

An example of micromodulation with speech stimuli may be found in McAdams (1989). Chords of three simultaneously sung vowels ([a], [i], and [o]) on the notes C3, F3 and Bb3 were modulated according to various patterns. In one pattern, one vowel was modulated according to a different pattern than the others, or that vowel was modulated while the others were held steady. A second pattern was the inverse of the first pattern, and a third pattern consisted of either no modulation at all, or all three vowels modulating together. In each condition, listeners were asked to listen to the sound of all three vowels contained in the chord and independently judge the prominence of each. The three vowels did not always appear at the same pitch: on any given trial, [a], [i], or [o] could appear on any of the notes C3, F3, and Bb3, so listeners could not simply rely on judging the prominence of the pitches.

In the condition where all vowels were modulated according to the same pattern, listeners tended to rate none of the vowels as more prominent than the others. Thus, having common modulation patterns or no modulation at all made the collection of vowels more uniform in quality and less distinct in the identifiability of its individual elements: they blended. This appears to be an inverse example of the paradigm "segregation by identification" mentioned earlier. In the similar case where none of the vowels were modulated at all, the vowels were similarly less distinguishable. Interestingly, in this condition, listeners also became more uncertain about the *number* of pitches present, reporting four to six when in fact they were told to expect only three; this may be a corroboration of the paradigm "blend by indistinct numerosity" mentioned earlier. The other conditions, on the other hand--either one vowel modulating according to a different pattern than the others, one vowel modulating with the others not modulating, or one vowel not modulating while the others modulated--resulted in strong differences in prominence between vowels. When one vowel was modulated according to a pattern different than the other two, that vowel separated out from the others; again, this is a case of "segregation by identification."

A similar finding appeared in the experiment by Rasch (1978) which was described earlier (detecting if notes in an upper voice lept upward or downward). On some of the trials, vibrato was added to the upper voice (a frequency modulation of 5 Hz with a depth of 4%). In this condition, listeners' accuracy was maintained even when the upper voice was reduced 17 dB, whereas, with no vibrato on the upper note, accuracy was maintained only if the voices were of equal amplitude. In other words, with vibrato on the upper note, listeners could tolerate a disparity of 17.5 dB (i.e., the top voice softer than the bottom) and still successfully judge whether notes ascended or descended. This indicates that the micromodulation was successful at causing the upper note to segregate from the vertical combination with the lower tone.

Common Temporal Envelope

Regrettably, there is still little research investigating the common fate mechanism when large-scale temporal changes are applied to a mixture of two or more complex sounds.[21] One explanation for this is that while temporal changes such as pitch trajectory and amplitude envelope are important for music, and most Auditory Scene Analysis research focuses on acoustic factors contained in artificial stimuli. In any case, there appears to be little work in this area that is applicable to the musical question of blend.

---

[21] One possible exception to this is Bregman and Chalikia (1989), who studied the fusion of tones "gliding" in parallel (i.e., changing in fundamental frequency according to straight-line functions).

Spectrum and Blend

It will be recalled that there was some evidence in the review of orchestration manuals that instruments having "dark" spectra were useful for promoting blend. Dark instruments such as the bassoon and French horn, for example, were frequently cited as blending well with a number of instruments, or effective at "bridging" timbres together. Some studies that offer some empirical support for this belief are discussed below.

Spectral Differences Between Solo and Choral Singing

It was mentioned earlier that the performer exerts an influence on whether or not a particular combination will blend. Goodwin (1989) studied the acoustical properties of sung female voices when performers consciously attempted to obtain a blend with a choral ensemble.

Goodwin defined "choral blend" as an ensemble sound in which individual voices are not separately discernible to a listener, and observed that choral singers intentionally alter their vocal production in the interest of achieving this quality. To investigate this, each singer in Goodwin's experiment (all sopranos) first produced sustained vowel sounds at a moderate dynamic level and in a manner typical of her solo singing, which were recorded and analyzed. Next, each singer attempted to blend with the sustained vowel of a well-blended unison ensemble of other soprano voices that was prerecorded on tape loops. The singer heard this tape loop through headphones, and while listening, sang the same vowel, which was recorded and analyzed. Through the headphones, however, she heard her voice mixed in with the voices on the tape loop. Thus, the situation closely matched what would occur in a real choral situation where a performer is trying to obtain a blend: the singer would sing in such a way that she could not discern her voice from the rest of the ensemble. The vowels [a], [o], [u], [e], and [i] were used, with fundamentals on c4, a4, and f5. Goodwin then compared the acoustical differences between solo and blended choral singing.

The results showed that choral singing, where a blend was desired, "tended to have slightly stronger fundamental frequencies in combination with fewer and weaker upper partials, and also slightly stronger first formants in combination with weaker second and third formants" (p. 25). Furthermore, in blended tones, (1) there tended to be fewer and weaker partials on frequencies above the first formant; (2) the partials in the area around the first formant were stronger in the blended tones than in the solo tones; and (3) the strengths of partials between the formant peaks tended to be reduced. Note that all of these transformations are such that they result in a lowering of the spectral centroid, or increased darkness of timbre. Goodwin observed that such transformations could relate to a number of intentional acts by the singer: (a) the singer could consciously be attempting to sing softer (since many musicians use a strategy of matching loudness with others in the ensemble to obtain a blend), a transformation that is usually accompanied by darker spectra (Clark & Milner, 1964; Meyer, 1978, pp. 30, 34); (b) the singer may be employing a strategy of consciously darkening the tone to obtain a blend; (c) the singer might have been avoiding distinct vowel pronunciation, contrary to what they are taught to do in solo singing (and which tends to de-emphasize the valleys between formants); or (d) the singer may have omitted vibrato (which is hypothesized to clarify vowels by tracing out the spectral envelope; see McAdams, 1984, and Marin and McAdams, 1991).

The same sort of inquiry was conducted by Rossing, Sundberg, and Ternstrom (1985), studying the difference between male solo and choral singing. Their study showed an increased use of the "singer's formant" in solo singing, a type of singing in which increased power in the 2-3.5 kHz spectral region is obtained by skilled adjustment of the vocalist's apparatus and taking advantage of particular resonant cavities. Choral singing showed an increased emphasis on the lower partials relative to solo singing, and glottal and vocal tract adjustment was used by singers to obtain these qualities. Similar findings were reported for soprano voices in Rossing, Sundberg, and Ternstrom (1987).

Spectrum and Fusion of Single Sounds

Dannenbring and Bregman (1978) provided some evidence that darker spectra aided in the fusion of collections of harmonically-related partials into single complex tones. It may be speculated that the findings may be applicable to the fusion of *multiple* complex tones as well, since the fusion of a single tone may contribute the global fusion of several tones.

One of the conditions in their study was designed to explore the role of the shape of the spectral envelope in contributing to the fusion of a complex tone as a single auditory image. The shapes were either flat (all harmonics having the same amplitude), sloped downward (each harmonic half the amplitude of the one below it in frequency), or sloped upward. The downward-sloping pattern is one that is typical of "darker" sounding spectra, while the upward-sloping pattern is an extreme form of a "bright" sounding spectrum. Results in their study showed that fusion was best for downward-sloping patterns, and poorest for upward-sloping patterns (p. 372).

Additionally, Dannenbring and Bregman (1978) found that strong amplitude differences among individual harmonics could affect the global fusion of the complex. Increasing the amplitude of a single high harmonic would cause it to segregate out as a separate object from the remainder of the complex. Related work investigating the tendency for a single harmonic to segregate as a function of its departure from exact tuning is reported in Moore, Glasberg, and Shailer (1984), Moore, Glasberg, and Peters (1986), and Hartmann, McAdams, and Smith (1990).

Masking

Spectral Masking

Many musicians have speculated that a factor in determining the quality of concurrent instrumental combinations is masking among harmonics (see Jeffress, 1970). Such masking might enable two very heterogeneous instruments to obtain a blend, for example. Recall, for instance, that especially "bright" spectral envelopes appear to be harmful for blend (see previous section); an instruments whose spectrum favors lower harmonics might provide the necessary masking capability to cancel out the offending high-frequency components of a bright sounds, and help promote fusion among them. It was shown earlier (Chapter 2) that orchestration manuals described a timbral transformation called *softening* which closely resembled a masking situation. So far, no one has researched this question in the laboratory. However, some speculation based on spectral data of musical instruments and masking data was offered in a short study by Pepinsky (1941).

Using spectra obtained by analyses of various brass instruments playing notes of a Bb major triad, Pepinsky sought to analyze the acoustics of combined instruments by summing them together into chords. The ensemble consisted of Eb tuba on Bb1, baritone on Bb2, trombone on D3, French horn on F4, and cornet on Bb4. He recognized that simply adding together the spectra of individual instruments was not an accurate representation of how the ensemble was heard. Rather, one had to consider the "accumulation of masking effects produced by the components of the complex tones." Pepinsky believed that an aesthetic end could be served by this knowledge, in that "both good and bad effects in orchestration" could be accounted for by these masking patterns (p. 405).

Applying masking data by Wegel and Lane (1924) to the spectra produced by the brass chord, Pepinsky speculated which components would mask others, and what the effects (positive and negative) such masking patterns would have on the quality of the chords. He also considered how a change in dynamic for individual instruments would improve or worsen a given pattern, observing that "a routined conductor would be expected to make just such an adjustment" in obtaining the optimum quality (p. 407). Although Pepinsky did not explicitly use the word "blend," its role is apparent in his suggestions to make this or that change to obtain more "balanced" sounds.

For example, when all instruments in his simulated ensemble played at the same *forte* dynamic, he predicted that the baritone's especially strong 19th and 20th harmonics would create an undesirable roughness. However, with all instruments marked *piano,* he speculated that this problem would not occur, yielding a "smooth" quality.[22] The baritone's roughness quality could be made even more acute when playing *forte* and the others *piano;* he advised that "a better balance is secured by raising the level of the trombone and french horn to a mezzo forte level" (p. 407) since they would mask out

---

[22] The spectral descriptions of instruments Pepinsky possessed comprised only one dynamic level. The different dynamic levels were obtained by estimate, by a applying linear change in amplitude up or down for louder and softer dynamics, respectively. Since such transformations are highly artificial and do not reflect the actual changes in spectra that accompanied in natural dynamic changes, Pepinsky's speculations must be regarded with some caution.

the undesirable partials.[23] Similarly, the combination of cornet *forte* and the rest *piano* would mask the trombone and French horn, with the result that the cornet would protrude from the ensemble. However, raising the trombone and French horn to *mezzo forte* "would produce a better balanced accompaniment" (p. 407).

Informational Masking

Conventional research in masking tends to be simple (using static noisebands and steady state tones as stimuli) because most paradigms for explaining masking are in terms of low level aspects of the peripheral auditory system (the physical behavior of neurons in the basilar membrane). There is a small body of research, however, concerned with a somewhat higher-level, or cognitive aspect of masking called *informational masking* or *recognition masking*. This subject concerns the fact that listening involving recognition and categorization of very brief sounds requires a certain amount of undivided attention by the peripheral auditory system, and that competition from other incoming sounds of other information-rich short sounds can result in sounds being heard but not processed (i.e., not recognized, categorized). In order to be given the necessary attention by the system, the sound must be allowed to spend a sufficient amount of time in a

---

[23] Note that practice of using instruments to mask out undesirable partials is also suggested by Piston (1955). In observing that minor chords contain an inherent conflict between the "major third" partial (5th harmonic) and the "minor third" fundamental, Piston remarked "Overtones from the lower notes are sometimes disturbing in a chord, and may need to be canceled by upper tones" (p. 447).

very short-term memory location that Hawkins and Presson (1977) calls the *precategorical acoustic store* (PAS):

> The PAS presumably holds a relatively unanalyzed echoic representation of acoustic input for some period of time following stimulus offset and continues to transmit information regarding this representation to the subsequent categorization process --- perhaps at the level of the PAS. The earlier the disruption occurs within the several hundred milliseconds interval of test tone processing, the less information will be available for test tone analysis, and therefore the less accurate categorization performance will be. (p. 198)
>
> To perform optimally under any given set of listening conditions, the categorization mechanism requires some minimum interval for sampling from the acoustic stimulus or its residual. If the wanted stimulus is followed by an unwanted input with a delay less than this minimum sampling interval, the unwanted input interrupts and *replaces* the first as the dominant source of information feeding categorization. As a consequence, the categorization decision will reflect the character not only of the wanted stimulus, but also that of the masker. (pp. 207-208)

This mechanism could play an important role how the auditory system employs the information of instrumental *onsets* when determining whether combinations of instruments are perceived as blended or segregated. It is widely known that instrumental onset portions as brief as 10 ms in duration convey a great deal of information about the identity of an instrument; the auditory system must have an undisturbed opportunity to process such fleeting sounds in order to correctly categorize according to its identity. If the attack form (or some other component) of a competing instrument interferes, then the original instrument's identity may not be apprehended; it is informationally masked. Since the attainment of the blend condition depends in part on instruments not distinguishing themselves from one another--recall the paradigm "segregation by identification"--simultaneous attack forms could play a role in causing sounds to blend. One might speculate that the optimal condition for attaining blend is when the two sounds mask one another; in this case, the best blend would be attained between instruments whose onsets were of equal duration, since neither would be given uninterrupted access to the precategorical acoustic store. The evidence of Stern (1972) regarding the effect of offsets in onset times in instruments, discussed earlier in this chapter, can be interpreted in this light.

Chapter 3:

Method for Investigating "Blend"

Goals, Preferences and Requirements

Psychophysical Investigation

The goal of this study, as indicated in the introduction, is to gain a better understanding of how concurrently-sounding musical instruments are perceived by musicians as *blending*. The approach taken by orchestration manuals, it was shown, is "instance based"--that is, their suggestions for obtaining blend are limited to specific illustrations of various instrument pairs. This binds the information to specific style period, instrument types, and instrument technologies, making it difficult to generalize their methods to different musical media (such as synthesizers, early music instruments, or non-western instruments). To achieve generalizing power, it is necessary to address the problem at a more abstract level, exploring the range of acoustical factors that are active when instruments are judged as blending or not blending.

The science of correlating the physical properties of a sound to the perceptual attributes they give rise to is known as *auditory psychophysics*. Psychophysics involves measuring the physical and perceptual properties of a stimulus and discovering a mathematical function for relating the two, with the goal that the physical properties be used as predictors for the perceptual properties. Physical properties of sound are obtained by acoustic measurement, usually along several dimensions; perceptual properties are derived by measuring a behavioral response to the sound. The researcher determines a mapping of a physical scale to a perceptual scale, using a statistical "best fit" criterion. Regression, for example, can be used when the mapping is many-to-one: this procedure finds a polynomial that enables one to predict the perceptual response from a weighted combination of the acoustical attributes.

Most auditory psychophysical studies investigate the scaling of sounds which are presented one at a time. For the purpose of this study it will be necessary to devise a method for investigating two or more simultaneously sounding events. When a stimulus consists of two timbres sounded at once, for example, the physical scale values should include both the properties of the independent constituent timbres and the properties that emerge as a result of their being combined. The perceptual scale values should represent the subjects' responses to the emergent, sum property alone.

It might be asked whether psychophysics can address matters of a musical nature as complex as those employed in judgments of orchestration.

Judgments about orchestration involve not only complex perceptual phenomena but also high-level *musical schemata.* This raises the following questions:

1. Given the restrictions of a controlled experiment, what level of realism in the representation of "musical instruments" is possible (or alternatively, what level of artificiality is tolerable)?
2. What level of acoustic description is desirable and informative?
3. What is an appropriate task to elicit a judgment of musical blend?

These questions are addressed below.

Realistic Stimuli

It is typical in the design of many auditory psychophysics studies to ask a question concerning specific auditory parameters of sound, and investigate this question using stimuli that vary only in those parameters. The advantage of constraining the stimuli to include only *known* variations is that the attribution of perceptual responses to specific changes in the individual acoustical parameters is very reliable. For example, consider the stimuli used in a study of timbre by Howard and Silverman (1976). The authors wanted to observe the degree to which each of four timbral dimensions (sound source, fundamental frequency, formant region and number of formants) determined the judged similarities between sounds. To this end, they employed sounds that varied in only these four dimensions, with two possible values in each dimension: sound source (square or triangle wave), fundamental frequency (90 or 140 Hz), region of formants (center frequencies of 600 or 1550 Hz), and number of formants (1 or 2). In every other way the sounds were equalized. All factorial combinations of these features led to exactly 16 different sounds. A analysis of the results showed that similarity judgments were made among three primary dimensions, which were, in descending order of importance: fundamental frequency, sound source, and formants (both region and number of formants playing equally strong roles).

A study in the perception of blend might similarly constrain the stimuli in a simple factorial fashion. For example, motivated by the information from Chapter 2, one might devise stimuli of concurrently-sounding complex tones in which the constituent complexes within each trial differ from one another in fundamental frequency, onset disparity, or brightness, but which in every other way are identical. Listeners would be asked to judge whether each pair was fused or separated.

There are two problems with this research agenda. First, stimuli that explore only a small set of very simple parameters typically sound unmusical, and the timbre space they explore is predictable and unexpressive. Since the mid-1960's a number of studies have shown that in order to engage the auditory system in a musical way timbres must be informationally rich: they must exhibit complexity in their attack forms and resonant structures, as well as temporal complexity in their amplitude and spectral envelopes (for a review of this literature, see Grey, 1975, pp. 1-15; Risset & Wessel, 1982; Freed, 1988, pp. 11-17; and Handel, 1989, pp. 226-263). Such features are present in most acoustic musical instruments; consequently, electronic or computer music sounds that exhibit such features are felt to be "natural sounding" by listeners. Since the objective of the present study is to explore how instruments blend in an orchestral context, it is important that the stimuli maintain a strong relationship to typical orchestral sounds.

Second, biasing this study in advance to produce an outcome that relates to only a restricted set of acoustical parameters will be informative only to the extent that the set accurately captures parameters of variation relevant to orchestration. In the blend experiment informally proposed above, even if the judgments showed effects pertaining to the devised features, the results would not assure their status as primary perceptual attributes of orchestration. On the other hand, they might be valuable if previous research had shown, for example, effects for fundamental frequency and onset in separate experiments, but whose interaction was not yet understood. The status of the current research project, in contrast, is more pristine: with the possible exception of Stern (1972), there exists no research previous to the experiments reported here concerning the judgment of blend for concurrently-sounding orchestral instruments.[24] By necessity then, the present study must include experiments of an *exploratory* nature. This in

---

[24] Stern (1972) can be considered to have explored blend indirectly, since one measure of blend is the degree to which the constituent sounds in a combination can be identified ("segregation by identification").

turn increases the importance of having informationally rich stimuli: the stimulus set should contain the natural acoustic variation found in actual musical instruments in order to elicit blend judgments based on as yet unidentified factors. It is anticipated that such exploratory experiments will identify salient parameters affecting blend. With such information, it will also be possible to pursue experiments of a non-exploratory nature that test hypotheses about these parameters, for example, exploring their range of applicability. This study will include some non-exploratory research as well.

Distinctive Features

Although a "closed set, factorial design" based stimulus space is inappropriate here, it is necessary nonetheless that the acoustic attributes of the chosen stimuli be characterizable in some dimensional way. Freed (1988) reviewed a number of methods for characterizing the acoustic properties of complex, time-varying sounds. One method he identified consisted of a strictly objective, physical measurement of acoustical properties, such as "a list of locations, levels, and bandwidths of resonant regions or formants" (p. 9). Another method he identified is based around properties of the peripheral auditory system, characterizing sounds by such factors as "critical bands, loudness functions, masking, lateral suppression, frequency-dependent thresholds, and resolution in time and amplitude" (p. 9). Another method is based on the assumption that the auditory system is "directly sensitive to parameters which carry environmental information--i.e., information about the object in the environment that produced the sensation" (p. 10); this is referred to as the "ecological approach," after Gibson (1966). In this approach, the focus is on those acoustical parameters which carry information about the object's physical features such as size, composition, density, mass, method of sound production, materials of construction, and so on. Freed's own study, for example, investigated the attributes of a set of percussive stimuli in terms of the *perceived mallet hardness* of the striking instrument.[25]

The use of any efficient, compact way to represent those aspects of the sound that carry perceptually relevant information may be called the *distinctive features* approach. *Distinctive features* originates in a speech research paradigm proposed by Jakobson, Fant, and Halle (1951), which

> codified certain long-standing observations of phoneticians by hypothesizing that the many sounds of speech can be placed in categories based on the presence or absence of certain distinctive features. Whether the mouth is open, whether there is a narrowing of the vocal tract at a particular place, whether a consonant is aspirated--properties such as these make up the features that characterize and distinguish the phonetic content of a language. (Slawson, 1985, p. 51)

Grey (1975), for example, used a similarity rating task and Multidimensional Scaling to uncover the distinctive features affecting the similarity of a group of realistic sounding syntheses of orchestral instrument tones.[26] The dimensions he found were: (a) the location of the spectral energy distribution (a "bright-dark" dimension), (b) the presence or absence of

---

[25] Freed's experiment is also described in Freed (1990).

[26] Grey's similarity study is also described in Grey (1977).

noise in the attack portion of the instrument, and (c) the presence or absence of synchrony among the amplitude envelopes of upper harmonics at attack time. An alternative account of the third dimension was (d) instrumental family membership (.e.g., strings, brass or woodwind). In other words, comparing two instruments with these features is an efficient way to represent the listener's perceptual process of comparison.

Selecting Natural, Complex Timbres for Research

The possession of synthesis specifications of orchestral tones allowing low-dimensional control of their parameters is the cornerstone to this project. The stimuli used in Grey (1975), described above, satisfy the requirements of this project. Their main advantages are that (a) they are realistic, natural sounding syntheses of orchestral instruments originally obtained from recordings of real performers, (b) they are specified with parameters that allow low-dimensional control and that enable the production of variants or modifications for experimental or compositional purposes, (c) they have been shown to be rich in distinctive features, some of which can be manipulated by the synthesis parameters.

Because the synthesis specifications for Grey's tones were available to the present author, it was decided to design experiments around these tones. The stimuli for the experiments in this study will therefore consist of concurrently-sounding presentations of these tones. For that reason, the origins or the tones, their parameters, and the discovered distinctive features will be covered in an extended section contained later in this chapter.

Rating Blend

A few remarks about the task that listeners will perform is necessary. The description of the task will be limited, for the present, to only general issues. More complete expositions of the exact task for each experiment will be included in the following chapter.

How will the blend judgment be obtained? It was shown earlier that experiments by other researchers employed an identification task (e.g., Stern, 1972, and McAdams, 1989), where ease of identification was presumed to indicate segregation, and difficulty indicated blend. Unfortunately this assumes the auditory objects of the orchestral medium all strongly internalized in listeners. On the contrary, not all instruments are equally identifiable even in ideal contexts (Saldanha & Corso, 1964; Strong & Clark, 1967; Grey, 1975). A violin tends to be more easily identified than an English horn, but it should not be concluded from this that the former blends poorly and the latter well. Although McAdams' vowel-segregation study (McAdams, 1989) could safely count on listeners to recognize the vowels [a], [i], and [o], expecting similar levels of performance for even musicians on orchestral instruments is somewhat premature.

It seems a better idea to employ a direct evaluation of blend as the task in these experiments. This now raises the question of whether the rating should be a simple bi-polar scale ("blends" vs. "segregates") or a continuum between these two poles. It would seem that, to be responsive to the needs of the current musical scene, a continuous rating scale is a more attractive proposition. Accordingly, the experiments should provide a heuristic not merely for attaining either blended or segregated timbres, but for obtaining any degree of blend along a continuous scale of possibilities.

The Stanford Tones

The synthetic musical instrument tones used by Grey (1975) will be described here. Since all the conclusions drawn in this study will be based entirely on correlations between blend ratings and the features of these tones, the origins and acoustic properties of the tones are of great importance. To avoid encumbering the experiment reports in the following chapter with long digressions, the physical properties of the individual tones will be analyzed in advance.

The Origins of the tones

The tones used by Grey were created at Stanford University, at the Center for Computer Research in Music and Acoustics (CCRMA), as part of a long-term analysis-by-synthesis project. They have been used in various studies by a number of researchers mostly associated with CCRMA (Grey, 1975; Grey, 1977; Grey & Moorer, 1977, Grey, 1978; Gordon & Grey, 1978; Gordon, 1984; and Gordon, 1987). They are hereafter referred to as the "Stanford tones."

The instruments were: flute, oboe, English horn, e-flat clarinet, bass clarinet, soprano saxophone, two alto saxophones (one played *forte,* the same instrument and same player playing *mezzo forte*), bassoon, French horn, trumpet, muted trombone, cello bowed *sul ponticello,* (at the bridge of the instrument), cello bowed normally, and cello bowed *sul tasto* (over the fingerboard).[27] The tones were syntheses of data originating from analyses of live recordings of performers. The process of recording and analysis, and the data reduction process will be summarized below; for further information on this subject, however, see Moorer (1973), and Grey (1975).

The performers recorded by the CCRMA researchers were requested to play a short note of pitch Eb4 (ca. 311 Hz). Recordings were made originally on analogue tape recorded at 7.5 ips, and then digitally sampled at 25,600 samples per second with a 12-bit digital-to-analogue converter. The digital soundfiles were then submitted to a heterodyne filter (Moorer, 1973) which analyzed the energy in the region of each harmonic multiple of the expected 311 Hz tone. The result was a representation of the original tone as the sum of several sine waves (one for each harmonic) that varied in amplitude and frequency over time. The temporal resolution was about 1 ms. The representation is sufficiently detailed that tones resynthesized from these parameters by additive synthesis sounded remarkably natural and faithful to the original recordings.

---

[27] John Gordon's dissertation (Gordon, 1984) employs the Stanford tones but refers to two of the saxophones as being *tenor* saxophones (p. 31). This is an error, according to Grey (personal communication, October, 1988): the saxophones were indeed altos.

Figure 7 illustrates this representation of the output for the second harmonic of one of the instruments (trumpet), showing the amplitude and frequency data in separate plots.[28] The erratic portions in the frequency plot between 0-60 ms and 400-510 ms are random patterns that the heterodyne filter produces when there is insufficient gain, so they did not include this in their representations of the tone. Figure 8 shows all the amplitude functions for the harmonic in a single plot.

Much of the fine temporal detail in both the amplitude and frequency functions shown in Figures 7 and 8 is perceptually superfluous. That is, the functions can be simplified to a certain degree without the human ear detecting the difference or noticing a degradation in quality. For example, the discontinuities contained in the initial rise in the amplitude of the trumpet's first harmonic (the amplitude change of 0 to 450 occurring between 60 to 90 ms) can be replaced by a single continuous linear change. The CCRMA researchers simplified all of the amplitude and frequency envelopes by a series of carefully selected contiguous straight line functions. They called a tone represented in this way a *line segment approximation,* while the original non-simplified representation was called a *complex synthesis.*

---

[28] The term *harmonic* will be used, by convention, to refer to a component of a tone even though none of the Stanford tones have strictly harmonic relationships among their partials.

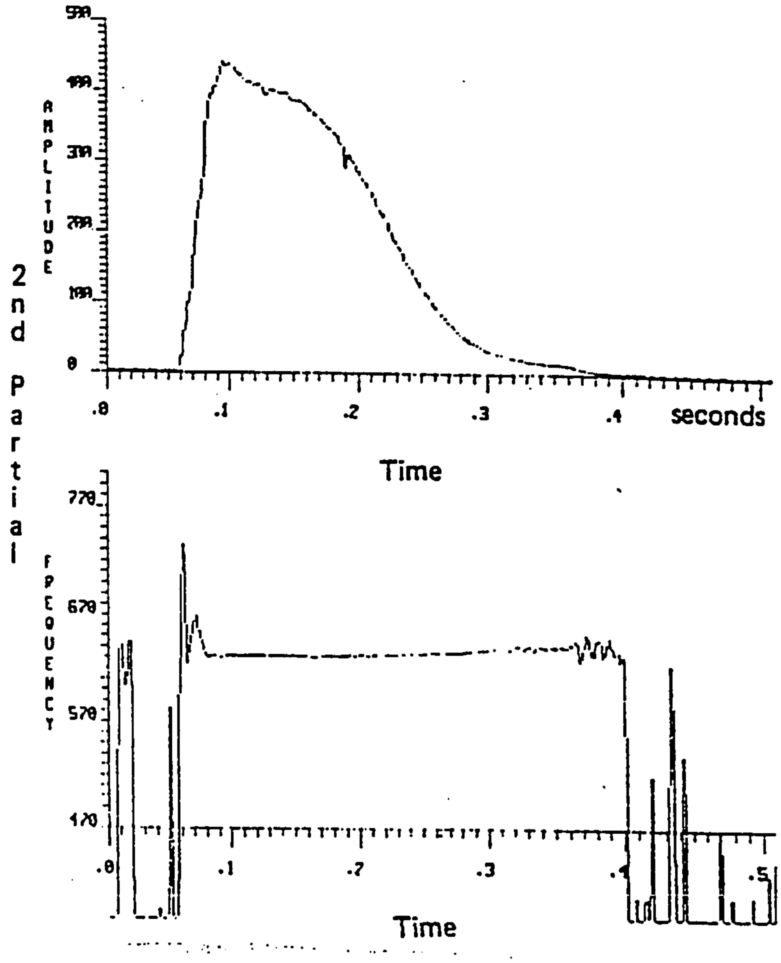


Figure 7. Amplitude and frequency functions from a heterodyne filter analysis of a trumpet tone, 2nd harmonic. Reproduced from Moorer, Grey and Strawn (1978).

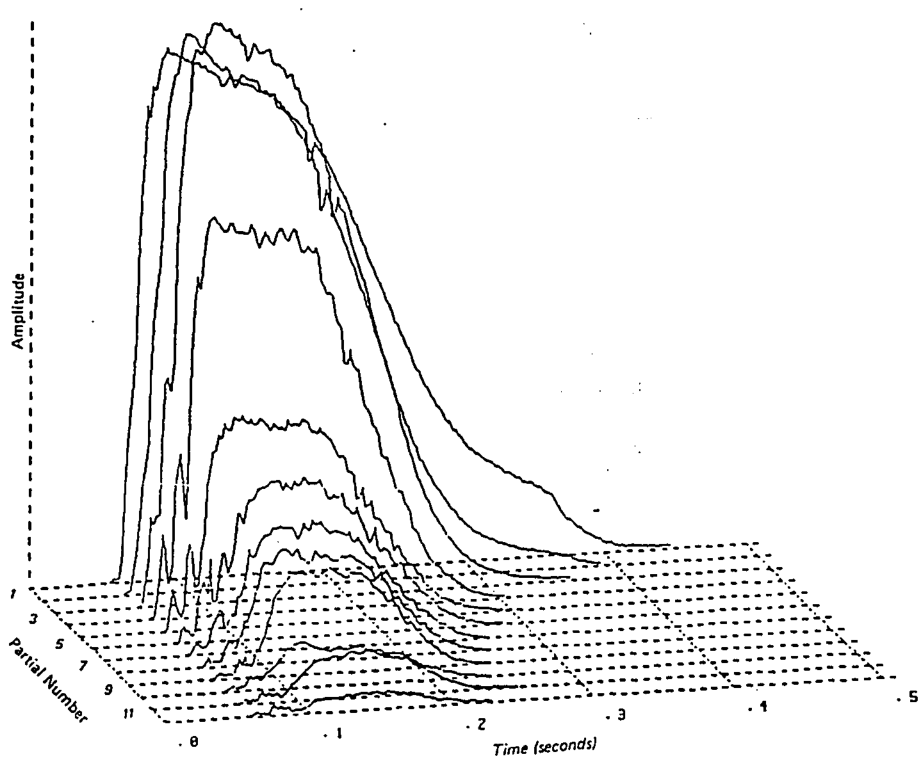


Figure 8. Amplitude functions from a heterodyne filter analysis of a trumpet tone, all harmonics. Reproduced from Moorer, Grey and Strawn (1978).

For all the tones they analyzed, the CCRMA researchers obtained the simplest possible line segment approximations that could be obtained for each harmonic without losing any audible detail. They also chose to limit the number of harmonics necessary for a given tone based on how many appeared to be necessary for accurately reproducing the tone. Apparently the researchers used trial and error and informal audition as a guide to what information was superfluous and what was not during this process.

The line segment approximations of the amplitude functions for all 15 tones are shown in Figures 9 through 23. As an example of how much reduction was obtained, the original representation of the trumpet (Figure 8) can be compared its line segment approximation (Figure 19).

Similarly, the line segment approximations of the frequency functions for all 15 tones are shown in Figures 24 through 38. Here the representation is slightly different from the amplitude functions. The harmonics are numbered from front to back rather than back to front as before. The y-axis now shows frequency; there is no amplitude information shown in these plots. A naming convention has been adopted in these graphs (e.g., "flute," "oboe2," "englhorn," and so on) which will be employed throughout the remainder of the study.

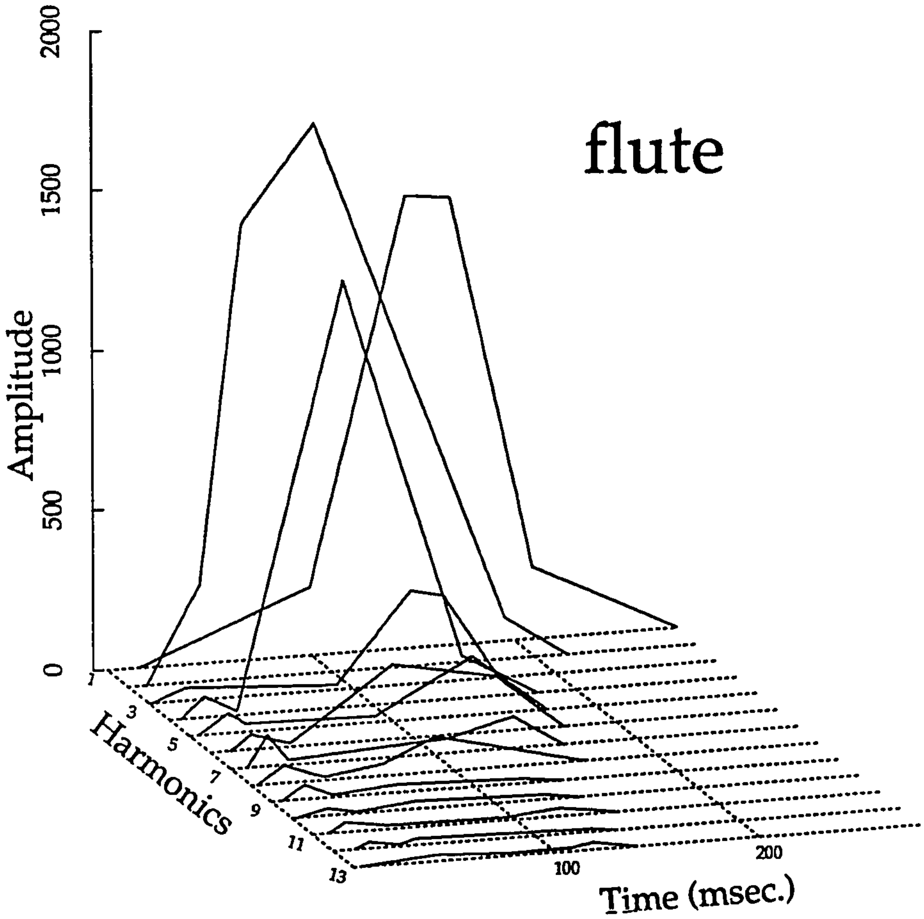


Figure 9. Amplitude functions of the Stanford flute tone, all harmonics, *line segment approximations.*

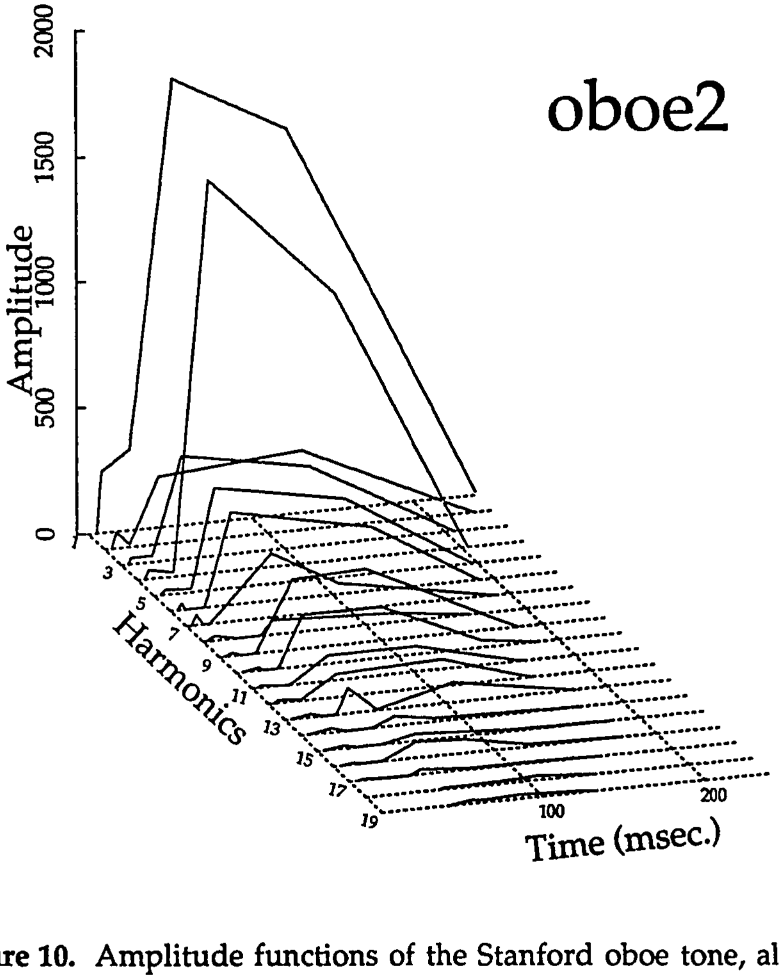


Figure 10. Amplitude functions of the Stanford oboe tone, all harmonics, *line segment approximations.*

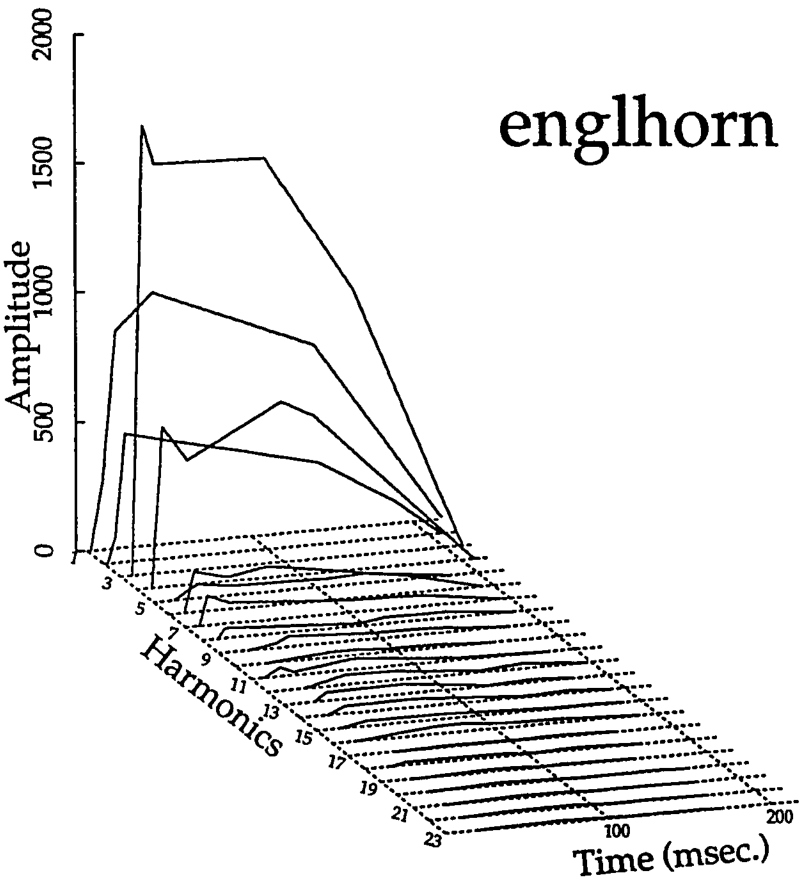


Figure 11. Amplitude functions of the Stanford English horn tone, all harmonics, *line segment approximations.*

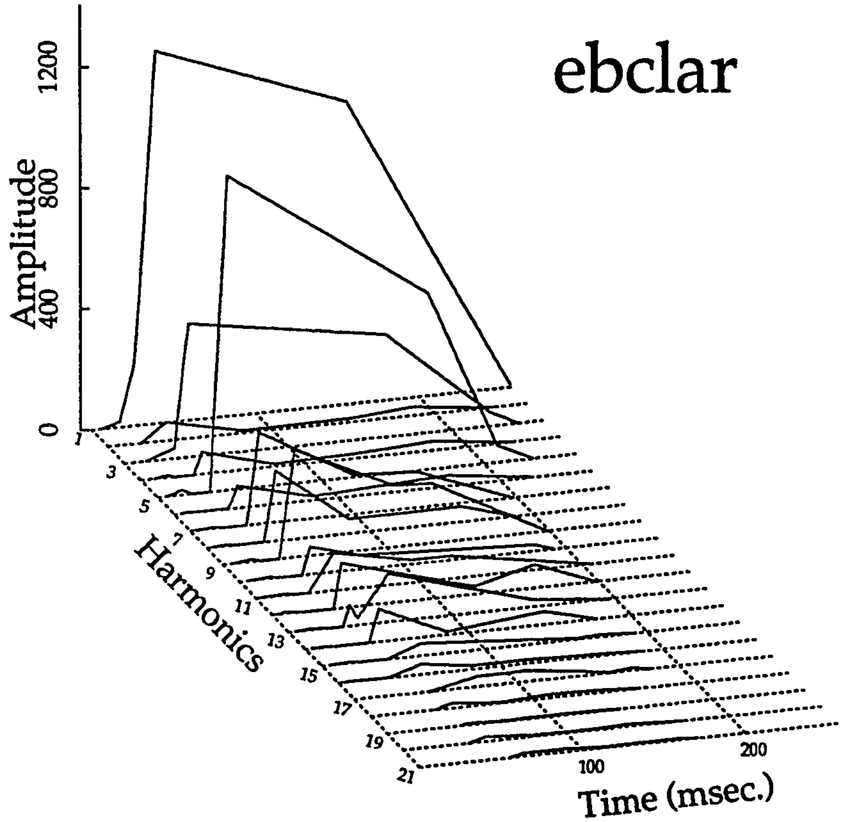


Figure 12. Amplitude functions of the Stanford Eb clarinet tone, all harmonics, *line segment approximations.*

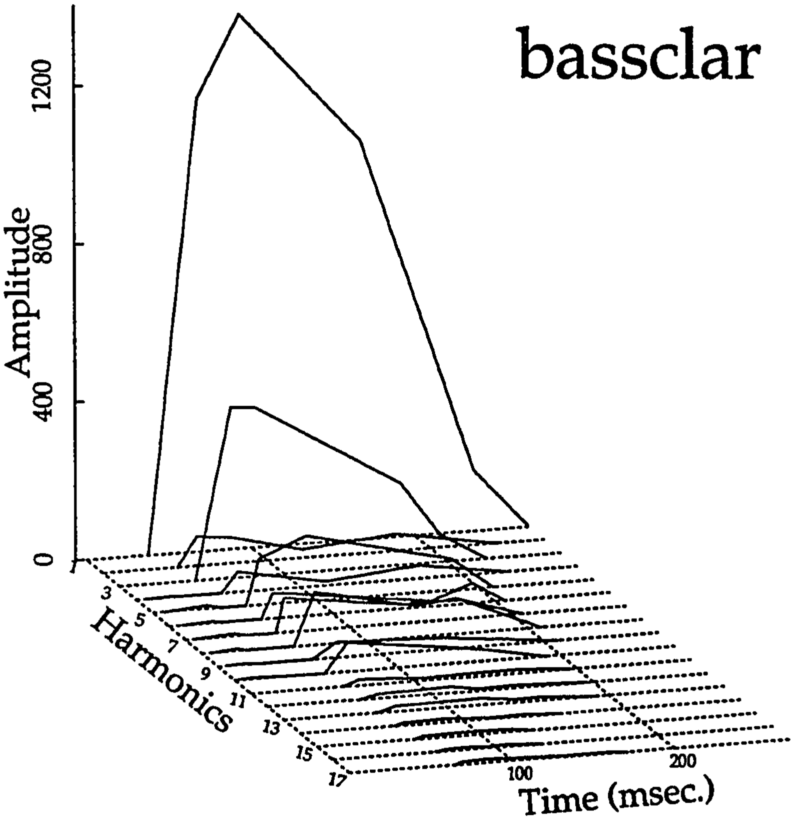


Figure 13. Amplitude functions of the Stanford bass clarinet tone, all harmonics, *line segment approximations.*

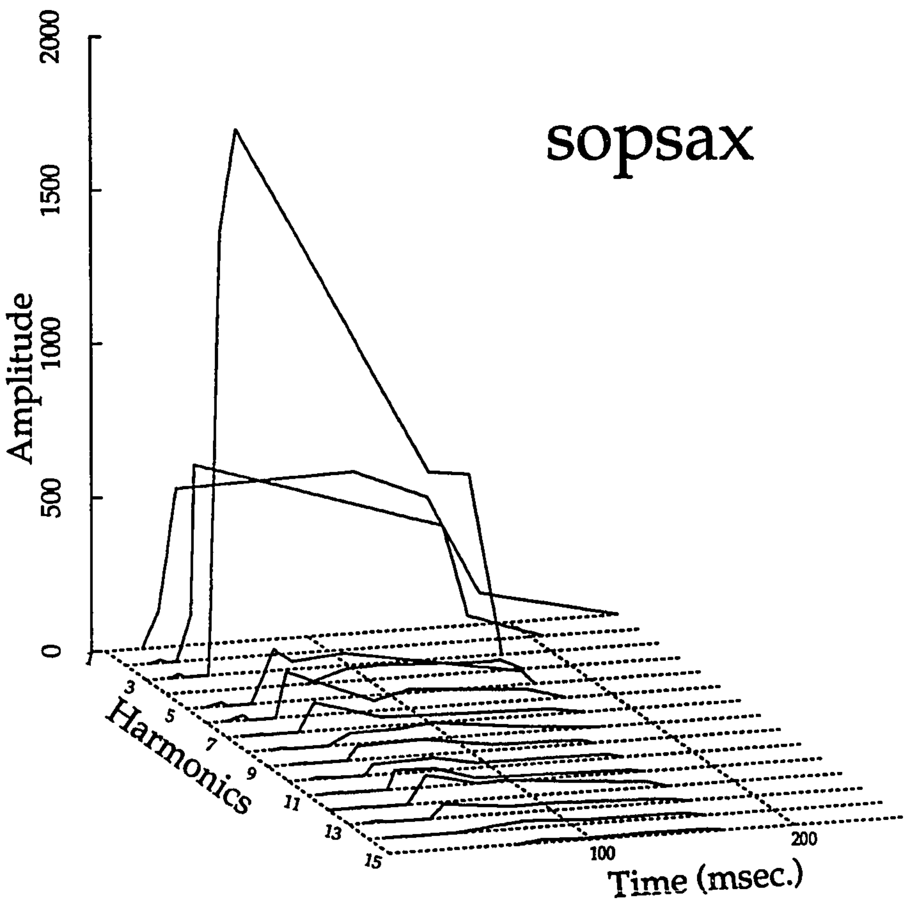


Figure 14. Amplitude functions of the Stanford soprano saxophone tone, all harmonics, *line segment approximations.*

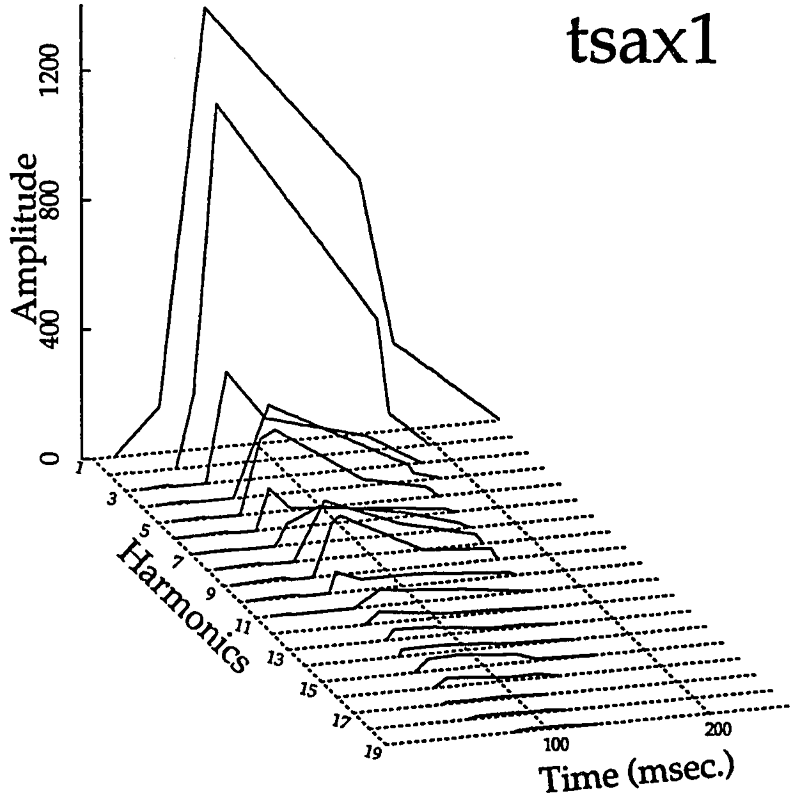


Figure 15. Amplitude functions of the Stanford alto saxophone (played *mezzo forte*) tone, all harmonics, *line segment approximations.*

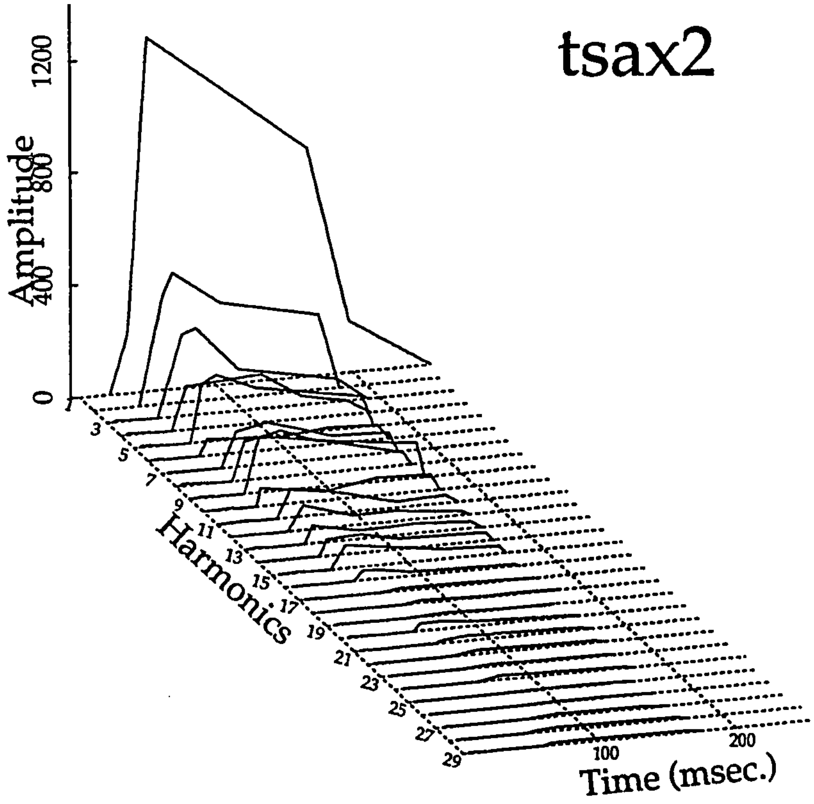


Figure 16. Amplitude functions of the Stanford alto saxophone (played *forte*) tone, all harmonics, *line segment approximations.*

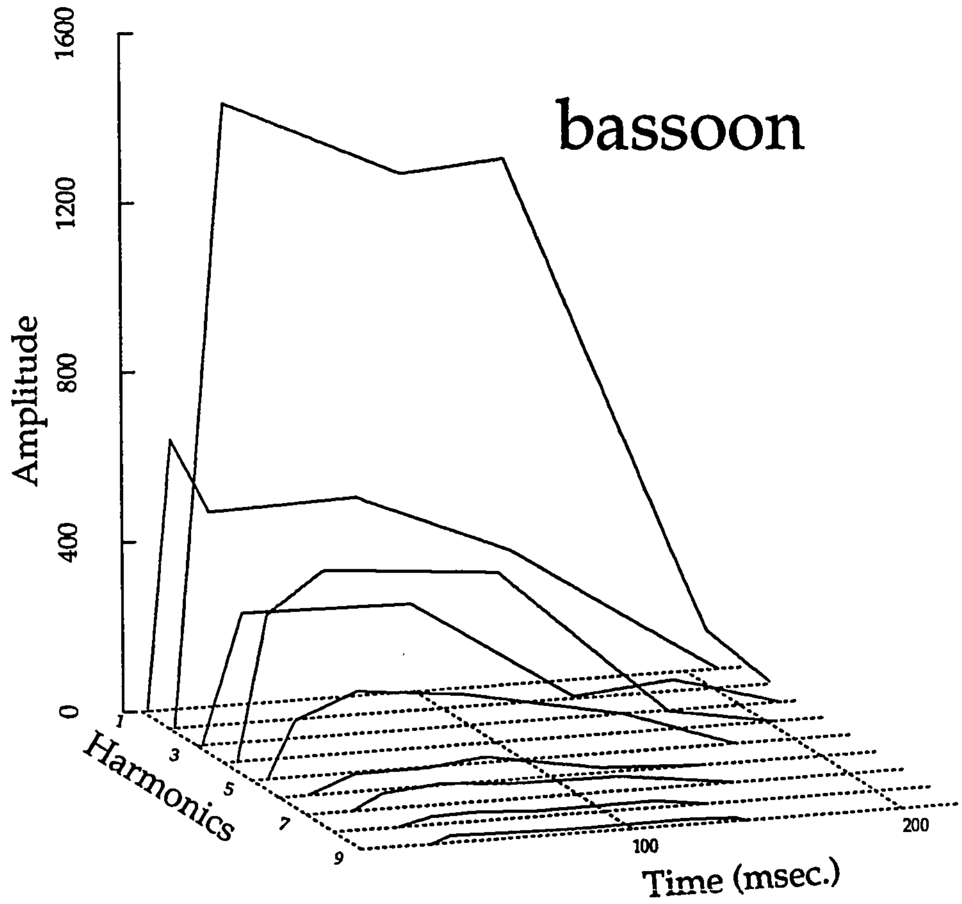


Figure 17. Amplitude functions of the Stanford bassoon tone, all harmonics, *line segment approximations.*

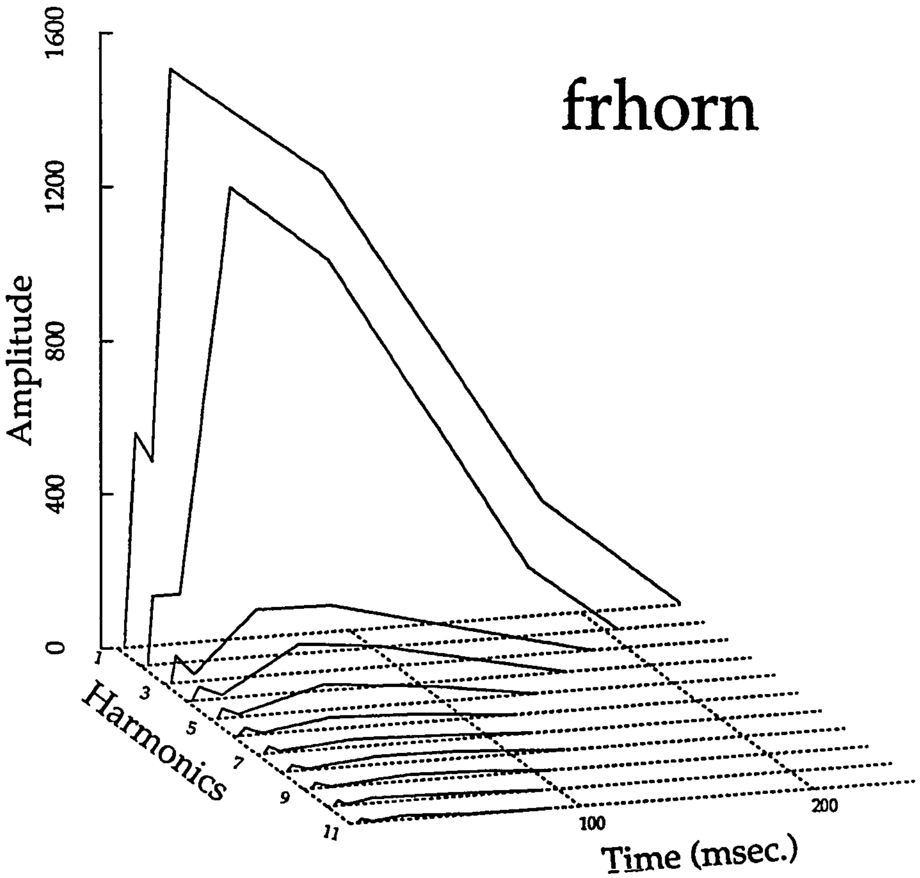


Figure 18. Amplitude functions of the Stanford French horn tone, all harmonics, *line segment approximations.*

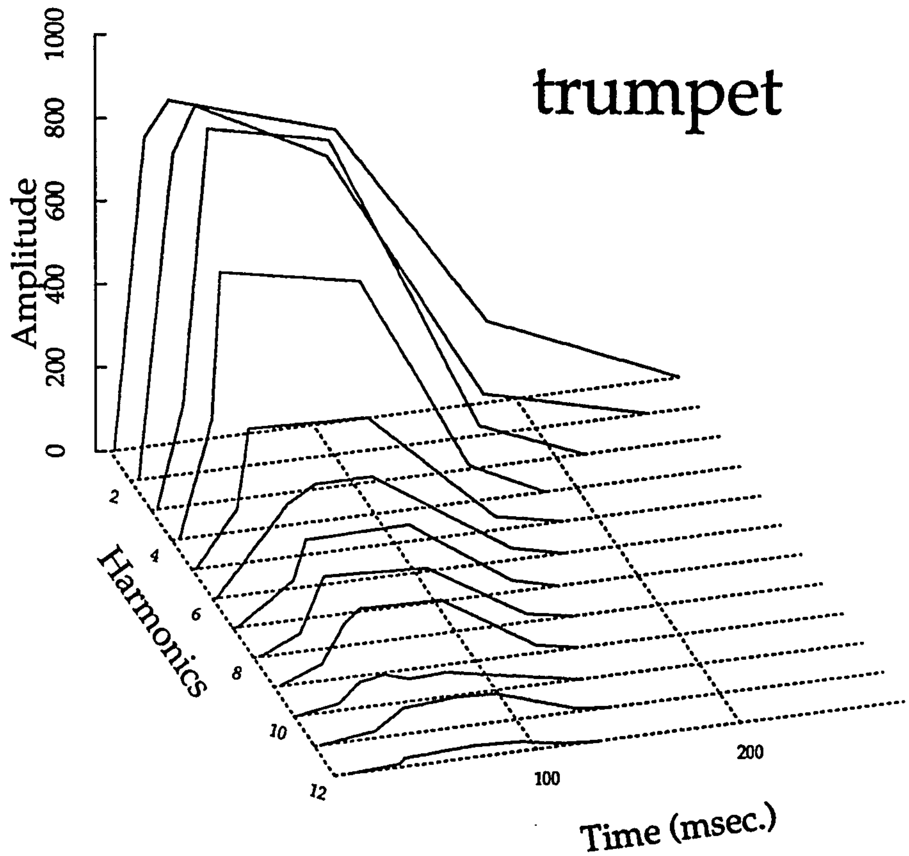


Figure 19. Amplitude functions of the Stanford trumpet tone, all harmonics, *line segment approximations.*

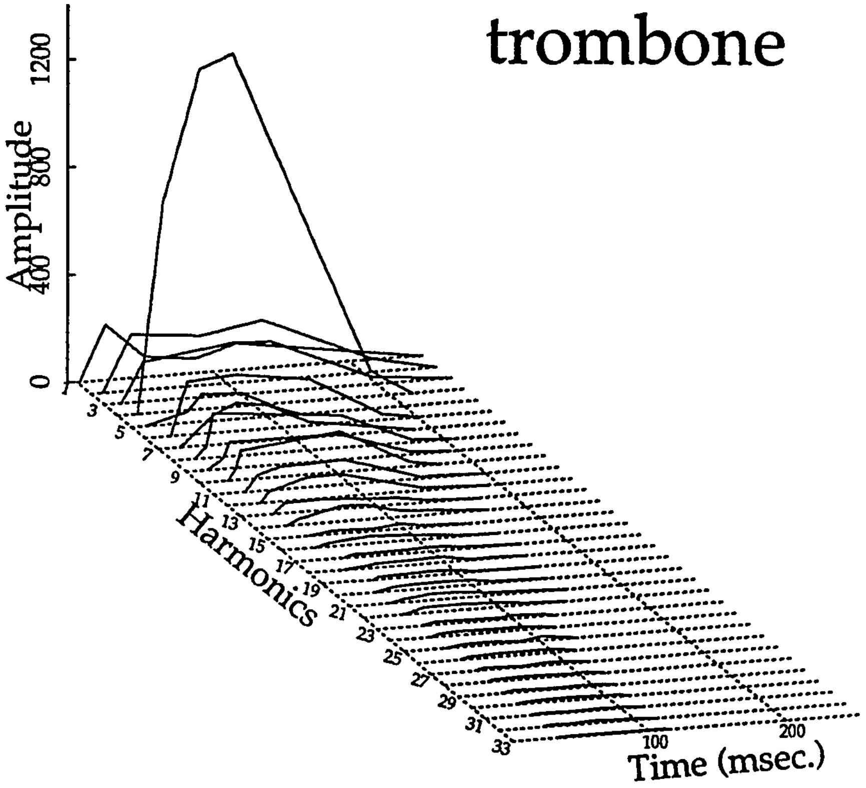


Figure 20. Amplitude functions of the Stanford trombone (muted) tone, all harmonics, *line segment approximations.*

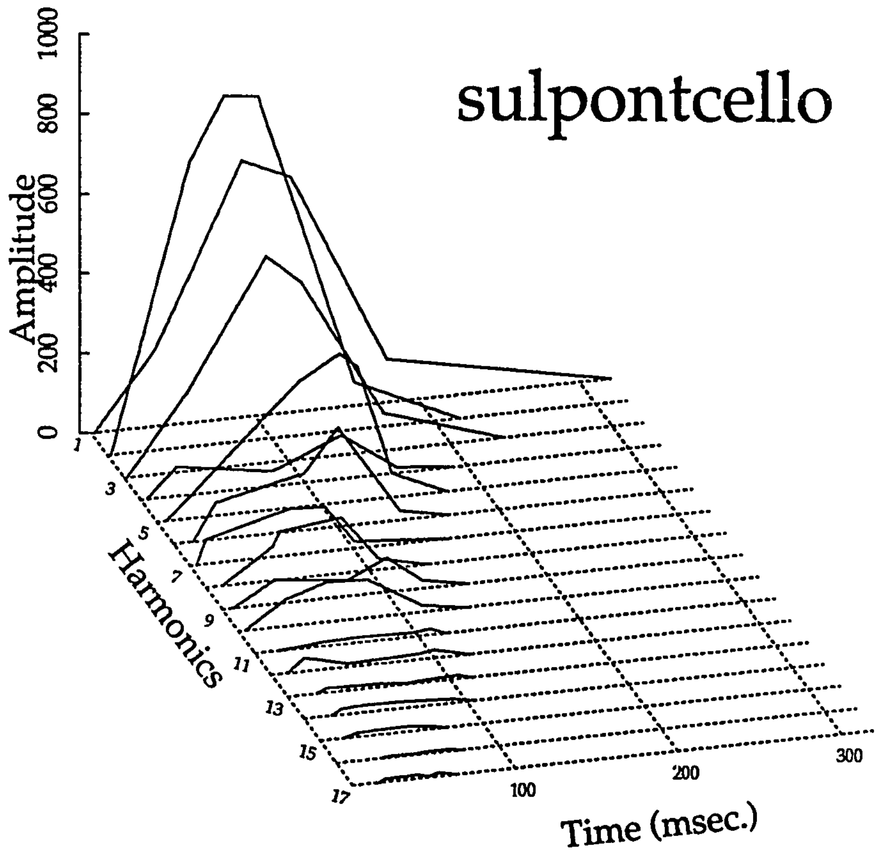


Figure 21. Amplitude functions of the Stanford cello (bowed *sul ponticello*) tone, all harmonics, *line segment approximations.*

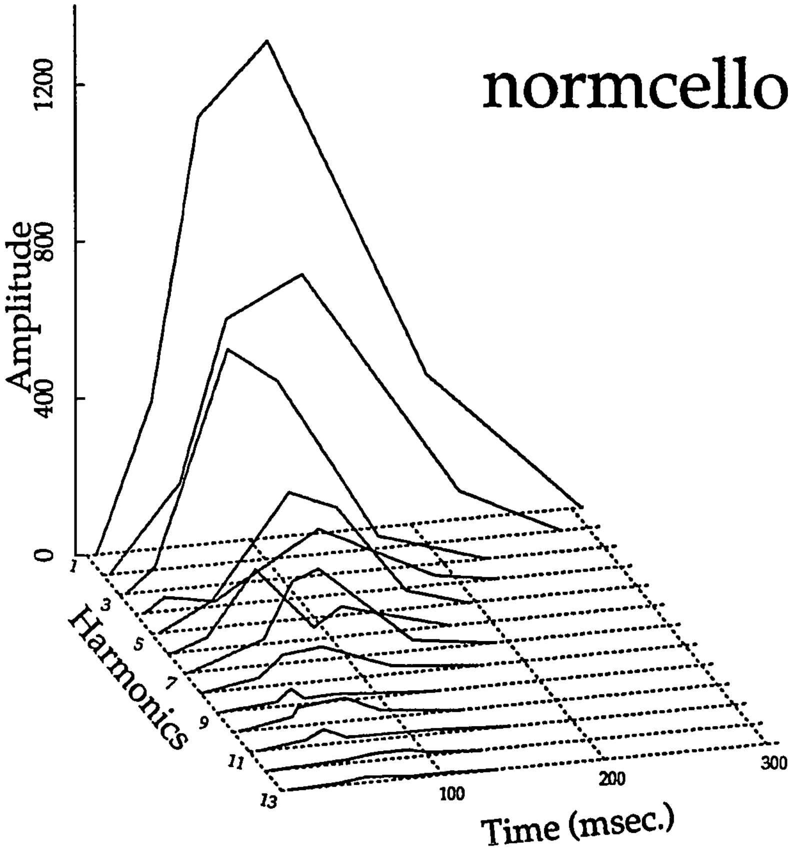


Figure 22. Amplitude functions of the Stanford cello (bowed normally) tone, all harmonics, *line segment approximations.*

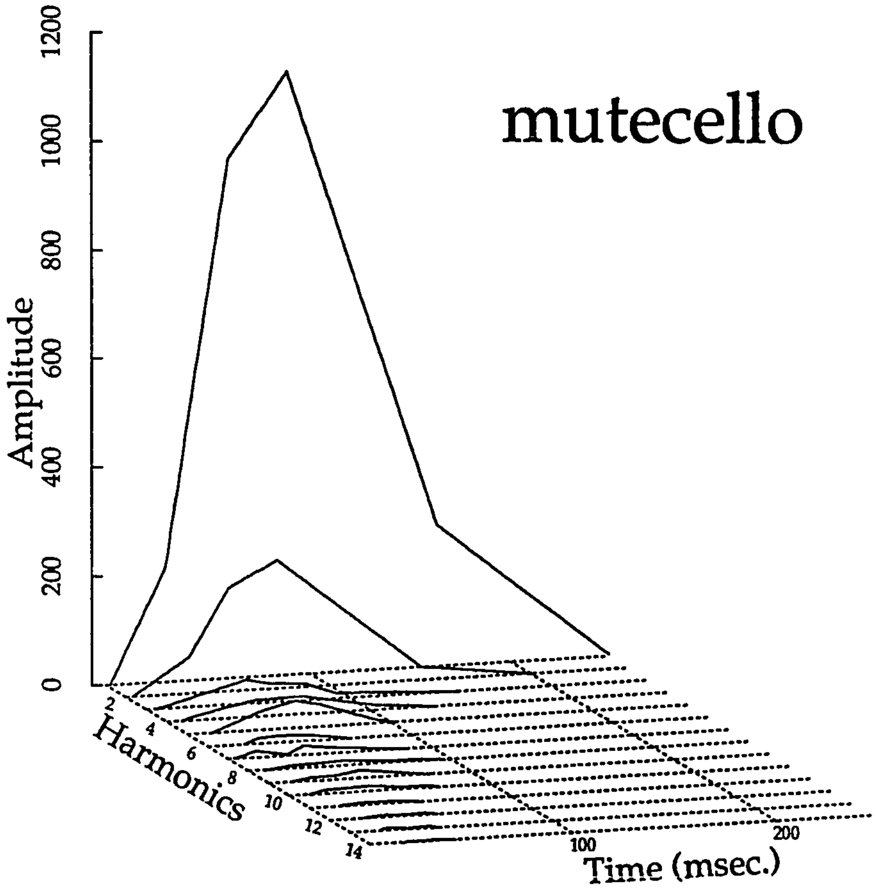


Figure 23. Amplitude functions of the Stanford cello (muted, bowed normally) tone, all harmonics, *line segment approximations.*

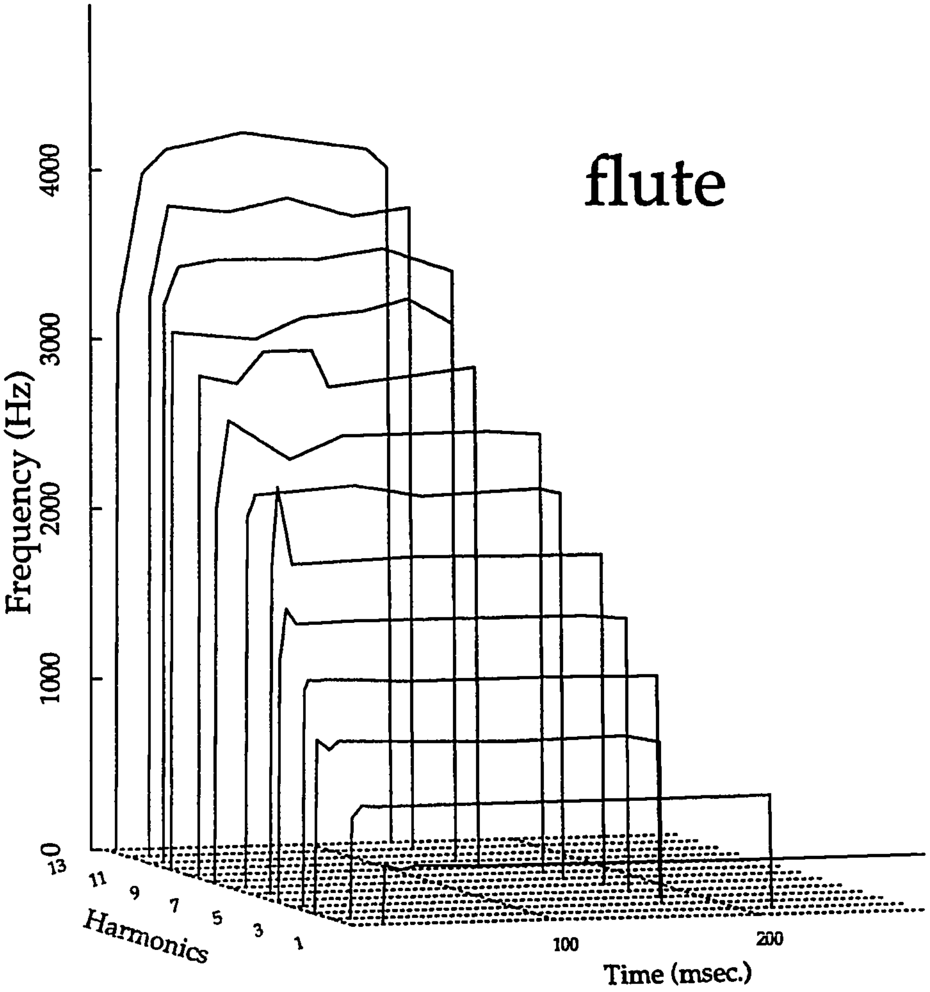


Figure 24. Frequency functions of the Stanford flute tone, all harmonics, *line segment approximations.*

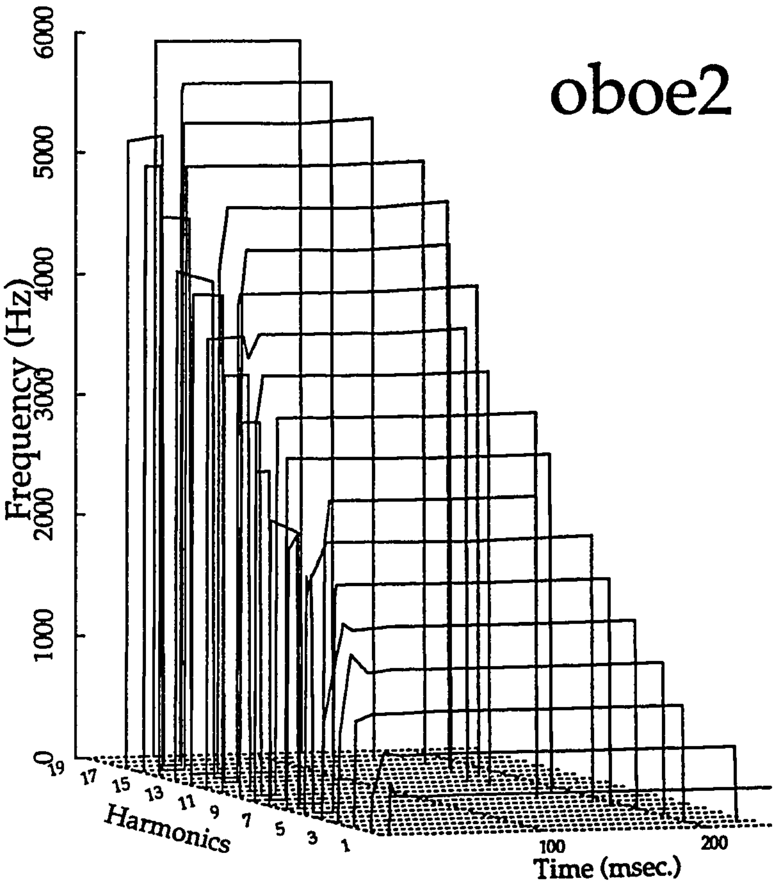


Figure 25. Frequency functions of the Stanford oboe tone, all harmonics, *line segment approximations.*

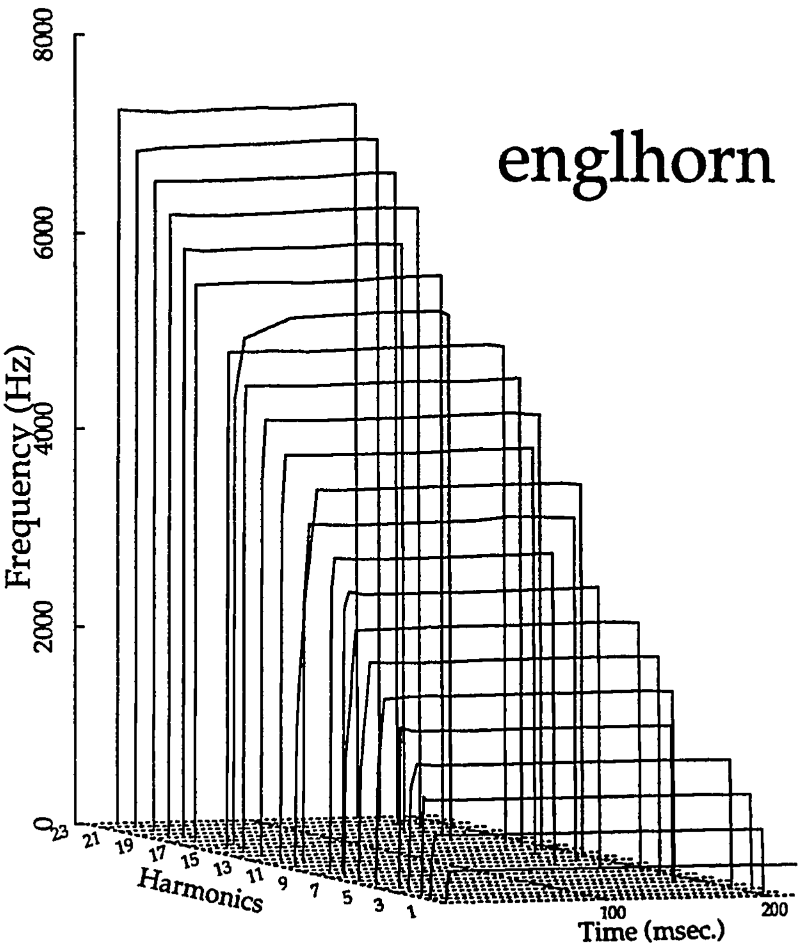


Figure 26. Frequency functions of the Stanford English horn tone, all harmonics, *line segment approximations.*

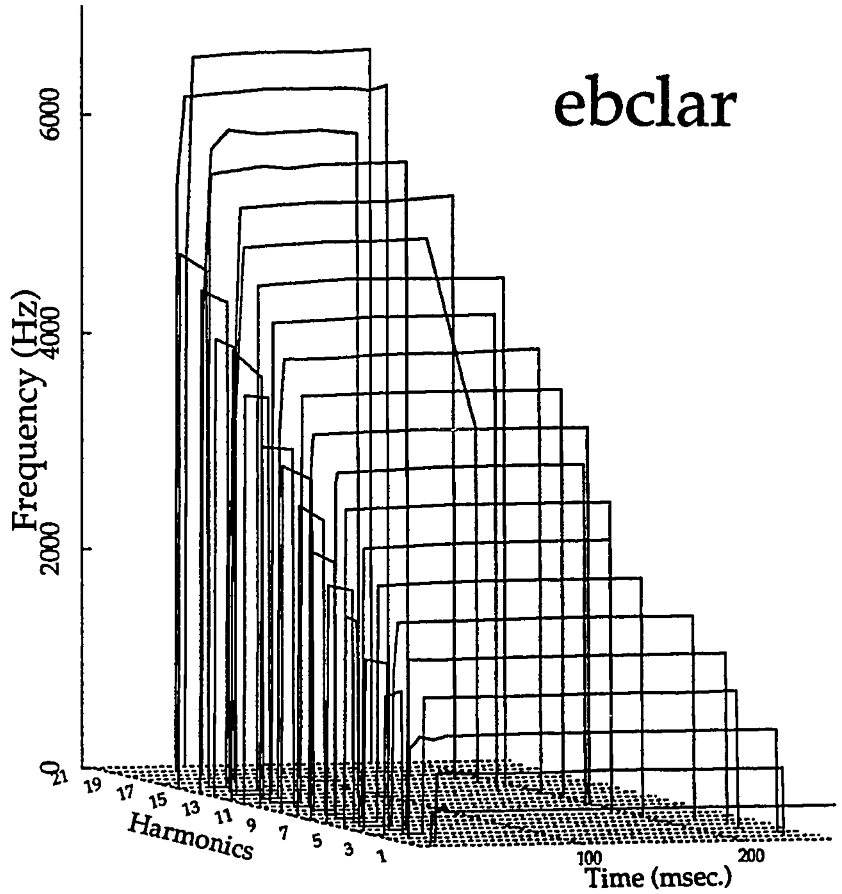


Figure 27. Frequency functions of the Stanford Eb clarinet tone, all harmonics, *line segment approximations.*

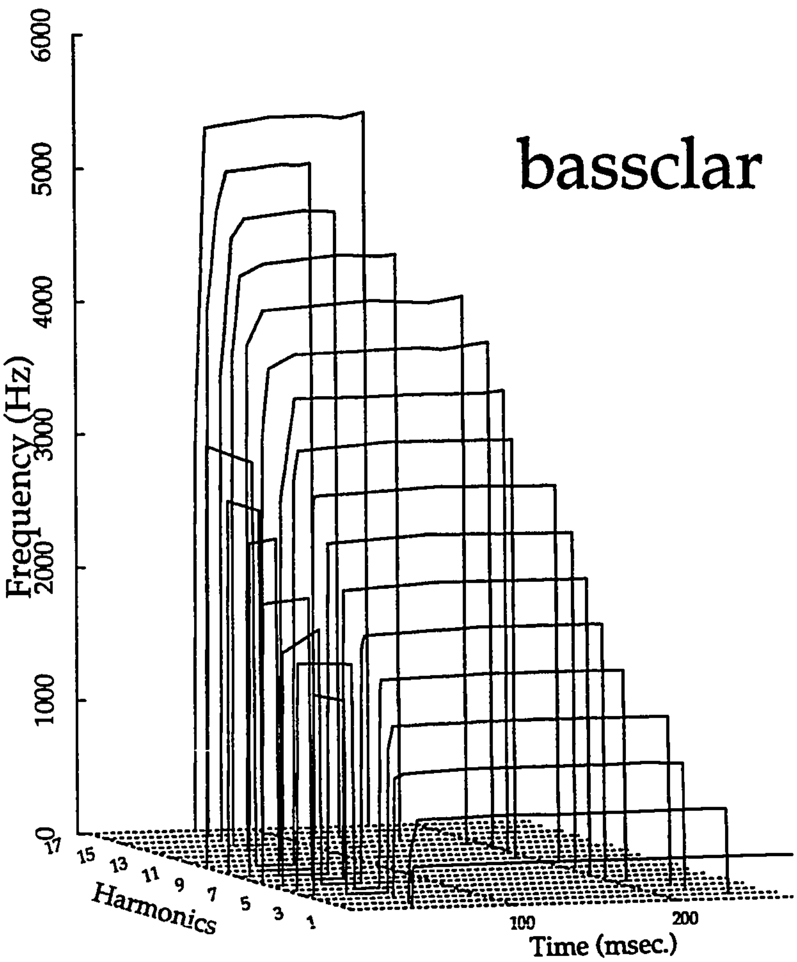


Figure 28. Frequency functions of the Stanford bass clarinet tone, all harmonics, *line segment approximations.*

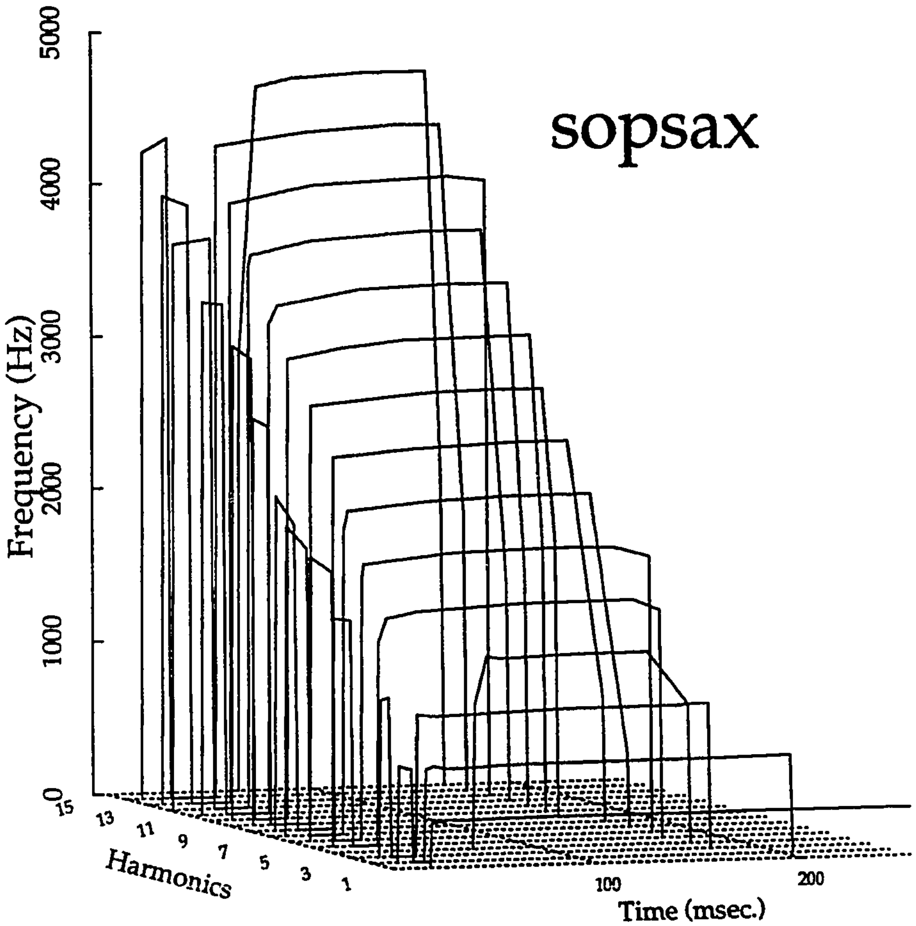


Figure 29. Frequency functions of the Stanford soprano saxophone tone, all harmonics, *line segment approximations.*

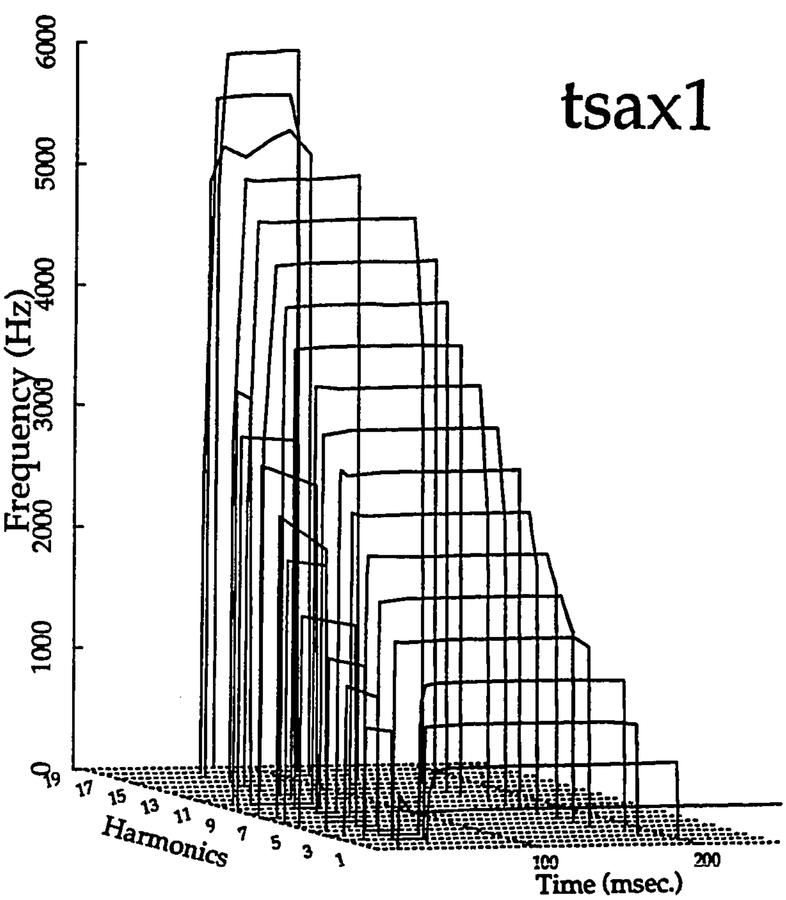


Figure 30. Frequency functions of the Stanford alto saxophone (played *mezzo forte*) tone, all harmonics, *line segment approximations*.

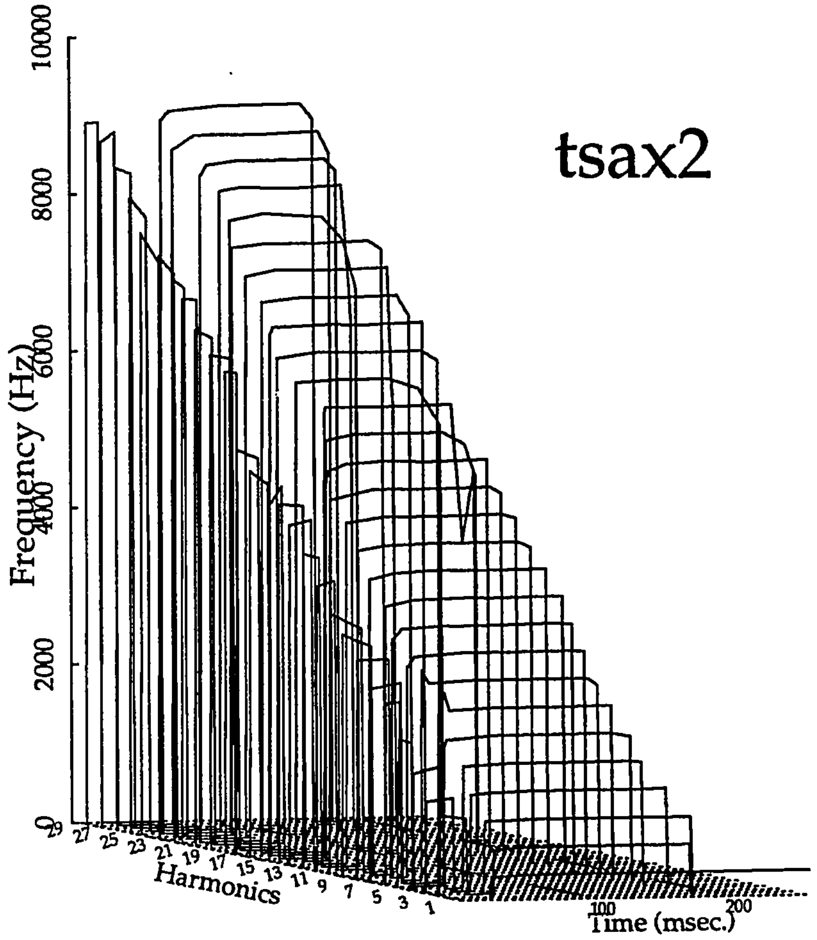


Figure 31. Frequency functions of the Stanford alto saxophone (played *forte*) tone, all harmonics, *line segment approximations.*

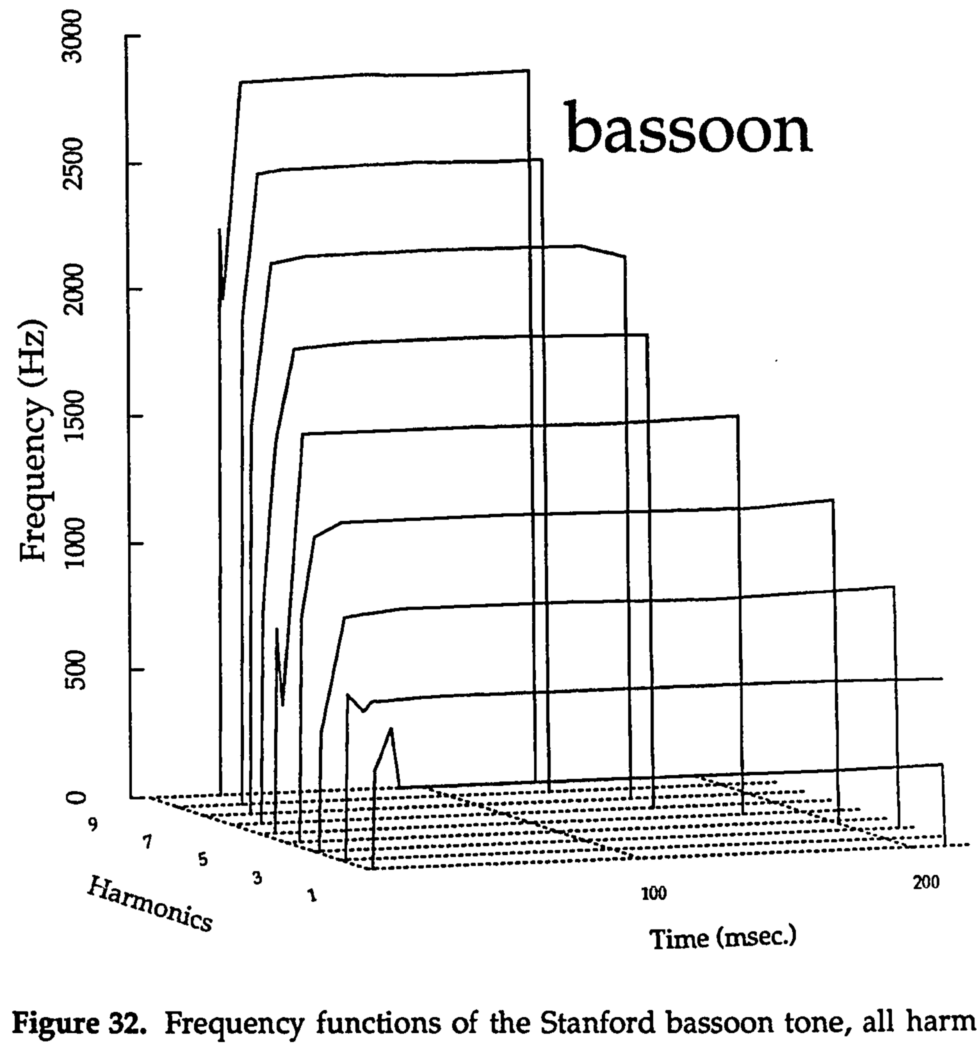


Figure 32. Frequency functions of the Stanford bassoon tone, all harmonics, *line segment approximations.*

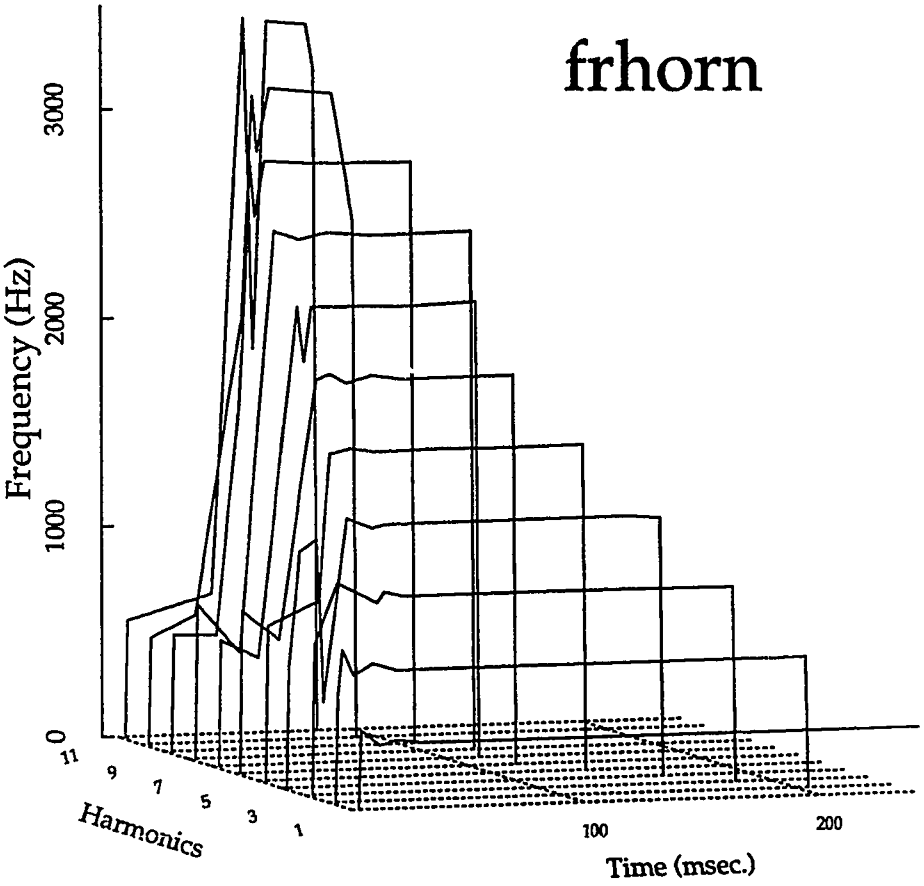


Figure 33. Frequency functions of the Stanford French horn tone, all harmonics, *line segment approximations.*

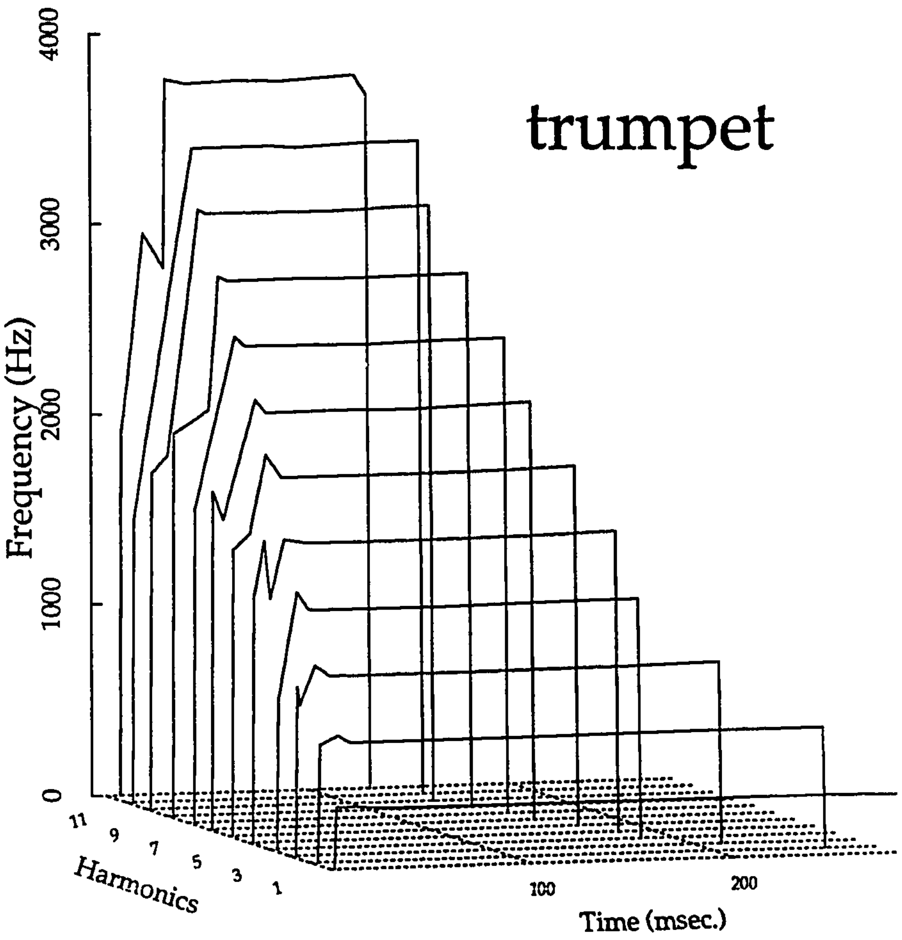


Figure 34. Frequency functions of the Stanford trumpet tone, all harmonics, *line segment approximations.*

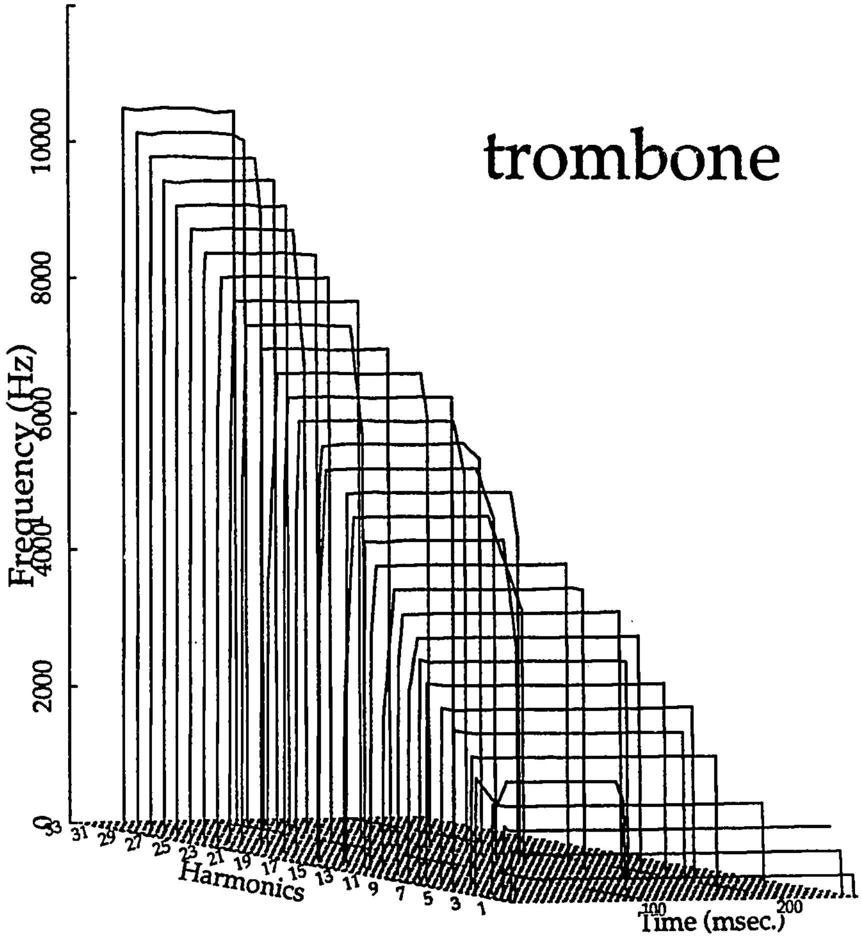


Figure 35. Frequency functions of the Stanford trombone (muted) tone, all harmonics, *line segment approximations.*

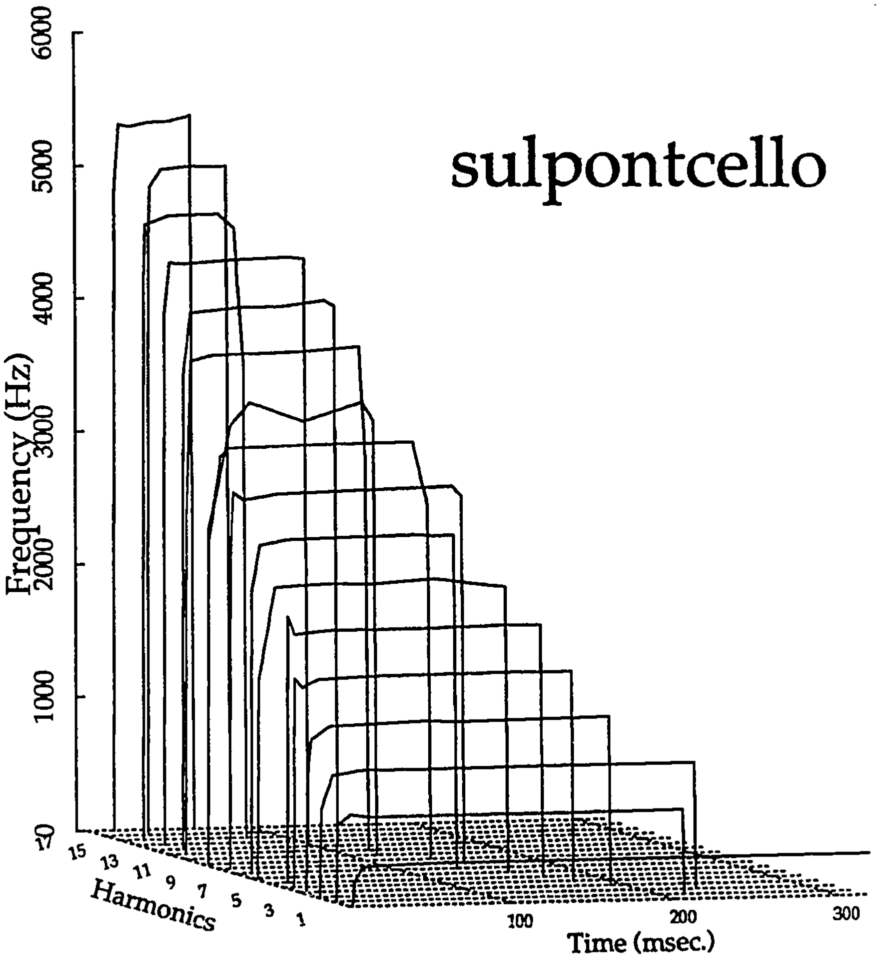


Figure 36. Frequency functions of the Stanford cello (bowed sul ponticello) tone, all harmonics, *line segment approximations.*

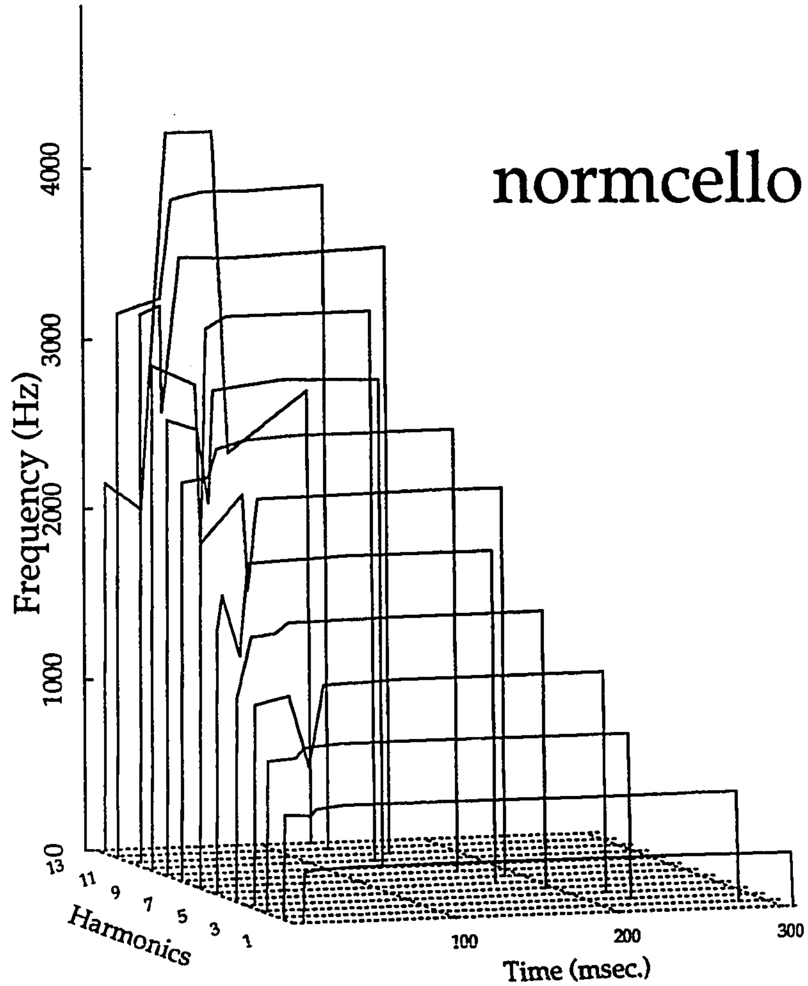


Figure 37. Frequency functions of the Stanford cello (bowed normally) tone, all harmonics, *line segment approximations.*

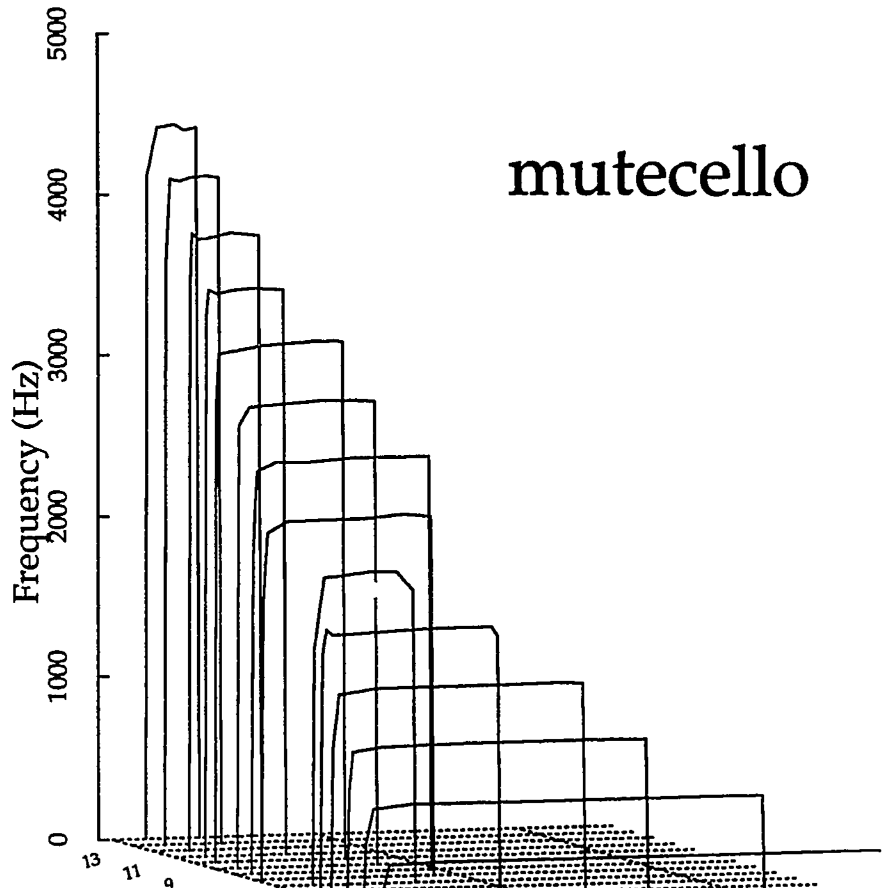


Figure 38. Frequency functions of the Stanford cello (muted, bowed normally) tone, all harmonics, *line segment approximations.*

Tests of Tone Naturalness and Identifiability

Grey ran experiments to examine the faithfulness of the line-segment approximations to their complex synthesis originals (Grey, 1975, Study A, pp. 25-41), and their recognizability as natural musical instruments (Grey, 1975, Study C, Experiment 2, pp. 70-74). Listener's abilities to distinguish the line-segment approximations from the complex synthesis was not far above chance level. Therefore there was little or no degradation in quality caused by the reduction process. Performance in an identification task (having previously been taught to associate each tone with a label) was also very good, which suggests that listeners find these tones to be natural-sounding instances of the instruments in question.[29] It is the present author's own experience that listeners generally find the tones to sound realistic. Musicians who hear these tones for the first time have little difficulty identifying them correctly from a list of the names of all fifteen instruments, with slight exceptions due to the uncharacteristic register for certain instruments (e.g., the bassoon's timbre on Eb4 is not a prototypical sound for the instrument) or easily confusable variants of the same instrument (the two alto saxophones or the three celli).

---

[29] The data from Grey's recognition study appears in Figure 56, and will be discussed at a later point.

Perceptual Equalization of the Tones done by Grey

An important objective of Grey's timbral similarity study (Grey, 1975, Study C, Experiment 1, pp. 58-69) was to discover perceptual factors of timbre perception that were not functions of the pitch, duration or amplitude of the tones. He therefore sought to eliminate as much as possible any differences between the tones on those three dimensions. Because the high-level nature of the additive synthesis and line segment approximations in these representations allow them to be easily manipulated, Grey was able to experiment with slight changes in these three attributes. Grey arbitrarily selected one of the tones as a standard (ebclar) and had listeners select one of several pitch instantiations of another tone as being closest in pitch (tones varied in 0.5 Hz increments). All 14 other tones were matched to the pitch of the standard. Similar studies were conducted so that tones could be equalized in loudness and duration as well. These equalized versions were what the listeners in Grey's similarity experiment heard.

Table 1 lists fifteen of the instruments used by Grey and some of their acoustical attributes. (Grey used a total of 16 tones; one of them, a second oboe, was not employed in the present study, so it is not listed here.) The physical differences in the tones that resulted from the perceptual equalizations can be observed in this table. The durations differed by as much as 101 ms. Comparing the tones by an average amplitude measure (RMS energy) shows large differences as well, and an average of the time-variant pitch analysis of each tone (using a pitch detector described by Gold

I.D.	Name	Grey's label	Gordon's label	Number of partials	duration (sec.)	RMS energy	PAT (sec.)	Avg. pitch (Hz.)
1	flute	FL	FL	13	0.280	.1893	.0109	310.5
2	oboe2	O2	O9	19	0.239	.1886	.0038	311.5
3	englhorn	EH	EH	23	0.218	.2182	.0007	313.3
4	ebclar	EC	EC	21	0.256	.1562	.0207	311.4
5	bassclar	BC	BC	17	0.270	.2069	.0427	309.6
6	sopsax	X3	SS	15	0.253	.1672	.0274	315.2
7	tsax1	X2	X6	19	0.250	.1483	.0497	314.9
8	tsax2	X1	X9	29	0.257	.1443	.0324	312.8
9	bassoon	BN	BN	9	0.221	.2084	.0020	309.4
10	frhorn	FH	FH	11	0.243	.1863	.0024	312.1
11	trumpet	TP	TP	12	0.280	.1117	.0021	315.5
12	trombone	TM	TM	33	0.255	.1485	.0074	313.6
13	sulpontcello	S1	V3	17	0.319	.0941	.0056	316.4
14	normcello	S2	V7	13	0.308	.1509	.0100	313.0
15	mutecello	S3	V6	14	0.248	.2057	.0097	311.6

Table 1. Acoustical properties and labelling system for the Stanford tones.

and Rabiner, 1969) shows differences as great as a quarter-tone (i.e., a frequency ratio of $\sqrt[24]{2}$).

Further Perceptual Equalizations of the Tones in the Present Study

Grey's similarity study involved comparisons between successively-presented tones. The tasks that accomplished the equalizations were well-suited for his experiment, since they too used successive comparisons. In contrast, the experiments to be reported in the present study had listeners evaluate tones that were combined into concurrently-sounding pairs. Equalization of pitch, loudness, and durations--and possibly other parameters--is of similar importance, so the suitability of Grey's equalizations is a concern.

Indeed, once tones had been mixed together into concurrent pairs it was clear that further equalizations were necessary. The main problems appeared to be pitch mismatch (in spite of Grey's equalization) and onset mis-alignment. Presumably in live performance situations, performers endeavor to correct these differences. Here it was desired that conditions be such that the instruments are presented in the most blended manner possible. Since the objective was to discover as yet *unknown* factors pertaining to blend, the presence of factors so obviously detrimental to blend would probably yield a trivial outcome, i.e., that blend was a function of how much the instruments differed in pitch or their onset alignment. Thus it was decided to try to minimize the role of these two parameters as much as possible without altering their identity or character.

The instruments' pitches were noticeably mismatched, resulting in the familiar warbling effect created by instruments that are not tuned to the same pitch. This resulted in the individual instruments sounding more distinctive, and hence less blended. The pitch differences could be broken down into differences in trajectory and average fundamental frequency or pitch. Since an instrument's pitch trajectory (the deviation from a perfectly constant pitch over the duration of the note) constitutes part of the musical signature and to some degree the timbre of the instrument as well, it did not seem appropriate to alter it. However, the *average* pitch is a factor more completely under the control of the performer and one which is expected to be altered in certain ways in ensemble playing. The logical choice was to to correct the tones so that they all had the same average pitch.

As mentioned earlier, the 15 tones were analyzed for their probable pitch by the Gold and Rabiner (1969) method. This analysis made a pitch estimate 4096 times a second, and checked for candidate pitches frequencies between a range of 20 Hz below and above the expected 311 Hz pitch. The averages are shown in column 9 of Table 1. To equalize the tones for pitch, the data for each of the tones was modified to correct for these pitch differences; for example, the bassoon was raised 1.9 Hz, and sulpontcello was lowered by 5.4 Hz, so each would have an average pitch of 311 Hz.

A second, similar problem pertained to the attack time of the tones. When combined in concurrent pairs, the tones in many pairs sounded as though they were not attacked at the same times, even though their physical onsets were simultaneous. Whereas this did not make a difference in Grey's successive-comparison tasks, here they had a detrimental effect on blend: the individual instruments in many cases would clearly stand out from one another. As before, the goal was to simulate what the typical performer would do to maximize attack-time unity in an ensemble, rather than alter the attack-envelope signature of the tone.

Each tone's attack-pattern causes it to have its own temporal *point of emphasis,* the moment of its "downbeat." When musicians play in an ensemble each instrumentalist delays or anticipates his or her time of onset (or adjusts the nature of the attack portion itself) relative to the conductor's beat, to create the sense of a common point of emphasis. The duration of time between the physical onset of a tone and the point of emphasis is called the *perceptual attack time,* or PAT (Gordon, 1984). Instruments with PATs of long duration, then, are begun slightly earlier than those with PATs with short durations, if the sense of a common attack point is desired. Some of the factors determining the PAT of complex, time-variant sounds are "the time of an absolute amplitude, relative amplitude, integration, or slope threshold crossed" (Gordon, 1984, p. i).[30] Gordon's study, coincidentally, investigated the PAT of the Stanford tones in particular. His study provides estimates of

---

[30] Gordon's PAT research is also described in Gordon (1987).

the temporal point of perceived attack for all of these tones, and one of his methods (p. 109, case 1 of Gordon, 1984) was used here to offset the start times of one tone from another. In other words, tone pairs were combined so that they did *not* begin at the same physical time. The PAT for flute, for example (see Table 1), indicates that when the flute tone is begun at time zero, its perceived moment of attack is at time 10.9 ms. To synchronize its attack with that of oboe2 (perceptual attack at 3.8 ms), the flute must begin 7.1 ms earlier than oboe2.

Figure 39 illustrates the nature of the offsets. The waveform of two instruments, tsax2 and bassoon, are aligned with each other to reflect their relative PAT offset. As close inspection will reveal, they do not begin at the same moment in time. Because tsax2 begins with a fairly lengthy attack noise, and bassoon has an abrupt attack, their moment of perceived emphasis occurs at very different times relative to their own objective beginnings.

These were the only equalizations that were added to the Stanford tones for the purpose of the experiments reported here. The matter of tone duration did not seem to be of great importance: in previewing the tones differences were not noticeable, and in the actual experiments, listeners never remarked about this being a factor in their judgments. Further equalization of amplitude was also not pursued.[31]

---

[31] The present author did, however, ask several other listeners to evaluate whether the amplitude levels of the Stanford tones seemed to be equal. For comparison, the listeners also evaluated an equalization method based on RMS energy. Opinions differed, but slightly more listeners felt that the Stanford amplitude levels obtained more uniform loudness.

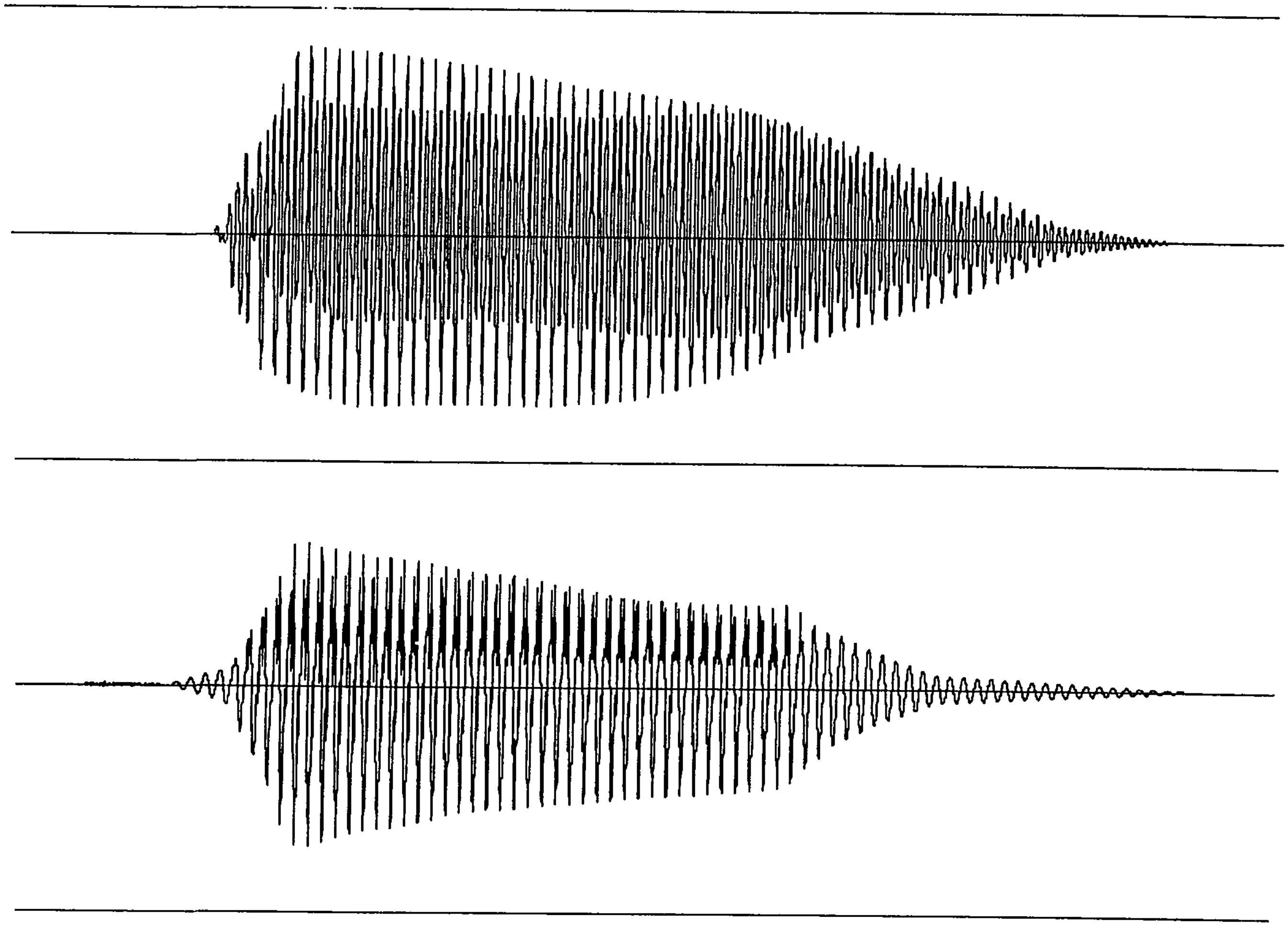


Figure 39. Waveforms for instruments tsax2 and bassoon, aligned in time to reflect their perceptual attack times (PAT).

The equalizations of pitch and attack time produced a remarkable improvement in the quality of the concurrently presented tones. The tones now sounded as blended as possible, at least as far as attack time, duration, pitch and loudness were concerned; this made it possible to examine other, non-obvious factors affecting their blend. All the experiments reported in this study used these equalizations.

Acoustical Properties of Single Tones

The objective of the present research is to identify physical factors determining the perception of blend when the Stanford tones are combined in concurrently-sounding pairs. In order to identify the acoustical properties of tones in combination, it will be valuable to identify their *individual* physical properties in as much detail as possible. Grey (1975) discovered a number of *distinctive features* which seemed to determine the degrees of similarity among the Stanford tones, relating to spectral concentration, onset, and harmonic amplitude envelope synchrony. For the purpose of the present study the 15 Stanford tones are analyzed for a set of ten acoustical properties, some of which are related to the distinctive features revealed in Grey's study. Those factors that are similar to Grey's are of special interest because it is anticipated that they will exert some influence on the perception of concurrently-presented tones as well. Other properties suggest themselves as important potentially relevant as a result of the literature review in Chapter 2. However, it is important to note that they themselves are not distinctive features (since they are not the outcome of an experimental study) but merely physical features of the tones. The ten acoustical properties are: centroid, acoustic dissonance, precedent noise, perceptual attack time, amplitude, harmonic envelope synchrony, peak synchrony, onset/offset synchrony, harmonicity, and pitch deviation. An eleventh property, obtained from some of Grey's own studies of the Stanford tones, is recognizability; it is distinguished from the other ten properties because it is not strictly an "acoustical" property. Later, physical properties arising from the interaction of tones will be considered.

1. *Centroid*

One factor expected to be important in the blend judgments pertains to the distribution of spectral energy in the tones. One method of characterizing this, as mentioned earlier, is the *centroid* formula. The centroid of a spectrum is calculated by weighting each frequency component by its linear amplitude, summing all such values, and dividing it by the sum of the amplitudes alone. The division step cancels out the amplitude units and leaves a single frequency which identifies the midpoint of spectral energy concentration, or "center of gravity." Two simple spectra are shown in Figure 40 to demonstrate centroid calculation. Four harmonics of a steady-state complex with a 100 Hz fundamental and using a simple linear amplitude scale ranging from 0 to 10 are shown in both spectra. The centroid calculation for a steady-state spectrum is shown in Formula 1. The centroid for the first spectrum is solved as follows:

$$\frac{\sum_{n=1}^{h} f_n a_n}{\sum_{n=1}^{h} a_n}$$

where

h = the number of harmonics

f_n = the frequency of harmonic n

a_n = the amplitude (linear) of harmonic n

Formula 1. Centroid formula for a steady-state spectrum.

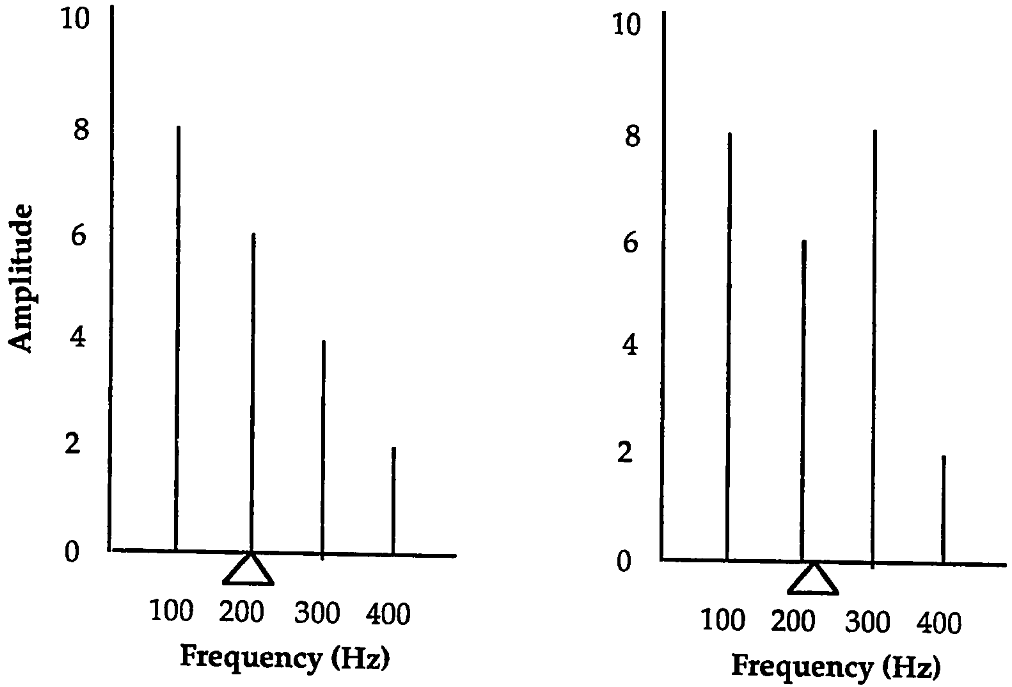


Figure 40. Two spectra with different centroids, the first at 200 Hz., the second at 216.7 Hz.

$$\frac{(8 * 100) + (6 * 200) + (4 * 300) + (2 * 400)}{8 + 6 + 4 + 2} = \frac{4000}{20} = 200 \text{ Hz}$$

The second spectrum, with its stronger third harmonic, shifts the spectral concentration of energy slightly towards the upper harmonics. This results in a slightly higher centroid:

$$\frac{(8 * 100) + (6 * 200) + (8 * 300) + (2 * 400)}{8 + 6 + 8 + 2} = \frac{5200}{24} = 216.7 \text{ Hz}$$

The "fulcra" under each spectrum in Figure 40 (intended to evoke the image of a seesaw), identifies the value of the centroid. The seesaw metaphor helps emphasize the notion of "center of spectral symmetry" and "point of spectral balance" which is inherent in the centroid concept.[32]

Calculating centroid for the Stanford tones requires an expansion of the formula to account for the temporally-varying nature of the harmonics. Essentially this can be accomplished by calculating frequency and amplitude weights over the duration of the tone at each millisecond in time, and then averaging into a single value, as Formula 2 shows. Calculated in this way, the centroids for the 15 tones result in the values shown in Figure 41.

Two further representations of centroid are shown in Figures 42 and 43. Using the seesaw metaphor mentioned earlier, each centroid in Figure 42

---

[32] The idea for the "fulcrum" illustration comes from Grey and Gordon (1978), p. 1497.

$$\frac{\sum_{t=0}^{d} \sum_{n=1}^{h} f_{tn} a_{tn}}{\sum_{t=0}^{d} \sum_{n=1}^{h} a_{tn}}$$

where

d = the duration of the tone
h = the number of harmonics
f_{tn} = the frequency of harmonic n at time t
a_{tn} = the amplitude (linear) of harmonic n at time t

Formula 2. Centroid formula for a tone with temporally variable amplitude and frequency functions for each harmonic.

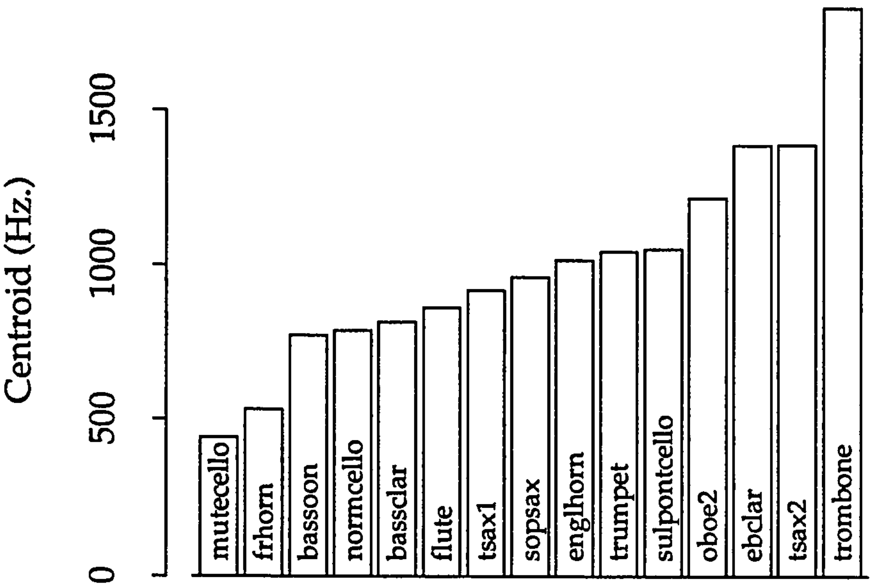


Figure 41. Centroids for the Stanford tones, calculated with Formula 2.

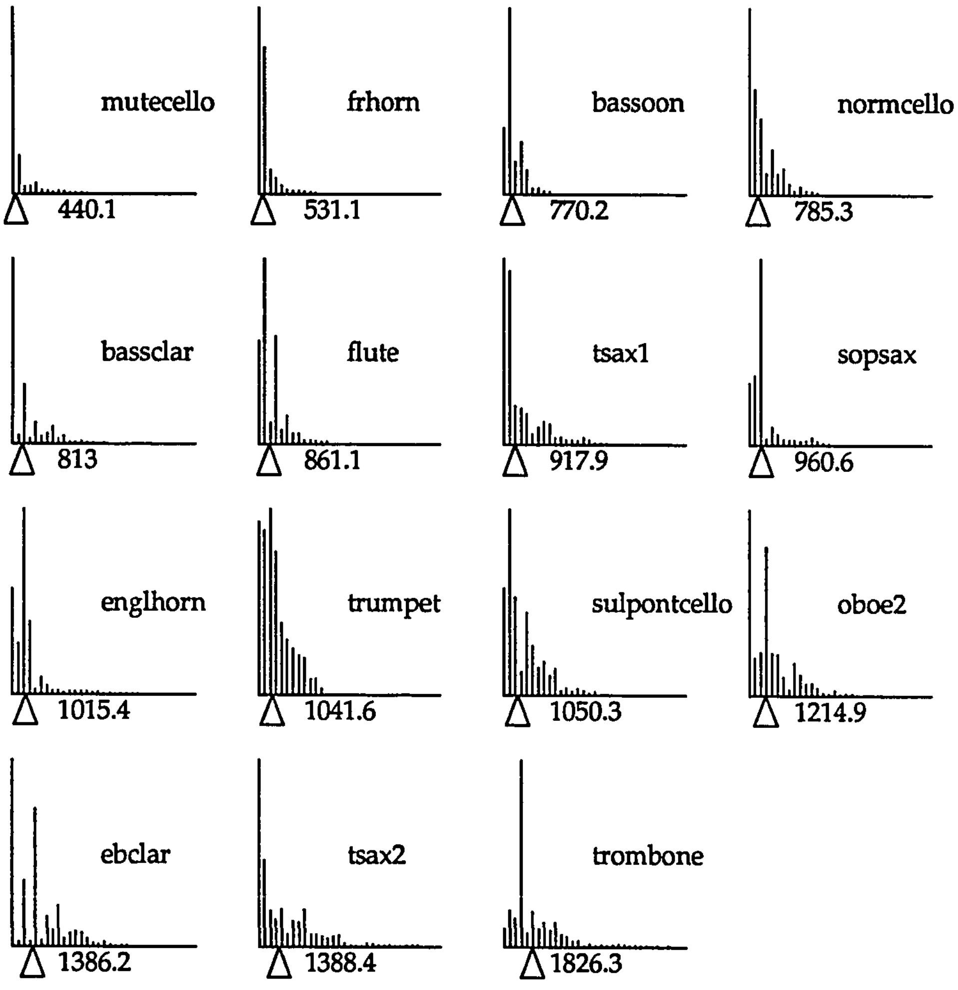


Figure 42. Spectral envelopes for each of the Stanford tones, with frequency on the x-axis, amplitude on y-axis. The ordering of instruments is according to centroid (see Figure 41); envelopes were calculated according to Formula 3. A triangle identifies the location of the centroid on the x-axis.

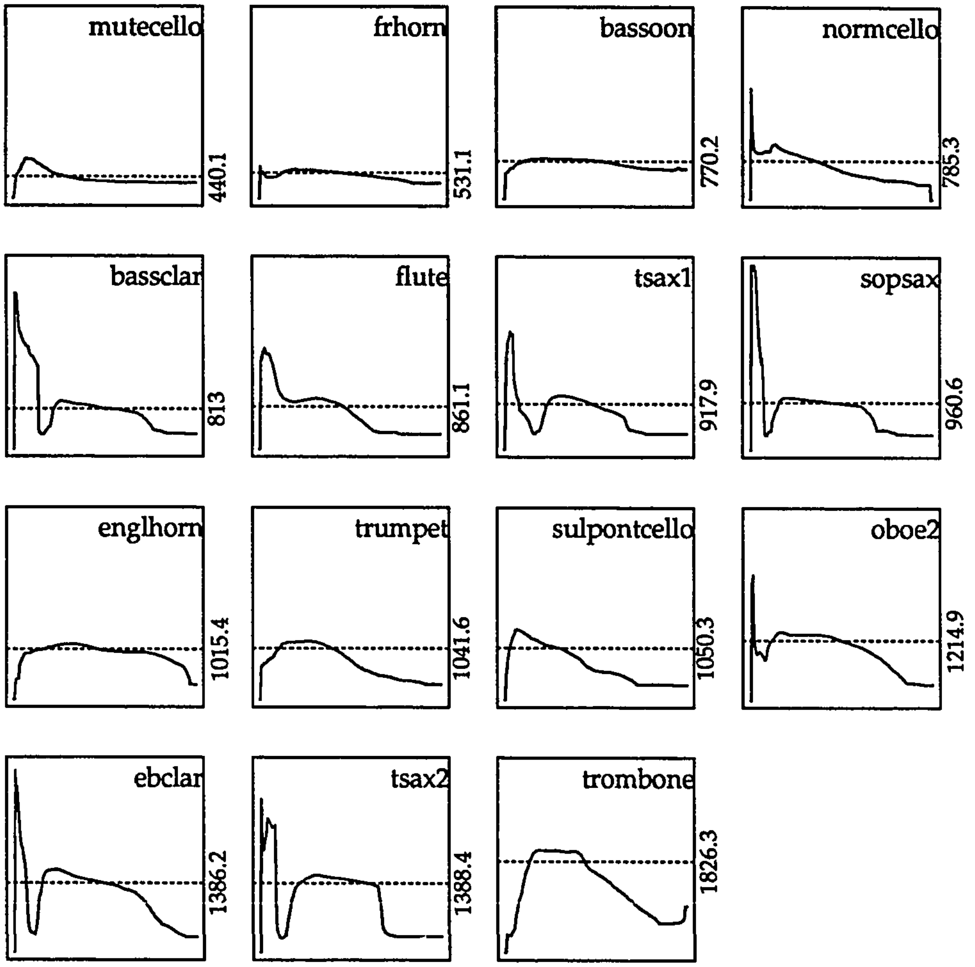


Figure 43. Centroid envelopes for each of the Stanford tones, with time on the x-axis and centroid on the y-axis, calculated according to Formula 4. The ordering is according to the centroids in Figure 41; a dotted line identifies the location of each centroid on the y-axis.

is shown with respect to the tone's spectral envelope (obtained by using Formula 3). Figure 43 shows the centroid temporal envelope of each tone; that is, the centroid calculated for each moment in time according to the method in Formula 4. Time is on the x-axis and centroid on the y-axis, while a dotted line identifies the location of the centroid.

2. *Acoustic Dissonance*

The usage of the word dissonance among western "classical" musicians can take on several meanings. It can refer to categories of musical intervals taught in tonal music theory (for example, minor seconds are more dissonant than major seconds), an attribution given to a harmony based on its syntactical function among other harmonies, or the quality of "smoothness" or "roughness" that arise from interference patterns between the components due to their closeness in frequency. The latter concept was explored by Plomp and Levelt (1965), who obtained consonance judgments of concurrently-sounding sine tones at various frequency separations from listeners. Their study showed that maximal dissonance is obtained when the absolute difference in frequency of two sines is 25% of the *critical band* (see Scharf, 1970) of their mean frequency. For example, given a mean frequency of 440 Hz (which has a critical bandwidth of 70.8 Hz), maximal dissonance is obtained for pair of sine tones having frequencies of 431.15 Hz and 448.85 Hz.

A useful application of this measure of dissonance is to characterize the quality of larger aggregates of concurrently-sounding sine tones, such as those making up complex musical timbres. Hutchinson and Knopoff (1978)

For each harmonic n of the tone, plot the value for:

$$\frac{\sum_{t=on}^{off} a_{tn}}{off - on}$$

where

on = the onset time of harmonic n
off = the offset time of harmonic n
a_{tn} = the amplitude (linear) of harmonic n at time t

Formula 3. Calculation of a steady-state spectrum envelope for a tone with temporally variable amplitude and frequency functions for each harmonic.

For each time t over the duration of the tone, plot the value for:

$$\frac{\sum_{n=1}^{h} f_{tn} a_{tn}}{\sum_{n=1}^{h} a_{tn}}$$

where

h = the number of harmonics

f_{tn} = the frequency of harmonic n at time t

a_{tn} = the amplitude (linear) of harmonic n at time t

Formula 4. Calculation of a centroid envelope for a tone with temporally variable amplitude and frequency functions for each harmonic.

proposed a method for calculating the dissonance for any collection of sine components by weighting the dissonance measurement of every possible pair of sine tones according to their amplitudes, summing all such weights, and then factoring out the amplitude to obtain a single dissonance value. Their method is shown in Formula 5.

This method was used to calculate the dissonance of the single Stanford tones; dissonance calculations were made at each millisecond in time over the duration of each tone, and averaged over the number of calculations. The results are shown in Figure 45. The range of dissonance goes from 0 to 1, representing complete consonance and maximal dissonance, respectively. Note that for the single tones, the dissonance levels are very small.

2. *Duration of Precedent Noise*

Another factor expected to play a role pertains to the attack of the tone. One such measure used by Grey (1975) in characterizing the Stanford tones was the presence of a noisy attacks, described as "precedent high-frequency, low amplitude energy, most often *inharmonic energy,* during their attack segment" (p. 67, emphases Grey's). For the present study, *precedent noise* was identified by noting when the amplitude of at least one harmonic rose and fell back down to zero at the beginning of the tone, preceding the principle onset of the tone. Figure 46 shows the approximate durations of precedent noisy portions of the various tones; durations were calculated by averaging

$$\frac{\sum_{i=1}^{h} \sum_{j=i}^{h} a_i * a_j * g\left(\frac{|f_i - f_j|}{\mathrm{CBW}\left(\overline{f_{ij}}\right)}\right)}{\sum_{i=1}^{h} a_i^2 * 2}$$

where

h = the total number of harmonics in one or more tones

a = the amplitudes (linear) of all harmonics

f = the frequencies of all harmonics

$|f_i - f_j|$ = the absolute difference of f_i and f_j

$\overline{f_{ij}}$ = the mean of f_i and f_j

$\mathrm{CBW}(x)$ = the critical bandwidth for frequency x. See Formula 6.

$g(x)$ = See Figure 44.

Formula 5. Measure of dissonance for an aggregate of harmonics from one or more complex tones, from Hutchinson & Knopoff (1978).

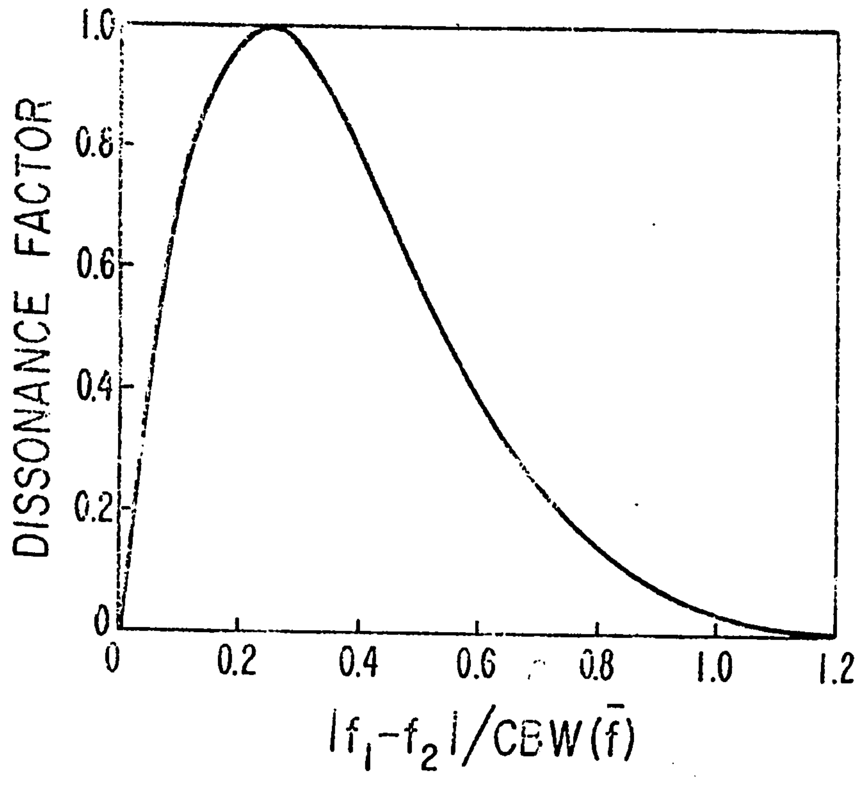


Figure 44. Function used for estimating the dissonance of two sine tones based on the critical band of their average frequency. From Hutchinson and Knopoff (1978).

$$\left(6.23 * f^2\right) + \left(93.39 * f\right) + 28.52$$

where

$$f = \frac{\text{frequency}}{1000}$$

Formula 6. Critical band formula from Moore & Glasberg (1983).

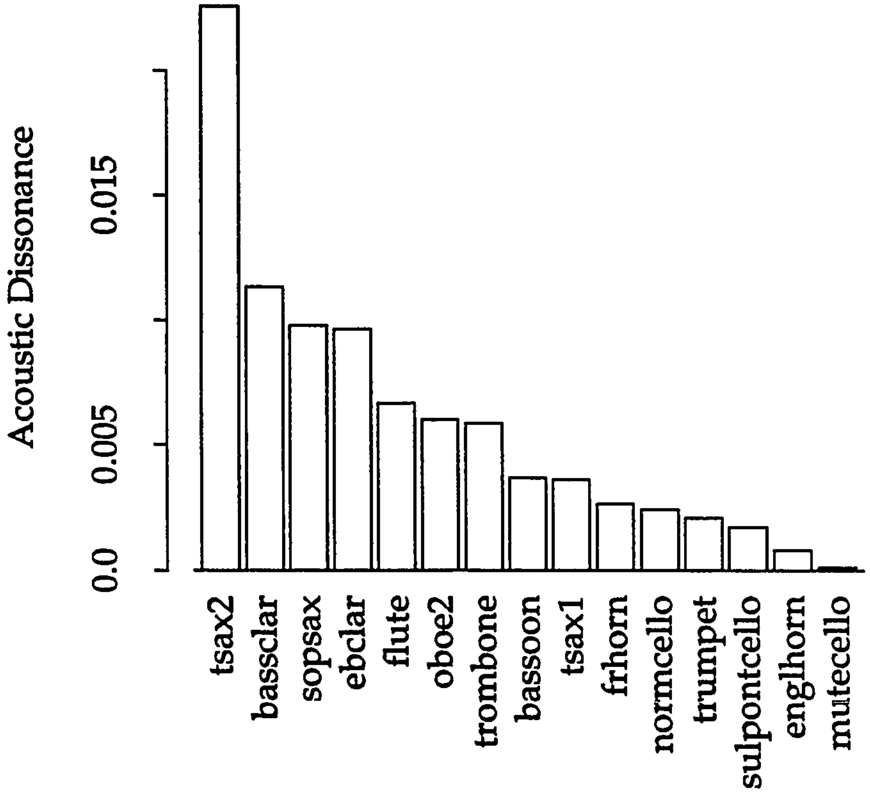


Figure 45. Dissonance measurements for the Stanford tones based on Formula 5. Larger values indicate greater amounts of dissonance.

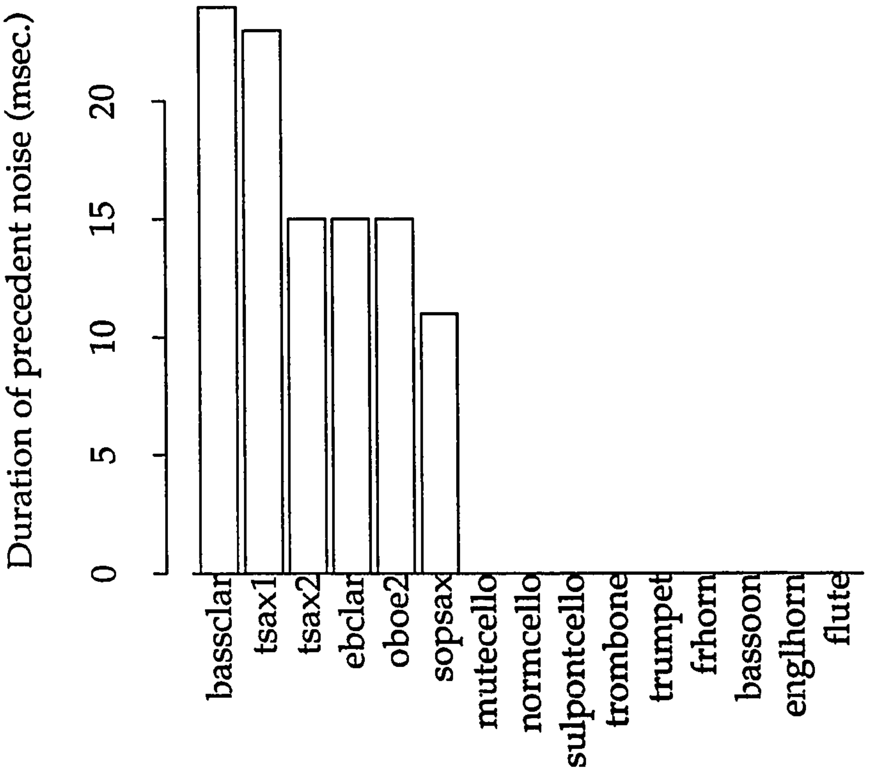


Figure 46. Duration of precedent noise for the Stanford tones.

over the durations of the portions of all harmonics that had a precedent portion as described above.[33]

3. *Duration of Perceptual Attack Time*

Another useful measure of the attack portion is the duration of Gordon's *perceptual attack time* (PAT): the amount of time it takes an instrument to arrive at its point of emphasis is to some degree the duration of its attack. To emphasize its interpretation as the *duration* of attack as opposed to *time* of attack, it is referred to from here on as DPAT. Figure 47 shows the DPATs for the fifteen tones in magnitude order. Figure 48 provides an additional visualization of the attack properties of the tones. Each curve shows the overall amplitude envelope of the tone, calculated by Formula 7. An "x" marks the point of the DPAT, while, where present, a small vertical line marks the time at which a noisy onset portion ended.

4. *Amplitude*

It was observed earlier that Grey's subjects' loudness equalizations resulted in, obviously, different physical amplitudes. The peak *amplitude* of each instrument are shown in Figure 49. The possibility that these equalizations might be a factor the blend judgments in the experiments, instead of being transparent to the listener as intended, cannot be ruled out.

---

[33] Note that only six instruments show any precedent attack noise at all. Although Grey (1975) identified the three string instruments (sulpontcello, normcello, mutecello) as having noisy precedent energy as well, no such noise was found for the strings using the present criteria.

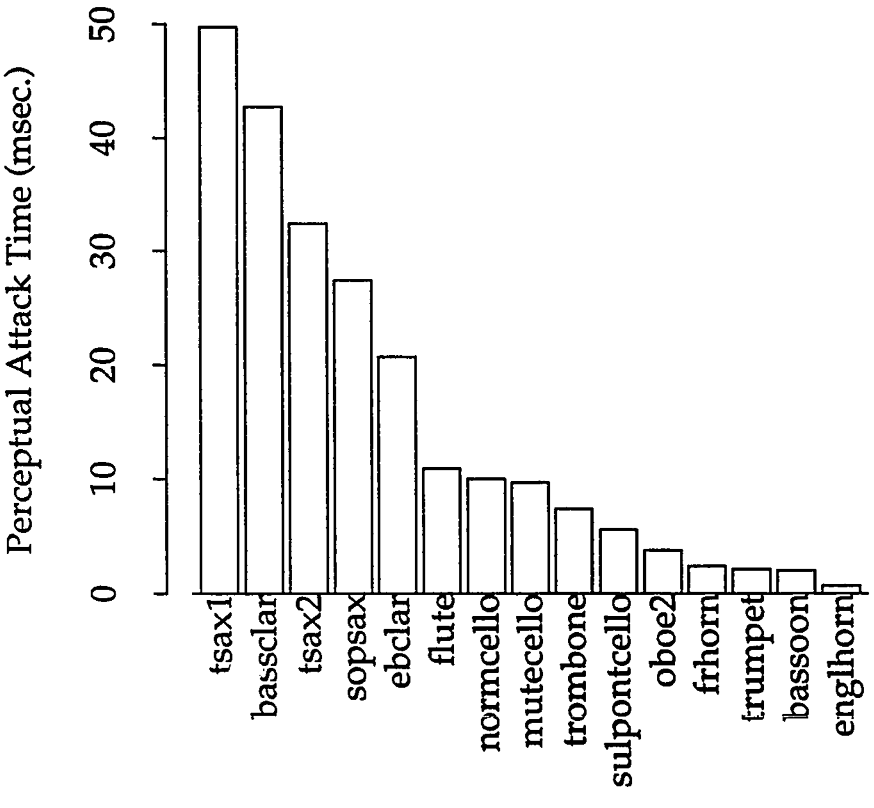


Figure 47. Duration of Perceptual attack time (DPAT) for the Stanford tones. Values are from Gordon (Gordon, 1984, p. 109, case 1).

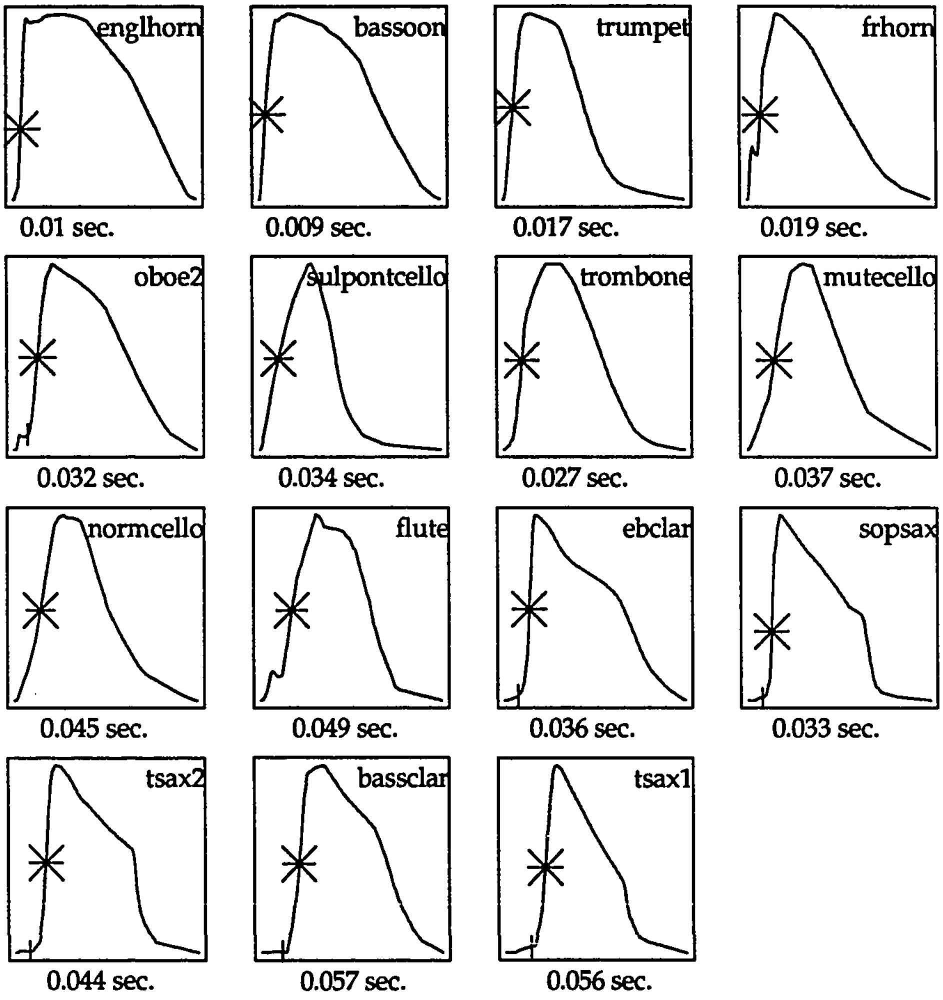


Figure 48. Amplitude curves for each of the Stanford tones, with time on the x-axis and amplitude on the y-axis, calculated according to Formula 7. Each curve is scaled to its own maximum (i.e., relative amplitudes of the tones are not apparent). An "x" marks the DPAT while a small vertical line marks the time at which a noisy onset portion ended (when one was present). Plots are ordered according to their DPAT value, which is indicated below each tone.

For each time t over the duration of the tone, plot the value for:

$$\frac{\sum_{n=1}^{h} a_{tn}}{h}$$

where

h = the number of harmonics

a_{tn} = the amplitude (linear) of harmonic n at time t

Formula 7. Calculation of an amplitude envelope for a tone with temporally variable amplitude functions for each harmonic.

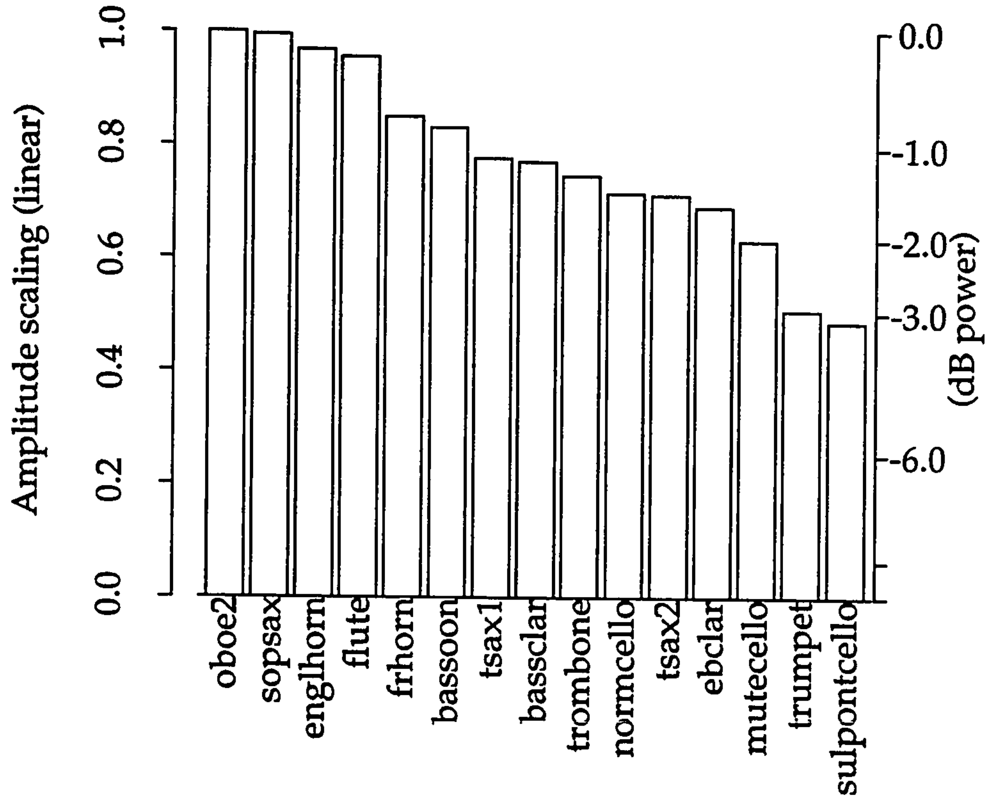


Figure 49. Peak amplitudes for each of the Stanford tones, calculated from peak values yielded by Formula 7.

Any loudness equalization is always an approximation; moreover because Grey equalized the tones in experiments using successive comparison, there is no certainty that this would result in levels that would be perceived as equal in concurrent presentation. Therefore, this measure can be used later to determine if the amplitude settings actually affected the blend judgments.

5 - 7. Harmonic Synchrony

The degree to which the amplitude envelopes of a single instrument match each other in shape is called *harmonic synchrony*. A cursory examination of Figures 9 through 23 reveals that the amount of similarity among a tone's individual harmonic envelope shapes is quite variable over the group of 15 tones. The harmonics of the brass tones, for example, show a great deal of synchronized movement, while the harmonics of the flute and strings (with the exception of mutecello) show more independence among patterns. There are also other aspects of asynchrony, for example, due to differences in onset times among the harmonics and the difference in time at which each harmonic arrives at its peak value. Unfortunately, there is no unique, single method of quantifying the synchrony of a particular tone. Grey had identified a kind of synchrony factor pertaining to "the presence of synchronicity in the collective attacks and decays of upper harmonics" (p. 61), but did not explicitly define how this was measured. For the purposes of the present study, three different measures of synchrony are defined:

(1) *Harmonic envelope synchrony:* the overall amount of amplitude envelope similarity among the harmonics. This was calculated by

correlating the amplitude envelope of each harmonic with every other, and then averaging these correlations together; thus, larger values mean greater amounts of harmonic envelope synchrony.[34] The values are shown in Figure 50.

(2) *Peak synchrony:* the variance in time at which all the harmonics arrive at their peak values. This is shown in Figure 51. This measure was obtained simply by collecting the points in time (in milliseconds) at which each harmonic arrived at peak amplitude, and taking their standard deviation. Lower standard deviations therefore indicate greater amounts of peak synchrony. However, in Figure 51 these values have been inverted so that this measure is analogous to the meaning of the previous one (i.e., greater synchrony is indicated by larger values).

(3) *Onset/offset synchrony:* the variance in time at which all the harmonics enter and exit. This is shown in Figure 52. This measure is similar to the one above, except that the standard deviation of the times of entry and the standard deviation of the

---

[34] The number of points compared in each correlation differed depending on which pair of harmonics were considered, since upper harmonics often enter later and end earlier than upper harmonics and hence have fewer points. Since each correlation involved different degrees of freedom, the probabilities for determining significance differed for each r value. This was accounted for by dividing each r by the necessary r to achieve a probability of $p < 0.05$. That is why the averaged correlation values in Figure 50 do not fall within the customary range of -1.0 to +1.0. For example, for twelve points (10 degrees of freedom), $p < 0.05$ when $r = .576$. A correlation with $r = .891$ would be recorded as .891/.576 = 1.547.

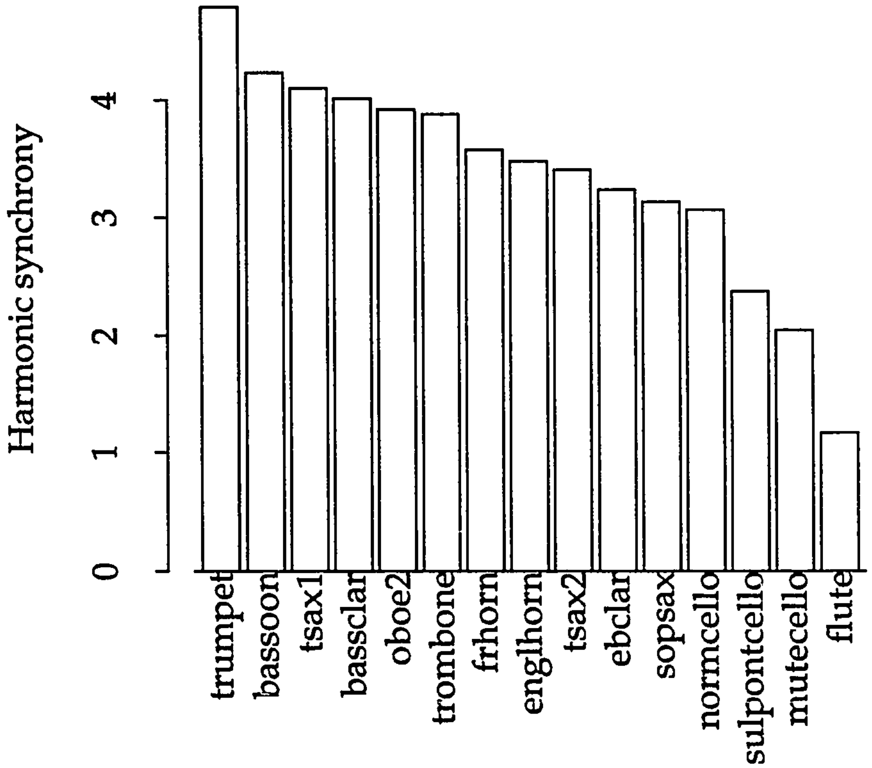


Figure 50. Measure of harmonic envelope synchrony for the Stanford tones, representing the amount of correlation among the amplitude envelopes of all harmonics in the tone.

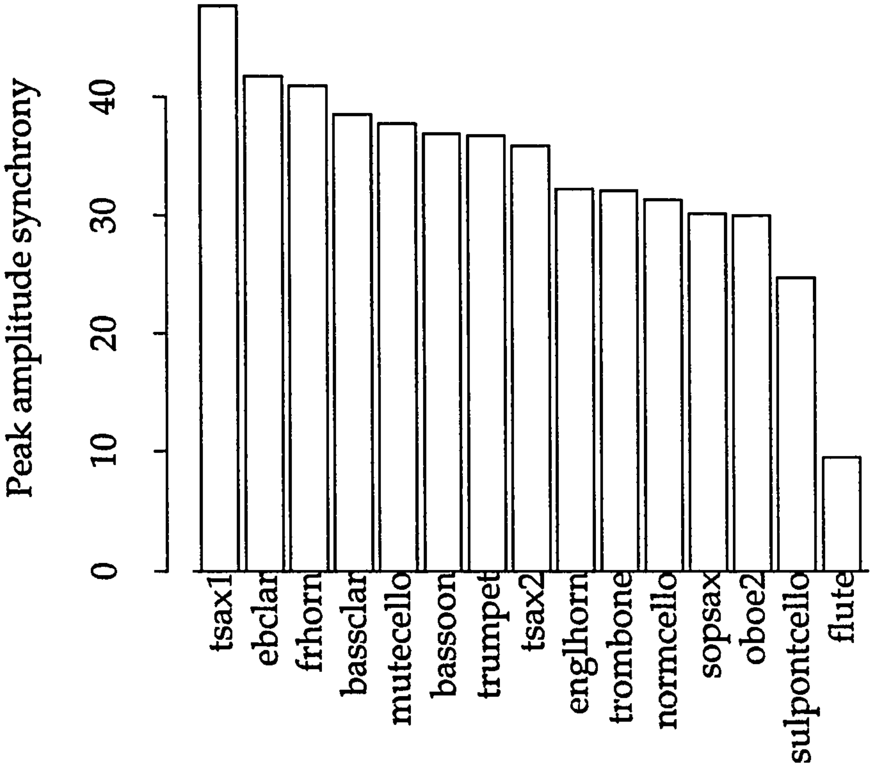


Figure 51. Measure of peak synchrony for the Stanford tones, showing the variance in time at which all the harmonics arrive at their peak values. Values are inverted so larger values indicate greater synchrony.

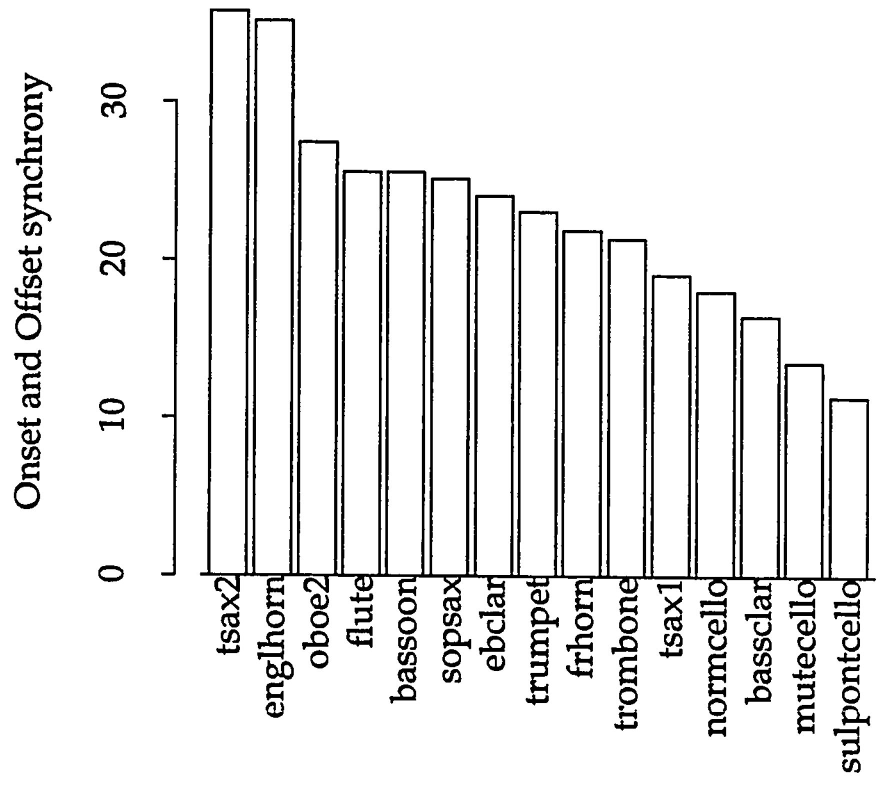


Figure 52. Measure of onset/offset synchrony for the Stanford tones, showing the variance in time at which all the harmonics enter and exit. Values are inverted so larger values indicate greater synchrony.

offset times were averaged together. For the same reasons as before, these values have been inverted in Figure 52.

8. Harmonicity

Another measure concerns the instrument's overall degree of *harmonicity*. None of the instruments had harmonics whose frequency ratios were perfectly integral throughout the duration of the tone. In fact, as Figures 24 through 38 illustrated, the dynamic nature of the individual frequency functions of the harmonics was such that no single set of ratios was ever constant throughout the duration of any tone. Therefore, any characterization of the degree of harmonicity for a tone had to in some way average over the duration of the tone. The following measure of harmonicity was devised, and the results are shown in Figure 53.

(1) For every possible pair of harmonics, at every millisecond in time, the ratio between the actual frequency values of the two harmonics and the ratio between the theoretical values of the two harmonics (i.e., their pure, integral-ratio frequency values) was calculated. Any time at which either harmonic was silent was ignored.

(2) The absolute differences of these two ratios were added to a running total of all absolute differences. At the end, this sum was divided by the total number absolute differences calculated. The result is a measure in which larger values indicates greater harmonicity.

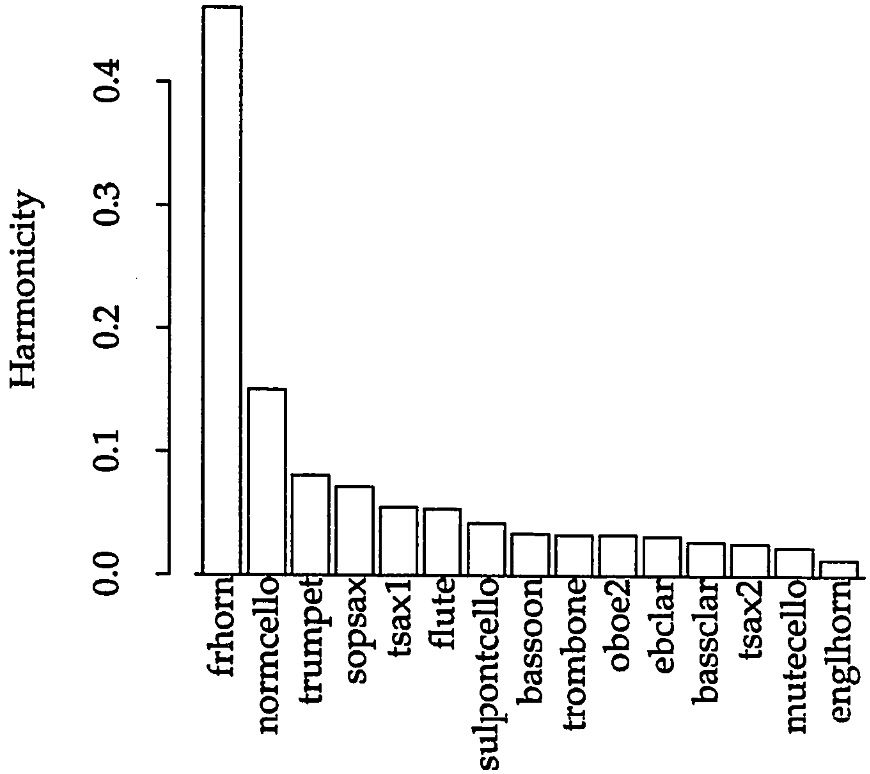


Figure 53. Measure of harmonicity for the Stanford tones, based on the amount of deviation from pure integer ratios among the frequency values of the harmonics. Larger values indicate that the relationships among the harmonics approach purer ratios.

9. Pitch Deviations

Tones were also not constant in fundamental frequency and hence, pitch; furthermore, different instruments showed different amounts of fluctuation in pitch over the duration of the tone. To measure this, the pitch curve of each tone was extracted by the pitch detection method by Gold and Rabiner (1969). The standard deviation of each pitch curve resulting from that analysis was used as the measure of *pitch deviation*; these values are shown in Figure 54. For further visualization, Figure 55 shows the entire pitch curve from which the standard deviations were calculated.

10. Recognizability

In Chapter 2 a paradigm for explaining the perception of blend called "segregation by identification" was mentioned. This states that the degree to which individual instruments in a concurrent arrangement can be recognized by a listener is one possible determinant of how well they blend. What makes an instrument more recognizable may be due to a naturally salient quality, cognitive familiarity, prototypicality, or numerous other factors. These same attributes could play a role in making an instrument stand out from the others in a concurrent presentation of instruments. Conversely, a pair instruments which are both hard to recognize in isolation may be harder for the ear to decompose into component sounds when

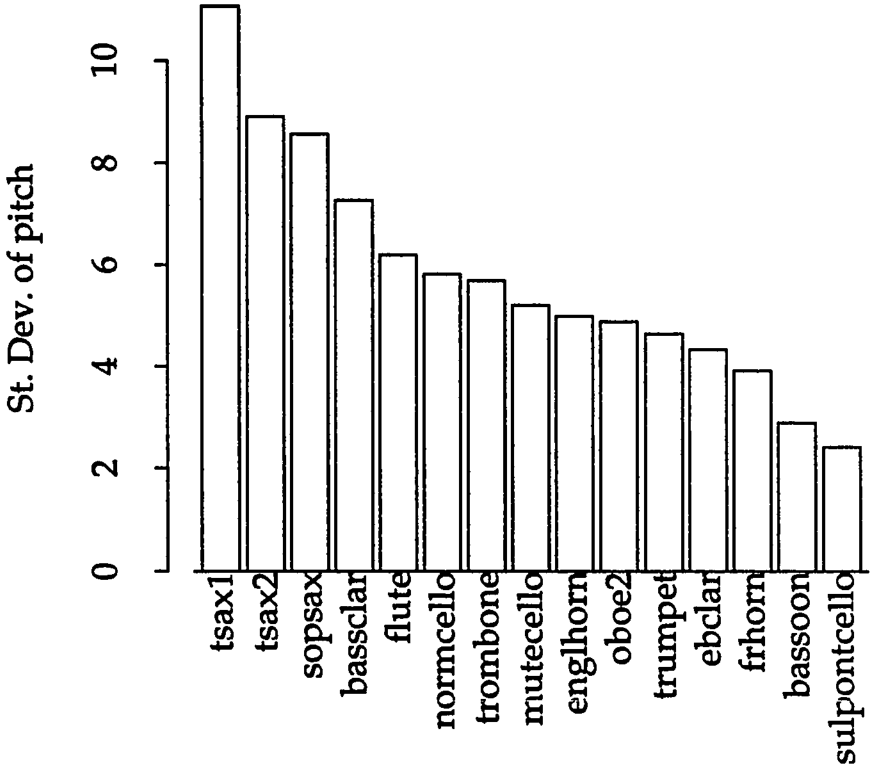


Figure 54. Measure of pitch deviation or "jitter" for the Stanford tones, calculated from the standard deviations of pitch-tracking analyses.

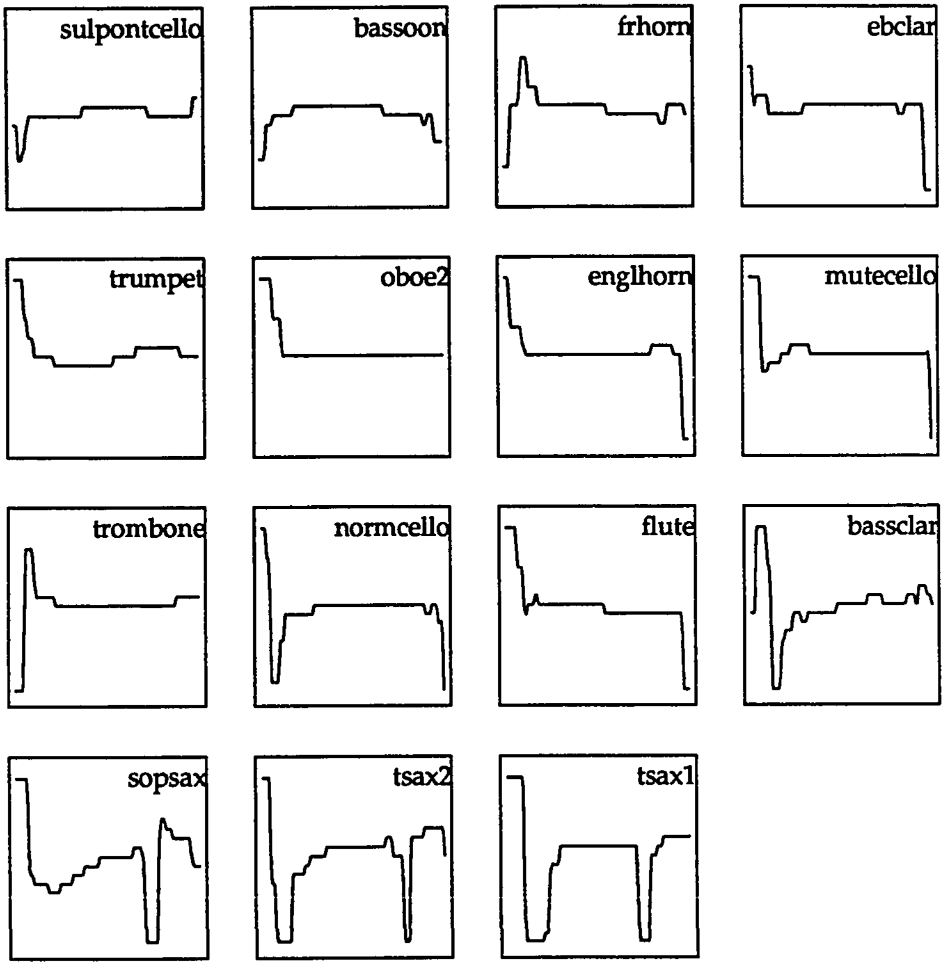


Figure 55. Pitch curves of the Stanford tones. Each plot displays a range from 291 to 331 Hz. Curves are ordered according to the standard deviation of the pitch (see Figure 54).

concurrently presented, and hence they will blend. A measure of *recognizability* for the Stanford tones can be inferred from the results of Grey's own instrument recognition task (Grey, 1975, Study C, Experiment 2, pp. 70-74). Listeners learned to associate a label with each instrument, and then identified instruments by their labels in an experiment. The percentages of judgments in which they correctly identified each instrument are given in Figure 56 (calculated from Table 6, p. 73 of Grey, 1975).

Correlated Factors

Eleven factors have been defined: centroid, acoustic dissonance, precedent noise, DPAT, amplitude scaling, harmonic envelope synchrony, peak synchrony, onset/offset synchrony, harmonicity, pitch deviation, and recognition. It should be noted that these are not all independent dimensions; a number of them are correlated. Some of the correlations are, expectedly, between similar measures: precedent noise and DPAT ($r = .861$, $p < .0001$), and two of the synchrony measures (peak synchrony and harmonic envelope synchrony, $r = .667$, $p = .007$).

Other correlated factors are harder to explain. Acoustic dissonance correlates both with DPAT ($r = .556$, $p = .031$) and precedent noise ($r = .573$, $p = .025$); this suggests that a large proportion of the acoustic dissonance in a tone can be attributed to inharmonic activity in attack portions. Yet precedent noise correlates with two other factors (pitch deviation: $r = .660$, $p = .009$; recognition: $r = -.648$, $p = .009$); perhaps the tendency for so many factors to match with this factor is due to its impoverished range of values (more than

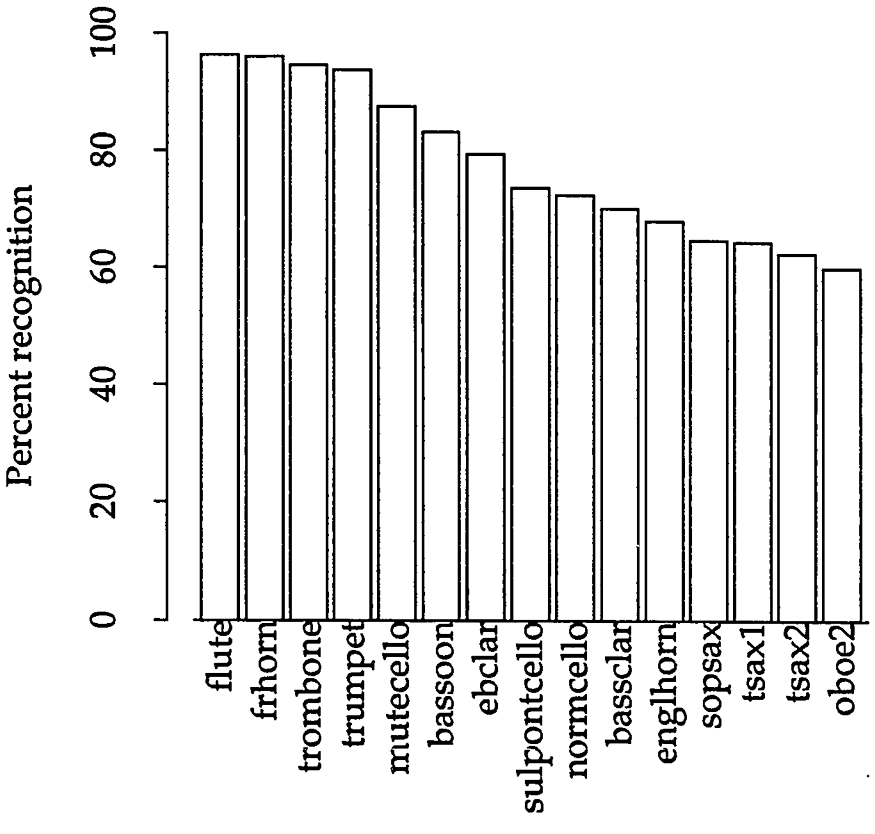


Figure 56. Percentage correct recognition for the Stanford tones from Grey's recognition study (Grey, 1975, Study C, Experiment 2, pp. 70-74).

half of the instruments have zero precedent noise, and the number of unique values is very small). Other correlations include pitch deviations and DPAT ($r = .844$, $p = .0001$); onset/offset synchrony and amplitude scaling ($r = .546$, $p = .035$).

Overall, the number of correlating factors is small (8 out of 55 possible correlations are significant). However, their existence raises a concern as to the interpretability of the succeeding analyses. It may be, for example, when looking at various aspects of the blend judgments, DPAT and precedent noise (correlating factors) are always found to be co-occurring: whenever DPAT correlates to the data, precedent noise does as well. In this case, it may be wise to consider the two factors to be part of a single phenomenon. However, it is not the case that all such correlating factors will co-occur: correlations are not transitive. For example, if A and B correlate, and A and C correlate, it is not necessarily true that B will correlate to C. Conversely, A and B might *not* be correlating factors, yet cases will be found where both correlate to some third phenomenon. It is not wise, then, to discard any of these measures at this time. Whether the two or more factors are independent vs. the same phenomena must be decided on the basis of accumulating evidence.

Grey's Results

Since the perceptual attributes of the Stanford tones have been studied previously by Grey (1975), it will be useful to review his findings here. His similarity study (Study C, Experiment 1, pp. 58-69) offered the most information about the tones. It is possible that some acoustic factors that were thought to play a role in Grey's listeners' similarity judgments may play a role in blend judgments as well.

Grey's method for interpreting his similarity data was mainly through various Multidimensional Scaling (MDS) techniques, primarily using the INDSCAL procedure. MDS is a powerful tool for analyzing any data involving psychological distance (Kruskal & Wish, 1978). In Grey's case, the rating scale consisted of values between 1 and 30 for every pairwise comparison of the instruments (flute-oboe2, flute-englhorn, and so on); listeners used low values to express perceived dissimilarity and high values to express perceived similarity. MDS was then used to calculate a representation of this network of similarities that could be mapped onto a three-dimensional spatial configuration, with closeness indicating larger degrees of similarity, and greater distances indicating lesser degrees of similarity; this yielded Grey's MDS solution which appears in Figure 57 (reproduced from Grey, 1975, Figure 11, p. 90).[35]

The relative proximity of the instruments in the three dimensions x (left-right), y (up-down) and z (front-back) conveys their similarity (he shadow-like projections of the data on the left wall and the floor help clarify their position). The exact locations of each item result in some pairs of items being simultaneously near in one dimension and far in another dimension.

---

[35] Table 1 should be consulted to make correspondences between Grey's labeling method and the one used in this study. Note that Grey's symbol "O2" stood for the oboe which was omitted from the present study.

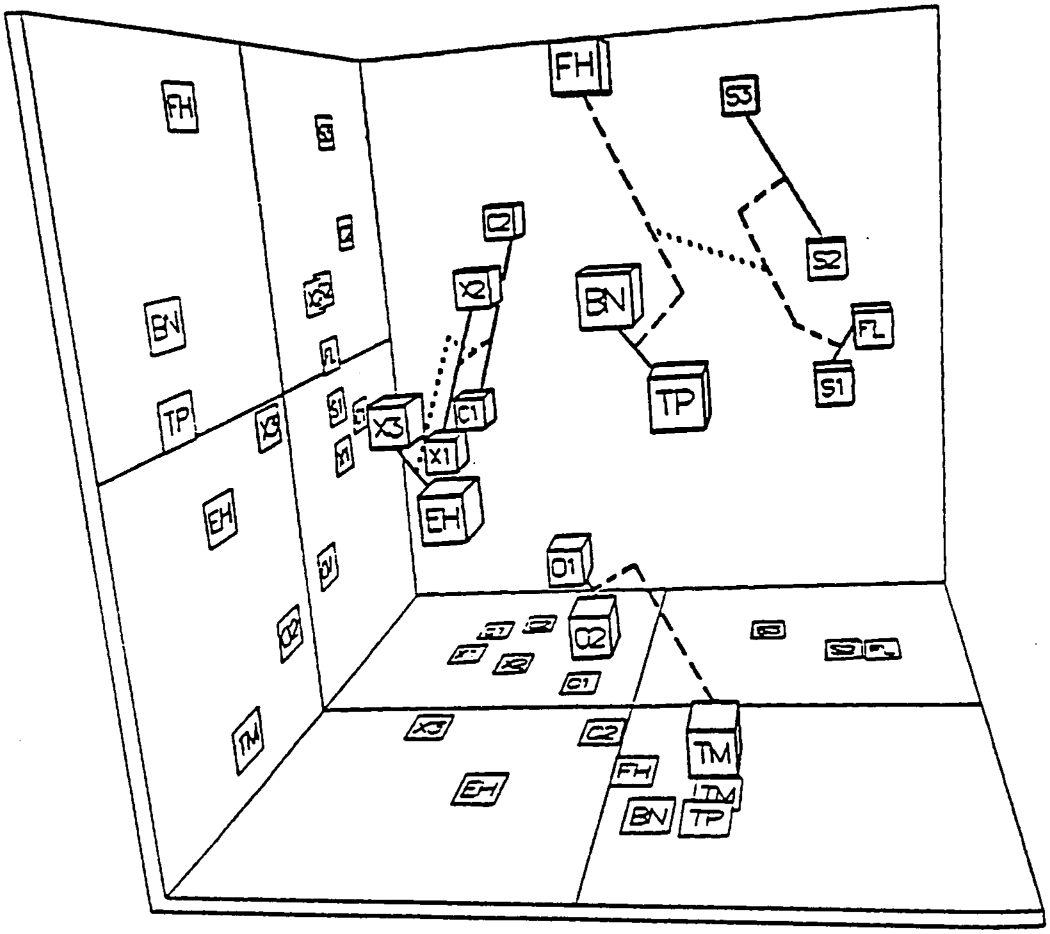


Figure 57. Multi-dimensional scaling solution from John Grey's similarity study for the Stanford tones (Study C, Experiment 1, pp. 58-69). FL = flute, O2 = oboe2, EH = englhorn, EC = ebclar, BC = bassclar, X3 = sopsax, X2 = tsax1, X1 = tsax2, BN = bassoon, FH = frhorn, TP = trumpet, TM = trombone, S1 = sulpontcello, S2 = normcello, and S3 = mutecello (O1 represents an additional oboe that was not used in the current study). Reproduced from Figure 11, p. 90 of Grey (1975).

For example, bassoon, trumpet and frhorn are equally similar in all dimensions except one: the frhorn distinguishes itself from the other two in the y-axis. This multifaceted portrayal of similarity provides the researcher with a clue to the perceptual bases on which judgments were made; that is, for an N-dimensional MDS solution, it is hypothesized that the judgments were driven by N interacting factors; the identification of those N factors are must be made by the researcher.

Grey evaluated various acoustical properties of the tones and identified which dimensions of the MDS solution seemed to correlate to these features. The x-dimension (left to right) was identified as capturing "synchronicity in the collective attacks and decays of upper harmonics" (p. 61), with more synchronous tones on the left and more asynchronous tones on the right. The y-dimension (low to high) was said to correspond to the spectral energy distribution, or centroid, with low centroids at top and high centroids at the bottom. The z-axis (front to back) was identified as corresponding to the presence or absence of "precedent high-frequency, low amplitude energy, most often *inharmonic energy,* during their attack segment," (p. 67) with greater amounts at the rear, less in the front.

Because the correspondences Grey discussed were apparently informal (no statistical procedure is cited), it is unclear how strong these correspondences are, or whether they are the only possible correspondences.[36]

---

[36] Grey corrected this problem with respect to centroid in a later presentation of his data (Grey & Gordon, 1978).

They can be examined, however, by correlating the coordinates for the plotting each of the instruments in the three dimensions with their corresponding acoustical measurements. Although the coordinates are not available in Grey's thesis, they can be extrapolated by measuring certain of Grey's graphs. Grey's Figures 12a and 12b, two-dimensional plots of the x-y and y-z dimensions (pp. 63-64), were measured by the present author with the aid of a computer scanning device, and from these points a likeness of the three-dimensional solution was created (Figure 58).[37] The acoustical measures discussed in the preceding section were compared to the coordinates used to produce Figure 58, and the results are enumerated below.

> Y-axis: Grey's analysis that the centroids of the tones were captured in this dimension was confirmed ($r = -.893$, $p < .0001$). In fact, none of the ten other factors correlated with this dimension.
>
> Z-axis: In contrast to Grey's analysis of this dimension, the measure of precedent attack noise discussed earlier (Figure 46) failed to correlate here. However, the z-axis does correlate to another measure of attack, DPAT ($r = .578$, $p < .05$), suggesting that DPAT may be a more

---

[37] The small discrepancies in the locations of instruments that may be observed between Figures 57 and 58 are because the datapoints for Figure 58 were extracted from Grey's two-dimensional graphs (Figures 12a and 12b in Grey, 1975, pp. 63-64) rather than Grey's three-dimensional plot (as reproduced in Figure 57). Possibly the perspective method of Grey's Figure 11 distorted the data slightly. Note also that the extra oboe ("O2" in Grey's study) is omitted from Figure 58.

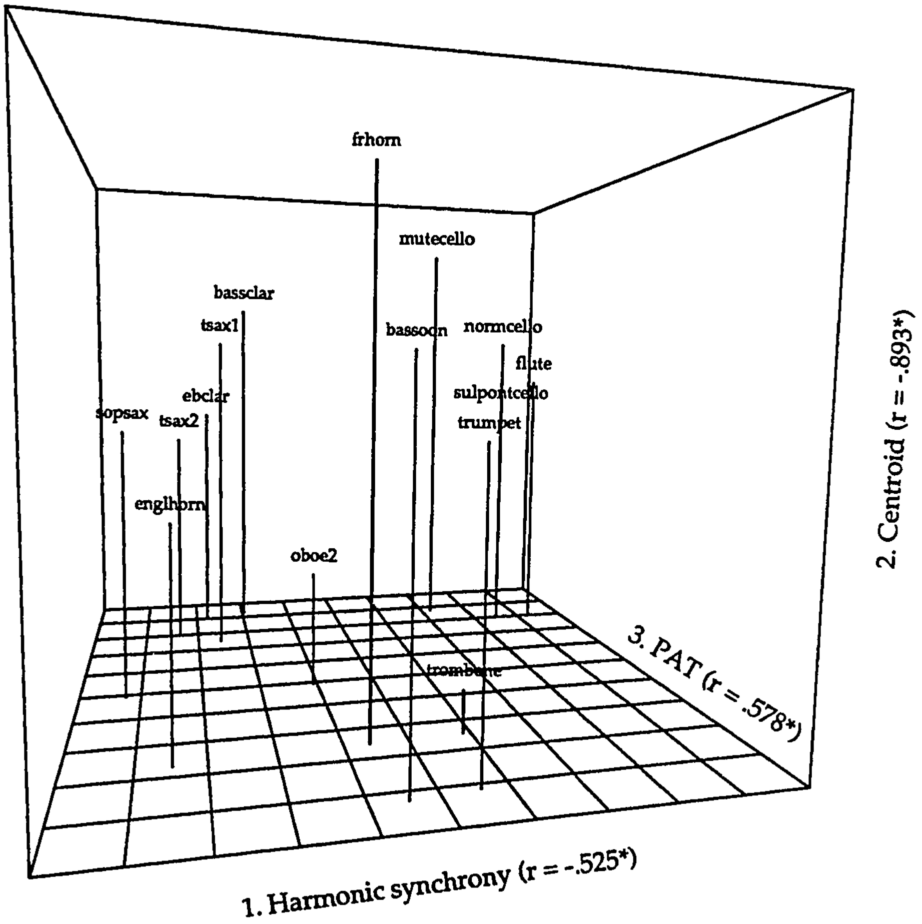


Figure 58. Reproduction of Grey's MDS solution (Figure 57) showing significantly correlating acoustical factors to each of the three dimensions.

salient attribute than presence of precedent noise. An additional measure, not explicitly related to attack features, correlated here as well: harmonic envelope synchrony. What is curious about this is that Grey identifies synchrony as an attribute of the X-axis, not this axis.

X-axis: Grey's identification of this axis as pertaining to synchronicity seems warranted: all three measures of synchrony proposed earlier show correlations (harmonic envelope synchrony: $r = -525$, $p < .05$; peak synchrony: $r = -.555$, $p < .05$; on/off synchrony: $r = -.5671$, $p < .05$). The measure of synchrony Grey used, pertaining to the onsets and offsets, is corroborated by the fact that the present study's on/off synchrony measure receives the highest correlation of all three synchrony methods with this dimension. Grey's measure, which was limited to the behavior of *upper* harmonics only, may have been unnecessarily restrictive since the X-axis appears to capture all aspects of synchrony for the tones. However, there are two other factors which correlated to this dimension: precedent attack noise and recognizability. It appears that this dimension is *less distinct* in its meaning than the other two. Indeed, Grey also noted that the dimension seemed to have alternative explanations; he observed that it captured a factor of instrumental family membership with "*woodwinds* . . . on the far left, the *brass* in the middle and the *strings* on the far right" (p. 67).

Despite the fact that the reader can "eyeball" the correlations between the MDS solution and the acoustical properties of the tones (using Grey's Figures 12-13, pp. 63-66), the failure to report any *quantitative* measure of correlating physical properties of tones to the MDS dimensions is a major shortcoming of Grey's study: no evaluation of how robustly each dimension matched the physical properties of the tones is provided.[38] While Grey's analysis indicates that dimensions Y and Z correlate to centroid and precedent noise, respectively, the present author's measurements show a strong correlation for the former but no correlation for the latter. Although the present author's methods for measuring acoustic properties may have differed from Grey's, this finding does indicate that the magnitude of the correlations was somewhat variable. This uncertainty is compounded by the absence of any evaluation of the success of the the MDS analysis, such as a Level of Stress measurement or a report of the percentage of the variance accounted for by each of the dimensions.[39] In the absence of any quantitative evaluation, the appropriateness of the number of chosen dimensions and their relative reliability must be considered an open question.

---

[38] Once again, this was corrected with respect to centroid in Grey and Gordon (1978).

[39] Grey (1975) indicates that steps were taken to obtain the most optimal solution (p. 127, second paragraph), but no details are given.

Chapter 4:

Experiments in Blend Perception

Experiment 1 (UnisonBlend)

Stimuli

The 15 Stanford tones were synthesized from the data as described in Chapter 3, using Grey's loudness and duration equalizations, and the author's PAT and pitch equalizations. Csound signal generation software (Vercoe, 1986) was used to generate NSS soundfiles (see Roth, Kendall, and Decker, 1985) with a 44.1 kHz sampling rate, using a single channel (mono). The control rate (responsible for determining the temporal resolution of the amplitude and frequency envelopes) was 4.41 kHz. Individual tones were merged into concurrently-sounding pairs (e.g. flute-oboe2, bassoon-englhorn), digitally mixing single soundfiles into stereo (one instrument in each channel), although the presentation to subjects was mixed down to mono through a mixing console. The amplitude differences between the tones were preserved as much as possible during all transformations of the soundfiles.

The musical interval between the tones was, of course, a unison, since all tones had the pitch Eb4. All possible concurrent pairs of the 15 tones were generated, including same-instrument pairs for a total of 120 trials. The lower part of Table 2 shows the systematic process that led to the generation of 120 trials.

Subjects

Eight listeners participated in the study, all of whom were Northwestern students currently involved in some kind of musical activity. Four were researchers in music (music theorist, or investigator of music-related auditory phenomena), three were performers enrolled in an instrumental music degree program, and one was a composer. Five of the listeners had some acquaintance with the Stanford tones and the synthesis method involved, and two of them were intimately acquainted with the methodology, development and purpose of the experiment. The initials of the eight subjects were GJS, MDS, PJD, WLM, JFK, MDW, MAN and BMD. Listeners were not paid.

Equipment

The experiment took place in a Northwestern Computer Music's sound studio, a room measuring 14' x 11' x 9' and designed specifically for listening to music (details of the room's design are given in Jones, Kendall, and Martens, 1984). One listener at a time ran the experiment in the closed room. The listener was seated eight feet from a pair of Urei model 809

flute-flute					
flute-oboe2	oboe2-oboe2				
flute-ebclar	oboe2-ebclar	ebclar-ebclar			
flute-bassclar	oboe2-bassclar	ebclar-bassclar	...		
.	.	.	...		
.	.	.	...		
.	.	.	...		
flute-normcello	oboe2-normcello	ebclar-normcello	...	normcello-normcello	
flute-mutecello	oboe2-mutecello	ebclar-mutecello	...	normcello-mutecello	mutecello-mutecello

15 + 14 + 13 + ⋯ + 2 + 1 = 120

Table 2. The process of generating all possible pairs of trials in Experiment Unisonblend.

loudspeakers, which were separated by about six feet, and faced the listener. Soundfiles were played through a custom 16-bit digital-to-analog converter and two QSC Model 1200 amplifiers. A computer display screen (a Sun II workstation) was positioned between the loudspeakers, facing the listener. The listener held the Sun II's mouse (the device for controlling the movement of a cursor on the screen) and its tracking surface in his or her lap. On the screen was a display indicating an eleven-point scale (0 through 11), as shown in Figure 59. The length of the line in the middle would extend either partially or completely to the right; the line's extent was controlled by horizontal movements of the mouse. In Figure 59, it is located close to the value of 3.5. When moved to the left, the line would shorten, and when moved to the right, it would lengthen. By this means the listener could indicate answers to questions on the eleven-point scale. The software that collected the responses recorded the precision of the rating indicator, and listeners were encouraged to use the "inbetween values" to express whatever values they thought were appropriate. The values actually saved by the program were floating point values between zero and one (that is, the values collected were divided by 11).

All listeners heard the experiment at the same gain setting on the amplifier. This was insured prior to each session by measuring the level of a 1000 Hz sine tone (played through the amplifier over the speakers) with a calibrated Brüel & Kjær Precision Soundlevel Meter (type 2203, with condenser microphone type 4145) and adjusting the gain to reach a reference level of 80 dB SPL.

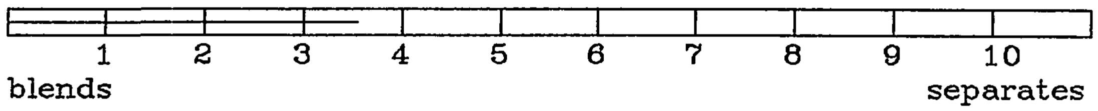


Figure 59. The rating scale that appeared on the computer screen, and which listeners used to make blend judgments.

Procedure

The experimenter gave the listener a demonstration of the use of the mouse and the method for responding. The listener was then given the following written instructions: "I. You will hear a short sound consisting of two instruments playing the same note together. II. When the computer prompts you, rate the blend of the pair of notes. If the two sounds seem to blend into one, unified sound, then you should give it a low number. If the two sounds seem to separate into two kinds of sounds, then you should give it a high number. Try to use the entire range of the scale to capture the varying degrees of blend and separation that you hear over the course of the experiment. III. Slide the mouse on the pad until the line moves to the rating you wish to enter. When you have made your choice, press the right button. If you wish to hear the notes again, press the left button."

Following this, the experimenter played 12 of the 15 Stanford tones individually over the loudspeakers while the listener followed a list of the names of the instruments in the order that they were sounding. The intent here was for the listener to grasp that the stimuli that they would hear in the experiment were indeed the composite of two realistic-sounding musical instruments. The three tones which were omitted were tones that sounded very much like others in the set: tsax2 was much like tsax1, while sulpontcello and mutecello were much like normcello. Omitting these instruments avoided the risk of misleading the listener that the task would involve making distinctions between variants of the same kinds of instruments. In addition the listener was urged not to let his tendency to recognize the instruments influence his judgements, and to make his answers without lengthy introspection or numerous replays.

Each listener ran a block of 120 trials a total of four times in a different random order for each block. Some listeners divided this into two sessions on different days. At the beginning of each new session, 60 arbitrarily selected practice trials preceded the actual experiment. The total duration of the experiment, including instruction, was typically 1.5 hours.

A Note on the Rating Scale

The system of using low numbers to indicate "more blended" and high numbers to indicate "less blended" was chosen to suggest a scale of "oneness" vs. "twoness," respectively, and to discourage listeners from using the blend rating scale as a kind of preference judgment.[40] This choice has a disadvantage in the presentation of data, however. Later, the reader will encounter many references to "increasing blend" and "decreasing blend," meaning larger and smaller amounts of blend, respectively. However, the numeric data these descriptions refer to is the *inversion* of this: decreasing blend values means increasing blend, and vice versa. This creates the possibility of confusion in the following discussions. One way to avoid the

---

[40] In fact, most subjects reported that they found it challenging at first to resist using the rating scale as a preference scale despite being fully aware of the operative definition of blend.

confusion would be to invert the blend data such that increasing rating values correspond to increasing blend, which would make graphic presentations somewhat more intuitive.[41] However, this has resulted in some confusion on occasion, so it was decided to present the data according to the manner in which it was collected.

Analyzing Blend: Preliminary Remarks on Methods

This experiment, and the two that follow it, pose special problems for data analysis. Although various methods will be introduced along the way, there are a few methods that will be used throughout the analyses. It is best to take a large detour to introduce those methods now, rather than burden the exposition with frequent small detours.

In some ways, the aim of the present study is similar to that of Grey (1975), suggesting that much of his methodology can be imported here. The most obvious way the two studies are similar is the employment of an experimental task to obtain judgments of instruments presented in pairs. Furthermore, both studies are concerned with the discovery of distinctive features for those instruments. However, this study contrasts with Grey's in that the *concurrent* properties of instrument pairs, as opposed to their similarities, is the object of study. Grey's similarity task was used as a method to reveal the ways in which tones perceptually distinguish themselves from

---

[41] This strategy was in fact taken in earlier presentations of Experiment UnisonBlend by the author (Sandell, 1989a, 1989b). The reader comparing the present study to the earlier ones must be careful to note that in the latter, the rating scale has been reversed.

one another, which in turn defined the properties that made them distinctive as *individual* tones. Similarly, a few of the analyses to be presented here will focus on the role that properties of single instruments appear to play in determining the blend for a given concurrent pair. The primary purpose, however, is to discover the distinctive features that arise from the pairing itself. This entails correlating the perceptual data with the *differences* or *interactions* between the physical properties of the instruments.

Overall correlations and Single Instrument Correlations

Two procedures that will be used throughout the discussion of all the experiments are *overall correlations* and *single instrument correlations*. The two are primarily distinguished in terms of the number of trials involved in the analysis. The former involves every trial in the experiment (for this experiment, n = 120), while the latter involves a subset of trials from the experiment, specifically all the trials that include a particular instrument (for this experiment, n = 15).

Table 3 illustrates both of these procedures with respect to the property of centroid, arbitrarily selected for purpose of demonstration. Columns 1-3 present an abbreviated list of all the trials in the present experiment, ordered according the process that was indicated in Table 2. Columns 4 and 5 show the centroids for each of the instruments involved in each trial. The *overall correlations* procedure correlates all 120 blend judgments for these trials to some aspect of the acoustical interaction between the two tones; in this case, the interaction concerns the centroid properties of the individual tones. Two

Trial #	Names		Centroids		Sum	Absolute Difference
	Inst. 1	Inst. 2	Inst. 1	Inst. 2		
1	flute	flute	861.107	861.107	1722.214	0
2	flute	oboe2	861.107	1214.9	2076.01	353.796
3	flute	englhorn	861.107	1015.45	1876.55	154.339
4	flute	ebclar	861.107	1386.23	2247.34	525.12
5	flute	bassclar	861.107	813.009	1674.12	48.098
6	flute	sopsax	861.107	960.601	1821.71	99.493
7	flute	tsax1	861.107	917.903	1779.01	56.795
8	flute	tsax2	861.107	1388.42	2249.53	527.316
9	flute	bassoon	861.107	770.167	1631.27	90.94
10	flute	frhorn	861.107	531.059	1392.17	330.048
11	flute	trumpet	861.107	1041.62	1902.72	180.508
12	flute	trombone	861.107	1826.33	2687.44	965.226
13	flute	sulpontcello	861.107	1050.34	1911.44	189.228
14	flute	normcello	861.107	785.344	1646.45	75.763
15	flute	mutecello	861.107	440.112	1301.22	420.995
16	oboe2	oboe2	1214.9	1214.9	2429.81	0
17	oboe2	englhorn	1214.9	1015.45	2230.35	199.457
18	oboe2	ebclar	1214.9	1386.23	2601.13	171.324
19	oboe2	bassclar	1214.9	813.009	2027.91	401.894
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
117	sulpontcello	mutecello	1050.34	440.112	1490.45	610.223
118	normcello	normcello	785.344	785.344	1570.69	0
119	normcello	mutecello	785.344	440.112	1225.46	345.232
120	mutecello	mutecello	440.112	440.112	880.225	0

Table 3. Illustration of centroid sums and absolute differences for trials in Experiment UnisonBlend.

interactions that will be frequently considered are given in columns 6 and 7. Column 6 is the sum of the two centroids, and column 7 is their absolute difference. The relevance behind the act of summing and subtracting physical properties is depends on the property being analyzed; for centroid they may be interpreted to indicate the overall centroid height[42] and centroid separation (as a simple measure of spectral difference), respectively. Depending on the sign and magnitude of correlations, different conclusions may be drawn, as suggested below:

1. Positive correlation with sums: blend decreases with increased overall centroid height.
2. Negative correlation with sums: blend increases with increased overall centroid height.
3. Positive correlation with absolute differences: blend decreases with increased separation between centroids.
4. Negative correlation with absolute differences: blend increases with increased separation between centroids.

Overall correlations, then, can be used to indicate the degree to which any of these four trends were operative over all trials (n = 120 in the present experiment). Frequently there are cases, however, in which trends are

---

[42] Summing the two values is like obtaining an average, but with the step of division by two removed. Since the correlation procedure ignores magnitude, the division does not add any relevant information.

operative over a subset of the data. Looking at only the trials involving a particular instrument is an instance of a subset which may have particular musical relevance. For example, trials 1-15 in Table 3 show all the flute trials in the experiment; all the oboe2 trials, on the other hand, are found in trials 2 and trials 16-29; and so on. Correlating the physical and perceptual attributes within such subset is called *single instrument correlations.* For example, the 15 ratings for the flute trials can be correlated with the centroids for the *non*-flute instruments (column 5). Here the emphasis is not on examining the interaction between tones, but the variation in blend caused by the changing instrument. The interpretation of the significance of obtaining positive vs. negative correlations, or looking at sum vs. absolute difference interactions, is the same as in *overall correlations.* However, the "control and variable" paradigm focuses on the phenomenon more closely than the more scattershot approach of *overall correlations.* That is, the unchanging instrument (the flute in the case of the flute trials) is the "control" and the changing instrument is the "variable"; what is being observed is how the change in the variable instrument *increases* or *decreases the blend* with the control instrument.

Average Instrument Correlations

A third procedure which will be frequently used is *average instrument correlations.* This investigates the possibility that blend was determined by more global effects of particular acoustic factors. The global effect is analyzed on an instrument-by-instrument basis, by executing the following steps:

1. For each instrument, the blend rating for all trials in which it was included are averaged into a single value. In the case of the flute, this would be the average of all blend ratings for trials 1-15 (see Table 3). This produces one value for every instrument (n = 15 in the present experiment).

2. The physical factor being investigated is collected for each instrument. For example, if the factor is centroid, this list consists of the centroids for flute, oboe2, englhorn . . . mutecello (n = 15 in the present experiment); that is, column 5 of Table 3.

3. (1) and (2) above are correlated.

The purpose of this procedure is to determine if a particular factor was so potent as to impose an *overall* degree of blend across multiple instruments. Consider the following example, involving centroid; for demonstration purposes, it is simplified to include only two instruments rather than all 15. The centroid for trombone is very high (1826 Hz), and the centroid for frhorn is very low (531 Hz). If it was found that the majority of trials involving the trombone blended poorly (as judged by the average of their ratings), while the majority of trials involving the frhorn blended well, one might propose that the centroid levels were imposing an overall influence on blend, with dark instruments tending to produce increased blend, and bright instruments tending to produce decreased blend.

Results of Experiment UnisonBlend

Variability of Responses

Since an important objective of the experiment was to determine if a diverse group of listeners could show a consensus on the perception of blend, it was necessary to establish that the eight listeners were consistent in their individual performances and showed agreement among themselves. With regard to the first consideration, within-subject correlations were very high: for each listener, each of the four blocks were compared with every other (six comparisons); without exception every correlation was significantly positive. 44 of these correlations had a probability of $p < .0001$ and none of the remaining four were worse than $p < .005$. *T* tests were also used to compare the means of each listener's group of four blocks. 40 of the 48 total comparisons (six for each listener) showed that means did not significantly differ (where $p < .05$). Six of the differences were because listeners MDS and MDW each had one block of trials which had a significantly different mean than all their other blocks. In view of the overall results, however, it is safe to conclude that listeners' responses were consistent. Each subject's four blocks of data were averaged together into a single dataset representing that listener's ratings.

Between-subject consistency was analyzed by correlating each listeners' averaged dataset with one another (28 comparisons altogether). The correlations were high, as before: each had a probability of $p < .0001$. All correlations were positive, with the exception that for one listener (MAN), for whom all correlations were negative. This meant that MAN showed the same patterns of blend ratings as the other subjects, but in "mirror image"; this interesting reversal will be discussed at a later point. Since the other seven subjects' data agreed so strongly, it was decided to discard MAN's data and average the others together. This is the dataset which will be used in all further statistical analyses of Experiment 1.

The distribution of ratings, in units of 0.1, is shown in Figure 60.[43] The averaged data to covers a fairly wide range of the rating scale, so there appears to be no impoverishment due to regression to the mean. It can also be seen that the data is slightly skewed towards the blended end: indeed, the mean rating is 0.444, somewhat below the midpoint of the scale. This reflects the experience of many listeners (conveyed in their reports after the experiment) that more of the trials seemed to blend than separate.

Preliminary Results

Multidimensional Scaling

As a preliminary inquiry, the experimental data was analyzed by a Multidimensional Scaling program (SINDSCAL, Pruzansky, 1987). The blend ratings were entered so as to be interpreted by the program as distances; thus,

---

[43] Note that in this graph the data is present in the 0 to 1.0 scale recorded by the computer rather than the 0 to 11 scale that the listeners were presented; all data analyses in this study will refer to the data in this form, and all further references to "the rating scale" refer to this 0 to 1.0 scale as well.

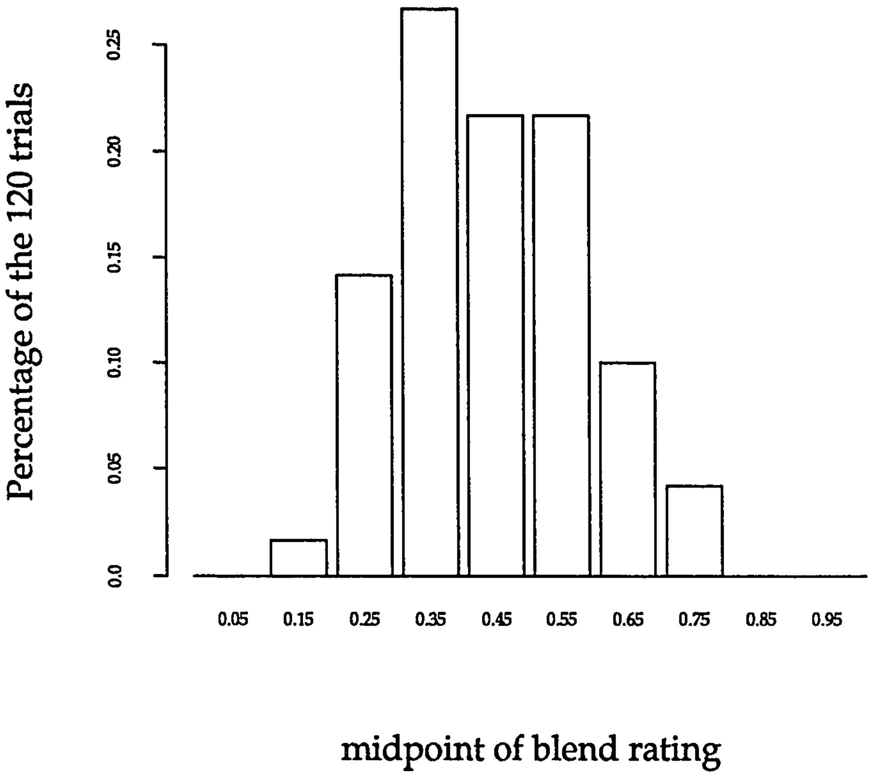


Figure 60. Distribution of blend judgments in Experiment UnisonBlend. Subject data was parsed into ten categories from the 0 to 1.0 scale (0 to .1, .1 to .2 . . . .9 to 1.0) and the proportion of total ratings in each category calculated.

the output represented poorly-blending instruments as geometrically distant, and blending instruments as geometrically close in an n-dimensional space. A three-dimensional analysis was attempted: Figure 61 shows dimensions one and two of the solution with correlating acoustical dimensions indicated. Dimension 1 correlates both with centroid ($r = .549, p = .034$) and acoustic dissonance ($r = .538, p = .039$), suggesting that some aspect of spectral quality underlay this dimension in general. Dimension 2 correlates both with perceptual attack time ($r = -.642, p < .01$) and the pitch variability of the tones ($r = -.698, p < .01$); the interpretation of this combination of factors is considered below. Altogether the three dimensions of the MDS analysis accounted for 51% of the variance (29.1%, 14.7% and 7.3% for dimensions 1 through 3, respectively).[44]

The MDS solution suggests that the two alto saxophones blended poorly with other instruments in general, because of their extreme locations at the edges of the solution.[45] In fact, for MDS to suitably distance the two, most of the remaining instruments have been clustered into a relatively small portion of the space. Since this results in an information loss, such solutions are often referred to as *degenerate* (see Shepard, 1974). The correlations with acoustical data on dimension 2 suggest two causes of the

---

[44] Only two dimensions are displayed here because a three-dimensional plot was unavoidably cluttered and difficult to read. Furthermore, the third dimension showed no correlation to any of the eleven factors.

[45] Referring to blend as "good" or "poor" is merely a semantic convenience to refer to the magnitude of the blend judgments (low and high values on the rating scale, respectively). No judgment as to the aesthetic value of the degree of blend is intended.

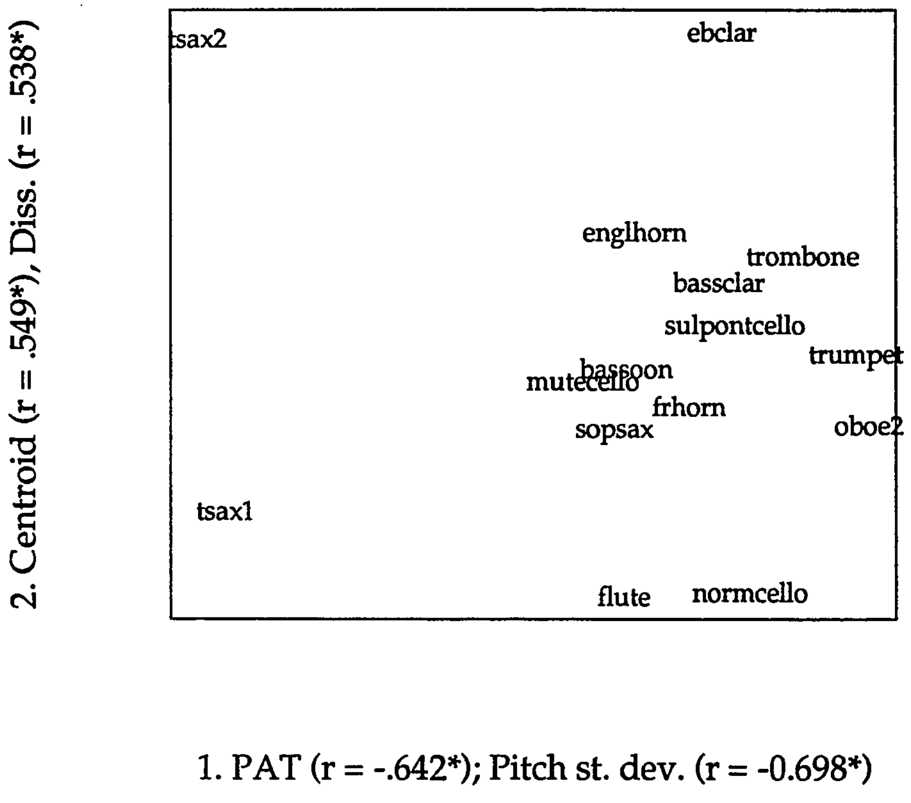


Figure 61. Multidimensional scaling solution of the data from Experiment UnisonBlend, with blends treated as distances (i.e., blended = close, separated = far). Correlating acoustical factors are shown. Dimensions 1 and 2 accounted for 29.1% and 14.7% of the variance of the data, respectively.

extreme differences, namely the difference in attack time and the amount of variability in their pitch curve (see Figures 47 and 54). Recall also from Figure 46 that tsax1 and tsax2 have lengthy precedent attack noise. From another perspective, one may also hypothesize that the two saxophones distinguished themselves in a stylistic sense from the remainder of the largely "classical" symphonic orchestra instruments; recall from Chapter 1 that orchestration manuals present varying opinions on the blending capacity of the saxophone depending on the perspective of the author (i.e., "classical" vs. jazz). The soprano saxophone, however, did not group itself with the other saxophones; possibly this is due to its similarity to the E-flat clarinet (see Figure 58).

Intrinsic "Blending Power" of Instruments

The position of the alto saxophones in the MDS solution suggested that they tended to be receive poor blend judgments, regardless of what they were paired with. Figure 62 shows each instrument's average blend rating; the value for tsax1, for example, is the average rating for all the trials which included tsax1 (n = 15). Not only do tsax1 and tsax2 have the highest means, as predicted, but a range of means (0.4 to 0.6) is observed across the whole collection. It appears that each instrument tended to impose a certain degree of blend regardless of what it was paired with.

To examine the possibility that the averages were so close that they were statistically identical, *t* tests were applied to all possible pairs (n = 105) of averages shown in Figure 62. In general averages that differed by at least 0.1

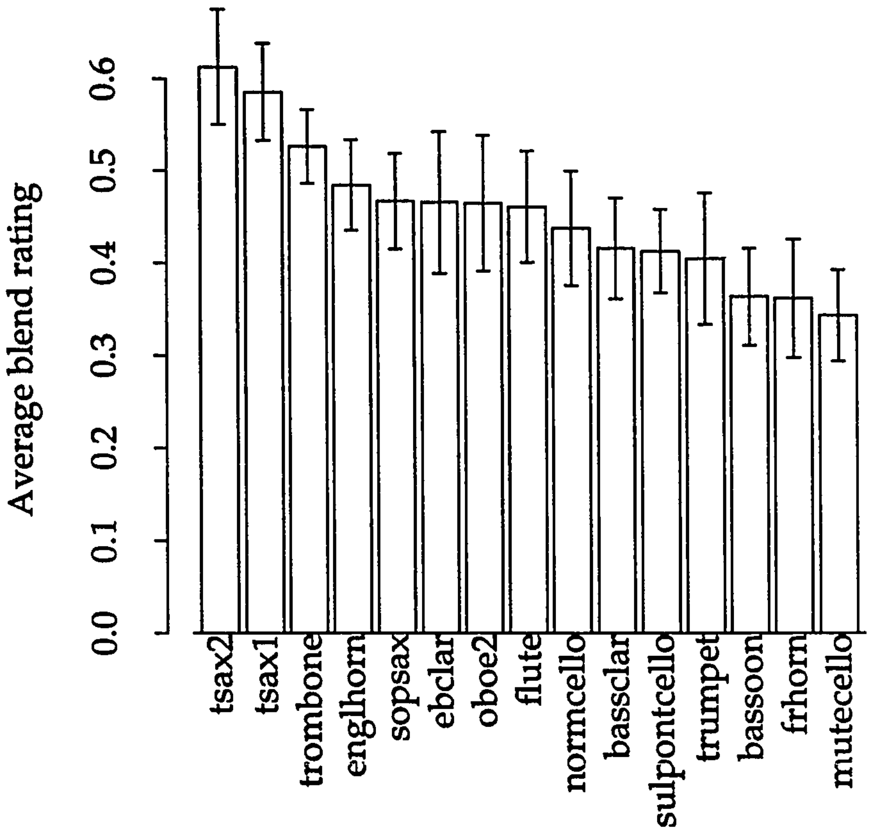


Figure 62. Blend judgments from Experiment UnisonBlend averaged by instrument. Each value shows the average for the 15 trials in which the named instrument was a member. Error bars indicate the standard deviations for the 15 values.

on the rating scale were found to be significantly different ($p < .05$) from one another; of the 45% that were found to be different from one another--47 out of the 105 comparisons--most of those (88%) had mean differences of at least 0.1. Given the requirement of $p < .05$, one would expect fewer than six of the comparisons to have been significant by chance alone, so those that differ by 0.1 or more can safely be considered different values. These findings show that the means for instruments tsax1 and tsax2 were different from all other instruments, while the group mutecello, frhorn and bassoon was different from all others with the exception of trumpet, sulpontcello and bassclar. It seems appropriate then, to say that tsax1 and tsax2 were "poor blenders" and mutecello, frhorn and bassoon were "good blenders." The group englhorn, sopsax, ebclar, oboe2, flute, normcello, bassclar, sulpontcello and trumpet, on the other hand, all had nearly-identical means. Considering this and their somewhat higher standard deviations (see the error bars in Figure 62), it appears that whether these instruments blend or not depends more on the instruments that they are paired with than with any intrinsic capacity for blend.

The possibility that "blending power" was due to specific acoustic factors was explored by correlating the averaged ratings of Figure 62 to each of the eleven factors. *Average instrument correlations* showed significant correlations, all positive with four of the eleven factors: centroid ($r = .653$, $p = .008$), DPAT ($r = .549$, $p = .034$), acoustic dissonance ($r = .585$, $p = .022$) and pitch deviation ($r = .716$, $p = .003$). These correlations suggest that greater "blending power" is obtained whenever pairs contained instruments with darker timbres (lower centroids) and shorter attacks; similarly greater blending power was obtained with tones were acoustically consonant and steady in pitch. However, these early indicators based on averaged groups of data must be corroborated by further analyses.

A few visualizations of these correlations are shown in Figures 63 and 64. Figure 63 illustrates the centroid correlation showing "seesaw" plots (first introduced in Chapter 3, Figure 42) to identify the centroid (x-axis) of the named instrument and the averaged blend judgment (y-axis) for that instrument. The seesaw plots emphasize that as centroid becomes higher (as indicated by the location of the fulcrum), blend gets worse. The significant correlation ($r = .657$, $p = .008$) proves that this interaction is statistically meaningful. The reader will recall that Goodwin's experiment with choral singers (Goodwin, 1989) showed that when blend was obtained, singers were using a "darker" tone; the correlation shown here also indicates that the presence of dark instruments improves blend.

Figure 64 illustrates the DPAT correlation in a similar way. Here, miniature amplitude envelopes are used to plot each datapoint. They are plotted on the x-axis at the value of the DPAT of the named instrument, and on the y-axis at the value of the average blend judgment for that instrument. It can be seen that many of the envelopes with long preliminary portions (precedent noise) are on the non-blended end. Again, the significant correlation ($r = .550$, $p = .034$) means that blend decreased as the DPAT was longer.

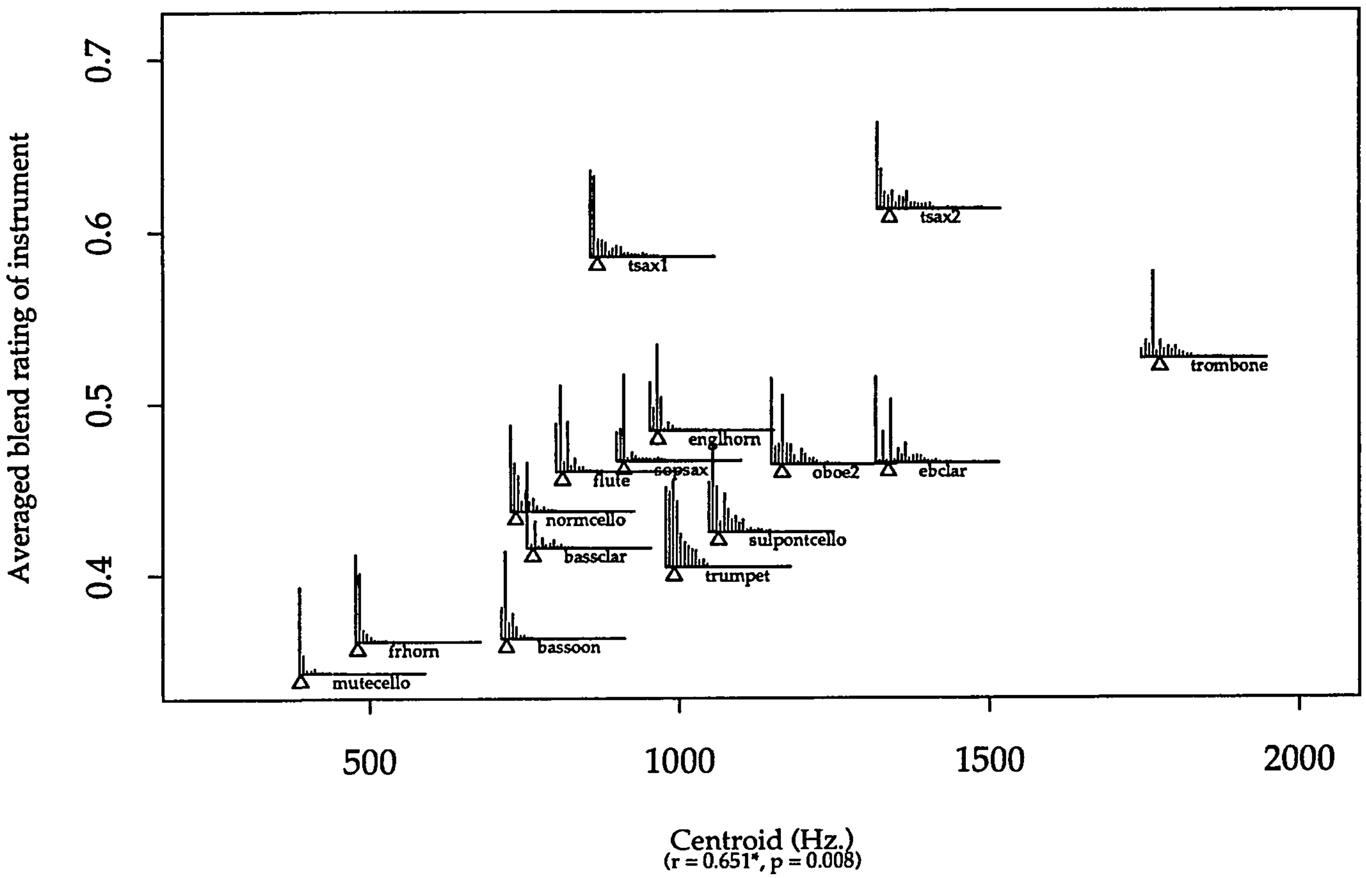


Figure 63. The average blend judgments from Figure 62 plotted against the centroids for the named instruments. Each datapoint is represented by an average spectrum envelope, anchored at the location of its own centroid.

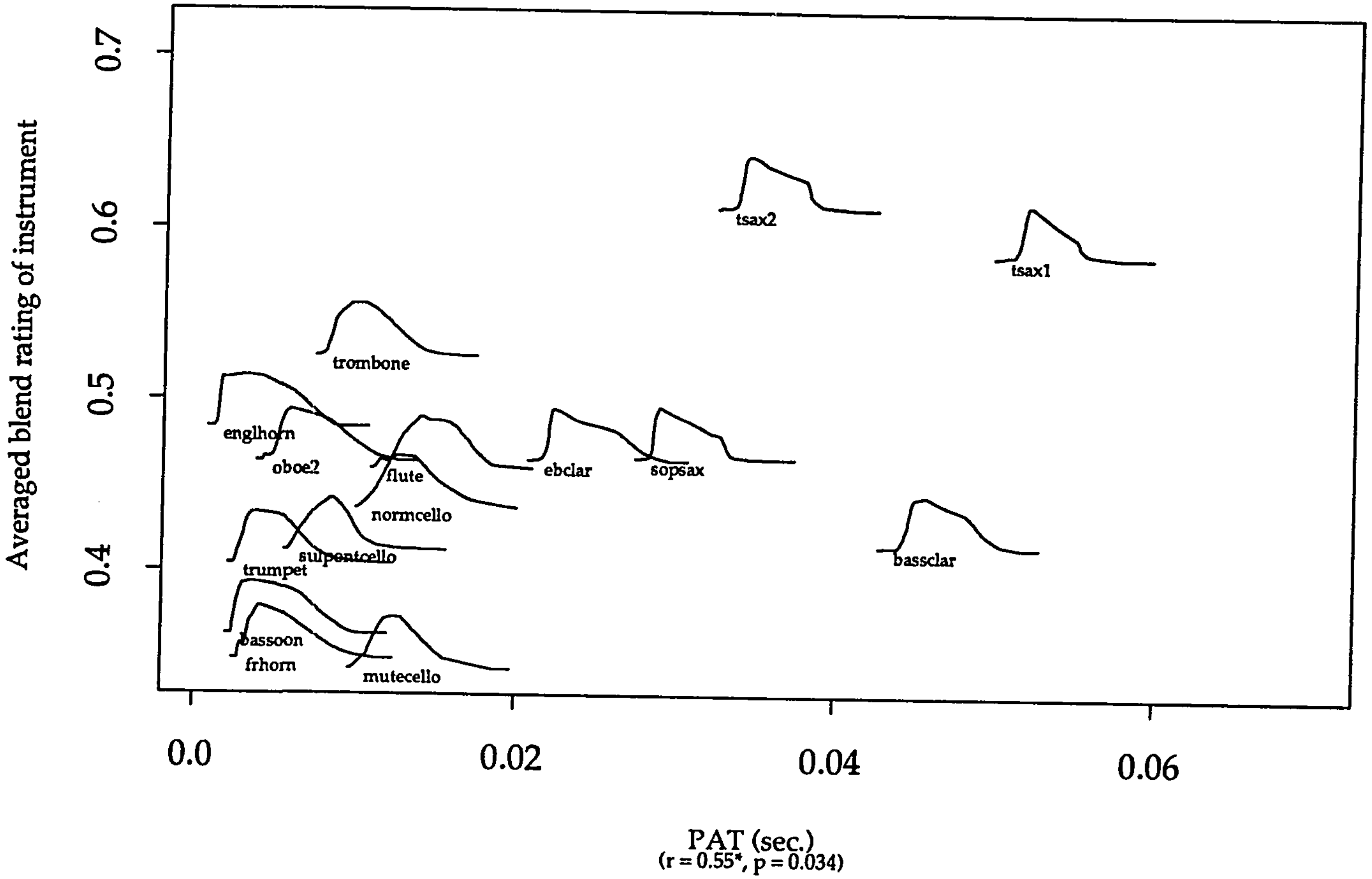


Figure 64. The average blend judgments from Figure 62 plotted against the DPATs for the named instruments. Each datapoint is represented by an amplitude envelope, anchored at the beginning of its line.

The MDS and the *average instrument correlations* give only a preliminary view of the interactions between the perception of blend and the acoustical properties of the instruments. To follow up on these findings, and to uncover information that may have been missed by these procedures, more detailed analyses of the various acoustical properties are discussed in individual sections below.

Spectrum and Blend

The MDS analysis and the *average instrument correlations* just mentioned present evidence that centroid was an important factor in the blend judgments. These analyses focused on the individual acoustical properties of instruments that were present in the pairs. Further analyses concerning centroid and other measures of spectrum, emphasizing interactions between the tones, will be considered here.

Earlier, procedures called *overall correlations* and *single instrument correlations* were proposed for analyzing interactions between tones. Recall that correlations with either *sums* or *absolute difference* can be investigated; each has different implications for centroid. These offer two contrasting ways of looking at the outcome; for the purpose of discussion, procedures using sum and difference correlations will be called *model 1* and *model 2*, respectively. *Model 1* suggests that the overall centroid height determine blend, whereas *model 2* suggests that the difference in centroid determines blend. Evidence for both is considered below.

To investigate *model 1, overall correlations* (n = 120) between the blend data and the sums of centroids were performed. This yielded a significant correlation ($r = .481$, $p < .0001$); hence, it does appear that the lower the overall centroid of the pair, the better the blend. To further investigate this, *single instrument correlations* were applied to each instrument. Several instruments (six) showed correlations with centroid using this procedure; the names of the instruments and their correlations are listed in column 2 of Table 4. It shows, for example, that blend judgments for the flute increase as it is paired with successively darker instruments.

Centroid Distance

To investigate *model 2, overall correlations* (n = 120) between the blend data and the absolute differences of centroids were performed. This yielded a significant result, as above, but with lower magnitude in this case ($r = .229$, $p = .001$). Then, *single instrument correlations* were applied to this question as well. In this case, fewer instruments (three) showed correlations (see Table 4, column 3). Thus some evidence in support of *model 2* was present, but not as much as in the case of *model 1*; it appears that for centroid, overall height is a better predictor for blend than centroid proximity. Furthermore, the particular instruments that showed these correlations--frhorn, bassoon,

Instrument	Centroid		DPAT		Precedent Noise		Temporal Envelopes		
	Sum	Absolute Diff.	Sum	Absolute Diff.	Sum	Absolute Diff.	Centroid	Amplitude	Pitch
flute	r = .755 *p* = .001				r = .524 *p* = .045	r = .524 *p* = .045			
oboe2			r = .669 *p* = .006	r = .682 *p* = .005	r = .526 *p* = .044	r = -.547 *p* = .035	r = -.650 *p* = .009	r = -.608 *p* = .016	
englhorn			r = .610 *p* = .016	r = .610 *p* = .016	r = .539 *p* = .038	r = .539 *p* = .038	r = -.526 *p* = .044	r = -.581 *p* = .023	
ebclar								r = -.536 p = .039	
bassclar									
sopsax	r = .625 *p* = .013								r = .527 *p* = .044
tsax1									
tsax2			r = .584 *p* = .022	r = .584 *p* = .022				r = -.585 *p* = .022	
bassoon	r = .695 *p* = .004	r = .531 *p* = .042	r = .558 *p* = .031	r = .565 *p* = .028	r = .518 *p* = .048	r = .518 *p* = .048			
frhorn	r = .611 *p* = .016	r = .602 *p* = .018	r = .689 *p* = .004	r = .701 *p* = .003	r = .637 *p* = .011	r = .637 *p* = .011	r = -.543 *p* = .037		r = -.526 *p* = .044
trumpet			r = .589 *p* = .021	r = .602 *p* = .017					
trombone			r = .658 *p* = .008	r = .661 *p* = .007			r = -.526 *p* = .044	r = -.564 *p* = .029	
sulpontcello			r = .546 *p* = .035	r = .546 *p* = .035					
normcello	r = .573 *p* = .025		r = .545 *p* =.036		r = .543 *p* = .036	r = .543 *p* = .037			
mutecello	r = .657 *p* = .008	r = .657 *p* = .008							

Table 4. Various *single instrument correlations* to data in Experiment UnisonBlend.

and mutecello--suggest that, because of their own very low centroids, this analysis merely reproduced the results of the sum analysis.[46]

Spectral Difference

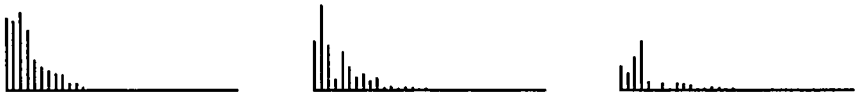
One of the purposes of using absolute difference measurements of centroid in the previous section is to compare the spectra of two instruments in a simple way. Centroid itself, in addition, is a simplification of a complex spectrum, meaning that the method of examining spectrum is "twice-removed." A more direct comparison of the spectra is desired; a method of characterizing the overall *difference* between spectra is proposed here. Since the two tones in this experiment had the same fundamentals, it is possible to directly compare each pair of harmonics. By using the average spectrum for each instrument (as was seen in Figure 42) absolute difference in power for each harmonic between the spectra can be taken to yield a "difference spectrum."[47] For demonstration purposes, the calculation of a difference spectrum is illustrated in Figure 65 for very similar and very different spectra, respectively.[48]

---

[46] That is, when the centroid of the fixed instrument is at an extreme end, all the differences have the same sign (positive or negative); in such a context, differences yield the same result as sums.

[47] I would like to thank David Wessel for suggesting the idea of difference spectra.

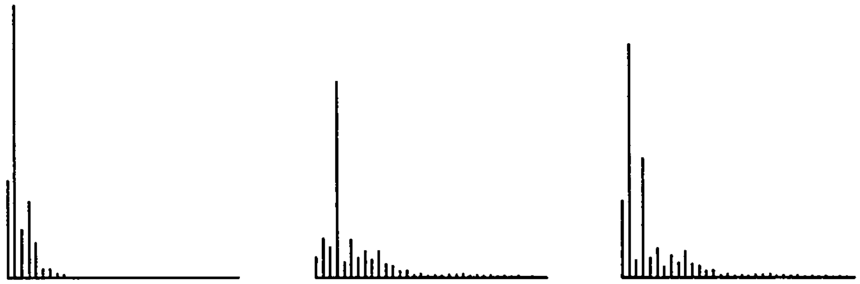
[48] The maximum number of harmonics in any instrument was 33 (trombone), so all instruments being compared are treated as though they have 33 harmonics (with a value of zero amplitude where that instrument had no energy at a particular harmonic).



trumpet

sulpontcello

difference (sum = 484)



bassoon

trombone

difference (sum = 1952)

Figure 65. An illustration of spectral difference calculation. Each row shows the average spectrum of two instruments followed by third a spectrum showing the absolute difference at each harmonic. A quantitative overall measure of the difference is obtained by the sum of the amplitudes in the third spectrum (shown next to the difference spectra).

From here, the overall amount of difference between the spectra can be estimated by summing the amplitudes of all the partials of the difference spectrum together (see the figures given in Figure 65). The total amount of difference in power among the individual partials, summed together, characterizes the amount of difference between the two spectra. Calculating all such sums for all pairs of tones (n = 120), and comparing it to the blend judgments with *overall correlations* yields a significant result ($r = .302$, $p = .0008$). Pursuing this further with *single instrument correlations,* however, yields only one instrument with a significant result (tsax2). In both cases, the sign of the correlation shows that blend decreases as the difference gets greater.

Although the evidence for the effect of spectral difference does not appear to be very extensive, a visualization of this interesting relationship is offered in Figure 66, showing the *single instrument correlation* for tsax1. Each datapoint is represented with a small spectral envelope representing the difference spectra between tsax1 and the named instrument. One can see that the spectra on the non-blended end of the scale have more energy at larger numbers of harmonics. The location of the points on the x-axis is with respect to the difference sum.

Tristimulus Representation of Spectral Difference

Although intended as an improvement to centroid as a measure for comparing spectra, the spectral difference approach offered above suffers

Instrument	Tristim. Spect. Diff.	Composite Dissonance	Pitch Deviations		Recognition	
			Sum	Absolute Difference	Sum	Absolute Difference
flute		r = .559 *p* = .030				
oboe2		r = . 817 *p* = .0002	r = .708 *p* = .003	r = .604 *p* = .017		
englhorn			r = .703 *p* = .003	r = .585 *p* = .022		
ebclar	r = .744 *p* = .026	r = .607 *p* = .017	r = .588 *p* = .021	r = .641 *p* = .010		r = .649 *p* = .009
bassclar		r = .599 *p* = .018	r = .514 *p* = .049			
sopsax						
tsax1						
tsax2						
bassoon	r = .828 *p* = .004	r = .529 *p* = .043	r = .648 *p* = .004	r = .658 *p* = .008		r = .523 *p* = .046
frhorn	r = .793 *p* = .010	r = .748 *p* = .001	r = .698 *p* = .004	r = .753 *p* = .001	r = -.639 *p* = .010	r = .641 *p* = .010
trumpet		r = .758 *p* = .001	r = .733 *p* = ..002	t = .715 *p* = .003		
trombone			r = .794 *p* = .0004			
sulpontcello		r = .610 *p* = .016	r = .661 *p* = .007	r = .661 *p* = .007		
normcello	r = .850 *p* = .002	r = .703 *p* = .003	r = .592 *p* = .020			
mutecello		r = .633 *p* = .011	r = .563 *p* = .229			r = .582 *p* = .023

Table 5. Various *single instrument correlations* to data in Experiment UnisonBlend. Spectral difference refers to the sum of all amplitude differences between the harmonics of a pair of tones. On/Off Synchrony refers to the sum of on/off synchronies between a pair of tones, while Pitch Deviations refers to the absolute difference in pitch deviation between a pair of tones.

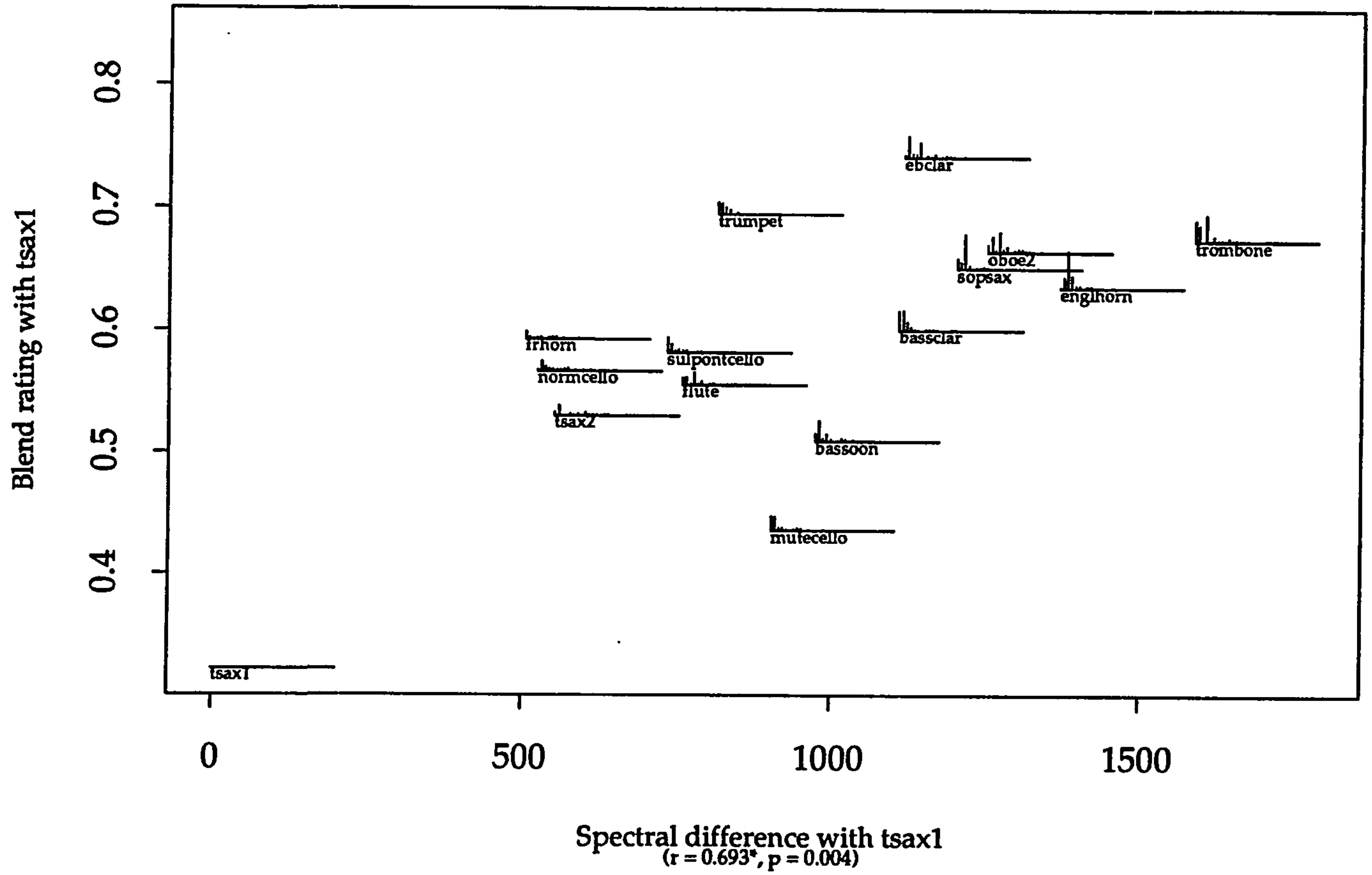


Figure 66. The blend judgments for all trials including tsax1 plotted against the spectral differences between tsax1 and the named instrument. Each datapoint is represented by the difference spectrum, anchored at its left- and bottommost point.

nonetheless by reducing spectral difference to a single value. A more useful representation would be to characterize the difference in spectrum with a higher-dimensional representation. For example, in the bottom group of spectra in Figure 65, the difference is mainly in harmonics 2 and 4, while in the top spectra the difference is more or less equally distributed over harmonics 1 through 4. A three-dimensional representation of difference was explored by modifying elements from Pollard and Jansson's *tristimulus* method (Pollard & Jansson, 1982) of representing single spectra. Their method represents a spectrum by the ratio among the amount of energy in three regions: the fundamental, harmonics 2 through 4, and the remaining harmonics. In the latter two categories, the contributions are simply summed across the member harmonics. These regions, the authors argue, represent broad categories of spectral information to which listeners attend to make distinctions between spectra.

Their method was modified here so that *two* spectra can be *compared* by the degree to which they differ in (a) the fundamental frequency, (b) the sum of harmonics two through four, and (c) the sum of the remaining harmonics. As in the previous method, this approach is dependent on the tones having the same fundamental frequency. Difference spectra were represented with this method by summing the amplitudes of the harmonics in each of the three regions, and converting each to ratios by dividing by the sum of all three. For example, the upper spectra in Figure 65 yield a ratio of about 14:61:25 (14% due to the difference in the fundamental, 61% due to the difference in harmonics two through four, and 25% due to harmonics 5 through 33), while the bottom spectra yield a ratio of about 12:57:31. The emphasis on the middle category (harmonics 2 through 4) can be seen in Figure 65.

With three dimensions of information, regression analysis is necessary to compare this representation to blend judgments. First this was done on an instrument-by-instrument basis. Predictions for blend ratings for all trials in which a given instrument was a member (n = 15) were obtained with the three values of each pair's difference ratios. Several individual instruments (six) showed significant correlations; Table 6 shows their correlations and resulting regression formulae. Interestingly, the beta weights suggest that the region carrying the most importance for determining blend varies depending on the instrument. Ebclar showed that blend could be best accounted for by the amount of amplitude difference between its middle harmonics (2-4) and those of the instruments they were paired with (as shown by the higher beta weights for this category). The same was found for tsax1 as well. This suggests that as the magnitude of difference in this region grew larger, blend decreased. The role of the fundamental seems to vary somewhat for different instruments: for ebclar and normcello it appears that larger differences in the amount of energy for the fundamental led to *better* blends, while the reverse was true for tsax1 and tsax2.showed. For the three dark instruments (bassoon, frhorn and normcello) larger amounts of difference in the high region (harmonics 5 and above) led to decreased blends.

Instrument	Regression		Beta-weights for Prediction of Blend			
	r	p	fund	mid	high	+ constant
ebclar	.744	.026	-.157	.564	.270	.155
tsax1	.718	.040	.144	.347	.255	.322
tsax2	.741	.028	.146	.022	.262	.230
bassoon	.828	.004	-.095	.032	.811	.205
frhorn	.793	.010	.069	.099	.591	.142
normcello	.850	.002	-.212	-.062	.512	.367

Table 6. Predictions of blend judgments for individual instruments from difference spectra using regression analysis, Experiment UnisonBlend. Spectra were represented by the tristimulus method (Pollard & Jansson, 1982), parsing differences into three frequency regions.

To investigate an overall effect, all 120 blend ratings were predicted from the three values of every pair's difference ratios. The result was quite high ($r = .516$, $p < .0001$):

$$\text{blend} = (-.067\ \text{fund}) + (.111\ \text{mid}) + (.415\ \text{high}) + 0.306$$

From this it appears that, in general, a large amount of difference in the high harmonics (5-33) is the most reliable predictor of blend, with larger differences producing poorer blend.

Acoustic Dissonance and Blend

The *average instrument correlations* shown earlier showed a tendency for acoustic dissonance to have an effect on blend judgments. However, the dissonances for *single* tones is quite small in magnitude (see Figure 45); the correlation between variations in blend judgments and these very tiny values is not very convincing. Calculating the composite dissonance of a *pair* of tones, however, yields somewhat higher values. To demonstrate this, Figure 67 shows the values of acoustic dissonance for an arbitrary group of instruments (flute, oboe2, bassoon, frhorn), both in solo and paired presentation; levels of dissonance are greatest in combinations of different instruments.

Correlating composite dissonance to blend judgements yields $r = .627$, $p < .0001$, very strong match. This suggests that greater dissonance among tones leads to poorer blends. Further evidence is provided by several instruments (ten) showing *single instrument correlations* (see Table 5).

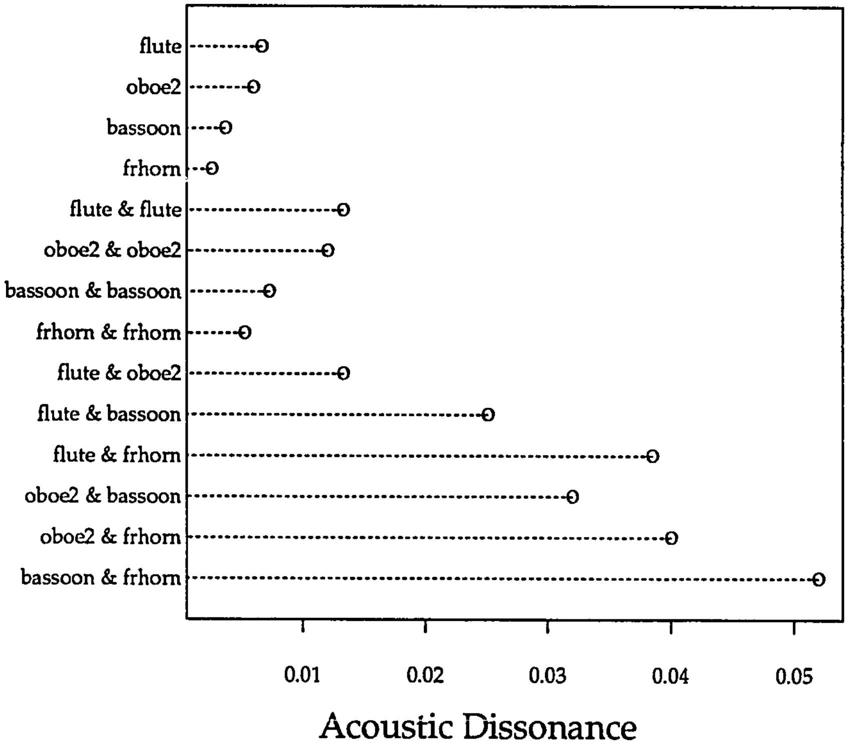


Figure 67. Levels of acoustic dissonance for the instruments flute, oboe2, frhorn and bassoon, comparing solo and paired presentations.

Attack and Blend

MDS analysis (Figure 61) and *average instrument correlations* presented evidence that the attack forms of the instruments played a role in determining blend. Further analyses concerning various measures of attack are considered here.

The *average instrument correlations* showed that a good blend tended to be obtained merely by the presence of at least one tone with a short DPAT (duration of Perceptual Attack Time). Presumably even better blends should be obtained when both tones have short DPATs; in that case summing the two DPATs (or using the other attack representation, precedent noise) of a pair would be a simple measure of the overall attack shortness of a pair. This is equivalent to the *model 1* which was used in the centroid discussion. Alternatively, the amount of difference between attacks may be an important determinant of blend; this could be represented by taking the absolute difference of DPATs for a pair. This is equivalent to *model 2*. This terminology will be borrowed here.

The two models suggest interesting interpretations of the role of onset in blend. It will be recalled that onset portions convey a great deal of information about the identity of instruments, and, when identities are more apparent, blend decreases (as suggested by the "segregation by identification" paradigm mentioned earlier). Hence blend will be aided when onsets are either masked, or when the cue that there are two onsets present is masked. This is a possible explanation for *models* 2 and *1*, respectively. *Model 1* suggests that if one of the onsets is short in duration, then the sound of the pair will be dominated by the identity of the instrument with the longer, more salient attack; hearing only one identity, the listener hears the sound as fused. The explanation for *model* 2, however, may be that the closer that the two onsets are in duration, the greater the possibility that they will mask each others' identities. The hypothesis of *informational masking* presented in Chapter 2 (Hawkins and Presson, 1977) suggests a mechanism underlaying either of these models. In *model 1*, the *precategorical acoustic store* (PAS) is dominated by the longer-lasting onset, while in *model* 2 both onsets mutually prevent each other from using the PAS.

Comparing all such sums to blend judgments in an *overall correlations* analysis produces a correlation of $r = .379$, $p < .0001$ for DPAT and $r = .334$, $p = .0002$ for precedent noise. *Single instrument correlations* with sums for both DPAT and precedent noise shows significant results for several instruments (nine for DPAT, six for precedent noise; see Table 4). Thus, *model 1* seems to supported: an overall shortness of attack duration in a pair of instruments seems to promote good blend.

With absolute differences, *overall correlations* produced significant results as well for both DPAT ($r = .503$, $p < .0001$) and precedent noise ($r = .329$, $p = .0002$). *Single instrument correlations* showed significant correlations for eight instruments in the case of DPAT difference and six in the case of precedent noise difference (see Table 4).[49] These correlations show that as the differences in attack duration between a pair of tones grew larger, blend decreased. Thus there appears to be support for *model* 2 as well. Judging from the magnitude of the correlations, in fact, *model* 2 appears to be slightly more robust than *model 1*.

Temporal Factors

It will be recalled from Chapter 2 that temporal properties play a very large role in explanations of fusion and segregation offered by the Auditory Scene Analysis paradigm. The most powerful temporal mechanism is the principle of *common fate,* which results in listeners hearing as fused sounds that change in similar ways (Bregman, 1990, pp. 248-292). Therefore it is of interest to examine possible ways in which the temporal properties of individual Stanford tones match and thereby promote blend. Some candidate factors are: harmonic synchrony (similarity in temporal harmonic activity among the two sounds) and various global envelope similarities such as pitch envelope, amplitude envelope and centroid envelope.

Harmonic Synchrony and Blend

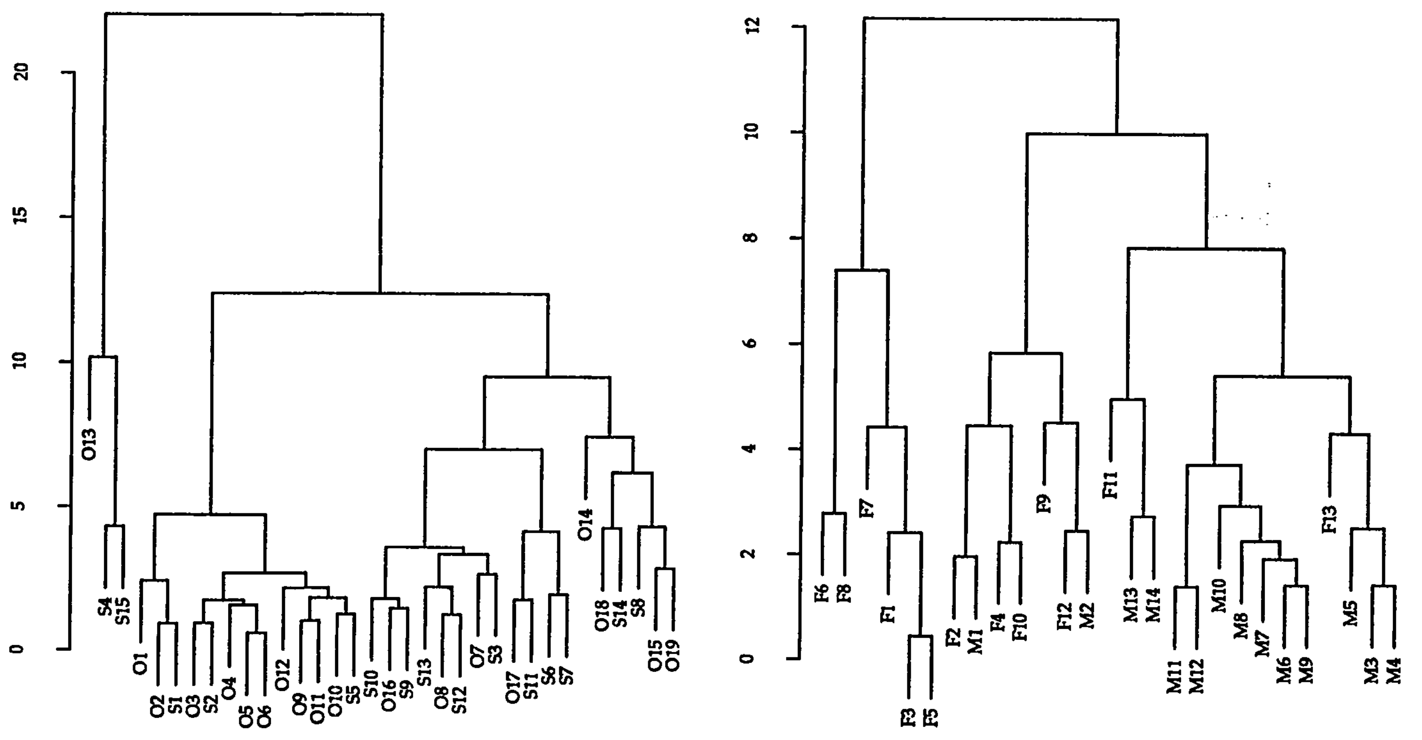
The three measures of harmonic synchrony that were proposed in Chapter 3 can be modified to address the *composite* synchrony of two

---

[49] Note that with the exception of oboe2, the correlations for precedent noise absolute differences are identical to the correlations for precedent noise sums (previous column). This is due to the fact the all the instruments in question have precedent noise durations of zero, except for oboe2.

instruments. For example, *harmonic envelope synchrony* for a single tone consisted of averaging the correlation of all pairs of harmonic envelopes within a single tone. To analyze the *composite harmonic envelope synchrony* for a pair of tones, all pairs of harmonic envelopes both within and across the two tones can be correlated.

Figure 68 shows a visualization of the harmonic envelope synchrony among pairs of instruments by means of *hierarchical clustering* analysis (Johnson, 1967). This method of representing data is similar to Multidimensonal Scaling in that psychological similarities are represented by physical proximity. Here, various levels of similarity are conveyed by the proximity of datapoints in a "hierarchical tree," where the nature of the branching organizes the datapoints into various groups, some of which are embedded in other groups. In this particular analysis, the proximities do not represent psychological similarity, but rather acoustical similarity according to the harmonic envelope correlation procedure. In Figure 68, the greater the number of strong harmonic envelope similarities across instruments, the more one should see harmonics from both instruments grouping in close proximity at the leaf nodes. The oboe2-sopsax pair shows this, and accordingly it has a high synchrony estimation. In the case where harmonic envelope synchrony is not great, the harmonics of the two instruments will be segregated, with entire branches consisting mostly of harmonics of one or the other instruments at its leaf nodes; this situation is obtained with the flute-mutecello pair, which accordingly has a low synchrony estimation (.49).



oboe2 and sopsax Harmonics

5.39

flute and mutecello Harmonics

0.49

Figure 68. Hierarchical clustering analyses showing the amplitude envelope similarities among all harmonics for two pairs of instruments. Harmonics are coded by the initial of the instrument and the harmonic number. The two pairs shown illustrate very high and very low levels of harmonic envelope synchrony, respectively.

Although the visualization shows striking evidence that there were strong similarities and differences between individual envelopes, statistically it did not appear to have an effect on the blend judgments. *Overall correlations* with this factor yielded non-significant results, as before ($r = .069$, $p = .456$). Furthermore, *single instrument correlations* with composite harmonic envelope synchrony turned up only one significant result (flute).

A composite measure of *peak synchrony* was obtained by averaging over the standard deviation of the peak amplitude times among the harmonics of both tones. This too produced no *overall correlations* but two *single instrument correlations* (flute and normcello). Thus peak synchrony among two tones was not a convincing predictor of blend.

A composite measure of *onset/offset synchrony* was the only measure to obtain significant *overall correlations* ($r = .294$, $p = .001$), but here no *average instrument correlations* obtained significance. Overall, apart from this weak suggestion of onset/offset synchrony, measures of synchrony in general did not seem to have been a meaningful determinant of blend.

Temporal Envelopes

The three measures of synchrony characterized temporal similarities between tones that emerged from the accumulation of envelope similarities among the various harmonics of the tones. Another class of temporal similarity is by means of more global time-varying properties of the tones, such as amplitude, centroid and pitch envelopes. If two instruments are very similar in any of these properties, the "common fate" principle may suggest the cue to the listener the two sounds arise from a single source, and therefore blend. Evidence for correlations between envelope similarities and blend judgments are considered below.

Amplitude Envelopes

Amplitude envelopes for the 15 Stanford tones were shown earlier (see Figure 48). Estimates of similarity between the two amplitude envelopes of a pair of tones were made by correlating the data for their curves.[50] To demonstrate the nature of this analysis, the lower half of Figure 69 illustrates a correlation between very similar amplitude envelopes (flute and oboe2), yielding a high correlation. Correlations were calculated for all the pairs in the experiment (n = 120) and the correlations themselves were used as a factor in *overall correlations*. This yielded a significant correlation: $r = -.359$, $p = .0001$. These effects can be observed in individual instruments as well: *single instrument correlations* yielded significant results for four instruments (englhorn, oboe2, frhorn and trombone). To illustrate the correlation, the results from englhorn's *single instrument correlation* is illustrated in Figure 70. In this graph, the amplitude envelope of the named instrument is plotted at the value of its correlation with englhorn's amplitude envelope (x-axis) and the blend rating for that pair (y-axis). The sign of all the

---

[50] All curves were time-aligned according to the PAT offsets that were used when they were synthesized.

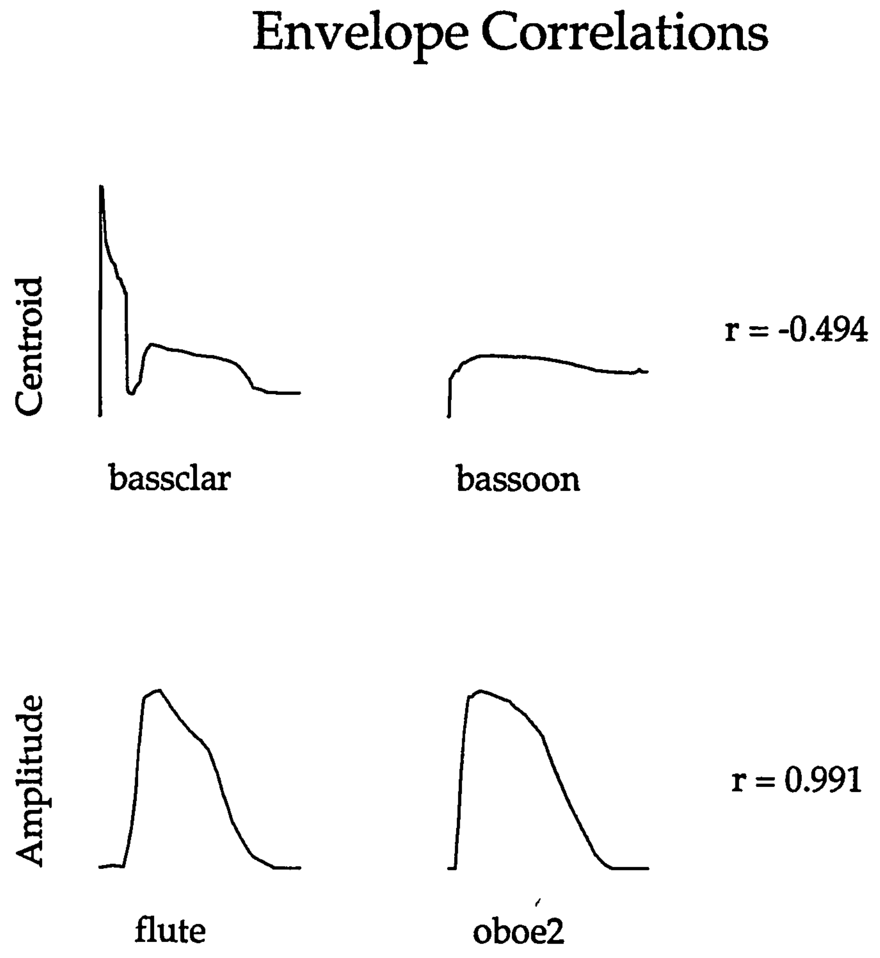


Figure 69. Demonstration of correlation between centroid and amplitude envelopes, showing very dissimilar and similar envelopes, respectively.

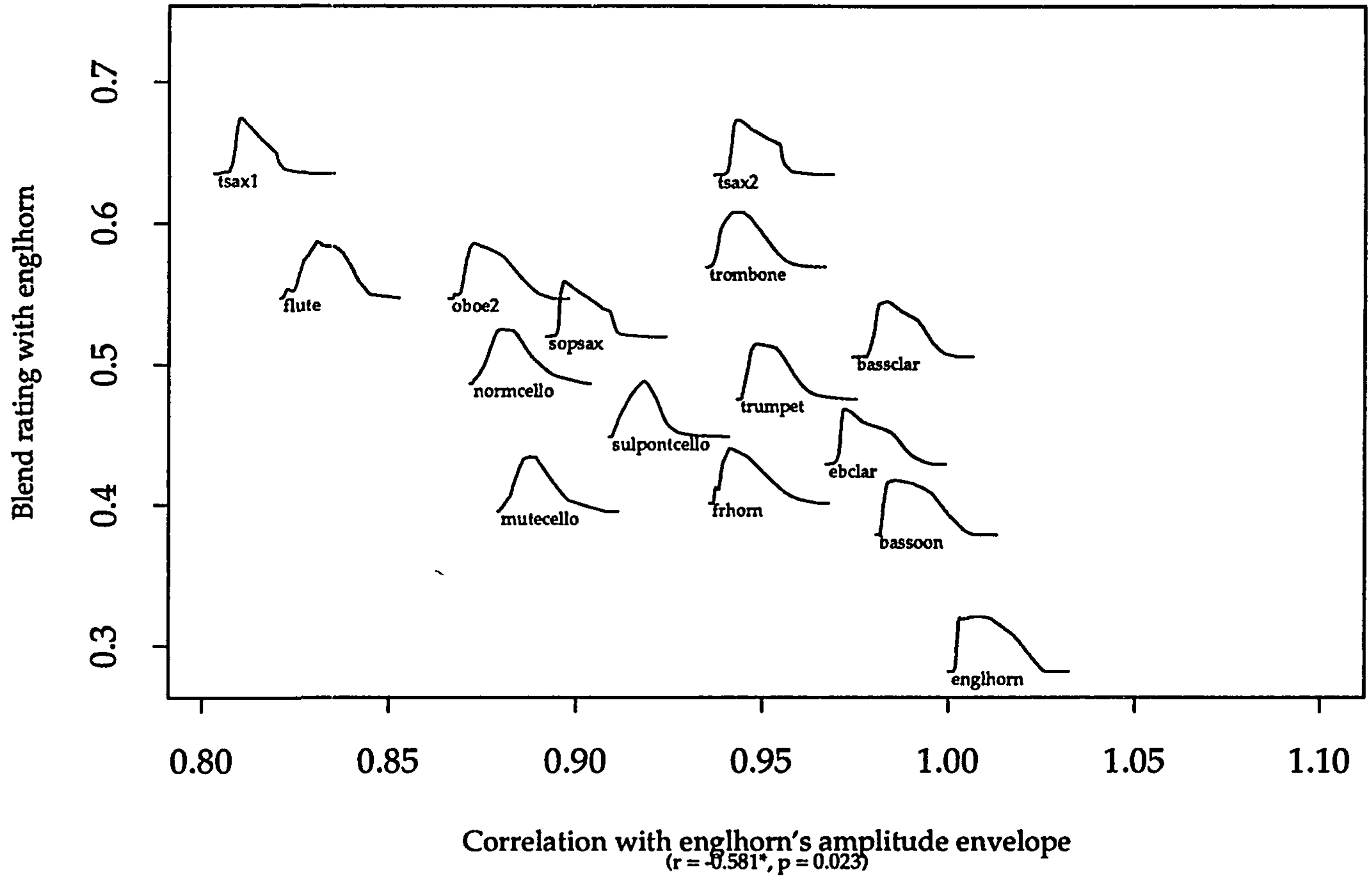


Figure 70. Relationship between blend and the amount of similarity between amplitude envelopes, englhorn trials only, Experiment UnisonBlend. Instruments are positioned on the x-axis according to the correlation of their centroid envelope with englhorn's. The y-axis shows the blend rating for the pair, and the datapoint is plotted as the amplitude envelope of the named instrument.

correlations shows that as similarity between amplitude envelopes grew larger, blend increased.

Centroid Envelopes

Earlier, *centroid envelopes* for the tones were shown (see Figure 43). The change in brightness over time is presumed to be a salient timbral property, since temporal changes in spectrum carry a great deal of information concerning the identity of a tone (Chowning, 1973), centroid envelopes. As with amplitude envelopes, the centroid envelopes for each pair of instruments were correlated to one another and the curves were time-aligned according to the PAT offsets that were used when they were synthesized. To demonstrate the nature of this analysis, the upper half of Figure 69 illustrates the correlation between two very different centroid envelopes, yielding a poor correlation. Performing correlations between all pairs of instruments yielded a list of 120 values; these correlations *themselves* were used in a correlation with blend judgments. An *overall correlations* procedure produced significant results: $r = -.394$, $p < .0001$. The sign of the correlations shows that as similarity between centroid envelopes grew larger, blend increased.

These correlations can be observed with several individual instruments as well. *Single instrument correlations* with this factor showed significant results with oboe2, englhorn, ebclar, tsax2 and trombone; the correlation coefficients are given in column seven of Table 4. For illustration, the results from oboe2's *single instrument correlation* is shown in Figure 71. In this graph, the centroid envelope of the named instrument is plotted at the point of its correlation with oboe2's centroid envelope (x-axis) and the blend rating for that pair (y-axis). The greater the centroid envelope similarity, the better the blend.

Pitch Envelopes

Curiously, correlation among pitch envelopes (see Figure 55), did *not* demonstrate itself to be an important determinant of blend. The *overall correlation* between the correlation coefficient of the two tones' pitch envelopes and the blend judgments was non-significant ($r = -.127$, $p = .166$) and only one significant *single instrument correlation* was found (sopsax). This is a surprising outcome, since pitch tends to be such a pronounced aspect of auditory perception. Perhaps the patterns of change between instruments were not contrasted enough for this to show an effect. However, there was a strong effect for the amount of *pitch deviation*, as shown in the following section.

Pitch Deviation

Evidence that pitch deviations had some impact on blend judgments was found in the MDS solution (Figure 61) and the *average instrument correlations* mentioned earlier, so it is expected that further evidence can be found for these factors. It is unclear how one should measure the *composite* effect of pitch deviation between two tones, or what such a measure would

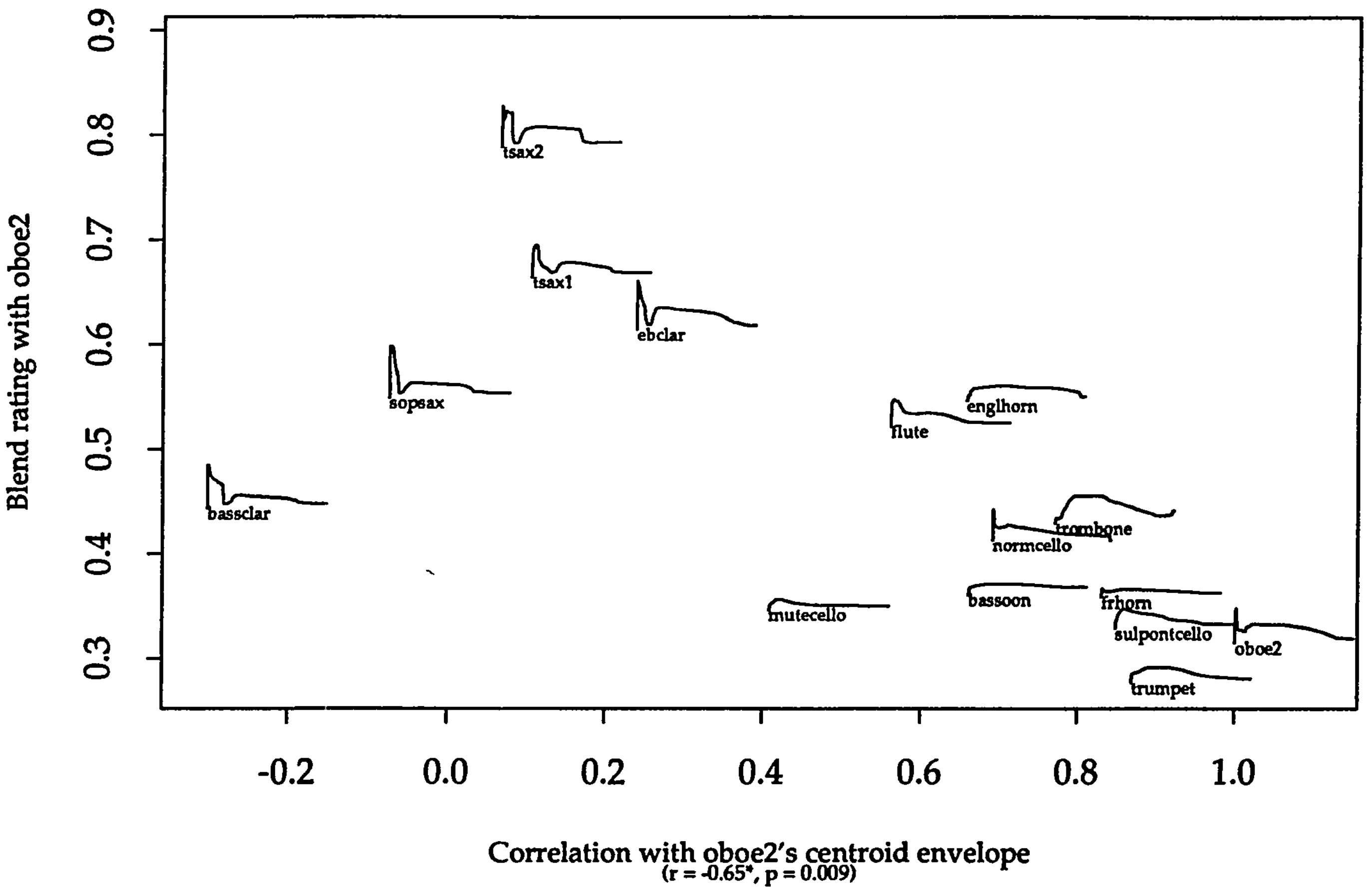


Figure 71. Relationship between blend and the amount of similarity between centroid envelopes, oboe2 trials only, Experiment UnisonBlend. Instruments are positioned on the x-axis according to the correlation of their centroid envelope with oboe2's. The y-axis shows the blend rating for the pair, and the datapoint is plotted as the centroid envelope of the named instrument.

mean. The methods used for the factors of centroid and onset--sum or absolute difference correlations--may be the best candidate for examining the interactions between levels of pitch deviations for a pair of tones. As before, it is necessary to distinguish between the meaning of outcomes in sum vs. absolute difference correlations, so a two-model explanation is offered. In both models the mechanism of *informational masking* seems relevant once again, because pitch deviation is considered to be a salient cue for identity. *Model 1* (sums) suggests that the greater the overall amount of pitch deviation in both tones, the more salient the instruments' identities are, which in turn makes them more distinctive, recognizable, and therefore leads to segregation of the two sounds. *Model* 2 (absolute differences) suggests that as pitch deviations are more similar to one another in magnitude, this information will mutually mask one another's identities, and the two instruments will blend.

Model 1 can be tested with sum correlations; a strong *overall correlation* with this factor suggested that the greater the overall amount of pitch deviation, the poorer the blend ($r = .509$, $p < .0001$), and 11 significant *single instrument correlations* lent further support (see Table 5). This suggests support for *model 1.*

Absolute difference correlations yielded equally powerful correlations, with an *overall correlation of* $r = .499$, $p < .0001$ and seven *single instrument correlations* (see Table 5). So there is strong support for *model* 2 as well.

Instrument Recognition

Grey's recognition measure showed results analogous to those of pitch deviations. Both sum and absolute difference correlations are found, but with less power and different directions of magnitude. Absolute differences showed *overall correlations* at $r = .239$, $p = .008$ and four *single instrument correlations* (see Table 5). However, sum correlations showed *negative* significant results: *overall correlations* at $r = -.338$, $p = .0002$, and one significant *single instrument correlation* (frhorn). The combined effect of these two kinds of correlations might suggest the following principle: blend is better when both instruments are highly recognizable, but worse when one instrument is not easily recognizable and the other is highly recognizable. Admittedly, this principle does not sound musically intuitive.

Modelling Blend

The objective of the current study is to arrive at a model of blend perception for concurrently-sounding orchestral tones. So far there are over a dozen factors which have been shown to have a relationship to blend judgments. Clearly then, no single factor can predict the blend of all possible pairs; some dimensional model involving multiple factors must be used to account for the variance of the data. Earlier, Multidimensional Scaling was used to estimate the primary dimensions. Here the method of *regression* will be applied.

If accuracy is the main criteria behind the model, then perhaps the best approach is one which combines all factors; this can be effected by means of a regression analysis with all factors predicting blend. If musical usefulness is the main criteria, then it might be best to try to come up with a low dimensional solution, based on perceptual variables that are cognitively familiar and available for conscious musical scrutiny. An attempt at each approach will be considered here.

Regression with All Factors

A regression with the 16 factors which showed significant *overall correlations* in the previous analyses resulted in a correlation of $r = .872$, $p < .0001$.[51] The resulting equation for predicting blend is shown below.[52]

blend = .146 diss + -.093 fund + -.064 mid + .231 high + .257 pitchdevs_sums + .252 PATs_absdiff + -.021 pitchdevs_absdiff + .245 cents_sums + .078 centenv_corrs + .039 PATs_sums + -.019 ampenv_corrs + -.074 recogntion_sums + -.184 noise_sums + -.085 noise_absdiff + .166 spectdiffsums + .040 onoffsyncs + .007 recognition_absdiff + -.059 cents_absdiff + .09

One informative aspect of the regression equation is that the beta weights suggest a prioritization of the various factors from most to least important. For example, pitch deviation sums are weighted more than any other factor. The relative contribution of each factor to the equation can be appreciated in Figure 72, showing a series of regressions with each factor

---

[51] The magnitudes of the *overall correlations* for the factors was used to decide the order in which they were entered in the input for the regression.

[52] All factors were scaled to a common 0 to 1 scale prior to running the regression. The purpose of this was so that the magnitudes of the beta weights could be evaluated on the same scale.

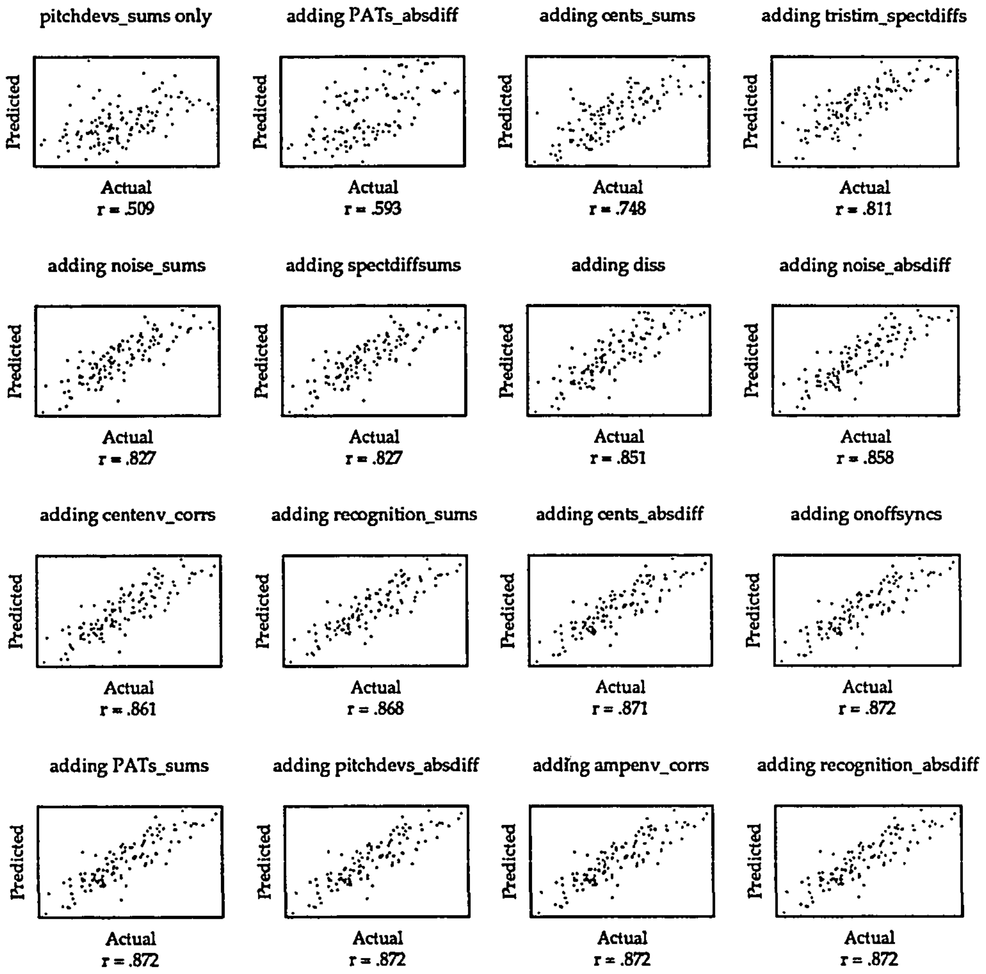


Figure 72. Scatterplots showing improvement of regression analysis to predict blend judgments as more acoustical factors are added, Experiment UnisonBlend.

added one at a time. The order of factors is based on the magnitudes of the beta weights in the formula above so that more important factors are modelled earlier. The first plot shows the prediction with one factor, the second plot with that factor plus a new one, and so on, successively adding each factor. With each successive regression the plot of predicted against actual blend values comes closer to approximating a straight line. This trend continues until about the 11th factor, where it ceases to improve (asymptotes). This suggests that these later factors are not as important in determining blend.

Figure 73 shows the actual vs. predicted data of the 16th regression in Figure 72 up close. This constitutes the "best possible" model for blend in this experiment, since it so closely approaches a straight line. However, it is interesting to try to evaluate in what ways this modelling process breaks down; after all, there are numerous "outliers" which apparently were not accounted for by the 16 factors. Examining these in detail may provide clues as to what factor was missing from the analysis process. Outliers that were beyond a particular threshold (arbitrarily declared to be those points falling outside a standard deviation of the mean error) are identified in Figure 73.

One can think of these outliers as the trials that "fell through the cracks"--trials that could not be accounted for by any of the 16 factors. One striking aspect is the number of single reed instruments: in the 23 outlier pairs, 15 include at least one single reed instrument. Since there were only 65 trials in the entire experiment that had a single reed, this means that

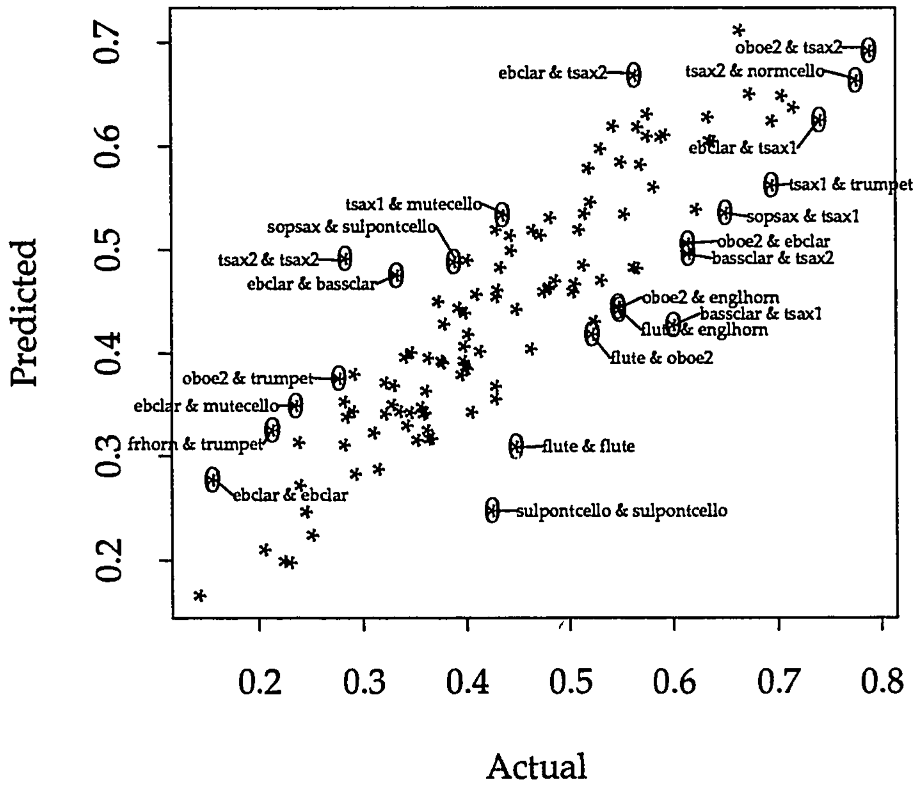


Figure 73. The scatterplot for the 14th regression shown in Figure 72, comparing the values produced by the regression equation with the actual blend judgments. The outliers (those that varied more than a standard deviation from the mean error) are labelled by their associated trials.

23% of the trials involving single reeds were poorly predicted by the regression, a fairly high percentage. Indeed, an examination of Tables 4 and 5 reveals that tsax1, tsax2 and bassclar showed fewer *single instrument correlations* than other instruments. Interestingly, the errors are approximately equally divided between both positive and negative errors. The most frequent "offenders" are ebclar, tsax1 and tsax2, and again, negative and positive errors are equally distributed. One conclusion that can be drawn from this is that single reed instruments tend to blend in less predictable ways.

A Low-dimensional Solution

The drawback of the regression approach taken above is that there were far too many relevant factors; among them seemed to be redundant factors, and even contradictory information. As a model that seeks completeness, it is permissible to not understand the meaning of all the dimensions or their interactions. However, the goal of a *musical* model of blend is to reduce the variables to a cognitively manageable set of distinctive features with musical relevance. To reduce the number of features, two main strategies will be adopted: reduce the number of factors, and weed out any redundant factors. Criteria for eliminating or retaining factors considered were:

(1) Retain factors found to correlate to the two dimensions of the MDS solution (Figure 61). This consists of the factors centroid, acoustic dissonance, DPAT and pitch deviations.

(2) Eliminate factors whose *overall correlations* were smaller in magnitude (see Table 7).

(3) Eliminate factors with few *single instrument correlations* (see Table 7, rightmost column). Factors with fewer than five include the three mentioned in (2) above, plus amplitude envelope correlations.

(4) Eliminate factors with small beta weights from the regression equation. For example, all the asymptotic factors in Figure 72 might be eliminated (recognition sums and absolute differences, centroid absolute differences, DPAT sums, on/off synchrony, pitch deviations absolute differences and amplitude envelope correlations).

(5) Eliminate some of the factors which are different versions of the same phenomenon: this includes the factors centroid (2 variants), attack (4 variants) envelope (2 variants), recognition (2 variants), pitch deviations (2 variants) and spectral difference (2 variants). However, this is desirable only so far as it can be shown that they are sufficiently redundant to warrant elimination. Regressions can be used to determine the relative importance of these similar factors: by examining the beta weights it can be determined which of the variants seem to have the most influence. For example, the

Factor			Blend Improves...	Correlations		
				r	p	# of SICs
Spectrum	Centroid	Sums	as the sum of centroids for a pair gets lower	.481	< .0001	6
		Diff	as the absolute difference of centroids for a pair gets lower	.229	.001	3
	Difference	Tri-stimulus	as the proportion of difference in harmonics five and above is minimized	.516	< .0001	6
	Dissonance		as the composite dissonance of a pair gets lower	.627	< .0001	10
Attack	PAT	Sum	as the sum of PATs for a pair gets lower	.379	< .0001	9
		Diff	as the absolute difference in PATs for a pair gets lower	.503	< .0001	8
	Precedent Noise	Sum	as the sum of the duration of precedent noise for a pair gets lower	.334	< .0001	5
		Diff	as the absolute difference in the durations of precedent noise for a pair gets lower	.329	< .0001	6
Temporal Envelope	Centroid		as the correlation of centroid envelopes of a pair gets higher	-.394	.0002	5
	Amplitude		as the correlation of amplitude envelopes of a pair gets higher	-.359	< .0001	4
Other	Recognition	Sum	as the sum of the recognizability for a pair gets higher	-.338	.0002	1
		Diff	as the absolute difference in the recognizability for a pair gets lower	.239	.008	3
	Pitch deviations	Sum	as the sum of the amount of pitch deviation for a pair gets higher	.509	< .0001	11
		Diff	as the absolute difference in the amount of pitch deviation for a pair gets lower	.499	< .0001	7
	On/Off Synchrony		as the composite on/off synchrony for a pair gets lower	.294	.001	0

Table 7. Various factors identified as correlating to blend judgments in Experiment UnisonBlend. (SICs = *single instrument correlations*)

regression equation below suggests that DPAT differences be retained and the others discarded:

$$\text{blend} = .074\ \text{DPAT\_sum} + .264\ \text{DPAT\_diff} + .019\ \text{noise\_sum} + -.070\ \text{noise\_diff} + .360$$

Considering all the candidate evidence for eliminating factors, it seems that the following five factors ought to be included:

centroid sums
composite dissonance
Tristimulus representation: region of harmonics 6 and above only
DPAT absolute differences
pitch deviations sums.

This model can be checked with regression. This results in a correlation of $r = .817$, $p < .0001$, and the following regression equation:

$$\text{blend} = .314\ \text{cents\_sum} + .189\ \text{diss} + .128\ \text{high} + .087\ \text{DPATs\_diff} + .157\ \text{pitchdevs\_sum} + .132$$

Summary of Findings, Experiment UnisonBlend

Seven subjects showed a great deal of consistency and agreement with one another on the judgment of blend. It appears that, for some western musical listeners at least, there is some consensus on what "blend" is.

Aspects of spectrum (including both centroid and acoustic dissonance), attack duration, temporal envelope similarity, recognizability, pitch deviation and harmonic synchrony (totalling to 15 factors) were found to relate to the blend judgments to varying degrees of magnitude. However, by eliminating some redundant and contradictory factors, a satisfactory model of the blend data could be found with just five factors. The factors determining blend (from most to least important) were the overall amounts of centroid height, composite dissonance and pitch deviation for a pair, the amount of difference between tones in the upper harmonics of their steady state spectra, and the amount of difference in the attack times of the tones, measured in terms of their Perceptual Attack Times.

Sums and Difference Correlations

In several of the factors, correlations could be found with both the sums and absolute differences of the values for the two tones: this was true with centroid, DPAT, precedent noise, recognition and pitch deviations. In the cases where both correlations were positive, this suggested a "gravitational principle" of blend. For example, with centroid, it was true both that the lower that each of the centroids was, and the closer the two centroids were in value, the better the blend. Thus the gravitational principle means that the best blends are obtained when the value of some factor is as low and close as possible for a pair of tones. The exception was with recognition, whose sum correlations were negative.

The magnitude of correlations tended to be greater, in general, for sums than absolute differences. This can be shown to be related to the concept of "blending power." When a particular instrument had an extremely high value for a given factor, this tended to make its sum with any other instrument high. For example, the centroid for the muted trombone is very high; even when paired with a dark instrument like mutecello, the centroid sum of the two is still relatively high, due to the extreme value of the trombone. For the same reason (but on the opposite end of the spectrum), very dark instruments such as frhorn, bassoon and mutecello were "good blenders."

In most cases, factors whose correlations produced significant results with the whole dataset (n = 120) also showed significance with subsets of the data in which a single instrument was considered (n = 15); this was demonstrated in the many *single instrument correlations* that were considered. These correlations demonstrated that the blend for a given instrument can increase or decrease by pairing it with instruments having lower or higher values for a given factor; for example, blend for bassoon increases as the instrument that is paired with it has a smaller amount of pitch deviation. For any given instrument, however, there was more than one factor determining its blend; bassoon, for example, was governed by pitch deviation and acoustic dissonance as well.

The multiple influence of these three factors on blend for the bassoon have been represented in a three-dimensional "blend space" for the bassoon in Figure 74. Each datapoint portrays an instrument which is paired with bassoon. In this space, blend improves in any direction out from the rear, lower right hand corner. In the centroid dimension, blend increases from left to right; in the pitch deviation dimension, blend increases from front to rear; and in the acoustic dissonance dimension, blend increases from top to

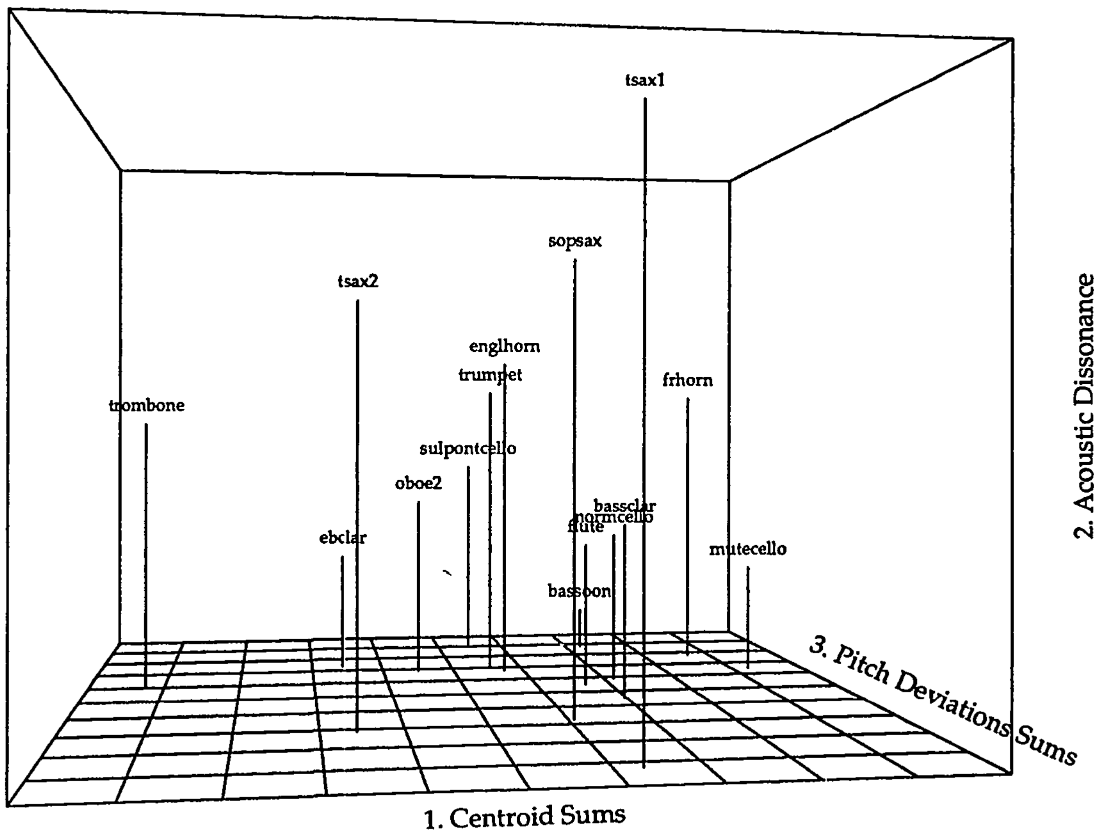


Figure 74. "Blend space" for bassoon. Three dimensions which had significant *single instrument correlations* with the bassoon trials are plotted against one another.

bottom.[53] This suggests compositional applications: one could explore a range of blends with the bassoon in a number of ways, by moving through the space in various dimensions. For example, the succession of pairs bassoon-tsax2 to bassoon-ebclar increases blend due to its change in only two of the dimensions (pitch deviations and acoustic dissonance), whereas the succession bassoon-tsax2 to bassoon-mutecello increases in all three dimensions.

No correlations for the factor of pitch envelope were observed. Even though the overall difference in amount of pitch deviation between a pair of tones was an important factor, whether the trajectories of the pitch curves matched or not did not appear to matter; this was surprising since the trajectories of amplitude and centroid curves *did* play a role. Also absent among any of the correlations was the factor of amplitude scaling; this suggests that Grey's loudness equalizations, used here, were successful. Harmonicity, surprisingly, also showed no correlations.

A note about the unusual data by subject MAN is warranted. MAN's data showed all attributes of that of other listeners: his data was very consistent, and his averaged data correlated highly with the others. However, all the correlations were negative. A later interview with MAN showed that he understood the task and the nature of the rating scale. Possible differences between this listener and the others were explored. MAN was the only

---

[53] The measure of acoustic dissonance is *composite* acoustic dissonance for the pairs in question; thus, this dimension corresponds to the values in Table 8 rather than those in Figure 45.

percussionist in the group a fact that might suggest that that he possessed a more heightened awareness of temporal synchrony among tones; yet it is hard to see how this would produce the data reversal that occurred. A more interesting avenue of explanation had to do with his musical background and career as a jazz drummer. According to his own account, MAN's exposure to jazz was rich and from an early age because his father was himself a jazz arranger (having studied with Henry Mancini at one time). It is possible that definitions of blend vary for different people and different musicians; recall that one jazz orchestrator advises that "The best and most colorisitically valuable pairing is obtained between two dissimilar instruments, at least one of which is highly colored by itself (the oboe) or used in a highly colored form (trumpet in Harmon mute)" (Russo, 1968, p. 562). However, an operative perceptual definition of blend was supplied to the listeners (detecting one vs. two sounds) was chosen so as to be free of any aesthetic implications. It seems unlikely that MAN could have made construed the instructions interpretation of the instructions at so fundamental a level, so the results remain a mystery.

Experiment 2 (m3Blend)

Preliminaries

The use of only a single pitch interval of presentation in experiment UnisonBlend (i.e., a unison, both instruments on Eb4) was a restriction imposed by the limitations of the Stanford tones, as well as practical limitations of experimental design; given the opportunity, a range of musical intervals should be explored. The unison, moreover, is an especially unique case, as Chapter 2 showed: more than any other interval it tends to naturally "fuse." In other words, the unison gave the tone pairs in this experiment an advantage in their tendency to blend that does not exist in all other musical intervals. Indeed, many listeners reported that with unisons it was difficult to detect that two instruments were even present in a number of the trials. It is of interest, then, to explore the factors that determine blend when the advantage of a highly consonant interval is not present.

One possibility would be to explore several different intervals and observe the different blends for each. However, to make the experiments of a practical size, the number of instruments would have to be limited to far fewer than the 15 used in UnisonBlend. The disadvantage of this approach is that a group of, say, only four instruments would create a very different timbral space than the full range of 15, and comparison of results to those of experiment UnisonBlend would be difficult. So it was decided to use a single interval and use the same (or similar) number of instruments.

The choice of interval was guided by the desire to avoid transposing the Stanford tones any farther than necessary (since any transposition departs from the acoustic reality of the given instruments), and the desire to choose an interval that was less naturally "fused" than a unison; the other perfect intervals (octave, fourth, fifth) tend to fuse as well (see Stumpf, 1890; DeWitt & Crowder, 1987), so they should be avoided. On the other hand, the other end of this spectrum (such as a minor or major second) should be avoided, since they will tend to easily separate out into two events. A minor third (i.e., fundamentals having a frequency ratio $\sqrt[12]{3}$) appeared to be a satisfactory interval of moderate consonance. This interval could be obtained with the Stanford tones by transposing one tone down to c# (two half steps) and the other up to e-natural (one half step).

To obtain the transpositions, the frequency content of the tones was shifted in these directions wholesale; that is, no attempt (which would have been theoretical at best) was made to adjust the tones to reflect the kind of spectra which would have occurred in authentically obtained tones at these pitch levels. The premise behind this decision was that (a) changes of such small distances would not result in spectra that were radically different than the spectra obtained from authentically played tones at those pitch levels, and (b) if wholesale transpositions did create non-natural spectra, the transgressions was probably greater in the upward transpositions than downward ones (which motivated the particular note choices for the minor third used here).

To see if (1) was indeed true, the spectra of the Grey tones were compared to steady state spectra of the same instruments and pitch levels collected from another source. Spectral analyses of orchestral instrument tones from the McGill University Master Samples (MUMS) compact disc collection (Opolko and Wapnick, 1989) were collected by the author (see Sandell, 1991 for further details) including those for the pitches c#4, eb4 and e4 for each of the instruments.[54] The important questions are: what is the nature of spectral change from eb down to c# and up to e, and how extensive is it? Centroids were used as the basis for comparison; Figure 75 shows the centroids for the MUMS tones on the right and for the Stanford tones on the left.

Looking at individual tones, the centroid changes across the three tones are fairly modest (with the exception of the normcello e), so for both sets of tones (Stanford and MUMS), there are no *dramatic* shifts in centroid from note to note. However, it is noted that the linear transpositions made to the Stanford tones produced a monotonicity between pitch/centroid which is not observed in the MUMS tones. There are also some differences in absolute and relative centroid height between the two collections, but this is a minor concern here. In any case, listening to the transposed Stanford tones confirmed that the perceptual changes were minimal and they remained identifiable. In addition, it may be noted that the Stanford tones have been transposed by Grey himself for use in other studies.[55]

Although the design of the new experiment was made to facilitate comparisons with data from experiment UnisonBlend as much as possible, it

---

[54] The one exception is bass clarinet; the MUMS collection did not include any pitches beyond c#4 for this instrument.

[55] In Grey (1978) the tones were transposed by five semitones (in both directions) for an experiment in musical patterns. Gordon (1984, p. 2) reports that Grey used the Stanford tones to make a synthetic realization of Raymond Erickson's *LOOPS* (Erickson, 1974). Judging from the musical example in Erickson (1982, p. 528), this would have required transpositions by at least a tritone in one or both directions.

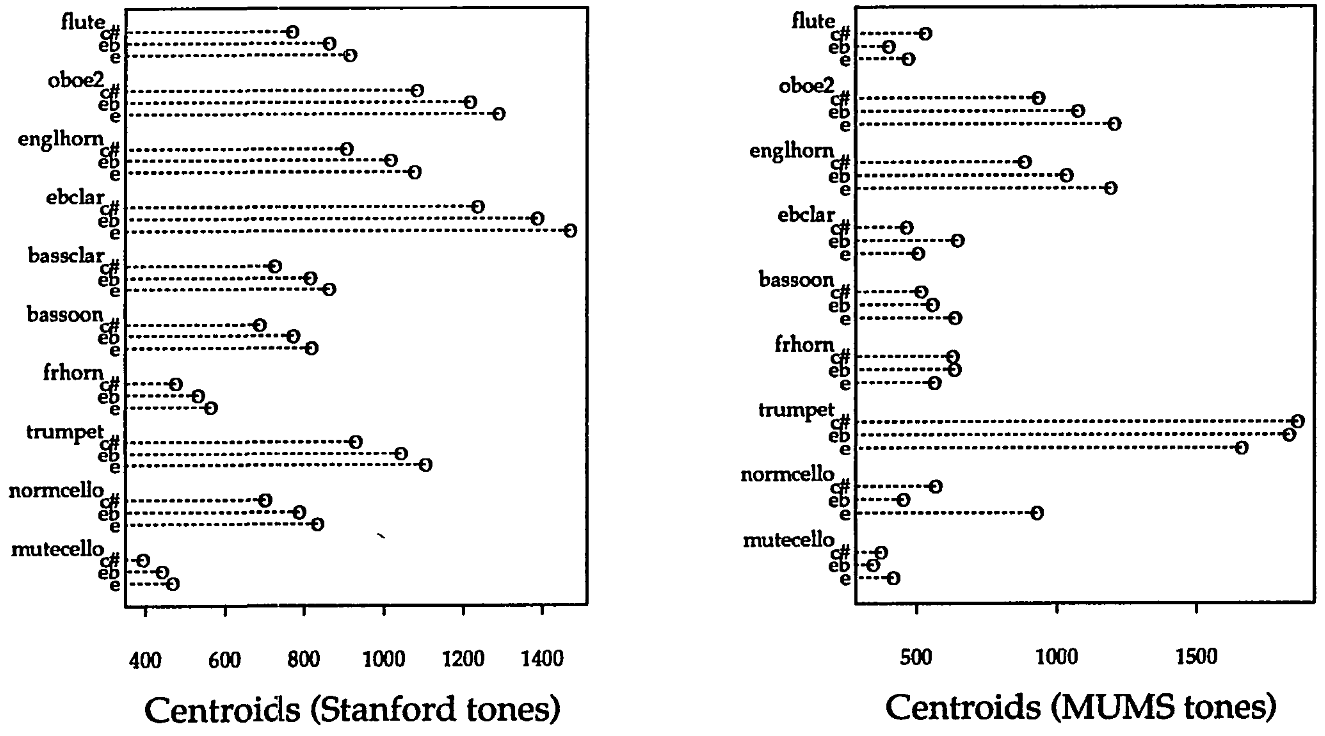


Figure 75. Centroids for the Stanford tones used in Experiment m3Blend, at their original (eb) and altered (c# and e) pitch levels, on the left; on the right, the same tones as analyzed from the McGill University Master Samples (MUMS) compact disc collection. The bassclar notes are omitted from the second graph because they do not appear in the MUMS collection.

was desired nonetheless that some instruments be eliminated. The main purpose was to keep the size of the experiment within practical bounds since the number of trials were roughly doubled by the new pitches, but there were other considerations motivating the specific choices. It will be recalled that recurring evidence showed that instruments tsax1 and tsax2 were highly differentiated from the remaining instruments, even to the point of some loss of information (cf. the MDS analysis, Figure 58). Because the strong differences in the data for the saxophones were hypothesized to be related to the fact that jazz instruments invoke different musical schema than "classical" instruments, the soprano saxophone was eliminated as well. The muted trombone was eliminated due to its extreme brightness (see Figure 41): without the trombone, the collection of centroid levels exhibited a more smooth continuum of values. Finally, the inclusion of three celli seemed redundant, especially since their ratings in experiment UnisonBlend show such strong similarities. Specifically, the means for all the sulpontcello trials and all the normcello trials in experiment UnisonBlend were not statistically different from one another ($t[28] = .631, p = .533$). It was decided, then, to eliminate sulpontcello and use only normcello and mutecello.

Stimuli

The ten instruments flute, oboe2, englhorn, ebclar, bassclar, bassoon, frhorn, trumpet, normcello and mutecello were generated at two new pitch levels each, c#4 (277.18 Hz) and e4 (329.63 Hz), for a total of 20 tones. The exact pitch levels of the syntheses were determined by the same method that was used in the previous experiment: adjusting them so the average frequency of their first harmonic fell at the frequencies for c#4 and e4.

Using these ten instruments at the two pitch levels, all possible pairings of one instrument on the pitch c#4 and another instrument (same or different than the one on c#4) on the pitch e4 were synthesized. For example, there were four different trials involving flute and/or oboe2: flute-flute, flute-oboe2, oboe2-flute, and oboe2-oboe2 (where the first named instrument is on c# and the second on e). This led to a total of 100 trials. The relative attack times of each pair were offset according to Gordon's PAT, as before.

When listening to the transposed tones using Grey's original amplitude levels (as used in the previous experiment) the tones seemed, unlike those in UnisonBlend, uneven in loudness. The e-natural transposition always seemed louder than the c# transposition (possibly this was due to the presence of higher frequencies brought on by the direct transposition). To fix this, the tones were equated in RMS energy. With this change the tones seemed much more uniform in loudness to the author, so these equalizations were used in the current experiment.

Subjects

Eight listeners participated in the study, all of whom were Northwestern students or faculty currently involved in some kind of musical activity. Three were researchers in music (music theorist, or investigator of music-related auditory phenomena), four were performers enrolled in or recently graduated from an instrumental music degree program, and one was a composer. Three of the listeners had some acquaintance with the Stanford tones and the synthesis method involved, and were intimately acquainted with the methodology, development and purpose of the experiment. The initials of the eight subjects were BLW, BOT, GJS, GSK, MRG, TMG, TMW and WLM. Two listeners, GJS and WLM had also participated in experiment UnisonBlend. Listeners were not paid.

Equipment

The room, playback equipment, level calibration method and rating method were the same as in experiment UnisonBlend.

Procedure

The instructions were identical to that of experiment UnisonBlend, except amended to refer to an interval of a minor third rather than a unison. To stress the importance of attending to the timbre of the stimuli, the experimenter suggested the analogy of hearing two singers and trying to determine if one or two vowels were sung.

As in experiment UnisonBlend, the experimenter played the ten Stanford tones individually over the loudspeakers while the listener followed a list of the names of the instruments in the order that they were sounding. The listener was urged not to let his tendency to recognize the instruments influence his judgements, and to make his answers without lengthy introspection or numerous replays.

Each listener ran the complete block of 100 trials a total of four times, and each heard the trials in a different random order in each block. At the beginning of the experiment, 50 randomly selected practice trials preceded the actual experiment. The total duration of the experiment, including instruction, was typically one hour, and was completed in a single session.

Results of Experiment m3Blend

Variability of Responses

All listeners showed consistent performance across their four blocks of trials: correlations among each subject's four blocks (6 correlations for each subject) yielded significant results in every case. 40 out of the 48 total correlations had $p < .0001$, and only three had probabilities as low as $p < .05$. The means of each listener's four blocks were compared using t tests: except for two cases, all 48 comparisons yielded non-significant differences ($p < .05$), meaning that listeners did not appear to change the magnitude of their judgments from block to block. Because of these consistencies, each listener's block of four trials were averaged together into a single dataset representing that listener.

The listeners' averaged datasets correlated with one another quite well, with the exception of one subject (TMG), who showed non-significant correlations for four out of the seven comparisons with other listeners. For the remaining seven listeners, however, all correlations were significant at $p < .0001$. It was decided then, to discard TMG's data and combine the remaining seven subjects' data into a single averaged dataset. The distribution of ratings in the dataset, shown in Figure 76 shows that a fairly wide range of values is demonstrated; as in experiment UnisonBlend, the data is skewed towards the "blended" end.

Terminology: A Preliminary Note

The new parameter (pitch-instrument attribution) in this experiment introduces new considerations in the data analysis. For clarity and brevity of discussion the following conventions will be adopted:

1. The indication *instrument1-instrument2*, as in "flute-oboe2" and "bassoon-bassoon," always refers to the specific trial in which the instrument 1 is on the note c# and instrument 2 is on the note e. A pair of instruments in brackets, such as [flute-oboe2], is used to indicate any unspecified combination of flute and oboe2, or in some contexts, to refer to both pairs. The use of the word "pair" will be reserved for the former case (e.g. "the pair flute-oboe2"), whereas the word "combination" will refer to unspecified pairs (e.g., "the combination [flute-oboe2]").

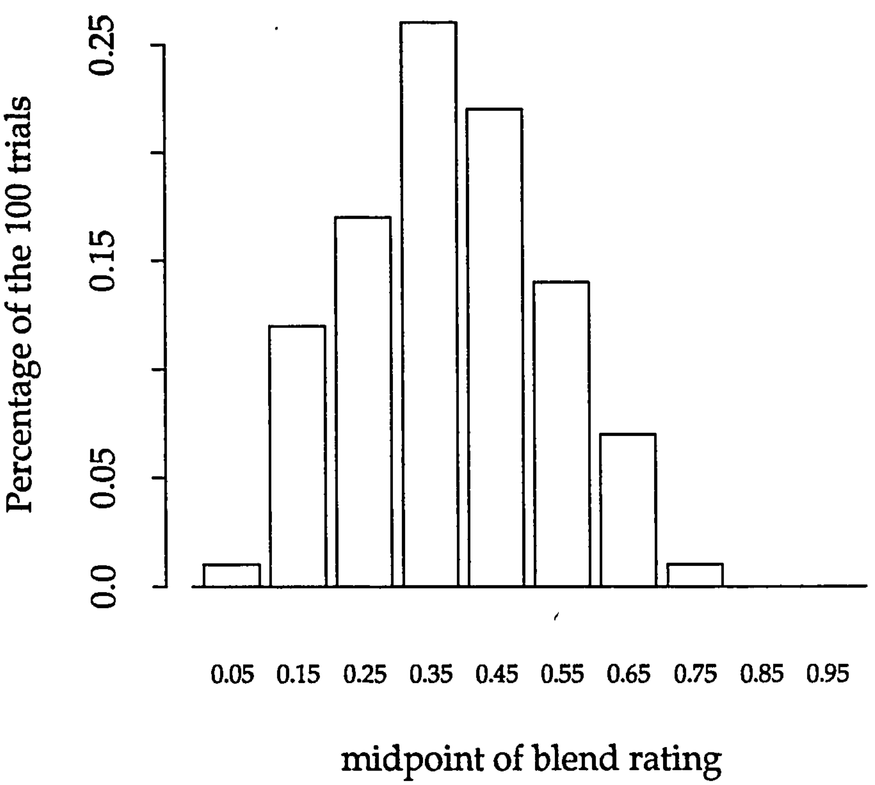


Figure 76. Distribution of blend judgments in Experiment m3Blend. Subject data was parsed into ten categories (0 to .1, .1 to .2, and so on) and the proportion of total ratings in each category calculated.

2. The "inversion" of a pair of instruments refers to the same instruments in a different position; flute-oboe2 is the inverse of oboe2-flute, and vice-versa.

3. The *"note* trials for the *instrument"* (as in "the c# trials for the flute") refer to all ten trials in which that instrument was playing the named note. The "c# trials for the flute," for example, refer to the trials flute-flute, flute-oboe2, flute-englhorn . . . flute-mutecello.

4. The *"factor* difference" (as in "centroid difference") for a pair refers to the subtraction "factor(e) - factor(c#)." The "centroid difference" for the trial flute-oboe2, for example, is 520, since the flute c# centroid is 767 Hz and the oboe2 e centroid is 1287 Hz. The "centroid differences for the flute c# trials" would be a list of the ten differences for the flute c# trials.

5. The names of analysis procedures used in experiment UnisonBlend will be used in the present discussion as well. For example, *"single instrument correlations* for the centroids of the flute's c# trials" would involve correlating the centroids for the instruments flute, oboe2, englhorn . . . mutecello on the note e with the blend rating for the trials flute-flute, flute-oboe2, flute-englhorn . . . flute-mutecello.

Differences between Experiments UnisonBlend and m3Blend

One question of interest was whether the use of a new musical interval introduced new tendencies in the blend judgments. Unfortunately, the study did not include both unisons and minor thirds within the same experiment; if it had, presumably, the magnitudes of blend ratings would be dramatically different. What can be examined, however, is whether the new interval changed the *patterns* of responses.

Figure 77 shows three sets of average blend judgments: two sets for each of the ten instruments corresponding to the c# and e trials of experiment m3Blend, plus equivalent data from experiment UnisonBlend for comparison.[56] Using *t* tests to compare the mean of each instrument's averages with their counterparts in experiment UnisonBlend showed no significant differences ($p < .05$) in any of the ten instruments, either for the c# trials or the e trials; the overall means of the two experiments were also statistically identical. So far, it would appear that the two experiments are fairly similar and that the minor third interval condition did not change the patterns of blend magnitudes. Further evidence for this is found in a

---

[56] An explanation of "equivalent trials" or "counterparts" between Experiments UnisonBlend and m3Blend is necessary here. Because each combination of tones (e.g., [flute-oboe2]) had two representatives in m3Blend (flute-oboe2 and oboe2-flute) but only one in UnisonBlend, the mapping between m3Blend and UnisonBlend is many-to-one. Hence the trials flute-oboe2 and oboe2-flute in Experiment m3Blend share the same counterpart in Unisonblend, the single trial flute-oboe2.

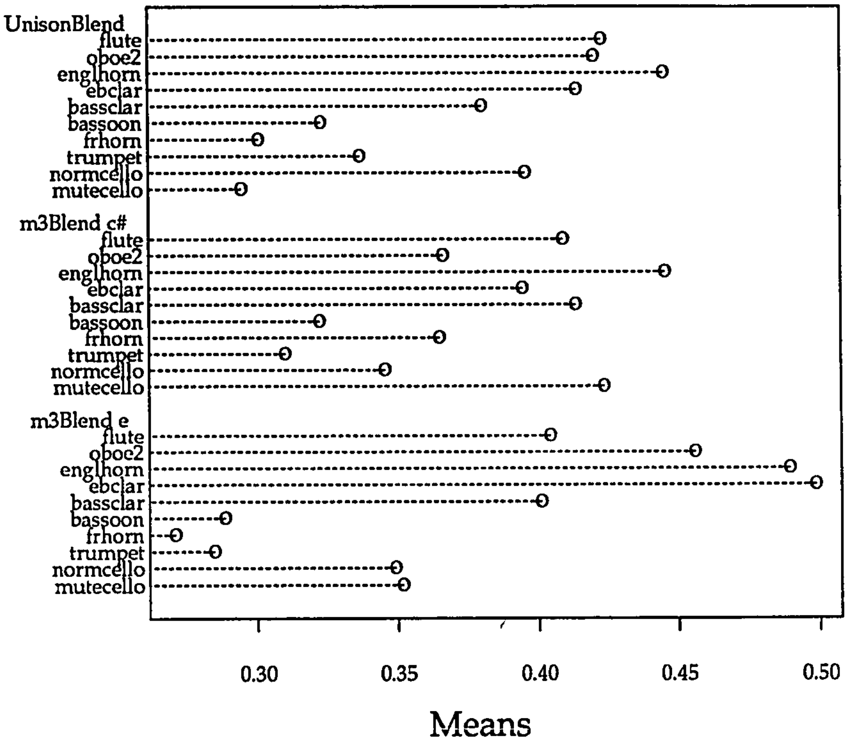


Figure 77. Data averaged across instruments in Experiments UnisonBlend and m3Blend. The UnisonBlend data is reproduced from Figure 62, showing only the 10 instruments used in Experiment m3Blend.

correlation of all 100 trials of experiment m3Blend with their counterparts in UnisonBlend: this results in a fairly strong match ($r = .497$, $p < .0001$).

The match between experiments is somewhat less pervasive when observed in further detail by performing *single instrument correlations* between ratings of the c# and e trials of individual instruments in m3Blend and their counterparts in UnisonBlend. The c# trials for bassclar, bassoon and frhorn show significant correlations with equivalent trials in UnisonBlend; so do the e trials for these instruments. However, for flute and normcello, only the c# trials show significant correlations, and for the instruments oboe2, englhorn, ebclar, trumpet and mutecello, no correlations are found. Thus, at the individual trial level, there is some evidence, albeit small, for a difference between the two intervals of presentation.

Effect of Note Position on Blend Judgments

The next question is whether a difference can be found between inverted combinations; for example, if flute blended differently with the oboe2 depending upon whether it was above or below. A glance back at Figure 77 shows that most of the means for the c# trials of an instrument are fairly close to the means of the e trials of the same instrument. *T* tests show in fact, that the differences between the means are nearly all non-significant ($p < .05$). The one exception to this is in the case of oboe2: its c# trials blend somewhat better ($m = .366$) than its e trials ($m = .456$; $t[18] = 2.232$, $p = 0.039$). This has interesting musical implications, since in orchestral music the oboe tends to play a soloistic, rather than accompanimental role.

Another demonstration of how small an impact position seems to have had is by examining all the differences in blend judgments between inverted combinations to see how large they are. Considering a few individual cases, for example, ebclar-trumpet is rated .312 and trumpet-ebclar was rated .539, a difference of .227. This in fact was the largest difference for all pairs (n = 100). The mean difference for all inverted pairs was less than a single increment of the rating scale, .087, and the standard deviation is .056; and only eight of all the possible inverted combinations had differences larger than a mean and standard deviation. Putting this into perspective, the difference in rating between inversionally-related combinations is rarely greater than 14% of the rating scale (.143), a very small amount. In fact, the maximal difference (.227, the ebclar-trumpet pair mentioned above), is barely 25% of the rating scale. In general, inversionally related pairs received very similar blend ratings.

A second question is whether the *patterns* of blend are similar between c# trials and e trials. *Single instrument correlations* between the two sets of trials for every instrument showed high positive correlations in all ten cases; this is illustrated in Figure 78.

Multidimensional Scaling

MDS was used once again to discover possible acoustical correlates to the blend judgments. A three-dimensional solution is shown in Figure 79. This analysis was made by putting the 100 data points of m3Blend into a 10x10

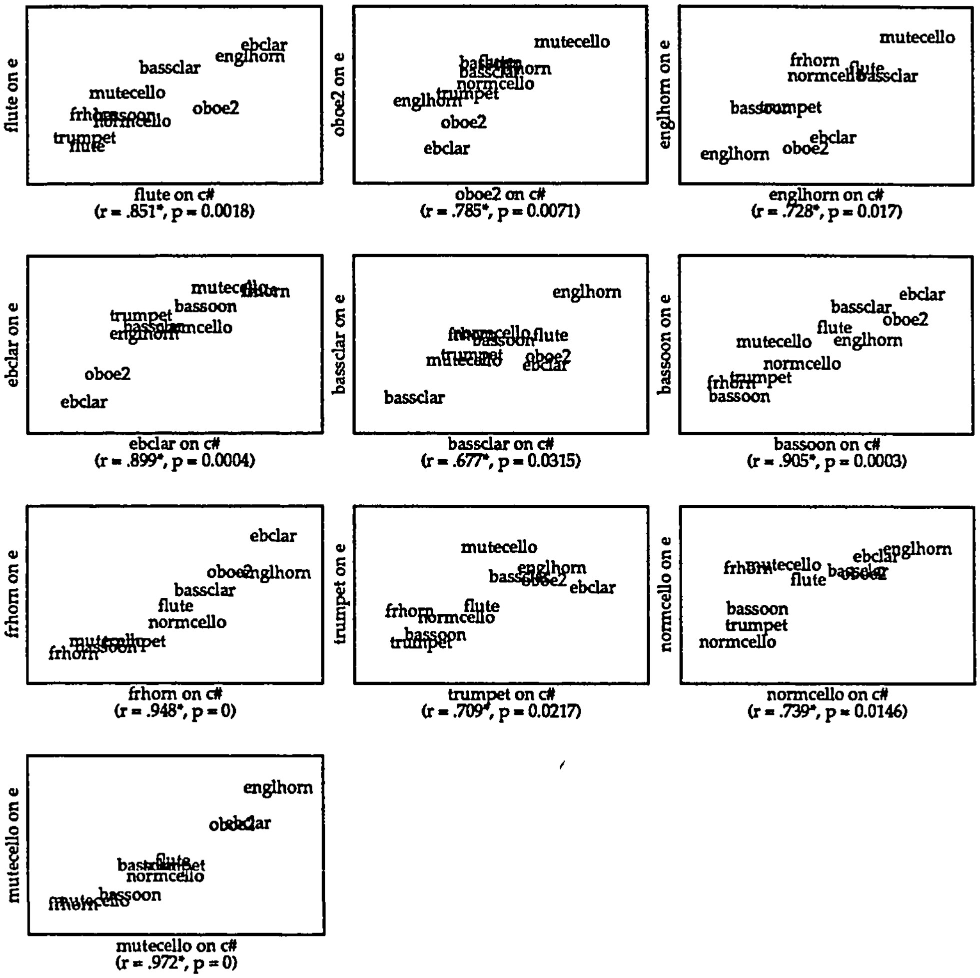


Figure 78. Single instrument correlations of data from Experiment m3Blend comparing c# trials with e trials. In the flute plot (top left), the point marked "oboe2" plots the rating for the trial flute-oboe2 (i.e., flute below) on the x-axis against the rating for the trial oboe2-flute (i.e., flute above) on the y-axis.

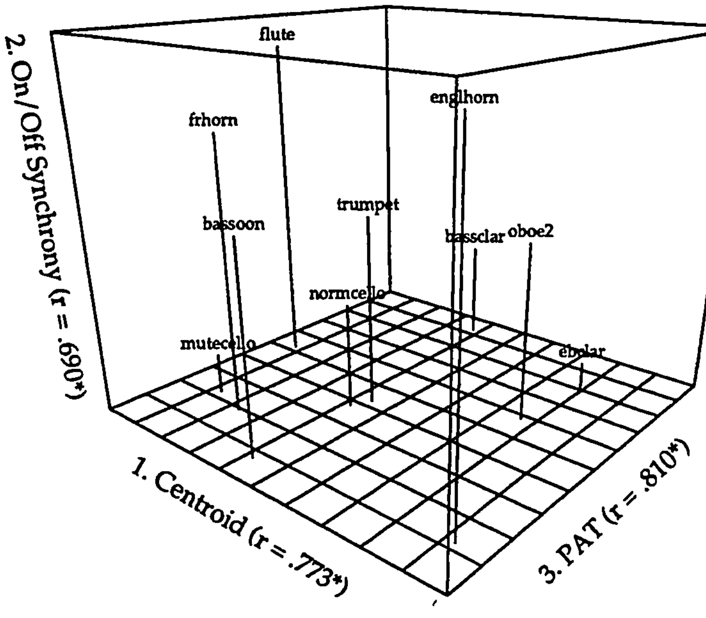


Figure 79. Multidimensional scaling solution of the data from Experiment m3Blend, with blends treated as distances (i.e., blended = close, separated = far). Inversionally-related trials (e.g. flute-oboe2 and oboe2-flute) are conflated into a single distance. Correlating acoustical factors are shown. Dimensions 1 through 3 accounted for 45%, 12% and 11.7% of the variance of the data, respectively.

full-matrix such that inversionally equivalent trials (e.g. flute-oboe2, oboe2-flute) were on opposite sides of the matrix. This means that the MDS solution averaged over the two inversional equivalents for a combination; for example, the single distance shown between flute and oboe2 in Figure 79 accounts for the blend of both the trials flute-oboe2 and oboe2-flute. Previously cited evidence suggests that this generalization results in little loss of information. This solution is somewhat more reliable than the MDS solution for experiment UnisonBlend: dimensions one through three accounted for 68.7% of the variance (45%, 12% and 11.7% for dimensions one through three, respectively).

Correlating acoustical dimensions are marked; the evidence of the effects of centroid, PAT and on/off synchrony are consistent with what was found in experiment UnisonBlend. However, there are even more correlating factors than are shown. Dimension 1 also correlates to recognition ($r = -.717$, $p = .019$), and on/off synchrony ($r = .657$, $p = .039$); and dimension 3 also correlates to precedent noise ($r = .861$, $p = .001$) and acoustic dissonance ($r = .889$, $p = .0006$). So the dimensions do not seem to sharply define any one acoustical phenomenon.

Factors Correlating with Blend Judgments

Average instrument correlations were performed separately on the c# trials and the e trials. That is, the data in Figure 77 for the 10 c# trials and the 10 e trials were each correlated with the eleven factors for each instrument which have been considered so far. This analysis makes it possible to see if "blending power," the phenomenon observed in experiment UnisonBlend, was operative here. The results were very weak. The only significant correlation was with centroid on the e trials: blend decreased as the centroid got higher. It appears that the phenomenon of "blending power" only weakly applies to this data. It is noted, additionally, that standard deviations over the averaged instruments are much higher than in experiment UnisonBlend (see Figure 80); larger standard deviations means less distinctly high or low values, which explains why "blending power" is not observed.

Sums vs. Absolute Differences

It will be recalled that many of the *single instrument correlations* in experiment UnisonBlend were considered from two different perspectives, depending on whether stronger evidence was found for *sum* correlations or *absolute difference* correlations (*model 1* vs. *model 2*). The equivalent analyses for the present experiment are listed in Table 8.[57]

Centroid

The table shows a large number of correlations by absolute differences and fewer by sums. Using the two-model approach once again, this finding supports *model 2* (that centroid closeness determined blend), as opposed to

---

[57] In contrast to the way *single instrument correlations* were presented in Experiment UnisonBlend, here they are listed only by the number of instruments whose correlations were significant, and by the sign of the correlation.

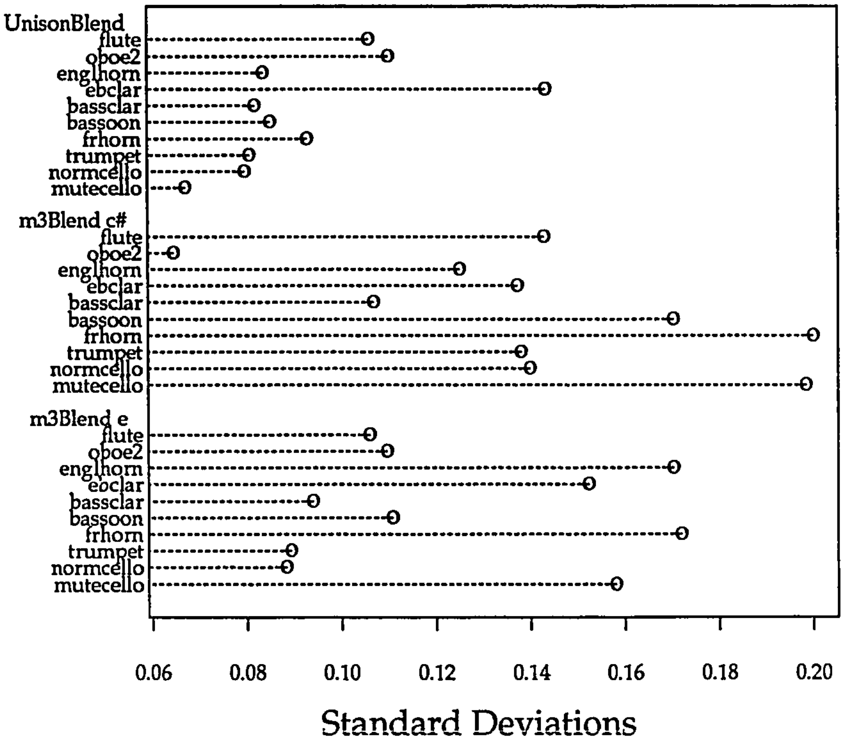


Figure 80. The standard deviations for instrument averages in Experiment m3Blend. The equivalent data from Experiment UnisonBlend is shown for comparison.

	sums					**absdiffs**		
Factor	overall corrs	number of c# SICs		number of e SICs		overall corrs	number of c# SICs	number of e SICs
		pos	neg	pos	neg			
Centroid	r = .268 *p = .007*	4	2	3	2	r = .594 *p < .0001*	6	
PATs	(n.s.)			1	1	r = .306 *p = .002*	(n.s.)	2
ampscales	r = .237 *p = .018*					(n.s.)	(n.s.)	
harmonicity	r = -.267 *p = .007*					(n.s.)	(n.s.)	
noisedur	r = .242 *p = .015*	1		1	2	r = .436 *p < .0001*	2	5
pitchdevs	r = .291 *p = .003*			1		(n.s.)		1
recognition	r = -.286 *p = .004*	1	3	1		r = .358 *p = .0003*	2	1
	Composite Correlations							
Centenvs	r = -.425 *p < .0001*		1		4			
Ampenvs	r = -.208 *p = .038*							
Pitchenvs	(n.s.)				1			
dissonance	(n.s.)				4			
harmsyncs	r = -.268 *p = .007*		3		2			
onoffsyncs	(n.s.)		2		1			
peaksyncs	(n.s.)		1		1			

Table 8. Various *overall correlations* from Experiment m3Blend, with counts of significant *single instrument correlations* (SICs).

model 1 (that overall centroid height determined blend). This is in interesting contrast to the findings of experiment Unisonblend, which showed more support for *model 1* and less for *model 2*. It would appear that the increased pitch distance changed the role of centroid in determining blend.

Table 8 shows that sum correlations exhibited both positive *and* negative outcomes. Figure 81 shows the *single instrument correlation* for the frhorn e trials, exhibiting decreasing blend with each increase in centroid for the instrument on c# (positive correlation). These results are consistent with those found in Experiment UnisonBlend. In contrast, Figure 82, showing the *single instrument correlation* for the ebclar c# trials, exhibits a negative correlation: decreasing blend with each increase in centroid for the instrument on c# (positive correlation). In general, positive correlations are found for the darker instruments, and negative correlations for the brighter instruments. This finding provides further illustration of the degree to which centroid closeness (*model 2*) was operative throughout Experiment m3Blend. That is, for a dark instrument such as frhorn, since its centroid is at "bottom," all other instruments are brighter than it; since blend decreases with increasing centroid difference, the brighter instruments blend poorer than the darker ones. For a bright instrument such as ebclar, since its centroid is at "ceiling," all other instruments are darker than it; in contrast, the brighter instruments blend *better* than the darker instruments.

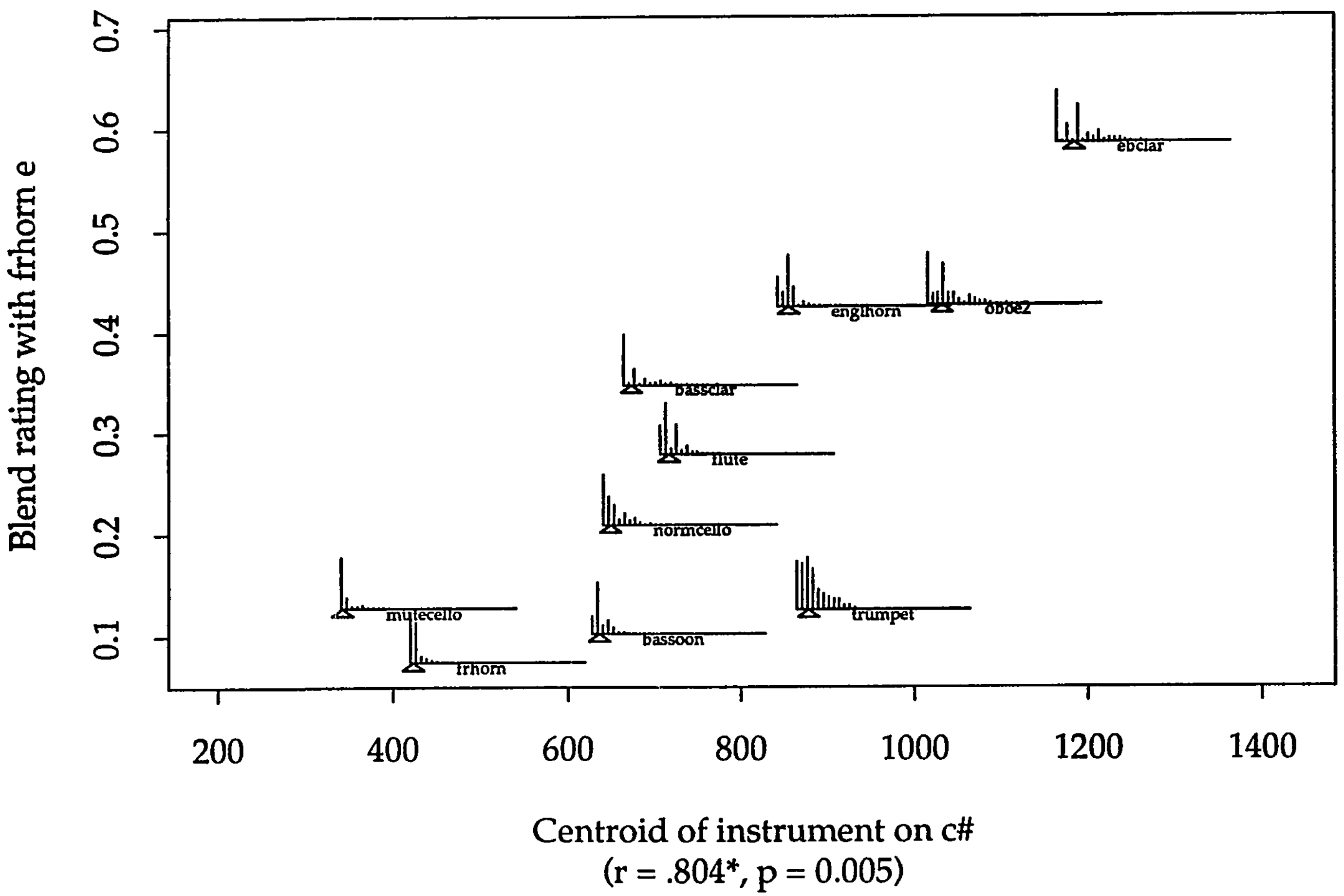


Figure 81. *Single instrument correlation* for the centroids of frhorn's e trials, Experiment m3Blend.

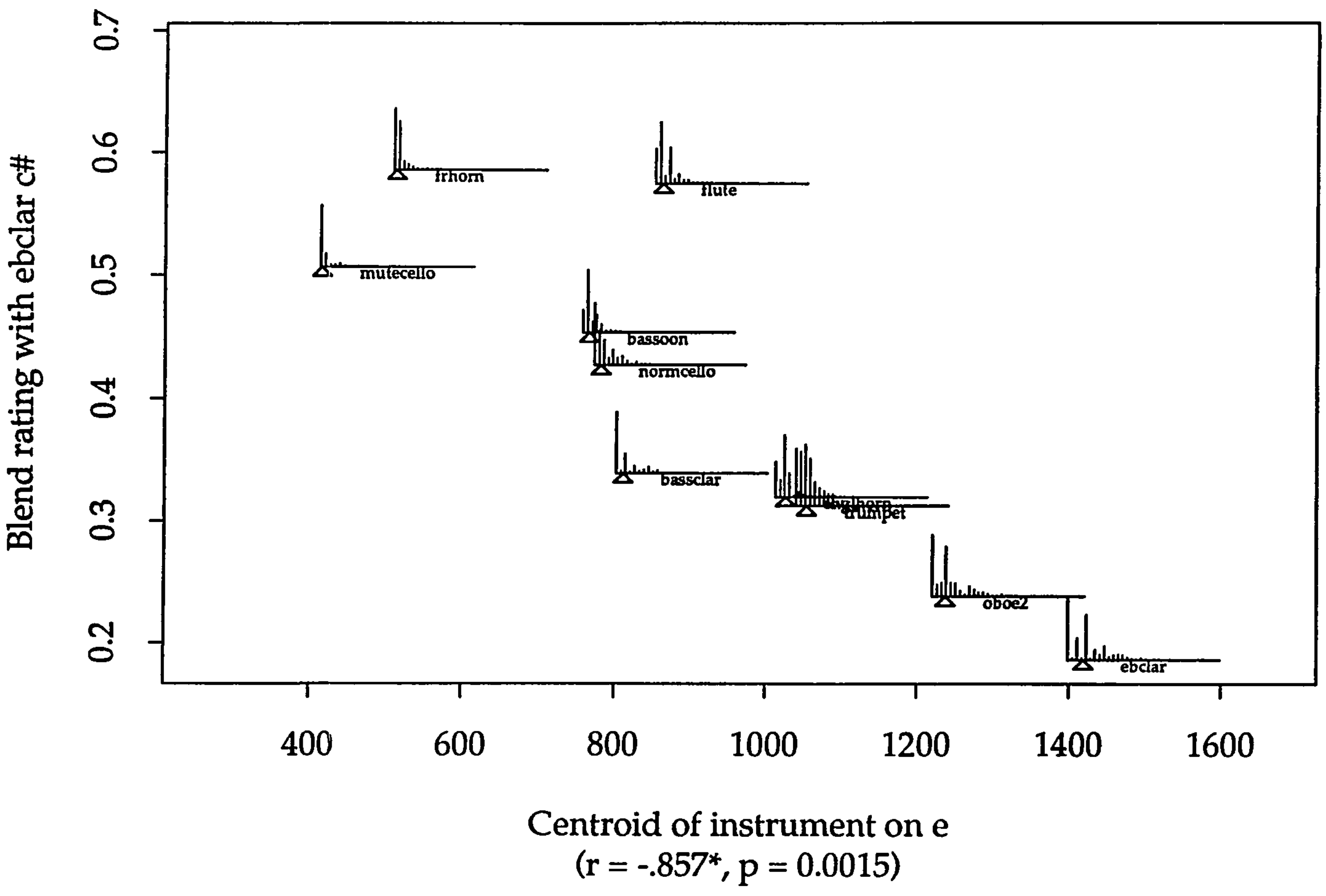


Figure 82. *Single instrument correlation* for the centroids of ebclar's c# trials, Experiment m3Blend.

Thus, as demonstrated by numerous instrument pairs, smaller centroid widths lead to greater blend. This suggests a powerful rule of thumb for orchestration; the question is, is the principle as general as it appears? One challenge that can be offered is the question, will the principle apply to changes in centroid width where the instruments *stay the same*? For example, if it were possible to make the oboe's centroid closer to that of the flute, would its blend with the flute improve? This would corroborate that the trend was truly related to centroid and not merely the byproduct of other acoustic features accompanying the change in instrument identity.

The opportunity for observing this situation will be provided in the following experiment. For now, however, something similar to it occurs in the trials for m3Blend. Changes in centroid width automatically accompany inverted combinations: for example, oboe2-flute, when inverted to flute-oboe2, narrows its centroid width. The blend judgments for these trials may be compared. Table 9 shows the proportion of instrument combinations whose blend was increased or decreased due to increased or decreased centroid widths across these two experiments. The tendency shown in all of these transformations confirms the hypothesis: increased blend with decreased centroid distance, and decreased blend for increased centroid distance. The table also offers an opportunity for testing the hypothesis by comparing these two trials to the single flute-oboe2 trial in experiment

Trials Compared	Change in Centroid Width	Number of Cases...		
		total	blend improves	blend worsens
Inversionally-related trials in m3Blend	Increased	12	4	8
	Decreased	33	27	4
Equivalent Trials between UnisonBlend and m3Blend	Increased	70	33	37
	Decreased	40	25	15

Table 9. Effects of change in centroid widths for a pair of tones on blend, comparing across Experiments m3Blend and UnisonBlend.

UnisonBlend, which has yet another centroid width. This too shows increased blend with decreased centroid difference.

Other Factors

Table 8 shows that the "closeness principle" (*Model 2*) also applies to DPAT, precedent noise, and recognition: when two instruments are close in value for these factors, they blend better. For DPAT, absolute difference correlations have completely taken over from sum correlations (although the magnitude is in fact lower than it was in UnisonBlend). For both precedent noise and recognition, where sum correlations had been of greater magnitude than absolute difference correlations in experiment UnisonBlend (see Table 7), the relative magnitudes have been reversed, as was observed with centroid earlier. The presence of both positive and negative *single instrument correlations* are found in the sum correlations for all three of these factors, presumably for the same reason that was discussed above with respect to centroid.

Pitch deviations, however, still show only sum correlations, albeit weakly. Two new factors appeared, amplitude scaling and harmonicity, both with sum correlations only, and both weak in magnitude. The sign of the correlations suggests that blend decreased for pairs that were of greater overall amplitude, and increased for combinations that had greater overall harmonicity. Acoustic dissonance, curiously, has largely vanished as a factor influencing blend judgments.

Temporal Factors

Both centroid and amplitude envelope correlation appears to have played a role in determining blend, while pitch envelopes played no role (see Table 8); all of this is consistent with the outcomes of experiment UnisonBlend. On/off synchrony has also vanished; although another synchrony factor replaces it (harmonic envelope synchrony) the sign of the correlation now suggests that greater asynchrony produces better blend. However, the magnitude of the correlation and the absence of supporting evidence calls this rather unintuitive finding into question.

Pitch Position and Negative vs. Positive Differences in Factors

In experiment UnisonBlend there was nothing intrinsic in the *presentation* of pairs which differentiated the individual tones from one another; if tones differed from one another, it was due to *acoustical* properties of the tones themselves. In experiment m3Blend, however, the tones differed in pitch. It is of interest to see if the difference in pitch between a pair of instruments somehow interacted with the difference in their acoustical factors.

One such interaction which was anticipated to have some impact was "pitch-centroid reversal": cases where the centroid of the lower note was below the upper one, and vice versa. Earlier evidence suggested that note position had little affect on the blend judgments for a given combination, so it is unlikely that large reversal effects are likely to be found; however, even small traces of such effects for any of the factors may be worth noting. One way to examine this is to compare the difference in the acoustical factors for one instrument pair with the difference in acoustical factors for the inversion of that pair, and compare this change to the change in blend ratings. To simplify the process, only positive and negative differences are distinguished. For example, in the pair flute-oboe2, the centroid difference is positive, while for oboe2-flute the difference is negative (meaning that the lower voice was brighter than the upper voice); the blend rating for these pairs is .538 and .379, respectively, showing a better blend for a negative centroid difference than a positive one. If this pattern were observed over a large number of pairs, one might hypothesize that listeners found blend to be better when the lower note was brighter than the upper note.

To discover if any such patterns really occurred, *t* tests were used to compare the mean blend judgments for all trials which obtained a positive difference for the factor in question with those that obtained a negative difference. Table 10 shows the results of this analysis, showing significant results ($p < .05$) for the factors of harmonicity, pitch deviation and recognizability. For example, the average rating for all trials in which the amount of harmonicity in the instrument playing e was *greater* (positive difference) than the harmonicity of the instrument playing c# was .351, while the average rating for trials exhibiting the reverse case (negative difference) was .413. *T tests* showed these means to be significantly different from one another ($t[98] = 2.131$, $p = .036$), meaning that blend was better for positive

Factor	Overall *T* Tests			number of signif. single instrument *t* tests	Conclusion
	Means		p		
	Pos. Diff.	Neg. Diff.			
Harmonicity	.351	.413	.036	2	Blend better when upper note is more harmonic
Pitch. Dev.	.351	.413	.035	4	Blend better when upper note has more pitch deviation
Recognition	.341	.426	.003	5	Blend better when upper note is more recognizable

Table 10. Comparing the mean blend judgments for pairs in which a factor had a positive difference with pairs with a negative difference, Experiment m3Blend. Factors showing significant differences ($p < .05$) are shown. Overall *t* tests involve the entire dataset (n = 100), whereas single instrument *t* tests involve only the trials for a given instrument (n = 10).

differences in harmonicity than negative differences. In addition, applying the analysis to subsets of the data (in a way analogous to *single instrument correlations*) yielded significant differences in two out of the ten instruments. In other two factors as well (pitch deviations and recognition), the negative differences led to poorer blends than the positive differences. In other words, blend is better for a combination when harmonicity, amount of pitch deviation and recognizability is greater for the upper note than the lower note.

Modelling Blend with Minor Thirds

Following the example of the previous experiment, the 13 *overall correlations* found to be significant (see Table 8) were submitted to a regression analysis in order to evaluate their relative importance. The result yielded the following regression equation, which had a correlation of $r = .858$, $p < .0001$:

blend = .397 cents_absdiff + -267 noisedur_sums + .234 pitchdevs2_sums
+ -.210 centenvcorrs + .167 ampscales_sums + .155 cents_sums +
-.091 harmcity_sums + .087 recognition_absdiff +
-.069 recognition_sums + .044 noise_absdiff + .037 harmsyncs +
-.009 PATs_absdiff + -.009 ampenvcorrs + .162

To observe the improvement of the model with each successive factor, a succession of regression analyses (analogous to the procedure seen in Figure 72) is shown in Figure 83. Once again, the final (13th) regression has been enlargened and shown in Figure 84 reveal the "outliers."

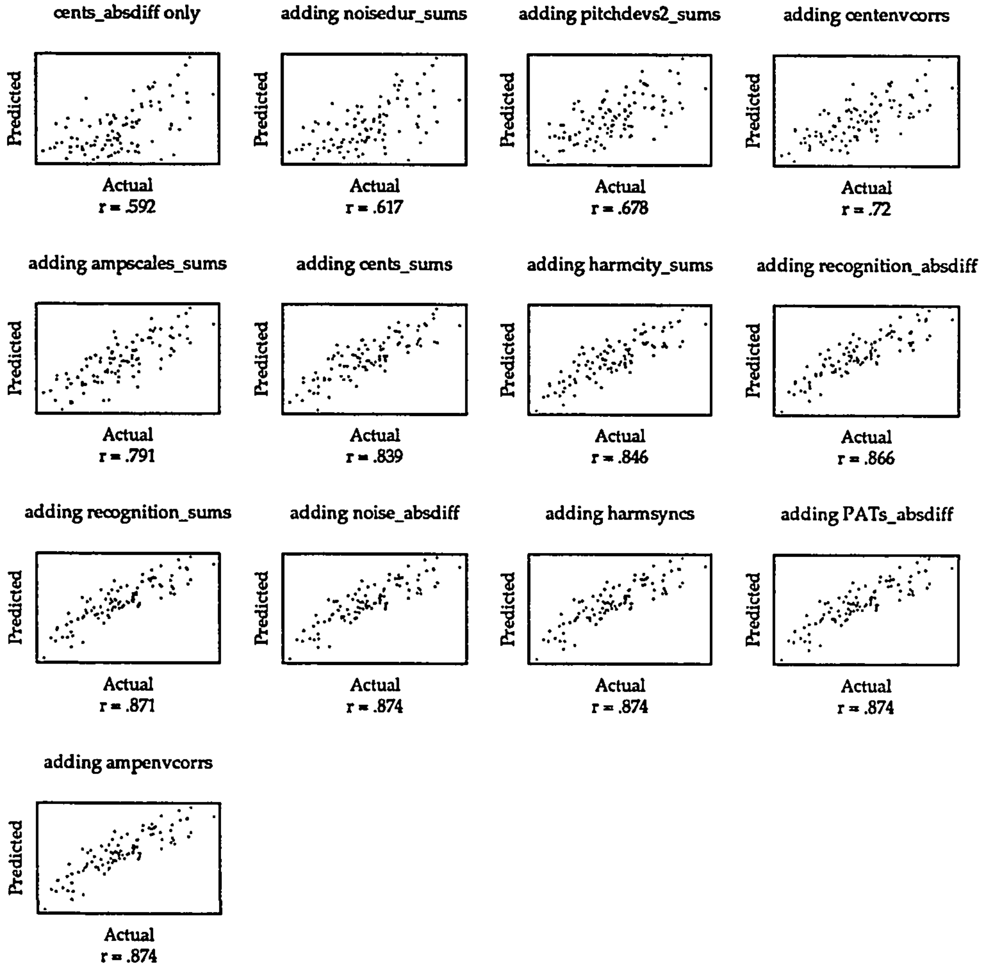


Figure 83. Scatterplots showing improvement of regression analysis to predict blend judgments from Experiment m3Blend as more acoustical factors are added.

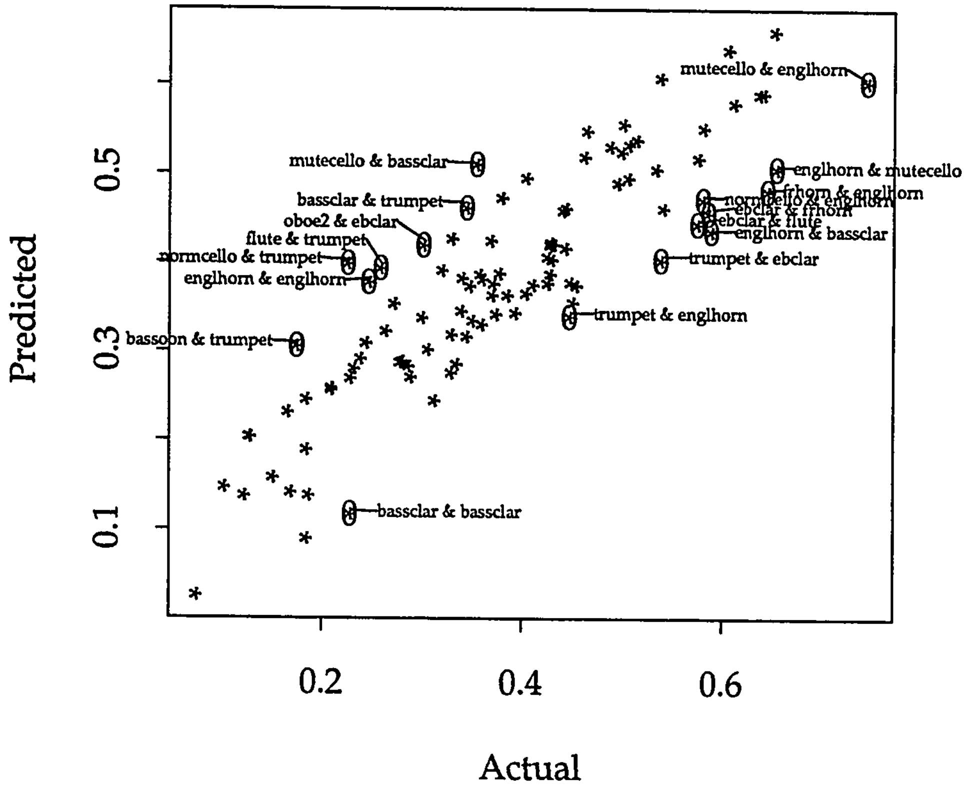


Figure 84. The scatterplot for the 13th regression shown in Figure 83, comparing the values produced by the regression equation with the actual blend judgments. The outliers (those that varied more than a standard deviation from the mean error) are labelled by their associated trials.

In experiment UnisonBlend the regression equation exhibited an inability to account for the blend of pairs involving single reed instruments. This is not repeated here to as great a degree. In place of this, however, is a general failure to correctly account for trials involving trumpet or englhorn: 32% of the trials involving trumpet and 26% of the trials including englhorn are found as outliers here. As before, the outliers are more or less evenly distributed between negative and positive errors, so one cannot simply conclude that the model simply underestimated or overestimated blend for either instrument.

To narrow the number of factors included in the model, the evidence so far can be compared and certain redundant, weak, or over-correlated factors can be eliminated, as was done in experiment UnisonBlend. The details of this reduction will be skipped, but using similar to the previous reduction criteria a model based on centroid envelope correlation, centroid differences, pitch deviations sums and absolute differences in recognition is suggested. The correlation is $r = .764$, $p < .0001$, and the regression equation is:

$$\begin{gathered}\text{blend} = .368\ \text{centdiff} + .165\ \text{pitchdevs\_sum} + -.130\ \text{centenv} + \\ .140\ \text{recognition\_diff} + .214\end{gathered}$$

Summary of Findings, Experiment m3Blend

For the most there were no dramatic differences in the patterns of blend ratings in experiment m3Blend and equivalent trials of UnisonBlend, although the variability was higher (see Figure 80). As a result, there was no evidence of a general "blending power" for specific instruments as before.

In comparison to experiment UnisonBlend, however, correlations based on the differences between the acoustical features of tones played a more important role. The most prominent case was centroid: the blend among a pair of tones was best when the centroids were close in value. While the overall height in centroid was a factor somewhat as well, the effect was not as great as in experiment UnisonBlend.

A reduction in the influence of DPAT (no sum correlations, smaller absolute difference correlations) suggests that attack factors played less of a role in the minor thirds condition. It would seem fitting that in a condition where detection of the presence of one vs. two notes was reduced to triviality, attack cues would not play as important a role in the judgments. Temporal factors in general still appear to be important, on the other hand: a heightened role in centroid envelope correlation was found.

It was anticipated that effects due to the relationship between pitch level and centroid for a pair of tones would be found; for example, that pairs in which the lower of two notes had the higher centroid would blend more poorly. In reality, "pitch-centroid reversal" was of little relevance to blend: because note-position had little impact, there was little evidence for this phenomenon. In fact, there were few indications of any difference in blend for combinations in different positions (e.g., flute-ebclar did not receive a markedly different blend rating than ebclar-flute). There was one exception:

when comparing all the trials in which oboe2 was present, blends were in general better when oboe2 was on the bottom (c#) than when on the top (e). This confirms a musical intuition about the soloistic nature of the oboe: when set against a body of strings, for example, the oboe will stand out distinctly.

A regression formula with five factors predicted blend moderately well. These factors (in order of descending magnitude) were centroid absolute difference, sum of pitch deviations, amount of centroid envelope correlation, and recognition absolute difference. Compared to the model used in UnisonBlend (the smaller of the two), the factors of centroid sum, acoustic dissonance and DPAT absolute difference are gone, while centroid envelope correlation and recognition absolute difference have been added.

Experiment 3 (AdjustBlend)

Preliminaries

Manipulating Centroid

Two experiments so far have shown aspects of centroid to be a primary factor in the judgments of blend. It has been observed that for minor third intervals, instruments closer in centroid blend better than those that are farther apart, and for unison intervals, blend improves as either instrument gets darker. It is the fact that the Stanford tones represent a more or less continuous range of centroid categories that makes it possible to draw such conclusions. To confirm them, however, some evidence is needed to show that when centroid *alone* changes, the blend increases or decreases according to the expected pattern. So far changes of centroid have been the fortuitous byproduct of an instrument identity change, or due to a change in pitch position for a pair. Unfortunately, however, in such cases one cannot rule out that the resulting blend judgments might have been due to other acoustical factors accompanying the instrument change (or pitch position).

A method for artificially changing the centroids of Stanford tones was explored in order to create an experiment in which instruments could appear at different centroid levels. The goal was to change in some way the proportion of energy in low vs. high harmonics while not distorting the patterns of energy among harmonics any more than necessary. A low-order digital filtering solution was explored but rejected because the results had an unnatural, "electronic" sound to them. The preferred choice was modifying the slope of the harmonics to favor the upper or lower harmonics as needed by altering the synthesis parameters.

Modifications in slope were made by applying a rate of attenuation or amplification on the steady-state spectra based on peak amplitudes of harmonics. An attenuation slope of 3 dB (power) per octave reduced the linear amplitude of the first, second, fourth and eight harmonics (i.e., octave-related harmonics) by approximately 0%, 50%, 75% and 87.5%, respectively, with the process continuing as necessary for all harmonics. Amplification slopes were obtained simply by magnifying the harmonics by the same values. Harmonics in between these octave multiples were attenuated or amplified by exponential interpolation between the new amplitudes of the flanking octaves.

The main advantage of using slopes to change the spectra is that, since the relative amplitudes of partials are largely maintained, the sound of the tones maintain their instrumental identity; also, with this method the increments of brightness sounded more or less similar across instrument. That is, a 3 dB per octave brightening of the bassoon sounded much like the same transformation of the frhorn, relative to their respective "normal" versions.

Note, however, that this method did not produce identical centroids at each level for each instrument: the dark versions of oboe and ebclar, for example, are brighter than the dark versions of frhorn and bassoon. At first, an attempt was made to select absolute centroid values at which each instrument would appear; this turned out to be unfeasible due to the difference in number of harmonics in instruments. A nine-harmonic bassoon tone cannot be practically made as bright as a 21-harmonic oboe with a 3 dB positive slope to its harmonics; indeed, only by a gross amplification of frhorn's upper harmonics could it approach the brightness of even the *normal* oboe2. So the idea of using a fixed set of slopes for all instruments was pursued: 0 dB (normal condition), -3 dB ("dark"), and +6 dB ("bright") per octave.

Figure 85 shows the attenuation and amplification slope changes applied to four of the Stanford tones. The tones were then regenerated simply by scaling all the amplitude envelopes of the tones according to the new peak values shown in these spectra. The centroids produced by each of the three slopes, for each of the two notes c# and e, are shown in Figure 86. To illustrate the nature of the changes, time-amplitude-harmonic plots (equivalent to those of Figures 9 through 23) for the dark, normal and bright conditions of the flute are shown in Figure 87.

Manipulating Other Factors

It will be recalled that certain factors in experiments UnisonBlend and m3Blend showed far less of an effect on blend judgments than did centroid (e.g., harmonicity, amplitude, harmonic synchrony). This does not necessarily mean that these factors are irrelevant for the perception of blend; it is possible, rather, that the range of variation for these factors in the Stanford tones was simply insufficient. Thus, it was of interest to find if there was a way to expand the effective range of variation in harmonicity, amplitude, and harmonic synchrony, and observe the effects on blend judgments. Each of the factors was considered as a candidate.

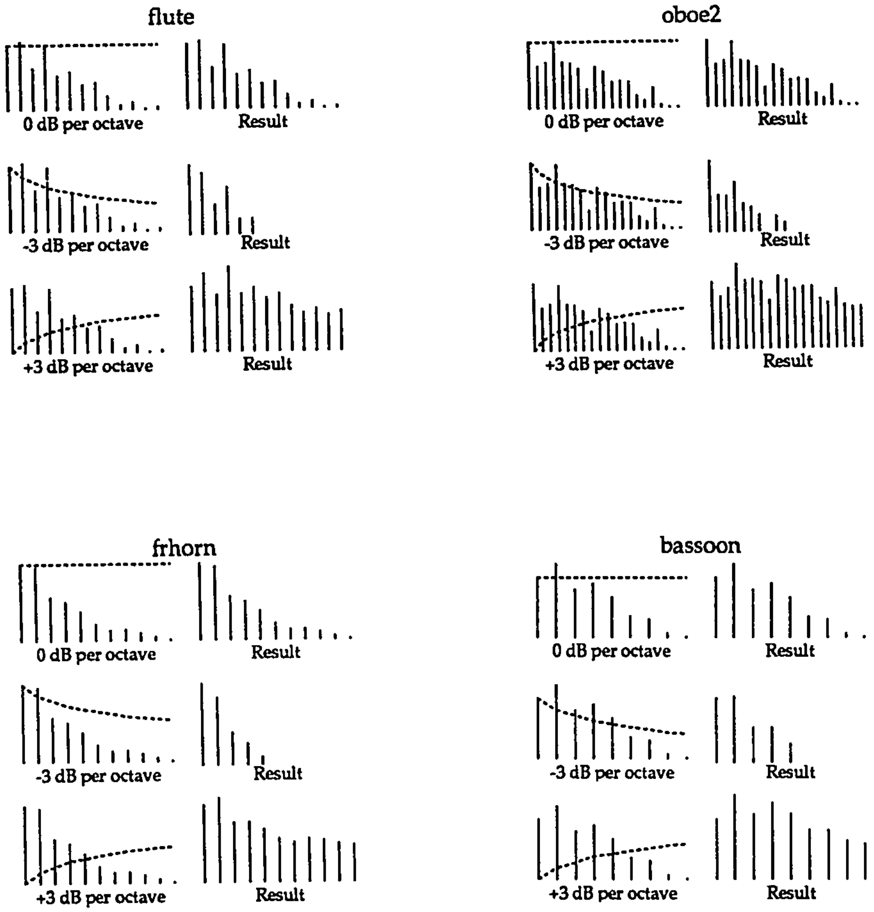


Figure 85. The process of changing centroids of Stanford tones by changing the slope of attenuation of the harmonics. Each spectra is based on peak amplitudes, and the amplitude scale is in dB power. The dotted lines indicate the amount of energy added or subtracted to the partials for each spectra.

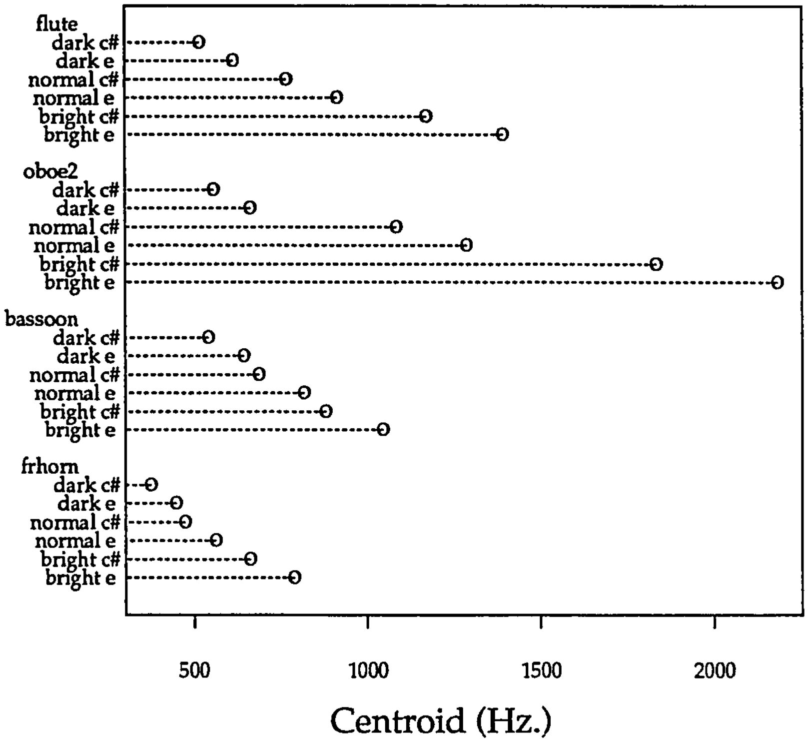


Figure 86. Centroids for the instruments flute, oboe2, bassoon and frhorn in normal, artificially dark and artificially bright conditions.

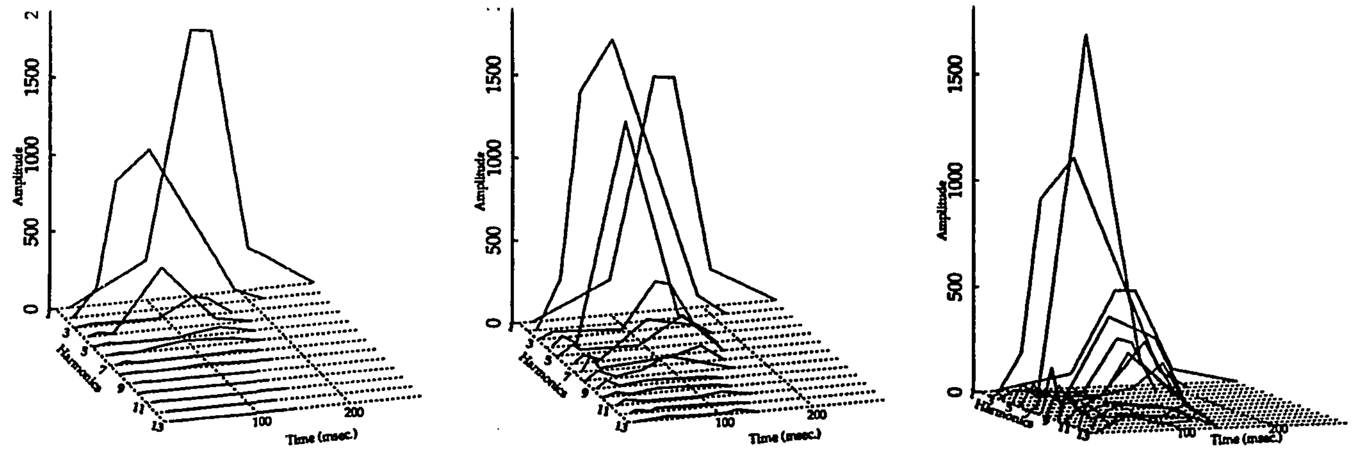


Figure 87. Time-amplitude-harmonic plots for the dark, normal and bright conditions of the flute in Experiment AdjustBlend.

Amplitude

Ensemble performers, when trying to obtain a blend among themselves, often adjust their dynamics to match one another in loudness, usually with the louder instrument matching to the softer one (Scherchen, 1933, p. 69-70). This suggests that manipulating amplitude and observing the changes in blend would be an interesting avenue of pursuit. Another reason for including this condition rests on more perceptual grounds. It is assumed that as one instrument is attenuated in level, its identity contributes less to the combination, thus the overall sound of the pair is dominated by the non-adjusted instrument. The paradigm of "segregation by identification" is premised on the idea that hearing *multiple* identities results in the sensation of non-blending. In an amplitude change in one tone resulted in the pair having one overall timbre rather than two competing ones, this would promote their blend. To this end, versions of the Stanford tones were generated at intensities 3 dB and 6 dB (power) down from the level of the normal tones, leading to a total of three amplitude levels for each tone.

A remark concerning the inclusion of both amplitude and centroid conditions is warranted here. In real acoustic conditions, the two factors are rarely orthogonal. Changes in dynamic level are often accompanied by changes in brightness (namely, brighter spectra corresponding to increased effort in performance); similarly, an effort made on the part of a performer to change the brightness of his sound often is effected by a change in the dynamic level (the study by Goodwin, 1989, examined in Chapter 2 presented compelling evidence for this in choral singing). Examining the two conditions independently, however, may be a way to measure their relative contribution to the perception of blend.

Harmonicity

The other factor that was chosen to manipulate was harmonicity. Harmonicity is a factor not typically under the control of the performer of a conventional orchestral instrument. However, this hardly eliminates it from musical interest, because composers of computer generated music often exploit this dimension. Furthermore, one would expect that harmonicity would play a large role in blend judgments: for example, it was suggested in Chapter 2 that inharmonicity promotes blend by confusing the segregating influence of the pitch perception mechanism.

The inharmonicity of the Stanford tones was manipulated by exaggerating the degree to which the frequency envelopes departed from their "correct" target frequencies. The normal variations from strict harmonic relationships can be observed in Figures 24 through 38. For example, for a first harmonic which overshoots its target by 15 Hz before settling on 311 Hz, a 100% positive exaggeration would change the overshoot to 30 Hz. The opposite transformation (removing all inharmonicity) would be to remove all overshoots and undershoots and make the frequency envelopes entirely static. Using these methods, three harmonicity levels were produced for each tone: harmonic (all inharmonicity removed), normal and inharmonic (100% exaggeration of natural inharmonicity). To illustrate the nature of the changes, time-frequency-harmonic plots (equivalent to those of Figures 24 through 38) for the harmonic, normal and inharmonic conditions of the flute are shown in Figure 88.

Thus, each of the three conditions (centroid, amplitude and inharmonicity) had three levels. For the centroid condition, the three levels will be called "dark," "normal" and "bright." For amplitude the three levels will be called "-6 dB," "-3 dB," and "normal" (-0 dB). For the harmonicity condition, the three levels will be called "harmonic" (all harmonics tuned to perfect ratios throughout), "normal" and "inharmonic" (all frequency overshoots and undershoots exaggerated 100%).

Experimental Conditions

In order to gauge the effects of these changes, the goal was to compare the change in blend for the same pair when one of the two tones varied in the various levels of the condition in question. Furthermore, (1) this test would be applied separately to the c# instrument and to the e instrument, and (2) the same suite of tests would be applied to the combination in inversion. In considering an experimental design based on the number of variables are suggested so far, one must beware of a combinatorial explosion which would lead to an experiment of impractical length; thus, some kind of economizing is necessary. The following limitations were adopted:

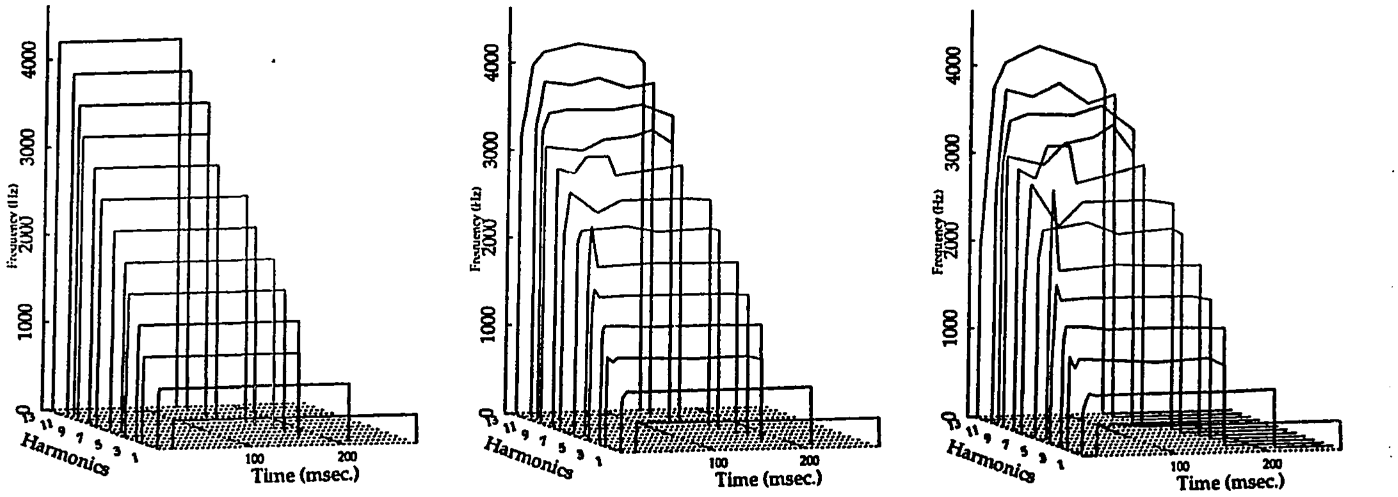


Figure 88. Time-frequency-harmonic plots for the harmonic, normal and inharmonic conditions of the flute in Experiment AdjustBlend.

(1) Combined factors will not be explored; only pairs in which one note is normal and the other is in one of the three conditions will be included.

(2) Only four instruments will be explored: the four instruments shown in Figure 85 were chosen because they represent both dark and bright centroids, and come from different instrumental families.

(3) Same-instrument trials (e.g., flute-flute) are omitted.

Given these restrictions, the combination [flute-oboe2] would produce all the trials shown in Table 11. However, 10 of these 36 trials can be eliminated since they are the same "normal-normal" pair over and over again, which can be represented by a single trial. Thus, for every combination of instruments, 26 trials are generated.

Stimuli

The instruments flute, oboe2, bassoon and frhorn were selected for investigation because they represented a variety of centroid and attack types. Combining them into all the combinations described above produced 156 trials. The amplitudes of tones in the normal, centroid and harmonicity conditions were first transformed according to their condition and then equalized in amplitude to an RMS value of .15, as in experiment m3Blend.

centroid		**amplitude**		**harmonicity**	
flute c#	**oboe2 e**	**flute c#**	**oboe2 e**	**flute c#**	**oboe2 e**
Normal	dark	Normal	-6 dB	Normal	harmonic
	normal		-3 dB		normal
	bright		normal		inharm.
dark	Normal	-6 dB	Normal	harmonic	Normal
normal		-3 dB		normal	
bright		normal		inharm.	
oboe2 c#	**flute e**	**oboe2 c#**	**flute e**	**oboe2 c#**	**flute e**
Normal	dark	Normal	-6 dB	Normal	harmonic
	normal		-3 dB		normal
	bright		normal		inharm.
dark	Normal	-6 dB	Normal	harmonic	Normal
normal		-3 dB		normal	
bright		normal		inharm.	

Table 11. All the trials for the combination [flute-oboe2] in Experiment AdjustBlend.

The tones in the -3dB and -6dB conditions were first equalized to the RMS value of .15 and then scaled down in amplitude accordingly.

Subjects

Twelve listeners participated in the study, who had similar backgrounds to the listeners that had participated in the two previous studies. The initials of the twelve subjects were BLW, GJS, GSK, TWM, MDS, EJB, TJP, ELI, MDW, KWS and DRC. Four of these listeners had participated in m3Blend, and four had participated in UnisonBlend. None were paid.

Equipment

The room, playback equipment, level calibration method and rating method were the same as in the two previous experiments.

Procedure

The instructions were identical to those of experiment m3Blend. As before, the experimenter played the four Stanford tones individually over the loudspeakers while the listener followed a list of the names of the instruments in the order that they were sounding.

Preceding the main experiment was a preliminary experiment, consisting of only the "normal-normal" conditions of the 156 trials (12 trials total). The purpose of this experiment was to see if these trials received the same kind of blend judgments as in the succeeding full experiment, or if the context of the altered instruments changed the responses. This block of 12 trials was run four times for each listener.

Each listener then ran the complete block of 156 trials a total of four times, and each heard the trials in a different random order in each block. The total duration of the experiment, including instruction, was typically 1 3/4 hours, and was (in most cases) completed in separate sessions on different days. At the beginning of each new session, 50 randomly-selected practice trials preceded the actual experiment.

Results of Experiment AdjustBlend

Variability of Responses

Performance on the main experiment was consistent. All listeners showed consistent performance across their four blocks of trials with the exception of ELI who showed two non-significant correlations among the six comparisons. Subject agreement (based on averaged datasets) was also good except for subject TWM, who showed non-significant correlations in three of the 11 comparisons, and lower correlations overall. The datasets for ELI and TWM were discarded, and a dataset based on the average of all the remaining subjects' datasets is used for the remainder of the study.

Performance on the preliminary experiment was less reliable. Three of the remaining ten listeners had significant correlations in only half or less than half of the comparisons. However, unlike the previous two experiments, no warmup trials preceded this preliminary portion. The lack of consistency while they were still becoming familiar with the task is not surprising. It was observed that listeners showed increased consistency with successive blocks. Between-subject correlations on the preliminary experiment was very good, with the exception of one listener (DRC).

Despite the different listeners and the changed context, the general patterns of blend judgments were maintained across both experiments m3Blend and AdjustBlend. That is, the preliminary experiment showed a fairly strong match to the identical trials in experiment m3Blend and the main portion of experiment AdjustBlend. The averaged data subset for m3blend correlated significantly with the equivalent subsets of both the preliminary and main portions of AdjustBlend. Moreover, the four subjects who participated in both m3Blend and AdjustBlend individually showed strong matches between the respective datasets.

Effects of Conditions on Blend Judgments

The main questions that experiment AdjustBlend were intended to answer were as follows:

(1) Could the blend for a pair of instruments increase or decrease by changing the centroid (or amplitude, or harmonicity) of one of the two notes?

(2) Was the outcome of (1) related to the particular instruments in the pair, which instrument was being modified, and whether the modified instrument was on c# or e?

These questions can be answered at various levels of specificity. The present section will proceed from more general to more specific approaches to these questions.

Effects for Conditions Alone

The first question is whether any given level--regardless of which instrument and which notes were being varied--were judged as blending differently than another level. For instance, consider the centroid condition. Were "dark" conditions (all trials in which one instrument was artificially lowered in centroid) were judged as blending differently than the "normal" and/or "bright" conditions? In other words, regardless of which of the two notes were altered in centroid and which instruments were present, did the different levels of centroid tend to produce different blend ratings? *T* tests showed that the mean ratings for dark conditions differed significantly from those of bright conditions (means of .538 vs. .385, respectively; $t[46] = 2.941$, $p = .005$). The normal conditions, however, did not differ from either the dark condition ($t[34] = 1.203$, $p = .237$) or the bright conditions ($t[34] = 1.234$, $p = ..226$). These particular means suggest that blend improved toward greater brightness--a finding counter to what is expected. However, in the analysis of the data in further detail, the facts that emerge reverse this and confirm the hypothesis after all.

The other two conditions (amplitude and harmonicity) showed no such overall trend as was the case with centroid. So it appears that centroid changes in general had the biggest effect on the blend judgments.

Effects for Conditions and Single Instruments

Next, did all "dark" conditions for a particular instrument (e.g., all the trials in which flute was made artificially dark) blend differently than the normal and bright conditions for that instrument? Here the factors of the note on which the adjusted instrument is sounding, and the identity of the other instrument is being ignored. Only the instruments oboe2 and flute showed effects. For oboe2, all three means differed significantly from one another, with blend decreasing from dark to normal to bright conditions. The flute showed a difference between the normal and bright centroid conditions, again decreasing in the brightness direction. These findings confirm the hypothesis that blend increases with increasing darkness of the instruments involved.

The other factors (amplitude and harmonicity) did not show the trend described above, with one exception for the oboe2's amplitude condition: the -6 dB level blended significantly better than the -3 dB level. This suggests that blend for the pairs including the oboe will increase if the oboe's intensity is lowered.

Effects for Conditions, Single Instruments and Notes

For a particular note position of an instrument, does blend increase in different levels of a condition? In this case the only remaining parameter which is ignored is the identity of the other instrument. The answer is yes in several cases, but almost exclusively with respect to centroid, and furthermore, only in the cases of oboe2 and flute. As before, *t* tests were used to determine when the means of one condition were significantly different ($p < .05$) from another. All trials in which a flute was on e decreased in blend when going from the normal condition to the bright condition. The only other instrument showing centroid trends was oboe2's c# and e trials: there were six different cases for oboe2 in which blend increased in the direction of greater darkness and decreased in the direction of greater brightness.

The only other condition showing trends similar to that described for centroid above was with amplitude, but only in the case of the oboe2 c# trials. Blend for oboe2, when on c#, increased from -3 dB to -6 dB conditions.

Effects for Conditions, Instrument Pairs and Notes

Finally, the changes in blend for changes in factors at the individual trial level can be observed. To clarify further discussion, the following conventions will be adopted here:

1. A "triad of *instrument1-instrument2* pairs in the *condition* condition," as in "a triad of flute-<u>oboe2</u> pairs in the centroid

condition," refers to the three pairs flute-oboe2(dark), flute-oboe2(normal) and flute-oboe(bright).

2. The underlined instrument indicates which instrument and note is being varied in the three levels of the condition. Thus, a triad of flute-oboe2 pairs in the amplitude condition refers to the three pairs flute(-6 dB)-oboe2, flute(-3 dB)-oboe2 and flute(normal)-oboe2.

The outcome that is anticipated that (based on the evidence so far) is that for any given triad in the centroid condition, blend will increase in the direction of levels *bright, normal, dark*. Table 12 shows whether this was true or not. A "yes" or "no" in column 2 indicates the cases where the anticipated relationship was found for triads of pairs in the centroid condition. The blend ratings for the triad flute-oboe2, for example, were .647 (bright), .619 (normal) and .394 (dark).

However, even when the magnitudes of blend ratings follow the expected progression as they do in this example, there needs to be some method of evaluating whether the increments are actually meaningful; namely, does the dark condition blend *significantly* better than the normal and/or bright conditions? For example, consider the flute-bassoon triad which follows the expected progression yet which is dubious because the variation in blend magnitude is very small: .332 (bright), .328 (normal) and .280 (dark), a range of only .051. It is not clear that these ratings indicate that the listeners truly heard an improvement in blend. The triad of the previous

Instrument Pair	Blend improved...		sig. diffs. ($p < .05$) between conditions	Range of means
	from bright to dark?	with narrower centroid difference?		
flute-oboe2	no	yes*	2	.254
flute-oboe2	yes	yes	3	.410
oboe2-flute	yes	no	2	.323
oboe2-flute	no	yes*	0	.055
flute-bassoon	yes	no	2	.159
flute-bassoon	no	no	0	.051
bassoon-flute	no	yes*	0	.048
bassoon-flute	yes	yes	3	.450
flute-frhorn	yes	yes	2	.274
flute-frhorn	yes	no	0	.038
frhorn-flute	no	no	2	.099
frhorn-flute	yes	yes	3	.347
oboe2-bassoon	yes	yes	3	.465
oboe2-bassoon	no	no	1	.089
bassoon-oboe2	no	yes*	2	.135
bassoon-oboe2	yes	yes	3	.471
oboe2-frhorn	yes	yes	3	.428
oboe2-frhorn	no	no	2	.135
frhorn-oboe2	no	yes*	0	.038
frhorn-oboe2	yes	yes	3	.352
bassoon-frhorn	yes	yes	2	.196
bassoon-frhorn	yes	no	3	.139
frhorn-bassoon	no	no	0	.053
frhorn-bassoon	yes	yes	2	.348

Table 12. All triads of trials for the centroid condition in Experiment AdjustBlend, showing correlation of blend ratings to the dark, normal and bright levels of centroid.

example (flute-oboe2), by comparison, shows a range of .253, indicating less ambiguity in the judgments.

T tests can be used to compare the three means in a triad if the individual subject data is used. This was accomplished as follows. Recall that the dataset for experiment AdjustBlend (n = 156) consists of data averaged over 10 subjects rating 4 blocks of data. Thus there are in fact a total of 40 ratings for each trial. Employing the individual datapoints in a *t* test makes it possible to statistically compare the means. For example, in the case of flute-oboe2, all three means (.647, .619, .394) significantly differed from each other ($p < .05$), showing clearly that blend increased for every increment in centroid towards darkness; this is indicated in column 4 of Table 12. The flute-bassoon triad mentioned above, not surprisingly, shows that none of the three means were different from one another, calling into question the claim that blend increased towards darkness. For some of the other triads listed in Table 12, only two out of the three levels produced significantly different means; for example, in oboe2-flute the dark mean is different from both the normal and bright means, but the normal and bright means themselves are not significantly different. Furthermore it shows that this centroid pattern occurred over the majority of triads (14 out of 24), with all but one of those showing significant differences among at least two of the three pairs of means.

Returning once again to the two-model framework of looking at the results, the findings for these 14 triads lend support for *model 1* (sums). That is, as the overall centroid of the pair became lower (darker), the blend improved. It is also of interest to see if the results offer support for *model* 2 (absolute differences) as well. Column 3 of Table 12 indicates which triads showed that as centroid distance became closer, blend improved. As it turns out a number of the triads that exhibit *model 1* also exhibit *model 2* (10 of them). Additionally it might be expected that of the triads which did *not* exhibit *model 1* behavior (10 out of 24), some of these may exhibit *model 2* behavior. This may be examined in Column 3 of Table 12: surprisingly, only 4 out of these 10 exhibit the behavior (indicated with an asterisk). It was the case, nonetheless, that 3 of these 4 triads are triads in which this might be expected to happen, that is, the instrument whose centroid was changing was a dark instrument (<u>bassoon</u>-flute, <u>bassoon</u>-oboe2, and <u>frhorn</u>-oboe2). It would appear then, that, unlike Experiment m3Blend, *model 2* does not dominate the findings.

For the amplitude condition, it is anticipated that blend will increase in the direction of levels normal, -3 dB, -6 dB. Indeed, this did occur in 10 out of the 24 triads, as Table 13 shows. However, only seven of these showed one or more significant differences among the three comparisons of means. Furthermore, column four of the table shows that range of means was never large. The effects for amplitude, then, appear to have been very weak compared to those for centroid.

The factor of harmonicity showed only minor evidence of patterns like those for the conditions of centroid and amplitude; few significant differences

Instrument Pair	Blend improved in progressions from loud to soft?	Number of sig. diffs. ($p < .05$) between conditions (out of three)	Range of means
flute-oboe2	no	2	.111
flute-oboe2	yes	2	.191
oboe2-flute	yes	1	.104
oboe2-flute	no	0	.081
flute-bassoon	yes	0	.064
flute-bassoon	no	0	.034
bassoon-flute	no	0	.083
bassoon-flute	no	1	.079
flute-frhorn	yes	2	.098
flute-frhorn	no	0	.046
frhorn-flute	no	1	.106
frhorn-flute	yes	1	.087
oboe2-bassoon	yes	2	.178
oboe2-bassoon	no	0	.053
bassoon-oboe2	no	0	.064
bassoon-oboe2	yes	1	.076
oboe2-frhorn	yes	1	.100
oboe2-frhorn	yes	0	.069
frhorn-oboe2	no	0	.044
frhorn-oboe2	yes	0	.053
bassoon-frhorn	no	0	.018
bassoon-frhorn	no	0	.036
frhorn-bassoon	no	0	.065
frhorn-bassoon	no	0	.036

Table 13. All triads of trials for the amplitude condition in Experiment AdjustBlend, showing correlation of blend ratings to the -6 dB, -3 dB and normal levels of amplitude.

among means were found, and the range of means was very small (hence no table is included to show these results). The pattern which was hypothesized to exist for harmonicity (that blend would increase with increased inharmonicity) did indeed occur to a small degree: two of the 24 triads showed this pattern with at least one significant difference among the three pairs of means: oboe2-flute and frhorn-flute. In general, however, there was very little effect for harmonicity. For example, the pair flute-oboe2 blended roughly the same regardless of whether the flute had normal inharmonicity, zero inharmonicity, or exaggerated inharmonicity.

Summary of Findings, Experiment AdjustBlend

Experiment AdjustBlend confirmed that for most instruments, when only the centroid changes, its blend with another instrument can be increased or decreased. Blend increases as one instrument is made increasingly dark, or lower in centroid.

A weak trend was found that confirmed the predicted effect for amplitude level changes, i.e., that blend increases as one of the two instruments is made lower in amplitude. Although blend increased in the order of levels normal, -3 dB, -6 dB, the magnitude of blend rating changes was very small.

It was mentioned earlier that in real acoustic instruments, amplitude and brightness tend to co-vary, and exploring these conditions separately were a way of gauging their contributions separately. If experiment AdjustBlend was a trustworthy measure of these two contributions, then it would appear that brightness is the more important of the when it comes to blend. Thus, when performers describe their strategies for blending with other players as "matching loudness," it is the resulting change in brightness rather than the change in intensity which has the greater impact on blend.

In the presence of even rather large changes in harmonicity, blend was largely unaffected. Whether the inharmonicity was grossly exaggerated or the tones made purely harmonic, the effect on blend was negligible. Ironically, the very slight effects that *were* observed for inharmonicity were *opposite* those which were hypothesized: blend increased as inharmonicity was increased. In retrospect, however, the present author's own impression that the presence of an inharmonic note in a pair tended to "blur" the sound in such a way that the distinction between timbres was harder to notice. A visual analogy might be smearing a freshly-painted canvas with a sponge, causing once-distinct lines to blur together.

Chapter 5:

Concluding Remarks

Summary

Orchestration manuals and treatises provided strong evidence for the suitability of selecting concurrently-sounding timbres as the focus for a perceptual study. Blend is a frequently identified and defined sonic effect in these sources, and authors show a consensus on its meaning: blend is achieved when two or more concurrently sounding timbres fuse as one to make a single timbral identity. The manuals offer several suggestions as to what conditions lead to blended combinations (e.g. choices of dynamics and articulations, voicing arrangements, performance practices) and provide several examples of instrument combinations which blend. There are some suggestions that attaining different degrees of blend may be useful compositionally, and that certain styles can be characterized in terms of the degree of homogeneity or heterogeneity of their orchestrations. Finally, there are numerous orchestral techniques in which obtaining blend is a prerequisite for their success: for example, augmenting timbres, inventing timbres, and imitating timbres.

Some of the suggestions that orchestration manuals give for selecting and arranging instruments to attain blend are: synchronizing attacks as much as possible; using soft dynamics; various strategies of chord-voicing, such as closely-spaced harmonies, "harmonic-series" chord spacing, following the natural order of register for instruments (i.e. using the standard orchestral score layout as a guide), or interlocking arrangements of chords; and limiting the number of timbres and narrowing the multiplicity of spatial origins.

The literature of psychoacoustics and auditory perception offers some clues as to the acoustical and psychological factors that determine blend. Various studies provided evidence that blend could vary as a function of pitch separation, inharmonicity, harmonic coincidence, common temporal elements, masking, location of spectral concentration, and onset synchrony among the constituent elements in a combination. Emerging from this review are a few candidate mechanisms for explaining timbre on a more cognitive level: segregation by identification (the more identifiable the constituent elements are in a combination, the less they blend), blend by indistinct numerosity (an ambivalent sense of the number of pitches present leads to blend), and blend by misattribution of pitch to timbre (uncertainty of which timbres belong to which pitches in a combination leads to blend).

The goal of the study was to obtain descriptions of the acoustical conditions that promoted blend among natural musical instrument sounds.

It was considered of primary importance to employ descriptions that pertained to salient high-level attributes, or "distinctive features" typical of acoustic instruments. This way it would be possible to obtain a metric for blend that could be transferred to contexts other than instrumental music, or to instruments other than the particular ones used in this study. It was also desired that the initial parts of the study be exploratory in nature. Hence, rather than creating artificial stimuli with limited parameters of variation, informationally rich, life-like stimuli were used. These stimuli were the same as those used by Grey (1975).

Judgments of blend for various pairs of instruments were obtained using a continuous scale ranging from "blended" to "separated." The data was compared to a number of acoustical properties of the instruments, including both solo properties and properties arising from their interaction. Two of the factors that emerged as major determinants of blend were the location of the midpoint of the spectral energy distribution (centroid), and the swiftness of the onset for the two tones in a pair. The behavior of the two factors would be characterized as a "gravitational effect": better blends were obtained as the values of the property for both tones were made both closer and lower. For spectrum, the presence of at least one "dark" instrument (low centroid) in the pair helped promote blend, and blend was increased on each trial in which the second instrument was lower in centroid; at the same time, better blends could be found among instruments that were close in centroid, regardless of the overall centroid height of the pair. For onset, the presence of at least one instrument with an abrupt, noise-free attack in the pair helped promote blend, and blend was increased on each trial in which the second instrument was swifter in attack; at the same time, better blends could be found among instruments that were close in attack duration, regardless of their location on the continuum of longer and shorter attacks.

The degree to which each of these dimensions (height and closeness) was influential was related to the musical interval of presentation. When the interval was a unison overall centroid height was the dominant factor, but when the interval was a minor third, the closeness of centroid was the dominant factor. For onset, overall shortness and closeness were equally important with the unison interval, but with the minor third closeness became the dominant factor. However, the effect for attack factors in the minor third condition were overall of smaller magnitude than in the unison condition. Beyond these features, there was little that differentiated the data between the unison and minor thirds condition, and consequently, little that differentiated the inversional pairings within the minor thirds condition. One exception to the latter was in the tendency of the oboe to show greater amounts of blend in the lower position than the upper position.

Another trend found in the data is what was called "blending power;" namely, that certain instruments tended to impose a certain degree of blend regardless of what they were paired with. In particular, tsax1 and tsax2 were "poor blenders" and mutecello, frhorn and bassoon were "good blenders." Acoustical factors contributing to this affect appeared to be related to centroid, onset, acoustic dissonance, and pitch deviation.

Temporal factors seemed to play a role in determining the blend of instruments as well. Blend increased as the amplitude and centroid envelopes of the instruments became more correlated. Curiously, pitch envelopes did *not* show a similar effect; however the general degree of pitch deviations (amount of jitter, or deviation from a steady tone) exhibited a "gravity effect" similar to that of centroid and onset: blend improved as the overall amount of pitch deviation got lower, or the instruments were closer in their levels of deviation. There were also slight effects for the amount of individual harmonic synchrony across two instruments in a pair. Curiously, however, they showed an opposite temporal phenomenon: blend increased as instruments' harmonics were *more* asynchronous with each other.

As a result of the findings in the first two studies, it was decided to investigate the factor of centroid somewhat further. In order to determine that the effects for the different levels of centroid were not simply artifacts of the different instruments which instantiated them, artificial changes in centroids for single instruments were devised. This way it was possible to observe the change in blend for the flute and oboe (for example) as the oboe was made brighter or darker. The results emphatically supported the "overall height" dimension of the centroid effect: namely, that blend increases for a pair with each successive darkening of one of the two instruments.

Similar conditions involving amplitude and harmonicity were evaluated as well. Similarly to centroid, but to a less emphatic degree, it was shown that blend increases for a pair as one of the instruments decreases in amplitude. The harmonicity factor showed few distinctive results, suggesting mainly that harmonicity has little effect on blend; however, there were a few instances of blend improving as the overall inharmonicity increased.

Contributions of the Study

The three experiments reported here suggest strategies for selecting concurrent timbres that blend. These strategies--using darker timbres, using instruments with correlated temporal envelopes, and so on--could be employed as *heuristics* for obtaining blended combinations. A heuristic is a principle which does not necessarily guarantee solutions, but offers helpful guidelines that lead to quicker solutions and reduce the problem space (i.e., eliminates obvious wrong solutions).

Musicians have from time to time wondered about the feasibility of a "systematic approach" to orchestration. Rimsky-Korsakov (1912) wrote in his introduction that his original intent was to frame a system of orchestration based on the work of acousticians such as Helmholtz (Rimsky-Korsakov, 1912, p. VII), while Schillinger (1941) speculated briefly about an "Acoustical Basis of Orchestration" (pp. 1603-1604). At the same time, these same authors and others reject the idea on the grounds that "art of poetic orchestration" cannot be learned by rule (Rimsky-Korsakov, p. 1) and that "any stated principle is immediately subject to qualification from so many angles that is usefulness is brought into question" (Piston, 1955, p. 444). On the other hand, Piston (1955) also observes that the chief value of a rule or recipe "lies in its utility as a commonplace from which to survey other and more interesting possibilities" (p. 450).

It is this latter viewpoint which characterizes the objective of this project's study of blend. The distinction between a "utility" and an "interesting possibility" is a matter of generality versus specificity. For most composers, "obtaining a blend" is rarely the primary objective of selecting instruments for a chord or melodic doubling; more likely, "blend" is used as a *means* for insuring the success of some other effect. The experimental tasks used in this study purposely centered on an objective, non-aesthetic definition of blend: whether one or two timbres are heard. Although the heuristics remain to be evaluated in actual orchestration applications, such knowledge could potentially be used as a utility for fine-tuning several "interesting possibilities." It has been asserted that perceptual research cannot answer questions about how music is heard, since the listener is not a passive receiver but an active participant in shaping the meanings of sounds in ways that depend on his or her mood, musical culture and previous musical experiences (Randall, 1965, 1967; see also Sloboda, 1989). The "one vs. two timbres" definition of blend captures a fundamental auditory percept that is not itself an aesthetic decision (such as whether a chord is "unstable" is "resolved" by some other chord), and is operative regardless of the aesthetic orientation of the listener. Even if a different meaning for the word "blend" is adopted by the listener due to aesthetic motivations, at some level the task of determining whether one or two timbres is present is fundamental to that meaning.

The chief value of the heuristics for blend suggested here is their generalizability, the fact that they are based on underlaying psychoacoustical principles rather than stated in terms of specific instances. For example, consider some ways in which the findings of this study suggest solutions to problems of orchestration. If blend cannot be obtained in a particular chord, then bright instruments such as oboe could be marked with a lower dynamic marking (because softer playing often leads to darker spectra), or extremely bright instruments such as muted brass might need to be eliminated entirely. Similarly, if instruments with lengthy, noisy onsets are involved in the ensemble, giving them articulations or dynamic markings that attenuate the contribution of their onset portions may improve the blend; or another solution might be to place such instruments in a higher octave, since the lower ranges of many instruments tend to produce lengthy onsets (Meyer, 1978, pp. 39-70).

Traditional orchestration manuals, on the other hand, tend to define both the problem and the solution in instrument-specific terms. For example: oboes often blend poorly in combinations, unless they are played softly; saxophones blend poorly in their lower register; and so on. Furthermore, this approach limits the applicability of the advice to only those instruments in currency at the time of the orchestration book's publication. The heuristics suggested here, however, are general enough to be applied to electronic music.

Relationships to Earlier Studies

A number of the findings in this study corroborate the statements of orchestration manuals. The finding about "blending power" appears to be a illustration of what orchestration manuals refer to as "naturally homogeneous" instruments. Indeed, two instruments that are cited most frequently as natural blenders in orchestration manuals are among those with the greatest blending power in this study, French horn and bassoon. This study's findings concerning the use of similarity of onsets and softer dynamics for obtaining blends also has correspondences in orchestration manuals.

The results concerning centroid closeness have an obvious parallel in the general recommendations given in orchestration manuals to combine instruments of "like qualities." Rogers (1951, p. 5) in particular refers to combining sounds according to bright and dark qualities. Parallels to the other result concerning centroid--overall darkness promotes blend--is surprisingly absent in orchestration manuals (with the exception of a brief allusion by Rogers). The only way this phenomenon is demonstrated is in the many instances of instrument combinations suggested to obtain blends that happen to have low centroids, such as cello, bassoon, trombone, bass clarinet, and French horn.

There are many ways in which this study corroborates auditory phenomena that were reviewed in Chapter 2. The effects pertaining to attack duration agree with the findings of Rasch (1978); his studies showed that listeners could more correctly identify the melodic direction of different voices in a polyphonic context when their attack times were offset from each other, and their performance improved as a function of the amount of offset time. Although listeners in this study gave direct judgments of blend, it is hypothesized that identifiability of constituent instruments in a pair is a measure of their segregation ("segregation by identification"). Thus, closeness in onset time promotes blend, while greater differences in onset time produce greater amounts of segregation.

The results in the present study regarding amplitude envelopes and centroid envelopes (greater correlation promoting increased blend) seem consonant with the "common fate" model of Auditory Scene Analysis, the principle that sounds changing according to similar temporal patterns are perceived as issuing from a single source. On the other hand, the results in the study concerning measures of harmonic synchrony yield conclusions contrary to what Auditory Scene Analysis would suggest. The study showed that greater asynchrony promoted increased blend, whereas other research suggests that asynchrony would promote segregation.

The results pertaining to centroid in the three experiments are notably compatible with Goodwin's investigation of blend among choral singers (Goodwin, 1989). His findings showed that, when a singer strives to obtain a blend, he or she produces "slightly stronger fundamental frequencies in combination with fewer and weaker upper partials." This is very similar to the observation of centroid "gravity," since the outcome of low centroids is due to energy weighted towards the fundamental and away from the upper harmonics.

The results of the present study are comparable to those of Grey (1975), since the two studies share the same stimuli. Two primary findings of his study were that timbral similarity could be well accounted for by closeness in centroid height and presence of attack noise energy. Of course, in this study they are primary determinants of blend, as well. A third dimension of Grey's was declared to pertain to a measure of the synchrony of onsets and offsets of harmonics in an instrument. In Experiment UnisonBlend, Multidimensional Scaling analysis similarly showed a correlation with on/off synchrony although this was fairly weak compared to the other dimensions. There is apparently a great deal of similarity in the factors determining both similarity and blend for these tones, even though the tasks that elicited these results were very different in the two studies (simultaneous blend vs. successive similarity comparisons).

Shortcomings of the Study

The central limiting feature of this study was the stimuli, which, although strongly musical in their timbre, were limited to single notes. In a more advanced musical presentation (melodies in thirds, a succession of chords) a listener could obtain more ecological information from hearing a wider range of pitches from the instruments involved. Recently Kendall and Carterette (1991) have presented some studies of concurrent timbres that use presentations of this sort. Also, in melodies, contextual, "prosodic variations" might have an effect (see Risset & Wessel, p. 42; Kendall, 1986) on blend judgments because every transient that occurs between a pair of notes is unique to the way those two notes are connected sequentially in a melody.

Studies from Auditory Scene Analysis indicate that fusion can be aided by sequential streaming principles. For example a pair of instruments might not blend in isolation, but as part of a horizontal streaming pattern, could be perceived as blending; or the reverse may be true. Obviously the presentation of single note pairs prevented the emergence of such a phenomena. Another fact pertaining to Auditory Scene Analysis is that fusion requires a certain amount of integration time (see Bregman, 1990, p. 333). Although the Stanford tones were sufficiently long to obtain some degree of fusion, it is possible that and increased fusion might have been obtained with longer durations.

The methods of acoustical analysis of the tones were far from exhaustive. Indeed, the scatterplots used to show the modeling of the data from UnisonBlend suggested that the analyses missed something important about the single-reed instruments. Additional work concerning the temporal properties of the instruments might prove worthwhile.

It may seem strange that aspects of attack played a prominent factor in the data when great pains had been taken to synchronize attacks by offsetting start times according to Gordon's PAT values. In fact, nearly all listeners mentioned that large differences in attack qualities in some of the pairs influenced their blend judgments. However, attack *duration* and attack *synchronicity* are different matters. Gordon's PAT offsets may reflect one type of adjustment made by a performer to induce the sense of blend; attack duration, while adjustable to some degree by performers, was not equalized here, and it is not surprising that this acoustical parameter played a role in the judgments.

Suggestions for Further Study

It would be useful to conduct follow-up work to explore the robustness of the present claims about blend. Exploring how instrumental performers effect blend (similar to the work of Goodwin, 1989) would be a way of finding out the degree to which they manipulate centroid, attack and temporal envelopes to attain blend. The performers could be asked to blend and segregate with each other, and recordings of these sounds could be compared acoustically.

The factors playing the largest roles in the blend judgments, centroid and attack, could be explored further. The change in the relative importance of *overall centroid height* and *centroid difference* with different interval conditions (unisons and minor thirds) found in Experiments UnisonBlend and m3Blend could be investigated systematically. Observing changes in blend judgments for trials that change only in centroid height vs. those that change only in centroid distance, with interval as an additional factor, might explicate the mechanisms that are suggested here. This could be accomplished by using the method of artificial centroid manipulation that was used in Experiment AdjustBlend. Similarly, the role of attack duration could be explored by artificial manipulations of the onset characteristics of tones; of course, the challenge would be to maintain naturalness, since the factors affecting attack duration are complex and not well understood.

The orchestration and psychoacoustic surveys suggest some entirely new topics for investigating orchestration as well. Some of the suggestions in orchestration manuals that gaps in chords be avoided, and the material from auditory perception on the differential fusion of intervals invite some possibilities. For instance, do chords with large gaps in them fuse poorly? What is the correlation between size of gap and amount of segregation, how does it relate to harmonic coincidence and traditional dissonance? Another question is also related to chord voicings: does interlocking improve the blend of a chord?

Some of the mechanisms nominated for explaining blend from a cognitive point of view (segregation by identification, blend by uncertainty of source numerosity, and blend by misattribution of pitch to timbre) might provide interesting material for research. For example, various concurrent combinations could be rated for accuracy of instrument identification, correct identification of number of pitches, or accuracy of matching each timbre to its pitch; these findings could be compared with blend judgments for the same stimuli to evaluate the importance of each of these cognitive mechanisms for determining blend.

References

Abeles, H. (Summer 1979). Verbal timbre descriptors of isolated clarinet tones. *Council for Research in Music Education* **59**, 1-7.

Adler, S. (1989). *The Study of Orchestration.* New York: W. W. Norton, Second Edition.

Apel, W. (1972). *Harvard Dictionary of Music.* Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Bacon, F. (1627). *Sylva Sylvarum, or A Natural History in Ten Centuries.* London: William Lee.

Beauchamp, J.W. (1982). Synthesis by spectral amplitude and 'Brightness' matching of analyzed musical instrument tones. *Journal of the Audio Engineering Society* **30**, 396-406.

Becker, H. (1964). The history of instrumentation. In K.G. Fellerer (Ed), *Anthology of Music; a collection of complete musical examples illustrating the history of music.* Cologne: A. Volk Verlag. Trans. by R. Kolben.

Belkin, A. (1988). Orchestration, perception, and musical time: A composer's view. *Computer Music Journal* **12** (2), 47-53.

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

Berlioz, H. (1856). *A Treatise on Modern Instrumentation and Orchestration.* Boston: Elias Howe. Trans. by Mary Cowden Clark.

Berlioz, H./Strauss, R. (1948). *Treatise on Instrumentation.* New York: Kalmus. Revised and enlarged edtion of Berlioz (1856). Trans. by Theodore Front.

von Bismarck, G. (1974). Timbre of steady sounds: a factorial investigation of its verbal attributes. *Acustica* **30**, 146.

Blatter, A. (1980). *Instrumentation/Orchestration*. New York: Longman.

Brant, H. (1967). Space as an essential aspect of musical compositions. In E. Schwartz & B. Childs (Eds.), *Contemporary Composers on Contemporary Music* (pp. 221-242). New York: Holt, Rinehart and Winston.

Brant, H. (1971). Orchestration. In J. Vinton (Ed.), *Dictionary of Contemporary Music* (pp. 538-546). New York: E. P. Dutton.

Bregman, A.S., & Chalikia, M.H. (1989). The perceptual segregation of simultaneous auditory signals: pulse train segregation and vowel segregation. *Perception and Psychophysics* **46** (5), 487-496.

Bregman, A.S. (1990). *Auditory Scene Analysis*. Cambridge, Massachusetts: MIT Press.

Brindle, R.S. (1966). *Serial Composition*. London: Oxford University Press.

Carse, A. (1941). Text-books on orchestration before Berlioz. *Music and Letters* **22**, 26-31.

Carse, A. (1964). *The History of Orchestration*. New York: Dover.

Clark, M., Luce, D., Abrams, R., Schlossberg, H., & Rowe, J. (1963). Preliminary experiments on the aural significance of tones of orchestral instruments and on choral tones. *Journal of the Audio Engineering Society* **11**, 45-54.

Clark, M., & Miller, P. (1964). Dependence of timbre on the tonal loudness produced by musical instruments. *Journal of the Audio Engineering Society* **12**, 28-31.

Cogan, R., & Escot, P. (1976). *Sonic Design: the Nature of Sound and Music*. New York: Prentice-Hall.

Cogan, R. (1984). *New Images of Musical Sound*. Cambridge, Massachusetts: Harvard University Press.

Cone, E.T. (1989). *Music: A View from Delft*. Chicago: University of Chicago Press.

Dannenbring, G.L., & Bregman, A.S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception and Psychophysics* 24, 369-376.

Ehresman, D.L., & Wessel, D.L. (1978). Perception of timbral analogies. Tech. Rept. 13. IRCAM, Paris.

Elliott, C. (1975). Attacks and releases as factors in instrument identification. *Journal of Research in Music Education* 23, 35-40.

Erickson, R. (1974). LOOPS: An informal timbre experiment. Tech. Rept. Q-037. Center for Music Experiment, University of California at San Diego, La Jolla, California.

Erickson, R. (1975). *Sound Structure in Music.* Berkeley, CA: University of California Press.

Erickson, R. (1982). New Music and Psychology. In D. Deutsch (Ed), *The Psychology of Music* (pp. 517-536). New York: Academic Press.

Forsyth, C. (1914). *Orchestration.* New York: Dover.

Fowler, L. (June 1980). How to let individual instruments speak for themselves. *Down Beat* 47, 68-69.

Freed, D.J. (1988). *Perceptual control over timbre in musical applications using psychophysical functions.* Master's thesis, Northwestern University, Evanston, IL.

Freed, D.J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America* 87, 311-322.

Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems.* Boston: Houghton-Millen.

Gold, B., & Rabiner, L. (1969). Parallel processing techniques for estimating pitch periods of speech in the time domain. *Journal of the Acoustical Society of America* 46, 442-448.

Goodwin, A.W. (1989). An acoustic study of individual voices in choral blend. *Journal of Research in Singing* 13 (1), 25-34. Reprint Goodwin (1980), same title, The Journal of Research in Music Education 28 (2), 119-128.

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

Gordon, J.W. (May 1984). *Perception of Attack Transients in Musical Tones.* Ph.D. dissertation, Stanford University, Stanford, CA. Issued through the Department of Music, Center for Computer Research in Music and Acoustics, Report no. STAN-M-17.

Gordon, J.W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America* **82**, 88-105.

Green, D.M. (1976). *An Introduction to Hearing.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Grey, J.M., & Gordon, J.W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America* **63**, 1493-1500.

Grey, J.M., & Moorer, J.A. (1977). Perceptual evaluation of synthesized musical instrument tones. *Journal of the Acoustical Society of America* **62**, 454-466.

Grey, J.M. (February 1975). *An Exploration of Musical Timbre.* Ph.D. dissertation, Stanford University, Stanford, CA. Issued through the Department of Music, Center for Computer Research in Music and Acoustics, Report no. STAN-M-2.

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

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

Gridley, M.C. (1987). Trends in description of saxophone timbre. *Perceptual and Motor Skills* **65**, 303-311.

Halikia, M.H. (1985). *The Perceptual Segregation of Simultaneous Sounds.* Ph.D. dissertation, McGill University, Montreal, Canada.

Hall, J.W. (1987). Experiments on comodulation masking release. In W.A. Yost & C.S. Watson (Eds.), *Auditory Processing of Complex Sounds* . Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events.* Cambridge, Massachusetts: MIT Press.

Hartmann, W.M., McAdams, S., & Smith, B.K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America* **88**, 1712-1724.

Houtsma, A.J.M., & Beerends, J.G. (1986). The role of aural frequency analysis in pitch perception with simultaneous complex tones. In B.C.J. Moore & R.D. Patterson (Eds.), *Auditory Frequency Selectivity* . New York: Plenum Press.

Howard, J.H., & Silverman, E.B. (1976). A multidimensional scaling analysis of 16 complex sounds. *Perception and Psychophysics* **19** (2), 193-200.

Hutchinson, W., & Knopoff, L. (1978). The acoustic component of western consonance. *Interface* 7, 1-29.

Jakobsen, R., Fant, C.G.M., & Halle, M. (1951). *Preliminaries to Speech Analysis: the Distinctive Features and Their Correlates.* Cambridge, Massachusetts: M.I.T. Press.

Jeffress, L.A. (1970). Masking. In J.V. Tobias (Ed.), *Foundations of Modern Auditory Theory* . New York: Academic Press.

Johnson, S.C. (1967). Hierarchic Clustering Schemes. *Psychometrika* **32**, 241-254.

Jones, D., Kendall, G., & Martens, W. (1984). Designing a sound room for auditory research using the TEF™. Unpublished. Poster presented at the 1984 International Computer Music Conference.

Kendall, R.A. (1986). The role of acoustic signal partitions in listener categorization of musical phrase. *Music Perception* 4, 185-214.

Kendall, R.A., & Carterette, E.C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception* 8, 369-404.

Koury, D.J. (1986). *Orchestral Performance Practices in the Nineteenth Century.* Ann Arbor, Michigan: UMI Research Press.

Kruskal, J.B., & Wish, M. (1978). *Multidimensional Scaling.* Beverly Hills, California: Sage Publications.

Krumhansl, C.L. & Iverson, P. (in press). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance.*

Lerdahl, F. (1987). Timbral hierarchies. *Contemporary Music Review* **2**, 135-160.

Lichte, W.H. (1941). Attributes of complex tones. *Journal of Experimental Psychology* **28** (6), 455-480.

Marin, C.M.H., & McAdams, S. (1991). Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width. *Journal of the Acoustical Society of America* **89**, 341-351.

McAdams, S. (May 1984). *Spectral Fusion, Spectral Parsing and the Formation of Auditory Images.* Ph.D. dissertation, Stanford University, Stanford, CA. Issued through the Department of Music, Center for Computer Research in Music and Acoustics, Report no. STAN-M-22. .

McAdams, S. (1989). Concurrent sound segregation. I: Effects of frequency modulation coherence. *Journal of the Acoustical Society of America* **86**, 2148-2159.

Melara, R.D., & Marks, L.E. (1990). Interaction among auditory dimensions: Timbre, pitch and loudness. *Perception and Psychophysics* **48**, 169-178.

Meyer, J. (1978). *Acoustics and the Performance of Music.* Frankfurt/Main: Verlag das Musikinstrument. English translation of Akustik und Musikalische Auffuhrungspraxis.

Miller, J.R., & Carterette, E.C. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America* **58**, 711-720.

Moore, B.C.J., Glasberg, B.R., & Shailer, M.J. (1984). Frequency and intensity difference limens for harmonics within complex tones. *Journal of the Acoustical Society of America* **75**, 550-561.

Moore, B.C.J., Glasberg, B.R., & Peters, R.W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America* **80**, 479-483.

Moore, B.C.J., & Glasberg, B.R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America* **74**, 750-753.

Moorer, J.A., Grey, J., & Strawn, J. (1977). Lexicon of analyzed tones (Part II: clarinet and oboe tones). *Computer Music Journal* **1** (3), 12-29.

Moorer, J.A., Grey, J., & Strawn, J. (1978). Lexicon of analyzed tones (Part III: the trumpet). *Computer Music Journal* **2** (2), 23-31.

Moorer, J.A. (1973). The Heterodyne Method of Analysis of Transient Waveforms. Tech. Rept.. Artificial Intelligence Laboratory, Stanford University, Stanford, CA.

Olson, H.F. (1967). *Music, Physics and Engineering*. New York: Dover, 2nd Edition.

Opolko, F., & Wapnick, J. (1989). *McGill University Master Samples User's Manual*. Montreal: McGill University Faculty of Music.

Ott, J. (April 1969). A new approach to orchestration. *The Instrumentalist* **23** (9), 53-55.

Pepinsky, A. (1941). Masking effects in practical instrumentation and orchestration. *Journal of the Acoustical Society of America* **12**, 405-408.

Piston, W. (1955). *Orchestration*. New York: Norton.

Plomp, R., & Levelt, W.J.M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America* **38**, 548-560.

Plomp, R., & Steeneken, H.J.M. (1971). Pitch versus timbre. In *Proceedings of the Seventh International Congress on Acoustics* (pp. 377-380). Budapest:

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

Plomp, R. (1976). *Aspects of Tone Sensation, a Psychophysical Study*. London: Academic Press.

Pollard, H.F., & Jansson, E.V. (1982). A tristimulus method for the specification of musical timbre. *Acustica* **51**, 162-171.

Pruzansky, S., *UNIX SINDSCAL User's Guide*. AT&T Bell Laboratories, Murray Hill, New Jersey, January, 1987.

Randall, J.K. (Spring-Summer 1965). A report from Princeton. *Perspectives of New Music* **3** (2), 84-92.

Randall, J.K. (1967). Three lectures to scientists. *Perspectives of New Music* 5, 124-130.

Rasch, R.A. (1978). The perception of simultaneous notes such as in polyphonic music. *Acustica* 40, 21-33.

Rasch, R.A. (1979). Synchronization in performed ensemble music. *Acustica* 43, 121-131.

Riddle, N. (1985). *Arranged by Nelson Riddle.* Secaucus, NJ: Warner Brothers.

Rimsky-Korsakov, N. (1912). *Principles of Orchestration.* New York: Dover. Trans. by E. Agate. Published by Dover in 1964.

Risset, J., & Wessel, D. (1982). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed), *The Psychology of Music* (pp. 25-88). New York: Academic Press.

Rogers, B. (1951). *The Art of Orchestration: Principles of Tone Color in Modern Scoring.* New York: Appleton-Century-Crofts.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104, 192-233.

Rossing, T.D., Sundberg, J., & Ternstrom, S. (1985). Voice timbre in solo and choir singing: is there a difference? *Journal of Research in Singing* 8 (2), 1-8.

Rossing, T.D., Sundberg, J., & Ternstrom, S. (1987). Acoustic comparison of soprano solo and choir singing. *Journal of the Acoustical Society of America* 82, 830-836.

Roth, J.M., Kendall, G.S., & Decker, S.L. (1985). A network sound system for UNIX. In B. Truax (Ed.), *Proceedings of the International Computer Music Conference.* Computer Music Association.

Russo, W. (1968). *Jazz Composition and Orchestration.* Chicago: University of Chicago Press.

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

Sandell, G.J. (1989a). Perception of Concurrent Timbres and Implications for Orchestration. In *Proceedings of the 1989 International Computer Music Conference* (pp. 268-272). Computer Music Association. San Francisco: Computer Music Association.

Sandell, G.J. (1989b). Effect of spectrum and attack properties on the evaluation of concurrently sounding timbres. *Journal of the Acoustical Society of America* 86 (Suppl.1), S89. Abstract of presentation to the 118th Meeting of the Acoustical Society of America, St. Louis, October 1989.

Sandell, G.J. (1991). A library of orchestral instrument spectra. In *Proceedings of the 1991 International Computer Music Conference.* The Computer Music Association.

Scharf, B. (1970), Critical Bands. In J. Tobias (Ed.), *Foundations of Modern Auditory Theory*. New York: Academic Press.

Scheffers, M.T.M. (1983). *Sifting Vowels: Auditory Pitch Analysis and Sound Segregation*. Ph.D. dissertation, Groningen University, The Netherlands.

Scherchen, H. (1933). *Handbook of Conducting*. London: Oxford University Press.

Schillinger, J. (1941). *The Schillinger System of Musical Composition.* New York: Fischer.

Schoenberg, A. (1911). *Theory of Harmony*. Berkeley, California: University of California Press (1978). Roy E. Carter, translator.

Schoenberg, A. (1931). Instrumentation. In L. Stein (Ed), *Style and Idea* (pp. 330-336). New York: St. Martin's (1975).

Shakespeare, W. (1605). *Othello,* Dramatic play.

Shannon, R.V. (1976). Two-tone unmasking and suppression in a forward-masking situation. *Journal of the Acoustical Society of America* 59, 1460-1470.

Shepard, R.N. (1974). Representation of structure in similarity data: problems and prospects. *Psychometrika* 39, 373-421.

Singh, P.G. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? *Journal of the Acoustical Society of America* **82**, 886-899.

Slawson, W. (1978). Review of R. Erickson's Sound Structure in Music. *Journal of Music Theory* **22** (1), 105-109.

Slawson, A.W. (1985). *Sound Color.* Berkeley, CA: University of California Press.

Sloboda, J. A. (1989). Music psychology and the composer. In S. Nielzén and O. Olsson (Eds.), *Structure and Perception of Electroacoustic Sound and Music* (pp. 3-12). Amsterdam: Excerpta Medica.

Solomon, L.N. (1958). Semantic approach to the perception of complex sounds. *Journal of the Acoustical Society of America* **30**, 421-425.

Stern, R.M. (July 15 1972). Perception of Simultaneously Presented Musical Timbres. Tech. Rept.. Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Massachusetts Quarterly Progress Report, pp. 202-210.

Stiller, A. (1985). *Handbook of Instrumentation.* University of California Press.

Strawn, J.M. (June 1985). *Modeling Musical Transitions.* Ph.D. dissertation, Stanford University, Stanford, CA. Issued through the Department of Music, Center for Computer Research in Music and Acoustics, Report No. STAN-M-26.

Strong, W., & M. Clark, J. (1967). Synthesis of wind-instrument tones. *Journal of the Acoustical Society of America* **41**, 39-52.

Stumpf, C. (1890). *Tonpsychologie.* Berlin: Springer.

Summerfield, Q., & Assmann, P.F. (1989). Auditory enhancement and the perception of concurrent vowels. *Perception and Psychophysics* **45**, 529-536.

Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acustica* **30**, 201-213.

Varèse, E. (1936). New instruments and new music. In E. Schwartz & B. Childs (Eds.), *Contemporary Composers on Contemporary Music* (pp. 196-198). New York: Holt, Rinehart and Winston (1967). Lecture given at Mary Austin House, Santa Fe, New Mexico, 1936.

Vercoe, B., *CSOUND: A manual for the audio processing system and supporting programs.* Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1986.

Walsh, S. (July 1972). Schumann's Orchestration: function and effect. *Musical Newsletter* **2** (3), 3-8.

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology* **13**, 228-240.

Wegel, R.L., & Lane, C.E. (1924). The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. *Physiological Review* **23**, 266-285.

Wessel, D.L. (1978). Low dimensional control of musical timbre. Tech. Rept. 12. IRCAM, Paris.

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

Wright, J.K., & Bregman, A.S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review* **2**, 93-92.