MODELING MUSICAL TRANSITIONS

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF MUSIC

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

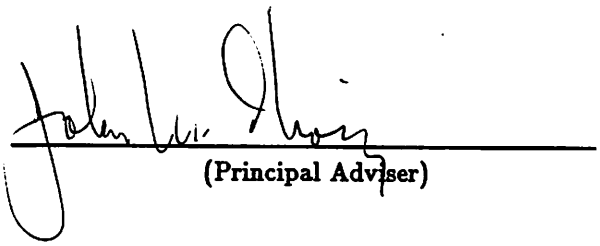
DOCTOR OF PHILOSOPHY

By

John Michael Strawn

June 1985

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

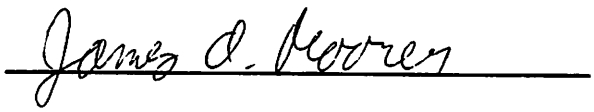


(Principal Adviser)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

(Hearing and Speech)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Approved for the University Committee
on Graduate Studies:

Dean of Graduate Studies & Research

Abstract

Due to limitations of available analysis techniques, musical instruments have to date been studied one note at the time. As a first step in exploring larger musical contexts, this dissertation examines what happens in the transition between notes. In a transition, the pitch changes, the amplitude momentarily drops, and there are spectral changes. These elements of a transition were the focus of this thesis, along with the length of time required for the transition.

Ascending and descending intervals, ranging in size from major second through minor seventh, were recorded on nine non-percussive orchestral instruments. Plots of many of these 212 transitions are presented in the text and appendices. Existing techniques were extended to permit time-varying power and spectral analyses of the transitions. The recorded transitions did not vary significantly with the size of the instrument (except to a certain extent for the strings), the size of the interval, or the direction of the interval. On the other hand, there were characteristic differences between tongued and untongued transitions in the gap time between notes, in the time-varying amplitude envelope of the transition, and in the time-varying spectral changes around the transition. Analysis of repeated performances of the same interval and playing style on a given instrument showed that the set of recordings was representative. The ascending third (tongued and untongued, with and without bow change) was chosen from the clarinet, trumpet, and violin recordings for further work.

The short-time Fourier transform (phase vocoder) was shown to be adequate for modeling a transition. New methods were developed for creating and editing line-segment approximations which were also shown to be adequate for modeling transitions.

A number of experiments were conducted on the time gap, amplitude dip, spectral changes, and slope of the amplitude envelope at the transition. Categorical perception of transitions could not be shown to occur. Some insight was gained into the perception of timbre in transitions.

Various useful signal processing techniques were developed, including methods for amplitude scaling and for extending waveforms. The dissertation concludes with a summary of methods, culled from the experimental work, for creating natural-sounding synthetic transitions.

Approved for publication:

Music Department

Dean of Graduate Studies & Research

CONTENTS

Chapter 1. Introduction

Chapter 2. Analysis of Physical Properties of Transitions

Chapter 3. Modeling Time-varying Spectral Cues

Chapter 4. Overall Spectral and Amplitude Cues

Chapter 5. Time-varying Amplitude Cues

Chapter 6. Categorical Perception

Chapter 7. The Last Chapter

CHAPTER 1

INTRODUCTION

"Everyone knows" that music is made of notes. Some music, called *monophonic*, is played on one instrument—and notes are strung together. When more than one instrument is playing simultaneously (*polyphony*), several notes are heard at the same time. Due to the problems involved in recording and analyzing musical instruments, most studies of the physics or the perception of musical sound to date have dealt with isolated notes. With the advent of digital audio and digital signal processing, these constraints no longer apply, so musical sound in a musical context can be explored more easily than before. The musical context chosen here is monophony. In particular, this study concentrates on the transition between notes, the region where successive notes are connected.

What is a Transition? (I)

> *The monophonic instrument is thus the most important musical instrument, and the starting point of all musical thinking* (LeCaine 1956, p. 465)

At least since von Helmholtz,* musical notes have been split into a central region called the *steady-state*, which is preceded by an *attack* and followed by a *decay* (see Figure 1.1a). This model still dominates thinking about the overall shape of a musical note.

The financial and organizational assistance of Lucasfilm Ltd. and its affiliate The Droid Works is gratefully acknowledged. "This research was supported by the National Science Foundation under grant NSF MCS 82-14350 and the System Development Foundation under grant SDF #345. The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of Stanford University, any agency of the U.S. Government or of sponsoring foundations."

*Von Helmholtz's work first appeared in 1862; I have used the posthumous edition of 1913.

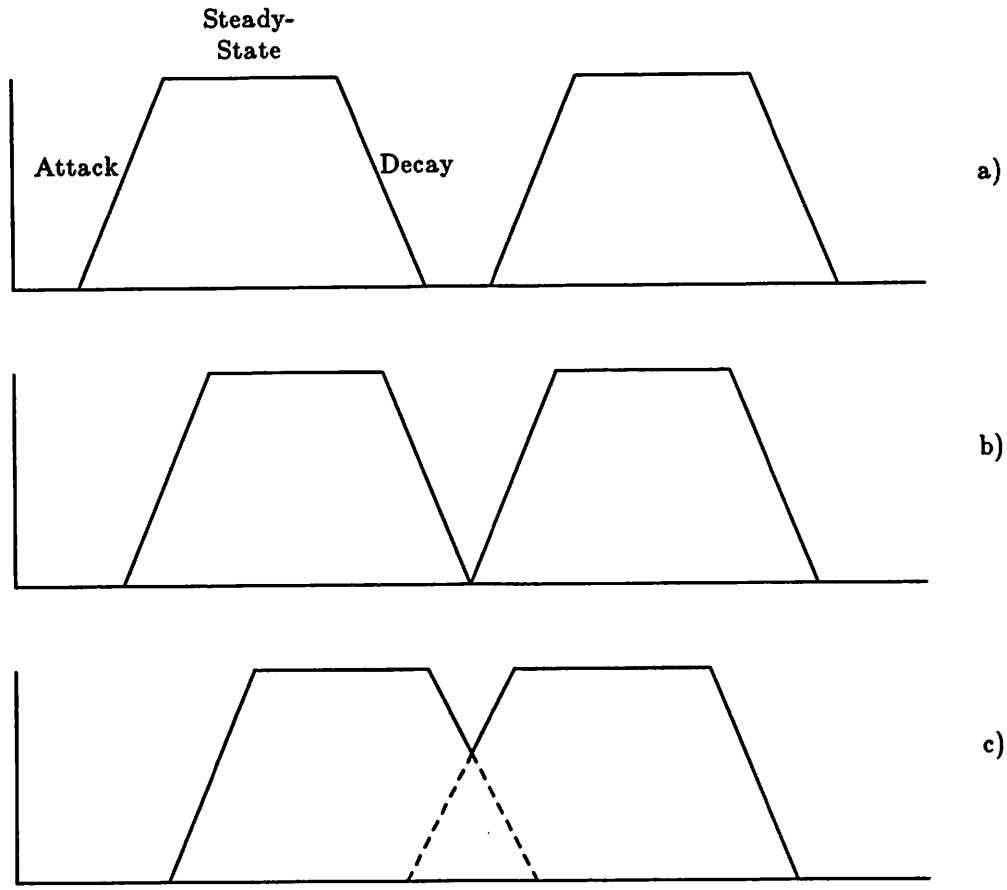


Figure 1.1. Possible methods for creating a transition between notes. a) No interaction occurs between the notes. b) The notes are abutted. c) The notes overlap.

The simplest way to model a musical line is to concatenate notes, as is the practice in many analog and digital synthesizers. There are several possibilities for connecting two successive notes: 1) there might be a gap between the decay of one note and the beginning of the next (Figure 1.1a); 2) the attack of the second note might start right when the decay of the first note finishes (Figure 1.1b); or 3) the decay of the first note might overlap the attack of the second (Figure 1.1c). This crude analysis helps refine the definition of transition given above. A *transition* includes the ending part of the decay of one note, the beginning and possibly all of the attack of the next note, and whatever (if anything) connects the two notes. It will become clear that a transition typically lasts on the order of tens to hundreds of milliseconds, i.e., a few tenths of a second or less.

Playing Methods vs. Perceived Articulations

Learning to control the juncture between notes is part of the training of a professional musician. It is not adequate to simply play notes one after another, as the model of the previous section would suggest. Successive notes must be purposefully joined in one manner or another. Inertia in the instrument plays a role too. In other words, the new note does not begin by itself; it must be helped along.

Articulation in Orchestral Instruments

Wind and brass players are taught the technique of "tonguing," in which a syllable such as "ta" or "da" is "spoken" inside the mouth right as the new note begins. Use of this technique is optional: The wind or brass player can start a note with no tonguing at all. There is a whole range of intermediate stages between tonguing and no tonguing.

The string player likewise has the choice of continuing to move the bow in the same direction when starting a new note, or changing bow direction. This is familiar to anyone who has watched a string section closely. In many ensembles, the strings bow together, following instructions from the section principal. Here, too, there are other options, such as varying the velocity of the bow while keeping it moving in the same direction.

Sometimes the composer gives explicit directions about how to play a passage. In general, when one slur joins several notes in a score, they are all played with one bow stroke on the strings, and are played without tonguing on the winds. However, in later practice, such as in Wagner, one very long slur indicates the desired *phrasing* rather than the bowing. Sometimes lighter tonguing or bowing is indicated with dots or dashes under a slur; see, for example, Seagrave and Berman (1976); Blatter (1980, p. 71); Adler (1982, p. 148); and Burton (1982, p. 2).

But such indications are often missing, or are changed by the performer. When to use these techniques, and how much separation to allow between notes, is a matter of applying long training seasoned with good taste to the particular musical passage at hand. For example, Forsyth writes: "In figures and melodic phrases the *legato* of the Flute is practicable throughout its entire compass, and is limited only by the capacity of the player's lungs. It is, however, more in the character of the Flute to break up a long series of notes into its component parts, to interpolate detached notes and groups of notes, and generally to substitute a vivid manner of performance for its more rapid *legato* style. [...] In single-tongued phrases the Flautist achieves a consonantal clearness and

distinction that is more akin to speech-in-song than to anything else. Each note can be made to sound like a falling hailstone or an unstrung pearl." (1936, p. 191)

There is a long tradition of written works on performance practice, most of which are not quite so poetic as Forsyth's. Playing manuals, even in the Renaissance, specifically discuss articulation (for an overview, see Donnington [1963]; Carse [1925] gives a good history of orchestration). For the purposes of the present study, it will be adequate to consider performance practice in modern orchestral instruments. It is generally accepted that the first modern treatise on orchestration is that by Berlioz (1948), originally published in 1843. His work was revised and expanded by Strauss, whose comments will also be cited here. Major works on orchestration since then have included those by Rimsky-Korsakov (1922), Forsyth (1935), Piston (1955), Kunitz (1961), and Kennan (1970). I also consulted the works of Bussler (1879), Humperdinck (1892), Kling (1905), Gilson (1922), Teuchert and Haupt (1924), Andersen (1929), Casella and Mortari (1950), Jacob (1962), Mancini (!—1962), Rauscher (1963), Blatter (1980), Del Mar (1981), Adler (1982), and Burton (1982). I did not find it necessary to consult playing manuals for individual instruments, for which Blatter (1980) gives especially good references; the works on orchestration provided adequate information on modern practice.

These books are filled with much information that is not relevant to the current study. There are, for example, detailed discussions of the ranges of instruments, which trills are difficult on a given instrument, when to use chromatic passages, when the high and low ranges of each instrument can be effectively exploited, and the like. It is of interest, however, to examine what the authors say about the techniques of tonguing and bowing and the impressions they evoke.

Tonguing: Many of the manuals (Berlioz is a notable exception) discuss when to use tonguing. In general, tonguing is appropriate when the note is to be set off slightly from the preceding note, or for fast passages. (There is even the technique of double- and triple-tonguing to make fast passages easier; Forsyth [1935, p. 98] has a good discussion.)

However, the situation is not quite so simple. First of all, the player is not limited to a choice between tonging and non-tonguing. Blatter writes (1980, pp. 71–72), "When placed over a series of notes, all of which are under a slur, tenuto marks [short horizontal dashes] indicate that the phrase is to be played legato, that is, as connected articulation in which each note is slightly stressed but no discernable [*sic*] separation is heard between notes." Thus, there is not necessarily a relationship between the technique used to produce the articulation and the impression that articulation evokes; the listener might or might not hear the tonguing. Blatter goes on to say, "Often the tongue is used to produce the stress, but due to the soft, quick stroke involved it may

not be perceivable, and for this reason the articulation is sometimes called *legato tonguing*." This is an important point for the current work, and will receive careful consideration here.

Concerning the trombone, Forsyth (1936, pp. 138–39) writes in a similar vein: "There is in general no true *legato* on the trombone at all. Each note has to be articulated at the moment that the slide changes its position. There is therefore a perceptible moment between each two notes when the air-column is not in vibration. Were this not so we should get a distressing *portamento* between the notes. In the *p* and the *pp* a good player reduces this moment-of-silence to [a] vanishing point, and produces an effect which Widor happily compares with the *sostenuto* of the Violin—a *sostenuto* which is scarcely interrupted by the turning of the bow. When playing a succession of notes from the same fundamental, that is to say, when playing without change of slide-position, the Trombone, of course, enjoys the same advantages of *legato*-playing as any slideless Brass Instrument." Blatter (1980, p. 126) writes along the same lines, and specifically says that a trombonist can create a perceived legato, even with tonguing. Likewise, Adler (1982, pp. 288–89) praises "[t]rombonists [who] have perfected the coordination of soft-tonguing with change of position to give an almost perfect impression of legato playing."

Bowing: As for the strings, Berlioz writes that "The manner of *bowing* is very important and greatly influences the sonority and expression of motives and melodies. It must be carefully indicated according to the nature of the idea to be rendered." Strauss adds, much in the same vein (p. 20), "For composers it is very important to consider carefully the problem of up-bows and down-bows when they want to achieve certain nuances."

As with the winds and brass, the player has a broader choice than simply whether to continue to bow in the same direction. Berlioz lists 5 kinds of bowing (détaché, slur, extended slur, staccato, *grand détaché porté*). Kling (1905, pp. 3–5) lists 13; Andersen (1929) gives 5; Forsyth (1936, pp. 339–47) has 8; Burton (1982) discusses 6; Adler (1982, p. 19) enumerates 7 and treats pizzicato and playing with or without a mute in the same breath; and so on. In short, there is no general agreement on the way of classifying kinds of bowing, but there are certainly many shades at the player's disposal.

As with the wind and brass instruments, the relationship between playing technique and perceived articulation is not always straightforward. Blatter (1980, p. 28) warns that "[t]he bowings illustrated [...] do not provide any information relating to *articulation* (i.e., the manner in which a note begins and ends). It cannot be ascertained from the locations of up-bow or down-bow, nor from the note groupings within a bow-stroke, whether or not the notes are to be staccato or legato."

Also writing about the violin, Forsyth (1936, p. 337) says: "When the bow is turned, that is to say, when a stroke is to be made in the opposite direction, the pressure of the fingers on the bow-stick is relaxed. In other words, the bow is 'lightened.' At that instant the bow is turned and a second (reverse) stroke is begun [...] It must be said that, though at the moment of lightening the bow the pressure is completely removed and no sound can therefore come from the instrument, this moment can be made by the most ordinary fiddler so minute in duration as to be inappreciable to the ear."

Burton (1982, p. 28) is even more explicit when he writes that "[...] the string player can change from one bow direction to another with no perceptible break in the phrasing. Therefore, even if a string passage contains many changes of bow direction, the effect can be of a continuous legato." In the same vein, Adler (1982, p. 19) says that "[e]ven though changes of bow direction occur between each one of the notes [...], one does not necessarily perceive these changes, since skilled performers can play the successive notes without a break or any audible difference between up- and down-bow." An experienced conductor whose ears and judgment I trust likewise warned me against the notion that a certain kind of articulation implied a certain kind of playing technique (Wyss 1984).

Thus, for both the wind and the string instruments, the perceived articulation is not necessarily a clue to the playing technique. (Even computer-based analysis systems experience the same sorts of confusion; Galler and Piszczalski [1978] report that their system would occasionally fail to detect "legato-tongued notes on the same pitch.") Furthermore, the fine spectral and amplitude clues for a given playing style are most likely to be buried in the sound of a whole orchestral section, except in the case of purposeful articulation, and are also likely to be inaudible to a listener in a concert listening situation due to the great distance between player and listener.

The ambiguous relationship between playing technique and perceived articulation has important implications for the current work, as will become clear in Chapter 2.

Analysis Techniques

Let us now turn our attention to the question of how instrumental sounds, once recorded, can be analyzed.

If the *Problems* of (pseudo-)Aristotle can be accepted as authentic, then enquiry into the makeup of sound can be traced to at least the ancient Greeks. Certainly that text was a major

source of inspiration for medieval and Renaissance scientists, who began modern enquiries into the nature of sound. Miller (1935) and Hunt (1978) give a fascinating account of the development of knowledge about the nature of sound from antiquity to the present. This history need not concern us here. Rather, it will suffice to examine modern methods. We will therefore consider (very briefly) time-invariant Fourier techniques, then the time-varying Fourier techniques which form the backbone of the current work, and finally several other relevant techniques not used here; the reasons will of course be discussed.

Analysis of Steady-state Tones

Modern work on the makeup of sound relies heavily on mathematical techniques developed by Fourier (1888, Vol. 1) in his research into the propogation of heat. His is a method for breaking a complex waveform into a number of harmonically related sinusoidal components. Fourier techniques are well understood (see, for example, Bracewell 1965). It will be adequate to simply state the equations defining the technique to pave the way for the discussion in the next section.

Assume that $f(x)$ is a periodic function. The Fourier transform of $f(x)$ is given by

$$F(s) = \int_{-\infty}^{\infty} f(x)e^{-jxs}dx \tag{1.1}$$

where $j^2 = -1$. The existence of $F(x)$ is subject to certain conditions which need not concern us here. It is possible to recover the original function from its Fourier transform by means of the inverse Fourier transform

$$f(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} F(s)e^{jxs}ds.$$

This is the mathematical basis for the synthesis technique known as additive synthesis.

In the discrete world of sampled (musical) sounds, corresponding transforms fortunately exist. If $x(n)$ is a signal N samples long, then the discrete Fourier transform (DFT) is given by

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}nk}, \tag{1.2}$$

where k is an index into the N (complex) transform points produced at frequencies evenly spaced from 0 Hz to the sample rate f_s. To recover the original signal, one may apply the inverse discrete Fourier transform

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{j\frac{2\pi}{N}nk}.$$

The DFT can be efficiently implemented with the fast Fourier transform (FFT) (Rabiner and Gold 1975).

Mechanical, electrical, and electronic devices were built in the 19th and 20th centuries to calculate the Fourier transform. For example, in his classic study of the composition of musical tones, von Helmholtz (1913) built a series of what are now called "Helmholtz resonators:" glass or metal spheres with two openings. Sound is admitted to one opening; the other sits in the ear (von Helmholtz used sealing wax to make the connection airtight). Resonators can be built that are tuned to the frequencies of the harmonics of a note; careful listening with various resonators can permit a coarse analysis of the relative amplitudes of the harmonics.

In 1931, Meyer and Buchmann published an astonishingly thorough study of steady-state spectra.* Recording sounds with a microphone, they used a tube rectifier to produce a side band by beating a reference wave against the waveform. For each harmonic, a different reference wave was chosen. This meant that the note had to be played once for each harmonic. As they say (p. 740, my translation), "The note was repeated, depending on the instrument, about once per second; the demands on the performer were thus quite large. The lower brass instruments, which require a lot of air, caused a few difficulties." Of course, since a different note was played to analyze each new harmonic, the results (as we know now) have to be viewed with caution. At any rate, they recorded the amplitude of the first-order lower sideband produced in the rectifier. In fact, careful reading of their p. 740 and their Figure 6 shows that two sideband peaks were recorded; the average of those two was then taken. In all, the measurements which they made are remarkable, but cannot be considered reliable for modern work.

Time-varying Analysis Techniques

By the early part of this century, it became evident, as will be discussed shortly, that musical tones varied significantly with time. It was thus not long before certain researchers attacked the problem of time-varying analysis.

Another astonishing study of musical instrument tones appeared in 1932, this one by Backhaus. He apparently recorded tones over a condensor microphone onto a recording device built by a Mr. H. Gerdien. The recording was captured on a drum turning on a spindle. Apparently (pp. 32–34) he then performed Fourier analysis on a period-by-period basis, using what I presume were the mechanical means (Henrici analyzer) available at the time.

---

*I am indebted to Mr. Folkmar Hein of the Technical University Berlin for making a copy of this article available to me.

Luce (1963) outlined a technique in which discrete Fourier analysis was applied at every period of a digital recording of a musical note. He encountered various difficulties with this technique, but was able to use it for his work. In the very first preliminary studies for this thesis, I tried using a similar method, but found it impractical for applications to the transition between notes, especially as it is difficult to delineate where the periods from one note stop and those from the next note start (this will be discussed more in Chapter 2).

Rösing's thesis (1967) was based on recordings taken from phonographs, which were analyzed with the Kay sonograph, "as other means were not available" (p. 22, my translation). The work by Rösing has lapsed into an obscurity which it does not deserve; Cogan (1984), for example, seems to be unaware of Rösing's work.

The Heterodyne Filter: This digital technique for analyzing time-varying spectra was pioneered by Freedman (1965, 1967, 1968) and has also been used extensively by Beauchamp (1969). Basically, each spectral component is "pulled out" of a complex signal by "heterodyning" the signal by a sine and a cosine at the frequency of interest. The products are filtered and converted to produce time-varying amplitude and frequency plots. Moorer (1973, 1977) gives a good review of this technique; Moorer's implementation was used by Grey (1975). Moorer (1975) also points out out some pitfalls and how to avoid them. Beauchamp (1981) prepared a tracking version of the heterodyne filter, which foreshadows Dolson's work with the phase vocoder (to be discussed below).

Gish (1978) presented criticisms of the heterodyne filter, and suggested a model which explicitly included inharmonicity. After all, Fourier analysis claims to capture only harmonically related components, although information on inharmonicity might be included in the Fourier *representation.* Thus, Gish added a noise term to his synthesis technique. I have to agree with Gish when he writes that he found "many more components than has been previously reported," if he is discussing early studies from the 1960s. But I disagree when Gish criticizes Grey along these lines; the number of harmonics found by Grey was not that much smaller than the number of harmonics which I used in my work (see Table 3.5).

As for other precursors of the phase vocoder, to which we will then turn, it should be mentioned in passing that in his work on violin tones, Beauchamp (1974) prepared a "line spectrum movie," showing the evolution of a violin spectrum; the development of a 1-sec note lasted about 20 sec. This is also the place to mention the methods developed by Freedman (1967, 1968), in which Fourier analysis is converted into a set of functions called a *non-Fourier representation.* The amplitudes of the harmonics are broken down into a series of exponential terms. This technique,

which was used to analyze notes from musical instruments, is not the same as the heterodyne filter, as some have assumed.

The Phase Vocoder: The term *vocoder* was coined from VOICE CODER, the name of a device designed to reduce the bandwidth needed for satisfactory transmission of speech over phone lines (see Dudley 1939). The idea was to pass the speech signal through a set of contiguous bandpass filters, such that the combined output of these filters at a given point in time would be a rough approximation of the spectrum of the signal. In theory, by transmitting a few filter coefficients, a savings could be achieved in terms of the transmission bandwidth required to transmit a given signal (see Schroeder 1966).

In practice a savings was not possible, because too many channels were needed to preserve speech quality. There was an additional problem in that only the magnitude of each filter output was being transmitted; phase information was thrown out. Thus, the speech resynthesized from the encoded version was never identical to the input, regardless of the number of channels.

The *phase vocoder* (Flanagan and Golden 1966), developed as an extension of the original vocoder concept, preserved phase information, allowing the input and output of the system to be identical. Schafer and Rabiner (1973), Portnoff (1976, 1978, 1980) and Holtzman (1980) improved the technique, with the result that the speed was increased (in terms of computation time), while still allowing the synthesized output to be identical to the input. In particular, Portnoff (1976) showed how to implement the phase vocoder with the FFT (see also Allen 1977a, 1977b; Allen and Rabiner 1977; Crochiere and Rabiner 1983). Portnoff (1978), Holtzman (1980), and Dolson (1983) developed a method of scaling the original in time without modifying its pitch or spectrum, which inspired a cruder method which I used (see Appendix 2).

The analysis side of the phase vocoder, which is technically known as the short-time Fourier transform (STFT), is defined by

$$X(n,k) = \sum_{m=-\infty}^{\infty} x(m)h(n-m)e^{-j\frac{2\pi}{N}mk}. \tag{1.3}$$

A new variable m has been introduced into Equation 1.2; m is a dummy variable which allows for the filtering operation with $h(n)$. In other words, $x(n)$ is windowed with the low-pass filter $h(n)$, on which there are certain restrictions. The spectrum $X(n,k)$ at point n is divided into N frequency bands equally spaced from DC to f_s and indexed by k. The quantity k is thus the index for the "channel" number in the phase vocoder. The original signal can be recovered with the

inverse short-time Fourier transform

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{m=-\infty}^{\infty} f(n-m) X(n,k) e^{j\frac{2\pi}{N}mk}$$

where $f(n)$ is also a filter.

When one is working with musical sounds, the idea is to align the analysis channels so that no more than one harmonic of the signal falls into a given channel. If $a(n)$ and $b(n)$ are the real and imaginary outputs of a channel, respectively, then the amplitude $A(n)$ of the harmonic in the channel may be recovered as follows:

$$A(n) = \sqrt{a^2(n,k) + b^2(n,k)}.$$

The instantaneous phase $\phi(n)$ is given by

$$\phi(n) = \arctan\left[\frac{b(n)}{a(n)}\right].$$

The frequency, calculated as the derivitave of the phase, is then given by

$$\frac{a(n)\frac{db(n)}{dt} - b(n)\frac{da(n)}{dt}}{a^2(n) + b^2(n)}$$

Details of this conversion process are given in Moorer (1978) and Gordon and Strawn (1985). This is the system that I used for the work described in Chapters 2 and 3 and Appendix 3.

To prepare for the discussion of Chapter 3, it is necessary to point out here that the spectrum $X(n,k)$ is not calculated for every sample of the original signal. Rather, every R points can be skipped; there are restrictions on the relationship between R and N which do not concern us here. Also, it should be pointed out that choosing $h(n)$ is an important issue. A longer analysis filter gives better frequency resolution but with a correspondingly coarser resolution in time; and vice-versa.

Dolson (1983) developed what he calls a "tracking phase vocoder." The output of a pitch detection algorithm directs the phase vocoder; the fundamental frequency of the signal being analyzed can vary widely. On the surface, this would appear to be an ideal analysis method for research on transitions. Unfortunately, this work reached me after almost all of the phase-vocoder work described here had been completed. Since the phase vocoder proved adequate, the extra effort needed for implementing Dolson's version could not be justified. However, for future studies exceeding more than one note, it should be examined very carefully.

Models not found Useful, and Why

Spectrographic plots: Cogan (1984) used an analysis system which prepares spectrographic plots of musical performance (see also Potter and Teaney 1981). This is the digital equivalent to the study of Rösing mentioned above. For the current study, this method is inadequate—the resolution is simply too coarse. Indeed, the spectrographic representation of analyses of individual instruments has not proven to be useful except for the most general sorts of work (Strawn 1985a).

Acoustic Modeling: It is possible to model musical sound by a set of mathematical formulae which model the behavior of an instrument. Hiller and Ruiz (1971) reported on their work with strings. The possibility of modeling in this manner the behavior of a musical instrument during the transition between notes seems remote indeed at the present time.

Formant Models: Besides the model based on Fourier analysis, there is another very powerful model for analyzing and synthesizing sound, which is especially popular and useful in work on speech. A signal is modelled as a "driving function," which is usually quite rich spectrally, passed through one or more time-varying filters. The driving function can be the bowed string, the action of the lips in a brass mouthpiece, a reed, or the human vocal chords. The filters can be the resonances of the vocal tract or those of an instrument body. Smith's thesis on strings (1983) is the most recent work on this method, and provides a good review of this topic.

There are several reasons why I did not pursue this model. One is the difficulty of deriving the driving function and filter coefficients in the short time which some transitions take. Another is the question of modeling individual harmonics. It might be possible to model each spectral component with, say, two poles and two zeros (see Kay and Marple 1981). But what happens when the spectral components start moving in frequency? In order to obtain a pole-zero model in this case, there must be an upper limit on compute time used. But then the spectral lines might be spread into spectral regions. This means that the pole-zero model would yield wide-band resonances, which would not be useful in the transition. Also, analyzing the pole-zero model is difficult when spectral components disappear or (re-) enter, as happens in a transition.

Then there is the question of formants. There is no doubt that formants occur (see Strong and Clark 1967b; Luce and Clark 1967; and Clinch et al. 1982)* and synthesizing musical notes

---

*The anonymous performers in Clinch's survey are probably the first known victims of radiation poisoning in the history of psychoacoustics: they *gargled with a barium compound suspension before being X-rayed* while performing. Imagine.

with misplaced formants can certainly cause amusing results. Still, the role that formants play in the synthesis and perception of musical tones in musical contexts remains cloudy. More on this topic will appear later in this chapter.

The Constant-Q Transform: A number of experiments have demonstrated that some aspects of auditory perception are subject to what are known as critical band phenomena (see Plomp 1976 for an excellent review). Briefly, a critical band is a region, about a third of an octave wide; certain aspects of hearing can vary depending on whether the spectral components (or bands of noise) being tested fall within one critical band. This is true, for example, for aspects of the perception of loudness, phase differences, and some forms of masking. Even a listener's ability to "hear out" partials of a tone can be related to critical bandwidths. These phenomena neatly match the structure of the basilar membrane, where equal linear distances correspond to equal log-frequency distances. It should be emphasized that critical bands do not form a fixed set of filters; as Zwicker and Feldtkeller say, "The ear can form a critical band at any arbitrary point of the frequency scale" (1967, p. 73, my translation).

A mathematical technique called the constant-Q transform has been developed which models aspects of this critical band nature of perception (see Youngberg [1979]; Schwede [1983] discusses implementation details). The quantity Q is a measure of the quality of a filter, and is defined as the center frequency divided by the bandwidth of the filter. With the phase vocoder, the Q for each channel diminishes with increasing center frequency, as the filters are equally spaced in linear frequency, and the filter bandwidths are a constant number of Hertz. One can also construct filters whose center frequencies and bandwidths model *some* set of critical bands; the values of Q for these filters is a constant, hence the name given to the technique. Petersen (1980; Petersen and Boll 1983) applied the constant-Q transform to modeling the frequency-selective nature of the auditory system, especially in the suppression of noise (see also Callahan 1976; Kajiya 1979).

Stautner (1983) has developed a variant of constant-Q analysis which he calls the *auditory transform*. With this he has produced some interesting results based on principal components analysis of the filter outputs. Although that method is not used here, his analysis of a tabla recording is of special interest in this historical introduction (the tabla is a set of drums used in the music of the Indo-Pak subcontinent). Stautner was able to associate certain principal components with certain features of the signal, such as the onset of notes, the reverberation at the end of an earlier note, overall resonances of the drums, and specific drum strokes. However, he did not specifically attempt to model the transitions in his recordings.

I believe that the constant-Q approach with a set of fixed filters can become too simplistic in musical contexts, where the ear can separate two or more notes played together, even those played in unison. First of all, if critical bands are approximately one-third of an octave wide, and if the lowest band is centered around the fundamental, then more than one harmonic will fall into one analysis band *starting with the fourth harmonic*; given that the filters cannot have rectangular frequency responses, perhaps even the second and third harmonics would interact in one band. But if several spectral components falling into one band are reduced to one signal, then isolating instruments playing the same pitch would be difficult for the ear. To name one example, consider the continuo bass line in Baroque music; based on listening experience, I am convinced that it is possible for the listener to separate the bassoon or cello playing in unison with the organ or harpsichord. The role of critical bands in perception of polyphony is even harder to imagine. Certainly some *time-varying* information, such as coordinated vibrato in all of the spectral components of one note, can and does pass through critical bands to provide the basis for identifying sources. (I am assuming here that critical band phenomena occur before spectral components are grouped into virtual sources, one for each instrument playing.) Otherwise the effects demonstrated by Chowning (1980)—in which a voice "crystallizes" out of mass of seemingly unrelated spectral components when vibrato is applied to a subset of them—would not occur. This implies that separate spectral components falling within one critical band can still be resolved by the ear for some purposes. Thus, I decided that it was premature to try to apply constant-Q analysis to transitions.

Be that as it may, examination of the spectra of several hundred notes has lead me to believe that some way can be found to gather certain harmonics, especially higher-order ones, into broader groups than those which I have used here. Typically, the amplitude envelopes of the higher harmonics—say, after the 24th or so—have an irritatingly similar shape. (Risset [1966] reports on grouping harmonics, with a single amplitude envelope provided for each group.) More work definitely needs to be done on this topic. A set of timbre experiments using constant-Q analysis along the lines of Grey's work (1975) would be a big step in the right direction—and just might prove me wrong!

The Wigner Transform: This technique, given a thorough treatment by Claasen and Mecklenbräuker (1980), produces a time-varying representation of the spectrum of a signal which can be windowed to provide on-the-spot control, so to speak, of frequency or time resolution. It did not prove necessary to invoke this technique for the work described here.

Other models: Models of sound are implied by a number of synthesis techniques which have become popular in computer music. For example, there is VOSIM (Kaegi and Tempelaars 1978), CHANT (Rodet 1984; Rodet et al. 1984), frequency modulation (Chowning 1973; LeBrun 1977; Schottstaedt 1977; Justice 1979), waveshaping (Arfib 1979; LeBrun 1979), granular synthesis (Roads 1978), discrete summation formulae (Moorer 1976), and the like. I did not undertake to analyse transitions in terms of these techniques. Experience has shown that results from analysis of the generalized additive-synthesis case can be applied to other synthesis techniques.

Analysis of Musical Instruments

The previous section introduced a variety of analysis techniques. Let us now turn our attention to studies which have applied those techniques to single notes of orchestral instruments.

Steady-state Tones

The work by Helmholtz is the direct modern ancestor of the current study. In his analysis of musical instrument tones, he explicitly excluded any time-varying parts, although he realized that they play a role in musical perception (for a discussion, see Strawn 1982). Following Ohm (1843), he decided that the ear performs a Fourier-style analysis of an instrumental tone, breaking down the complex waveform into sinusoidal components. The number, frequencies, and relative amplitudes of these harmonics determine what he called "musical tone color" (*musikalische Klangfarbe*), meaning the timbre of a steady-state sound. In general this model and the controversies surrounding it dominated research into musical sound for the next century or so.

Meyer and Buchman (1931) analyzed vowels, several pianos (also with varying dynamics, with and without dampers), a *Hammerklavier*, harpsichord, clavichord, electric piano, harp, zither, lute, banjo, all the orchestral strings, two kinds of flutes, piccolo, bass flute, two clarinets, tenor and alto saxophone, contrabass bassoon, bassoon, English horn, oboe, oboe d'amore, some organ pipes, the usual brass instruments, timpani and various drums, cymbals, tam-tam, castanets, triangle, glockenspiel, xylophone, tubular bells, and singing saw (!). They used the entire note in their analysis, but gave only steady-state (line) spectra. This was remarkable work at the time, and has often been cited (e.g., Winckel 1960).

Von Bismarck (1974a, b) created 35 different signals, some based on filtered noise, some based on (filtered) spectra of periodic tones. To measure the timbres of these signals, he used 30 verbal

scales such as "smooth-rough" or "compact-scattered." Test subjects were asked to rate each of the test tones on each of the scales. This is probably the definitive study on the perception of steady-state spectra.

Time-varying Tones

Backhaus (1932) conducted analyses of the transients in various musical instruments (and speech), using the improbable apparatus described earlier. He showed (p. 40) that playing just the steady-state portion of the recordings led to confused identification. Identification was improved if such features as vibrato, which he found to be characteristic for each instrument, were included. He examined the attack times of each instrument analyzed, defining the end of the attack as that point at which energy of the instrument "essentially no longer rises." (p. 40, my translation). He found that this attack time was regular for a given instrument across several recordings, and therefore considered it to be a characteristic of the instrumental sound. Now in all of his recordings, the instruments started playing from silence; Backhaus assumed that the attack time was quicker when the instrument was already playing some other note than when the instrument started from silence. Examining the manner in which different partials enter, he found patterns in those entrances which he thought to be significant; he assigned verbal labels to the attacks, based on these patterns. In general, this was pioneering work, and has also been cited frequently.*

LeCaine (1956) wrote an interesting overview of the electric and electronic music at the time. He emphasized the time-varying nature of musical sound, and showed (p. 465) the attack waveforms for two different playing styles—on an organ. He also discussed how the attack of a note might happen on a monophonic instrument.

Winckel (1960) devoted an entire chapter to a discussion of the time-varying characteristics of musical sounds. Much of what he presented is based on speech or non-orchestral instruments. Unfortunately, his discussion of instrumental sounds seems to be based on the work of Backhaus. Still, this is a good review of the state of the art at the time.

---

*At this juncture I would like to correct an error in a book review which I wrote, and in the book itself. In my review of Wayne Bateman's *Introduction to Computer Music*, published in *Computer Music Journal* 5(1), 1981, I pointed out on p. 71 that the illustrations printed by Bateman (his pp. 80–81) were attributed by Bateman on p. 79 to Meyer and Buchmann. I questioned whether such dated work was reliable, and pointed out that more recent work lead to conclusions different from those reached by Bateman based on the figures. It turns out that the amplitude curves in Bateman's book are not from Meyer and Buchmann after all but rather from Backhaus, of which there is no mention in the text nor a citation at the end of the chapter. The trumpet plots on p. 80 of Bateman's book are from the lowest part of Backhaus' figure 20, p. 41, and the violin plots (Bateman, p. 81) are from the left-hand part of Backhaus' figure 25, p. 43. Of course my error was unfortunate, but as I pointed out in the review, at the time I had no copy of Meyer and Buchmann's article.

Luce (1963) analyzed 14 orchestral instruments, and presented data on the temporal evolution of the first eleven harmonics. He found a number of time-varying inharmonicities in the attacks and steady-states of some tones. His analysis also showed formant effects in some instruments. Luce's work, like that of Meyer and Buchmann, is a model of perserverance, and a monument to the amount of high-quality work which can be done in the face of inadequate equipment and mathematical techniques.

Luce and Clark (1965) measured the duration of attack transients, and found that they vary with pitch, performer, and instrument, but not with loudness. They define the attack time as spanning from the onset of the sound to -3 dB from the steady-state; but they do not specify how the beginning of the steady-state is determined.

Working with monophonic musical fragments recorded on tape, Risset (1966; Risset and Mathews 1969) digitized sounds at 10 kHz (12-bit accuracy). They used a quasi-Fourier analysis performed on a period-by-period basis, and found that the overall spectral curves of the trumpet vary significantly with amplitude. They also found that higher-frequency partials appeared later in the attack and disappeared earlier in the decay; also, the lower-order partials built up faster in the attack. Examining the amplitudes of the attacks, they found a wide variation in attack times, and observed further that the attack often proceeded in steps; I assume that the pre-attack noise which they found was tonguing noise. Looking at individual harmonics, they found what I call "blips" (their term was "hollows"), but could not determine the effect of blips in resynthesized tones. Most importantly, they created line-segment approximations of the amplitudes of partials, and synthesized tones from them which turned out to be quite good. Between successive notes, they found a frequency glide lasting about 0.05 sec.; they presented a few time-domain plots of some tongued transitions.

Nine instruments were analyzed by Strong and Clark (1967a). They found three groups of spectral components and discussed the spectra of each instrument in terms of these groups. Using a modified Fourier method, they synthesized tones on the basis of their analysis data, and found that some of the tones could be identified by some listeners.

In another study, Luce and Clark (1967) analyzed chromatic scales played on 14 instruments—but with a rest between each note. Two players were recorded on each instrument, for a total of 3100 (!) recordings. These analog recordings were digitized to 11 bits at a sample frequency of 20 833 Hz. The amplitudes of 11 harmonics were analyzed using a modified Fourier technique, in which each successive period was assumed to be stationary; they analyzed the attacks as well as the steady-states. Their results, based on my experience, were remarkably accurate. For example,

they found blips in the brass attacks. Furthermore, their observations on the relative entrance times of harmonics and on the change of frequency during an attack were consistent with what I observed.

Meyer (1972) approaches a dictionary of analyzed tones. For each instrument of the orchestra, he cites research on the instrument and discusses the instrument's dynamic range, overall spectrum, and attack characteristics. In discussing these attacks, he sometimes gives some general observations on the transition between notes, based on research done by others. In many cases he reproduces the time-domain waveforms of instrumental attacks.

Fifty-seven violin tones at a variety of amplitude levels were analyzed by Beauchamp (1974). He measured the attack and decay times for individual harmonics, and discussed line-segment approximations to those tones.

Meanwhile, some interesting work was conducted on the relative importance of attacks in instrument identification. Saldanha and Corso (1964) conducted what is perhaps the key work here. Risset (1966) and Wedin and Goude (1972) also performed studies on this topic (Grey 1975 gives a good overview). In general, these studies showed that listeners could identify an instrument on the basis of the attack; but listening to just a segment from the steady-state lead to confusion. From this work, it is commonly assumed that the steady-state and decay of a tone are less important in forming the listener's impression of timbre. This question will be re-examined in Chapter 5.

Grey's thesis (1975) was a turning-point in the study of the timbre of musical instruments, and is the direct precursor of the current work. Grey showed that the heterodyne filter was adequate for analysing time-varying musical tones, including their attacks and decays. Furthermore, he showed that the detailed microstructure of the amplitude and frequency traces could be omitted (using line-segment approximations) without significantly degrading the timbre of the tone. He did find that it was important to retain a certain independence in the frequency traces in order to avoid an "electronic" or "artificial" percept; and his work confirmed the importance of the attack in timbre. (Grey's work on categorical perception will be discussed in Chapter 6).

Returning to the question of delineating the attacks and decays of notes, Moorer (1977) presented waveforms and spectral analyses of several notes, and discussed the difficulties in finding the boundaries of the steady-state.

Charbonneau (1981) found that reasonable tones could be generated by simplifying Grey's line-segment approximations even further:

1. A normed amplitude curve was calculated from the amplitude curves of the various harmonics; for synthesis, this normed curve was scaled to reach the maximum

amplitude of the original amplitude curve for each harmonic. (This method was foreshadowed by Luce and Clark [1967].) Also, the original begin and end times for each harmonic were preserved.

2. The frequency trace for the fundamental was multiplied by the harmonic number.
3. The start and end times for the amplitudes of the harmonics were approximated by a polynomial.

Each of these was judged to be quite close but still discernibly different from tones generated using Grey's original line-segments (in some cases, Grey's cut-attack approximation was used). The timing simplification had the least effect on timbre, with the frequency and amplitude simplifications having increasingly greater effects. He also found that the different instruments reacted differently to a given simplification.

Physical Properties of Musical Transitions

> *The musical properties that fall into the general area of articulation have not been studied extensively as to their physical correlates.* (Howe 1975, p. 29)

Having examined previous work on individual notes, we can now complete this introduction by discussing a number of studies which have more or less directly examined the transition between notes.

Most analyses of the physics of musical instruments are concerned with how to *maintain* an oscillation and are therefore not much help here. One finds discussions of topics such as how the driving force interacts with the rest of the vibrating system, how the room reacts with the instrument, and the like. There is an emphasis on the more or less stable behavior of an instrument. It is typical that when works on the physics of instruments (Benade 1976 is one example) go over to discussing successive tones, they immediately switch to reverberation, or tuning and temperament, without discussing transitions.

If one considers musical sounds as sum of sinusoids, it might be useful to examine the behavior of a single sinusoid as it starts and stops. Benade (1976, pp. 153–56) analyses the initial transient of a system driven by a sinusoid as being the sinusoid plus an exponentially decaying sinusoid at some frequency which is a natural mode of vibration of the driven system. At the other end of a note, it is common to think of the decay of musical tones as a superposition of exponentially decaying sinusoids.

Benade also analyzes the startup of a brass tone in some detail. Of interest here are his comments that "[a]ssuming the player has buzzed his lips accurately for the desired note, the air column is happy to begin collaboration as soon as there has been time for the initial sound to make at least one complete round trip of the air column. Several more round trips are required before the regime of oscillation has set itself up completely. In a fast running passage, there is barely time for one regime of oscillation to be set up before it must give way to the next" (1976, p. 425).

Acoustical Studies Spanning more than One Note

Spectrographic Studies: In the largest non-computer study of musical sound that I have encountered, Rösing (1967) used the Kay Sonograph to analyze many orchestral instruments (violin, viola, cello, bass, flute, piccolo, clarinet, oboe, bassoon, trumpet, horn, tenor trombone, and tuba) as well as various percussion instruments. He also examined gagaku (a type of Japanese ensemble music), gamelan (the music of Indonesia), and other oriental works. All of his recordings were taken from phonograph records, with the individual instruments coming from DG 13910 (*Musikkunde in Beispielen*).

Rösing found characteristic attack times for each instrument, and discussed how his differ from those of Reinicke (1953) and Backhaus. Following the tradition of assigning verbal attributes to timbre, he gave qualitative judgments about various kinds of transitions: quick, raw, smooth, that sort of thing.

Examining the attack of notes, he found that many were "noiseless", by which he meant that inharmonic components were absent; for example, he found this to be the case with the flute. Now my experience leads me to conclude that there is considerable inharmonic activity in the attacks of notes, especially the flute. I attribute this anomaly in Rösing's data to the lack of resolution in the Kay Sonograph. Rösing noted further that a dip in amplitude in the middle of a note is simultaneously accompanied by a spectral rolloff. The amount of rolloff that he observed there is surprising. Again, I attribute this artifact in the analysis output to the limitations of the Kay Sonograph; in this case, I believe that the limited dynamic range of the analyzer caused some weak higher-order harmonics to simply disappear.

Even more interesting for the current study is Rösing's analysis of musical transitions. In contrast to earlier work, he felt (pp. 27–28) that "in the course of musical events the attack (i.e., the start of sound from a state of rest) and the decay (i.e., dying out to a state of rest) [...] have little meaning. In the melodic continuum there is rarely a complete attack or decay; rather, there

are places where the pitch changes, i.e., [places where the instrument] changes from one state of vibration to another. Backhaus discussed this, and presumed 'that the change to the new state of vibration happens then [i.e., in a transition] more quickly than when the sound is generated from rest.' This assumption was in general confirmed" in his work.*

Continuing his analysis, Rösing found three phases in a transition (p. 28):

1. Phase of preparation for the change in pitch.
2. Phase of the transition between pitches.
3. Phase of the formation of the (next) note.†

Phase 1 occurs at the end of the decay of one note; phase 3 at the beginning of the attack of the next, with the three phases perceived as a single unit. He found that in each instrument the total duration of the transition as well as the relative lengths of individual phases had distinctive, "near constant" (p. 28) characteristics. He decided that these affected the subjective impression of the transition, mentioned earlier.

In general, he found that the spectrum of the steady-state was not affected by the changes in the spectrum during the transition. Contrary to the observations of Raman (1918) and Backhaus, he found that in all instruments the strongest components always enter first, to be followed by the weaker; and the weaker components are the first to decay. Reverberation might be in part responsible for this in the decay. He found that reverberation made it difficult to decide where the exact point of transition falls: if the second note is separated right at a supposed transition point, and then played, the end of the first note would still be heard in the reverberation left in the recording. Be that as it may, he found in the transition "a wedge-like funnel, open toward the higher frequencies," in the spectrum at the transition (p. 36).‡

Turning now to specific instruments, Rösing found four kinds of transitions in the violin:

---

* "Innerhalb des musikalischen Geschehens haben [...] die Ein- (Anklingen aus dem Ruhezustand) und Ausschwingvorgänge (Ausklingen bis zum Ruhestand) wenig Bedeutung. Im melodischen Kontinuum kommt es nur selten zum vollen Ein- bzw. Ausschwingvorgang, sondern zu Tonwechselvollzügen, d.h. dem Wechsel eines Schwingungszustandes in einen anderen. Darauf hat bereits Backhaus hingewiesen und vermutet, 'daß dann der Übergang in den neuen Schwingungszustand schneller erfolgt, als wenn der Klang aus der Ruhe heraus erzeugt wird.' Diese Annahme wird generell bestätigt." The translations of the next few German passages are all mine.

† 1. Phase der Tonwechselvorbereitung; 2. Phase des Tonüberganges; 3. Phase der Tonbildung.

‡ "Bei fast jedem Tonwechsel verschwinden die höheren schwachen Teiltonkomponenten des Spektrums, und zwar um so eher, je schwächer sie sind. Entsprechend dauert es auch um so länger, bis sie nach dem Tonwechselvollzug wieder auftreten. Es entsteht ein keilartiger, zu den hohen Frequenzen hin geöffneter Trichter."

1. *Legato.* As an example of a legato transition, he cited an excerpt from m. 55 of the first movement of the Beethoven *Violin Concerto*, in which "[a] small reduction in intensity and a lessening of the vibrato (if any is present at all) occur before the change in pitch. The transition to the new note is seamless; the new note arises directly from the old, and reaches the required pitch only after about 0.004 sec. The total amplitude [during the transition] varies only slightly." (p. 55)*
2. *Glissando*, a transition whose nature is implied by its name.
3. *Normal.* I assume that this is a transition with bow change. Rösing says that in this kind of transition, "[t]he notes are clearly separated from each other; the change in pitch happens with a jump. Usually the preparation for the change in pitch manifests itself in the form of a reduction in intensity, which results in the removal of the weaker spectral components and a light reduction in the overall amplitude." (p. 35)†
4. *Staccato*, i.e., total separation of the notes.

Examining his analyses of the trumpet, Rösing found differences in the amplitude dip between notes for different kinds of articulation. For a legato transition, he found that the total amplitude remained almost constant (p. 87). However, he found a large jump in amplitude associated with what he called the "normal attack," which I interpret to be the tongued manner of performance.

The computer-based successor to this study is that by Cogan (1984), which reached me after my work was finished. Although I find many of his comments about the nature of music to be overextended, his work needs to be mentioned here. The book presents spectrographic plots of entire musical works. In such plots, the resolution is not adequate for analysis on the detailed scale attempted here. Even though Cogan does not discuss the transitions between notes in general, he points out a certain correspondence between playing styles (e.g., bowed vs. pizzicato) and the overall shapes on his spectral plots. Still, his work is of relevance to this study because, in some plots (e.g., in the Webern example, p. 63), what appears to be pitch jumps between notes can be seen.

---

*Vor dem Tonwechsel tritt eine schwache Intensitätsreduktion und Dämpfung des Vibratos (sofern vorhanden) ein. Der Übergang zum neuen Ton vollzieht sich nahtlos, der neue Ton geht sachte aus dem alten hervor und erreicht erst nach etwa 0,004s. die eigentlich geforderte Tonhöhe. Die Gesamtamplitude unterliegt nur leichten Schwankungen.

† "Die Töne sind deutlich voneinander abgesetzt, der Tonwechsel geht sprunghaft vor sich. Die Tonwechselvorbereitung macht sich meist in Form einer Intensitätsreduktion bemerkbar, die den Abbau der schwächeren Teiltonkomponenten und ein leichtes Schwanken der Gesamtamplitude bewirkt."

Timbre in Musical Contexts: Using the clarinet, trumpet, and bassoon, Grey (1978) studied the influence of musical context on the perception of timbre. Test tones were synthesized from complete heterodyne filter analysis data, or from line-segment approximations. The sets of resynthesized tones were optionally changed in the middle of a musical passage consisting of one or more lines. Listeners were asked to determine if the two halves of the passage were played by the same instrument. In general, Grey found that with the clarinet and trumpet in polyphonic contexts, the judgements were not as reliable as those for monophonic settings; the results for the bassoon were different. The resolution of that dilemma is not of interest here; but his results do serve to point out that detailed differences in timbre which may be audible in isolation may well be lost in real musical contexts.

The Legato Transient: Let us return, then, to the importance of the attack already mentioned in the discussion of single-note studies. Campbell and Heller (1978) analyzed analog recordings of six instruments (clarinet, flute, oboe, piano, trumpet, and violin) playing the interval F349 to A440. In addition to the usual attack, steady-state, and decay sections of a note, they identified a legato transient, defined as "the transition between two notes in a legato passage played on a continuous tone instrument. It is initiated when the performer interrupts an existing standing wave and ends when a new standing wave has been established" (p. 1). The four regions (attack, steady-state, decay, legato transient) were spliced apart with voltage-controlled amplifiers and triggers from a Moog synthesizer. They found that subjects could identify instruments better from the legato transient than from using the attack transient alone. Although their work does not contradict what one would expect, it should not be accepted without reservation, due to the possible limitations of their equipment. (The work of Cutting and Rosner [1974] ultimately suffered from the use of the Moog synthesizer as a signal processor—see Chapter 6).

Registers: The role of registers is closely related to this line of research.

For example, the clarinet range is traditionally divided into several registers with names like chalumeau and clarino. Benade (1980) discusses the register changes in the clarinet, and presents a recorded example of a skip in registers using an interval of a twelfth. Limacher (1979) analyzed eight clarinet tones and came to these conclusions (pp. 16–17):

> The lower even-numbered partials are consistently weaker than the corresponding odd-numbered partials in the chalumeau register. The second partial is especially weak in the notes of the chalumeau register. However, in the clarino and extreme high registers,

the second partial becomes relatively strong. The higher even-numbered partials don't exhibit any consistent relationship to the odd-numbered partials. They are sometimes weaker, sometimes stronger and often just as strong as the odd-numbered partials.

As to the difference between registers, I can see no sharply defined distinctions. The chalumeau register appears to exhibit fairly strong first and third partials, along with a very weak second partial. The only generalization I can derive from the two clarino register tones is that the second partial is stronger than in the chalumeau register. The extreme high register exhibits a very strong first partial. The second partial is stronger than the third. There are few partials of consequence in the extreme high register.

Working with the mezzoforte recordings of isolated clarinet tones prepared at IRCAM, I examined Fourier analyses of the steady-state of clarinet tones from D3 through A♯6, at every chromatic step. There are certainly differences in the spectral envelopes between the high and the low ends of the clarinet range. However, I did not find any noticeable change in the spectral envelope at or near the traditional boundaries for the clarinet registers. Perhaps this is a measure of the clarinettist's success in bridging these register boundaries.

In the current study, I did not encounter difficulties due to a change in registers on the clarinet or any other instrument, and so this topic will not be considered further here. Still, more work needs to be done on this question.

Reverberation: As for the effect of reverberation, Meyer. (1972) discusses the performance of articulation in a real hall, and presents (pp. 210–11) a plot showing how even in staccato violin passages (Mozart *Symphony* K. 319, first movement) the amplitude between the notes at the back of the hall can be smoothed by reverberation.

Auditory Streaming: Considerable work has been done on the question of how the auditory system groups simultaneously sounding spectral components; McAdams and Bregman (1979) give a good review. In general, streaming can be asumed to happen for the cases examined in the current work; "degenerate" cases (not meant in a pejorative sense), such as those covered by Steiger and Bregman (1981), are not likely to occur and were in fact not encountered. Thus, a close examination of issues in streaming is not necessary here. It is of interest to note that streaming is more likely to occur when the pitch trajectory at the end of one note matches the pitch trajectory at the beginning of the next note; I have found some evidence that this occurs almost naturally in performance on orchestral instruments (see the frequency plots in Appendix 5).

Melodic Studies: There is a large body of work on the performance and perception of melody, some of it dealing with higher-level cognitive or aesthetic issues and none of which will be dealt with here, as it is not relevant for the current discussion of transitions. The contributions in (Deutsch 1982) provide a good starting point for those interested in this area.

Seashore (1938, pp. 200–203) reproduced analyses by Small of two performances of *Ave Maria* on violin. Both a pitch curve and an amplitude curve were shown. Seashore noted a characteristic drop in amplitude for transitions with bow change, but did not remark on the amplitude behavior without bow change. The amount of tremolo registered in the amplitude traces was as great as the dip in amplitude that I have come to expect at a bow-change transition. Seashore did remark that differences in amplitude dip "indicate a characteristic difference in bowing." He also found that some notes begin with "a clean attack in pitch."

In a study of phrasing, Morrill (1980) recorded trumpet melodies; every note in the melody was tongued (Morrill 1982). From the graphs of amplitude envelopes (which of course included the amplitudes of the transitions), he found that there was an overall phrase envelope which modified the middle section of the amplitude envelopes of individual notes.

Beauchamp (1981) included a preliminary report on synthesizing a melody—a two-bar Messiaen passage for solo clarinet—using his brightness matching technique (a form of waveshaping). The resynthesis involved eight time-varying amplitude traces plus a single frequency trace.

Dolson (1983, p. 107) presented an analysis of one legato transition played on the violin (see the figure on his p. 105); the pitches were apparently the second G and A♭ above middle C. Dolson showed the output of the channel of his tracking phase vocoder which is following the fundamental. He found "slowly decaying amplitude and frequency modulation" in the attack of the new note, lasting about 300 msec, which he ascribed to the effects of room reverberation (causing the notes to overlap). He conjectured that legato transitions might be "effectively simulated with simple overlapping." During the transition, it was hard for his technique to track the higher harmonics; the reasons will become clear in Chapter 2.

Sundberg et al. (1983, p. 39) found that "[i]n instrumental music, particularly that played on bowed instruments, wide melodic leaps are often performed with a very short pause just between the two tones." Although they presented no research to support this assertion, this corresponds to the well-trained musician's "common-sense" feeling for bridging a large gap; but I will show in Chapter 2 that this may not be as pronounced in the physical signal as one would have hoped.

Gordon (1984) studied the perceived attack time of musical notes. In particular, he synthesized a musical sequence (*Twinkle, Twinkle, Little Star*) with tones played on several different

instruments, each with different physical attack time. Working with the results of his research into perceived attack time, he was able to arrange the physical onsets of the notes so that they were perceived to fall in a regular rhythm. However, he was not concerned with the transition region itself; any overlap between consecutive notes was coincidental.

That leaves the study by Mathews and Miller (1982), which dealt with the effects of length of splice, abruptness of attack and decay, and abruptness of pitch change between notes, on a listener's judgments of slurring. Their study used artificial stimuli with simple linear attacks and decays, and its emphasis was on *producing* a *slurring* effect. Its implications for the current study will be mentioned later in this document.

Overview of the Current Study

Scope

It should be clear from the foregoing that a number of issues surround the exploration of sound in musical contexts. The current study will be limited to work with common non-percussive orchestral instruments. The only analysis techniques applied will be a form of the phase vocoder as well as a form of power measurement, both to be discussed in Chapter 2.

As a trustworthy set of recorded transitions was not available when this study commenced, a set of recordings was made and analyzed; the results are also presented in Chapter 2.

At the same time, it must be emphasized that this is a study of the physics and perception of transitions, but not a study of articulation. It should be clear from the foregoing that a player can use a variety of performance techniques to achieve a variety of perceived articulations. This study will not attempt a definitive statement of what makes a *bowed* transition sound different from a transition with no bow change. To be sure, much information on these differences will be presented, but the goal here in such cases is to demonstrate what makes it possible for the listener to tell apart two separate articulations, and not to demonstrate how to synthesize specific playing techniques.

Initially, the attack–steady-state–decay model will be adequate for modeling individual notes; a monophonic phrase will be modelled as a concatenation of such notes. Incidentally, this study will not examine a melodic fragment consisting of two notes on the same pitch; this is a subject worthy of a study by itself. There will be no attempt to examine differences in performers, tempi,

or dynamics, nor to examine the differences due to the quality of the instruments used, different fingerings for same pitch on a given instrument, formant effects, and the like.

Organization of this Document

Chapter 2 discusses the transitions which were recorded, and presents an initial analysis of them. Some plots are given there of the power and time-varying spectrum of transitions; the plots in Appendices 4 and 5 complete the set. A number of informal studies and formal experiments are given in Chapters 3–6. Chapter 7 summarizes the results. Certain methods which were developed in the course of this study, such as for amplitude scaling, are presented in Appendices 1 and 2. Appendix 3 presents details of experimental procedure which may be of interest to the reader but which are not necessary for explaining the focus and results of the experiments.

CHAPTER 2

ANALYSIS OF PHYSICAL PROPERTIES OF TRANSITIONS

This chapter presents a generalized framework for specifying transitions. Then the process of recording transitions played on various instruments is outlined. Based on analysis of changes in amplitude, power, and spectrum in the transitions, a representative set of transitions is selected, upon which the experiments of Chapters 3–6 will be based.

Parameters of a transition

The discussion in the previous chapter showed that there is no general agreement on what consitutes a transition. As a step toward defining a transition (which will happen near the end of this chapter), it will help to list the parts of a transition that might be subject to variation in controlled experiments. Figure 2.1a shows a simplified view of the time-varying amplitude in a transition, following the tentative definition of Chapter 1 (the reason for the choice of letters will soon become clear). The line AB is the steady-state of a note, BD is the decay. The actual transition happens at some point E in the general area DF. To simplify terminology, I propose to adapt Rösing's term *Tonwechselvollzug*, calling point E the *point of pitch change*. (In general, I found that E lies just before F.) This contrasts with the longer *transition*, which includes at least DF and at most BK. The attack of the second note spans FJ, with a steady-state reached at K and continuing to L and beyond.

Even a cursory examination of actual transitions will show that this model does not suffice in many cases. As shown in Figure 2.1b, the decay BD is often convex or concave. Even more complicated shapes are of course possible; for example, BC may be convex and CD concave, or there may be a sort of "plateau" at C. The reader may wonder that the segment BD is not an exponential decay, as acoustical theory would suggest. There are several reasons for this. Not only is a theoretical exponential decay modified by irregularities in the instrument and by

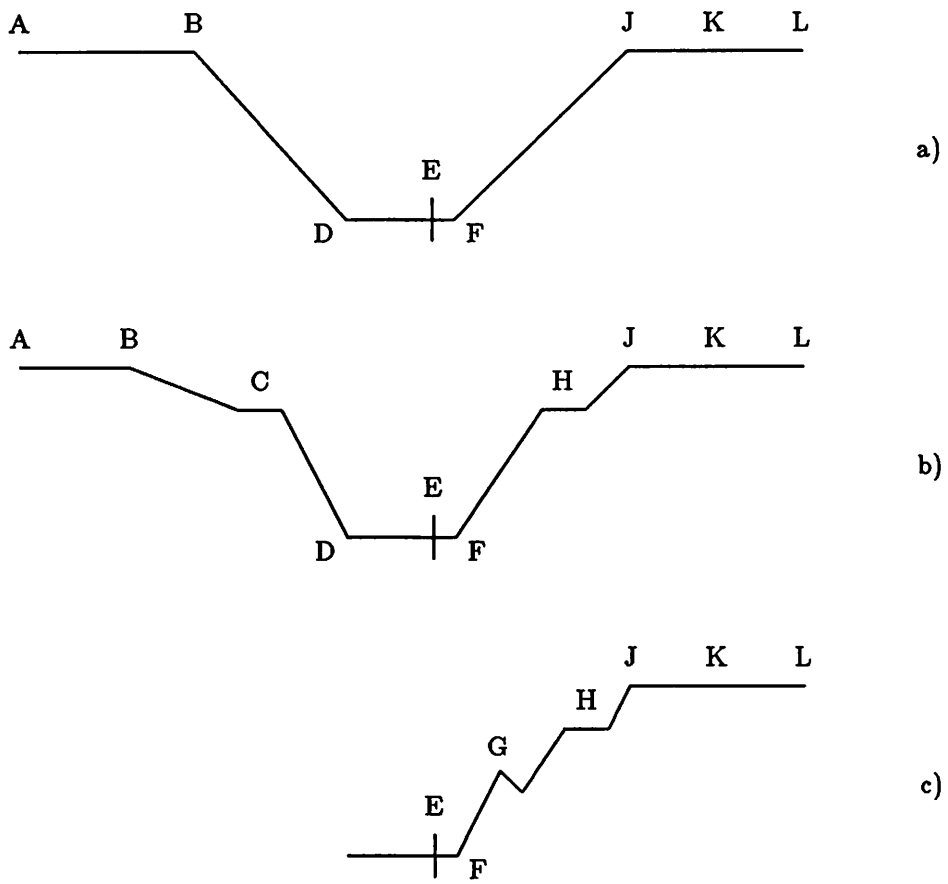


Figure 2.1. Amplitude during a transition. a) simplified model. b) More detail added to decay of first note and attack of second. c) Possible perturbations in attack of second note.

reverberation in the room, but the player may also continue to add energy to the vibrating system in the instrument. Likewise, the attack FJ may be more complicated than a straight line; FJ may be concave or convex, and a plateau may be found at H. In fact, the attack in some instruments may include characteristic features. Figure 2.1c shows a simplified version of the "blips" (G) commonly seen in brass attacks, already mentioned in Chapter 1. There may in fact be several blips in such attacks.

Consider, then, the decay of the first note. The parameters necessary for specifying this part of the transition include at least the amplitude levels of the breakpoints, the shapes of the lines connecting them, and the amounts of time between adjacent points.

The spectral characteristics also need to be determined. Do some harmonics drop out? If so, do they drop out in synchrony with each other? If they roll off asynchronously, is there a pattern

to be found? Is that pattern perceptually significant? Does the excitation source continue to supply energy through, say, BE? Are there resonances that ring in that region?

Does the amplitude at DF reach 0.0?* What is the shape of DF: a line, concave, convex, or irregular? Where within DF does E fall? Is there a discontinuity at E, or is the crossover smooth from one state of vibration to the next?† How long does this changeover take—that is, how much of EF appears to be unstable? Indeed, can DE and EF be characterized as periodic signals? Are there blips, bumps, grinds, wheezes, or other artifacts in DF? If so, of what shape, how large, and how long are they?

In the spectral domain, do some harmonics "persist" through DE? If so, is there any characteristic pattern determining which ones persist? How do their amplitudes change? How are the frequency traces joined at E, and do they move in synchrony with the amplitude traces? Are there specific spectral cues, such as pre-attack noise?

The parameters isolated for the decay of the first note also apply to the attack of the second. The position of any blips relative to the entire attack can be investigated. In addition, one often finds an "overshoot" in the area JK; if so, its amplitude and duration might be of interest.

Finally, there is the general relationship between the two notes. Is there a connection between the shapes of the decay and the attack, or between any spectral changes which might occur in them? How are the attack and decay times related? What about the overall amplitudes of the two notes: Are they the same, and if not, how does that affect the transition?

Recording Sample Transitions

> *Enormous difficulties were encountered in making this system operational.*
> (Luce 1963, p. 42)

To answer these questions, it is necessary to examine recordings from a wide variety of instruments—the answers to these questions might vary from instrument to instrument. For example, I would not expect to find "blips" in the attacks of string instruments; and the size of the instrument might affect the rate at which a note dies away.

---

* †Grey (1975, p. 110) quotes Moorer's discussion of these two questions. Neither Moorer nor I can find Grey's quotation in Moorer's thesis, which Grey cites with the date 1974; maybe someone else can find the passage. Or, since the correct date for Moorer's thesis is 1975, perhaps Grey's quote came from an earlier version.

From 1979 through 1983 I recorded transitions played on nine instruments at CCRMA: flute, piccolo, bass flute, clarinet, oboe, bassoon, trumpet, violin, and cello.* Some trombone recordings were also made, but were ultimately not analyzed here, for reasons which will be given presently. It seemed wise to select at least one instrument from each of the traditional families of orchestral instruments (wind, string, brass). Among the wind instruments, at least one from each kind of reed (air, single, double) was included. The reasons for selecting small and large instruments in these families will be discussed later.

Each player was recorded in a room measuring approximately 20' by 24' at CCRMA.† The walls were treated with absorbent material to reduce reverberation in the room. Its isolated location made it adequate for undisturbed recording.

Some recordings were made directly onto the mainframe computer using the 14-bit DACs installed on the PDP-10 of the Artificial Intelligence Laboratory, which occupied the building at the time. This recording setup is discussed in the earlier version of (Moorer 1977). The remaining recordings were digitized directly (16-bit) using a Sony F1 recorder; these recordings were digitally transferred into the CCRMA Foonly computer and stored on disk. For some recordings, I used a B&K 2619 microphone, and a Crown PZM for others. All of the recordings were resampled to 25.6 kHz, which seemed (and proved) to be low enough for practical work but high enough to ensure adequate fidelity. There were some low-frequency artifacts in the recording, such as a DC component, which were removed by high-pass filtering the recordings, typically with an 8th-order Butterworth filter with -3 dB point at 50 Hz or so. The recordings were spliced apart into individual (monaural) files, each containing a two-note segment.

The noise level of the recordings turned out to be around -60 dB. Given the theoretical limits of 84 dB (for the 14-bit DAC) or 96 dB (for the Sony 16-bit DAC), this may seem surprising. However, monitoring amplitude in the direct-to-computer-disk method (discussed in the earlier version of Moorer 1977) was clumsy at best, so recording at close to full amplitude was ill-advised; and experience with the trumpet showed that the "overload" light on the Sony F1 recorder was not as reliable as one needs when working with very tight headroom. Thus, the trumpet recordings were clipped in a few places in the steady-states of the tones (not in the transitions). Fortunately,

---

*The help of the following performers is gratefully acknowledged: Yvonne Kendall, Emily Bernstein, David Burkhardt, James Matheson, Angela Sohn, Gregory Dufford, Leland Smith, Dexter Morrill, David Jaffe, Chris Chafe, Stephen Harrison, and Pat Spurling.

†This room, variously known as "the recording studio," "the piano room," and "the pit," was the same room used by Borish (1984); the equipment shown on p. 42 of his thesis led to the name of "the dungeon." Many of the experiments by Grey and Gordon were also conducted in this room.

Table 2.1. Instruments Recorded.

Family	Instrument	Base Pitch
Air Reed	Flute	A220
	Piccolo	A1760
	Bass flute	A220
Single Reed	Clarinet	A220
Double Reed	Oboe	A440
	Bassoon	A220
String	Violin	A220
	Cello	A220
Brass	Trumpet	A220

the clipping lasted for only two or three samples each period. I was able to remove the clipping by low-pass filtering the recordings before doing the downsampling. Still, this warned me against trying again to exploit the full dynamic range of the F1; for the other recordings, more headroom was alotted, with a resulting reduction in dynamic range. This posed no further problems in any of the work presented here. In particular, the transitions all lay well above the noise floor, as will be shown in Table 2.2.

The Choice of Intervals

As discussed in Chapter 1, "common sense" suggested that the transition for a narrow interval might be different from the transition in a wide interval. Also, it seemed reasonable that a descending interval on an orchestral instrument might operate differently from an ascending interval. For example, with the oboe and bassoon, Forsyth (1936, p. 236) says that their "best slurred skips—that is to say, slurs merely between two notes as opposed to extended *legatos*—are those taken upwards. This point, however, is not of great importance unless the skips are very wide."

The following intervals were finally selected: major second (M2), major third (M3), perfect fifth (P5), minor seventh (m7). All four intervals were recorded ascending and descending.

Following the lead of Grey's work, it seemed wise to record as many instruments as possible playing the same pitches. This might make cross-instrumental comparisons easier; but proved to be impossible for the full set of the instruments listed, as their normal playing ranges do not overlap. The next best solution was to have many of the instruments play the same pitches, and

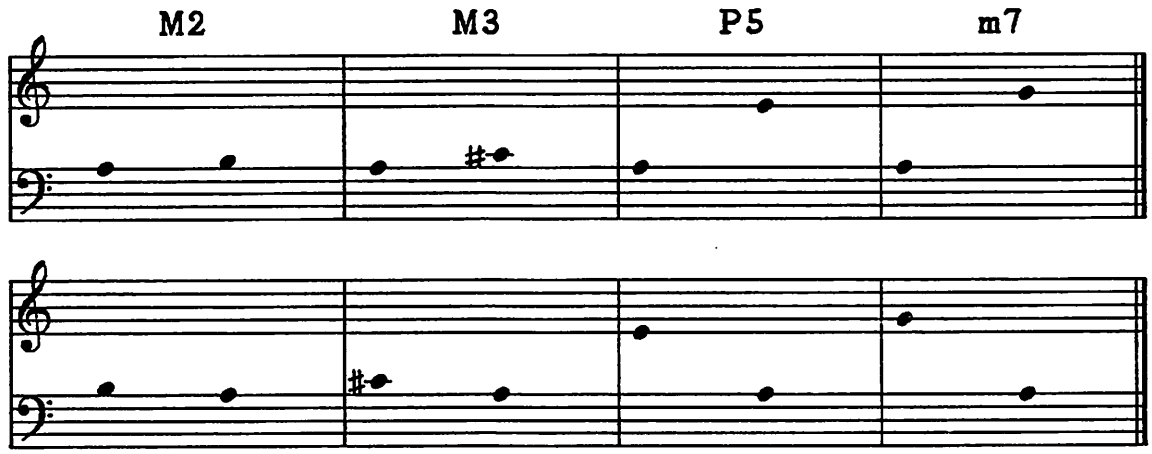


Figure 2.2. The ascending (top staff) and descending (bottom staff) intervals recorded. For some instruments, these notes were transposed up one or more octaves. (Typeset by Amnon Wolman using Leland Smith's MS music printing facility).

to have other instruments play an octave above or below. Another constraint was to avoid open strings on the string instruments; all pitches were to be played on stopped strings. In the end, the "base" pitches given in Table 2.1 were used for each instrument; that is, these were the lower notes for both ascending and descending intervals. Figure 2.2 shows the complete set of intervals (based on A220).

The Choice of Articulations

It was also necessary to choose more than one style of articulation for each interval. For the stringed instruments, articulating with or without bow change was the obvious choice. Likewise, with the woodwinds and brass, playing with or without tonguing would be adequate, it seemed.

But experience in recording sessions with the musicians showed that the choice was not so clear, as the readings cited in the last chapter had warned. One of the trumpet players whom I recorded stated explicitly, "It's possible to create the illusion of a slur with the 'ta'", where "ta" is the syllable which produces the most pronounced tonguing. To give another example, the following dialogue spontaneously took place during the recording session with the oboist:

[Strawn]: ... And there's still quite a good gap there.

[Oboist]: Well, I was gapping it somewhat; it's sort of a quasi-separated legato. I was experimenting with the attack in there aways. Now let me see if I can blend it clear in. (plays)

[Strawn]: (surprised) *That* was tongued?

[Oboist]: (plays) That one is there. The other one ... you can get so it's light enough you don't hear it. You have to sort of back off from that in order for it to sound articulated. Let me give you another example. (plays) That was tongued there. In one way, if you coincide it right on the nose when you're shifting to the new note and without the break in the middle, you can sort of hide it in the shock of the changed pitch.

[Strawn]: Even though you're tonguing?

[Oboist]: Yeah, yeah, to a certain degree.

[Strawn]: Some other instrumentalists have told me the same thing about their instruments too.

[Oboist]: Well, some are easier than others. It's mainly an individual ability, whether they develop it or not, or whether the schooling they had emphasized that to one degree or another. But it's not that hard to get to the point where it sounds slurred.

[Strawn]: Even when it's tongued?

[Oboist]: Yeah, even though it's tongued.

I thus instructed the performers to play a normal tongued transition, without trying to "hide" the tonguing; and to contrast this with a normal, untongued legato. The same was true for the strings, with and without bowing. Thus, all eight intervals in Figure 2.2 were recorded twice, with two different articulations. It is important to remember that this provided a "quick and dirty" method of obtaining two articulations which are well-known in the musical community, and which can be distinguished easily; but the goal of this work does not include trying to specifically model the method of performance.

To simplify matters in the rest of this document, "tonguing" (abbreviated T) will be understood to include the performance "with bow change" on the string; and "untongued" (U) also includes performance "with no bow change."

Equalizing the recordings

All of the recordings were equalized in amplitude so that the stronger of the two notes reached an amplitude of 75% of full scale. This was a nice round number to work with, allowed some headroom for later experimentation, but used a goodly amount of the dynamic range available.

Grey (1975) equalized his test tones for loudness, duration, and pitch. It proved impossible to equalize my two-note recordings along these lines. In order to equalize the recordings for pitch, the notes would have to be treated separately—which would destroy the very transitions to be examined. Given the variety of shapes and sizes in the transitions, as discussed in the beginning of this chapter, it was also impossible to find a way to equalize, say, the loudnesses of the tongued transitions, without again distorting the transition as recorded. Therefore, each instrument is studied first by itself. Still, the lack of an equalized set of recordings will not rule out making certain cross-instrument comparisons, as will become clear in later chapters.

Recorded Transitions

A sample set of transitions is shown in Figures 2.3–2.8. Each two-page spread contains amplitude plots for one instrument: the clarinet, trumpet, and violin are shown. On the left-hand page, there are four tongued intervals (three for the clarinet, as the ascending seconds were lost); the right-hand page shows the same intervals, played without tonguing. For the clarinet we thus have, from the top, major third, perfect fifth, and minor seventh. In each case, the decay of the first note is shown at the left, followed by the transition, then the attack of the second note. Amplitude is plotted on a linear scale, with 1.0 representing the full 15 (positive) bits available. Time is shown in seconds; each plot shows 300 msec. The "waves" in the clarinet recordings are artifacts of the display process and should be ignored; they disappear when a smaller time range is displayed. The trumpet has the same intervals, plus the major second. Only ascending transitions are shown for these two instruments. A smaller time range is shown for the trumpet in order to avoid certain artifacts in the display. Since the larger two intervals were not analyzed in the strings, the violin transitions shown here include the major seconds and thirds, both ascending and descending. The time range shown varies from 100 to 200 msec.

(Text continued on p. 42)

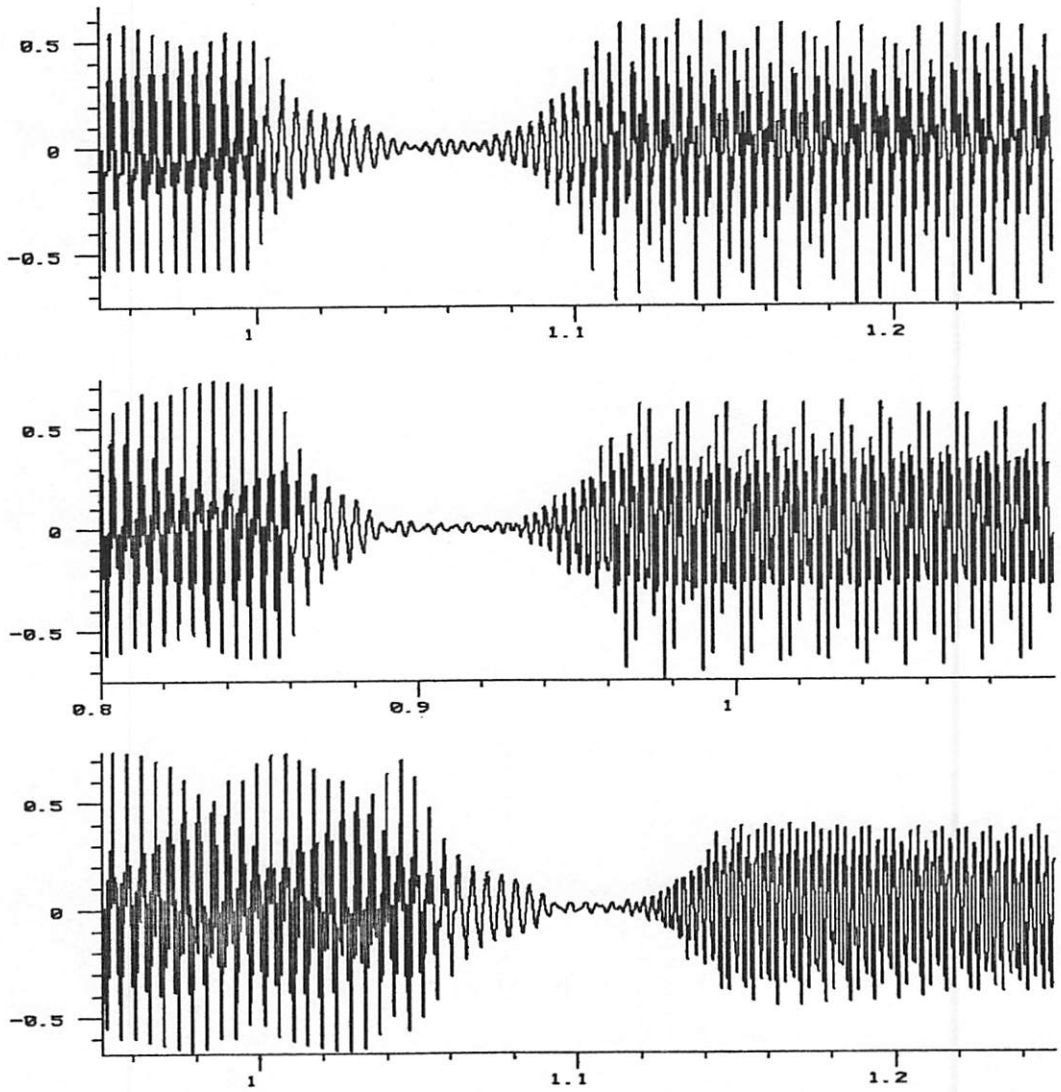


Figure 2.3. Tongued transitions between two notes on the clarinet. The lower (first) note is A220 in each case. From the top: ascending third, ascending fifth, ascending seventh.

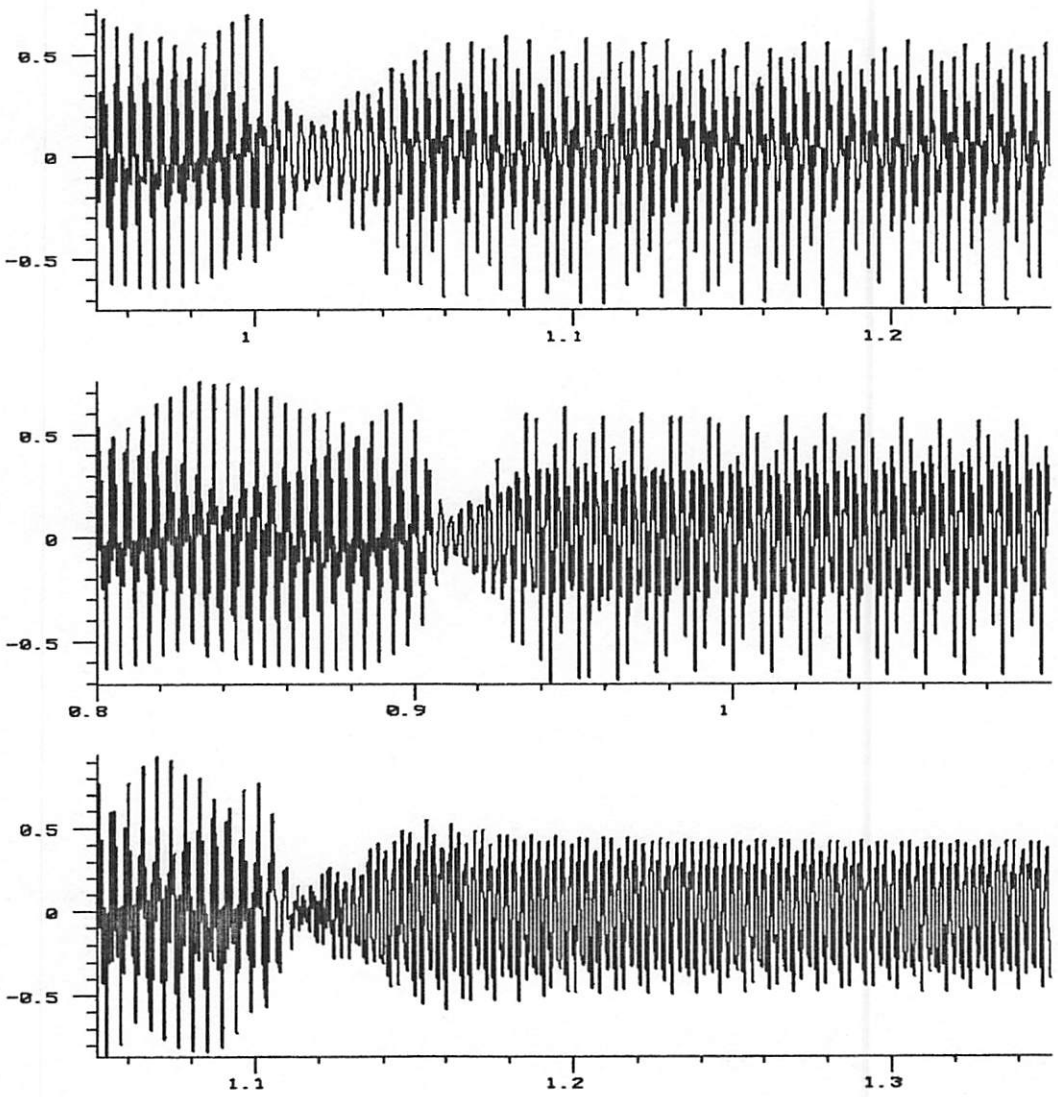


Figure 2.4. Untongued transitions between two notes on the clarinet. The lower (first) note is A220 in each case. From the top: ascending third, ascending fifth, ascending seventh.

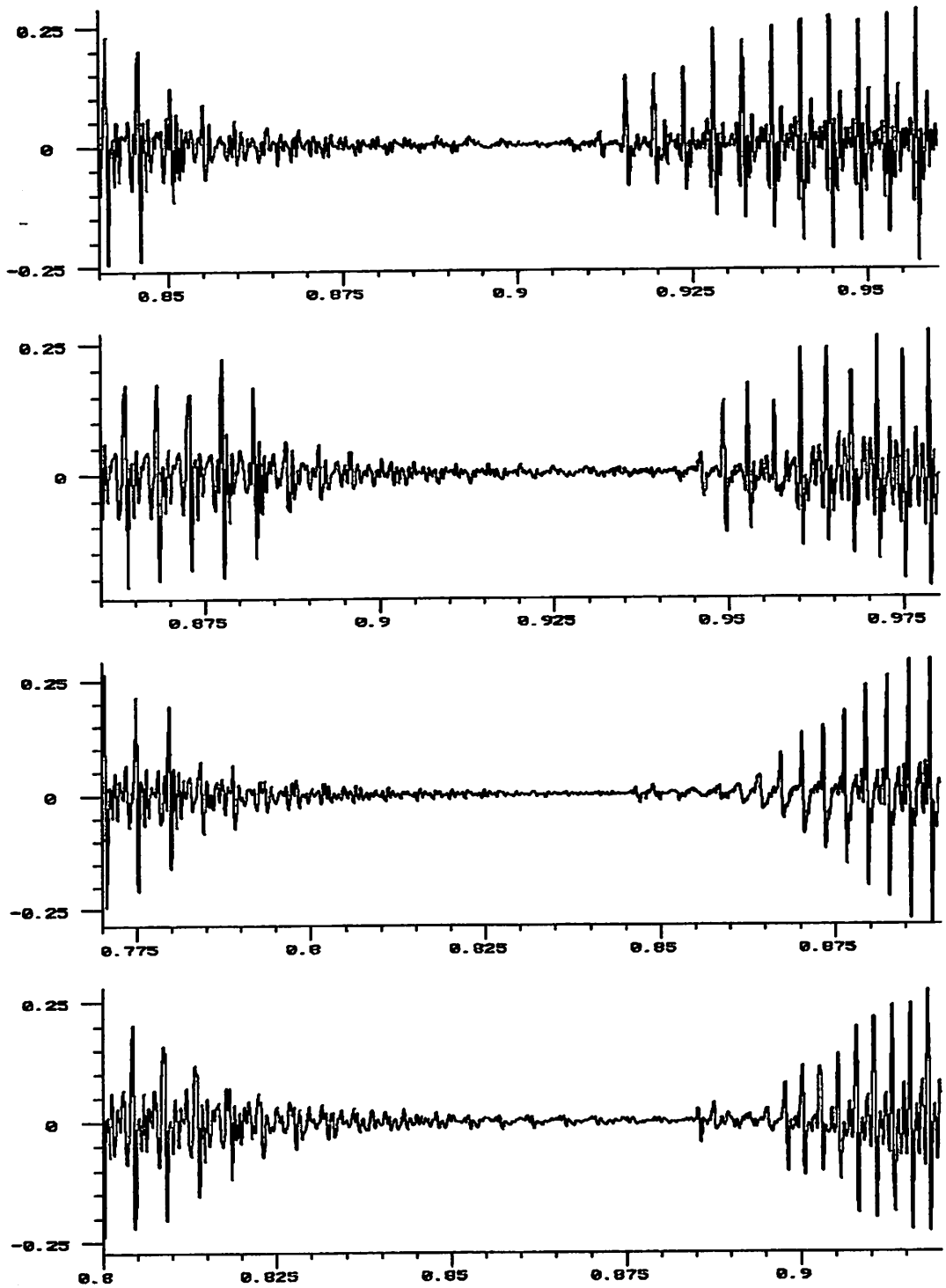


Figure 2.5. Tongued transitions between notes on the trumpet. The lower note is A220 in each case. From the top: ascending second, ascending third, ascending fifth, ascending seventh.

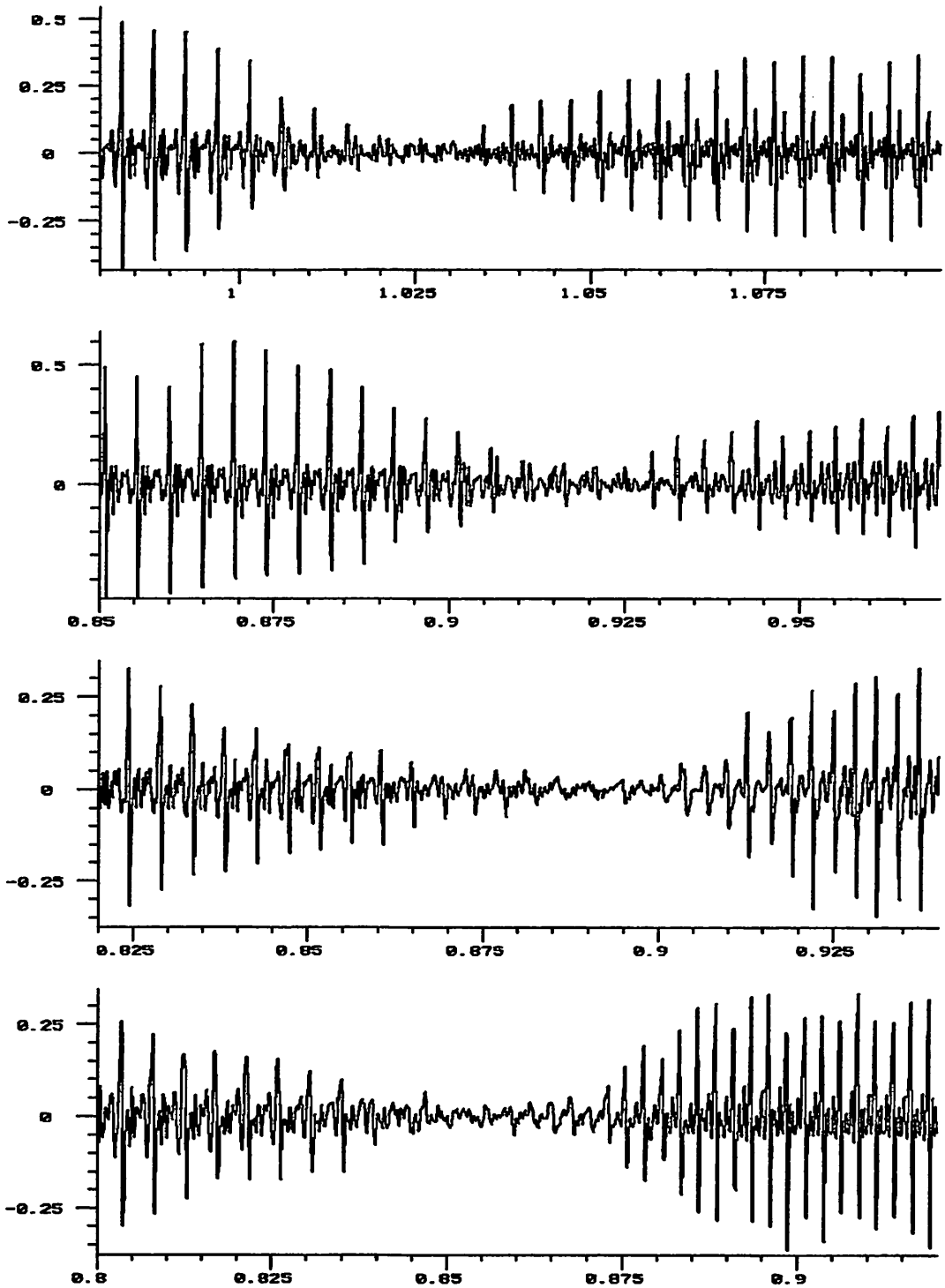


Figure 2.6. Untongued transitions between notes on the trumpet. The lower note is A220 in each case. From the top: ascending second, ascending third, ascending fifth, ascending seventh.

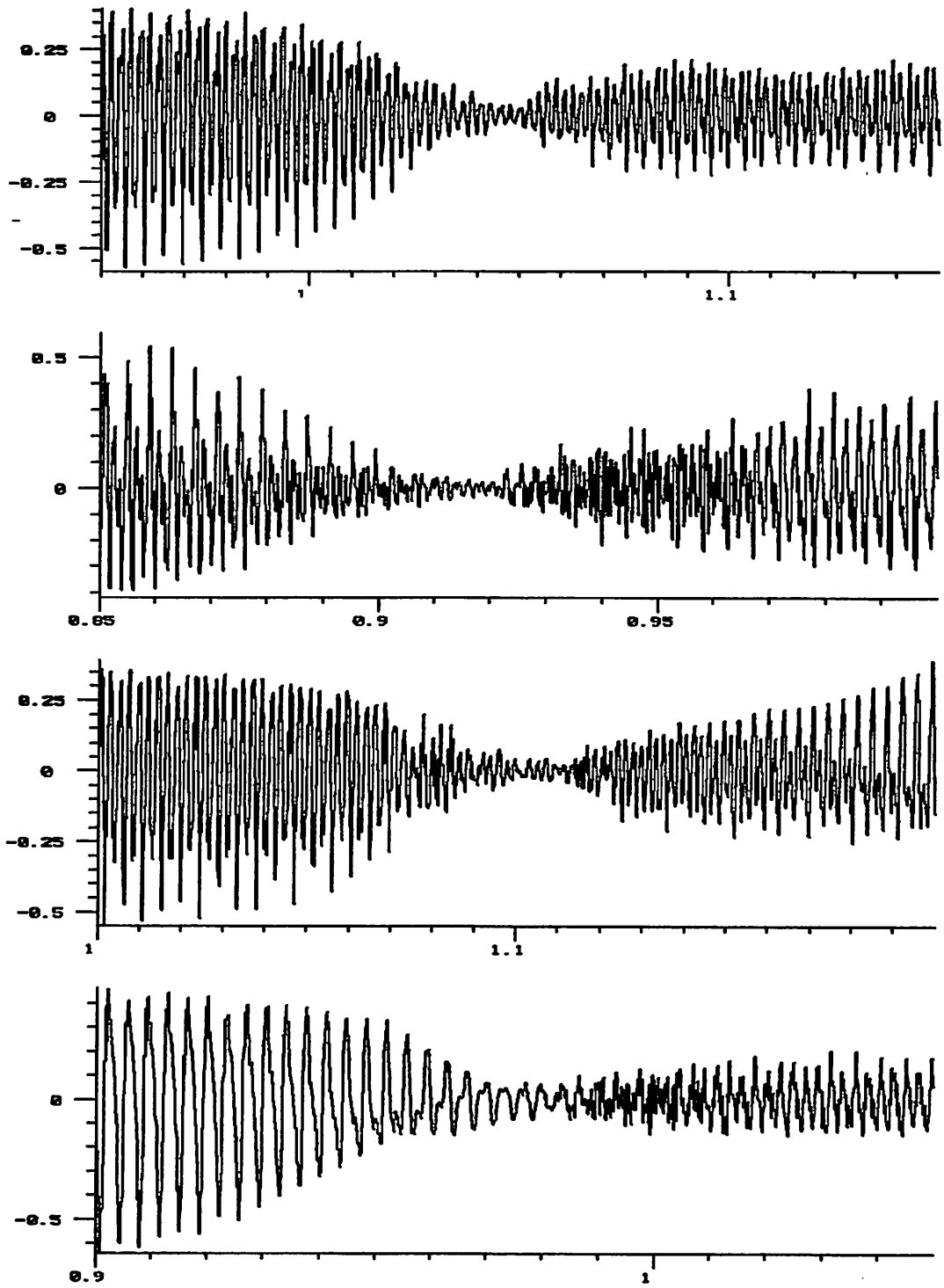


Figure 2.7. The transition between notes on the violin, played with bow change. The lower note is A220 in each case. From the top: ascending second, descending second, ascending third, descending third.

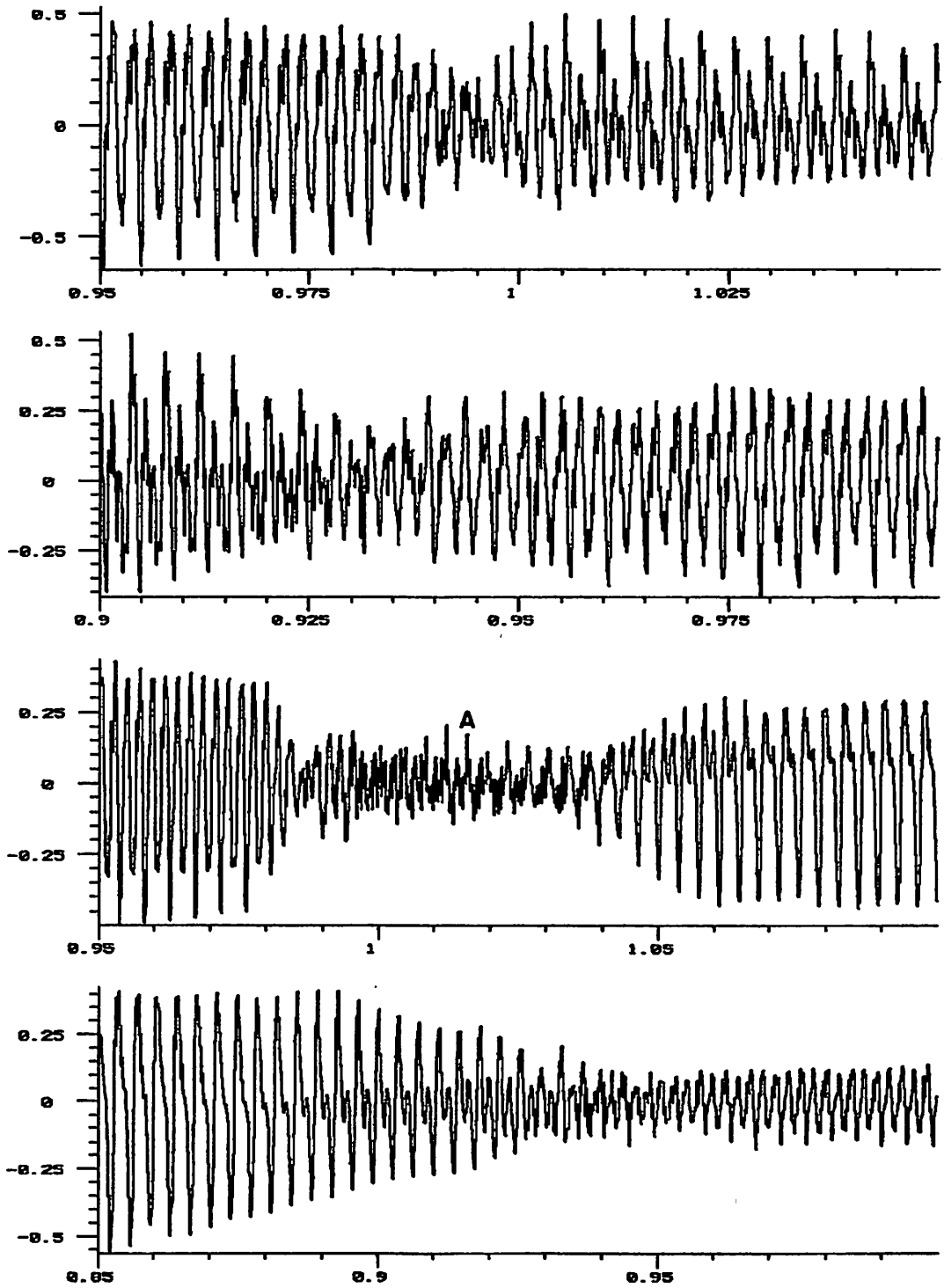


Figure 2.8. The transition between notes on the violin, played without bow change. The lower note is A220 in each case. From the top: ascending second, descending second, ascending third, descending third.

The General Nature of Musical Transitions

The conclusions in this section are based on the plots shown in Figures 2.3–2.8 as well as on similar plots, not reproduced here, for the other instruments recorded.

There is a characteristic drop in amplitude between the two notes surrounding a transition, as one would expect. This turned out not to always be the case for the cello, which will be discussed in more detail later.

The tongued case often exhibits a wider gap between the two notes, and a greater dip in amplitude, than for the nontongued case. To answer Moorer's question quoted earlier in this chapter, in no case did the amplitude between notes for the tongued case fall into the noise level of the recordings (in other words, drop to 0); this may be due in part to room reverberation. (One would not expect such a large drop in amplitude for the untongued case, nor did such a drop occur there).

The amplitudes of the two notes are sometimes quite different. The descending major third with bow change at the bottom of Figure 2.7 is one example, although the difference may be difficult to see in the plots given here; the power plots in the next section will make this clearer.

The decay time of the first note is often different from the attack time of the second note. Indeed, as was anticipated in the discussion earlier in this chapter, the attacks and decays often include a short "plateau;" see, for example, the attack in the tongued ascending fifth played on the clarinet (second from bottom, Figure 2.3), or the decay in the ascending fifth in the untongued trumpet (second from bottom, Figure 2.6). Some instruments showed a swelling on some notes, such as on the ascending third played with bow change on the violin (Figure 2.7, second from bottom). Some transitions, like the ascending second at the top of the same figure, show "shoulders" in the decay (or the attack); these will be modelled explicitly in Experiment 5. Only some of the decays follow the exponential path which one would expect; the tongued clarinet tones in Figure 2.3 are perhaps the "closest to theory." The reasons for this non-exponential behavior have already been given.

Although the amplitude plots are unclear in this matter, the change in pitch can be shown to occur at the earliest in the middle of the transition, and sometimes right at the attack of the second note. In none of my recordings did the change in pitch occur during the decay of the first note.

The tongued transition in the woodwinds and brass has, as one might expect, a small amount of noise right at the attack of the second note; this is easier to see in the trumpet plots than in the clarinets. In the strings, one might expect "bow change" to produce a more abrupt attack on

the second note; certainly some bow noise can be seen in the plots given here. However, the "no bow change" performance produces its own abrupt attack on ascending notes, because the finger "thwacks" the string to make it shorter, producing a characteristic sound which is probably not noticed by any listener in a real listening environment (see point A in Figure 2.8). A microphone near the instrument picks up this sound, of course. Still, this "thwack" is not prominent enough to allow the analyses presented later to distinguish between the ascending and descending cases, so it will not be considered further here. It would have to be taken into account if one were attempting to model specifically the "no bow change" playing style. In general, I found at least a little bit of noise in all the attacks; this is contrary to what Rösing observed. Probably the frequency resolution and dynamic range of the Kay Sonograph were not adequate to make this noise visible in his plots. Also, I often found the noise in the very high frequency ranges, which may have been off the scale of what was available to Rösing.

The transition from one pitch to the next occurs very quickly. In nontongued cases, one can often follow the peaks of the periods of the first note forward, and the peaks of the periods of the second note backward, until they overlap for just a few periods. Some of this overlap is no doubt due to room resonance, as others have noted. In other cases a few unstable "periods," the periodicity of which cannot be tracked easily, suffice for the transition. A good player can thus make the transition between notes in the time required for just a few periods; this matches well Benade's remarks, quoted earlier, about how the instrument can quickly reach a stable state of oscillation. At the same time, it shows the difficulty of pinpointing exactly where the point of pitch change lies.

The situation is not so simple, of course, when the player purposefully includes noise, such as can be seen in some of the violin plots. Thus, we can only offer a partial answer to Moorer's question about whether a discontinuity occurs. In smooth, more-or-less noise-free transitions, a discontinuity does not have to occur; many of the transitions shown here do not exhibit a discontinuity. The question of discontinuity in the midst of noisy transitions will not be examined further here.

Analysis of Time-varying Power in Transitions

It is misleading to base detailed analysis of time-varying signals solely on amplitude plots such as those just given. The waveshapes produced by musical instruments can vary widely, so that the peak values of successive periods are not a good measure of loudness. When a new note

starts on an instrument which is already vibrating, a large "phase shift" can occur between the notes (see, for example, the ascending third in Figure 2.8, second from bottom). Furthermore, the positive-going peak excursion is significantly different from the negative-going in some recordings. Another measure of the signal's amplitude is therefore needed. RMS power is widely used; but here a time-varying measure is called for.

To this end, I adopted an algorithm developed by Smith (1984), which identifies the peak of each period in the waveform (see also Hutchins 1975). It was possible to tune this algorithm so that it worked for both notes and the transition in almost all of the recordings. In some cases, however, the results had to be adjusted by hand (Beauchamp [1981] also had to correct his frequency estimates by hand). For this, a special software editor had to be written (this was a variant of the cursor-based editor described in [Strawn 1985a]). Those working with other period-driven algorithms, such as tracking phase vocoders, might find this approach useful. I should mention in passing that this algorithm sometimes worked better with an inverted signal (which of course sounds identical to the original); Risett (1966, p. 7) had similar experiences with his peak-detection algorithm.

After the period peaks have been identified, the power of the signal can be calculated on a period-by-period basis according to

$$P(n) = \frac{1}{T(n)} \sum_{t=n}^{n+T(n)} y^2(t)$$

where $T(n)$ is the length of a period (peak-to-peak) beginning at sample number n and $y(t)$ are the samples in the recording. $P(n)$ is thus measured once per period. I call this *period-synchronous* power, contrasting it with various well-known *pitch-synchronous* measures.

All of the recordings were analyzed for period-synchronous power. Figures 2.9–2.14 show time-varying power plots for the two-note pairs given in Figures 2.3–2.8; the power plots show the entire two-note recording in each case. All of the power plots are shown on a 60 dB scale. (The plots given here for the tongued trumpet agree well with those given in Morrill 1980). To facilitate comparisons, the layout of these power plots matches exactly the layout of the earlier amplitude plots. Appendix 4 contains power analyses for some of the other the instruments recorded, except for the cello and flute, which will be presented later in this chapter.

The plot of the attack of the first note and the decay of the second note should not be taken literally in these figures. For example, in the fifth shown in the middle of Figure 2.9, the line sloping downward just before the first note actually begins represents a small amount of noise in the recording which is not really as loud as the area under that line might imply. Likewise, the

representation of the end of the second note in the same recording is misleading—the note did not stop abruptly (the same is true of the violin plots). The lowest plot in Figure 2.9 shows how the begins and ends of all the plots "should" look.

Sometimes the amplitudes of the two notes vary significantly. This can be seen most clearly in the bottom violin plot of Figure 2.13. Notice the "swell" on the second note in the top three plots of the same figure.

Some plots include "burrs" along the power curve; Figure 2.14 is a case in point. These are areas where the peak-tracking algorithm was confused, often by a large-scale shift in phase due to time-varying spectral changes. It is possible to remove these "burrs" by correcting the locations of the peaks and then re-calculating the power curves; but I have done so only in the worst cases. Experience shows that the burrs accurately follow the outline of the curve; and no burrs occur here in the transition regions anyway, which is the subject of interest.

For some instruments, the power for a given playing style seemed quite consistent across interval size and direction. To give one example, the differences among the three plots in Figure 2.9 on the one hand, and the differences among the three plots in Figure 2.10 on the other, are not nearly so large as the differences *between* the two figures. In Figure 2.9, the amplitude dip is deeper, and the time gap between the two notes wider, than in Figure 2.10. The same generalizations can be made about the trumpet power plots (Figures 2.11 and 2.12), although the differences between the two ascending sevenths are not as pronounced. With the violin, the same sort of trend can be found, but not always as pronounced as in the clarinet or trumpet. Examination of the plots in Appendix 4 shows that this pattern holds for almost all of the instruments.

Furthermore, this pattern seems to hold no matter whether the intervals are ascending or descending. Thus, the power traces for the ascending and descending seconds in Figure 2.13 are quite similar, as are the two top traces in Figure 2.14; the largest difference occurs between the two figures. To give more examples, ascending and descending pairs are included for the piccolo and bass flute in Appendix 4 (see Table A4.1; cello plots will be given later in this chapter). Of course, there are some exceptions to this pattern, but they seem to be caused by idiosyncracies of playing a given interval on a given instrument, as in the second plot from the bottom of Figure 2.14.

I have thus concluded that *time-varying power varies more with the playing method than with size of the interval, the direction of the interval, or the instrument used.* The cello was the only major exception; a special section will be devoted to that instrument presently.

(Text continued on p. 52)

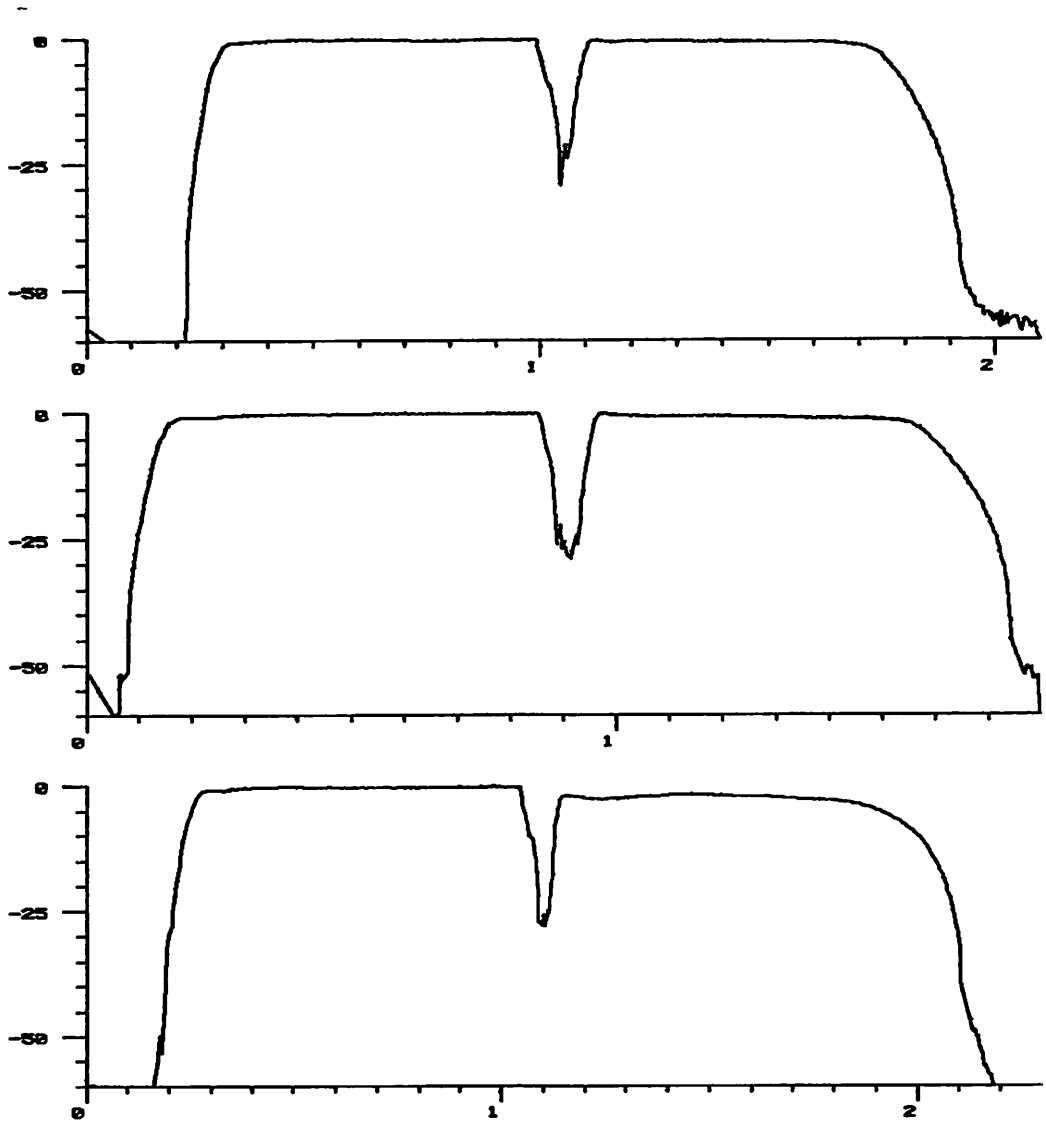


Figure 2.9. Period-synchronous power of two notes on the clarinet, with the second note tongued. The lower (first) note is A220 in each case. From the top: ascending third, ascending fifth, ascending seventh.

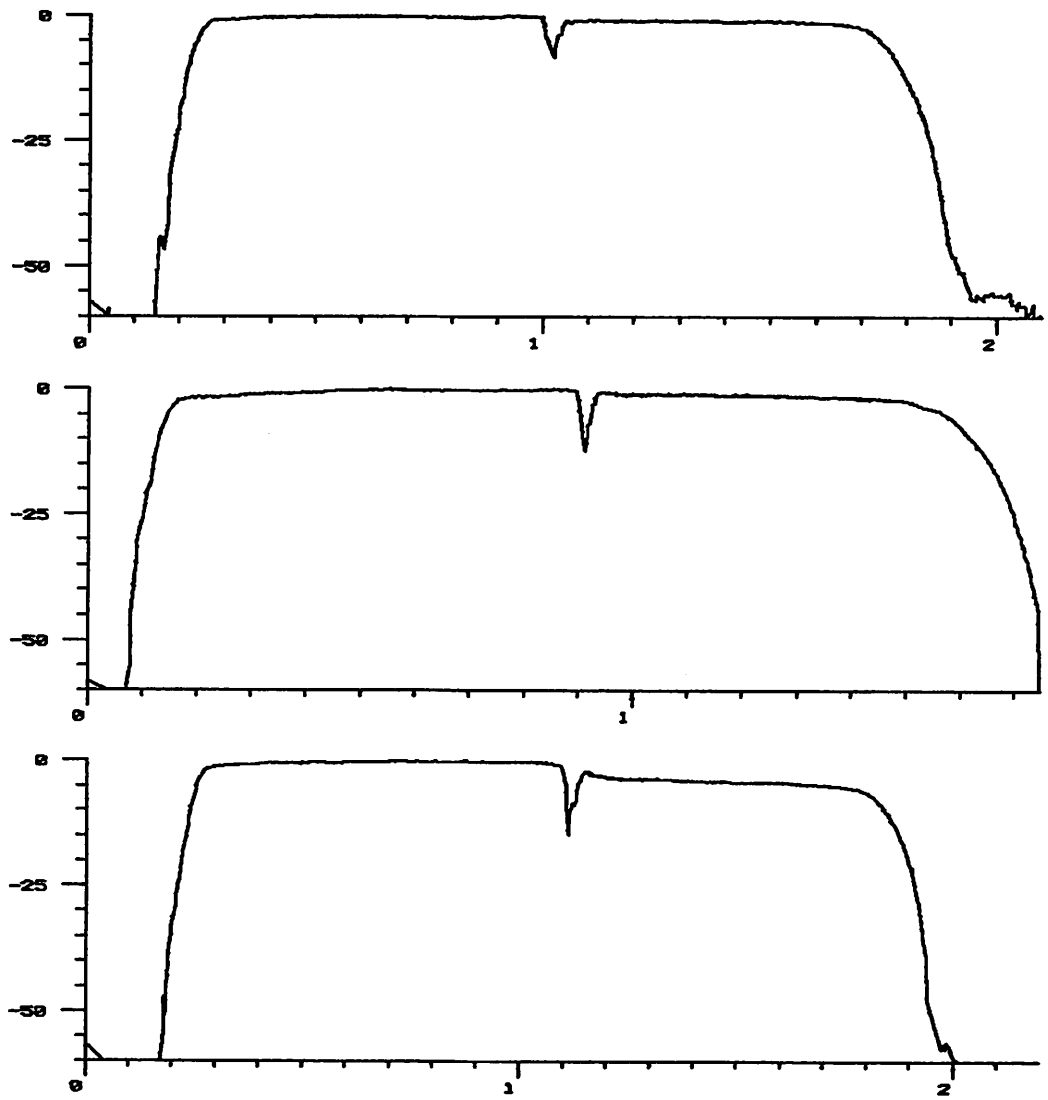


Figure 2.10. Period-synchronous power of two notes on the clarinet, with the second note untongued. The lower (first) note is A220 in each case. From the top: ascending third, ascending fifth, ascending seventh.

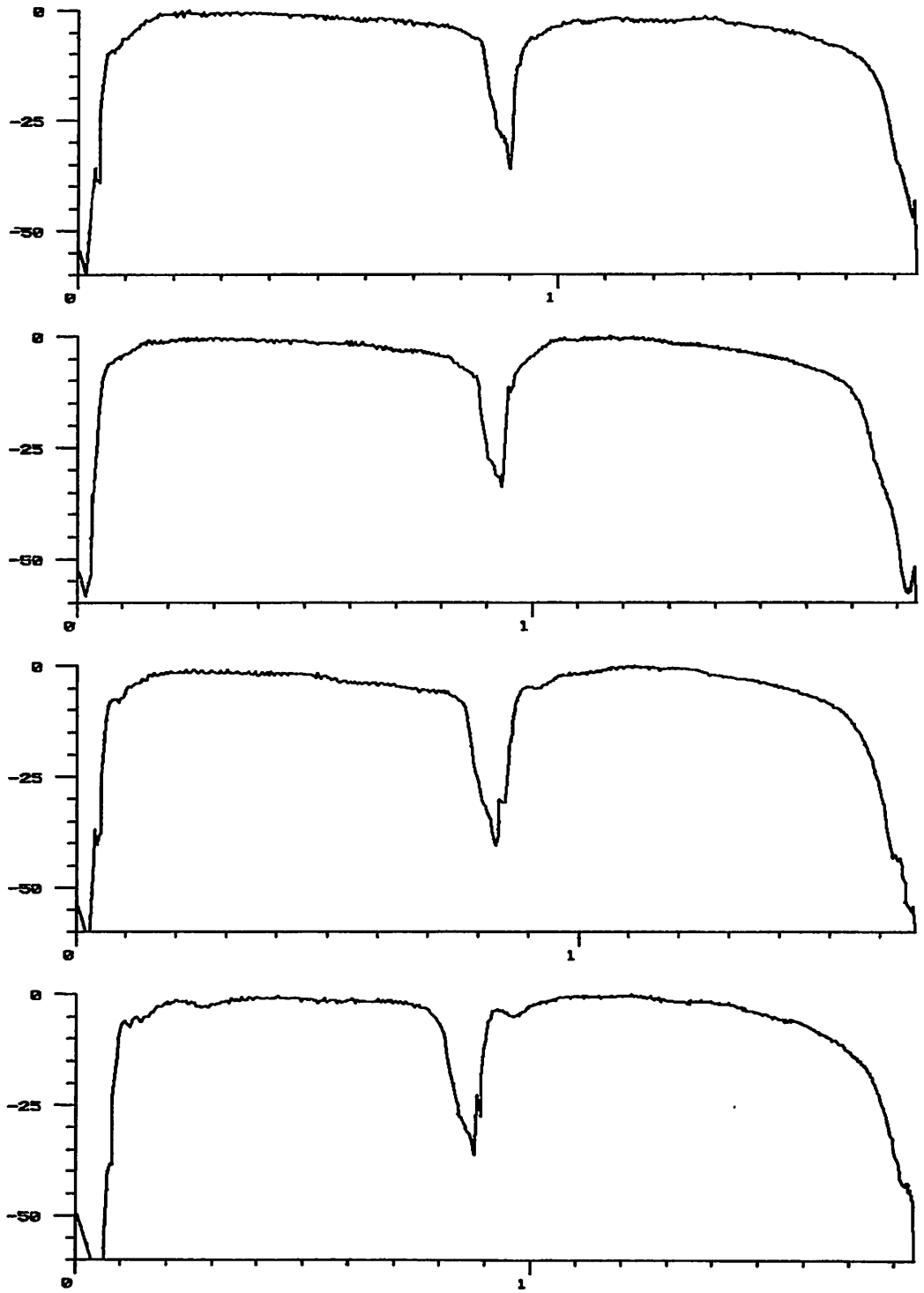


Figure 2.11. Period-synchronous power of two notes on the trumpet, with the second note tongued. The lower note is A220 in each case. From the top: ascending second, ascending third, ascending fifth, ascending seventh.

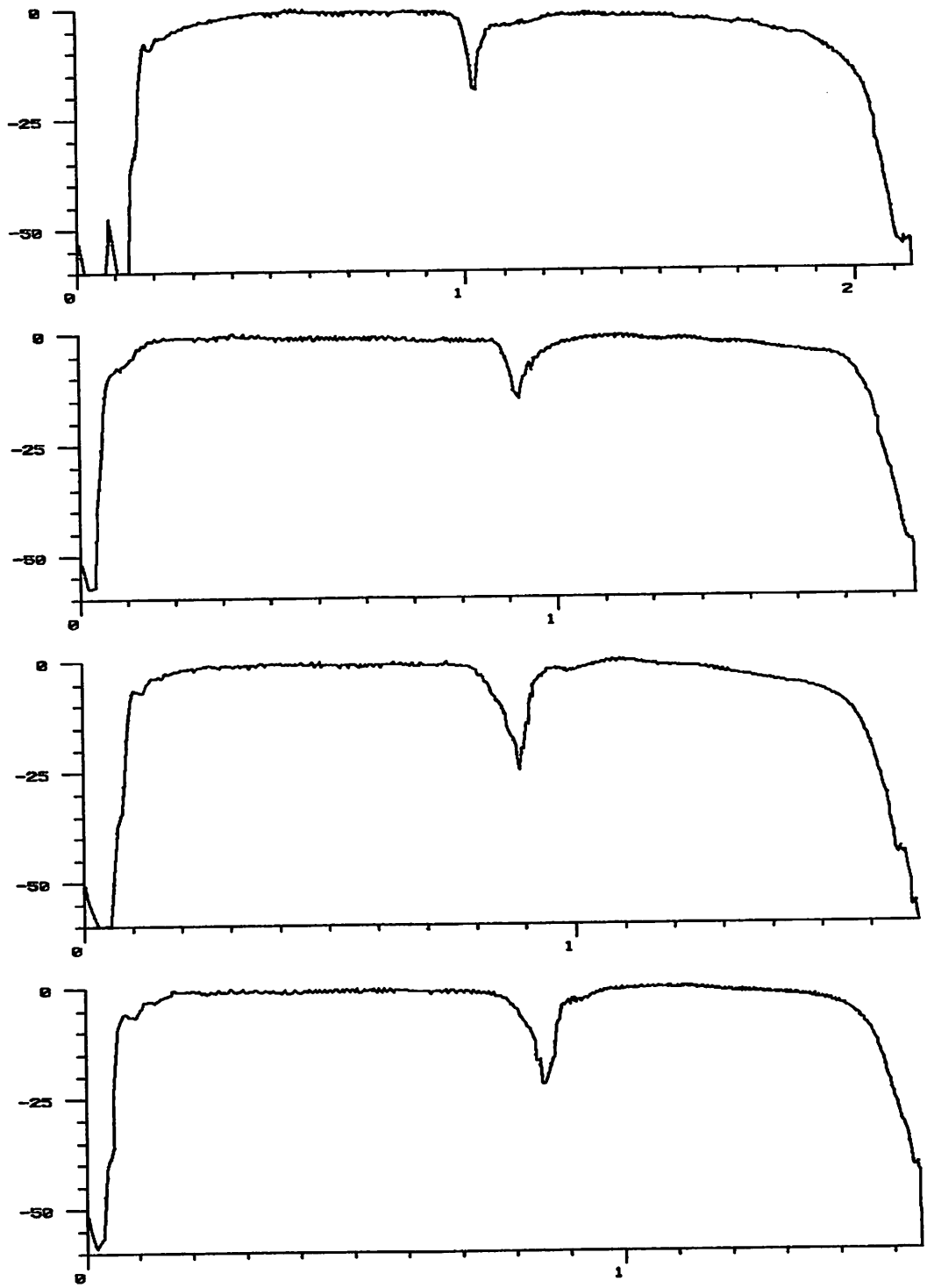


Figure 2.12. Period-synchronous power of two notes on the trumpet, with the second note untongued. The lower note is A220 in each case. From the top: ascending second, ascending third, ascending fifth, ascending seventh.

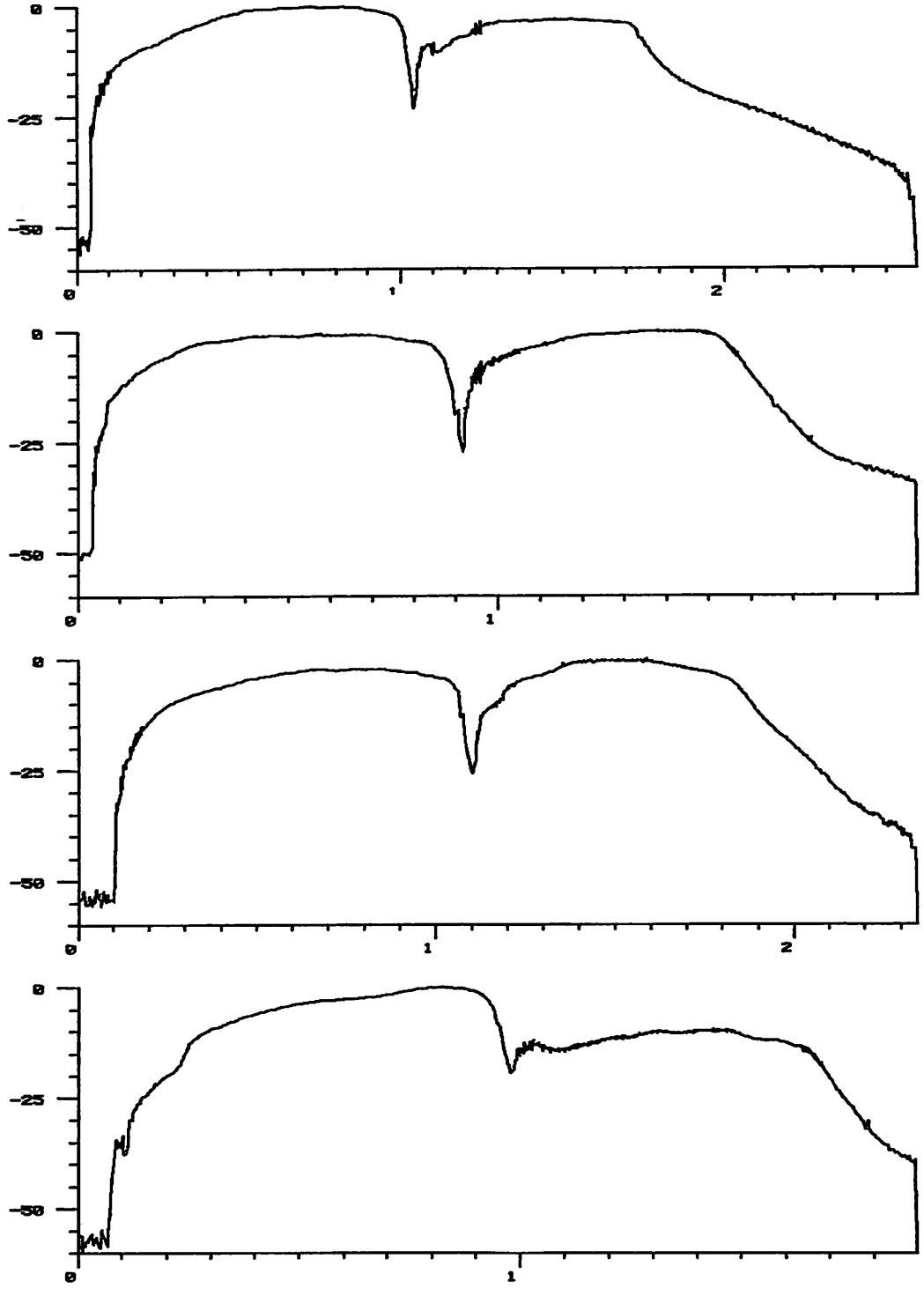


Figure 2.13. Period-synchronous power of two notes on the violin, with change of bow direction for the second note. The lower note is A220 in each case. From the top: ascending second, descending second, ascending third, descending third.

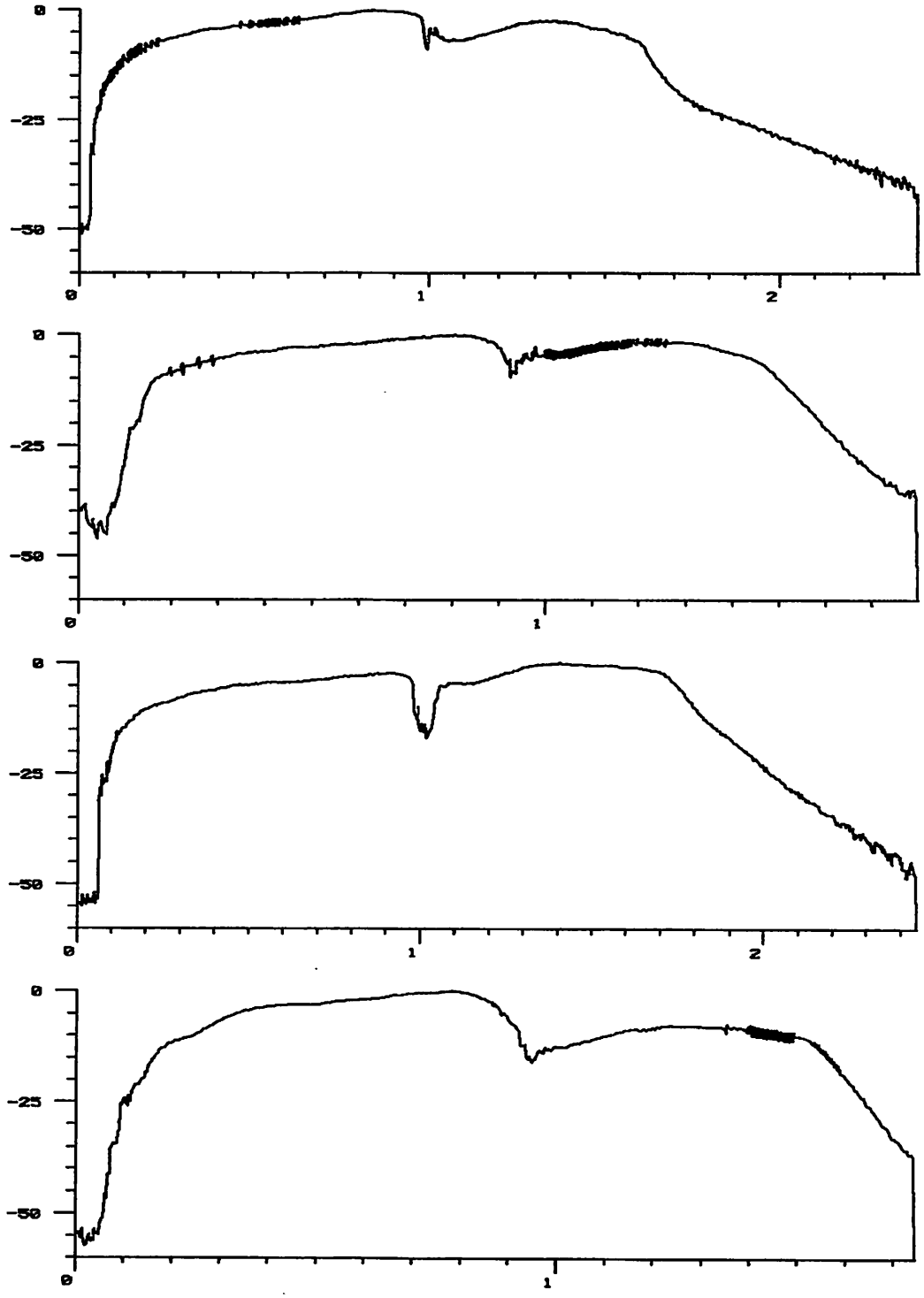


Figure 2.14. Period-synchronous power of two notes on the violin, with no change of bow direction for the second note. The lower note is A220 in each case. From the top: ascending second, descending second, ascending third, descending third.

One might apply some statistical techniques, such as analysis of variance, to these power plots to determine whether, say, the differences between several intervals were statistically significant on a given instrument. To do so would require an *even larger* data base than what I had gathered (this is not recommended for the casual researcher). Furthermore, any such recordings would have to be equalized to remove as many inessential variables as possible; but the difficult of doing so has already been discussed.

Time-varying Analysis of Spectrum in Transitions

Preliminary analysis of the amplitude plots (Figures 2.3–2.8) showed that the spectrum was changing in the transition. Measuring that spectrum and how it changes is not easy.

Problems with the Phase Vocoder

Reliability of Frequency Traces: A perennial difficulty with the phase vocoder is interpreting its output. Since the amplitude of the signal drops several tens of decibels during many transitions, the frequency traces are especially difficult to interpret, because the frequency trace becomes unstable at very low amplitudes.

Reliability of Amplitude Estimates: Recall that the phase vocoder, in effect, places a bandpass filter around each harmonic. Another problem with using the phase vocoder for analyzing transitions is that the center frequencies of the filters remain fixed once set, so that the harmonics of the new note no longer fall onto the analysis channels in a useful way. If two harmonics fall into one channel, a characteristic beating in both the amplitude and frequency traces for that channel is the result. (Dolson [1983] gives an especially clear account of this phenomenon). If some channels capture no harmonics, then their outputs must be selectively ignored.

Furthermore, as the signal leaves one channel and enters the neighboring channel, the amplitude of the signal is subject to distortion. The bandpass filters used to realize the phase vocoder have a rolloff of their own (see Figure 2.15). As long as a spectral component remains in the region shown at A–B in the figure, the magnitude output from the corresponding channel can be trusted. But the spectral component at C has its magnitude modified by the filter's own rolloff. Dolson

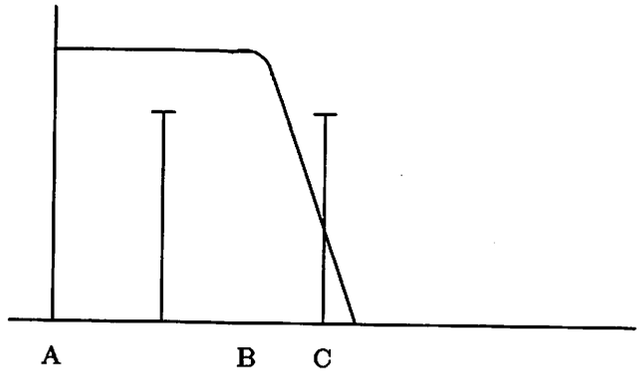


Figure 2.15. For a given channel of the phase vocoder, the amplitude of a component falling in the range A–B is unaffected by the analysis filter's frequency response. This is not the case for a component at, say, C.

(1983, pp. 37–38) suggested* that the magnitude of a signal at C could be scaled by the inverse of the filter's frequency response to recover the true magnitude. My own analysis and tests showed that this works well for steady-state signals, if the original magnitude is known.

The situation becomes more complicated with time-varying signals. Even when the frequency is moving slowly with respect to the filter bandwidth, this method works well. However, for pitch changes such as those found in the recordings used for the current work, the phase vocoder apparently does not track accurately enough to permit this amplitude compensation.

Mark Dolson and I have discussed this problem extensively; I ran a number of tests of the phase vocoder to convince myself that a problem might really exist. The simplest method to demonstrate the extent of the problem is to reproduce a few figures from Dolson's thesis, to which he has graciously consented.

Figure 2.16a shows the frequency response of a low-pass filter used to design the passbands of a phase vocoder analysis stage (taken from Figure 5b in Dolson's thesis, p. 39). In the same figure, c) shows the frequency of a sinusoidal test signal passing through the filter; the frequency of this test signal is swept from the center of a phase vocoder passband to well past the edge of the passband. The middle plot shows the amplitude *measured* at the output of the corresponding phase vocoder channel. (Figures 2.16b and 2.16c are Dolson's Figure 8, p. 42). The x-axis in a) is frequency offset from the channel's center frequency; 0 on this x-axis corresponds to 2000 Hz on the y-axis of the plot in c). The solid lines constructed on this figure are my own accretion.

*This idea was developed independently at CCRMA in conversations with Julius O. Smith and James A. Moorer, whose contributions are gratefully acknowledged.

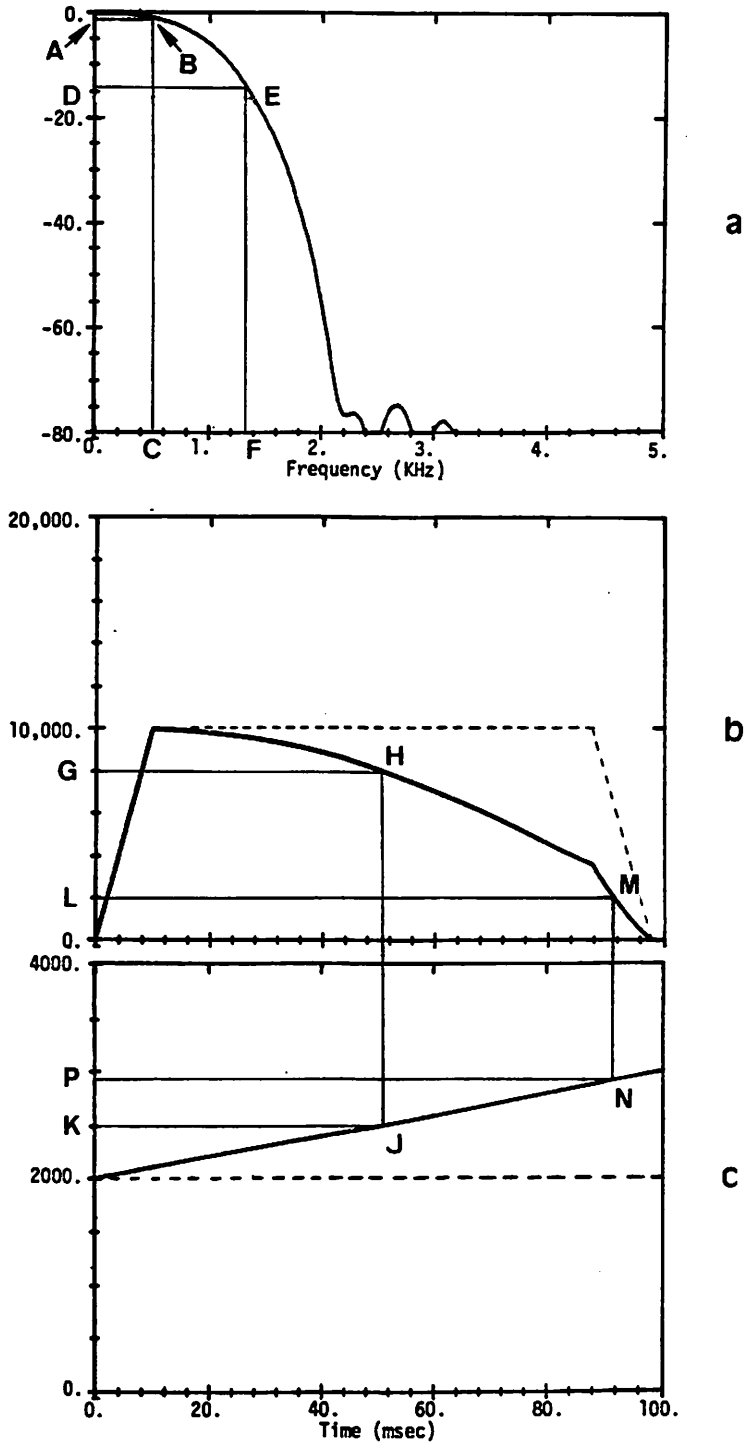


Figure 2.16. The response of a phase vocoder analysis channel to a test signal rapidly moving from the center past the edge of the passband (after Dolson [1983], pp. 39 and 42; reprinted with permission). a) Frequency response of the prototype analysis filter. b) Amplitude output by the channel swept by a sine wave whose frequency changes as shown in c).

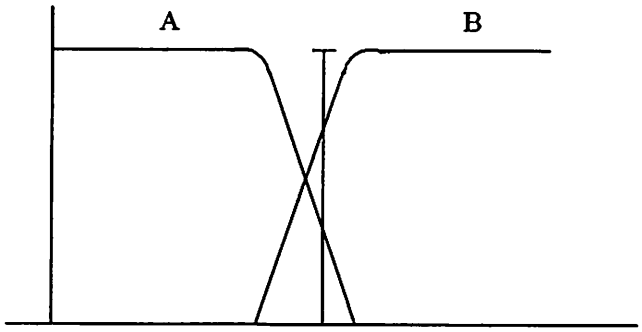


Figure 2.17. The amplitude of a spectral component lying between two adjacent channels of the phase vocoder may in some cases be calculated by combining the amplitudes from the two channels.

Point K in the lowest plot marks a displacement of 500 Hz from the center frequency of the channel, corresponding to point C in the top figure. By constructing the lines KJ, HJ, and GH, the magnitude of the ouptut in the middle plot can be read off as approximately 8000. Now $20\log(8000/10000)$ is approximately equal to -2 dB, which correctly corresponds to point A in the top figure. However, as the test signal continues its sweep, the output of the channel becomes less reliable. The magnitude value of 2000 at point L in the middle figure is reached when the frequency is at about 2850 Hz on the y-axis of the lowest figure; this is determined by constructing the lines LM, MN, and NP. The value of 2000 corresponds to about -14 dB; but in the uppermost plot, -14 dB (shown at point D) falls at a frequency of about 1400 Hz (found by constructing DE and EF), which translates to a frequency of 3400 Hz in the lowest figure—a frequency which the test signal never even reaches.

In actuality, the signal at point C in Figure 2.15 falls within two overlapping filters; this situation is shown in Figure 2.17. It turns out that the correct magnitude can be recovered even in the degenerate case being considered here by combining the magnitude of the output of the filter labelled A in Figure 2.17 with the magnitude of the B filter. This would work, of course, only if no other spectral components had entered the passbands of either filter; but since harmonically related spectral components would presumably be moving together, there would most likely be a time when every filter would fall over two spectral components: one at each edge of the filter. Based on observations of test signals, an even larger problem appeared. The method for combining the outputs of A and B varies with the length of the analysis filter h in Equation 1.3, the number of channels N, and the like. In one instance, the outputs of the two filters could be combined linearly; for another parameter setting, the filters had to be combined according to $\sqrt{A^2 + B^2}$;

and so on. For practical work, it would be unreasonable to have to determine this method for every new set of analysis parameters.

Further work with real musical tones suggested that none of this would cause problems; it appeared that the frequency and amplitude traces produced by the phase vocoder were adequate for resynthesizing the transitions. Experiment 1 in Chapter 3 will show that this is indeed the case.

Creating Spectral Plots (Amplitude)

The problem remained of how to produce useable spectral plots. It was certainly possible to run the phase vocoder twice: once for each note, with the filter center frequencies adjusted accordingly. The end of the first note analyzed in this manner looked reasonable; and so did the beginning of the next.

The only way I could find to make useful spectral plots was to splice together these two analyses in a three-dimensional representation. It was necessary to expand my spectral editor (Strawn 1985a) to handle these two analyses properly. Figures 2.18 (the tongued clarinet ascending major third) and 2.19 (the same interval, played untongued) show a sample of the result. These are the same ascending thirds already presented in Figures 2.3, 2.4, 2.9, and 2.10.

In these plots, time runs from left to right. The fundamental is at the top of the plot; higher-order harmonics are plotted along their own axes, which are arranged below the fundamental on the page. One should imagine this spectral plot as "coming out toward" the viewer from the fundamental "at the back". Each harmonic is plotted on a scale of 0 to −60 dB, with 0 dB being the maximum of the strongest harmonic in the entire plot. At the point specified in the caption, the plotting program switches from the phase vocoder analysis for the first note to that of the second; this is approximately the point of pitch change.

Clearly, there is a spectral rolloff at the end of the first note in the tongued transition of Figure 2.18; of the thirty harmonics shown here, perhaps the top 20 drop out. Note that the pattern with which the harmonics drop out and re-enter is not entirely regular. However, in general the higher-order harmonics leave sooner and re-enter later than their lower-frequency counterparts. This corresponds to the "wedge-like funnel, open toward the higher frequencies," found by Rösing (see Chapter 1). Comparison of Figure 2.18 with Figure 2.19 shows that this change in the spectrum is not so pronounced for the untongued case, where fewer harmonics drop out and the gap width is shorter (both plots show 300 msec). The fact that the upper harmonics

experience such a strong drop in amplitude might explain why tracking vocoders have trouble following them in a transition (Dolson 1983, p. 107).

To cite some more examples, Figures 2.20 and 2.21 show the tongued and untongued transitions, respectively, for the ascending major third on the trumpet. These transitions were already shown in Figures 2.5, 2.6, 2.11, and 2.12. As the attack of the second note begins, there is some noise visible in the higher-order harmonics. This would have been lost in Rösing's plots, I suspect. The irregularity in the pattern according to which the harmonics drop out and re-enter is striking. Another set, this time for the violin, is given in Figures 2.22 and 2.23; here we have the ascending third, already met in Figures 2.7, 2.8, 2.13, and 2.14.

There are sometimes spectral cues specific to one instrument. For example, the "blips" associated with the attack of the brass can be (barely) seen in the attack of the trumpet notes. However, the following generalization applies to all of the instruments analyzed: *the spectrum of the transition before the point of pitch change can be conveniently characterized as a low-pass filtered version of the spectrum at the end of the steady-state of the first note.*

More plots of this kind have been relegated to Appendix 5 (except for the cello and flute, to which we will turn shortly). The short introductions to Appendices 4 and 5 give an overview of why certain intervals were selected for inclusion there.

The following conclusion is based on an analysis of three-dimensional time-varying spectral plots for all (!) of the recorded transitions: *Fewer harmonics drop out in the untongued transition, and the gap in the spectrum is shorter in duration, than in the tongued transition. I found this to be a general principle, no matter the size of the interval performed, the direction of the interval, or the instrument playing.* This matches closely what was observed previously for time-varying power.

Plots of the Frequencies

For a single note, it is possible to create three-dimensional plots of the time-varying frequency traces similar to those for the amplitude traces. Three-dimensional plots of the frequencies in a transition did not prove to be useful. If one plots a large number of channels, then the frequency resolution is too coarse; this problem is compounded when the interval between the notes is wide. Plotting a few channels at a time fails to give information about the overall spectrum, which was so important with the amplitude plots. For the sake of completeness, a few plots of this kind are given in Figures A5.18–A5.20 in the appendix.

(Text continued on p. 64)

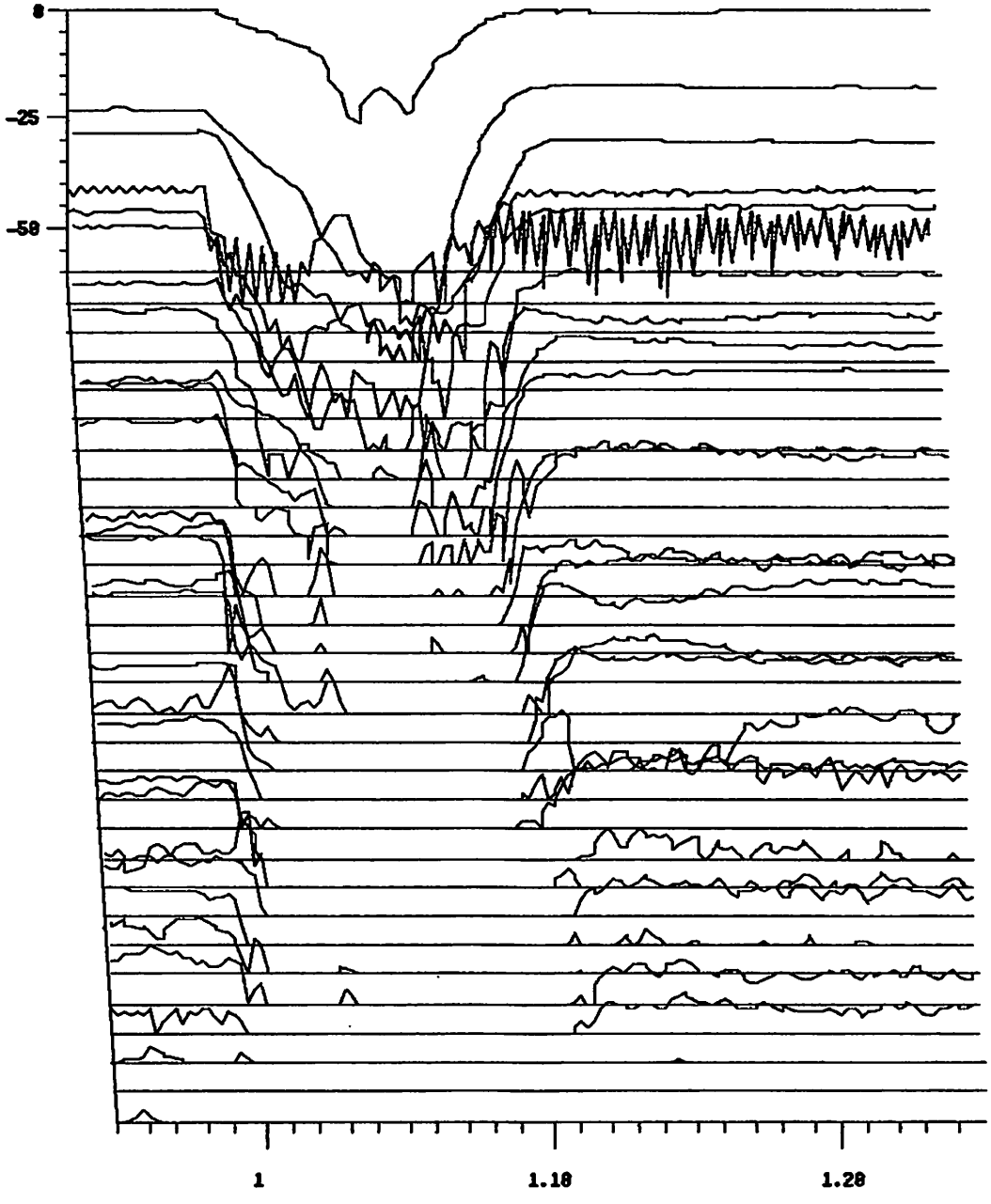


Figure 2.18. Time-varying spectral analysis (30 harmonics) of a tongued ascending major third played on the clarinet. The lower note is A220; the splice point is at t=1.05 sec.

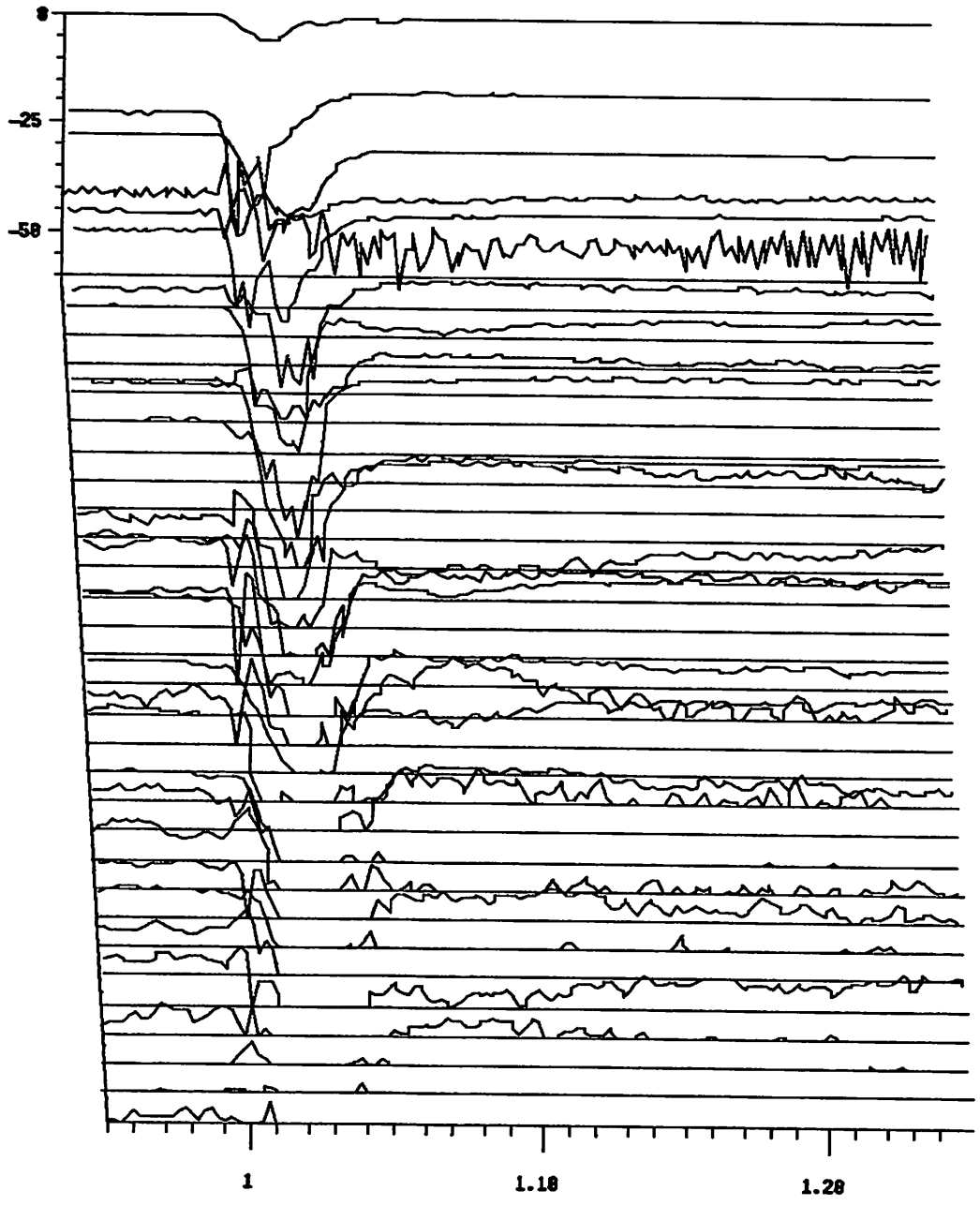


Figure 2.19. Time-varying spectral analysis (30 harmonics) of an untongued ascending major third played on the clarinet. The lower note is A220; the splice point is at t=1.02 sec.

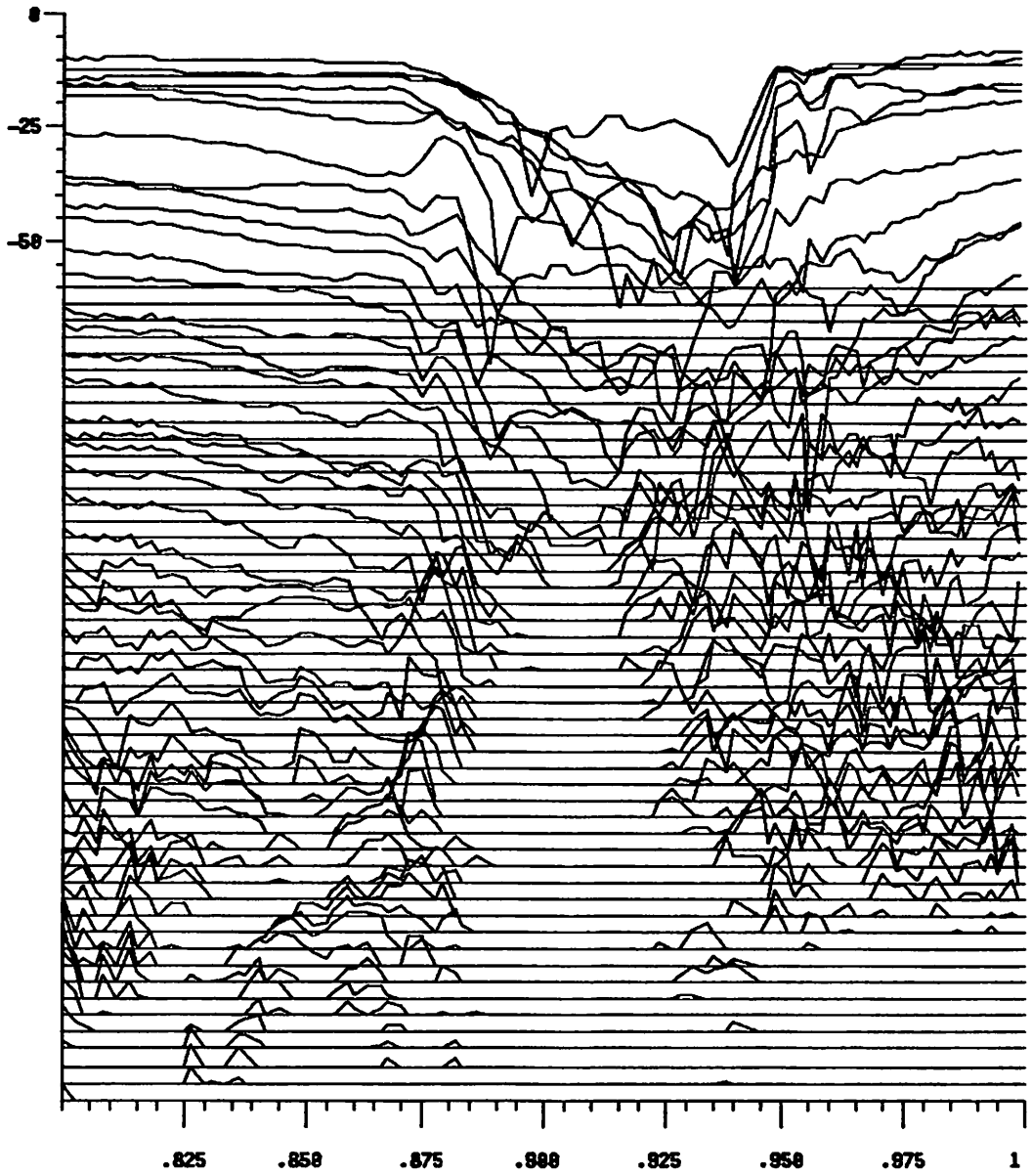


Figure 2.20. Time-varying spectral analysis (50 harmonics) of a tongued ascending major third played on the trumpet. The lower note is A220; the splice point is at t=0.95 sec.

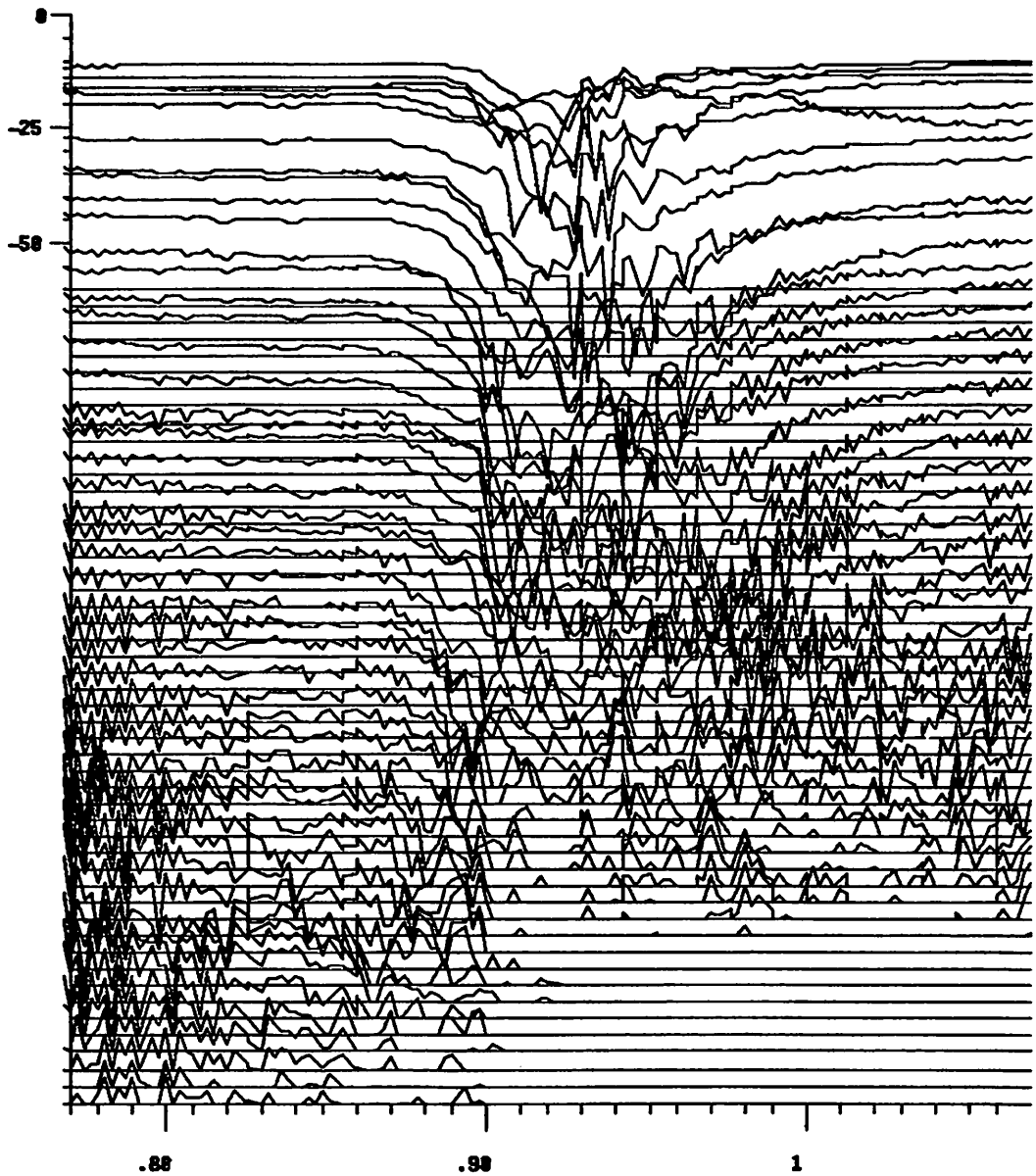


Figure 2.21. Time-varying spectral analysis (50 harmonics) of an untongued ascending major third played on the trumpet. The lower note is A220; the splice point is at t=0.925 sec.

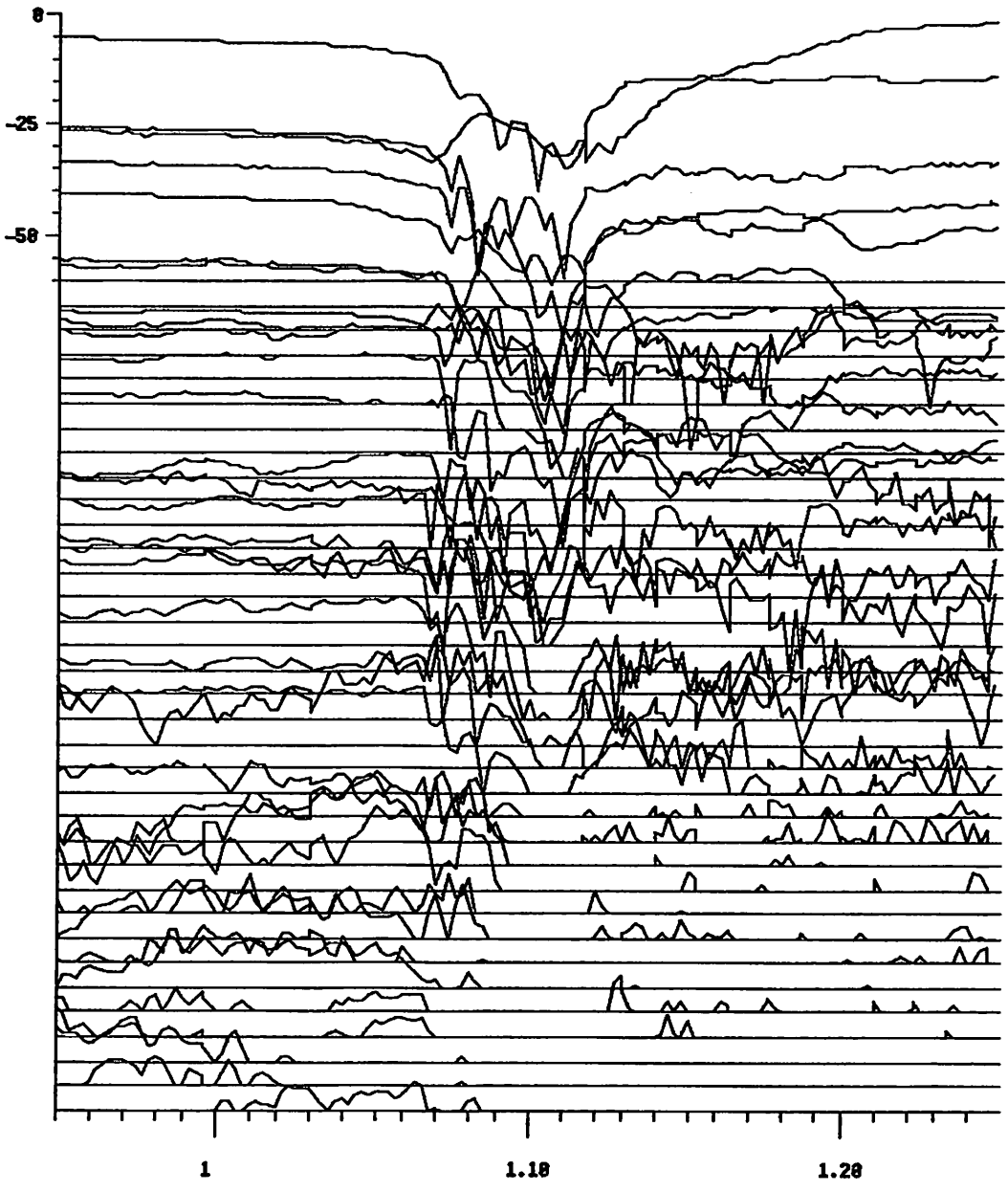


Figure 2.22. Time-varying spectral analysis (35 harmonics) of an ascending major third played with bow change on the violin. The lower note is A220; the splice point is at t=1.11 sec.

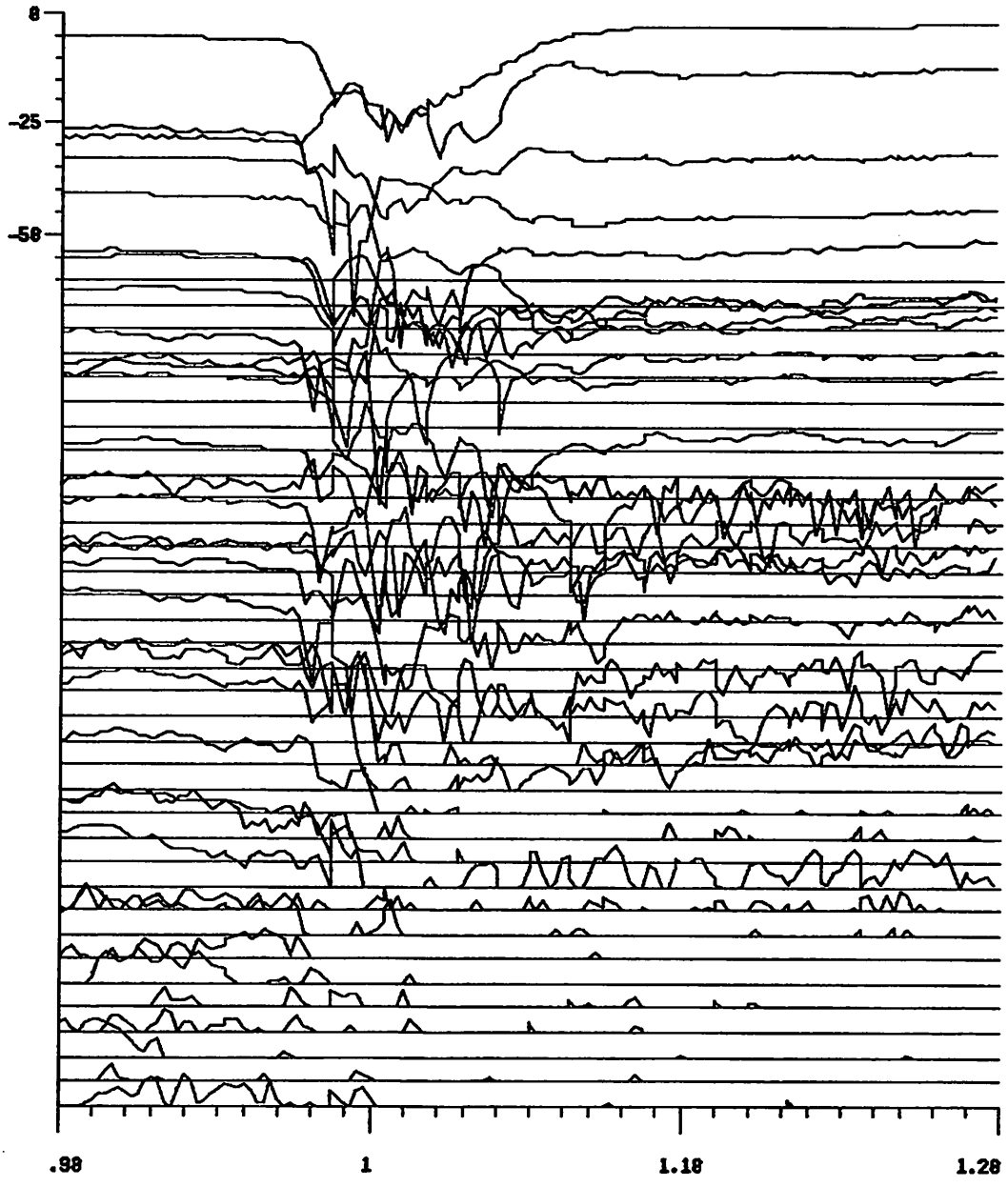


Figure 2.23. Time-varying spectral analysis (35 harmonics) of an ascending major third played with no bow change on the violin. The lower note is A220; the splice point is at t=1.00 sec.

Masking Effects in the Transition

The three-dimensional amplitude plots might imply with their visual impact a difference between the playing styles that does not correspond to their audio differences. It must be remembered that in the three-dimensional plots already given, the magnitudes output by the phase vocoder do not drop to 0.0 when the harmonic drops out of the picture; rather, those harmonics continue to be present but at a very small amplitude. Perhaps those low-level harmonics really "fill in" the tongued transition. Or, seen from another perspective, it might be possible that the amplitudes of the low-frequency harmonics are so loud that in the nontongued cases they mask the high-frequency components, making the distinction shown in the plots misleading.

To test this, I modified the phase vocoder analysis outputs of the trumpet ascending M3 (for which plots have already been given in Figures 2.20 and 2.21) so that the amplitude traces were forced to 0.0 at some threshold below the maximum of each note: −30 dB, −40, −50, and −60.

From these "squelched" analyses, I created new three-dimensional plots, still using the 60 dB range of the earlier plots. The results are shown in Figures 2.24 and 2.25 (tongued and untongued, respectively, squelched to −30 dB); Figures 2.26 and 2.27 (−40 dB); and Figures 2.28 and 2.29 (−50 dB). Retaining the −60 dB dynamic range gives these plots an unusual appearance: the harmonics appear to be "standing on stilts." Still, this facilitates comparison among the three "squelching" threshold and the originals.

These "squelched" analysis data were used to synthesize test stimuli. It turns out that even for the tones squelched to −30 dB, the difference between tonguing and nontonguing is still audible in the resynthesis, and visible in the plots. Admittedly the auditory difference is not so striking as the visual difference. Still, masking does not seem to play a role in the listener's ability to distinguish these transitions. Incidentally, these resyntheses are unsuitable for further experimentation, as they are characterized by "breebles" that occur when one attempts arbitrary filtering of time-varying spectral analysis data. In other words, the isolated "mountain peaks" in the plots cause problems. Besides, the ear's overall impression is that the signal sounds low-pass filtered.

Variation from one Performance to the Next

What I tell you three times is true. (Lewis Carroll, *The Hunting of the Snark*, 1876)

The possibility remained that, due to the quirks of fate, time-varying power and spectral characteristics of the recorded tongued and untongued transitions fell only coincidentally into the patterns which seemed to occur. To examine this, as many as five separate recordings were made (by the same performer) for a given instrument, interval size, interval direction, and playing style. One example of such a set of duplicate recordings, from the flute, will be presented here.

Figures 2.30 and 2.31 show three recordings each for the tongued and untongued (respectively) ascending major thirds played on the flute; we see here the time-varying power, analyzed in the manner discussed earlier in this chapter. The differences *between* the tongued and untongued cases are much greater than the differences *among* the repeated recordings for each case. The same conclusion is supported by the time-varying spectral plots for these same transitions. Figures 2.32, 2.34, and 2.36 show the tongued transitions of Figure 2.30; the untongued transitions are given in Figures 2.33, 2.35, and 2.37. Such multiple sets of amplitude and spectral plots were collected for many of the transitions recorded here. (More are given in Appendix 4). Thorough examination of this data quickly led to the conclusion that the performers could reliably reproduce a given transition, and that the transitions illustrated in the figures in this chapter and in the appendices were representative performances.

Incidentally, there were 212 recordings made and analyzed, including these repeated cases.

On the Effects of Instrument Size

Analysing power plots for the cello, given in Figures 2.38 (with bow change) and 2.39 (no bow change), proved to be a problem. Recall that on the violin only the seconds and thirds were recorded because it is awkward to finger a larger interval on one string; the same problem of course occurs in the cello. These plots should be compared with the violin recordings in Figures 2.13 and 2.14, respectively.

(Text continued on p. 84)

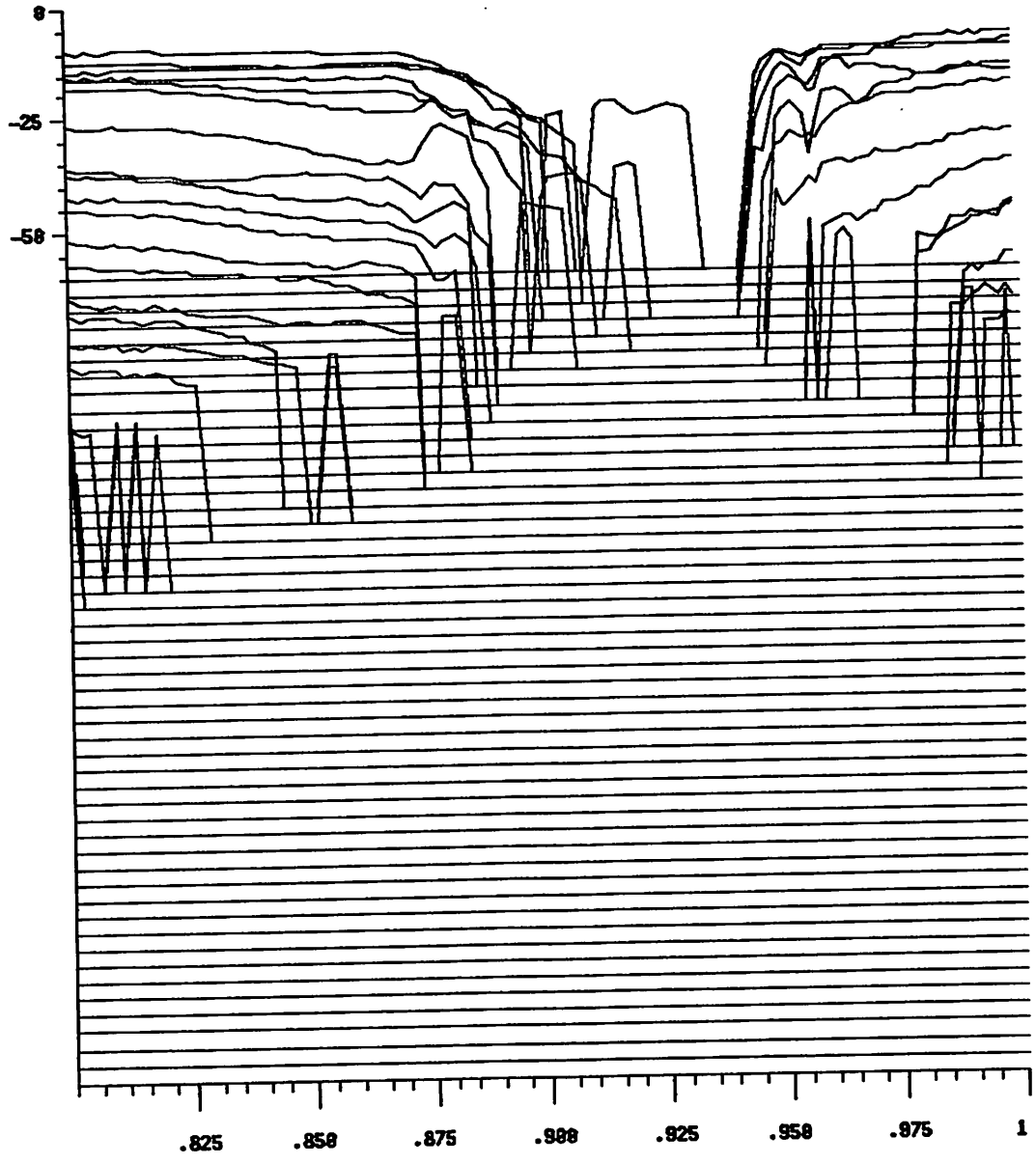


Figure 2.24. Time-varying spectral analysis of a tongued ascending major third played on the trumpet, as in Figure 2.20, but with the amplitude traces squelched to 0 when they fall below −30 dB. The splice point is at t=0.95 sec.

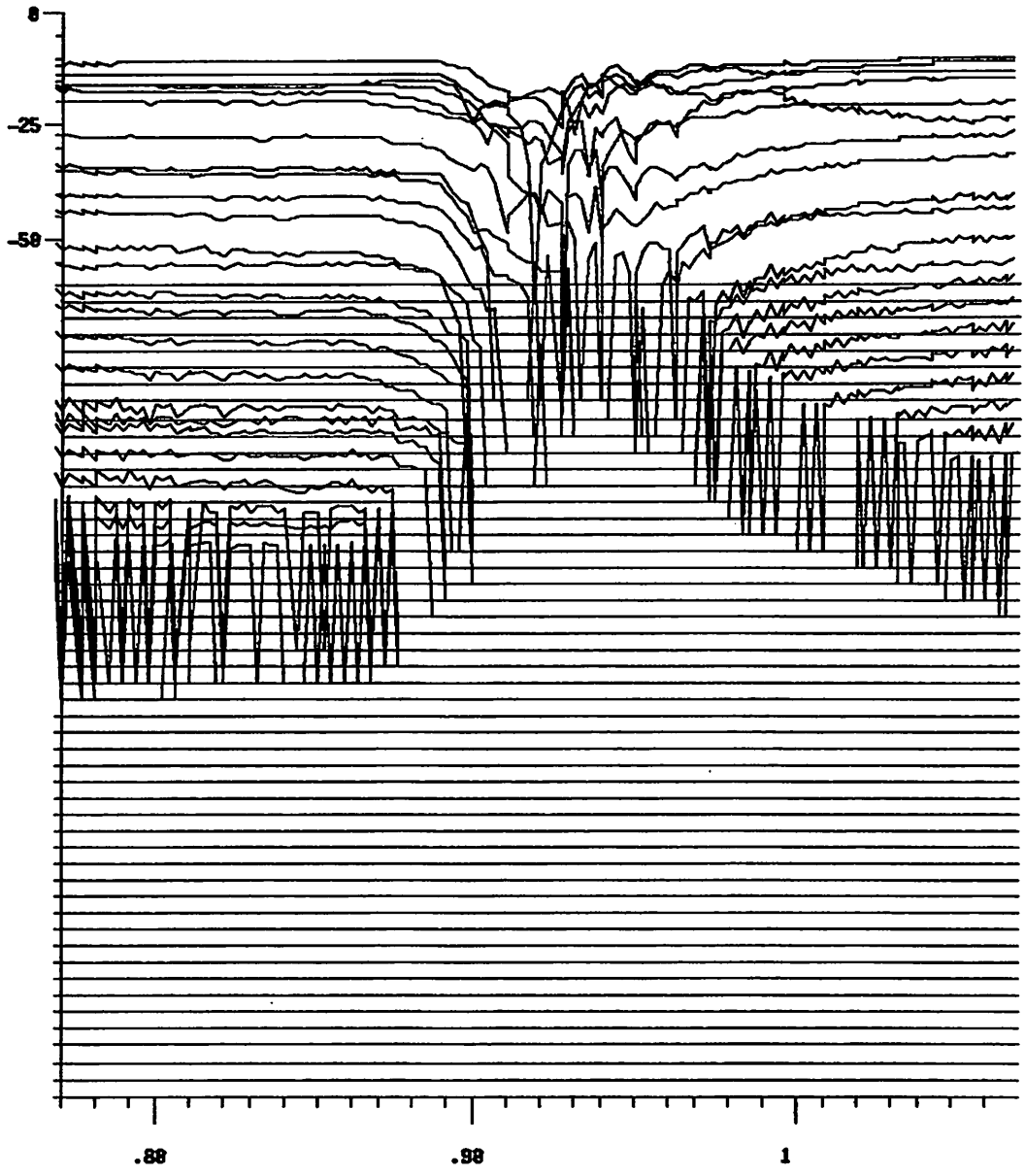


Figure 2.25. Time-varying spectral analysis of an untongued ascending major third played on the trumpet, as in Figure 2.21, but with the amplitude traces squelched to 0 when they fall below −30 dB. The splice point is at t=0.925 sec.

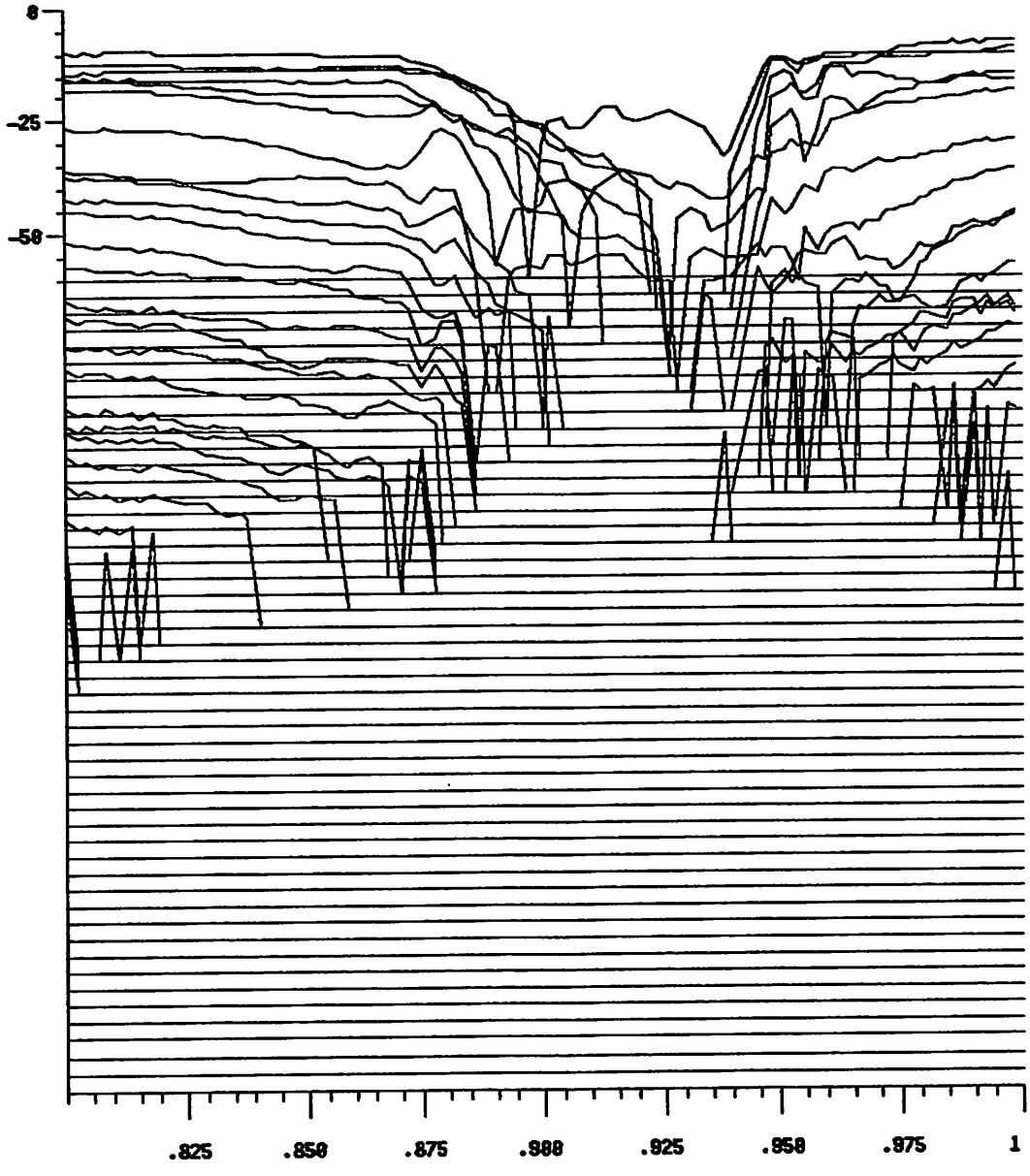


Figure 2.26. As in Figure 2.24, but with the amplitude traces squelched below −40 dB.

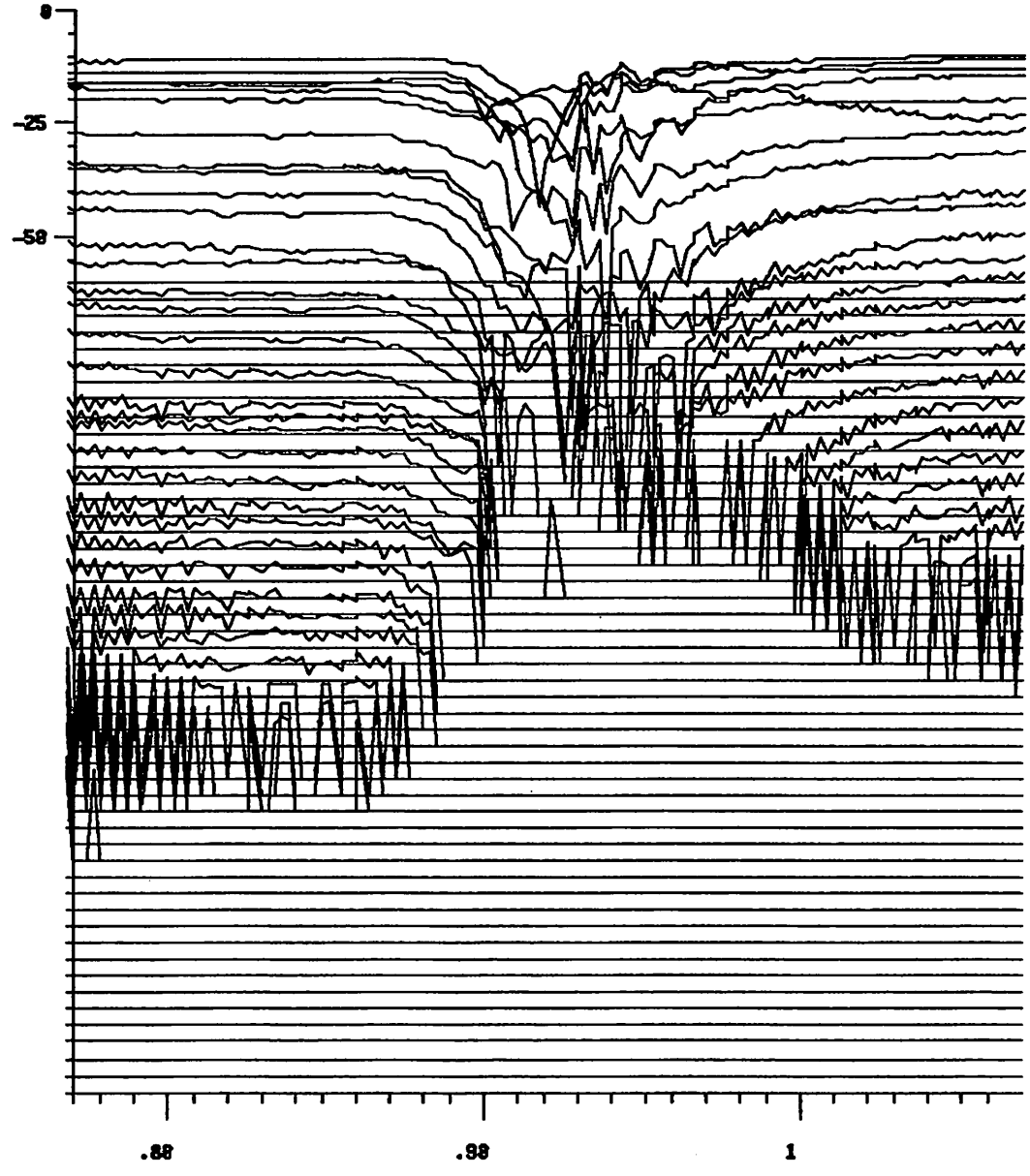


Figure 2.27. As in Figure 2.25, but with the amplitude traces squelched below −40 dB.

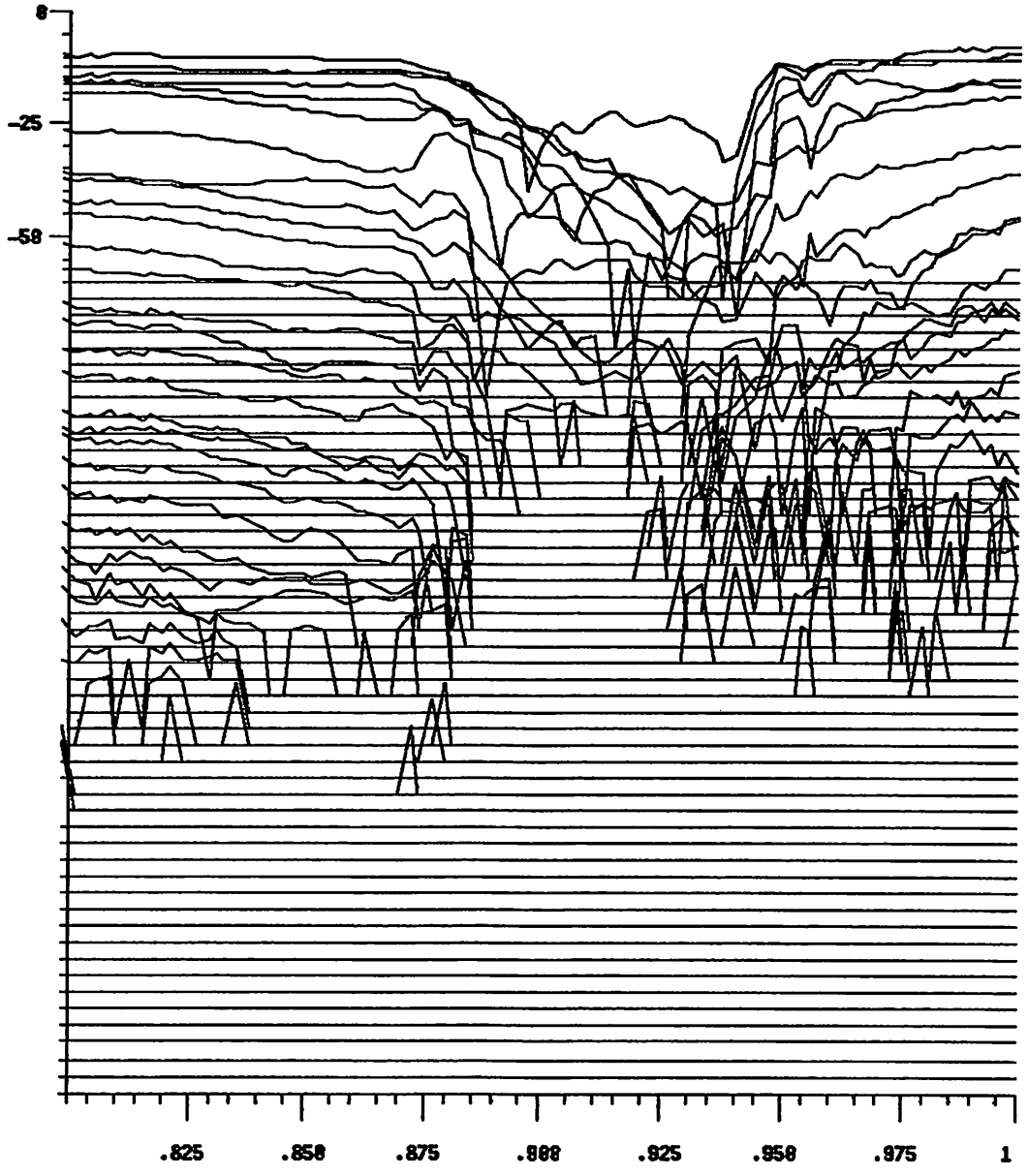


Figure 2.28. As in Figure 2.24, but with the amplitude traces squelched below −50 dB.

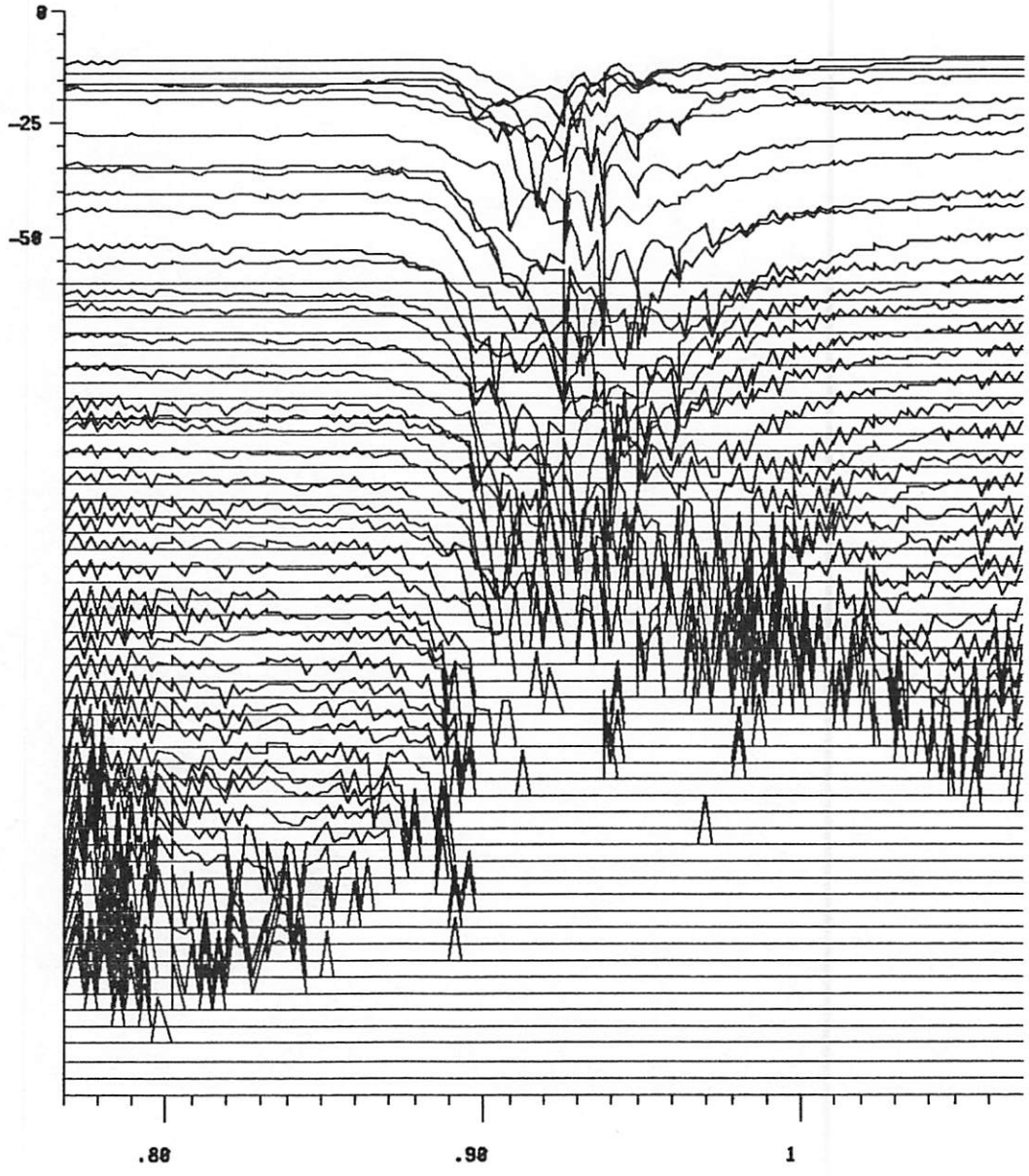


Figure 2.29. As in Figure 2.25, but with the amplitude traces squelched below −50 dB.

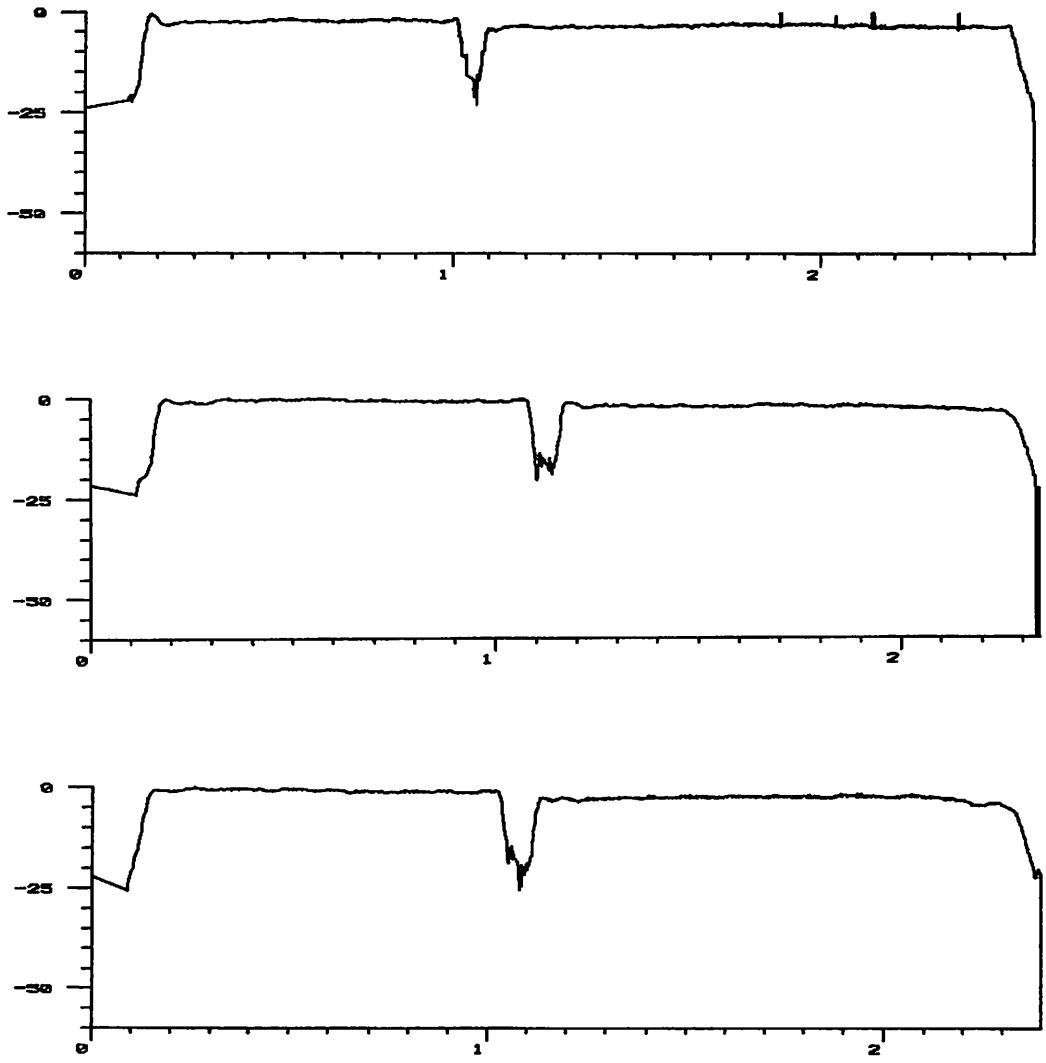


Figure 2.30. Three different recordings of ascending third tongued transitions on the flute. The lower (first) note is A220.

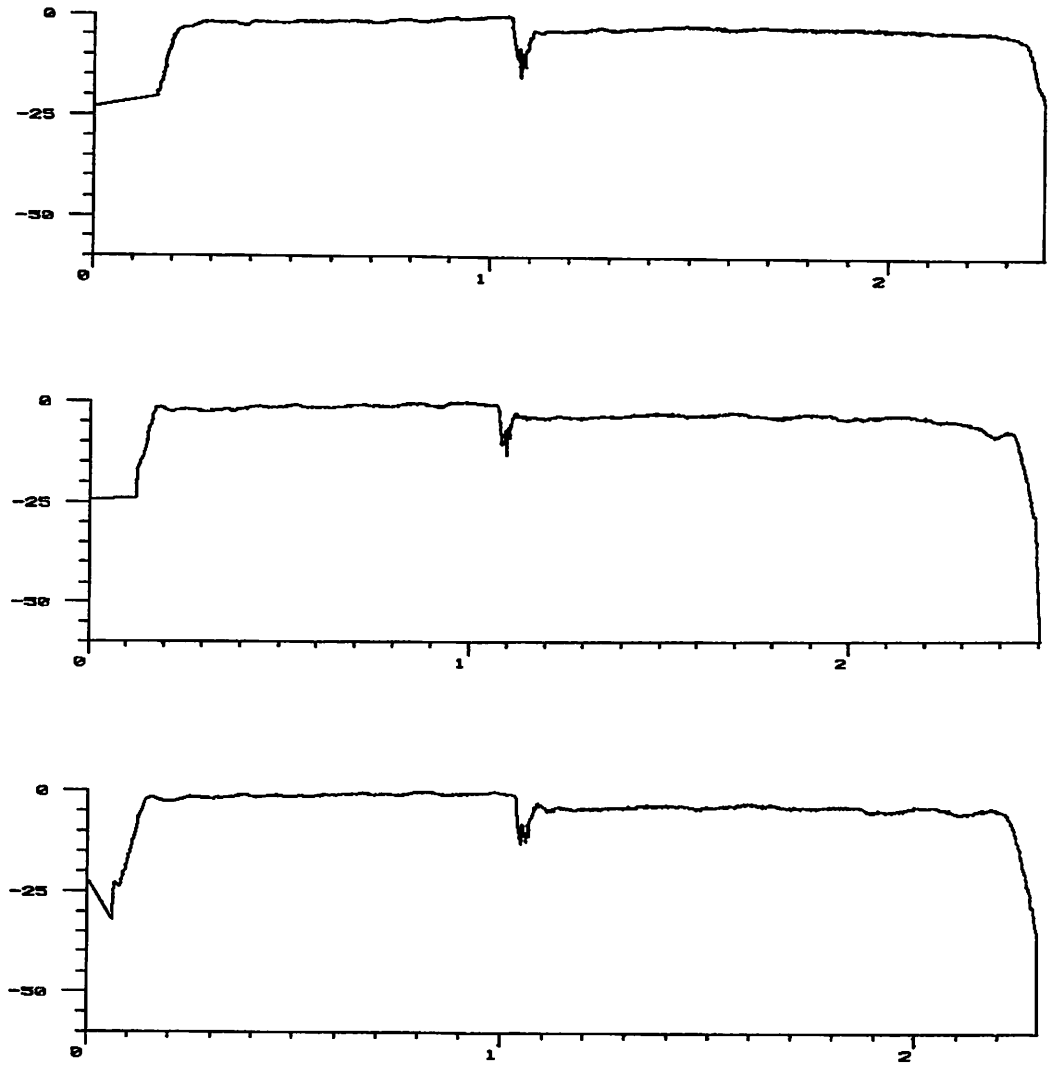


Figure 2.31. Three different recordings of ascending third untongued transitions on the flute. The lower (first) note is A220.

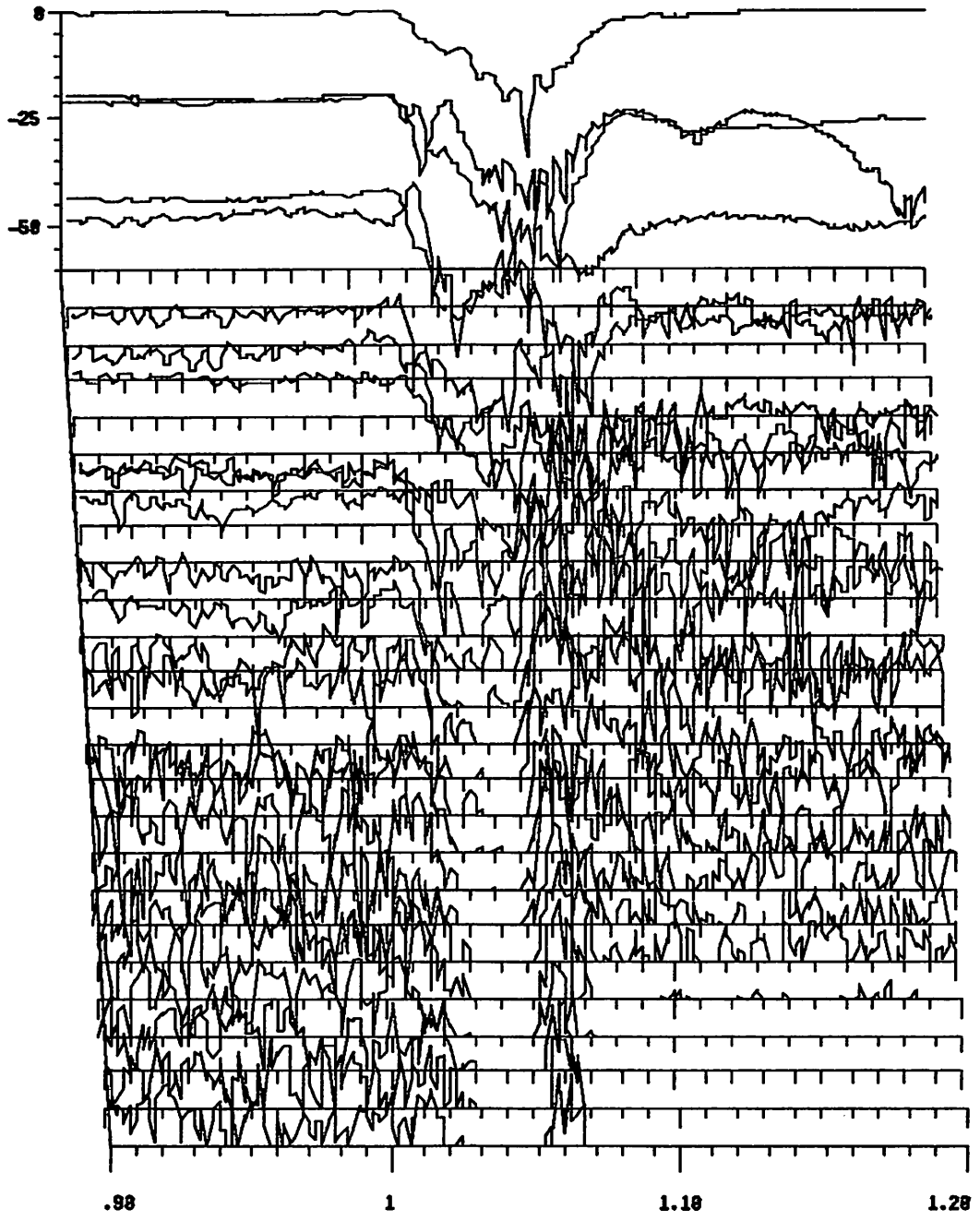


Figure 2.32. Time-varying spectral analysis (25 harmonics) of a tongued ascending major third played on the flute (cf. the top plot in Figure 2.30). The lower note is A220; the splice point is at t=1.065 sec.

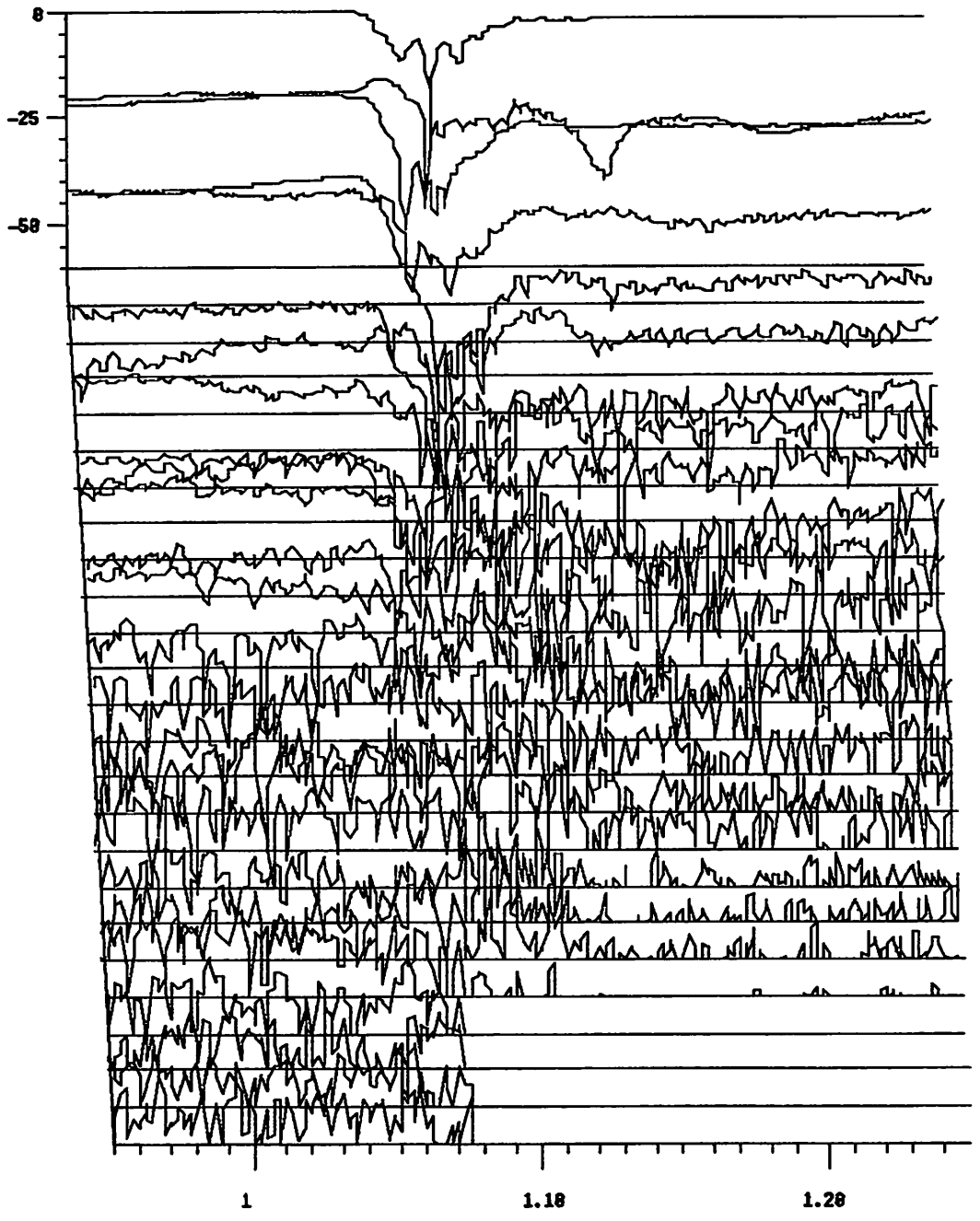


Figure 2.33. Time-varying spectral analysis (25 harmonics) of an untongued ascending major third played on the flute (cf. the top plot in Figure 2.31). The lower note is A220; the splice point is at t=1.075 sec.

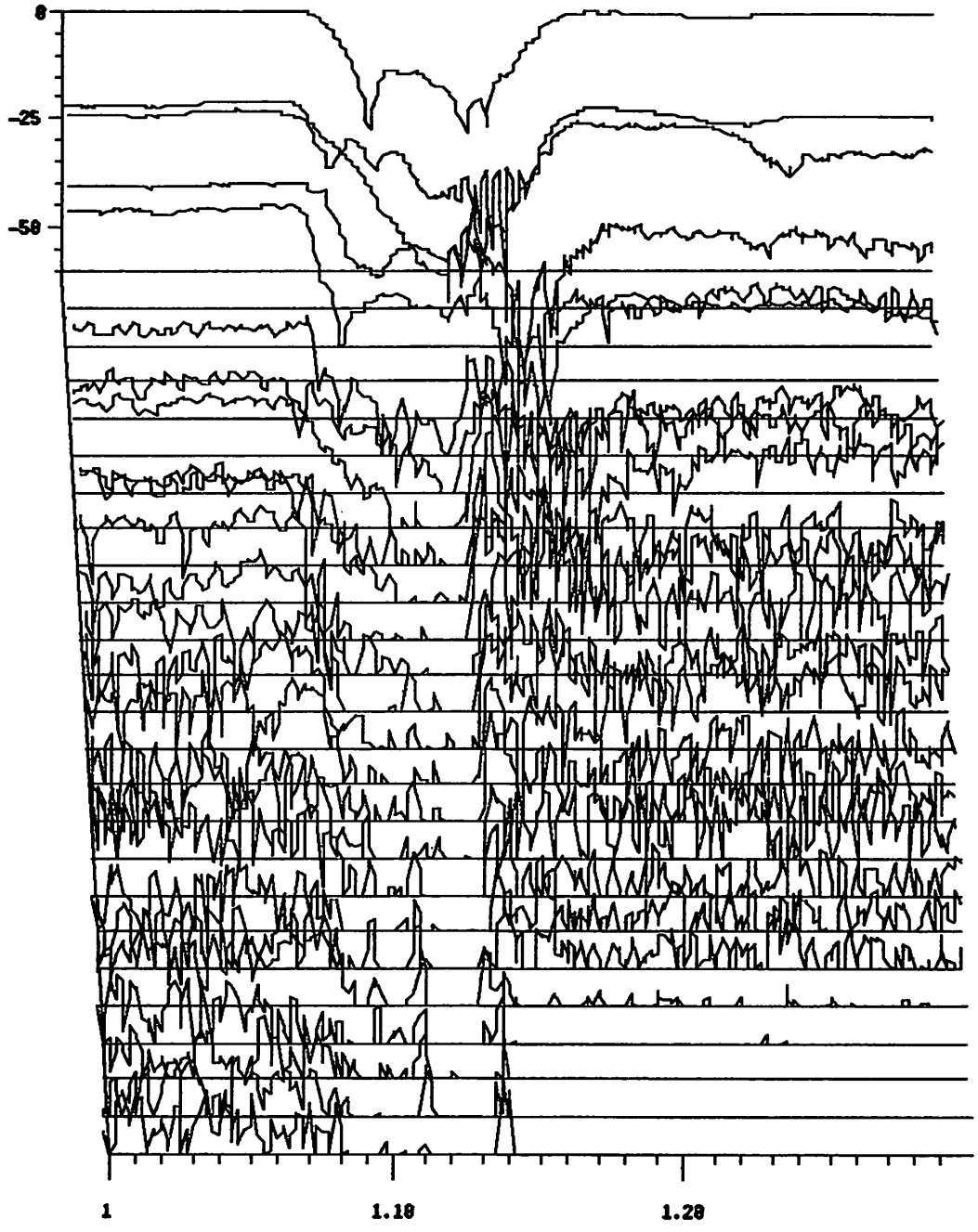


Figure 2.34. Another time-varying spectral analysis (25 harmonics) of a tongued ascending major third played on the flute (cf. the second plot in Figure 2.30). The lower note is A220; the splice point is at t=1.14 sec.

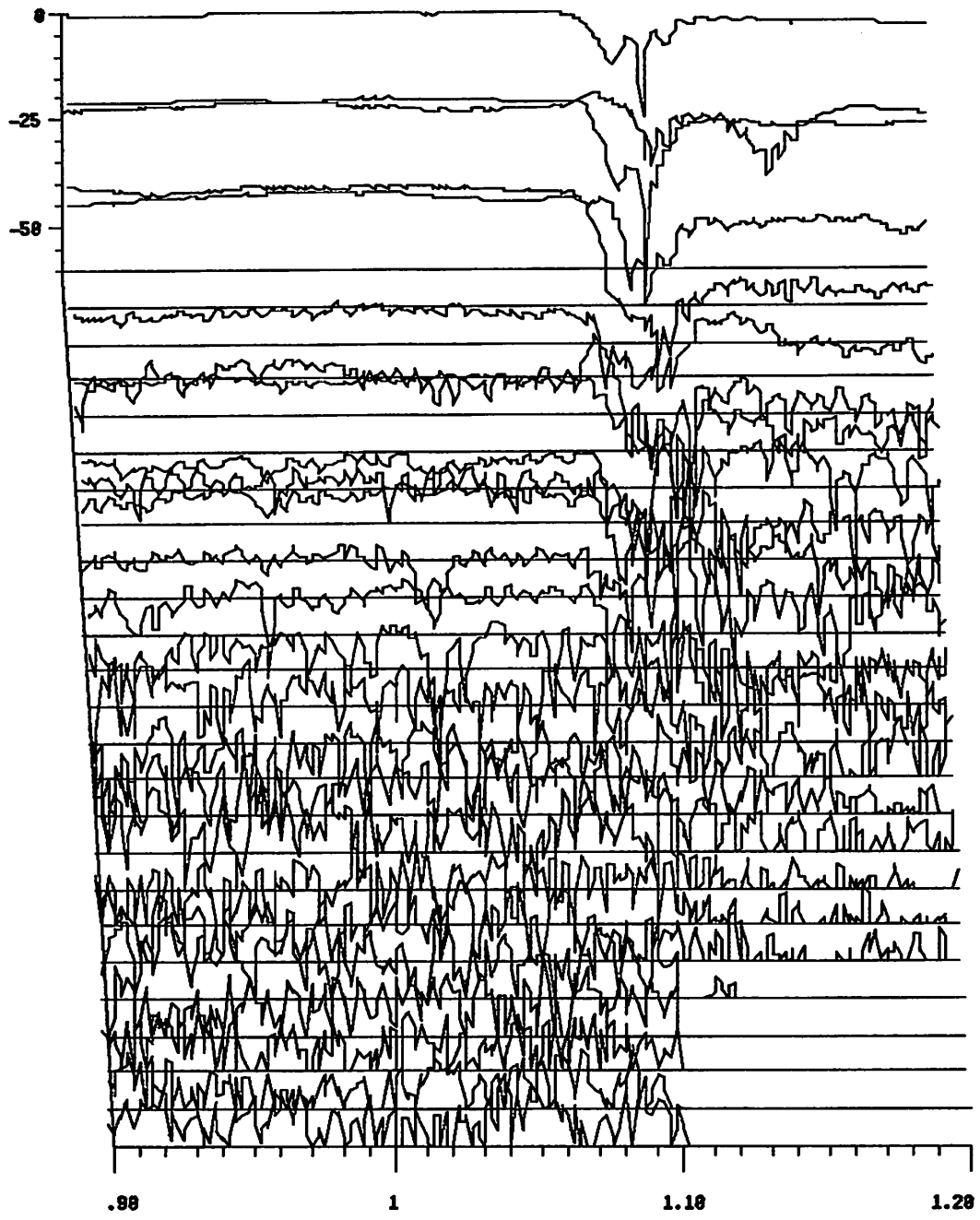


Figure 2.35. Another time-varying spectral analysis (25 harmonics) of an untongued ascending major third played on the flute (cf. the second plot in Figure 2.31). The lower note is A220; the splice point is at t=1.10 sec.

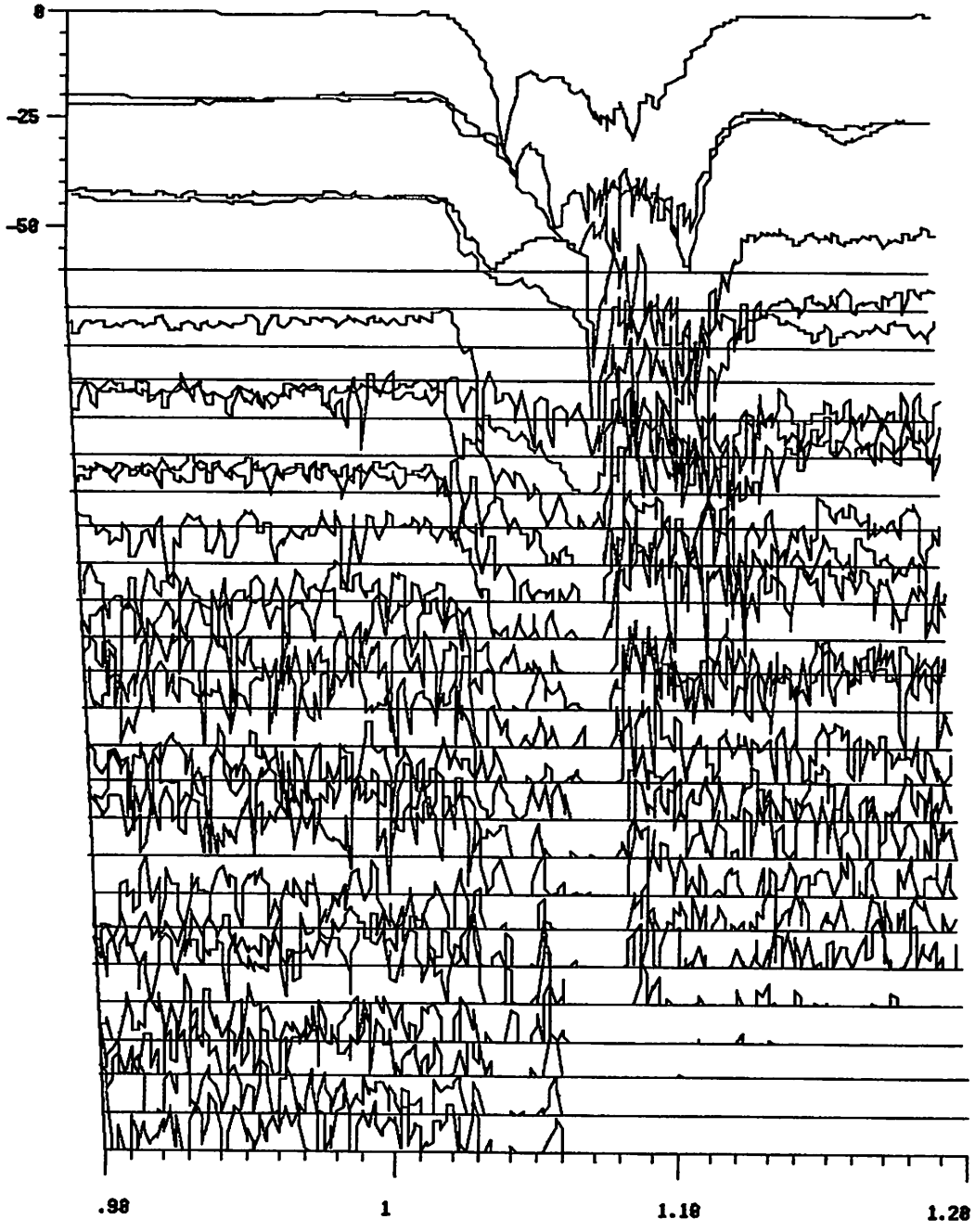


Figure 2.36. Yet another time-varying spectral analysis (25 harmonics) of a tongued ascending major third played on the flute (cf. the third plot in Figure 2.30). The lower note is A220; the splice point is at t=1.08 sec.

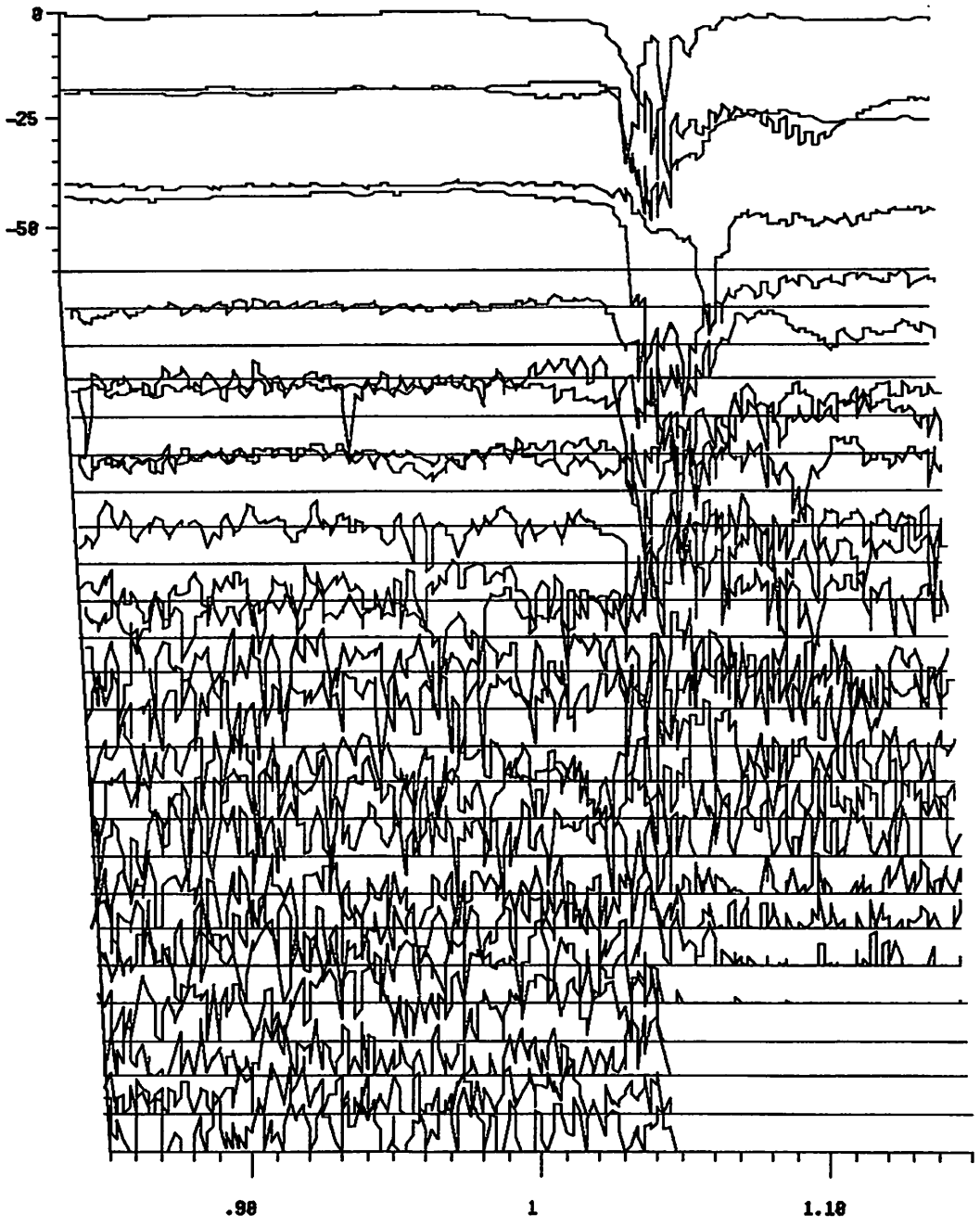


Figure 2.37. Yet another time-varying spectral analysis (25 harmonics) of an untongued ascending major third played on the flute (cf. the third plot in Figure 2.31). The lower note is A220; the splice point is at t=1.045 sec.

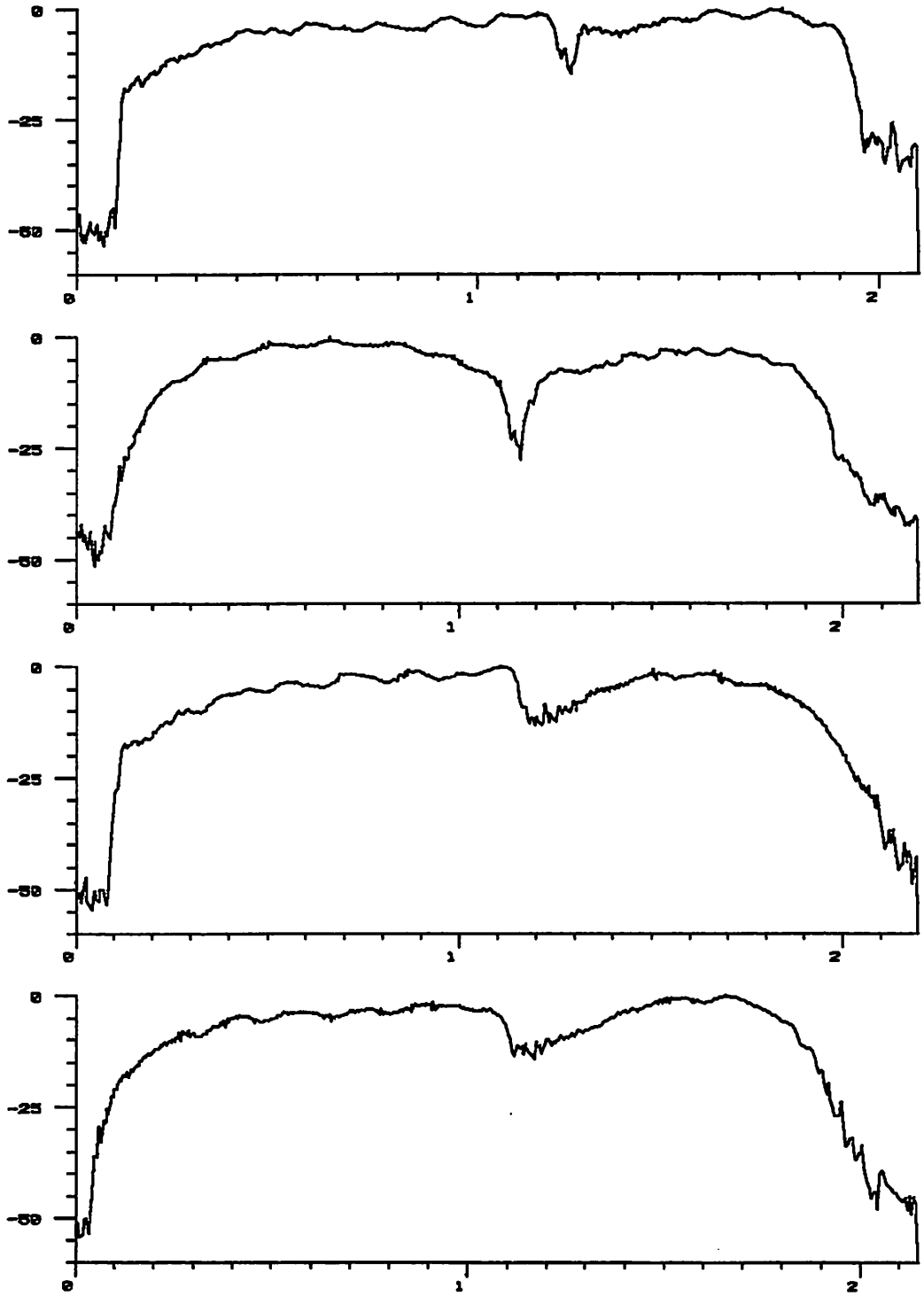


Figure 2.38. Period-synchronous power of two notes on the cello, with change of bow direction for the second note. The lower note is A220 in each case. From the top: ascending second, descending second, ascending third, decending third.

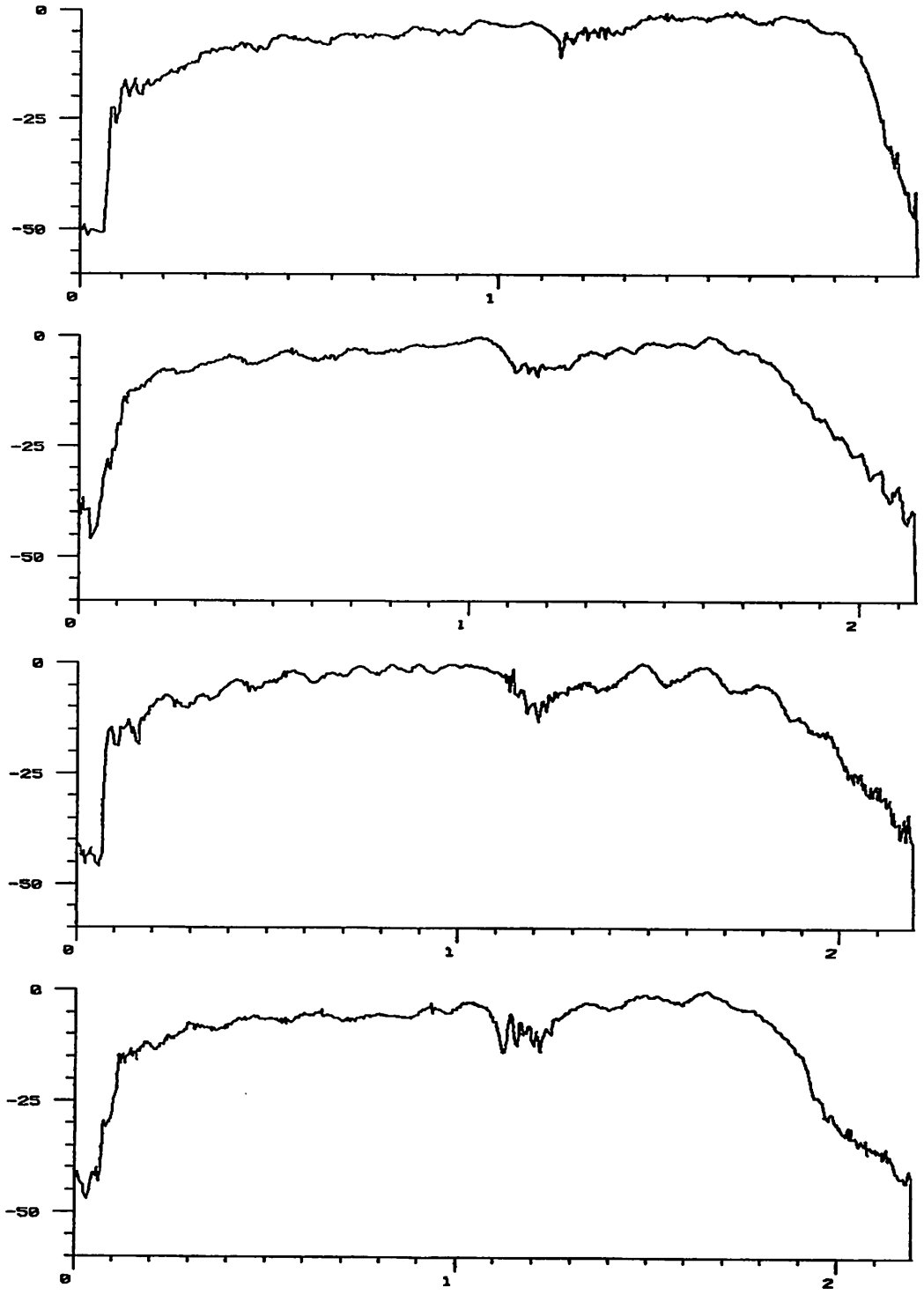


Figure 2.39. Period-synchronous power of two notes on the cello, with no change of bow direction for the second note. The lower note is A220 in each case. From the top: ascending second, descending second, ascending third, decending third.

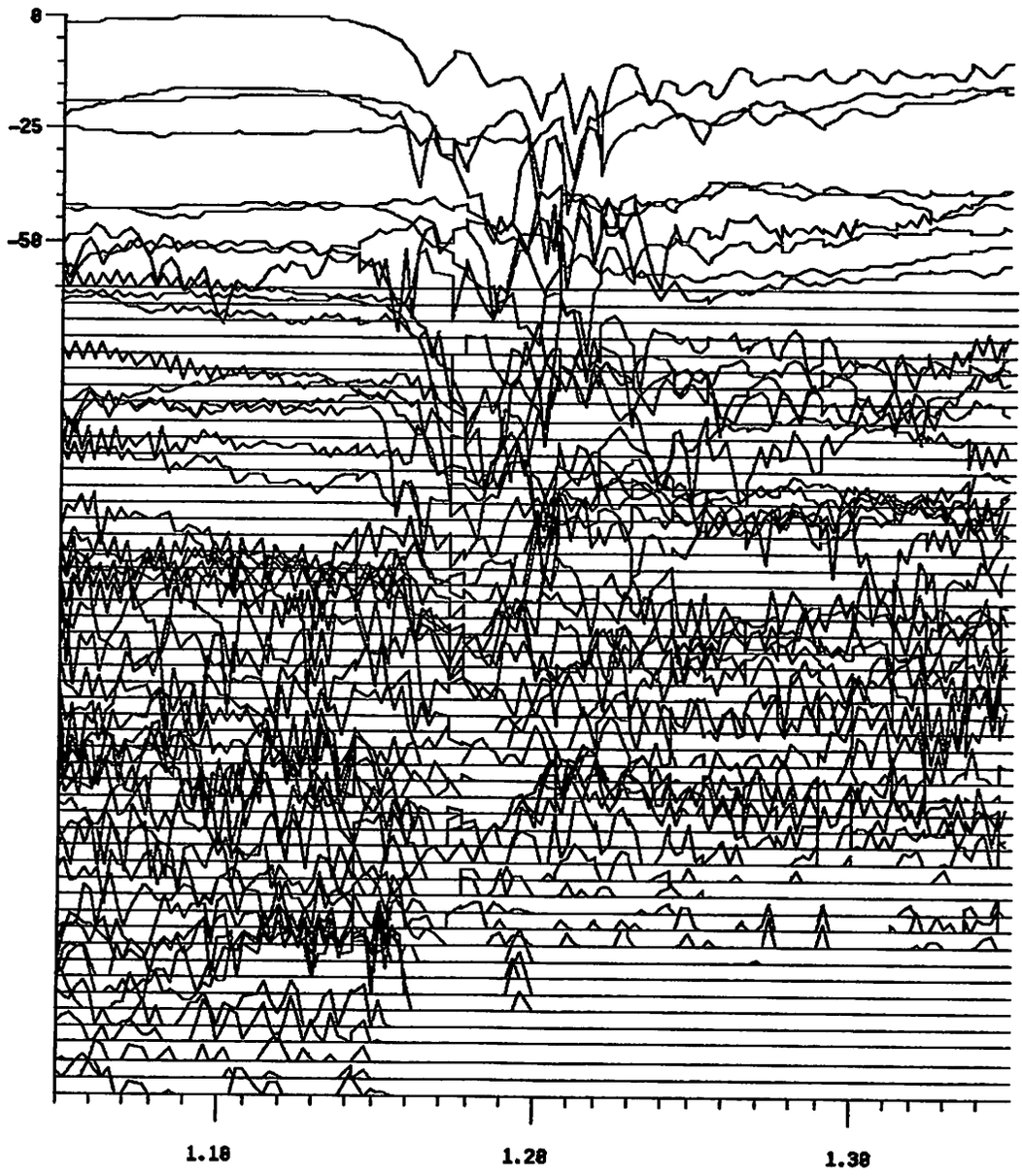


Figure 2.40. Time-varying spectral analysis (50 harmonics) of an ascending major third played with bow change on the cello. The lower note is A220; the splice point is at t=1.20 sec.

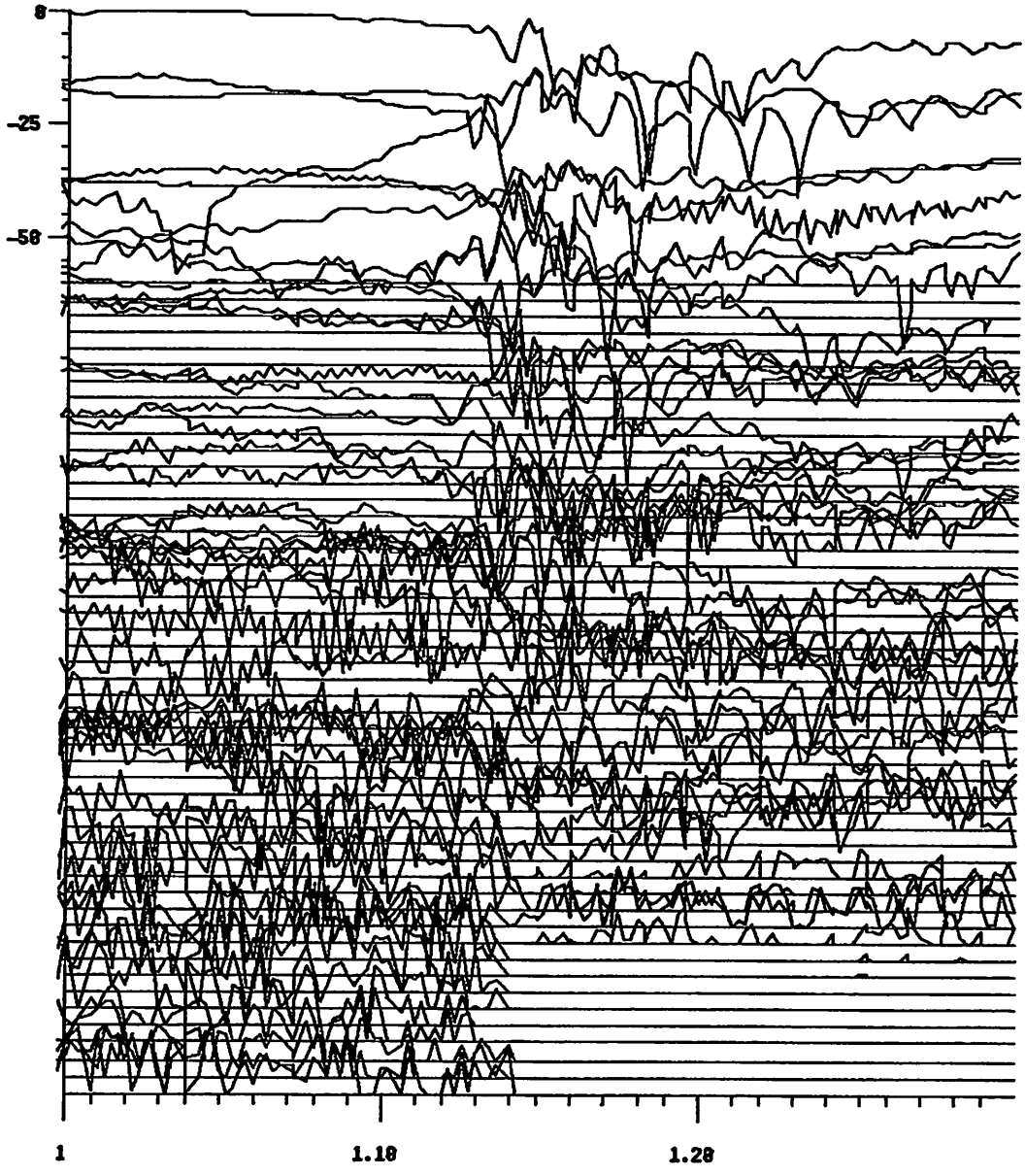


Figure 2.41. Time-varying spectral analysis of an ascending major third played with no bow change on the cello. The lower note is A220; the splice point is at t=1.14 sec.

Now the power plots for the seconds on the cello follow the pattern established previously. The amplitude dip for the bow change is deeper, and the time gap is wider, than without bow change. The distinction seems to disappear with the thirds, however. Perhaps the mass of the thicker cello strings is large enough, or the resonances in the cello body last long enough, that the expected gap is "blurred" in the bow-change recordings. Another interpretation might be that this is one of those cases where the cellist was able to create an extremely smooth bowed legato.

Examination of the spectral plots helps clarify the situation. Figures 2.40 and 2.41 show the time-varying spectrum of the ascending thirds with and without bow change, respectively. (These plots should be compared with Figures 2.22 and 2.23 for the violin). Compared with Figure 2.41, Figure 2.40 shows the dip in the spectrum found earlier for transitions with bow change. The effect is not as pronounced in the cello as in the violin; but it is still there. Thus, the power plots alone should not be taken as a measure of the parameters of a transition.

Still, the possibility remained that the size of the instrument might play a major role in shaping transitions. The situation was complicated for the brass instruments. The trombone, for which a few recordings had been made, was deemed to be unsuitable for comparison because of the difference between the slide and the valve mechanism. The closest analog to a cello in the brass family would be a valve trombone—but a valve trombone player could not be found. As for the woodwinds, recordings were made of the bassoon (for comparison with the oboe—both are double-reed instruments) and of three flutes: piccolo, bass flute, and regular flute. Examples of power and spectral plots for these instruments are given in this chapter and in the appendices. Examination of these plots (and those for the other intervals, not reproduced here) leads to the conclusion that in the woodwind family, the size of the instrument does not affect the evolution in power and overall spectrum at the transition; in the strings, the power plots can be obscured in some cases with larger instruments, but the overall spectral pattern follows that of the woodwinds; and the brass are assumed to function in the same way as the woodwinds.

What, then, is a transition? (II)

For the purposes of this study, a transition is a region of change between two notes performed on an orchestral instrument. This is another way of stating the definition given in Chapter 1 (see also Chapter 7). Compared with a transition, the surrounding notes themselves are relatively stable. Reducing the long list of possible parameters given at the beginning of this chapter, a transition can be coarsely characterized as shown in Figure 2.42 by the following "parameters":

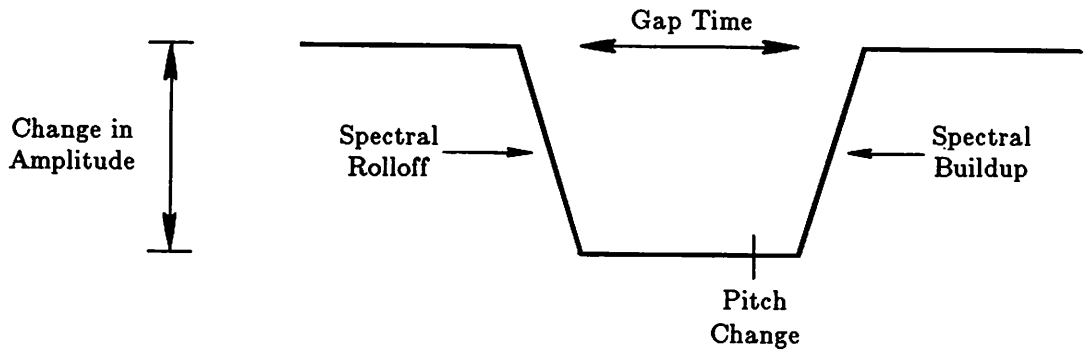


Figure 2.42. The parameters of a transition.

1. There is some change in amplitude: the first note dies out, after which the second note starts up.
2. As the first note dies out, its spectrum "falls off" (that is, the higher-order spectral components disappear); as the second note enters, its spectrum is enrichened. There may be spectral cues in the attack of the second note which depend on a given playing style.
3. There is a change in pitch which falls closer to the attack of the second note than to the decay of the first.
4. There is some amount of time between the two notes; the decay of the first note does not coincide with the attack of the second.

To summarize the results of this chapter, there is furthermore a regular, significant difference in time, power, and spectral rolloff between the tongued and nontongued cases across all the instruments. In the tongued case, the notes are generally farther apart, the amplitude dip in the transition is lower, and the spectral changes are more extensive, than in the nontongued case. On the other hand, there is no systematic difference between ascending and descending intervals for a given instrument, nor for intervals of varying sizes. The player can easily replicate a given articulation. Only in the string instruments does the size of the instrument show any difference in power analyses.

Table 2.2. Summary of representative transitions.

Transition		Maximum peak amplitude of steady-state		Minimum peak amplitude in transition
		First note	Second note	
Clarinet	Tongued	−0.76	0	−36.8
	Untongued	0	−0.14	−13.0
Trumpet	Tongued	0	−0.98	−38.7
	Untongued	0	−2.61	−20.7
Violin	Bow Change	−0.98	0	−30.4
	No Bow Change	−2.82	0	−18.7

Note: all values in dB, relative to linear amplitude of 0.75 (all two-note pairs were scaled to have their maxima at that value).

Choosing Representative Cases

The set of 212 recordings was much too large to permit design of rigorous experiments. It was thus necessary to isolate a set of representative transitions, not too many and not too few, which could be manipulated and studied further.

Clearly more than one instrument should be represented—at least one of the strings, brass, and woodwind. For the strings, the violin presented fewer problems, as its power plots were clearer. The trumpet was the only brass instrument seriously considered here. For the woodwinds, the clarinet was selected.

If size and direction of the interval performed do not play a major role, then only one interval in one direction need be chosen. That interval should be represented in the recordings for all of the instruments chosen. This ruled out the ascending major second, for which the clarinet recordings had been lost. The ascending major third was then chosen.

Obviously both tongued and untongued cases must be included, given that significant differences between them were found.

Some of the parameters of the resulting six transitions (two playing styles times three instruments) are summarized in Table 2.2. (Information about timing will be given in Table 5.4). These are the six transitions which have been analyzed for amplitude (Figures 2.3–2.6), power (Figures 2.9–2.14), and spectral changes (Figures 2.18–2.23). These six transitions form the data base for most of the experiments discussed in the rest of this document.

CHAPTER 3

MODELING TIME-VARYING SPECTRAL CUES

An important step which must be taken in methodological development is the analysis and synthesis of connected notes. This necessarily will precede carefully controlled work on the perception of timbre in musical contexts. Recall that the analysis techniques employed to date for timbre have been inherently designed for and applied to isolated notes. There exists no analytic scheme which is directly applicable to musical phrases. (Grey 1975, p. 109)

Experiment 1: Phase Vocoder Analysis/Resynthesis

Background

Chapter 2 discussed possible shortcomings of the phase vocoder in analyzing transitions between notes. In particular, the frequencies in a transition change too quickly for the phase vocoder to be able to track the spectral components accurately. This experiment is designed to show that these shortcomings are of such small magnitude as to be imperceptible, and that the phase vocoder is in fact an adequate technique for performing time-varying spectral analysis of such transitions.

If the phase vocoder were inadequate, one would expect the same problems to occur regardless of the instrument being analyzed. Therefore it is sufficient to experiment with just one instrument; the trumpet was arbitrarily chosen here. Also, any inadequacies of the phase vocoder should be more apparent with a larger interval between the notes played; therefore, the largest interval available (m7) was chosen. Finally, there should be no differences between ascending and descending intervals in observed behavior of the phase vocoder nor in the audibility of any distortion produced by the phase vocoder. In short, the ascending major seventh played on the trumpet formed the basis for this experiment.

At least two possible sources of distortion in the analysis can be identified:

1. As mentioned before, the phase vocoder might not be able to track signals adequately in the transition region. One would expect this to be most prominent in the untongued case, since the frequencies are shifting so rapidly (sometimes across just a few periods).
2. The phase vocoder models the signal as a group of harmonically related sinusoids. It might not be able to emulate the "puff" of noise at the beginning of a tongued attack.

Preliminary work demonstrated that neither of these produced audible distortion in the transition. Thus, this experiment will use only the untongued case.

Creating the Stimuli

The recording of the trumpet untongued ascending m7 was resampled to 26040 Hz, to provide a sampling rate into which the fundamentals of both notes would divide easily. Both notes in the stimulus were analyzed using two sets of phase vocoder parameters. The settings appropriate for the lower note were $N = 100$ and $R = 5$, with $N = 56$ and $R = 8$ for the upper note. As noted before, R was kept smaller than $N/2$.*

The note pair was resynthesized twice, using both analyses. Both resynthesized note pairs sounded completely natural, although slightly low-passed when compared with the original. (This problem will be discussed in more detail for Experiment 2).

To make a test stimulus, it proved impossible to splice the analysis data from the first note directly onto the analysis data from the second note, a procedure implied by the figures of Chapter 2 and Appendix 5. After resynthesis using this method, the second note still sounded quite natural, but underwent some severe phase distortion which made it unsuitable for use in an A/B experiment. The phase distortion apparently occurred because of the abrupt change in analysis parameters.

Thus, the final test stimulus was created with a 20-msec cross-fade from the end of the first note as analyzed with $N = 100$ to the beginning of the second note as analyzed with $N = 56$.

---

*Q, mentioned in (Gordon and Strawn 1985) is set to 1, effectively turning off any further interpolation of the data points as suggested in (Moorer 1978). Some modification to the code given in (Gordon and Strawn 1985) is necessary to make this work right.

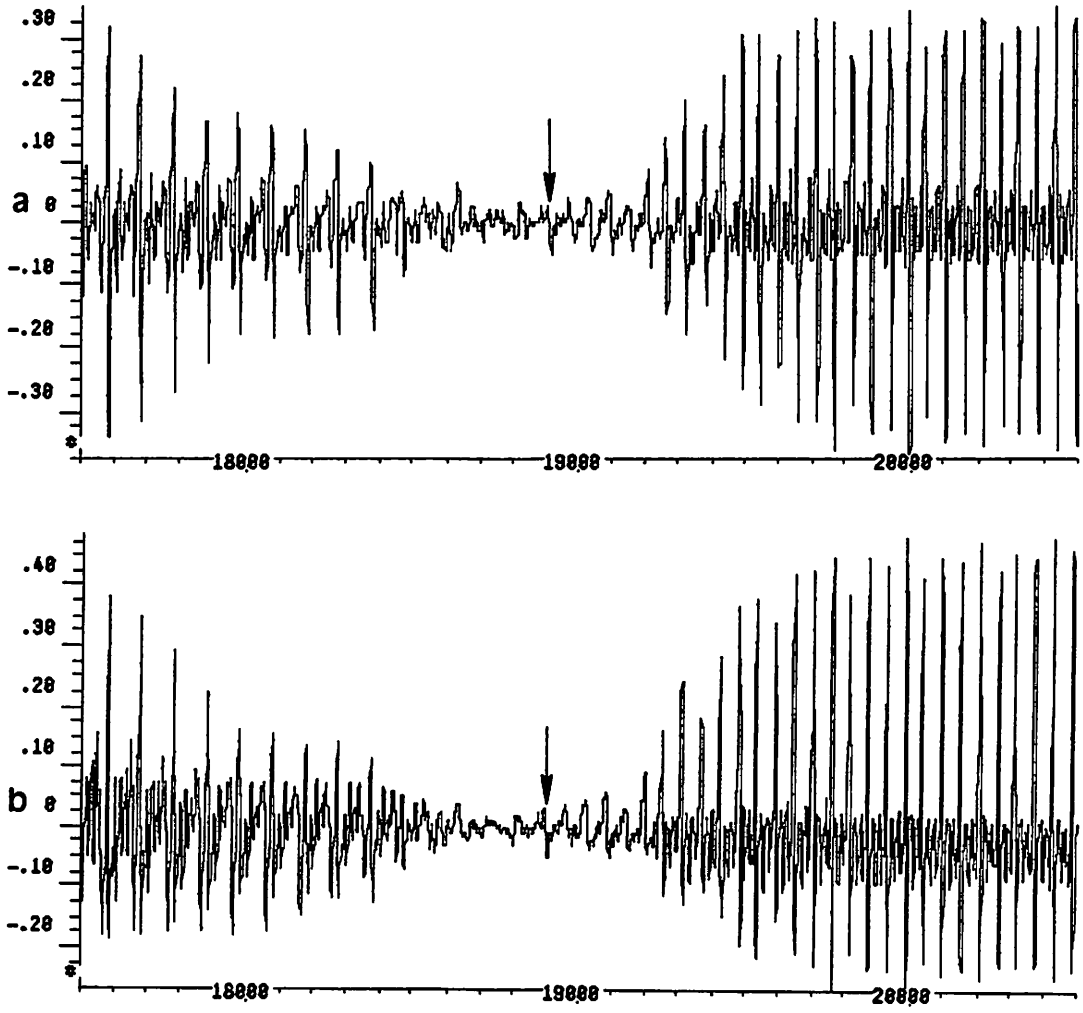


Figure 3.1. Stimuli for Experiment 1. a) Untongued trumpet transition. b) Each note has been analyzed and resynthesized separately. At the point marked by the arrow, a 20-msec cross-fade joins the resynthesized notes. (The x-axis shows time in samples, at a sample rate of 25 600.)

The cross-fade occurred at the point shown in Figure 3.1. The resulting transition is shown in Figure 3.1b. (This procedure also worked for the tongued case.)

The control stimulus was the original recording, shown in Figure 3.1a.

Experimental Procedure

Experience has shown that a note resynthesized on the basis of phase vocoder analysis is physically slightly different from the original (this will be discussed more under Experiment 2). It is not necessary here to show that the original and resynthesized tones are physically identical. However, it *is* necessary to show that the listener cannot reliably distinguish between the two. As stated above, the test stimulus sounded slightly low-passed when compared with the control stimulus. (Attempted solutions to this problem will also be discussed under Experiment 2). Thus, it proved impractical to conduct an experiment in which the subject decides whether two stimuli are identical, because the notes surrounding the transitions proper are themselves different in the two stimuli. Therefore, each subject was asked to state a *preference* for one of the two stimuli. If the subject has no clear preference for one of two quite similar stimuli, then we can conclude that the two are perceptually interchangeable. Indeed, they might even be identical for all practical purposes.

In such preference tests, it is important to account for any order effects. That is, if A and B are different stimuli in a given case, then the preference for A followed by B must be compared to the preference for B followed by A. Also, comparing each stimulus with itself (A:A and B:B) checks for subject bias toward the first or the second stimulus of each pair; such tests are sometimes called *Vexierversuchen*. The subjects in this experiment thus heard four cases, numbered 1–4 in the list below. Each case consisted of one stimulus, followed by a short pause, followed by another stimulus:

Comparison cases

1. Original (control stimulus) vs. resynthesized (test stimulus)
2. Resynthesized vs. original

Identical cases

3. Original vs. original
4. Resynthesized vs. resynthesized

The comparison cases (1 and 2) were presented 3 times each; the identical cases (3 and 4) were presented twice each. (Details on the presentation of the stimuli are given in Appendix 3).

Table 3.1. Average Preference for First or Second Two-note Pair in Experiment 1.

	Subject Number									
Case	1	2	3	4	5	6	7	8	9	10
Comparison cases										
1	1.67	2.00	1.33	1.67	2.00	1.67	1.67	1.67	1.00	1.67
2	2.00	1.67	1.00	2.00	2.00	1.00	1.67	1.33	1.33	1.00
Identical cases										
3	2.00	2.00	1.50	1.00	2.00	1.50	2.00	1.50	1.00	1.00
4	2.00	1.50	1.50	1.00	2.00	2.00	1.00	1.50	1.00	1.50

Note: Values may range from 1.0 to 2.0. A value of 1.0 in cases 1 and 2 means that the original is preferred over the synthetic; in cases 3 and 4, that the first of two identical stimuli is preferred.

Results

Table 3.1 shows the responses for all 10 subjects (information on the subjects is given in Appendix 3). Each entry in the table shows the average preference (across three presentations per subject for the comparison cases, across two presentations for the identical cases). The only intermediate values possible were thus 1.33 and 1.67... for the comparison cases, and 1.50 for the identical. Some order effects were apparent with the identical cases on a subject-by-subject basis. For example, subject 1 always selected the second of two identical stimuli, whereas subject 9 always chose the first. However, it was impossible to find any meaningful overall pattern in the identical cases.

For the comparison cases, examination of the data showed that subjects 3 and 9 preferred the original over the resynthesis; subjects 1, 2, 4, 5, and 7 preferred the resynthesis over the original (!); and subjects 6, 8, and 10 seemed to prefer neither. Although it might appear surprising that *any* subject would prefer the synthetic stimulus (if this preference were not due to random variation in the responses), this might occur because these particular test subjects were used to working with synthesized sound. At any rate, no consistent pattern appeared in the data for the comparison cases either.

This preliminary conclusion was borne out in Table 3.2, which shows the means of all of the responses for each case averaged across all subjects. It is reasonable to conclude that the subjects could not accurately distinguish between the two stimuli when the mean is close to (case 3) or at (cases 2, 4) the value of 1.5.

Table 3.2. Preferences for First or Second Stimulus in Experiment 1, Averaged across all Subjects.

Case	Mean	t
Comparison cases		
1	1.63	1.49
2	1.50	0.00
Identical cases		
3	1.55	0.44
4	1.50	0.00

Note: A value of 1.0 in cases 1 and 2 means that the original is preferred over the synthetic; in cases 3 and 4, that the first of two identical stimuli is preferred.

The value of case 1 might seem far from the expected mean. It is more meaningful to examine the combined mean of the comparison cases (1 and 2) across all subjects, which is 1.57. This value suggests that the subjects could not accurately distinguish between the original and the resynthesized stimulus; if anything, subjects showed a slight preference for the *resynthesized* stimulus.

Whether any of the means of Table 3.2 indicates an actual preference for the original or simply chance variation around the "no preference" mean of 1.5 is a question answered by the well-known t test. The results are also shown in Table 3.2. The t values indicate that none of these means was statistically different from what one might expect from a population of subjects with no preference for one signal over the other.

The question then arose as to whether the means for the four cases were significantly different from each other. If not, then it can be asserted that the seemingly large mean for case 1 is no more significant than the other, smaller means. Table 3.3 gives the analysis of variance for the data in Table 3.1, following the tabular organization given by Hays (1963, p. 372). The original data (not given here) were used to calculate the values of the sums of squares given in the SS column. The df column shows the "degrees of freedom" which, in the first line, is one less than the number of cases, and in the second line is the number of responses across all subjects and cases (100) minus the number of cases. The MS columns shows the values of the mean squares, each formed by dividing the SS value by df in the same line. The MS values in Table 3.3 are quite small; this is due to the fact that some subjects consistently favored the first or the second pair within a case, as was mentioned above. The F value is found by dividing the upper MS value by the lower. Any F value close to 1 suggests that there is no significant difference among the means of Table 3.2.

Table 3.3. Analysis of variance for Experiment 1.

Source	*SS*	*df*	*MS*	*F*
Cases	0.33	3	0.11	0.44
Error	24.42	96	0.25	
Totals	24.75	99		

Indeed, standard statistical tables shows that $F = 0.44$ with 3 vs. 96 degrees of freedom implies $p > 20\%$; that is, the probability that the given data would occur due only to chance is greater than 20%. In other words, there is no statistically significant difference among the means of Table 3.2, QED. Thus, none of the means varies significantly from the value of 1.5, and I conclude that the subjects showed no preference for either the original or synthetic stimulus, also QED.

Indeed, such a small F value suggests that the variation in the data is *less* than one might expect from pure chance. However, the value $1/F(= 2.29)$ is still less than the 5% F value for 3 vs. 96 degrees of freedom (= 8.55), so no significance is attached to the smallness of F. Indeed, this small F value can be attributed to the tendency of certain subjects to pick the first or second transition within a case.

All of this statistical sophistication may appear to be overkill when one reads the written comments of the subjects (see Appendix 3), of which these are typical:

> Subject 3: "They all sounded rather similar."
>
> Subject 5: "I was not able to hear any differences in any of these pairs (nor between one pair and another)."

Conclusion

The subjects showed no clear preference for either the original nor the resynthesized transition. The transition resynthesized on the basis of full phase-vocoder data is therefore perceptually interchangeable with the transition in the original.

Experiment 2: Line-segment Approximation

> *It remains possible that many of the properties necessary for the* simulation *of connected passages will be amenable to simplification. This will greatly reduce the process of stimulus specification in the physical domain, and lend greater control to the synthesis method.* (Grey 1975, p. 110)

Background

Phase vocoder analysis provides too much data for practical work in sound synthesis, and for controlled timbral studies. It is commonly accepted that line-segment approximation of the amplitude and frequency traces can produce individual resynthesized tones which sound quite close to the original (Risset 1966; Risset and Mathews 1969; Grey 1975; Moorer 1977; Moorer, Grey, and Strawn 1977, 1978; Charbonneau 1981). Experiment 1 showed that the phase vocoder adequately represents the time-varying spectrum in the transition. The question remains as to whether line-segment approximations are likewise adequate for synthesizing musical transitions.

Methods for creating reasonable line-segment approximations remained primitive (Risset 1966; Beauchamp 1969; Grey 1975) when this work started. (As Risset wrote (p. 29), "simpler methods have to be found for use in computer music.") A search of the literature on approximation theory and pattern recognition showed that several algorithms can be useful (Strawn 1980). That work also resulted in a data structure and paradigm for syntactic, hierarchical analysis. This data structure has the advantage that it allows for control *across* the entire spectrum, so that one can add or delete features in all of the harmonics with one operation.

Creating the Stimuli

For the current work, the Split-Merge Algorithm combined with the Adjust procedure, both due to Pavlidis (details and references given in Strawn 1980), were used. The algorithm was applied to amplitude data in the following way:

1. From the phase vocoder analysis, create a "spectral average" by averaging the phase vocoder data over a specified amount of time. A sample of such a spectral average is given in Table 3.4 (another example is given in [Strawn 1985a]). For amplitudes, this is equivalent to taking the Fourier transform over the time in question, which is selected from the "steady-state" of each note.

Table 3.4. Spectral Average of Steady-State of Violin Tone

Channel	Amplitude	(dB)	Freq. (Hz)	$freq_n/(freq_1 \times n)$
1	0.5392	−25.69	221.0782	1.00000
2	10.3786	0.00	442.0194	0.99969
3	1.7760	−15.33	663.0221	0.99968
4	1.5451	−16.54	884.0308	0.99968
5	1.1551	−19.07	1104.9779	0.99963
6	0.3557	−29.30	1326.0262	0.99967
7	0.5108	−26.16	1547.1213	0.99972
8	0.1608	−36.20	1767.9410	0.99961
9	0.4111	−28.04	1989.4891	0.99989
10	1.1457	−19.14	2210.2055	0.99974
11	0.4452	−27.35	2431.2068	0.99973
12	0.2809	−31.35	2652.6103	0.99988
13	0.1195	−38.78	2872.7697	0.99957
14	0.1521	−36.68	3093.1538	0.99937
15	0.1664	−35.90	3314.8350	0.99960
16	0.1051	−39.89	3535.6710	0.99955
17	0.1492	−36.85	3756.7322	0.99958
18	0.2259	−33.24	3977.9933	0.99964
19	0.1522	−36.68	4199.1868	0.99969
20	0.0565	−45.28	4420.1436	0.99968
21	0.1154	−39.08	4639.9356	0.99942
22	0.1104	−39.47	4861.1602	0.99947
23	0.0277	−51.49	5080.3468	0.99912
24	0.0333	−49.87	5307.1811	1.00025
25	0.0437	−47.52	5527.4550	1.00009
26	0.0333	−49.86	5749.9391	1.00033
27	0.0143	−57.21	5966.8107	0.99961
28	0.0106	−59.85	6191.0150	1.00013
29	0.0113	−59.25	6408.6716	0.99960
30	0.0138	−57.50	6629.8061	0.99962
31	0.0082	−62.04	6852.8772	0.99992
32	0.0092	−61.01	7060.0781	0.99796
33	0.0077	−62.55	7286.8273	0.99880
34	0.0073	−63.06	7501.3884	0.99797
35	0.0096	−60.65	7738.4663	1.00009
36	0.0068	−63.71	7941.5504	0.99783

Note: This average spectrum was calculated over 0.1 sec. The frequency of channel n is $freq_n$; $freq_1$ is the frequency of channel 1 (the fundamental).

2. Multiply the averaged amplitude from each harmonic by some small constant, say 0.001. This constant varies with instrument, sample rate, N, and R.
3. Use the resulting number as a threshold for the Pavlidis algorithm, with the integral error norm given in (Strawn 1980).
4. The resulting line-segment approximation, typically a dozen segments per harmonic per note, must usually be cleaned up slightly by hand. An editor for this purpose has been written, which displays both the original and the approximation (Strawn 1985a).

Note that this process must be done twice to create a single test stimulus for this experiment—once for each of the two notes surrounding the transition.

These two sets of amplitude traces must then be joined by hand on a harmonic-by-harmonic and point-by-point basis. I have written an editor which displays the phase vocoder analyses for both notes along with a composite function created by splicing the line-segment approximations from the two notes at the point of pitch change; this is an extension of the editor described in (Strawn 1985a). For each harmonic, the user creates a final transition function by hand. Figure 3.2 shows the tenth harmonic taken from the two phase vocoder analyses of the tongued ascending third on the trumpet. In Figure 3.2a, the parameters of the phase vocoder were set for the frequency of the first note. The phase vocoder analysis is shown along with the raw output of the Pavlidis Split-Merge algorithm. Figure 3.2b shows a similar analysis, but with the phase vocoder set up for the second note (C♯). This part of the figure must be carefully interpreted; the "beating" at the left of the figure results when two harmonics of the first note fall into one analysis band of the phase vocoder. Incidentally, the editor allows the user to view either or both phase vocoder analyses along with either or both of the original approximations as well as the approximation which the user creates by hand (Figure 3.2c; this is the actual function used in synthesizing the tenth harmonic for the tongued trumpet stimulus). Editing in this manner is not as easy as it might sound. Once the software works, several minutes of console time are needed for each harmonic. For a stimulus with 30 or so harmonics, an hour can be quickly consumed.

The result of this editing is a set of line-segment approximations which more or less accurately capture the amplitude characteristics of the harmonics in the transition. Figure 3.3 shows the approximations which were created for the clarinet transition originally shown in Figure 2.18.

Risset (1966, p. 36, p. A–9) was not able to show that including either the "blips" in the trumpet attack nor the slight burst of noise at the beginning of the note had any effect in his resyntheses. My experience is that both of these features make an important difference in how

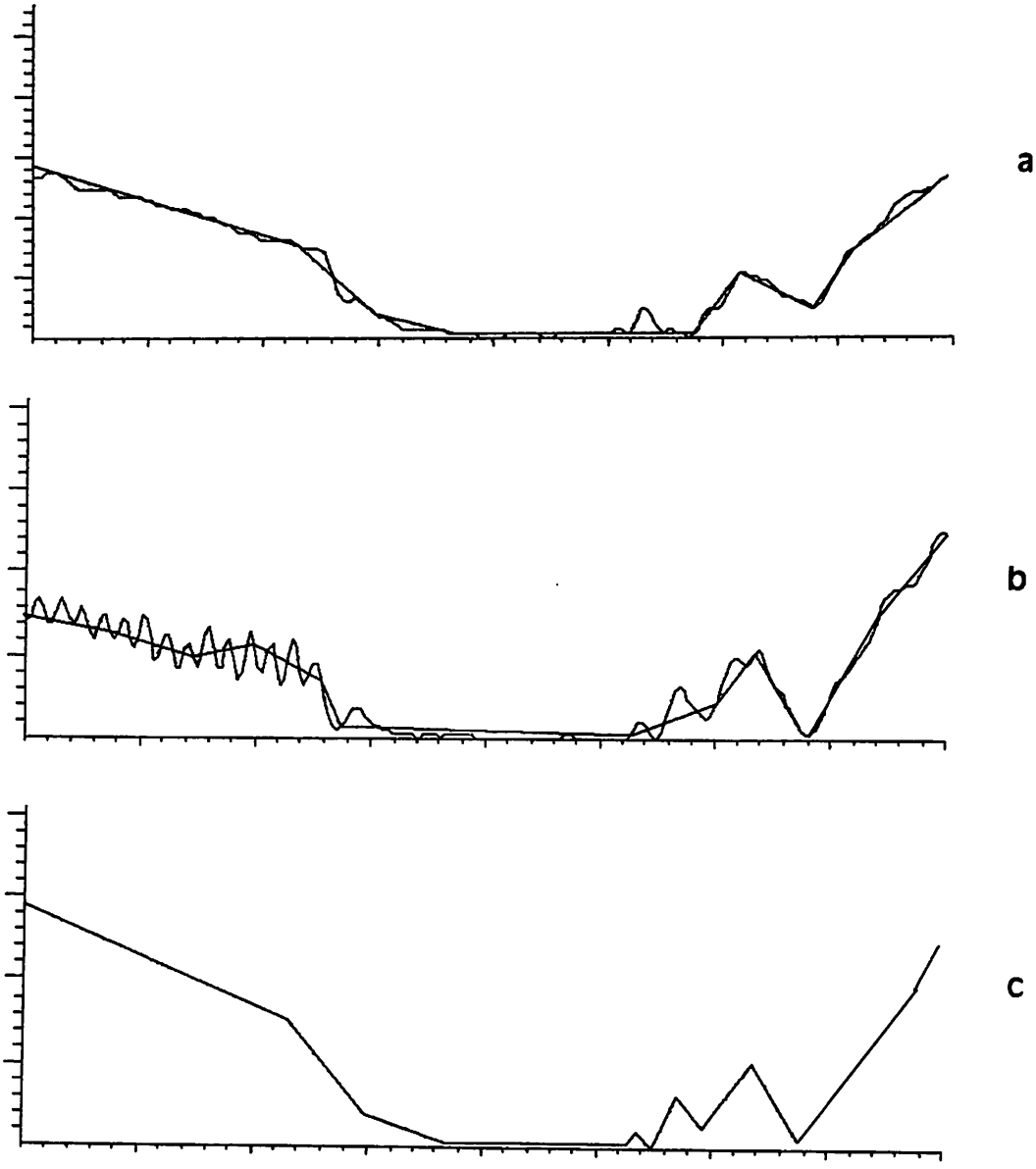


Figure 3.2. Editing the amplitude traces in a transition for the tenth harmonic of the ascending third tongued trumpet transition. a) Phase vocoder analysis parameters set for the first note, with line-segment approximation created with the Pavlidis Split-Merge algorithm. b) As in a) but with phase vocoder parameters set for the second note. c) The line-segment approximation created by hand from the approximations in a) and b).

the test stimulus sounds. Much of the time spent in refining the trumpet test stimuli for this experiment was in fine-tuning the blips in the attacks of the first dozen partials or so, and in adding small amounts of amplitude to the higher harmonics right at the attack, to simulate the tonguing noise.

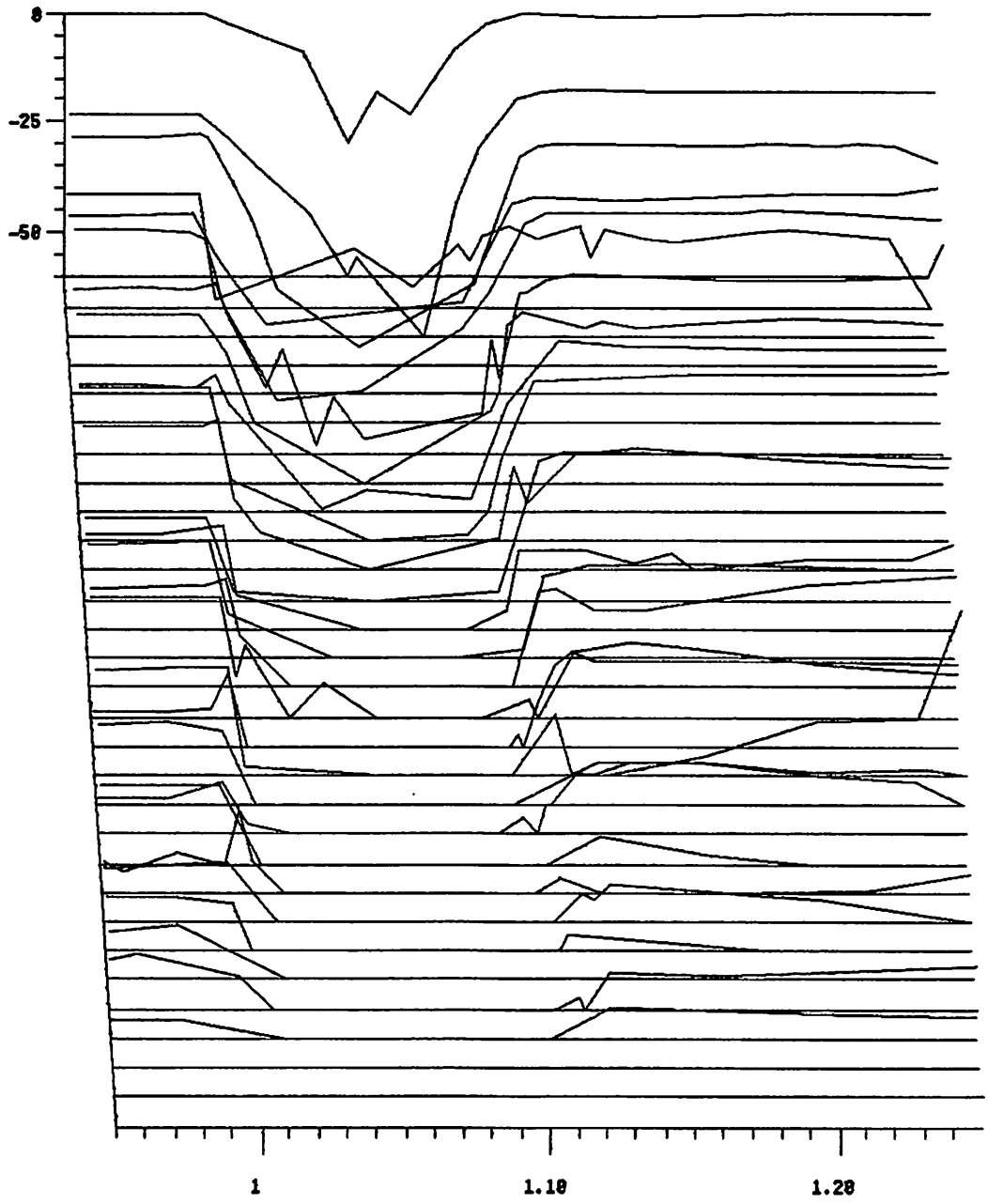


Figure 3.3. Line-segment approximations for the clarinet tongued ascending third transition in Experiment 2. Cf. Figure 2.18.

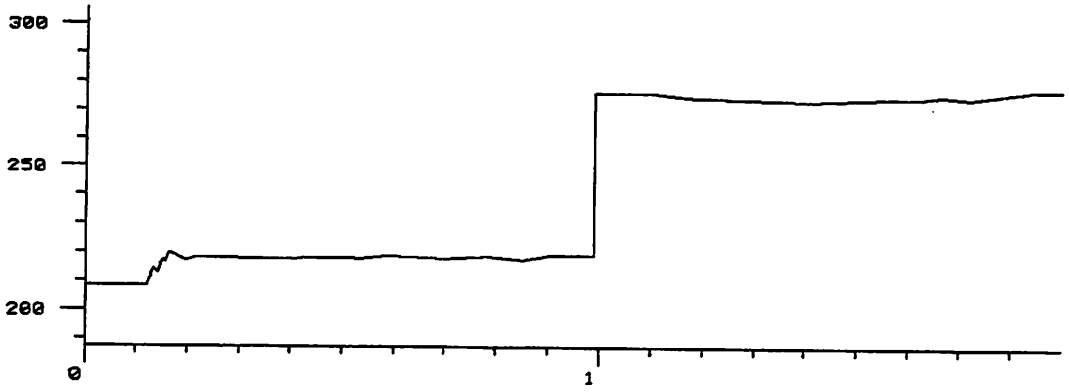


Figure 3.4. The fundamental frequency trace for the two-note untongued trumpet test stimulus from Experiment 2. The *y*-axis is frequency in Hertz; time (sec) is the *x*-axis.

For frequency traces, I found that it was adequate to create one line-segment approximation from the fundamental of each note, using the editor just mentioned. The spectral average of Table 3.4 also contains values (in the right-hand column) for what I term the "relative harmonicity" of the spectral component—how far it deviates from being an exact multiple of the fundamental. For each harmonic, each point of the hand-made fundamental frequency trace is multiplied by the harmonic number times this relative harmonicity value. This is slightly different from the work by Grey (1975), who used a constant-frequency approximation for some experiments, and from Charbonneau's tones (1981), where the fundamental frequency trace was multiplied by the (integer) harmonic number. In some cases, a slightly richer tone results by using the inharmonic case. In particular, the straight-line frequency approximation of Grey is noticeably enriched.

Again, this process must be followed for both notes in the test stimulus. The frequency traces for the two notes are simply spliced, using a vertical transition, at the appropriate point (Mathews and Miller [1982] suggested this step-function frequency transition independently.) Figure 3.4 shows the frequency function used for the fundamental of the untongued trumpet test stimulus for this experiment. Some activity in the attack of the first note is retained; its aural effect is not as pronounced as the illustration would suggest. I found that as long as the amplitude of the signal is low enough at the point of pitch change, the abrupt transition between the notes is never audible as such. One listener, not a test subject here (but an experienced woodwind player) claimed that he could detect the sudden frequency jump—but only after he knew the details of the synthesis process. Likewise, in discussing commercial synthesizers, Kaplan (1981) found that an abrupt frequency jump was audible. However, his remarks do not apply to the tones which I generated. For my tones, the amplitudes of each harmonic varied independently; in Kaplan's work, only the

Table 3.5. Bandwidth of stimuli used in Experiment 2.

Instrument	Cutoff Frequency (Hz)	Number of Harmonics			
		Tongued		Untongued	
		A	C♯	A	C♯
Violin	7000	32	25	40	32
Clarinet	7500	34	27	35	28
Trumpet	10500	40	38	40	38

Note: each stimulus consists of two pitches, A followed by C♯.

overall amplitude envelope was controlled. Kaplan also pointed out that the attack of the second note could be altered by the jump in frequency. In my work, the jump in frequency occurred *right at the beginning* of the attack, leaving the rest of the attack of the second note unharmed. At any rate, none of the test subjects in this experiment complained about the quality of the transition synthesized in this manner.

As with most of the experiments in this work, the tongued and untongued ascending thirds from the clarinet, trumpet, and violin were used. For each test stimulus, a two-note pair was created using additive synthesis of the amplitude and frequency functions just described. The control stimuli were the corresponding six original recordings.

The line-segment approximations included harmonics whose amplitudes were above approximately −60 dB from the note's maximum, as shown in Table 3.5. It was impractical to include harmonics with amplitudes much lower than this, the amplitude and frequency traces being badly degraded by noise. Also, the overall signal-to-noise ratio of the originals was about 60 dB.

The transitions in the synthesized stimuli sounded very close to those in the original recordings. However, as in Experiment 1, the notes in the test stimuli were and sounded slightly band-limited. I spent considerable effort trying to solve this pesky problem:

1. I tried splicing from the original phase vocoder data to the line-segment approximation at the very end of the first note, then splicing back to the original phase vocoder data (for the second note) at the very beginning of the second note. Resynthesis using CCRMA's Samson Box proved impractical because of the resulting high command rate. Also, even when the notes were resynthesized in software (prohibitively expensive for the amount of computation needed for this experiment), the problem mentioned in the discussion of Experiment 1 occurred here as well—there was a nasty phase shift where the splice occurs.

2. Using a very short cross-fade, I spliced the resynthized transition into the original recording, splicing at the end of the first note and again at the beginning of the second. For short cross-fade times (20 msec or so), a perceptible phase shift occurred at each splice. Due to the short duration of the transition, longer splice times proved impractical.
3. Following a suggestion by Portnoff (1983), I examined the difference signal between the original and the synthetic tones. This proved fruitless—the difference signal turned out to be a waveform almost identical to the original, except for a "phasing" throughout the duration of the note. Gish (1978) included an explicit noise term $n(t)$ in his synthesis model (his equation 1), and claimed: "When the residual error, or noise, $n(t)$, is listened to, it usually sounds just like tape hiss." Ideally, one would like to be able to characterize this noise signal. More work on the time-domain difference signal needs to be conducted.
4. I tried calculating the difference signal by subtracting, on a harmonic-by-harmonic basis, the amplitude and frequency traces of the line-segment approximation from those of the original analysis data.* These difference signals were used to synthesize a time-domain signal which was then added to the synthesized signal in order to make it sound closer to the original. (Beauchamp [1981] also developed a method for approximating the difference signal in this fashion; but he used only the error from the amplitude traces). The results were inconclusive. This approach needs to be explored further.
5. I tried, without success, to find a way to filter the original to match the quasi-low-passed nature of the synthesized tone. What one really needs here is a time-varying band-reject filter, because it turned out that the spectral differences could not be characterized by a time-invariant low-pass filter alone.
6. I tried to add low-amplitude white or colored noise to the synthesized signal to make it sound closer to the original.

Regarding item 6, Grey (1975, p. 37) also found that tape hiss present in the original recording but missing in the line-segment approximation could allow the listener to distinguish the two. Beauchamp (1981) likewise reported similar problems, in that in his resynthesized tones, key clicks and "a certain roughness" (p. 323) were missing. In Experiment 1, this was not a problem, as

*This is not recommended for those using slow computers.

resynthesis with full phase vocoder data captured all of the noise in the original. For Experiment 2, I had mixed success (as did Grey) with trying to add background noise from the original recordings into the synthetic stimuli. For the violin, I ultimately added white noise at −60 dB from the maximum of the two notes (noise limited to an absolute value of 0.001), which at least simulated a certain "scratchiness" missing from the line-segment approximation. The lack of noise was not so noticeable in the artificial stimuli for the clarinet and trumpet anyway; that is, adding in white noise did not help the "low-passed" sound of those two instruments. The obvious disadvantage of this method is that the amount of noise added is an experimental variable over which one has no systematic control.

The slight low-passed nature of the synthetic tones thus made it impossible to design a same/different experiment, or an experiment in which the subjects rate how different the resynthesized tone is from the original tone. The notes surrounding the transition in the test stimulus were themselves slightly different from those in the control, which might confuse the listener in such an experiment.

Experimental Procedure

Therefore, the preference test already discussed under Experiment 1 was used here as well. The subjects in this experiment heard four cases, numbered 1–4 as before. The comparison cases (1 and 2) were presented 3 times each; the identical cases (3 and 4) were presented twice each. (Details on the presentation of the stimuli are given in Appendix 3). These four cases were presented for each of the three instruments, using both playing styles (tongued and untongued) for each instrument. It seemed necessary to test more than one instrument, as any failing of the line-segment approximation might well show up for one instrument or playing method but not for another.

Results

A detailed table of the subjects' responses will not be given here. As in Experiment 1, examination of the raw data for the individual subjects' mean responses showed no clear preference for either the original or the resynthesized tones. This conclusion is supported by the subjects' written comments, of which these are typical:

Table 3.6. Subjects' Preferences in Experiment 2.

Case	Clarinet T	Clarinet U	Trumpet T	Trumpet U	Violin T	Violin U	Maximum possible
Comparison							
1	17	14	15	14	15	15	30
2	17	14	20	19	17	21	30
Identical							
3	7	10	9	12	7	5	20
4	11	6	9	8	8	8	20

Note: T = tongued (with bow change), U = untongued (without bow change).

Subject 2: "In a number of cases I heard no difference or at any rate had no preference..."

Subject 7: "Often hard!"

Subject 10: "Impossible!"

Table 3.6 shows *how often* the subjects preferred the original (cases 1 and 2) or the first stimulus (cases 3 and 4). The right-hand column gives the maximum score possible, derived from the number of subjects (10) times the number of presentations (3 for the comparison cases, 2 for the identical). This maximum score would be reached if all subjects preferred the synthetic stimulus in case 1, the original in case 2, or the second of the two identical simuli in cases 3 and 4. Here again no clear-cut pattern was discernible which might suggest whether the synthesized transition was preferred over the original. For the trumpet and violin in case 2, there is a slight tendency to pick the original over the synthesized tone; recall that in case 2, the original was played second. There seems to be no particular significance to this pattern in the data.

Table 3.7 gives the mean, standard deviation, and t value for each of the four cases, three instruments, and two playing methods. Only in three instances does the t value imply a probability less than 0.05, which means that for all instruments and playing styles except the violin with no bow change, it is safe to conclude that the observed mean does not vary from the expected mean of 1.5 any more than one would expect from random variation. The p value for case 2 on the tongued trumpet is not considered to be of significance, as case 1 for the tongued trumpet shows no deviation at all from the expected mean of 1.5. Therefore, I conclude that in five of the six instrument/playing method combinations the synthetic cases are essentially identical to the originals.

Table 3.7. Statistical Analysis of Experiment 2.

Case	Mean	s.d.	t	p	Mean	s.d.	t	p
Clarinet		Tongued				Untongued		
1	1.43	0.50	−0.72		1.53	0.50	0.36	
2	1.60	0.49	1.10		1.47	0.50	−0.36	
3	1.65	0.48	1.37		1.50	0.50	0.00	
4	1.45	0.50	−0.44		1.70	0.46	1.90	
Trumpet		Tongued				Untongued		
1	1.50	0.50	0.00		1.53	0.50	0.36	
2	1.67	0.47	1.90	<0.10	1.63	0.48	1.49	
3	1.55	0.50	0.44		1.40	0.49	−0.89	
4	1.55	0.50	0.44		1.60	0.49	0.89	
Violin		Bow Change				No Bow Change		
1	1.50	0.50	0.00		1.50	0.50	0.00	
2	1.60	0.49	1.10		1.70	0.46	2.35	<0.05
3	1.65	0.48	1.37		1.75	0.43	2.52	<0.02
4	1.60	0.49	0.89		1.60	0.49	0.89	

Note: A mean value of 1.0 in case 1 means that the original is preferred. The same value in case 2 means that the synthetic is preferred. In cases 3 and 4, this value means that the first of two identical stimuli is preferred. Values may range from 1.0 to 2.0. All t values imply $p > 0.10$ unless shown otherwise.

Analysis of variance of the other exception (violin, no bow change) is given in Table 3.8. (For an explanation of the entries in this table, see the discussion of Experiment 1.) The responses for case 2 were first "flipped" before this analysis of variance, so that Case 2 now represents the preference of the original over the synthetic, as does Case 1. Thus, there were only three cases considered for this analysis of variance. The F value of 4.25 in Table 3.8 implies that the variation of the means in Table 3.7 for the violin with no bow change was not just random ($p < 2.5\%$).

It is easy to accept the large amount of variation in Case 2, as Case 1 showed the expected behavior (i.e., the mean for Case 1 in Table 3.7 was 1.5), and both cases tested the preference of the original over the synthetic. Examination of the original data (not given here) for Case 3 showed that six of the ten subjects chose the second tone in both trials, which accounts for the large amount of variation seen there. Such a large bias did not occur in any other instance in this experiment. Thus the apparently large variation in the data for this one instrument and playing style is shown to have no real significance.

Table 3.8. Analysis of Variance for Experiment 2.

Source	*SS*	*df*	*MS*	*F*
Case	2.04	2	1.02	4.25
Error	22.95	97	0.24	
Totals	24.99	99		

Conclusion

The subjects showed no clear preference for either the original nor the resynthesized transition. The transition resynthesized using line-segment approximations to phase-vocoder data, with the frequency traces connected by a straight vertical line,* is perceptually interchangeable with the transition in the original.

Overall Conclusion

The model of time-varying spectra based on Fourier methods and implemented as the phase vocoder is adequate for analyzing and resynthesizing transitions between notes, using either the full analysis data or line-segment approximations.

*This should answer any questions raised in *Computer Music Journal* 8(2), p. 11, right column, second full paragraph, last sentence.

CHAPTER 4

OVERALL SPECTRAL AND AMPLITUDE CUES

Experiment 3: The overlapped transition

Background

From the initial data set of recordings of nine instruments, it was possible to make the general observations already given in Chapter 2: A transition between notes involves a dip in amplitude as well as certain spectral changes, all occuring across a given amount of time. This experiment examines whether the amplitude and spectral changes are *necessary* to create a usable transition. It also examines the amount of the time needed to change from one note to the next.

In general, to remove the amplitude and spectral changes, the end of the first note is overlapped with the beginning of the second note, with the amount of overlap time being the experimental variable. The change in pitch alone indicates the occurrence of the transition. Clearly, it is difficult to create a *tongued* transition without some sort of clue other than the pitch change. Thus, this experiment deals only with the untongued transition.*

Preliminary work showed that it made no sense to ask the subjects to judge if a given test transition were *acceptable,* as the range of acceptable transitions is wide indeed. Instead, the subjects judged whether the test stimuli represented acceptable *legato*. Examining "legato" transitions is merely a means to an end. If some transitions are shown to be acceptable legatos, then we can conclude that they are acceptable transitions. If none of the test transitions are acceptable legatos, more work may have to be done. The possibility remains that some test stimuli might

*Mathews and Miller (1982) attempted to create a *slurring* effect (which is presumably closest to the untongued case here) with synthetic tones by overlapping them. Their work will not be considered further in detail, as the experiment described here works with natural tones.

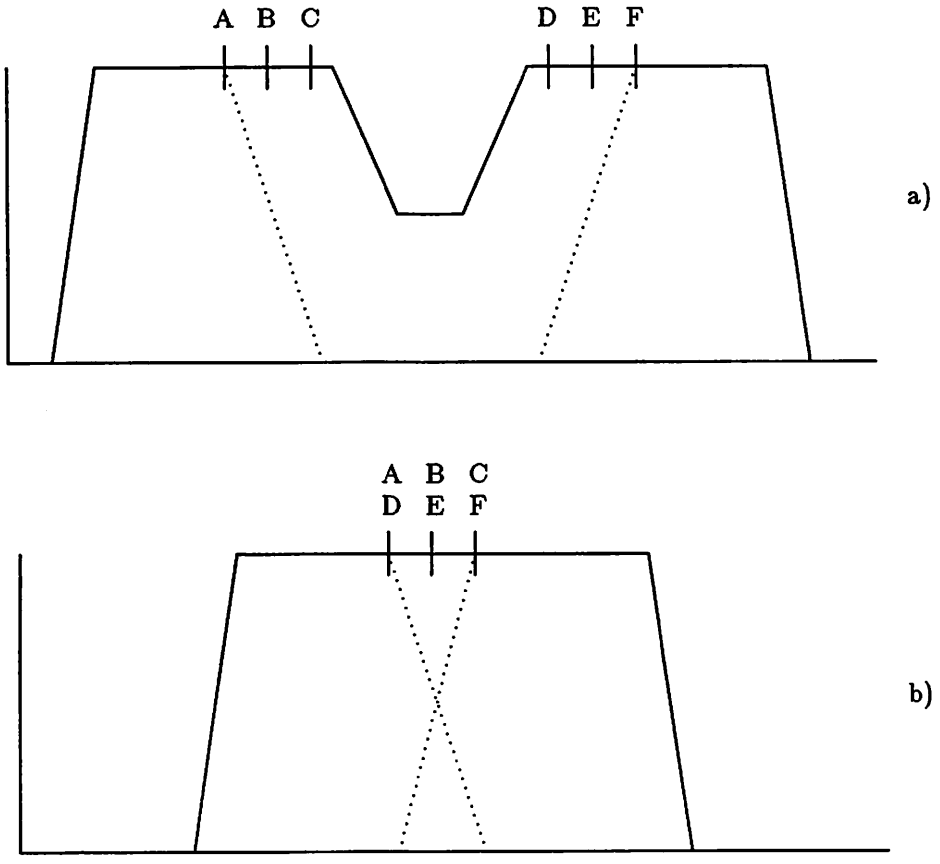


Figure 4.1. Experiment 3. a) Original recording of two notes. b) The end of the first note is overlapped with the beginning of the second note.

evoke a "tongued" (or portato, or detached) percept; however, none of the test stimuli sounded "tongued" to me, and none of the test subjects made comments to that effect in their written notes.

Creating the Stimuli

The ascending M3 untongued recordings from the clarinet, trumpet, and violin were examined to determine the boundaries of useful "steady-state" times in both notes. In general, the beginning of the steady-state of the second note was spliced onto the end of the steady-state of the first note. (These tones are thus similar to the "overlap" tones used by Mathews and Miller [1982].)

Figure 4.1 diagrams this process in detail. Point C is the end of the steady-state of the first note; the steady-state of the second note begins at point D. The durations AC and DF are equal to

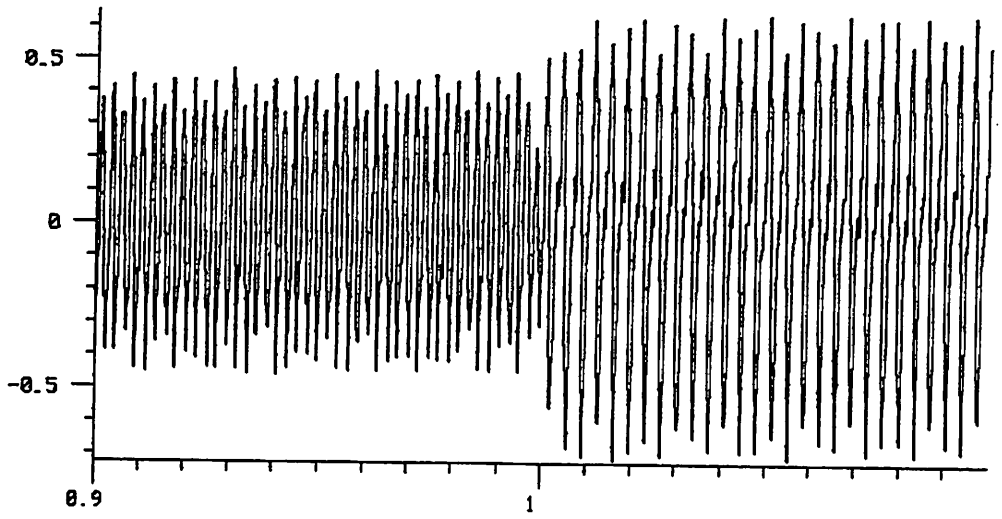


Figure 4.2. Two violin notes, at A220 and C♯ above middle C. The end of the first note has been overlapped with the beginning of the second note using the procedure given in the text. The transition time is 10 msec.

the desired cross-fade time. Point B lies half-way between points A and C; E is half-way between D and F. In order to avoid nasty phase jumps in the output, points B and E are corrected to the closest adjacent respective peaks in the waveform. The end of the first note (AC) is multiplied by a raised cosine wave (scaled to fit within the range [+1, 0], taking the cosine from 0 to π.) The beginning of the second note (DF) is multiplied by a corresponding sinusoidal fade-in function. (This method was developed by Loren Rush [1982].) The scaled waveforms are shown by dotted lines in the figure. Points B and E are aligned, and the scaled waveforms are summed. If the two notes are at the same amplitude, then there is no major change in amplitude during the overlapped transition. If the peaks of the waveform are aligned as described, then the phase shift at the transition remains unobtrusive. Figure 4.2 shows one example. The resulting two-tone pair is shorter than the original (see Figure 4.1b). To correct for this, the steady-states of the original recordings were extended using the methods of Appendix 2. In many cases, this whole process proved to be more difficult than it might sound; details are given in the appendix to this chapter.

The resulting test stimuli had a transition with a change in pitch but with no appreciable changes in amplitude or bandwidth such as one normally encounters. Six stimuli for each of three instruments, with cross-fade times of 10, 20, 40, 80, 160, and 320 msec, resulted in a total of eighteen test stimuli.

Table 4.1. Analysis of Experiment 3.

Transition	Clarinet		Trumpet		Violin	
Time	Mean	s.d.	Mean	s.d.	Mean	s.d.
10 msec	1.48	0.55	1.40	0.53	1.86	0.44
20	1.22	0.45	1.40	0.53	1.64	0.54
40	1.12	0.37	1.34	0.52	1.27	0.48
80	1.10	0.35	1.34	0.52	1.22	0.45
160	1.36	0.52	1.62	0.54	1.34	0.52
320	1.70	0.52	1.72	0.52	1.66	0.53

Note: 1.00 = acceptable, 2.00 = unacceptable

Experimental Procedure

At the beginning of this experiment, the subjects heard the original untongued recordings for each of the three instruments, played once each. The instructions stated: "You will first hear three examples, one on each instrument, illustrating a class of acceptable legato." For the actual trials, the subjects were asked to judge whether the test stimulus constituted an "acceptable legato." (Details on the presentation of the stimuli are given in Appendix 3).

Results

Table 4.1 gives the means and standard deviations for all three instruments and six transition times. (For the purposes of numerical analysis, the subjects' answers of "acceptable" and "unacceptable" were changed to 1.0 and 2.0, respectively). These data are shown in a graphic representation in Figure 4.3. Initial inspection of this data suggested that

1. it is indeed possible to create an acceptable legato transition (and therefore an acceptable transition, QED) while omitting all spectral and amplitude cues at the transition. This conclusion is justified because some of the transitions were rated as acceptable for each of the instruments.
2. some overlap times are more acceptable than others for this kind of overlapped transition. Extremely long or short crossfade times were rated unacceptable. These results agree with those given by Mathews and Miller (1982) for their overlapped tones.

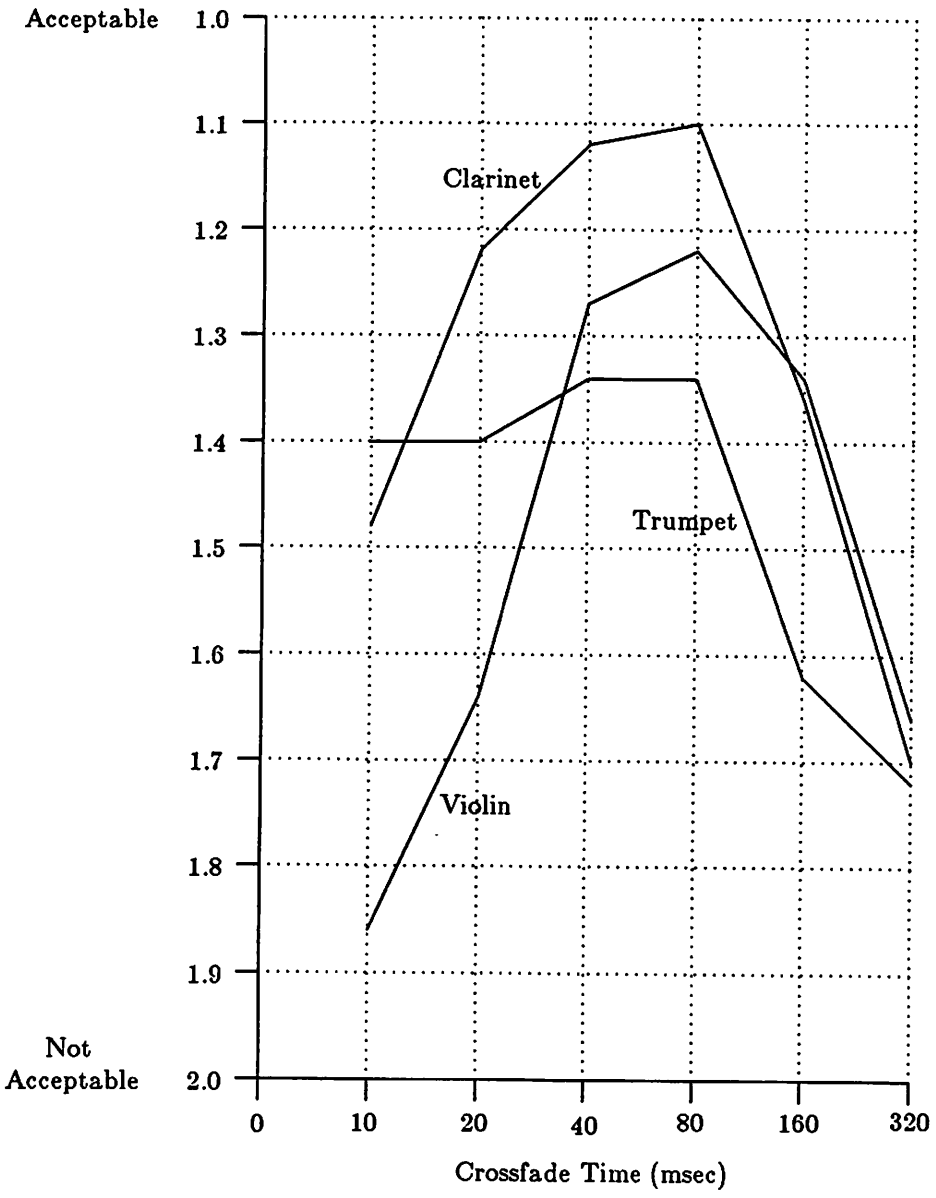


Figure 4.3. Results from Experiment 3, taken from Table 4.1.

Table 4.2. Analysis of Variance for Experiment 3.

Source	*SS*	*df*	*MS*	*F*
Instruments (I)	25.53	5	5.11	25.55
Crossfade times (C)	4.80	2	2.40	12.00
Instruments × Crossfade times	10.84	10	1.08	5.40
Error (I × C × S)	179.33	882	0.20	
Totals	220.50	899		

Initial audition of the test stimuli had led me to expect these results. Especially for the clarinet, with the very short transition in the untongued case, it seemed reasonable that a merely overlapped transition would be adequate. Also, I expected the extremely short and long times to be rated as unacceptable. The very long transition times sounded "muddy", almost as though there were an echo of the first note in the second.

The question then arose as to whether the curves in Figure 4.3 varied significantly from each other. Analysis of variance was again used to answer this question; but in this case, two-way analysis of variance was needed. The presentation of the results in Table 4.2 follows that given by Hays (1963, p. 402). From this analysis I draw the following conclusions:

1. The curves in Figure 4.3 vary from each other with more than random variation, since all three F values in the table are fairly large ($p < 0.1\%$ for all of them).
2. Since the F value in the "Crossfade times" row is so large, each curve is not "flat": that is, its variation from "unacceptable" to "acceptable" and back can be taken as statistically significant.
3. The F value in the "Instruments" row is fairly large. This means that the relative positions of the curves on the graph are significantly different. It would thus appear that the overlapped clarinet transitions are more likely to be judged acceptable than those for the violin or trumpet. This is consistent with the remarks of subject 7 (a violinist), who wrote: "[I] felt like I didn't *really* like any of the violin examples. Nor did I like many of the trumpet examples. The clarinet examples seemed best."
4. Along these lines, the value of 5.40 in the third line in Table 4.2 suggests that the shapes of the curves for each instrument are significantly different from each other. This means that as the crossfade time varies, the effect on the resulting transition varies from one instrument to the next.

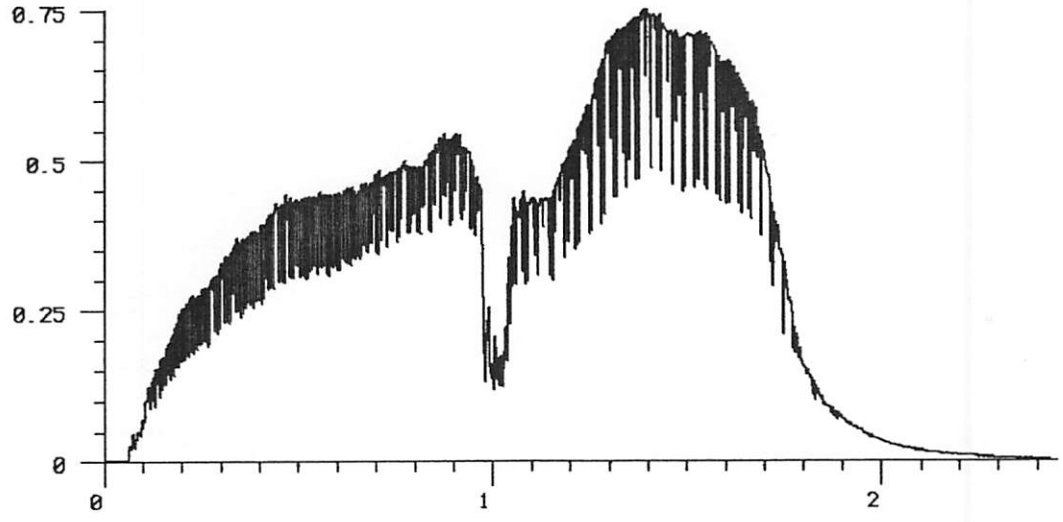


Figure 4.4. Original recording of the ascending M3 on the violin, with no bow change.

Conclusion

The spectral and amplitude changes observed in recordings of instruments are not always necessary for achieving an *acceptable* transition. In some cases, a simple overlap from one note to the next will suffice. The range of acceptable overlap times varies from instrument to instrument, as does the quality of the resulting transition.

Appendix: Creating the Test Tones for Experiment 3

Violin

There were several difficulties in creating test stimuli from the original violin recording, shown in Figure 4.4:

- The amplitude varies widely in each note, making it difficult to find a useable steady-state. The amplitude of the note should be more or less constant throughout the scaled transition region (AC or DE of Figure 4.1).
- The overall amplitudes of the two notes are different.

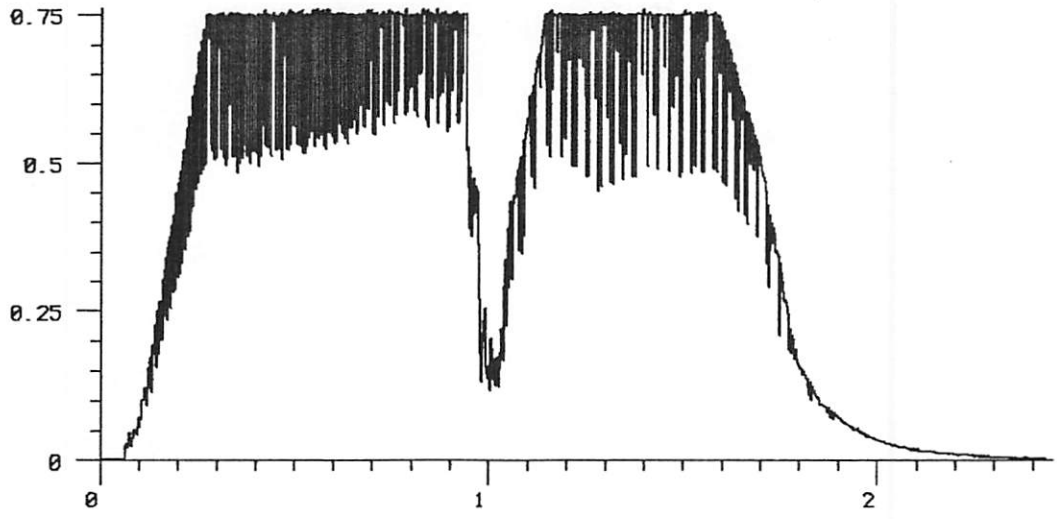


Figure 4.5. The recording of Figure 4.4, scaled in amplitude in an attempt to facilitate creating an overlapped transition.

- A "phase shift" occurs between the first and second notes. This can be seen in Figure 4.2; the relative positions of peaks and valleys shift considerably between the notes. This makes it more difficult to avoid a "phase jump" when the notes are overlapped, especially for longer transitions.

I first tried solving these problems by scaling the original recording to "force" both notes to have a steady-state, using the methods of Appendix 1. The scaled waveform is shown in Figure 4.5. The idea was to create the test stimuli using this scaled waveform, then scale the amplitudes of the attacks and decays of the test stimuli back to those of the original. This proved to be impractical.

Using the phase-vocoder-based method of Appendix 2, I extended the first note between 0.5 and 0.9 sec by a factor of 2, and extended the second note between 1.4 and 1.6 sec by a factor of 4. The resulting signals are shown in Figure 4.6. These time-extended signals were then used to create the test stimuli.

Clarinet

The major problem with the clarinet was in creating test stimuli with longer transition times, because the notes were short. I first attempted to extend the steady-states using methods 1–3 of Appendix 2; but, for the reasons given in that appendix, I could not make these methods work satisfactorily. After the phase-vocoder method was perfected, I used it to extend the steady-states

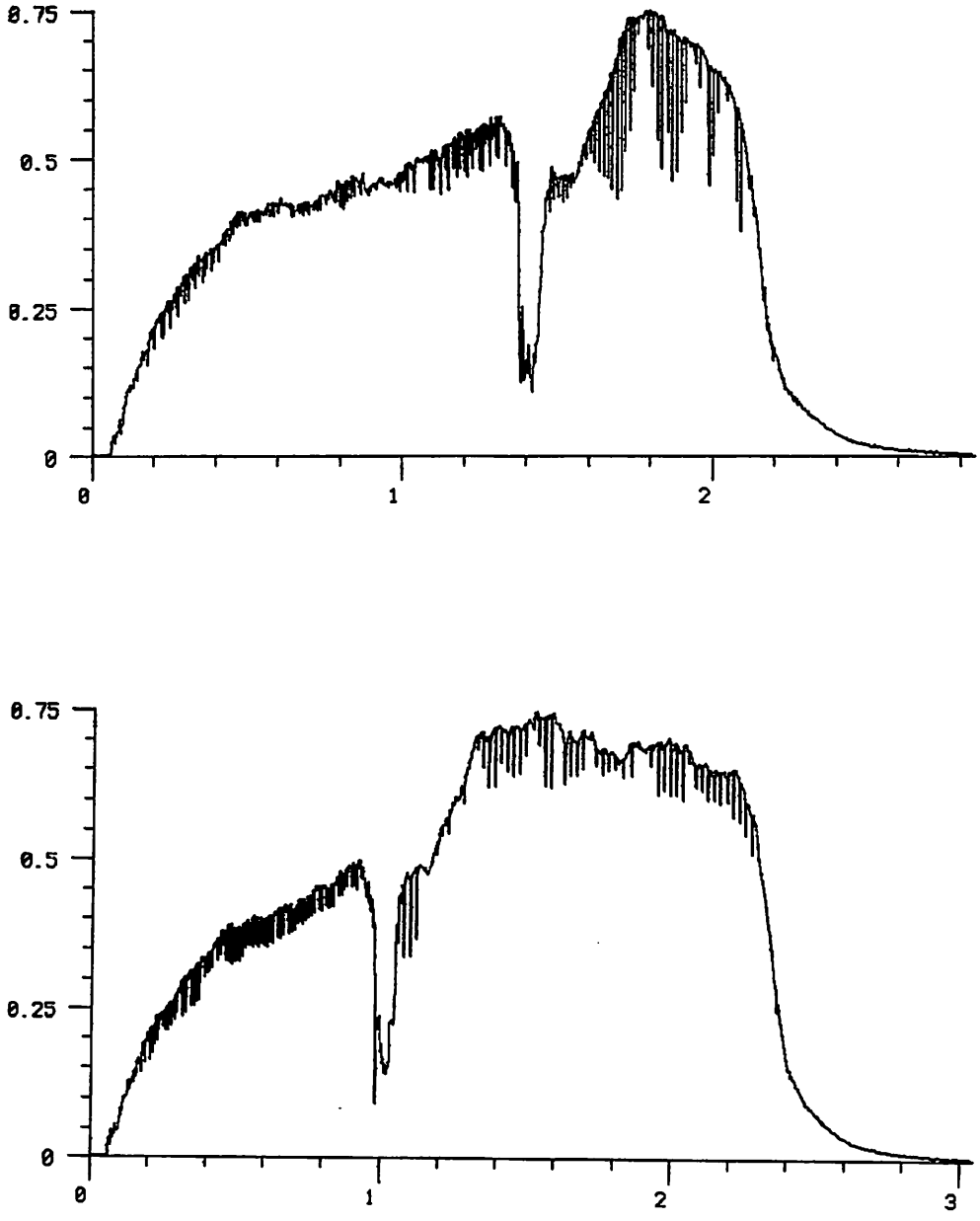


Figure 4.6. Extended violin tones. (top) The length of the first note (between 0.5 and 0.9 sec in Figure 4.4) has been doubled. (bottom) The length of the second note (between 1.4 and 1.6 sec in Figure 4.4) has been quadrupled.

of each of the tones, and proceeded as with the violin tones. Each test stimulus had the same duration as the original recording.

Trumpet

Here, too, it was necessary to first extend the steady-states of the notes in the original recording, using the phase-vocoder-based method of Appendix 2.

Neither the trumpet nor the clarinet presented the problems of widely varying amplitudes or of large phase shifts between notes, as were encountered with the violin.

CHAPTER 5

TIME-VARYING AMPLITUDE CUES

Introduction: Isolating the Components of a Transition

Experiment 3 in Chapter 4 showed that it is possible to create at least one sort of acceptable transition with only the pitch change to mark where the transition occurs. This model will clearly not work for all types of transitions. Of the four parameters of a transition (see Figure 2.42), it was easy to show (Chapter 2) that the *amount* of pitch change could be omitted from further study here. Considerable time and effort were spent in this study attempting to isolate the other three parameters in order to study them individually in a controlled manner. It proved possible to isolate amplitude, and this chapter is devoted to research into the role of time-varying amplitude in creating a transition. On the other hand, it proved impossible to find ways of varying spectrum or time in a rigorous manner without varying other parameters. This will become clear from the discussions in this chapter.

On the Relative Importance of Amplitude vs. Time

Background

In a study lasting several months, I attempted to create test stimuli (based largely on the clarinet and trumpet) which would answer the question of whether the gap time or the amplitude dip in the transition is perceptually more salient. (An interactive experiment to this end was ruled out from the start, due to the burden which such studies place upon the CCRMA system.) An answer

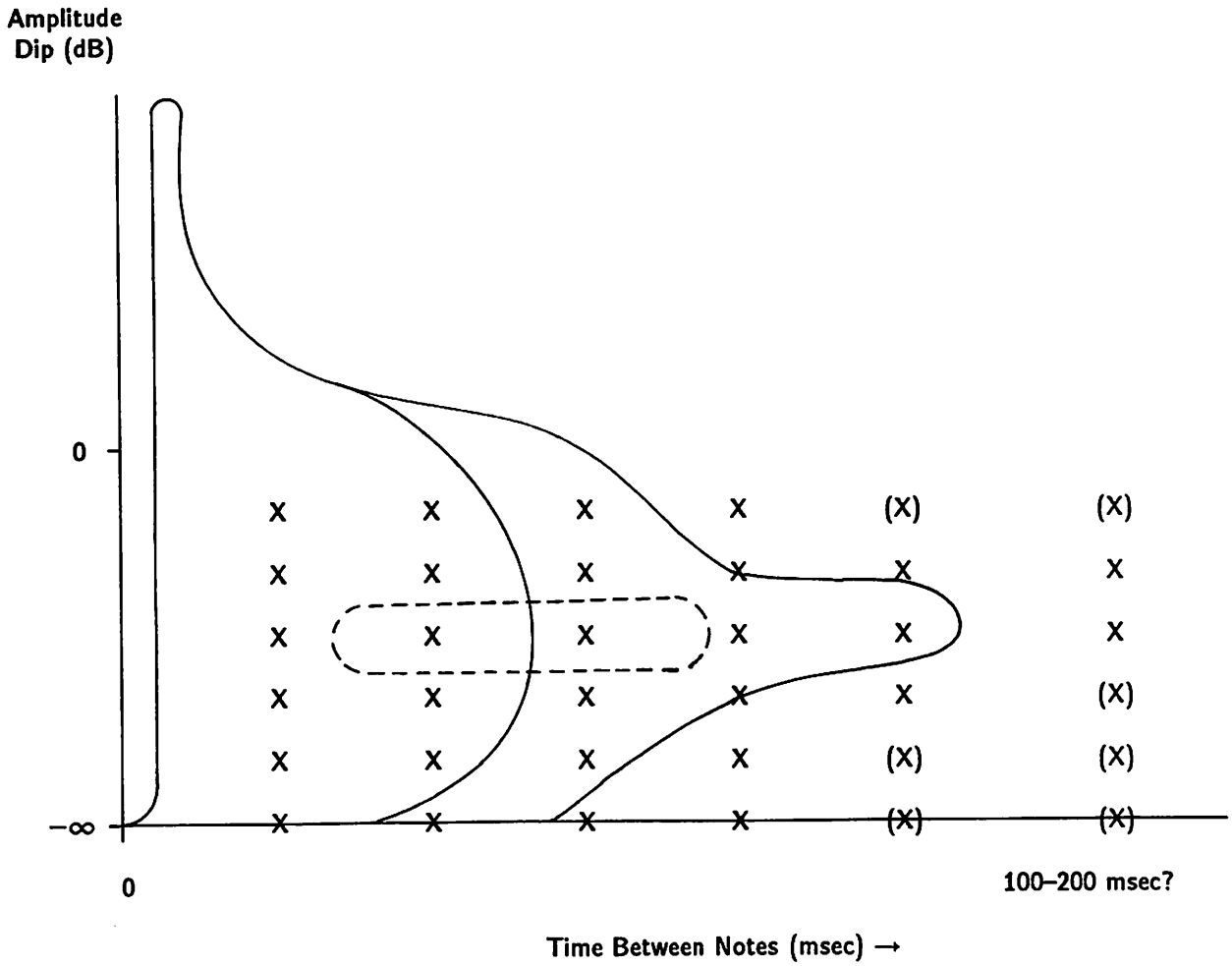


Figure 5.1. Hypothetical two-dimensional surface representing variations in time gap and amplitude dip in the transition between notes.

to this question might be of use to synthesists—it might suggest whether more attention must be paid to one or the other factor in creating musical lines. In a larger context, this inquiry might show what the the auditory system follows more closely at a "higher" level: amplitude or time.

It seemed reasonable to select points on a two-dimensional plane which would have gap time as one axis and amplitude dip as another (see Figure 5.1). Using the original tongued and untongued pairs as a starting point, the two notes could be pulled apart or moved together, and the amplitude of the dip could be scaled up or down. The dashed oval shows the expected region in which the original recordings (tongued and untongued) would fall. The solid line shows the hypothetical bounds of acceptable transitions—the right-hand line with the bulge for tongued

notes, the other line for untongued notes. X's mark points where test stimuli might reasonably be generated. (Stimuli at locations marked with X's in parentheses might not have to be generated, as preliminary studies would indicate the reasonable limits within which test stimuli should fall.) At each X, both the original tongued and untongued recordings would be used as a basis for creating a stimulus. With such test stimuli, it should be possible to design experiments to delineate the region of acceptable tongued and untongued transitions. Analysis of the data should also answer the question mentioned earlier of whether amplitude or time is more important. For example, there might be a wide range of amplitudes in which the transitions were acceptable, but only a narrow range of times.

Creating the Stimuli

It was comparatively simple to make stimuli with gap times shorter than those of the original recordings. To do so, some part of the transition was excised, and the gap was closed by simply abutting the end of the first note with the beginning of the second. Since the durations of the omitted signals were on the order of tens of milliseconds, and the amplitude in the transition was quite low anyway, the splice was usually not audible. It was of course necessary to splice right at the peak of a period (or some other reasonable juncture), in order to avoid gross phase discontinuities; this is discussed in more detail in Appendix 1. To extend a transition, Method 1 of Appendix 2 was adequate; that is, part of the transition was simply duplicated. Again, given the short times and the low amplitudes, the splice could not be heard as such, as long as care was taken to match period peaks at the splice.

Lowering the amplitude of the original recording or of the test stimuli with modified gap times presented no problems; it was at this stage that the techniques of Appendix 1 were developed.

Raising the amplitude, on the other hand, proved very difficult. The splices in the shortened or lengthened transitions, inaudible at normal or reduced amplitudes, suddenly became clearly audible.

Further difficulties for experimental design arose from the fact that the transitions were not equalized, as discussed in Chapter 2. Therefore, one transition could be shortened by an amount which was simply not available for another transition; and the amplitude dips of the tongued and untongued originals were not matched, making the placement of stimuli on Figure 5.1 difficult. Some of these problems were solved by limiting the amount of contraction and extension in time, and by calculating (separately for the tongued and untongued recordings) the change in amplitude relative to the original amplitude level of the dip in the transition.

Results

An initial set of test stimuli were presented to professional clarinet and trumpet players. Much to my surprise and disappointment, each complained strongly about the quality of the resulting transitions for his instrument. Indeed, the clarinettist crossed his legs and folded his arms as he made comments such as:

- "Sounds like a tape splice."
- "Strange ... sounds like a 'hoo' attack on the second note—which wouldn't work. I couldn't duplicate that."*
- "Sounds like an extra pop on the second note after it's started."
- "Pretty good but not quite right."

Figure 5.2 shows these and other comments by the clarinetist placed according to the arrangement of Figure 5.1. Comments preceded by T are for transitions created from the original tongued recording; U shows where the untongued recording was the starting point. The change in dB is relative to the amplitude of the transition in the original recordings. (Recall from Table 2.2 that these two transitions differ by about 24 dB). The disparity of responses at 30 msec between the original U and the foreshortened T is striking. What's worse, for all of the stimuli derived from the untongued notes, the clarinettist remarked that "they all sound tongued."

The trumpeter offered similar comments, paraphrased here:

- I can hear a discernible attack on the first note but not on the second.
- Students sometimes overdo tonguing with a syllable when their embouchure is weak.
- Students will lift off the air to get over a wide interval. I tell them to "keep the air going."

Of course, both players found some of the tones to be acceptable or even quite good. Indeed, the trumpeter remarked that some differences between adjacent stimuli would be hard to duplicate, even for a good player. But the the locations of acceptable tones on the plane of Figure 5.1 varied widely, depending on the instrument and which original recording was used. The solid line in Figure 5.2 shows a tentative "acceptable" region for modifications to both the tongued

*Piston (1955, p. 120) refers to this as a "htu" attack.

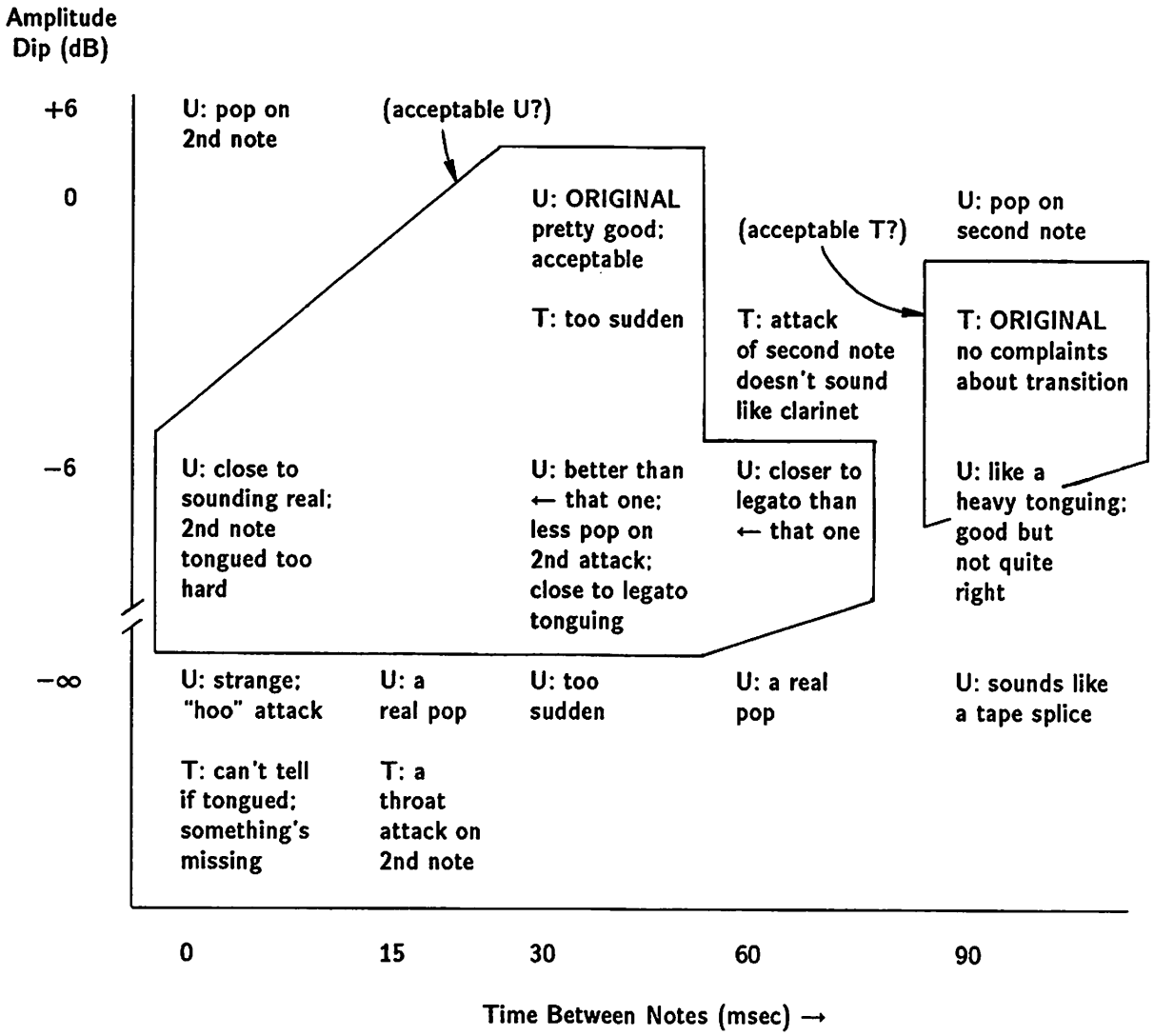


Figure 5.2. Remarks on clarinet test stimuli created to fall on selected points of Figure 5.1. (T: created from the tongued recording; U: from the untongued recording).

and untongued recordings. Contrary to what was expected (see Figure 5.1), these regions do not even overlap! This experience lead to the conclusion that it was impossible to systematically vary both amplitude and time independently without incurring unpredictable changes in the perceived transition. No other experimental paradigm could be found to test the question of the relative importance of time vs. amplitude, so further work on this question was abandoned. Still, the effort was not a complete loss, since this work was a direct precursor of some of the experiments which follow in this chapter and the next.

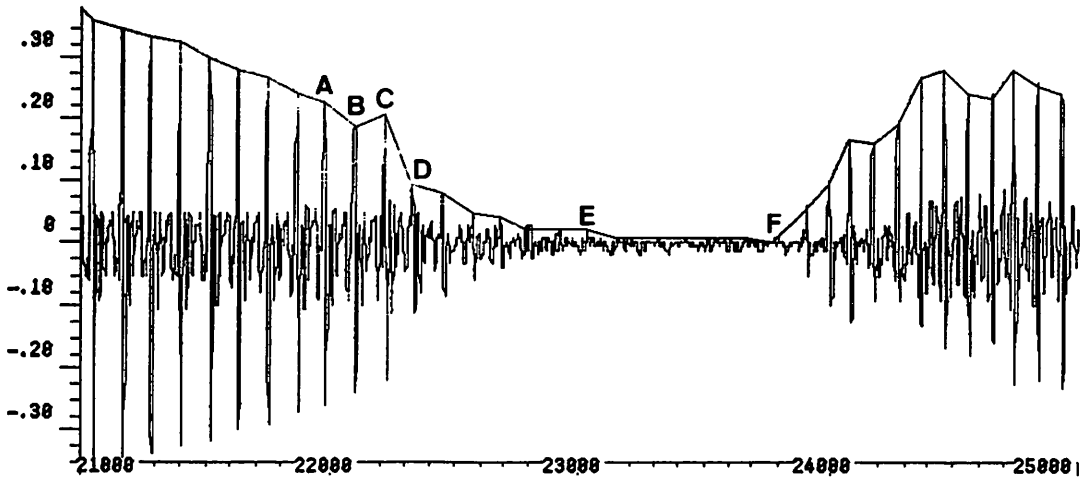


Figure 5.3. Tongued trumpet transition, with amplitude envelope. (X-axis: number of samples, at sample rate of 25 600 Hz)

Extending the time between the notes

The gap time between notes has been listed (Chapter 2) as one of the parameters of a transition. It would be interesting to leave the amplitude and spectral cues in place and vary only the gap time.

Creating the Stimuli

In an attempt to study this, I worked with the trumpet tongued ascending M3. The transition, shown here again in Figure 5.3, lasts approximately 60 msec. Initially, I moved the notes apart to see whether at some point they would sound completely detached. By detached, I mean a transition which sounds as though the player has deliberately performed two separate notes; two eighth notes, perhaps, separated by an eighth note rest, or two notes performed portato.

Of course, when the notes are moved apart, something must be supplied to fill the resulting gap. Using Method 2 of Appendix 2, I tried extending the transition at several places. It was impossible to isolate usable "periods" in any part of the region D–F in Figure 5.3, which seemed the logical place to start (F is the point of pitch change). Extending the period at E in Figure 5.3,

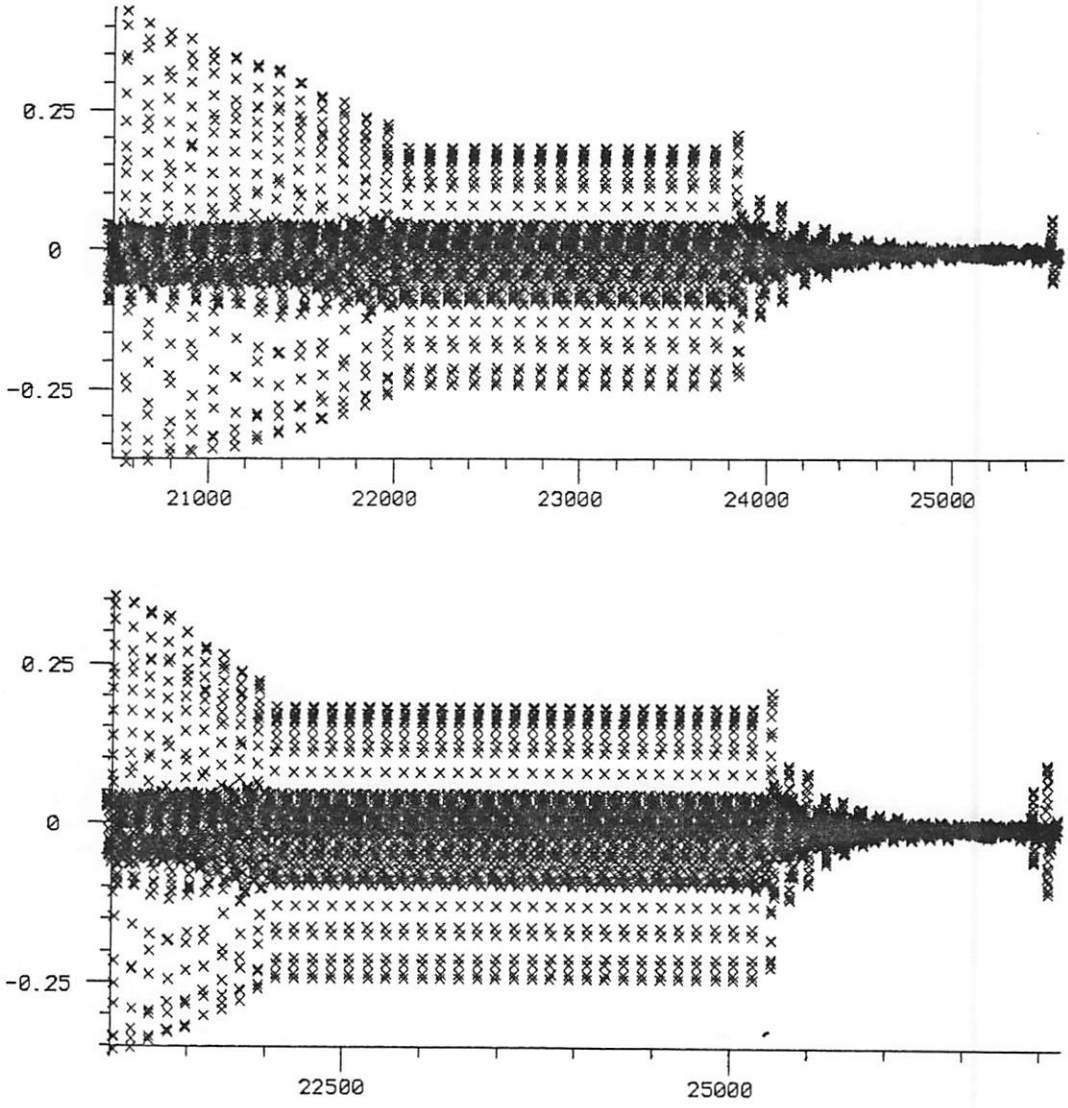


Figure 5.4. (top) The transition of Figure 5.3 has been extended at point B in that figure by about 60 msec, effectively doubling the length of the transition between the notes. (bottom) The extension here is 120 msec.

for example, produced a signal which was unusable because the periodicity implied by the period peaks in that region produced a pitch different from that of the first note.

So for this preliminary study I used the periods delineated by the peaks A–B–C in Figure 5.3. Extending this region by 60 and 120 msec, to double and treble the transition gap time, produced the transitions shown in the top and bottom of Figure 5.4, respectively. (This representation was chosen to highlight the location of individual samples, each of which is marked by an "x". In this manner, the exact location of each period peak is made clear. The dark region in the center of

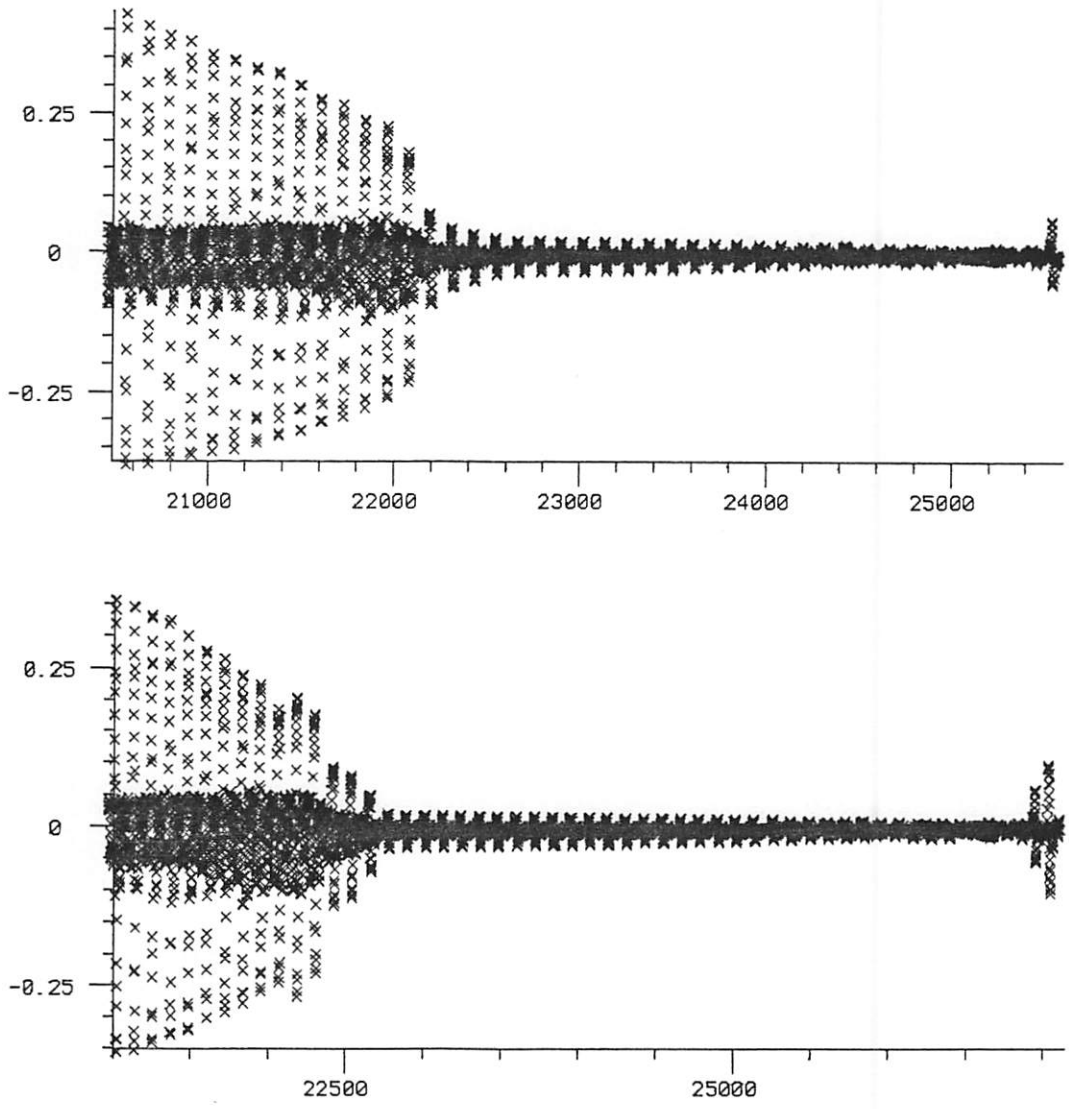


Figure 5.5. (top) The upper waveform in Figure 5.4 has been scaled in amplitude to match the amplitude envelope given by points C–F in Figure 5.3. (bottom) This same envelope has been applied to the bottom waveform from Figure 5.4.

each plot of course results when many x's overlap.) Neither of the extended transitions in the figure sounds particularly detached.

Using the methods discussed in Appendix 1 (where another example from this study is presented), I then applied the amplitude envelope of the original decay to the "barrel" extensions shown in Figure 5.4. The resulting transitions are shown in Figure 5.5. Figure 5.6 shows a different envelope that was applied to the 60-msec case of Figure 5.4a. Attempts to produce a

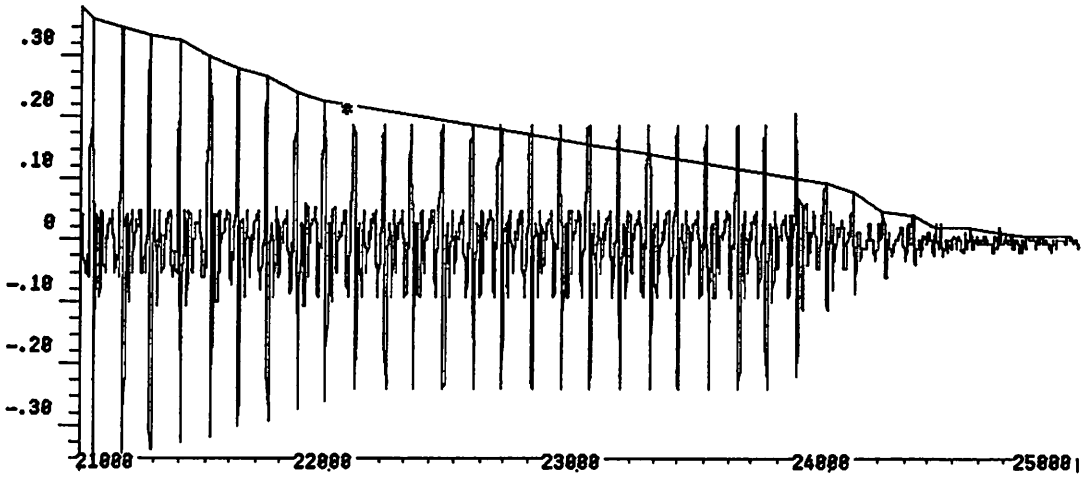


Figure 5.6. Another representation of the extended waveform from the top of Figure 5.4. An amplitude envelope different from that used in Figure 5.5 is also shown (solid line).

tone with a similar "softer" decay for the 120-msec extension proved impossible, as artifacts of the extension process became audible, and/or the results simply did not sound natural. At any rate, one would expect any amplitude-envelope-dependent effects to be even more pronounced for a longer extension.

Results

Both of the transitions in Figure 5.5 sounded detached. The "softer" envelope of Figure 5.6 did not sound detached. This produces the unexpected and interesting result that *in extending the region between two tongued notes, the shape of the decay plays a very important role.*

Similar studies of the violin bow-changed transition (shown in Figure A1.3) as well as of the clarinet (Figure A2.2) showed that the same effects occured for different instruments. From this work I have concluded that it is impossible to isolate the length of time between notes from the shape of the amplitude envelope of the decay of the first note (and probably of the attack of the second note as well). These factors interact in ways which made empirical study of the role of time impossible here.

Experiment 4: Amplitude Dip without Spectral Cues

Background

Experiment 3 demonstrated that it is possible, within certain bounds, to create an acceptable (legato) transition without any spectral or amplitude cues. Perhaps it would be possible to create other kinds of transitions simply by introducing an appropriate amplitude dip while still omitting the spectral cues. This experiment is designed to examine that possibility.

Creating the Stimuli

The overlapped untongued tones from all three instruments, with 20-msec overlap times, were used from Experiment 3. Similar tones with the same overlap time were created from the original tongued recordings of all three instruments as well.

These stimuli were prepared before the results for Experiment 3 were available. Thus, I had to choose one of the transitions from that experiment to use as a starting point for this experiment. At the time, it seemed reasonable to use the 20-msec crossfade, even though Experiment 3 later showed that this transition was not as acceptable as transitions with longer crossfade times. The procedure described below was tried with an 80-msec crossfade using the no-bow-change violin recording, but no difference between it and the test stimulus prepared from the 20-msec transition could be heard, so further attempts with other overlap times were not pursued. In general, the transitions of Experiment 3 changed remarkably for the better once the amplitude envelope was applied.

The six overlapped tones were extended to the length of their respective originals, so that the change in pitch occurred at the same time as in the original. In all cases, the steady-state portion of *both* notes had to be extended, as described in the Appendix of Chapter 4. Unfortunately, neither of the first two methods of Appendix 2 could be made to work for the second note of the trumpet tongued case (and Method 5 had not yet been developed); I tried two different recordings. These methods apparently failed because the steady-state of the second note was so short; the phase drift even across a few periods was so great that a reasonable-sounding extension was impossible to achieve with the methods then available. Pursuit of this question is beyond the scope of this dissertation; this topic was not explored further. Thus, for this experiment (and for Experiment 5), only the following five transitions were included: clarinet tongued and untongued; violin with and without bow change; and trumpet untongued.

The amplitude envelopes of the originals were calculated using the peak-finding method discussed in Chapter 2. These amplitude envelopes were applied to the overlapped transitions. It is necessary to be extremely careful with the placement of the amplitude dip relative to the center point of the overlapped transition (the nominal point of pitch change). That center point must fall where the amplitude of the attack of the second note begins to rise. The result was a set of five two-note test stimuli with amplitude dip and pitch change in the transition matching those of the original, but without the spectral cues in the decay and attack of the original. The control stimuli were of course the original recordings.

It should be noted in passing that some of the test stimuli sounded a little unnatural. For example, in brass tones generated in this manner, there was an effect which I heard as "reverberation" (and others heard as "comb filtering") at the end of the first note. Recall from Chapter 2 that the ear expects the spectrum to fall off with the drop in amplitude. My interpretation of this unexpected percept is that the ear "hears" the fall in amplitude without concomitant change in spectrum as reverberation. In a mildly reverberant environment, one would expect a drop in amplitude without a gross falloff in spectrum for several tens of milliseconds, which is approximately what happens here.

Experimental Procedure

The subjects heard four cases, each consisting of two stimuli separated by a short pause:

Comparison cases

1. Original recording vs. synthetic (amplitude envelope applied to overlap)
2. Synthetic vs. original

Identical cases

3. Original vs. itself
4. Synthetic vs. itself

The subjects were asked to rate these cases on a scale of 1 to 7, with 1 meaning that the stimuli were identical, and 7 representing the greatest perceived difference. A series of training examples presented at the beginning of the experiment included what I felt were the greatest possible differences. The comparison cases (1 and 2) were presented 3 times each; the identical cases (3 and 4) were presented twice each. (Details on the presentation of the stimuli are given in Appendix 3). As before, the identical cases are included to check for subject bias toward the first or second stimulus in each pair.

Table 5.1. Analysis of Experiment 4.

Case	Clarinet				Trumpet		Violin			
	U		T		U		U		T	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Comparison										
1	3.10	1.73	4.30	2.92	4.90	1.86	4.80	2.12	4.33	1.89
2	3.30	1.87	5.93	1.64	2.93	1.58	4.67	2.05	4.57	2.12
1 & 2	3.20		4.67		3.92		4.73		4.45	
Identical										
3	3.00	1.47	2.60	1.77	2.32	1.59	2.75	1.68	2.60	1.25
4	2.58	1.40	2.35	1.34	2.61	1.48	2.70	1.44	2.65	1.58

Note: U means untongued (no bow change), T means tongued (with bow change). "1 & 2" show the combined mean of Cases 1 and 2.

Results

Table 5.1 shows the means and standard deviations across all subjects for the five transitions studied here. The means for the identical cases (3 and 4) show that for all transitions, the subjects showed some bias toward hearing identical stimuli as different. This result, although unexpected, does not prohibit analysis of the data, as will be shown below.

Some order effects are noticeable. The comparison cases (1 and 2) are especially different for the trumpet untongued transition. Still, no consistent pattern for order effects can be found in the data, and so the combined comparison cases (in the row labelled "1 & 2" of Table 5.1) will be used in what follows.

Examination of the means for the comparison cases shows that they seemed to be larger than the means for the identical cases, indicating that the subjects seemed to find a noticeable difference between the original and synthetic cases. The untongued clarinet transition was the one possible exception.

The t test and analysis of variance were applied to see if the combined mean for the comparison cases differed significantly from the means of the identical cases. The results are given in Tables 5.2 and 5.3, respectively. The analysis of variance presented here follows the format given by Hays (1963, p. 372) already encountered in Chapter 3. (The values for df vary because some responses were missing; see Appendix 3.) The t value for the clarinet untongued transition implies $p > 5\%$. The same tendency ($p > 10\%$) is shown by the analysis of variance for that transition (see Table 5.3). Thus, it remains unclear whether the subjects truly found that the synthetic stimulus was significantly different from the original stimulus for the untongued clarinet. This is not too surprising, as the untongued clarinet transition (Figure 2.4) was perhaps the smoothest and

Table 5.2. *t* test for Experiment 4.

Case	Clarinet				Trumpet		Violin			
	U		T		U		U		T	
	μ	t	μ	t	μ	t	μ	t	μ	t
Comparison										
1 & 2	2.80	1.86	2.48	11.56	2.46	6.14	2.73	8.29	2.63	7.68
Identical										
3		0.71		0.34		−0.43		0.07		−0.10
4		−0.75		−0.46		0.46		−0.07		0.08

Note: μ, the expected mean of the overall population, is taken as the mean of cases 3 and 4 for each instrument, given in Table 5.1. $N = 100$.

quickest of the five transitions used here. For all other transitions, however, the difference in the means between the comparison cases and the identical cases seems to be statistically significant, given the high t and F values in Tables 5.2 and 5.3, respectively.

Discussion

No particular significance can be attached to the absolute value of the means for the comparison cases (1 and 2). It is surprising that these means did not take on a higher value. Examination of the raw data shows that all but one subject used the full range available. That subject (4) was not able to fit the stimuli onto the scale of 1–7; in other words, he found the scale to be too large for the stimuli. Here are his written comments:

> I listened to the test tones (twice) and tried to find your "ruler". In other words, as you will note by looking at my resulting scores on the "training tones," I did not perceive too much change in any of the 8 two-note pairs. I simply feel (and with some embarrassment) that your perception of transition is much more finely attuned [than mine].
>
> In the actual experiment, I think you'll see a range of scores between 1 and 4. Can you convert this to your 1 to 7 scale?

As the results of the experiment were clear enough, I did not attempt to rescale the responses of this or the other subjects, nor did I attempt to weed out "unsatisfactory" subjects before subjecting the data to further analysis.

Another factor may have been that this was the first experiment on the tape. It was apparently hard for some subjects to adjust to the kind of listening required, and perhaps this experiment showed such difficulties more than the ones which followed. For example, subject 1 reported:

Table 5.3. Analysis of Variance for Experiment 4.

Clarinet Tongued

Source	*SS*	*df*	*MS*	*F*	*p*
Presence of Spectral Cues	115.91	2	57.96	26.95	<0.1%
Error	208.68	97	2.15		
Totals	324.59	99			

Clarinet Untongued

Source	*SS*	*df*	*MS*	*F*	*p*
Presence of Spectral Cues	5.61	2	2.81	1.18	>10%
Error	228.23	96	2.38		
Totals	233.84	98			

Trumpet Untongued

Source	*SS*	*df*	*MS*	*F*	*p*
Presence of Spectral Cues	50.84	2	25.42	8.92	<0.1%
Error	273.49	96	2.85		
Totals	324.33	98			

Violin Bow Change

Source	*SS*	*df*	*MS*	*F*	*p*
Presence of Spectral Cues	78.72	2	39.36	14.97	<0.1%
Error	249.70	95	2.63		
Totals	328.42	97			

Violin No Bow Change

Source	*SS*	*df*	*MS*	*F*	*p*
Presence of Spectral Cues	96.83	2	48.42	16.41	<0.1%
Error	285.68	97	2.95		
Totals	382.51	99			

I personally feel better about my judgement on the last 20 or so examples. Took a while to start concentrating on the transition ...

Conclusion

Except for the untongued clarinet transition, the subjects definitely noticed a difference between the original recording and the synthetic transition containing the original amplitude changes but with the spectral changes in the transition removed. Therefore, the subjects noticed the absence of spectral cues except when the transition was very smooth or very quick (as in the untongued clarinet transition).

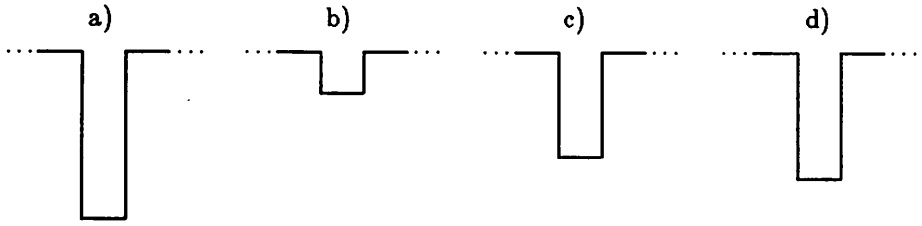


Figure 5.7. Experiment 5. a) Idealized amplitude envelope of transition. b)–d) Variations in the minimum amplitude of the transition.

Experiment 5: Variations in Amplitude Dip

Background

Ideally, one would like to vary the amplitude of a transition while keeping changes in spectral and timing information to a minimum. The discussion earlier in this chapter showed that this is difficult to do. Another difficulty arises with natural tones: If the amplitude of a transition is *raised*, then the amplitude of the spectral cues in the transition will be raised as well. It quickly became clear that amplifying, say, the sound of the bow change in the middle of the transition was not going to be useful for experimental investigations, as the results were ugly indeed.

Still, the question remains: For both the tongued and the untongued transitions, how far can the amplitude dip be raised or lowered from what occurs in nature, while still producing an acceptable transition? In particular, does lowering the amplitude to 0 result in the percept of separated notes? This has immediate practical interest, of course, for composers and performers using digital synthesizers. To answer this question, this experiment (like Experiment 4) uses the overlapped transitions of Experiment 3, that is, a transition with no special spectral cues at all. In this manner, it is possible to raise the amplitude of the transition without introducing the kinds of distortion just mentioned.

Initially, it seemed reasonable to assume that the transition followed a shape such as given in Figure 5.7a. Test stimuli could be created by raising and lowering the amplitude floor, as shown in Figure 5.7b–d. Recall that the recordings had an effective noise floor of about −60 dB. It seemed reasonable to create test tones with amplitude dips at −12 dB intervals. One extreme would be a test tone with no amplitude dip (as in Experiment 3). The other extreme, substituting for a test tone with a −60 dB dip, would be a test tone with 0 amplitude in the transition (this case will be called $-\infty$ dB here).

Preliminary work showed that the vertical edge of Figure 5.7a was inappropriate when the amplitude dip grew large—a click was of course produced. Even worse was the fact that for some instruments, the transitions of 0 dB and −12 dB produced inadequate upward variation in the amplitude of the transition; as shown in Table 2.2, the minimum of the untongued clarinet transition lay at a mere −13 dB.

Creating the Stimuli

Therefore, amplitude dips were selected for the following values, relative to the amplitude of 0.75 to which all two-tone recordings were equalized:

0 dB (no dip)
−7.5 dB
−15 dB
−30 dB
−45 dB
$-\infty$ dB

Thus, six test stimuli were created for each of the five transitions. (As in Experiment 4, the trumpet tongued transition was omitted).

Furthermore, the simplistic model of Figure 5.7 was modified, as shown in Figure 5.8. For the first note, all amplitude values before and including point 1 in Figure 5.8 remained unchanged in each test stimulus; the same was true for all amplitude values including and after point 6 for the second note. Points 3 and 4 in the figure took on the required amplitude value from the list just given. As for the other points, consider point 2. In the 0 dB case, it lay on the same horizontal line as the other points. For the other cases, point 2 moved down as point 3 moved down, but point 2 never moved down farther than the amplitude of the original in the first note at that point. For the second note, point 5 behaved in the same way. Points 1 through 6 differed in time and amplitude for each transition, and were determined by inspection. The times for each of the transitions are summarized in Table 5.4. The six amplitude curves for each of the five transitions are shown in Figure 5.9. To create the thirty test stimuli, then, the envelopes in Figure 5.9 were applied to the corresponding (20-msec) overlapped transitions already encountered in Experiments 3 and 4. The control stimuli were the original recordings of the five transitions in question.

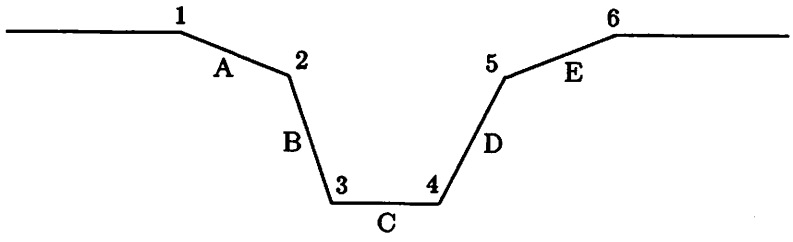


Figure 5.8. Model of amplitude during transition between notes, adopted for Experiment 5.

Table 5.4. Transition times for Experiment 5.

		A	B	C	D	E
Clarinet	tongued	348	1040	567	1082	463
		14	41	22	42	18
	untongued	116	321	266	642	554
		5	13	10	25	22
Trumpet	untongued	819	588	257	1430	3560
		32	23	10	56	139
Violin	bow change	811	682	479	1484	907
		32	27	19	58	36
	no bow change	1386	301	1118	387	459
		54	12	44	15	18

Note: For each transition, the first line gives number of samples; the second line gives time in milliseconds. The letters refer to the segments in Figure 5.8.

Experimental Procedure

As in Experiments 1 and 2, the subjects were asked to state a preference for one of two stimuli. For each transition, the six test stimuli and the control stimulus were arranged in the 19 cases given in Table 5.5. For each of the five transitions, the comparison cases (1–12) were presented three times each; the identical cases (13–19) were heard twice each. (Details on the presentation of the stimuli are given in Appendix 3.)

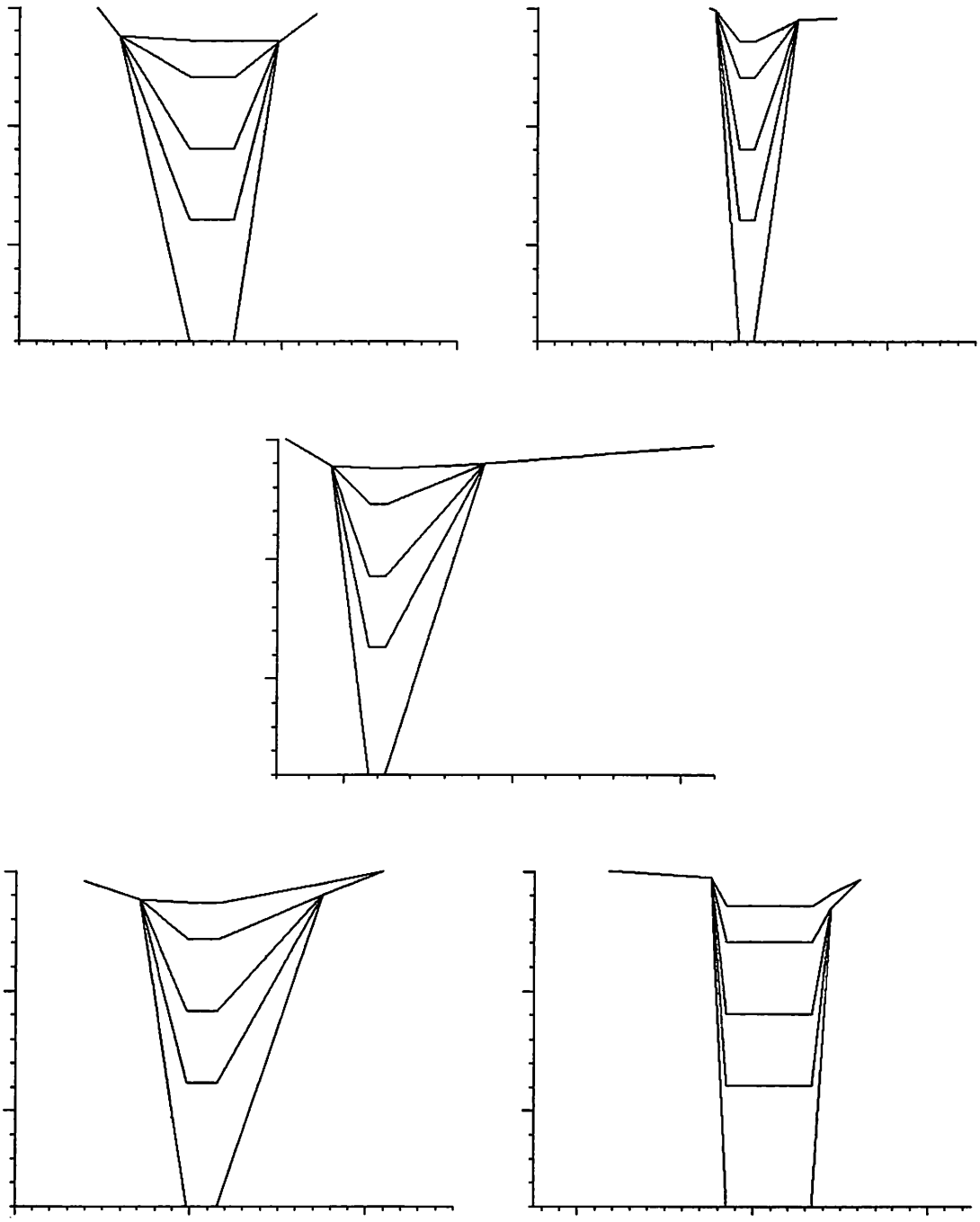


Figure 5.9. Amplitude envelopes used to generate test tones for Experiment 5. Top: Clarinet tongued and untongued. Middle: Trumpet untongued. Bottom: Violin with and without bow change. All plots show 250 msec on a 60 dB scale.

Table 5.5.
The Cases for Experiment 5

Comparison cases	
1.	0 dB test simulus vs. original
2.	−7.5 dB test simulus vs. original
3.	−15 dB test simulus vs. original
4.	−30 dB test simulus vs. original
5.	−45 dB test simulus vs. original
6.	$-\infty$ dB test simulus vs. original
7.	Original vs. 0 dB test simulus
8.	Original vs. −7.5 dB test simulus
9.	Original vs. −15 dB test simulus
10.	Original vs. −30 dB test simulus
11.	Original vs. −45 dB test simulus
12.	Original vs. $-\infty$ dB test simulus
Identical cases	
13.	0 dB test simulus vs. itself
14.	−7.5 dB test simulus vs. itself
15.	−15 dB test simulus vs. itself
16.	−30 dB test simulus vs. itself
17.	−45 dB test simulus vs. itself
18.	$-\infty$ dB test simulus vs. itself
19.	Original vs. itself

Results

For numerical analysis, the values of 1.0 and 2.0 were assigned to the subject's preference for the first and second stimulus, respectively. Examination of the responses showed no significant differences between comparison cases 1–6 on the one hand and comparison cases 7–12 on the other, so any order effects in the first 12 cases will be disregarded. For the (combined) comparison cases, a response of 1.0 indicates that the synthesized stimulus was preferred. Table 5.6 shows the means and standard deviations for all five transitions.

Recall from Experiments 1 and 2 that a mean response of (in this instance) 1.5 showed that the subject had no preference for either stimulus, or perhaps could not distinguish between the stimuli. In the cases of two identical stimuli, a mean of 1.5 was expected.

In the data for the identical cases in Table 5.6, some order effects can be seen. That is, in some instances the subjects consistently preferred the first or the second of two identical stimuli. However, no pattern in this data can be found by inspection of the raw data. This conclusion is supported by analysis of variance of the identical cases (Table 5.7). As in Experiment 3, two-way

Table 5.6. Results of Experiment 5.

Case	Clarinet				Trumpet		Violin			
	T		U		U		T		U	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Comparison										
1 & 7	1.85	0.43	1.80	0.47	1.95	0.34	1.68	0.52	1.85	0.43
2 & 8	1.97	0.31	1.67	0.52	1.93	0.36	1.77	0.49	1.90	0.39
3 & 9	1.87	0.42	1.67	0.53	1.90	0.39	1.70	0.51	1.85	0.43
4 & 10	1.72	0.51	1.65	0.53	1.87	0.42	1.67	0.52	1.93	0.35
5 & 11	1.58	0.54	1.73	0.50	1.83	0.45	1.68	0.52	1.95	0.34
6 & 12	1.63	0.53	1.70	0.51	1.83	0.45	1.63	0.53	1.93	0.36
Identical										
13	1.55	0.62	1.40	0.60	1.50	0.62	1.40	0.60	1.50	0.62
14	1.40	0.60	1.50	0.62	1.40	0.60	1.45	0.61	1.35	0.58
15	1.45	0.61	1.50	0.62	1.35	0.58	1.60	0.62	1.70	0.61
16	1.30	0.56	1.65	0.62	1.45	0.61	1.55	0.62	1.20	0.49
17	1.60	0.62	1.55	0.62	1.45	0.61	1.45	0.61	1.25	0.53
18	1.30	0.56	1.50	0.62	1.25	0.53	1.65	0.62	1.40	0.60
19	1.45	0.61	1.40	0.60	1.65	0.62	1.60	0.62	1.70	0.61

Note: T means tongued (with bow change), U means untongued (no bow change). Results for cases 1–6 have been combined with those for cases 7–12. A value of 1.0 in the comparison cases means that the synthesized stimulus was preferred; 2.0 shows a preference for the original. In the identical cases, 1.0 and 2.0 indicate a preference for the first and second stimulus presented, respectively.

Table 5.7. Analysis of Variance of Identical Cases in Experiment 5.

Source	*SS*	*df*	*MS*	*F*
Amplitude Dip (A)	1.64	6	0.27	1.08
Transition (T)	1.11	4	0.28	1.12
Transition × Amplitude Dip (TA)	8.67	24	0.36	1.44
Error (A × T × TA)	162.95	665	0.25	
Totals	174.37	699		

analysis of variance (Hays 1963, p. 402) is required. All of the F values imply $p > 20\%$. Thus, order effects in the identical cases will be disregarded here.

A graphic representation of the (combined) data for the comparison cases is given in Figure 5.10. All of the points for all of the transitions lie in the range of 1.5–2.0, which means that the subjects preferred the transitions which included spectral cues in all cases, for all instruments.

The x's in the figure show the amount of amplitude dip in the original transitions, taken from Table 2.2. For three of the five transitions, the subjects preffered transitions with amplitude dip

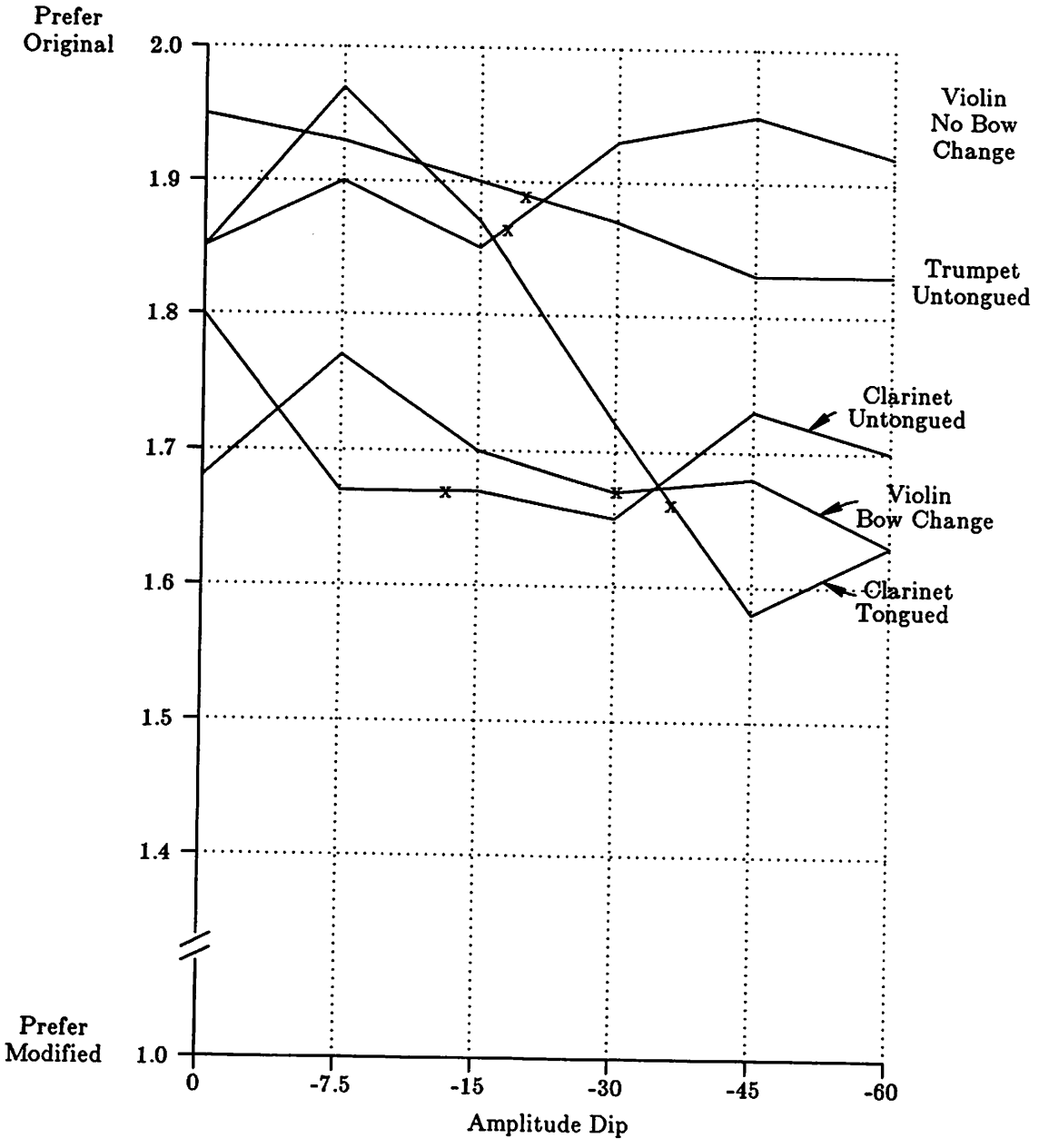


Figure 5.10. Mean responses for combined comparison cases in Experiment 5.

at or lower than that of the original. That is, three of the curves show a downward trend toward the x's in the figure. This effect is especially striking for the tongued clarinet transition, and may explain the unusual shape of the tongued clarinet curve in the figure.

Table 5.8. Analysis of Variance of Comparison Cases in Experiment 5.

Source	*SS*	*df*	*MS*	*F*	*p*
Amplitude Dip (A)	2.43	5	0.49	3.06	<0.5%
Transition (T)	15.02	4	3.76	23.50	<0.1%
Transition × Amplitude Dip (TA)	7.43	20	0.37	2.31	<0.1%
Error (A × T × TA)	275.48	1770	0.16		
Totals	300.36	1799			

The question then arises as to whether the curves in Figure 5.10 vary significantly from each other. Two-way analysis of variance (Table 5.8) was again used to answer this question. From this analysis I draw the following conclusions:

1. Given the *F* value in the "Amplitude Dip" row, the curves change more than one would expect from random variation; that is, the curves are not "flat."
2. The mean values for each transition are significantly different from each other, as the *F* value in the "Transition" row is so large. Again, as in Experiment 4, the untongued clarinet appears less susceptible than the other instruments to the modifications made in this experiment.
3. The *F* value in the third line suggests that the shapes of the curves in Figure 5.10 for each instrument are significantly different from each other. Thus, each instrument reacts to the change in amplitude dip in its own way.

Still, no pattern could be found which characterizes the behavior of all three instruments as the amount of amplitude dip changes. (Recall that in Experiment 4, all of the instruments followed more or less the same pattern). The possibility remained that one or more subjects might have given biased answers which would then cloud the data, making a pattern unrecognizable.

Examination of the raw data suggested that this might indeed be the case. Table 5.9 shows the data for the comparison cases. The entries in the table show how often the subject preferred the original over the synthesized transition; the maximum possible score was 6.

The asterisks mark data which might be questioned. Consider subject 4's responses for the tongued clarinet transition. In all but 3 of the 36 possible responses, he preferred the original, no matter how the transition was modified.

A column of data for a given transition has been marked with an asterisk if the responses satisfy two criteria: 1) At least 3 of the 6 entries for a transition must be 6, and 2) the other 3

Table 5.9. Preference for Original in Comparison Cases of Experiment 5.

	Subject Number									
Case	1	2	3	4	5	6	7	8	9	10
Clarinet tongued										
				*		*	*		*	
1 & 7	6	5	3	5	2	6	6	6	6	6
2 & 8	6	6	6	6	4	6	6	6	6	6
3 & 9	6	6	4	6	0	6	6	6	6	6
4 & 10	3	4	5	6	1	6	6	3	6	3
5 & 11	3	3	4	5	0	6	5	3	6	0
6 & 12	5	3	3	5	1	6	6	1	5	3
Clarinet untongued										
						*			*	
1 & 7	6	6	2	5	1	6	6	6	5	5
2 & 8	3	4	3	2	2	6	5	5	6	4
3 & 9	1	5	3	5	1	5	3	4	6	4
4 & 10	3	4	4	5	0	6	4	3	6	4
5 & 11	4	4	5	6	1	6	4	4	6	4
6 & 12	5	5	5	5	3	6	4	2	5	2
Trumpet untongued										
	*	*				*	*		*	*
1 & 7	6	6	6	6	3	6	6	6	6	6
2 & 8	6	6	3	5	6	6	6	6	6	6
3 & 9	6	5	5	6	2	6	6	6	6	6
4 & 10	5	6	4	6	3	6	6	5	6	5
5 & 11	6	6	3	4	2	6	6	5	6	6
6 & 12	6	6	5	5	1	6	6	4	6	5
Violin bow change										
						*	*		*	
1 & 7	5	5	6	3	3	6	6	0	6	1
2 & 8	6	6	1	4	5	6	6	3	6	3
3 & 9	6	5	5	5	2	6	6	2	5	0
4 & 10	6	5	1	2	5	6	6	2	6	1
5 & 11	6	5	1	3	5	5	6	3	6	1
6 & 12	4	6	0	4	3	6	6	2	6	1
Violin no bow change										
	*	*				*	*	*	*	*
1 & 7	6	6	4	4	3	5	6	5	6	6
2 & 8	6	6	4	5	4	6	6	5	6	6
3 & 9	6	5	5	5	2	6	6	5	6	5
4 & 10	5	6	3	6	6	6	6	6	6	6
5 & 11	6	6	4	6	6	6	6	6	6	5
8 & 12	6	6	3	5	6	6	6	6	6	6

Table 5.10. Results of Experiment 5, Omitting Subjects 6, 7, and 9.

Case	Clarinet				Trumpet		Violin			
	T		U		U		T		U	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Comparison										
1 & 7	1.79	0.50	1.74	0.52	1.93	0.40	1.69	0.54	1.83	0.47
2 & 8	1.95	0.37	1.55	0.56	1.91	0.42	1.74	0.52	1.88	0.44
3 & 9	1.81	0.49	1.52	0.56	1.86	0.46	1.69	0.57	1.81	0.49
4 & 10	1.67	0.54	1.57	0.56	1.83	0.47	1.62	0.55	1.91	0.42
5 & 11	1.48	0.56	1.67	0.54	1.79	0.50	1.64	0.55	1.93	0.40
6 & 12	1.62	0.55	1.69	0.54	1.81	0.49	1.57	0.56	1.91	0.42
Identical										
13	1.71	0.67	1.57	0.67	1.71	0.67	1.43	0.65	1.50	0.67
14	1.50	0.67	1.57	0.67	1.50	0.67	1.50	0.67	1.36	0.62
15	1.50	0.67	1.57	0.67	1.50	0.67	1.57	0.67	1.79	0.65
16	1.29	0.59	1.50	0.67	1.50	0.67	1.57	0.67	1.14	0.48
17	1.71	0.67	1.64	0.67	1.57	0.67	1.43	0.65	1.29	0.59
18	1.29	0.59	1.57	0.67	1.36	0.62	1.86	0.63	1.50	0.67
19	1.50	0.67	1.50	0.67	1.57	0.67	1.57	0.67	1.71	0.67

Note: U means untongued (no bow change), T means tongued (with bow change). Results for cases 1–6 have been combined with those for cases 7–12. A value of 1.0 in the comparison cases means that the synthesized stimulus was preferred; 2.0 shows a preference for the original. In the identical cases, 1.0 and 2.0 indicate a preference for the first and second stimulus presented, respectively.

entries must be 5 or 6. The subject's data was removed for further analysis if his responses for at least four of the possible five transitions satisfied these criteria. This was the case for three subjects: 6, 7, and 9. Table 5.10 shows the means and standard deviations for all five transitions, omitting these three subjects. As in Table 5.6, no order effects can be seen for the identical cases (13–19), so these cases will not be discussed further here.

A graphic representation of the data for the (combined) comparison cases, omitting subjects 6, 7, and 9, is given in Figure 5.11, which should be compared with Figure 5.10. The effects of removing the three subjects can be seen in Figure 5.11; each of the curves shows a wider variation, although the overall shape remains the same for each curve. In fact, the two clarinet curves lie much closer to (and in one instance cross) the mean value of 1.5, indicating that for this population the preference of the original over the synthesized transition was not as strong. Still, all but one of the points for all of the transitions lie in the range of 1.5–2.0. Thus, even after those subjects who consistently chose the original have been removed, I conclude that the subjects still preferred the transitions which included spectral cues in almost all cases, for all instruments.

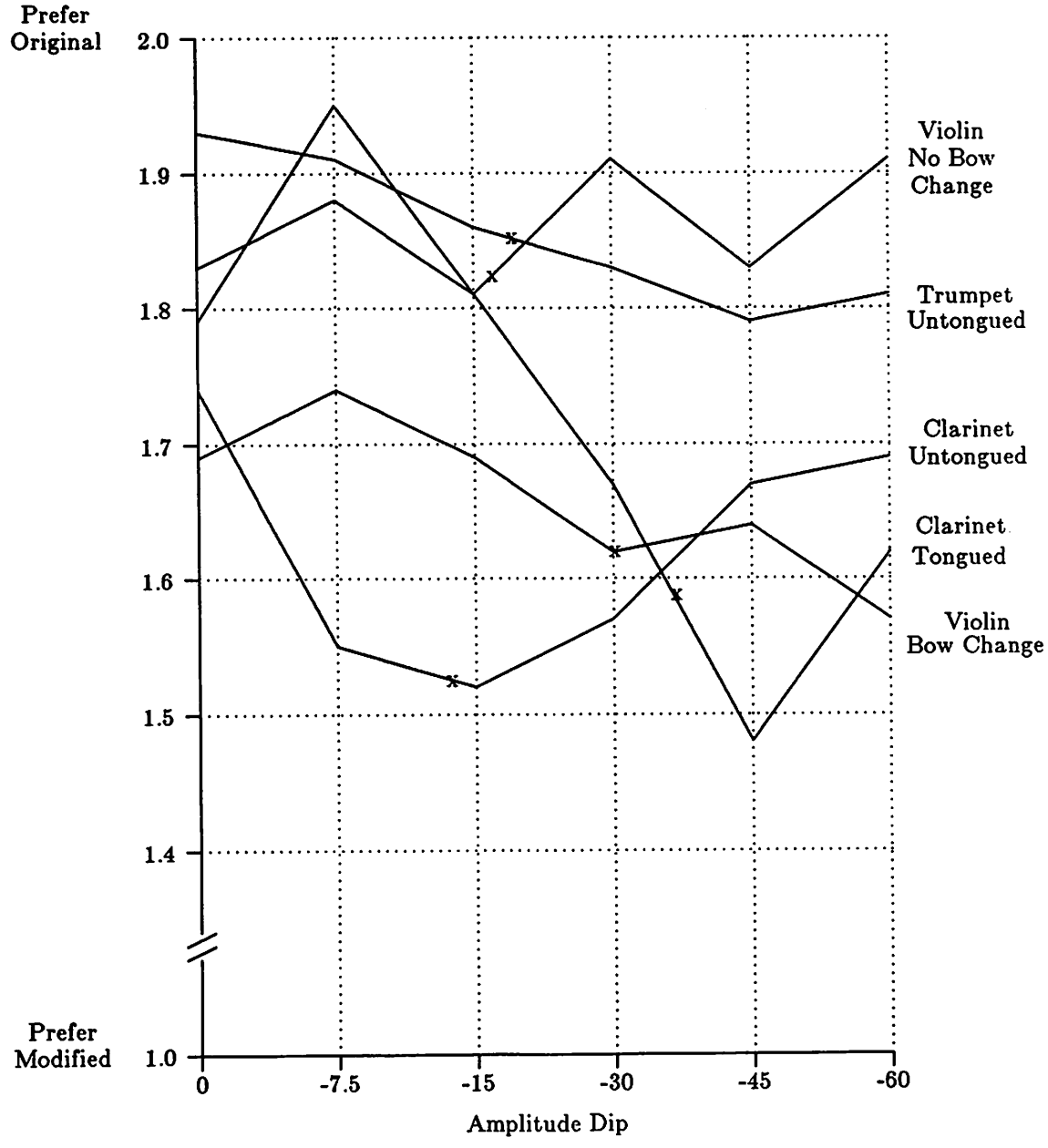


Figure 5.11. Mean responses for combined comparison cases in Experiment 5, omitting subjects 6, 7, and 9.

Table 5.11. Analysis of Variance of Comparison Cases in Experiment 5, Omitting Subjects 6, 7, and 9.

Source	*SS*	*df*	*MS*	*F*	*p*
Amplitude Dip (A)	2.00	5	0.40	2.22	<5%
Transition (T)	13.14	4	3.29	18.28	<0.1%
Transition × Amplitude Dip (TA)	7.26	20	0.36	2.00	<0.5%
Error (A × T × TA)	216.33	1230	0.18		
Totals	238.73	1259			

The question then arose as to whether the curves in Figure 5.11 vary significantly from each other. Two-way analysis of variance was once again used to answer this question. Table 5.11 shows the results. Comparison of this table with Table 5.8 suggests that removing the data for subjects 6, 7, and 9 does not modify the conclusions reached earlier.

The possibility still remained that a pattern in the data might be obscured because some subjects might prefer, say, the synthesized transition for one instrument but not for another. To test this, all of the data marked with an asterisk in Table 5.9 was removed. In other words, only the following subjects were included for each transition:

Clarinet tongued:	1 2 3 5 8 10
Clarinet untongued:	1 2 3 4 5 7 8 10
Trumpet untongued:	3 4 5 8
Violin bow change:	1 2 3 4 5 8 10
Violin no bow change:	3 4 5

Analysis of this data led to the same conclusions as those already reached, so a detailed discussion of this reduced data set will be omitted here.

Conclusions

When presented with the choice between the original transition or a transition containing only some change in amplitude, some subjects always preferred the original transition, and other subjects in general preferred the original. As in Experiment 3, the effects observed were weakest for the untongued clarinet transition, probably because of its extremely short duration.

This conclusion serves to modify the conclusions for Experiment 3. Recall that in that experiment, the subjects found that a transition with no amplitude dip and with no spectral cues was *acceptable*. Here, the subjects showed that they *preferred* the original in such instances (top

line in Table 5.6, left-most data points in Figure 5.10). Thus, even though a transition without spectral cues may be acceptable, a transition with spectral cues is preferred.

The effect of raising or lowering the amplitude dip in the transition is different for each instrument. No matter how large the dip in amplitude in the synthetic stimulus, subjects consistently preferred the original transition, which suggests that amplitude cannot be adjusted to compensate for missing spectral cues. For practical work, the time-varying spectral changes associated with a transition should be included, at least for transitions from instruments that are familiar to the subject.

In general, the amplitude dip for both tongued and untongued transitions should lie in the range of about -10 to -40 dB found in the original stimuli (see Table 2.2). Raising the amplitude above about -10 dB (relative to the maximum of either note) is not recommended, based on the data analyzed here.

The transitions in Figure 5.10 fall into two convenient groups for the $-\infty$ dB transition: The two tongued transitions (violin with bow change plus clarinet tongued) lie together, with the three untongued transitions lying closer to the top of the figure, indicating a preference for the original stimulus. It may be that tongued transitions can be successfully created with absolute silence between the notes, without causing the notes to be heard as separate; but this is not recommended for untongued transitions.

Experiment 6: Slope

Experiment 5 examined the effects of changing the amplitude in the transition, especially at the bottom of the "dip" in the transition. The discussion on pp. 121–124 examined the effects of moving the two notes apart. It would also be interesting to examine what happens when the dip at the bottom part of the note remains unchanged, but the "shoulders" in the transition are moved back and forth in time. By "shoulder", I mean lines A, B, D, and E in Figure 5.8. Initial investigations failed to find a useful way of using the *times* of the "shoulders" as an experiment variable. The slope of the decay of the first note and of the attack of the second thus became the focus of this study. As is often the case, the search for an experiment followed a twisted path, which will be traced here before the experiment itself is presented.

Preliminary Studies

Figure 5.12 shows an enlarged view of the clarinet tongued and untongued ascending M3 transitions. Point 1 in the figure was chosen to mark the start of the decay of the first note; point 2 marks the end of the decay. The attack of the second note begins at point 3 and finishes at point 4. As in other studies, these points were chosen to lie on the peak of a period. Table 5.12 summarizes the characteristics of the points and the lines which connect them. At this stage, the choice of points 3 and 4 was influenced in part by a desire to have them more or less equal in amplitude, facilitating a comparison between the tongued and untongued transition.

The first step is to investigate what happens when the slopes of lines A and C are modified. This model has the advantage of leaving the region between points 2 and 3 unchanged as the slopes of lines A and C vary. Figure 5.13a shows the amplitude envelope which would result if point 1 were moved to the right by 1/2 of the time between points 1 and 2, and point 4 were moved to the left by 1/2 of the time between points 3 and 4. Since the original duration of line A was about 20 msec, point 1 was moved by 10 msec; by the same reasoning, point 4 moved by about 6 msec. Given the short durations of lines A and C, it seemed impractical to attempt finer adjustments of points 1 and 4. If these points are moved by the same respective amounts in the other direction, the result would be as shown in Figure 5.13b. More drastic changes can of course be made; Figures 5.13c and 5.13d show the results when these points are moved by 50 and 100 msec, respectively.

Working in this manner, I settled on the following displacements for points 1 and 4:

A. +10 msec or 1/2 of the duration of A or C in Figure 5.12, whichever is smaller.
B. 0 msec (i.e., original)
C. −10 msec
D. −50 msec
E. −100 msec
F. −200 msec
G. −300 msec

The letters match those in Figures 5.14 and 5.15, which show the resulting amplitude envelopes for the tongued and untongued clarinet transition.

The methods discussed in Appendix 1 were used to create test stimuli whose amplitude envelopes followed those shown in Figures 5.14 and 5.15. Stimuli with envelopes A and C were difficult if not impossible to distinguish from the originals. Envelope D produced a slight softening

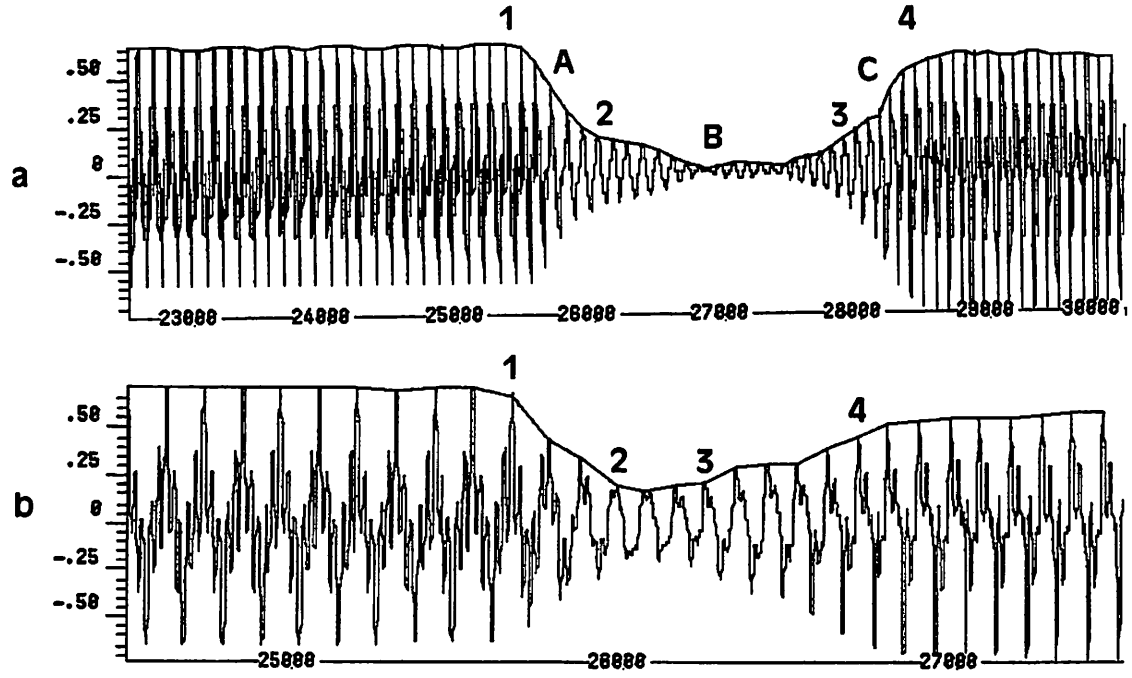


Figure 5.12. Preliminary study for Experiment 6. a) Detail of the tongued clarinet transition of Figure 2.3, with amplitude envelope (solid line). b) Detail of the untongued clarinet transition of Figure 2.4.

Table 5.12. Clarinet Transitions in Experiment 6.

	Amplitude (dB)	
Point	Tongued	Untongued
1	−1	−1
2	−12	−1
3	−14	−1
4	−2	−3

	Tongued		Untongued	
Line	Duration (sec)	Slope (dB/sec)	Duration (sec)	Slope (dB/sec)
A	0.023	−478	0.012	−916
B	0.072		0.010	
C	0.025	480	0.021	381

Note: Points 1–4 and lines A, B, and C are shown in Figure 5.12. Amplitude values are relative to two-note maximum of 0.75.

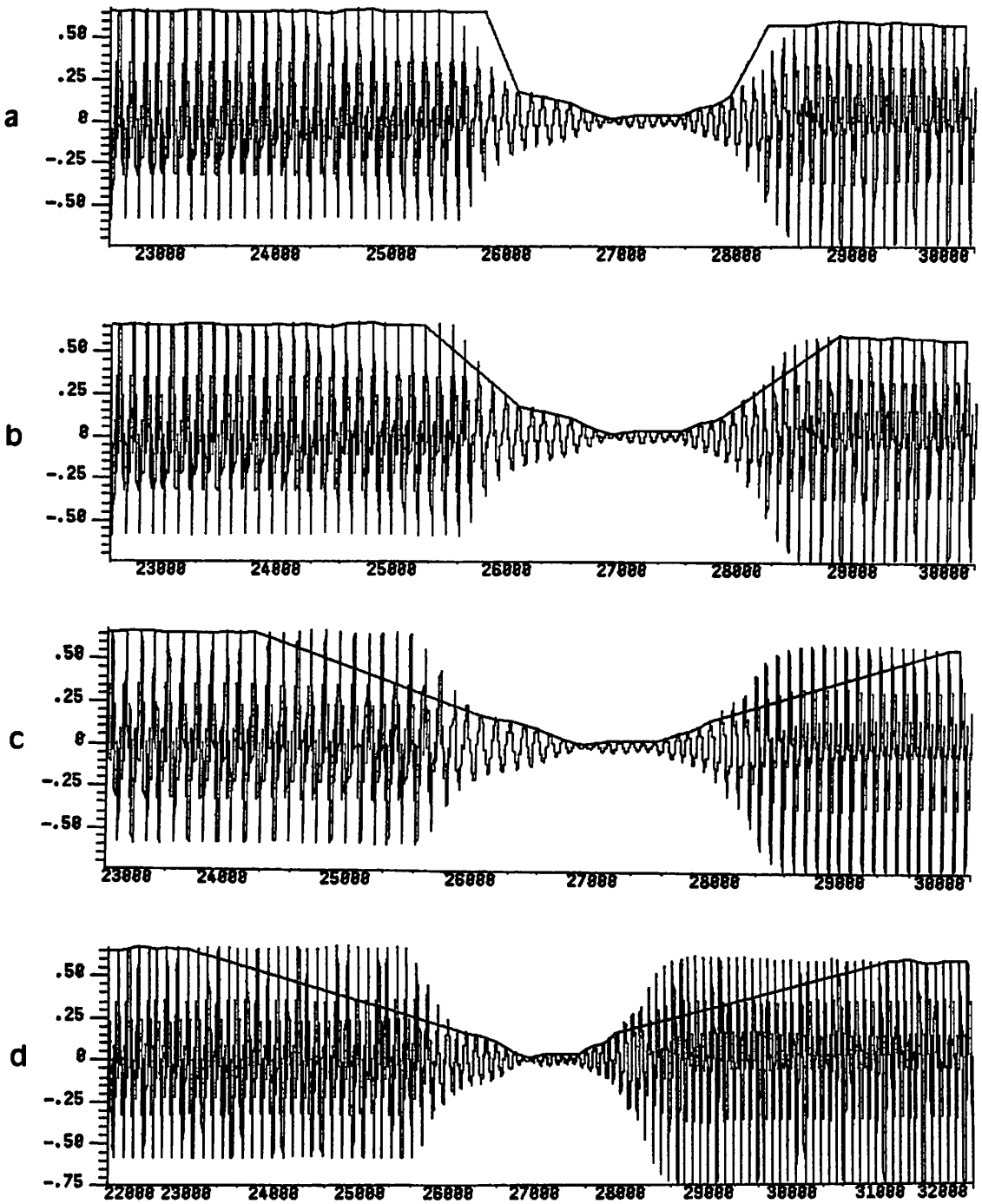


Figure 5.13. Modification of the tongued clarinet transition of Figure 5.12a.

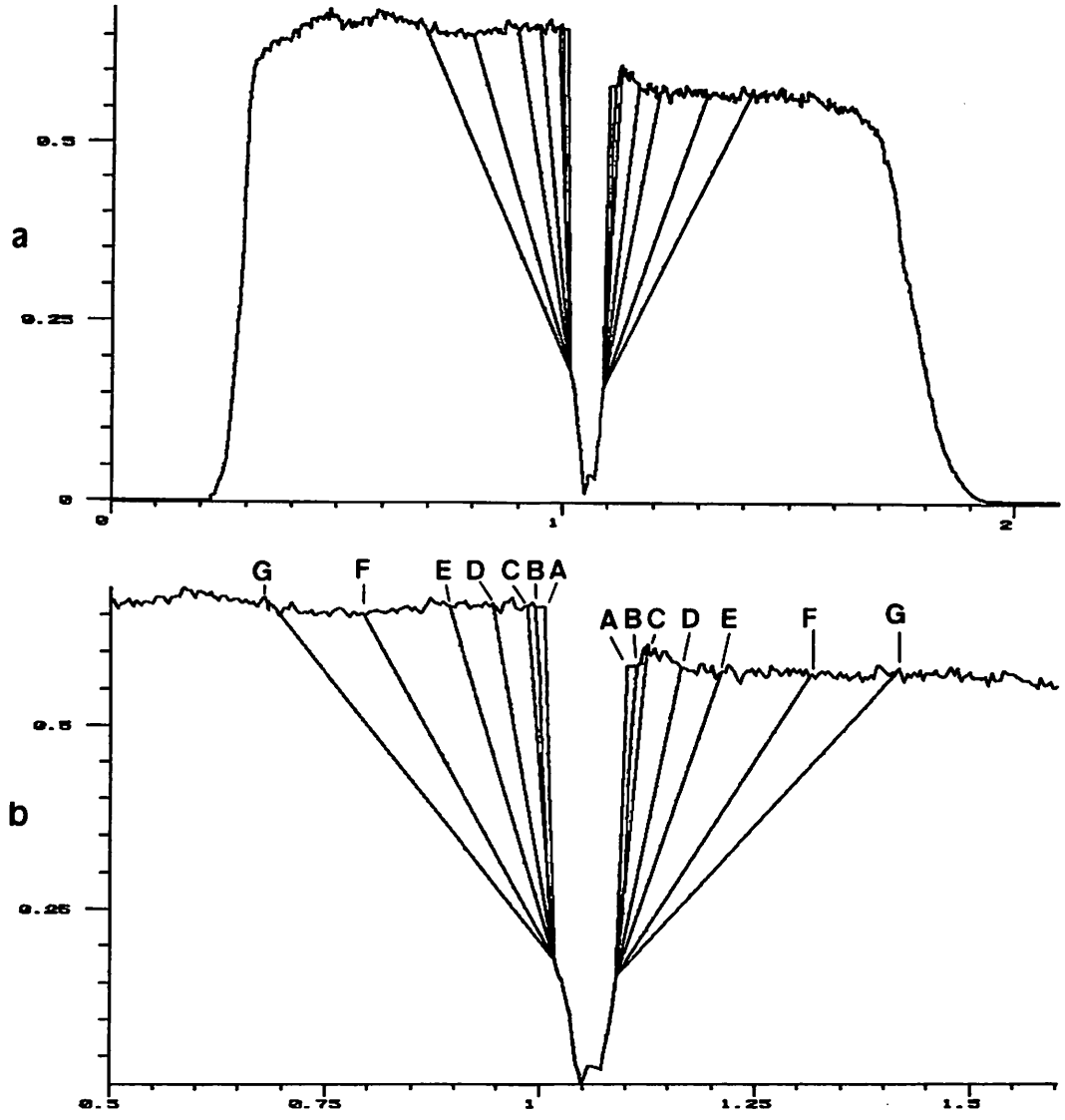


Figure 5.14. Amplitude envelopes for modifying tongued clarinet transitions. b) is a closeup of a).

of the attack and decay. The most pronounced effect occurred with envelopes E, F, and G; for the decay of the first note, they produced a diminuendo effect, as though the player were purposefully softening the end of the note. A corresponding effect, like the "reverse envelope" in rock guitar, occurred for the attack of the second note—as though the player were "swelling into" the note. Furthermore, the change in slope did not cause the untongued transition to become tongued, as one might expect, say, for envelope G in Figure 5.15, with its wide gap between the notes.

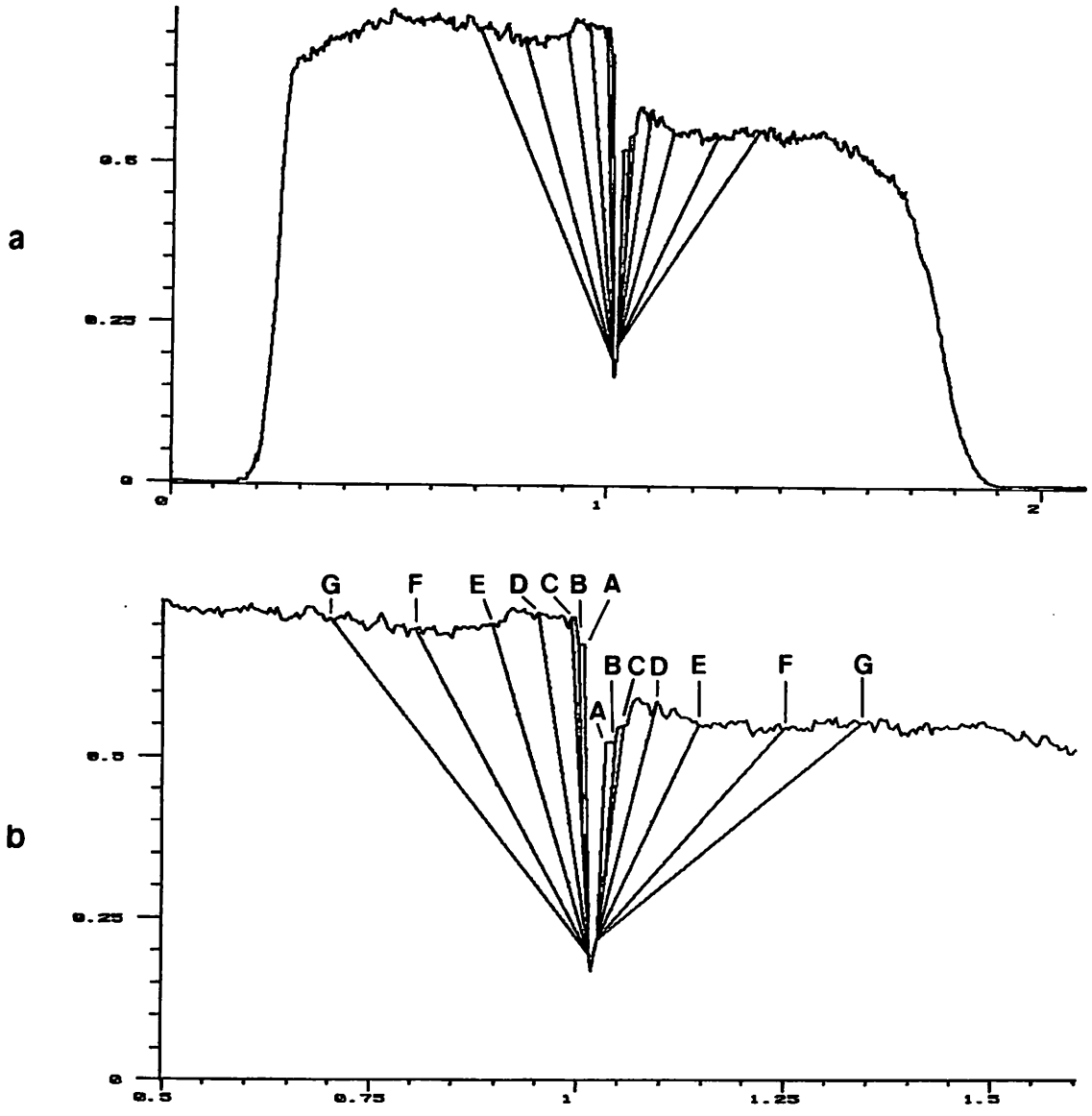


Figure 5.15. Amplitude envelopes for modifying untongued clarinet transitions. b) is a closeup of a).

Lengthy consideration of these test stimuli led to the conclusion that it was impossible to design any experiment using these stimuli which would answer any useful questions. The possibility remained that an experiment could be designed to test whether the subject was more sensitive to changes of this kind in one area than in another. The overall effect on the attack of the second note seemed to be the same as the effect on the decay of the first note, so the next study was limited to just the attack of the second note. Five groups of three amplitude envelopes each were generated (see Figure 5.16). The timing of the center envelope in each group was the same as that

used earlier; the ends of the other two envelopes were shifted by ±10 msec from that center time. This resulted in envelopes for which point 4 of Figure 5.12 varied from its original value by

A. −10 msec
B. 0 (i.e., no change)
C. +10 msec
D. 40
E. 50
F. 60
G. 90
H. 100
J. 110
K. 190
L. 200
M. 210
N. 290
P. 300
Q. 310

(The letters correspond to those in Figure 5.16). Informal listening tests with the clarinet stimuli quickly showed that it was impossible to distinguish the three members of a group.

Furthermore, this line of work ran into an unexpected stumbling block with the other instruments. Figure 5.16b shows a set of amplitude envelopes for the tongued trumpet transition, using the same times just given. (For clarity, lines B, D, E, and F are omitted.) Unfortunately, the envelope of the original attack is shaped such that it "undercuts" the envelopes A through J. Even worse is the behavior of lines G, H, and J themselves: The amplitude of the original varies so much that the line ending at H actually has a flatter slope than the line ending at J. This also happens with K, L, and M. The situation with the violin is even worse. As shown in Figure 5.16c, some of the lines in the attack of the second note lie on top of each other, due to the vagaries of the attack.

This line of inquiry was therefore abandoned. The question still remained as to whether flattening the slope of the *tongued* transition's attack would lead to the percept of an *untongued* transition. The rest of the discussion of Experiment 6 will focus on answering that question.

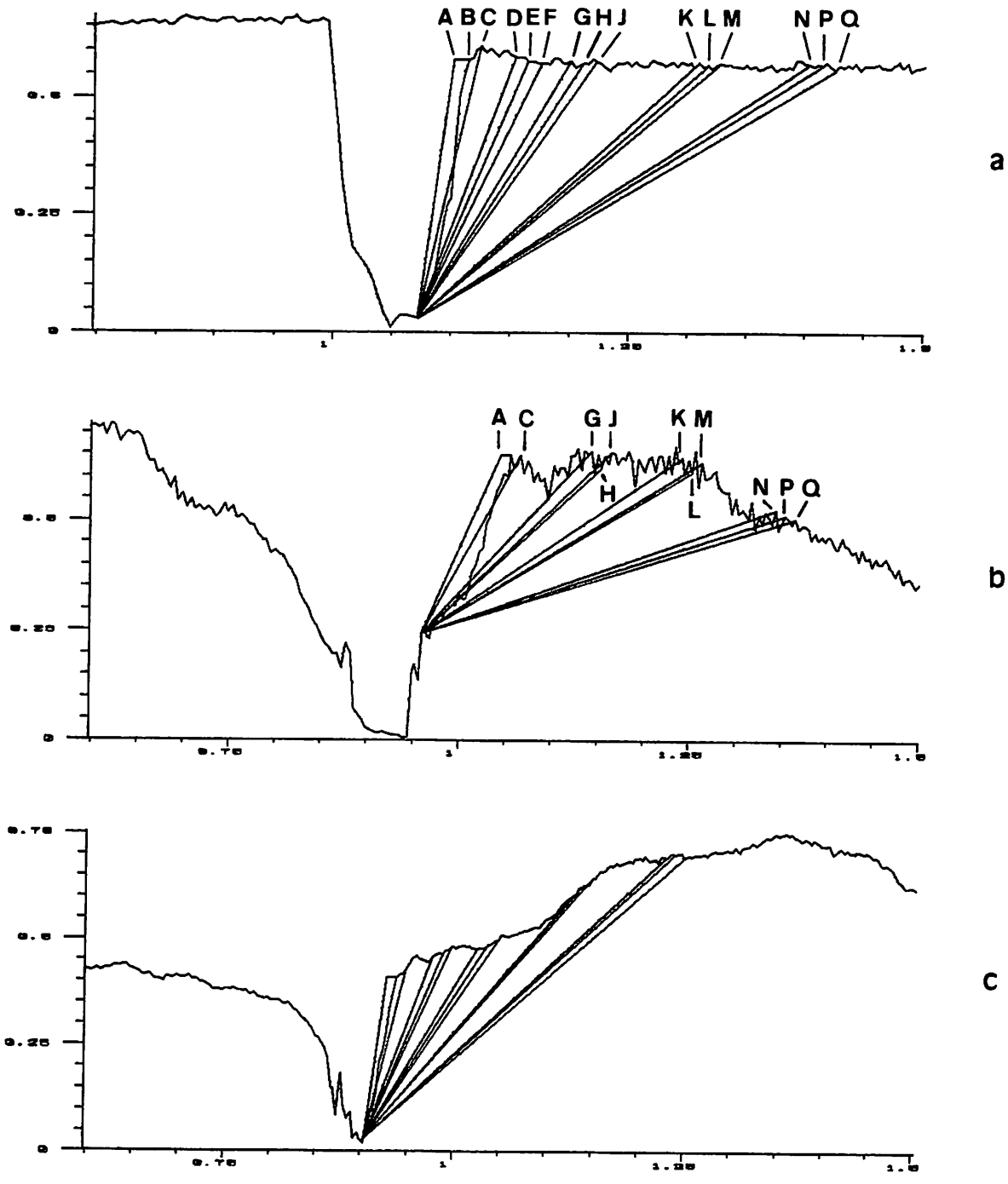


Figure 5.16. Amplitude envelopes for modifying various transitions for Experiment 6. a) clarinet. b) trumpet. c) violin.

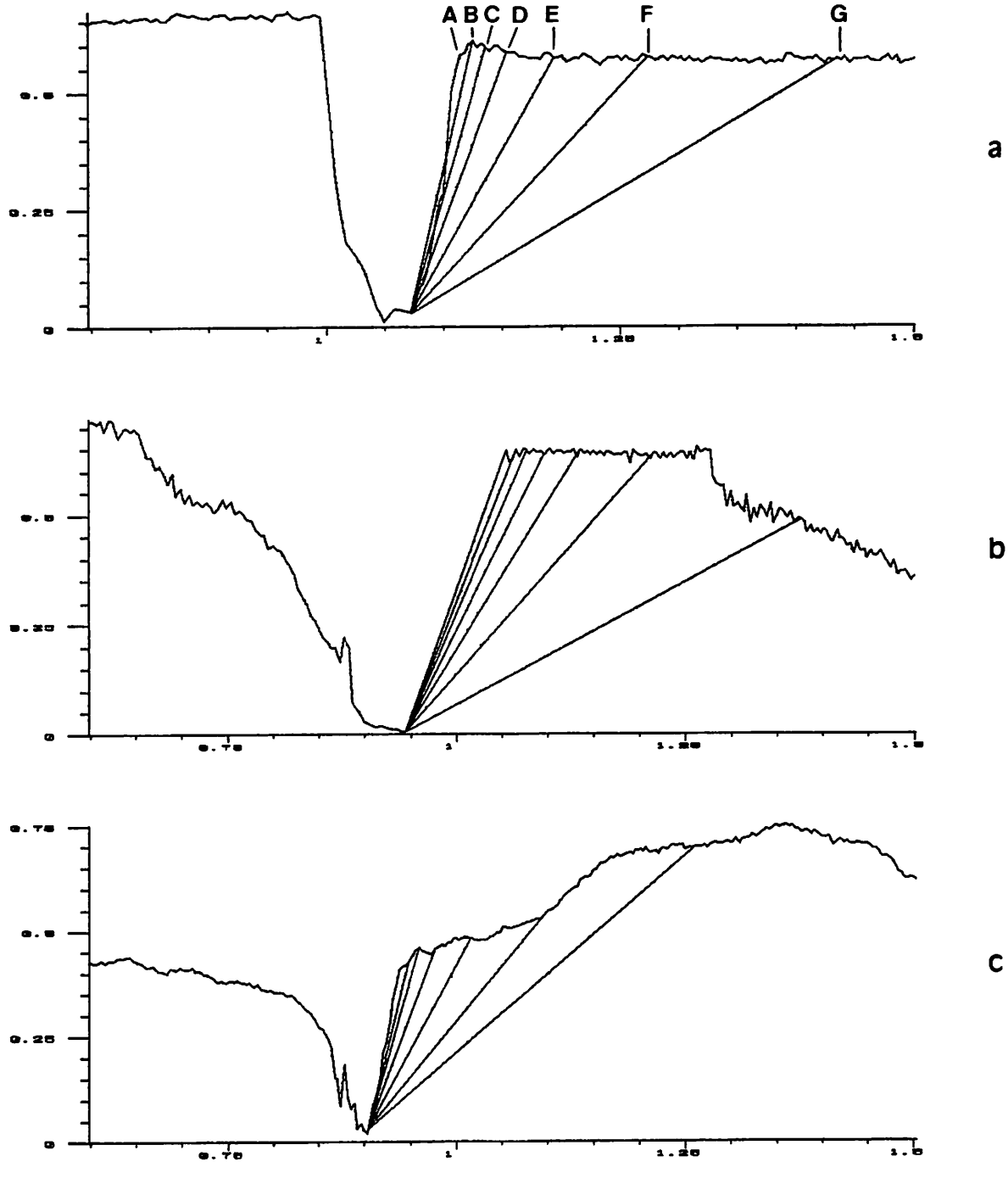


Figure 5.17. Amplitude envelopes actually used for creating transitions in Experiment 6. a) clarinet. b) trumpet. c) violin.

Preparing the Stimuli

The amplitude envelopes shown in Figure 5.17 were applied to the tongued transitions of the three instruments. The decay at the end of the first note was not varied. The end of the attack of the second note varied from that of the original by the following amounts:

A. 0 msec (i.e., the original recording unmodified)
B. +10 msec
C. 20
D. 40
E. 80
F. 160
G. 320

There were thus seven test stimuli for each of three instruments.

Experimental Procedure

The procedure for this experiment varied significantly from that for the other experiments. Therefore, the experimental procedure will be covered completely in this section (rather than in Appendix 3, as for the other experiments).

Seven test subjects participated in this experiment: 1, 2, 4, 5, 8, 9, and 10. This number of subjects was judged, and proved, to be adequate. The test subjects heard the test tones in one of the listening stations at CCRMA; all test subjects heard these tones in the same room (this was not the room used for the other experiments). The test subjects heard this experiment before taking the other experiments. For a given instrument, I played the seven tones in the series given above, A through G; the stimuli were played directly from computer disk through the A/D converter in the Samson box, connected to the speaker in the listening room. The subjects could hear the series for each instrument as often as they desired. The order of instruments was varied with each new subject, so six possible orders of three instruments were presented.

The subject was instructed to describe in his own words, after hearing the series, what if anything changed across the seven stimuli. As he spoke, I typed his comments into a file in the computer. If the subjects noted that the percept changed from tongued to untongued, we could reasonably expect them to remark on this in their comments. This was, in a word, a "sneaky" way of discovering the subject's response without asking leading questions.

Table 5.13. Order of Presentation of Instruments in Experiment 6.

Subject	First	Second	Third
1	clarinet	trumpet	violin
2	clarinet	violin	trumpet
4	clarinet	trumpet	violin
5	violin	trumpet	clarinet
8	trumpet	violin	clarinet
9	trumpet	clarinet	violin
10	violin	clarinet	trumpet

Results

The data consisted of the subjects' responses, which will be summarized here (and quoted verbatim where appropriate) not on a subject-by-subject basis but rather grouped according to content.

Order effects: In general, the subjects' responses depended to a certain extent on the order in which the instruments were presented; the order of presentation is given in Table 5.13. Except for subject 4 (more about him below), the subjects seemed to understand most quickly what was happening to the notes when the clarinet was played first; the effect was also noticeable for the trumpet, but hearing the violin first seemed to leave subjects uncertain of what they were hearing.

Change from tongued to untongued: A few subjects did suggest a change from tongued to untongued, but always with reservations and qualifications. Moreover, no one subject heard a change from tongued to untongued for all of the instruments; so the conclusion that such a change is generally heard cannot be supported by this data, QED.

Subject 1 remarked about the trumpet:

> [on first hearing the trumpet series, after having heard the clarinet]: Again, we have the same thing. I hear a little tonguing glitch at the begin of the second note—then it drops amplitude and come back up [across the seven stimuli]. It's not quite as thorough a job as on the clarinet of cutting the amplitude of the attack—it misses some microseconds.
>
> [after hearing the trumpet series again]: It might be just a valve being pushed down [that I hear at the begin of the second note]. It doesn't really sound like a tongued note, but there is some noise in there. I don't think they're tongued. There is a noise there, but the second note is probably not tongued.

On first hearing this series, subject 10 said that he "didn't hear anything"; after hearing the series again, he reported: "... the second note is trying to be more legato, as though it doesn't come as strongly or as quickly as in the beginning [stimuli]—sort of making [the second note] more legato. But [the player] didn't do a very good job of it."

As for the violin, subject 2 (a string player) reported:

> It sounds like there are two different bows in the first [stimuli]. By the end [of the series], it sounds almost like it's fingered [...] with two dots and a slur. There is still a little bit of a gap between the notes. It sounds like it's one bow rather than two bows.
>
> By "slur and two dots", I mean that for repeated notes, [you play] both notes [with the bow moving] in the same direction, but stopping the bow in between subtly. This is usually used only for the same note on the same string. That kind of bowing is usually not used for a new pitch.*
>
> The other way to hear that is with a bow change, but the person starts very slowly on the up bow [for the second note]; as though coming very gently into the up bow. The first [stimuli] sound like a definite down bow/up bow. At the end of the series, if there is a bow change, then there's almost a crescendo on the second note.
>
> In the first [stimuli], it's more possible for this [to have been played on] separate strings than in the later [stimuli].
>
> The same thing is happening in the clarinet: The last [stimulus] sounds synthetic because it's been changed so much.

Subject 9, also a string player, made similar remarks:

> The amplitude, i.e., the bow velocity, is being reduced at the beginning of the second note. By the last [stimulus], there is an attempt at legato, but we also get a dip in amplitude. It gives the impression of a legato with an accent on the second note. So the point of transition between the notes equals the point of lowest bow velocity.

Subject 10 heard similar effects for the violin:

---

*This notation is discussed, for example, in Burton (1982), pp. 31–32, and Blatter 1980, p. 71. See also the discussion of "getrennte Noten in einem Bogenstrich" in Humperdinck (1892), p. 26.

> "I could even say that the first [stimulus] seemed like a change of bow, but the later ones not. So there is some energy that stops in the first [stimuli], but the later ones seem more connected."

This same subject was the only one to comment along these lines for the clarinet:

> The first [stimulus] is very clear—two separate notes, two breaths. Later, it's more legato. It's untongued in the later [stimuli], at least. The later [stimuli] sound more artificial.

Thus, we see that none of these comments is a ringing indictment for hearing a change from tongued to untongued across the series of stimuli; and these were all of the comments in which this possibility is even mentioned.

Indeed, some subjects were confused about what was happening in the stimuli. Subject 4, mentioned above, heard the clarinet series first, and had this to say:

> The decay time is increasing on the first note [incorrect], and the attack time is increasing on the second note [correct].
>
> The first note became longer than the second note [incorrect]; in the beginning [of the series], the two notes were equal in duration [correct].
>
> [after hearing the series again] At the end, the duration of the first note is longer than for the second note. [incorrect]

For the trumpet, this subject first reported:

> The attack time on the second note is steadily increased. [correct]

but after he heard the same series again, he felt that

> it sounds like no change occurs, except at the end of the series, where the attack time of the second note becomes shorter. [incorrect]

Similar problems occured with the violin for this subject.

Subject 5, a string player, was likewise confused. In order to avoid belaboring the point, I will mention simply that in the violin, he did not "hear much changing; all sound pretty realistic; all sound very much alike;" and he felt that in the trumpet series, "it seems as though the two notes are getting closer together."

Subject 8 felt that the clarinet series went "in the opposite direction as in the trumpet." In fact, for the trumpet he said that the two notes "become more separated, less legato," contradicting the notion that the tongued transition might become untongued.

Naturalness: Another theme running through the comments dealt with the artificial impression evoked by the stimuli at the end of each series. Subject 10, for example, said that in the violin "there is something wrong with the last one". Subject 9 felt that the trumpet "gets electronic at the end ... it's been smoothed in some way that's not natural." With the clarinet, listen 2 said that "... by the last [stimulus], it sounds synthetic."

The overall impression created by these series of stimuli is best summarized by subject 2, who remarked, speaking of the trumpet:

> None of these sounds like a legato. The same is true in the clarinet and the violin [which he had heard earlier]. It sounds like it's trying to get to a legato, but it gets synthetic instead of getting to a true legato.

Conclusions

It is possible to change the slopes of the amplitude envelopes surrounding a transition over a fairly wide range. Steep slopes, even steeper than those found in nature, do not seem to be troublesome. As for flatter slopes (extrapolating from the comments of the subjects), an attack time in the range of 10 to 100 msec or so seems to be acceptable; longer attack times (and by inference, longer decay times) seem artificial, and can even confuse the subject's impression of the articulation. Most importantly, the slope is not a clear-cut cue for the kind of articulation used. This topic will be dealt with further in Chapter 6.

Experiment 7: Swapping Amplitude Envelopes

Background

This experiment is designed to shed further light on the relative importance of spectral vs. amplitude cues in allowing the subject to identify playing styles. It also examines the role of specific amplitude envelopes; that is, if the amplitude envelopes for tongued and untongued notes are swapped, does the percept swap as well?

As mentioned in the introduction to this chapter, it would be useful to be able to isolate specific physical cues, and to vary them independently. It seems reasonable to assume that musically trained subjects can reliably distinguish between tongued and untongued transitions (this experiment, along with Experiment 8, addresses that question). If the cues for those specific transitions can be isolated, perhaps they can be varied and analyzed experimentally. Of the four parameters of a transition listed in Chapter 2, pitch (Experiments 1 and 2) and timing (pp. 116–124) have been discussed.

Isolating spectral cues proved to be impossible. Careful examination of a number of transitions (such as in Figures 2.3–2.8) showed that it was often easy to *identify* a set of a dozen or so periods at the beginning of the second notes, which seemed to include major spectral cues for the transition. On the string instruments, for the transition with no bow change, there was a distinct "thwack" as the left-hand finger hit the string to shorten it; and one can see in the plots where this happens. Or, for the bow change, the waveform turned "noisy" for a little while. Indeed, if these few periods are spliced out, then the transition really sounds quite different, even though the changes in amplitude and timing are imperceptible.

However, it proved impossible to work with these short signals. Merely swapping a few periods between the two transitions produced unusable results. When the "thwack" of the no-bow-change transition replaced the "bow change" cues, the "thwack" was artificially amplified. Also, it proved impossible to create the necessary faultless seam when working on this small time scale. Merely splicing a few periods from one recording into another of course produced pops at either end of the splice. Cross-fading from the one signal to the other and back again was unsuccessful with cross-fade times on the order of a few milliseconds, because the abrupt change in phase was audible, again at both ends of the splice. Use of longer cross-fade times obliterated the very spectral signals which were being spliced in. These problems occurred for all of the instruments being examined.

The only avenue left was to isolate the amplitude envelopes, while keeping pitch change, timing, and spectral cues unchanged.

Preparing the Stimuli

Recall from Chapter 2 that the recordings were not equalized for duration, loudness, and the like. For this experiment, it was necessary to align the points of pitch change. Figure 5.18 shows the alignment of the original recordings. One of the two recordings from each instrument was shifted to produce the control recordings shown in Figure 5.19; the point of pitch change matched on both the untongued and tongued cases. Two control recordings (called T and U in the following discussion) for each instrument resulted in a total of six control recordings.

Two test stimuli, called TU and UT, were also created for each instrument. For the UT stimuli, the amplitude envelope of the *untongued* transition was modified between points A and B in Figure 5.19 to match the amplitude envelope of the *tongued* transition. The stimulus TU was created by scaling the tongued envelope to match the untongued envelope. (Methods for amplitude scaling are discussed in Appendix 1). The beginning and ending of the scaling were of course imperceptible. This resulted in a set of six test stimuli.

Experimental Procedure

As training tones, each of the control and test stimuli was presented once in a randomized order. For the actual experiment, each of the control and test stimuli was presented three times. (Details on presentation of the stimuli are given in Appendix 3). The subject was asked to rate each stimulus as "tongued" or "untongued."

For the most part, the TU stimuli for all three instruments sounded unnatural, since the major change was the amplification of the tonguing noise (or the bow change noise). In their written comments, some of the subjects commented on the poor quality of these transitions. For example, subject 5 wrote:

> This [experiment] includes examples of unrealistic "overtonguing" for which no response category was provided. Unfair! These notes are *not* "tongued," but neither are they "untongued."

Likewise, subject 7 felt that

> The "bloopy" clarinets sounded neither tongued nor untongued.

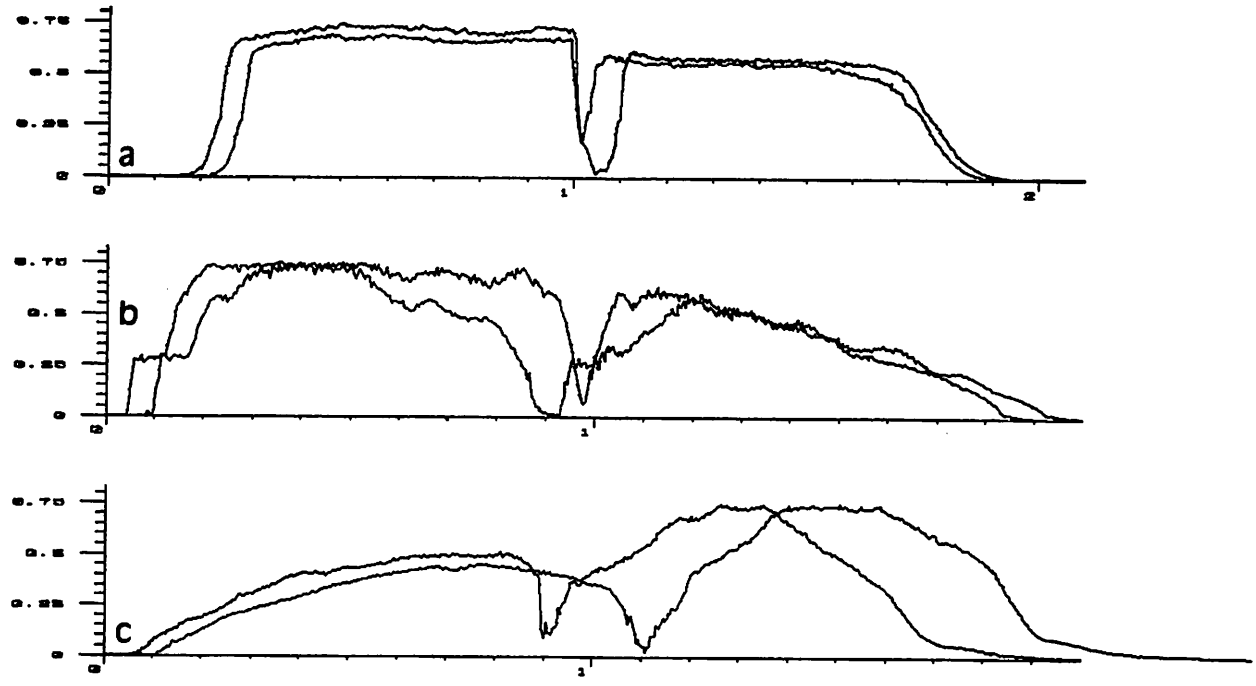


Figure 5.18. Amplitude envelopes of stimuli for Experiment 7. a) Original clarinet tongued and untongued recordings. b) Original trumpet tongued and untongued recordings. c) Original violin recordings, with and without bow change.

These sorts of remarks were expected. In fact, in designing this experiment, I considered omitting the TU stimuli altogether. Still, it seemed best to let the subjects make their own judgments about whether the transition was tongued or not.

Results

Table 5.14 lists the "raw data": of the possible three responses for each stimulus and instrument, how often each subject identified the stimulus as "tongued".

The first question to be answered was whether the subjects could correctly identify the originals as "tongued" and "untongued." The data show that this was not always the case. For the clarinet, only subject 7 performed exactly as one would hope. The answers for subjects 3, 5, 6, and 10 also seem more or less reasonable. Subjects 1, 4, and 9 were clearly confused by what they heard. The situation is not so bad with the trumpet, except for subjects 1, 3, 5, and 6. Except for subject 9 (himself a string player), the answers for the violin follow the desired pattern. In

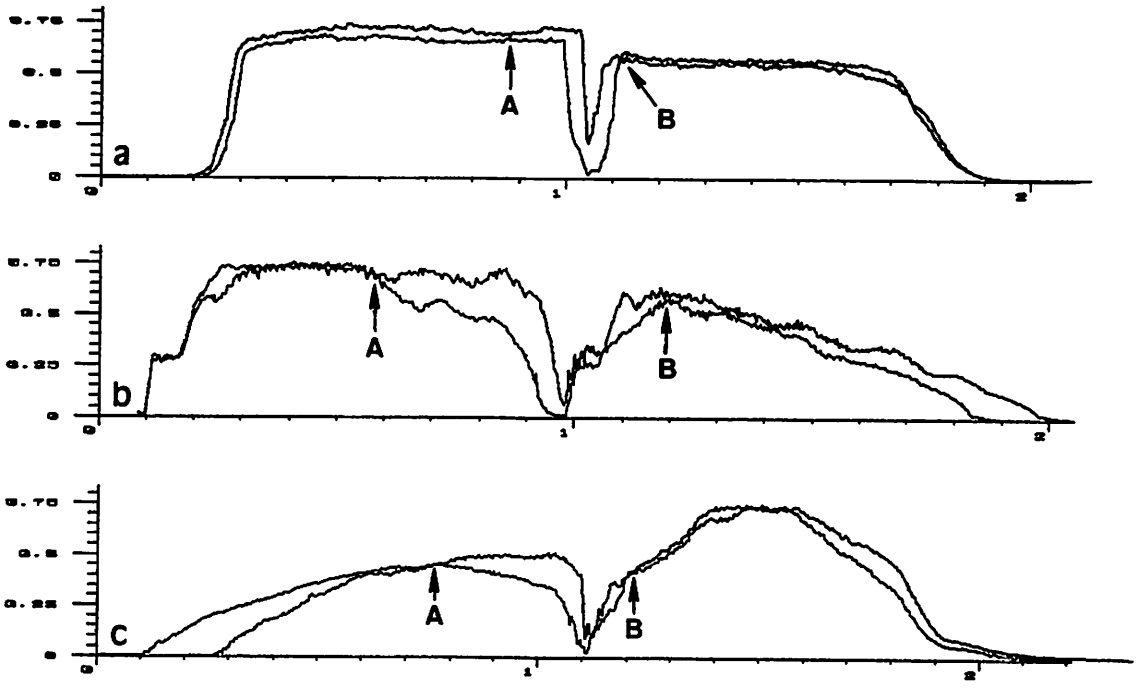


Figure 5.19. Amplitude envelopes of stimuli for Experiment 7. a) The clarinet recordings of Figure 5.18a, with the untongued stimulus shifted by 812 samples (= 32 msec). b) The trumpet recordings of Figure 5.18b, with the tongued stimulus shifted by 1382 samples (=54 msec). c) The violin recordings of Figure 5.18c; the no-bow-change case has been shifted by 5365 samples (=210 msec).

experiments of this kind, it would thus appear that the subjects need to be more carefully trained. Indeed, subject 4 wrote in his comments:

> Because I don't play wind instruments I am unaware of more subtle techniques of producing untongued/tongued attacks. I think you might derive better results if you gave the answers to the "training tone" examples. This would help as a point of reference (particularly for people, like me, who lack trumpet, clarinet, and violin playing experience).

Still, the data from this experiment produced useable results, as the following discussion will show.

For numerical analysis of data, the value of "0" was assigned to each "untongued" response, with "1" meaning "tongued." The means and standard deviations of the subjects' responses are given in Table 5.15. This same data is shown as a bar graph in Figure 5.20. The mean values for the T and U stimuli fall into the pattern one would hope for, in spite of the problems mentioned in discussing the raw data.

Table 5.14. Number of "Tongued" Responses (Experiment 7)

	Subject Number											
Stimulus	1	2	3	4	5	6	7	8	9	10	Row total	
Clarinet												
T	1	1	2	2	2	3	3	1	1	2	18	(30)
U	3	0	1	2	1	1	0	0	1	0	9	(30)
UT	0	0	1	0	0	3	3	0	0	0	7	(30)
TU	3	3	3	2*	3	3	2	3	3	3	28	(29)
Trumpet												
T	1	3	3	3	3	3	3	3	3	3	28	(30)
U	3	0	3	0	2	3	0	0	1	0	12	(30)
UT	0	3	2	2	0	3	3	2	1	1	17	(30)
TU	3	3	3	3	3	3	1	3	3	3	28	(30)
Violin												
T	3	3	3	3	3	3	3	2	3	1	27	(30)
U	0	0	1	0	0	0	0	1	2	0	4	(30)
UT	2	1	2	1	2	1	1	3	3	3	19	(30)
TU	3	3	3	1	3	3	3	3	3	3	28	(30)

Note: Highest possible score for each subject and stimulus is 3. Numbers in parentheses show total number of responses collected.

*Only 2 responses collected

Table 5.15. Analysis of Experiment 7.

	Clarinet		Trumpet		Violin	
Stimulus	Mean	s.d.	Mean	s.d.	Mean	s.d.
T	0.60	0.49	0.93	0.25	0.90	0.30
U	0.30	0.46	0.40	0.49	0.13	0.34
UT	0.23	0.42	0.57	0.50	0.63	0.48
TU	0.97	0.18	0.93	0.25	0.93	0.25

Note: A mean value of 1.0 shows that the subjects responded "tongued" for all presentations of the stimulus; a mean of 0.0 shows that all subjects responded with "untongued."

As for the others, all three TU stimuli were labelled as tongued, as expected. I believe that this happened because, as stated above, the tonguing noise was amplified in all three instruments.

For the UT stimuli, the trumpet and violin showed a shift "toward" tongued, as expected. The clarinet, on the other hand, was rated even *less* tongued than the original untongued stimulus.

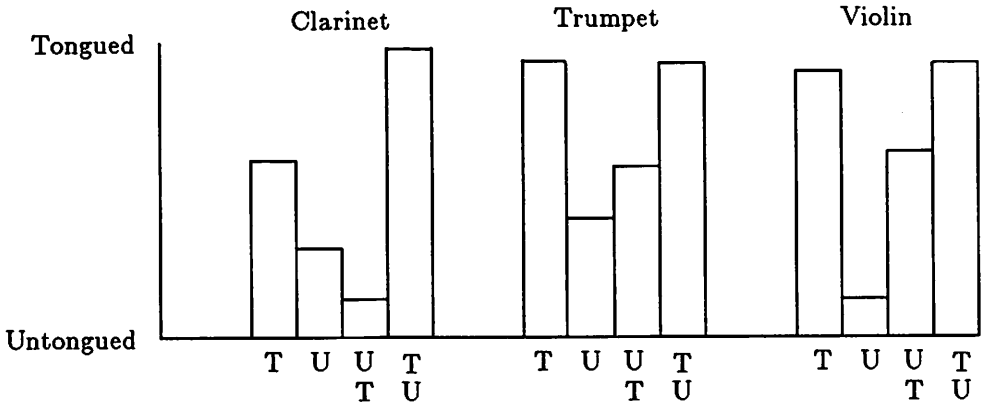


Figure 5.20. Bar graph representation of the data in Table 5.15.

Table 5.16. Analysis of Variance for Experiment 7.

Source	*SS*	*df*	*MS*	*F*
Modification (M)	24.88	3	8.29	55.27
Instrument (I)	2.18	2	1.09	7.27
Modification × Instrument (MI)	3.70	6	0.62	4.13
Error (M × I × MI)	53.23	348	0.15	
Totals	83.99	359		

The reason for this perplexing behavior of the data was found by listening to the U and UT stimuli again. In the U clarinet stimulus, there was a noticeable spectral cue which made it easy to distinguish from the T clarinet, and which I hadn't noticed as I was making the test stimuli. It is therefore not surprising that the UT stimulus for the clarinet would remain in the vicinity of the original U stimulus.

The question remains whether these variations in the mean responses were significant. To test this, the by now familiar two-way analysis of variance (Hays 1963, p. 402) was applied; the results are given in Table 5.16. All three F values imply $p < 0.1\%$. Clearly, most of the variance is due to the modifications made to the envelopes of the original tones. The large F value in the first line of the table suggests that the change from T to TU, or from U to UT, is indeed statistically significant. The F value for the "Instrument" term in the table shows that each instrument reacts to the modifications to a different degree. As for the F value in the third line, it seems reasonable to conclude that the instruments react to the modifications in a different way; which is not surprising, especially given the behavior of the clarinet just discussed.

Table 5.17. Analysis of Experiment 7, Omitting certain Subjects.

	Clarinet		Trumpet		Violin	
Stimulus	Mean	s.d.	Mean	s.d.	Mean	s.d.
T	0.80	0.40	1.00	0.00	0.96	0.20
U	0.20	0.40	0.06	0.23	0.08	0.28
UT	0.47	0.50	0.67	0.47	0.54	0.50
TU	0.93	0.25	0.89	0.31	0.92	0.28

Note: Only the following subjects were included:
Clarinet: 3, 5, 6, 7, 10
Trumpet: 2, 4, 7, 8, 9, 10
Violin: 1–8

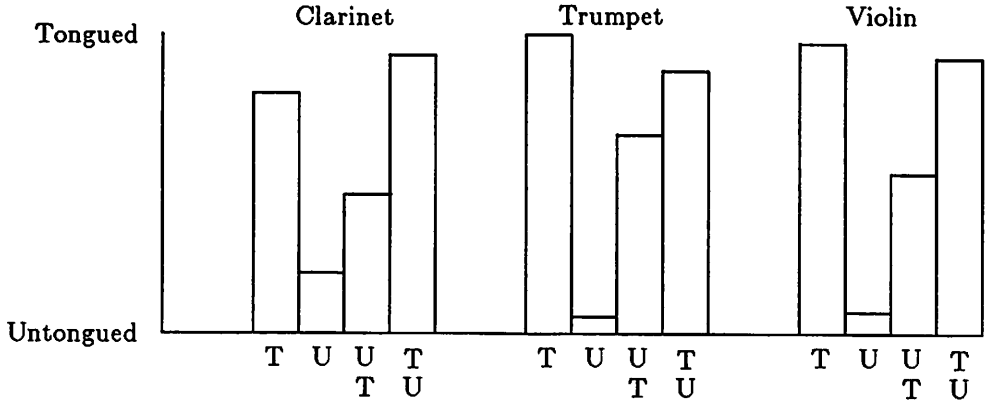


Figure 5.21. Bar graph representation of the data in Table 5.17.

As stated before, some of the subjects were confused by the original T and U stimuli; therefore, their responses to the TU and UT stimuli might be questionable. Table 5.17 shows the means and standard deviations for the responses when these subjects are removed; Figure 5.21 shows the corresponding bar graph. The overall shape of the graphs remains unchanged, except that the puzzling behavior of the clarinet between the U and UT cases has disappeared. For all three instruments, the distinction between T and U increases dramatically from Table 5.15 and Figure 5.20, on the one hand, and Table 5.17 and Figure 5.21 on the other. The behavior of the TU case remains unchanged.

Conclusion

When the amplitude envelope of an untongued transition is changed to that of a tongued transition, the percept tends to change as well. The opposite transformation—changing the tongued envelope to the untongued—produces a transition which can be difficult to classify either as tongued or untongued. Still, it seems safe to conclude that the shape of the amplitude envelope in the transition, and the dip in amplitude between notes, influences what is heard by the subject. This agrees with the conclusion drawn on pp. 121–124. It is thus clear that spectral cues are not the sole determinant of the kind of articulation perceived by the subject.

Overall Conclusions

In creating a transition, one must pay attention to the shape of the amplitude envelope connecting the two notes. It is possible to create an acceptable legato transition with a quick cross-fade between two notes, with no dip in amplitude. Other kinds of transitions require an amplitude dip of some sort. If, in such cases, a mere overlapped transition is used as a starting point, a wide range of overlap times is available (20–100 msec or so).

Subjects prefer transitions in which some spectral cues are present. Thus, spectral cues are at least as important as the amplitude dip in creating the percept of a natural transition. It is impossible to adjust the amplitude of a transition to completely compensate for missing spectral cues.

If both spectral and amplitude cues are present, a wide range of acceptable transition times and amplitude dips is available. In general, a dip of at least 10 dB or so is recommended, relative to the maxima of the surrounding notes. Raising the transition amplitude above the levels found in nature is not recommended.

Each instrument seems to react in its own way to changes in the transition.

CHAPTER 6

CATEGORICAL PERCEPTION

Historical Introduction

The issue of categorical perception lies at the center of a long-standing controversy in psychophysics. (Macmillan, Kaplan, and Creelman [1977] provide a good overview.) One of the themes running through this research is the question of how many stimuli can be discriminated from each other, as opposed to how many stimuli can be assigned a specific *label*. In "continuous" perception, "the process of discrimination is independent of the process of identification" (Studdert-Kennedy et al., 1970, p. 236). In "categorical" perception, on the other hand, stimuli can be assigned to only one of the available perceptual categories. It can be shown that speech sounds, such as pairs of consonants like "t" and "d", are perceived categorically. This tendency of the speech system is explained by a "motor theory" of perception, in which the procedure for forming the sound, itself categorical, molds the manner in which they are perceived. This in turn has led to the notion of "speech" vs. "nonspeech" modes of perception (Mattingly et al. 1971; Schouten 1980). Music is an obvious candidate for research into non-speech auditory perception, and has obligingly served as a testing ground for various experiments on categorical perception. This historical introduction will trace in particular the development of research on categorical perception of attack time.

It all started, it seems, with Cutting and Rosner (1974). They used sawtooth waveforms at frequencies of 294 and 440 Hz from a Moog synthesizer to create test stimuli. In particular, amplitude envelopes with a rise time of 0, 10, 20, 30, 40, 50, 60, 70, and 80 msec were applied to the Moog sawtooth waveforms.

Cutting and Rosner assert: "The rapid-onset stimuli sounded like the plucking of a stringed instrument, whereas the slower onset stimuli sounded like the playing of the same instrument with a bow." This is of course preposterous; but this fallacy in Cutting and Rosner's assumption seems to have escaped the other authors who have analyzed Cutting and Rosner's work.

The subjects were 20 Yale undergraduates not selected according to musical ability. (In a study involving the categorical perception of the middle note of a chord, Locke and Kellar [1973] found that musicians were more likely to show categorical perception than nonmusicians. It is thus perhaps a more rigorous test to use a mixed group.) In an identification test as well as a discrimination test, Cutting and Rosner found categorical perception of attack time, with the categorical boundary lying at a rise time of about 40 msec.

In a follow-up study, Cutting, Rosner, and Foard (1976) used sawtooth waves from the same Moog, this time only at a frequency of 294 Hz. The attack times were the same as before. However, the decay of each stimulus was clipped off at 250 msec; as they say, "every stimulus had an abrupt offset" (p. 363). Using these stimuli, they failed to find categorical perception. However, with longer "offset" times (750 msec), categorical perception did reappear.

Jusczyk et al. (1977) investigated auditory perception in two-month old (!) infants. Attack times of 1, 30, 60, or 90 msec were applied to a 440 Hz sawtooth waveform from the by-now familiar Moog; each stimulus lasted about 1 sec. As the tones were played, the sucking rate of the infant was measured, from which the discriminability of stimulus pairs can be deduced. These researchers claimed to find categorical perception of rise time in two-month-old infants.

Remez, Cutting, and Studdert-Kennedy (1980) investigated adaption of the categorical boundary, and concluded that both speech and non-speech sounds are processed by the same feature detectors. Although their work is not germane to this chapter, I must object to their characterization of Cutting and Rosner's tones as "synthetic violin sounds" (p. 524). It is further unfair for them to characterize their stimuli as forming "a synthetic stimulus continuum of violin sounds ranging from plucked string to bowed string." In discussing the "categorical perception of violin 'articulation'", they failed to appreciate the distinction between methods of sound production on the one hand and the percepts invoked on the other, as has so carefully been laid out in Chapter 1.

Rosen and Howell (1981) conducted some insightful experiments which also cast a new light upon the work of Cutting and Rosner. As in earlier studies, Rosen and Howell formed nine stimuli by applying envelopes with attack times from 0 to 80 msec in 10-msec steps to a low-passed 312 Hz square wave; the stimuli were generated digitally. They found that discrimination works best for stimuli with the shortest rise times, and discrimination between adjacent stimuli decreases

monotonically with increasing rise time. In a word, they failed to find categorical perception with their tones. Cutting was kind enough to supply the test tapes used in the original Cutting and Rosner study (1974). Measurements by Rosen and Howell of the rise times of Cutting and Rosner's stimuli showed that the stimuli of the latter did not have the desired rise times. Rosen and Howell decided that the categorical perception found by Cutting and Rosner was due to a nonlinearlity in the sequence of rise times in the tones used by Cutting and Rosner.

Finally, Kewley-Port and Pisoni (1984) used digitally generated sawtooth stimuli at 300 Hz, with rise times of 1, 10, 20, 30, 40, 50, 60, 70, and 80 msec. They trained their subjects to use the same "plucked" and "bowed" labels as in Cutting and Rosner (1974). Kewley-Port and Pisoni failed to find categorical perception with their stimuli.

Meanwhile, other aspects of music have been used to test for categorical perception. Locke and Kellar's work has already been mentioned. In his classic study on timbre, Grey (1975) investigated the categorical perception of interpolated timbres. Grey's original recordings were equalized for pitch, duration, and loudness. The reference tones in his categorical perception experiment (clarinet, horn, oboe, and cello) were created from a constant-frequency approximation (Grey 1975, p. 77) to the time-varying Fourier analysis (heterodyne filter) of the original tones. Test stimuli were created by interpolating between these reference tones. Grey failed to find categorical perception for tones interpolated between two instruments.

Experiment 8: Categorical Perception of Transitions

Background

This experiment examines whether transitions between notes are perceived categorically. The centuries-old tradition of distinguishing various kinds of playing styles, already discussed in Chapter 1, provides a solid framework for picking stimuli which are commonly accepted to be distinguishable, and which have convenient labels. These kinds of articulation can easily serve as endpoints between which test stimuli can be interpolated.

Besides, even a two-note melodic snippet matches more closely what happens in connected speech than does the attack of an isolated note. Furthermore, I believe that perceptual studies run the risk of reductionism if they limit themselves to studies with artificial tones (more on this can be found in [Strawn 1982]).

It is still important to remember, as discussed in Chapter 1, that a given percept can be evoked by a number of different playing styles. This lack of a simple correspondence will be made clear in the remarks of the subjects quoted below. For the purposes of this experiment, however, it is reasonable to assume that the tongued and untongued playing styles result in readily distinguishable percepts. The results of Experiment 7 showed that, with certain exceptions, the subjects could readily distinguish the two.

Creating the Stimuli

As in Experiment 7, the tongued and untongued recordings for each instrument were aligned so that the transition occured at the same time in both recordings (see Figure 5.19). For each instrument, new amplitude envelopes were created by interpolating between the tongued and untongued amplitude envelopes, using seven equal linear steps.

These steps were calculated by interpolating "vertically" between the original tongued and untongued envelopes for the clarinet. Figure 6.1 shows the resulting set of amplitude envelopes. The untongued original is the top trace, and the tongued envelope is at the bottom of the set of traces. The vagaries in the interpolated envelopes can be seen more easily in Figure 6.2, which shows the envelopes on a dB scale. The "waves" in the interpolated envelopes derive from local peculiarities in the original envelopes, such as shown at points C and D in the figure. I considered attempting to interpolate along some other line; one possibility might be to construct a series of time-varying normals to each original curve, and to try to interpolate along some sort of averaged normals. The ensuing headache was sufficient warning against such an undertaking.

For each instrument, the amplitude of the *untongued* original was scaled by the middle seven envelopes to form seven stimuli with interpolated amplitude envelopes. The two original recordings completed the set, for a total of nine stimuli for each instrument.

For the clarinet shown in these figures, the attack of the first note and the decay of the second note matched quite nicely in both recordings. In fact, inspection showed that the principal differences in the transitions occurred between points A and B in the figures. This is easier to see in the detail of the transition only (Figure 6.2). In fact, for this instrument it proved reasonable to change the beginning time for scaling, depending on where the interpolated envelope approached or touched the original untongued envelope; two possibilites are shown in Figure 6.3a and 6.3e. In short, the untongued clarinet recording was modified by one of the interpolated waveforms, with scaling starting between $t = 0.8495$ sec and $t = 0.8780$, and ending at $t = 1.1165$.

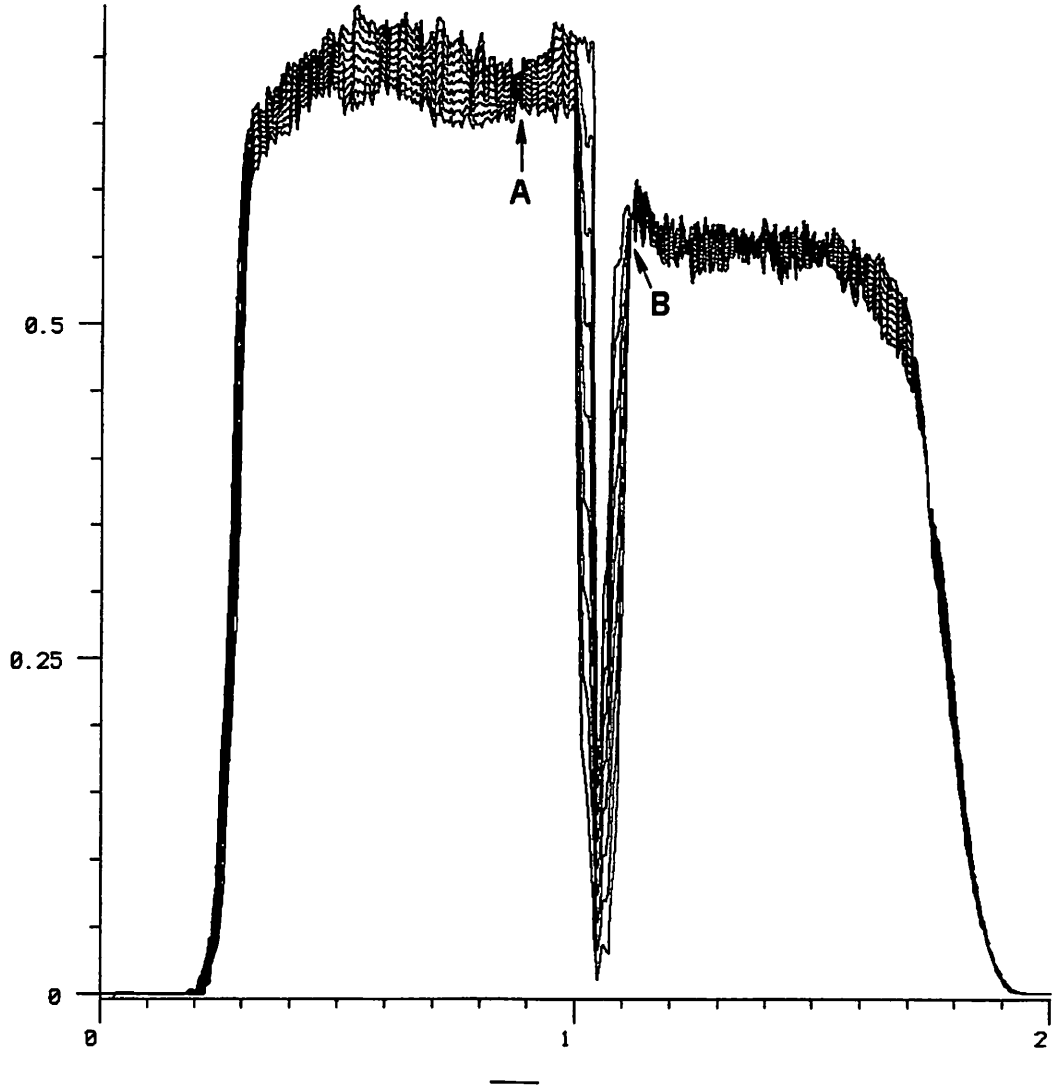


Figure 6.1. Clarinet tongued and untongued envelopes, with seven amplitude envelopes interpolated between them.

Figures 6.4 and 6.5 show the corresponding sets of envelopes for the trumpet and violin, respectively. In general, the stimuli for these two instruments were created in exactly the same manner as for the clarinet, with scaling limited to the region between points A and B shown in each figure. It was necessary to scale the amplitude of the decay of the second tongued trumpet note to make it match the decay of the untongued note. As shown in Figure 5.19b, those two decays were quite dissimilar, and the difference was definitely audible.

The attack and decay times for the trumpet transition varied over a range of about 50 msec. In the violin and trumpet, the decay of the first note varied over a range of about 300 msec, but

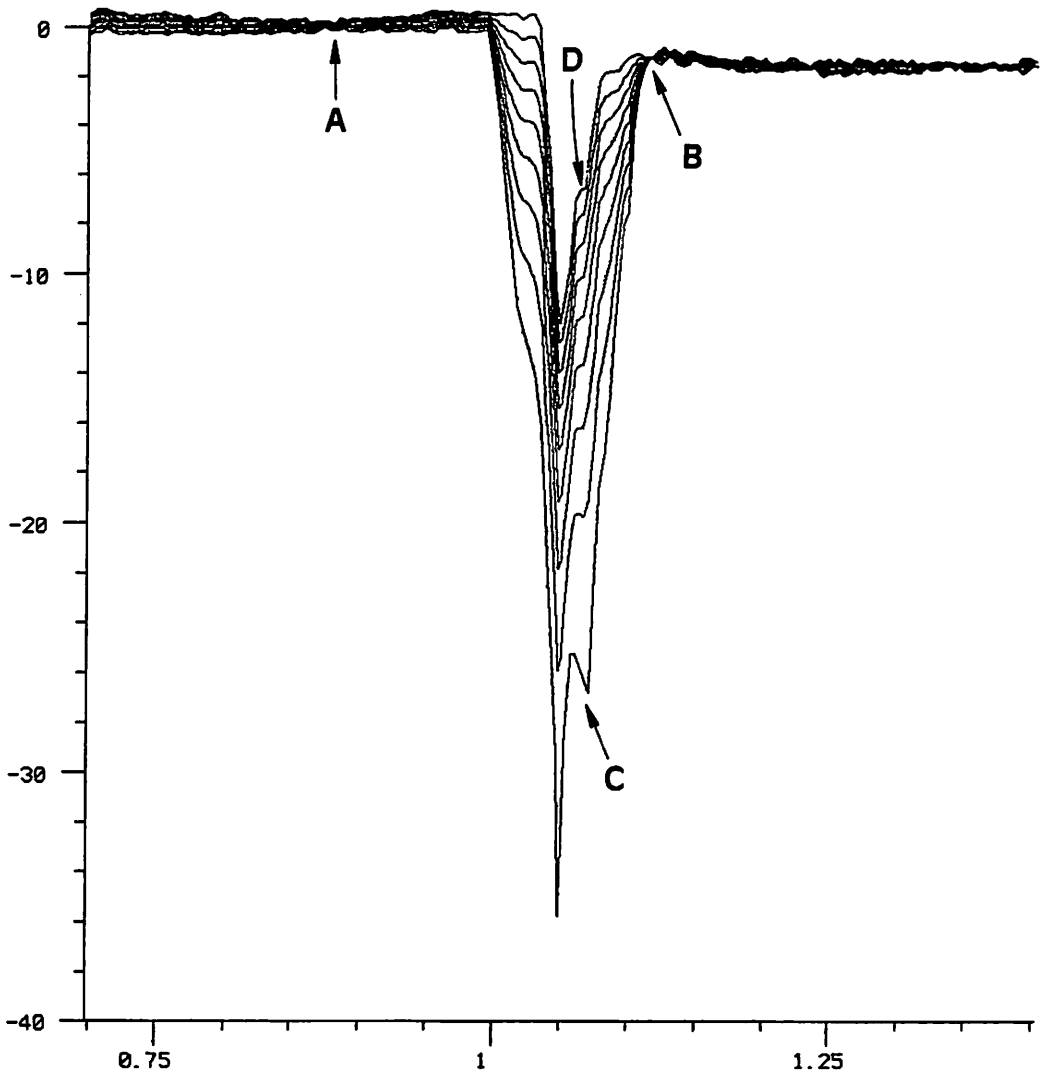


Figure 6.2. Detail of Figure 6.1, showing the boundary and interpolated amplitude envelopes at the transition. Here the *y*-axis is in decibels.

the variation of the attack *times* of the second note is difficult to describe, as the attacks of the tongued and untongued cases varied more in their form than in their rise time. Recall from the discussions in Chapter 5 that the shape of the attack of the first note can play a role in determining the percept. Thus, it is difficult to isolate the attack time of a second note and vary only it to conduct studies on the categorical perception of performed transitions. This is another instance in which studies of isolated notes can prove to be reductionistic.

The second stimulus (counting from the bottom of Figures 6.2, 6.4, and 6.5) had an amplitude envelope very close to that of the tongued transition, the first stimulus. If categorical perception

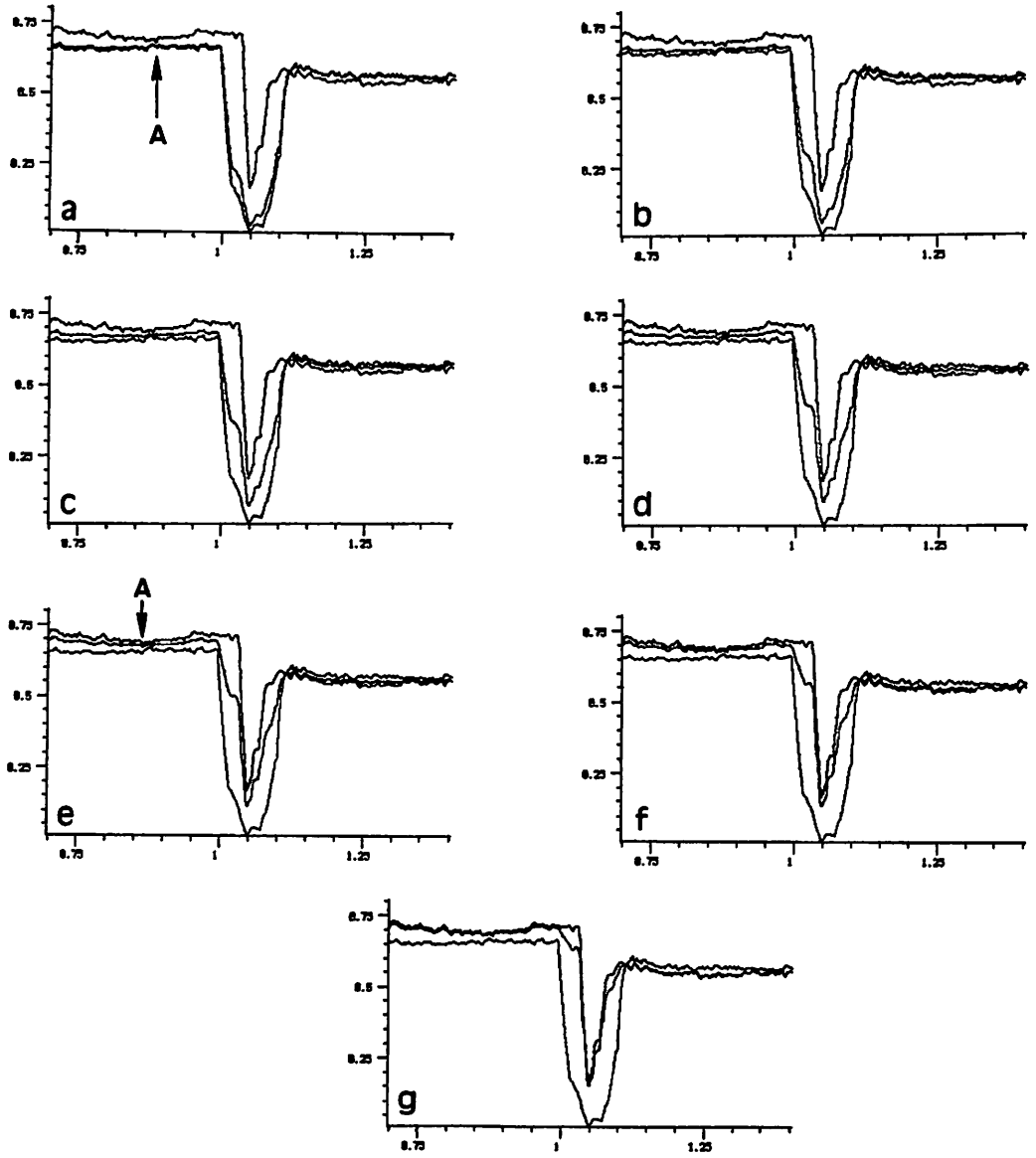


Figure 6.3. The individual amplitude envelopes of Figures 6.1 and 6.2. Each plot shows one of the seven interpolated amplitude envelopes along with the boundary (original tongued and untongued) envelopes. The *y*-axis is again linear.

occurs based on amplitude envelope alone, a categorical boundary should be found somewhere between stimuli 2 and 9 (the untongued original). If categorical perception based on spectral cues takes place, then the categorical boundary should fall exactly between stimuli 1 (derived from the original untongued recording) and 2 (the original tongued recording).

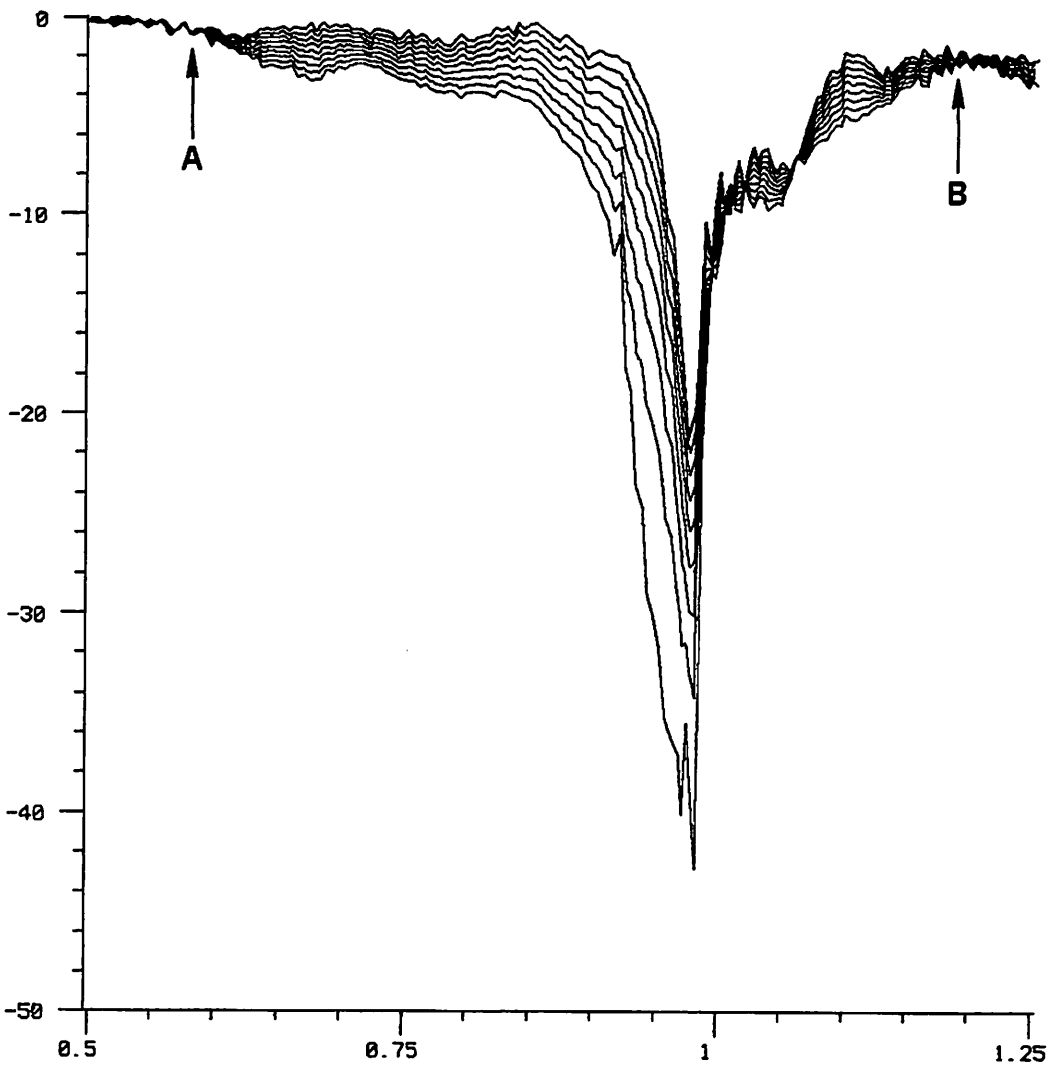


Figure 6.4. Boundary and interpolated amplitude envelopes, just at the transition, for the trumpet tones in Experiment 8.

Experimental Procedure

As in Experiment 7, the subjects were asked to judge each of the stimuli as "tongued" ("with bow change") or "untongued" ("no bow change"). Each stimulus was presented three times. (Details of the presentation of stimuli are given in Appendix 3).

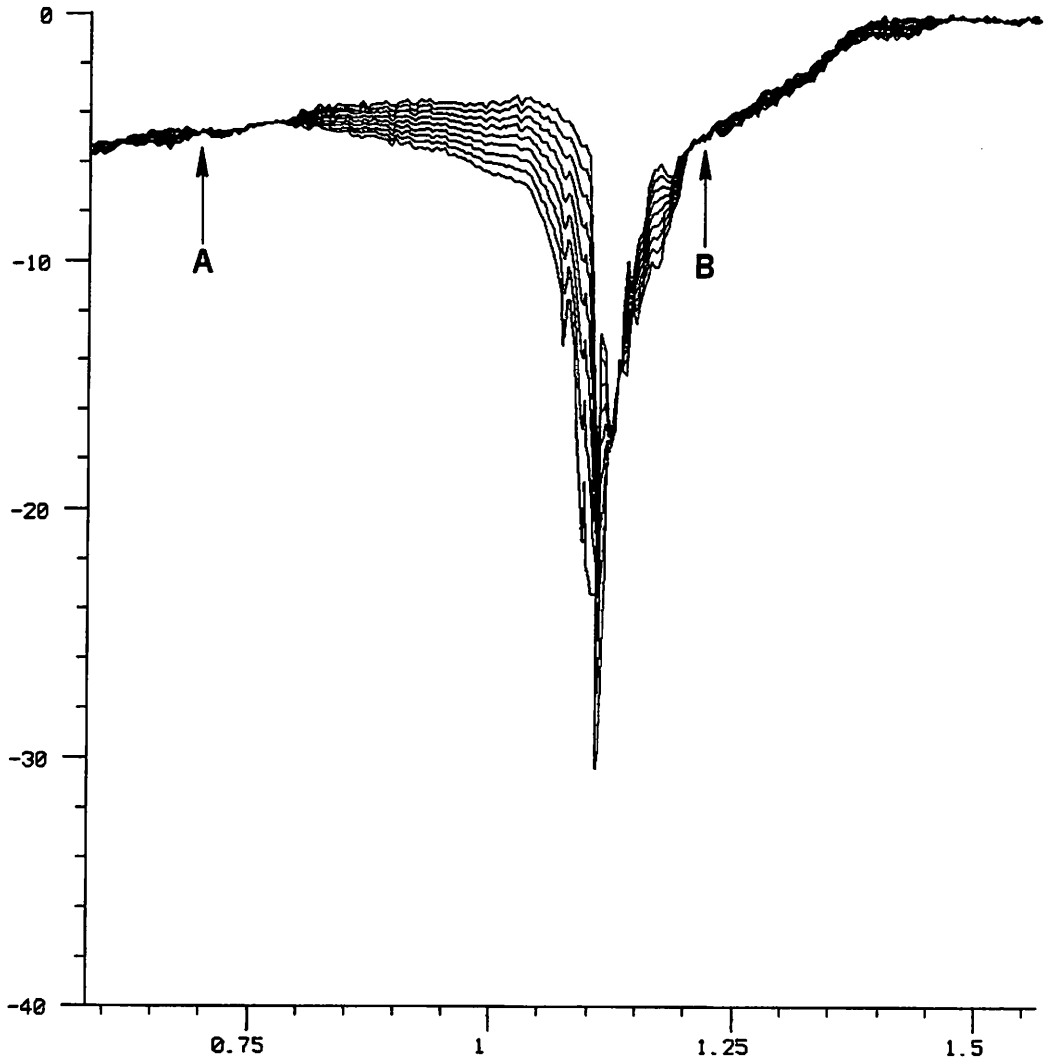


Figure 6.5. Boundary and interpolated amplitude envelopes, just at the transition, for the violin tones in Experiment 8.

Results

Table 6.1 shows how many times each subject labelled each stimulus as "tongued." For further numerical analysis, a response of "untongued" was assigned a value of 0.0, with "tongued" set to 1.0. The mean responses across all subjects are given in Table 6.2.

No clear pattern of categorical perception can be seen from these responses. There seems to be a jump in the mean scores for the clarinet between stimuli 1 and 2, indicating categorical perception based on spectral cues; however, the behavior of the means for the other two instruments is not

Table 6.1. Number of "Tongued" Responses (Experiment 8)

	Subject										
Stimulus	1	2	3	4	5	6	7	8	9	10	Total†
Clarinet											
1 (T)	1	1	3	3	3	3	3	3	1	3	24
2	0	0	1	0	1	3	0	1	0	2	8
3	0	1	2	1	1	1	0	0	0	2	8
4	0	0	1	1	0	2	0	1	1	0	6
5	1	0	1	1*	1	1	0	0	0	0	5
6	1	1	1	2	3	0	0	0	1	0	9
7	3	0	2	2	1	2	0	0	0	1	11
8	1	0	0	2	2	0	0	2	0	0	7
9 (U)	1	0	0	1	3	1	0	0	0	1	7
Trumpet											
1 (T)	2	3	3	3	3	3	3	3	3	3	29
2	0	1	3	3	1	3	3	1	3	3	21
3	0	3	3	2	2	3	3	0	3	2	21
4	0	2	2	2	2	3	3	2	3	1	20
5	0	1	3	3	3	2	3	0	3	3	21
6	1	2	3	2	3	3	3	1	2	3	23
7	1	1	2	3	3	3	2	1	3	1	20
8	2	2	3	3	3	3	2	0	2	2	22
9 (U)	2	1	3	2*	3	3	2	0	2	1	19
Violin											
1 (T)	2	3	2	3	3	3	3	3	3	3	28
2	3	3	2	2	3	1	0	1	3	2	20
3	2	2	1	0	3	0	0	2	3	2	15
4	2	2	2	1	1*	0	0	3	2	1	14
5	2	2	1	2	0	0	0	0	1	1	9
6	2	0	2	0	0	0	0	0	1	1	6
7	1	1	1	1	0	0	0	0	1	0	5
8	1	0	1	1	0	0	0	1	1	0	5
9 (U)	2	1	1	2	0	0	0	1	0	1	8

Note: Highest possible score for each subject and stimulus is 3.

*Only two responses collected

†Across all subjects (maximum = 30)

Table 6.2. Analysis of Experiment 8.

	Clarinet		Trumpet		Violin	
Stimulus	Mean	s.d.	Mean	s.d.	Mean	s.d.
1 (T)	0.80	0.40	0.97	0.18	0.93	0.25
2	0.27	0.44	0.70	0.46	0.67	0.47
3	0.27	0.44	0.70	0.46	0.50	0.50
4	0.20	0.40	0.67	0.47	0.48	0.50
5	0.17	0.38	0.70	0.46	0.30	0.46
6	0.30	0.46	0.77	0.42	0.20	0.40
7	0.37	0.48	0.67	0.47	0.17	0.37
8	0.23	0.42	0.73	0.44	0.17	0.37
9 (L)	0.23	0.42	0.66	0.46	0.27	0.44

Note: a mean value of 1.0 shows that the subjects responded "tongued" for all presentations of the stimulus; a mean of 0.0 shows that all subjects responded with "untongued."

so clear-cut. The totals column in Table 6.1 also fails to show any clear-cut patterns. Indeed, the total number of "tongued" responses does not monotonically decrease for any of the three instruments from stimuli 1 through 9. In the same table, subject 7 clearly shows categorical perception for the clarinet and violin; but not for the trumpet.

As in Experiment 7, some subjects were confused even by the original recordings. Subject 5 could not distinguish the tongued and untongued clarinets; the trumpet confused subjects 1, 3, 4, 5, 6, 7, and 9; and subjects 1 and 4 were also confused by the violin. Removing the responses of these subjects does not produce a set of data with any clearer results than those already presented.

Again, the written comments of the subjects provide some insight into their responses. For example, subjects 2 and 4 felt that the test tones were presented too quickly.

Another theme running through their comments is the difficulty of making this kind of judgement at all. Subject 7 wrote:

> I don't feel confident in my ability to tell apart a *real* violin bow change vs. non-bow-change if the player is very good. Similary, I don't think I could tell apart a "slightly tongued" real trumpet or clarinet from an untongued real trumpet or clarinet. So this made the synthetic ones even harder.

Be that as it may, this was the one subject who consistently picked tongued or untongued for two of the three instruments (see Table 6.1). Along the same lines, subject 9 remarked that "some weren't tongued or untongued, ...", and from subject 10 we have: "Many were not clear; $\Rightarrow$ random choice."

What I find so perplexing, in comparing Table 6.1 with Table 5.14, is the fact that some subjects performed differently in Experiments 7 and 8 when they were judging the same original recordings. A given subject performed perfectly in one experiment, but consistently mislabelled the same recordings in the other experiment. For the most part, the responses to the original stimuli in Experiment 7 were more reliable. This is especially the case for the trumpet; none of the means in Table 6.2 is as close to "untongued" as one would have experimented. Perhaps the effects of aural fatigue are visible in the data for Experiment 8, since it was played first on all three tapes.

Be that as it may, if categorical perception of the transitions of real musical instruments is to be perceived by anyone, it should be perceived by highly-trained musicians. Furthermore, if categorical perception occurs in 2-month-old infants, it should also occur in musically experienced adults without the need for carefully training the test subjects.

Conclusions

Categorical perception cannot be conclusively demonstrated for amplitude envelopes interpolated between recordings of tongued and untongued transitions performed on musical instruments. (Of course, the possibility remains that categorical perception might occur if the amplitude envelopes were more strikingly different.)

Incidentally, this conclusion strengthens the assertion made in Chapter 5, that changes in the amplitude envelope alone cannot compensate for missing or modified spectral cues in the transition.

CHAPTER 7

THE LAST CHAPTER

Summary

A few hundred transitions on nine non-percussive orchestral instruments were digitally recorded and analyzed. It became clear that a performer could repeat a performance with a high degree of precision; so the set of recordings was judged to be representative. This document contains amplitude, power, and spectral plots of many of these transitions.

What is a Transition? (III)

The work in Chapters 2–6, based on modifications of those recordings, requires only one small change to the definition of transition given in Chapter 1, which read, "A transition includes the ending part of the decay of one note, the beginning and possibly all of the attack of the next note, and whatever (if anything) connects the two notes." The "(if anything)" should be struck; if there is truly a transition between notes (as opposed to what happens when notes are purposefully detached), then something does exist in the transition, even for tongued transitions.

Amplitude

This definition begs the question of determining where the decay of one note begins and where the attack of another ends. Observation of the recorded transitions showed that the decay and attack surrounding a transition can have highly irregular shapes.

The amplitude in a transition drops 10 to 40 dB from the maxima of the surrounding notes. A consistent pattern was found in which the amplitude dip for tongued notes was greater than that for untongued. But the amplitude never drops to the background noise, even for tongued notes; such a low amplitude value might well be reached for notes purposefully detached.

The dip in amplitude found in nature should be included, and should not be raised. Raising the amplitude floor of the dip in performed transitions has the effect of artificially amplifying any noise which may be in the transition. The amplitude may be lowered without causing such problems, although the perceived articulation may be changed as a result.

The amplitudes of the notes surrounding the transition can often be quite different. This does not seem to have an effect on the area immediately surrounding the point of pitch change. Obviously, more potent articulations must be molded to mesh with their surroundings.

It became clear that the decay of a note may have a greater role than previously assumed, once the note is placed into a musical context. If, as I discussed in (Strawn 1982), timbre is broken down into two broad categories of instrument identification and quality judgment, then the results here do not contradict what is commonly assumed about the role of the decay in identifying an instrument. However, the shape and duration of the decay can have a significant role in determining the perceived articulation.

Time

On the other hand, there are bounds on the shape and duration of the "flanks" surrounding a transition. A decay that is too long sounds like a purposeful dimenuendo; an attack that is too long sounds like a "swell" on a note. In other words, flattening the slope of the "flanks" will not change a tongued percept into untongued.

The gap time between tongued notes is longer than for untongued transitions.

The point of pitch change occurs right at the beginning of the attack of the second note.

Pitch

The shape of the frequency glide between the two notes does not seem to be critical, unless it lasts too long; in such a case, a "glissando" or "sliding" effect can be heard. In the recorded transitions, the change in pitch occurs quite quickly, sometimes within a few periods.

Waveshape

Only in transitions where the attack of the second note shows significant noise can a discontinuity be said to occur; for legato transitions, no discontinuity at the point of pitch change or anywhere else in the transition was found.

Spectrum

There are characteristic spectral changes associated with a transition. As the amplitude drops at the end of the first note, the spectrum rolls off; and the spectrum builds up again as the second note enters. An adequate transition can be created without such spectral cues; but transitions containing them were preferred by listeners.

The spectrum of the transition region can be modelled as a low-passed version of the end of the first note, at least until the point of pitch change. Often, only the first 10 or 20 harmonics of the first note are left in the transition. The lower-frequency components in the transition region are not strong enough to mask the weaker higher-frequency components. More work should be done to determine how many of those higher-frequency components need to be retained to create a convincing articulation.

The attack of the second note may include instrument-specific features (such as the "blips" of the brass), which, at least in monophonic passages, should be retained. The amplitude characteristics of the transition cannot be modified to compensate for missing or modified spectral changes.

As with the amplitude dip, there is a difference in the spectral evolution between tongued and untongued transitions. A more detached articulation produces a deeper notch in the spectrum, and the notch itself lasts longer. There are certainly spectral cues peculiar to given playing styles; but these have been noted here only in passing.

Articulation

Such spectral cues are not the only determinant of the perceived articulation; amplitude cues play a very important role too. Changing the amplitude envelope of the untongued transition to that of the tongued transition produced a tongued percept. (The opposite case produced inconclusive results).

It must be remembered that a given perceived articulation can be achieved with a variety of playing techniques.

Modeling Transitions

I have shown that the phase vocoder (the short-time Fourier transform) is adequate, even if a little clumsy, for emulating a transition. It is able to track frequency adequately around the point of pitch change, and it is able to model any non-harmonicity in the transition adequately, even though the phase vocoder was not designed to do so. Line-segment approximations to the outputs of the phase vocoder adequately capture the characteristics of the transition. Again, any peculiarities of the attack, such as blips in the trumpet, should be retained. On the other hand, it seems adequate to connect the frequencies of the two notes with a simple vertical line at the point of pitch change.

Transitions and Timbre

It was not the goal of this work to use transitions to investigate timbre, even though that will undoubtedly become a fertile area of research. Still, some conclusions about the perception of timbre can be drawn based on the results presented here. The full-data representation of Experiment 1 is equivalent to Grey's *complex synthesis* (1975, p. 26). His *line-segment approximation* matches what was created for Experiment 2; and his *cut-attack approximation* is equivalent to the test stimuli of Experiments 4 and 5. In addition, I created, but did not use, transitions between tones which themselves matched Grey's *constant-frequency approximation*. My results agree completely with Grey's conclusions (1975, pp. 40–41). The complete resynthesis as well as the line-segment approximation adequately capture the timbre of the original transition. Since listeners consistently *preferred* the transitions with spectral changes to those without, the cut-attack approximation is once again seen to be less desirable. Finally, the "electronic" effects produced by the constant-frequency still occur, at least in the two-note snippets that I used here. Of course, I have not investigated that part of timbre perception which involves the *identification* of instruments.

Patterns in Transitions

The size of the interval performed (within the range examined here: a minor second through a minor seventh), the direction of the interval, and (except for the strings) the size of the instrument do not have as large an influence on the shape of the transition as does the intended perceptual result. On the other hand, my work showed that a given modification to the signal in the transition can affect each instrument of the orchestra in a different way. This is consistent with the findings of a number of timbre studies cited in Chapter 1.

Categorical Perception

Furthermore, it was not possible to demonstrate the categorical perception of articulations. The foregoing discussion of the importance of the decay in a transition suggests that research on categorical perception involving only attack times may be oversimplified.

Et Cetera

Finally, along the way we have taken a glance at some interesting issues, such as how to extend a musical signal in time; how to measure the time-varying power of a signal; how to make line-segment approximations easily; how to examine and edit time-varying spectra; and how to scale the amplitude of a signal more or less with impunity.

How to Make a Transition

Chapter 1 introduced the traditional model of individual notes, which breaks them down into attack, steady-state, and decay regions. The work discussed here does not require any modification of the essence of that model (no matter how inadequate that model might be). Thus, the shape and size of the transition seems to have no direct effect on the steady-state of the note. The attacks and decays have to be modified appropriately where they join a transition, which is what one would expect. Thus, the starting-point for making a good transition is still a pair of good notes.

This work has examined four different ways for joining those two good notes:

1. Cut off the end of the first note and the beginning of the second note. Overlap the two notes, and create a quick crossfade between them, perhaps on the order of 60 msec or so. Be sure to align the signals to avoid any gross phase perturbations. This will produce the world's cleanest legato; but remember that listeners really preferred transitions with some spectral cues.
2. Starting with the transition of the last paragraph, apply some amplitude dip, say 10 to 40 dB or so, lasting perhaps 10 to 100 msec, depending on the effect desired. In this case, the overlap time at the point of pitch change is not so crucial, since its amplitude is now reduced; shorter or longer amplitude times seem to work too.

However, the point of pitch change should occur right as the amplitude begins to rise for the second note.

3. The decay of the first good note presumably undergoes a spectral rolloff. Extend the "low-passed" end of the first note for the appropriate amount of transition time (again, perhaps 10 to 100 msec, depending on the effect desired). At the point of pitch change, crossfade from the extended end of the first note into the attack of the second note. Recall that this was done in Experiment 1, except that the extended transition was already available. The crossfade can be quick—20 msec should be adequate.
4. After the notes have been analyzed with the phase vocoder or some other suitable technique, create a transition by extending, say, the lowest 10 or 20 harmonics of the first note; their summed amplitude should be 10 to 40 dB down. At the point of pitch change, splice these harmonics onto the amplitude traces of the attack of the second note. The frequencies can jump vertically from the first note to the next note, right at the point of pitch change.

Suggestions for Future Work

The definition of transition given here has been kept purposefully vague. Many of the aspects of a transition listed at the beginning of Chapter 2 have not yet been explored. Doing so would lead to a better understanding of what is important in a transition, and what is not. On the other hand, coming up with a hard-and-fast rule for delineating the bounds of a transition may prove impossible.

Some more tools need to be developed to facilitate working with musical contexts. It would be useful to be able to equalize note pairs without distorting the transitions between them. Along these same lines, a measure of the "similarity" of two amplitude envelopes would be helpful. In the spectral domain, someone should investigate the differences between the original signal and the signal resynthesized from phase vocoder analysis. This applies to resynthesis from the full data as well as to resynthesis from the line-segment approximations. Surely some way can be found to reduce the number of channels needed for additive synthesis of high-quality musical sounds. Also, spectral editing could use considerable improvement. There should be a way, for example, to edit "across" a spectrum easily; I still think that the data structure proposed in (Strawn 1980) would be a good starting point.

This study has paved the way for work on the perception of timbre in multi-note monophonic contexts. This could follow several paths. One might use melodic contexts to throw light on human auditory processing. After all, an entire musical context makes higher requirements on human memory and attention than do individual notes. Or, one might use experiments on timbre perception to sharpen our understanding of what in the physical signal is important for generating real musical melodies. Another possibility would be to tackle the question of identifying musical instruments in a melodic context; if you splice, say, an attack from one instrument onto a note in the middle of a melody played by another instrument, can the difference be heard? Are there instrument-specific signatures which occur in a transition?

Leaving monophony, it would also be of interest to examine how transitions are affected by their "vertical" context. For example, what happens to the musical signal when two players "synchronize" their articulations?

APPENDIX 1

AMPLITUDE SCALING

For the studies and experiments discussd in Chapters 3, 4, and 5, it was necessary to perform amplitude scaling, often on a very small scale in time and/or amplitude. For example, in some preliminary work not discussed in this document, I found it necessary to raise the amplitude of a few periods by just 6 dB. Figure A1.1a shows a typical if highly simplified amplitude envelope. The idea is to scale a) by some function c) to make the final amplitude look like b).

At CCRMA, Loren Rush (1982) had developed a method of using raised sinusoids for crossfades. The function

$$\frac{\cos(\theta)+1}{2}, \quad 0 \leq \theta < \pi$$

is used for fadeout, and the same function in the range $\pi \leq \theta < 2\pi$ is used for fadein. It can be shown that the sum of the function at θ and $\theta + \pi$ equals 1.0, making this function useful for crossfading. Indeed, I used this facility, for example, in trying to splice spectral cues from one transition into the attack of a note in another transition, in the preliminary studies discussed under Experiment 7. However, this technique did not prove to be general enough for the amplitude scaling which I needed. Obviously, Figure A1.1c differs from a raised sinusoid.

Simple line segments such as shown in Figure A1.1c also proved inadequate in many cases, as a click occurred in the scaled waveform, especially at points A and D in the figure. It might be possible to modify the scaling function of c) to avoid clicks; but doing so by hand for each new case would be too clumsy.

I finally adopted cubic splines (which I had earlier rejected for line-segment approximation of individual amplitude and frequency functions produced by the phase vocoder—see [Strawn 1980], pp. 5–6). The function of Figure A1.1c is approximated with piecewise cubic splines, which have the property that at the breakpoints between each spline, the values and the slopes are constant, which means that a "phase pop" is avoided at each breakpoint. By "phase pop," I mean a sudden

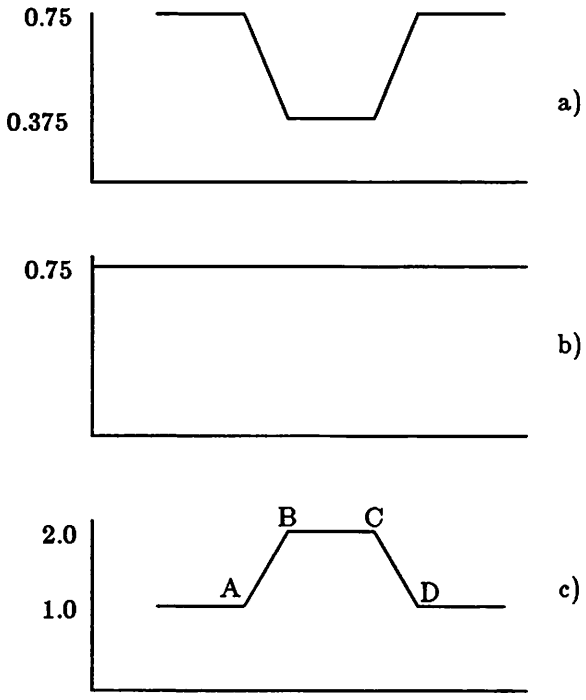


Figure A1.1. Typical amplitude scaling. a) Original envelope. b) Target envelope. c) Scaling function to convert amplitude of a) to that of b).

jump in the amplitude of a waveform, usually from one sample to the next. (Strong and Clark [1966a, p. 41] also reported problems with this). A phase pop can be caused when a scaling function suddenly starts, stops, or changes direction; improper amplitude scaling of this kind causes such pops, which are easily audible and a great source of irritation. Another advantage of using cubic splines is that scaling can be limited to an arbitrarily small part of the signal, resulting in a savings in computation time.

One must be careful here, though. If the function of Figure A1.1c is blindly approximated with cubic splines, the "Micky Mouse ears" of Figure A1.2b are the amusing result. It is therefore necessary first to interpolate the scaling function of Figure A1.2a as shown in Figure A1.2c; each dash connects explicit breakpoints. For most of my work, I used a spacing of 50 samples (=1.95 msec at my working sample rate of 25.6 kHz). When cubic splines are fit to *these* points, the resulting scaling function still has small "ears," but at this resolution in time, their effects are negligible. It is necessary to include an "extra" point with the value of 1.0, shown at E and H in Figure A1.2c, along with the true endpoints of the scaling function, shown at F and G in the

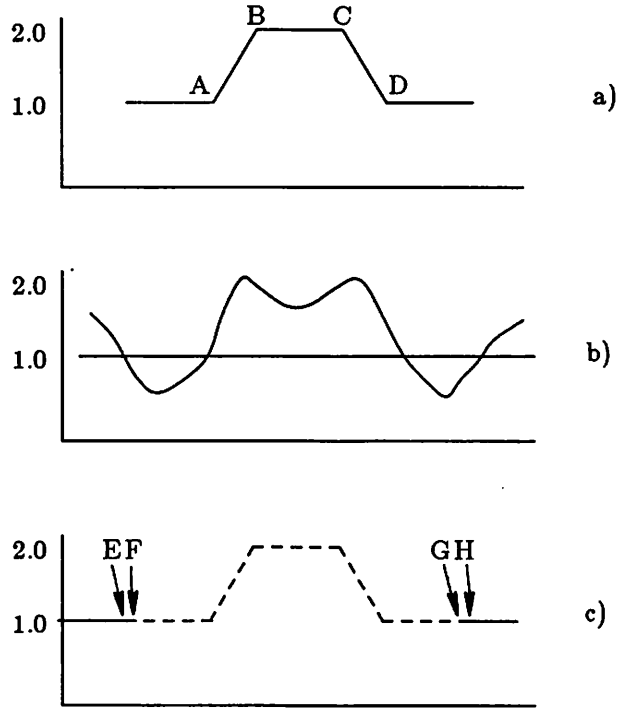


Figure A1.2. Amplitude scaling with cubic splines. a). The scaling function of Figure A1.1c, repeated here for clarity. b) Cubic spline interpolation of a). c) The function of a) linearly interpolated to a much finer resolution.

figure. This constrains both ends of the scaling function to a smooth transition to or from 1.0, again avoiding phase glitches in the output.

One small detail will complete the explanation. Values for the cubic spline do not have to be calculated at every sample time. I found it adequate to calculate 3 or 4 points for the cubic spline between two adjacent points in Figure A1.2c; the intervening points in the scaling function can then be linearly interpolated with impunity.

To give a concrete example, Figure A1.3 shows the amplitude envelope of the ascending M3 from the violin, played with bow change. The end of the decay of the first note has been extended. (This is another example of the work discussed on pp. 121–124 in Chapter 5). The goal is to provide a smooth decay at the end of the first note. Figure A1.4a shows the amplitude envelope of the first note and the transition; the upper line in the decay of the first note is the extended section. The lower line shows the target amplitude envelope. The scaling function is shown in the bottom half of the figure. It is this scaling function which is interpolated as shown in Figure A1.2c.

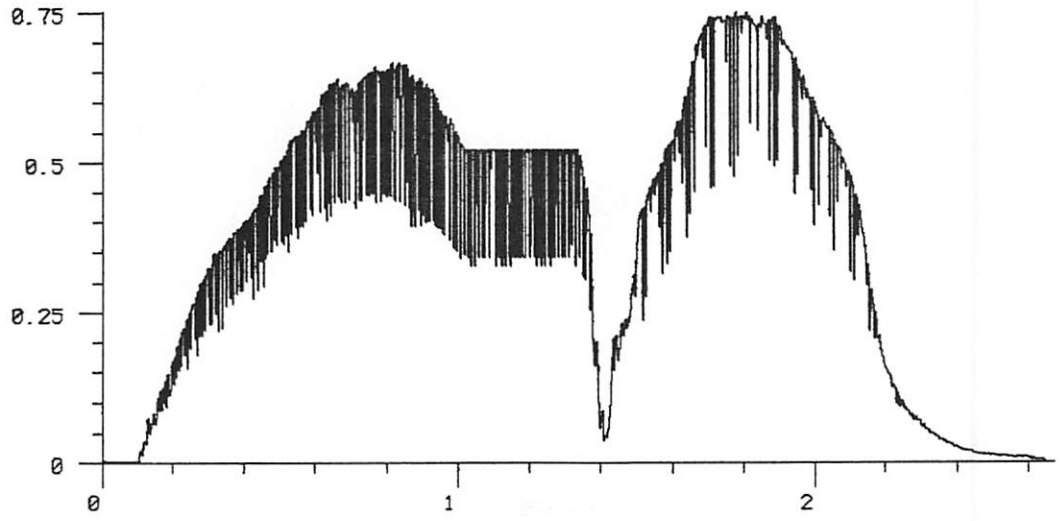


Figure A1.3. The end of the first note has been extended (violin ascending M3, with bow change).

By way of footnote: It might be possible to scale the magnitudes of the individual harmonics to achieve the same effect in a signal resynthesized with additive synthesis. The problem is that the magnitudes of the harmonics do not monotonically increase or decrease during a transition, nor do they increase and decrease in synchrony. Furthermore, it remains unclear how one could control the time-varying amplitude of the resynthesized signal.

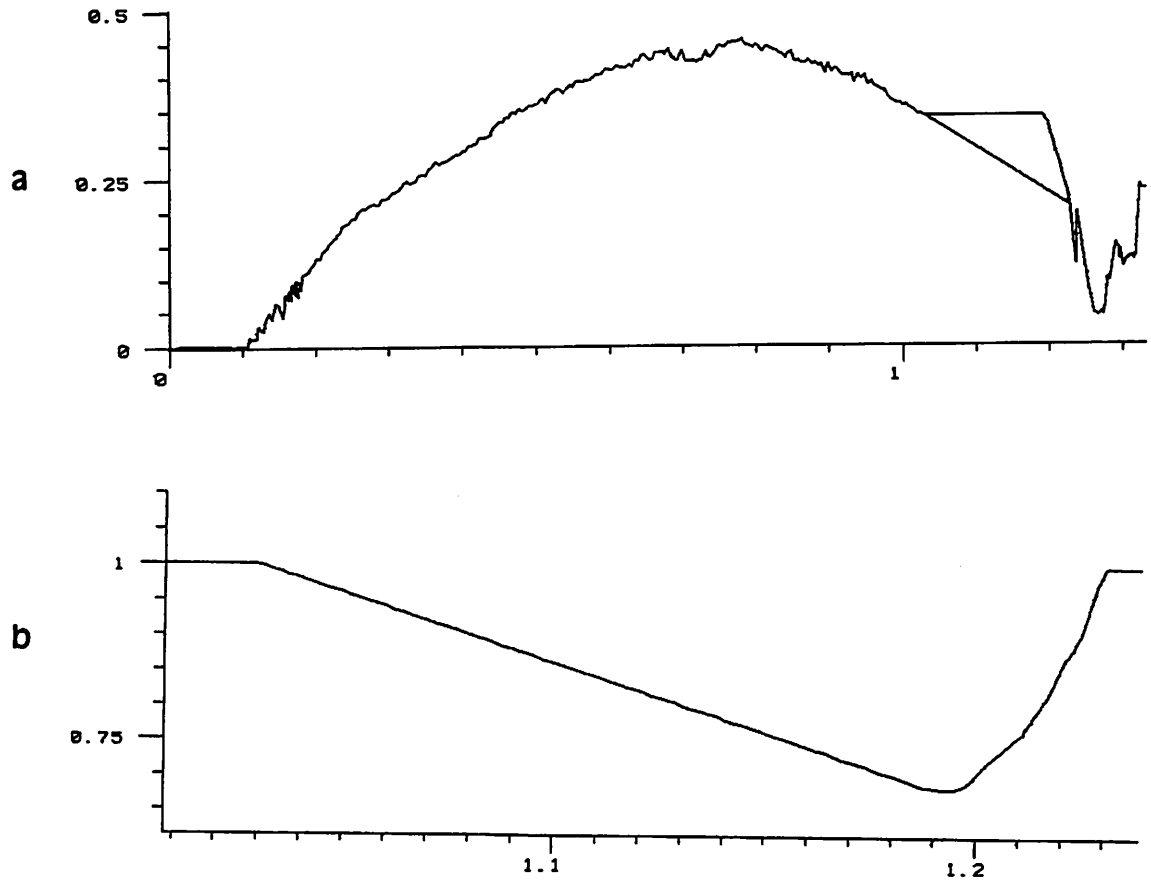


Figure A1.4. a) Amplitude of part of Figure A1.3. showing original and target envelopes. b) Scaling function (detail) for converting original envelope to target envelope.

APPENDIX 2

METHODS FOR EXTENDING WAVEFORMS

For the work reported here, as well as for any number of preliminary studies mentioned only in passing or not at all, it was necessary to extend a recording in time—sometimes by a few milliseconds, sometimes by a second or so. This appendix discusses three time-domain methods and two frequency-domain methods which I examined for this purpose.

Method 1: Repeating a Large Section of a Note

This is the simplest of the methods I developed. In Figure A2.1a, we see a schematic representation of the amplitude envelope of a note. A large section—500 milliseconds or so—of the steady-state is selected, as shown by B and C in the figure. Part b in the figure shows how this section can be duplicated and spliced back into the note, using the cross-fade procedure given in Appendix 1. The splice points E and F must be carefully chosen to produce a seamless splice. At E, some period peak in the signal from AB must align with some period peak from CD. The same holds true for point F. If the peaks are not aligned in this manner, a noticeable phase "glitch" is the usual result. I typically used splice times (AB or CD) on the order of 20 msec.

In order for this to work, the steady-state portion of the note to be extended must live up to its name: There cannot be a gross drop in amplitude across BC, nor can there be a large change in frequency. This method worked well for my clarinet recordings, but not so well for the violins or trumpets. It does have the advantage that the entire duplicated portion automatically has all of the "lively" quality necessary for creating natural-sounding musical tones, which is a problem with some of the other techniques discussed in this appendix.

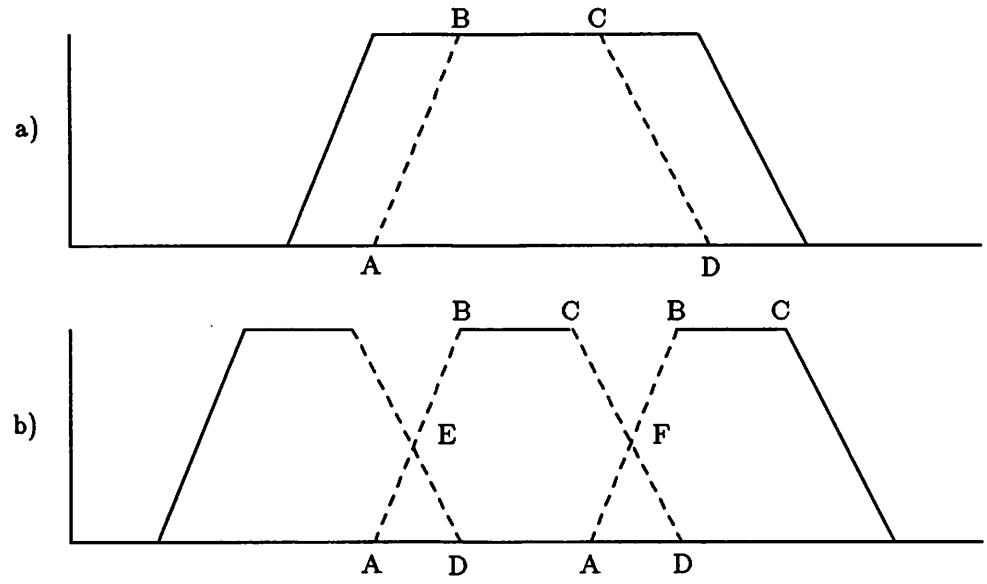


Figure A2.1. To extend the note shown in a), part of the steady-state (BC) can be duplicated and spliced back in, as shown in b).

Method 2: Concatenation

This is again a crude and simple method which works only in certain cases. The basic idea is to splice out one period of a waveform and duplicate it. As certain synthesizer manufacturers have discovered, this is not always as simple as it sounds. The so-called steady-state of a note can be changing so fast that a "phase pop" occurs as one samples from the end of such an excised "period" back to the beginning. (Methods 3 and 4 discussed below address that problem). At any rate, once a suitable period is found, it is duplicated some arbitrary number of times; the extension is spliced back into the original right where the isolated period came from.

It seems natural to excise a period from a signal by looking for zero crossings which delineate a period. I had good luck with picking off period peaks instead. This is one of the methods I used for preparing the test tones described on pp. 121–124. Figure A2.2a shows a detail of the tongued clarinet transition already presented in Figure 2.3. A period suitable for extension is shown at A in the figure. The lower half of the figure shows the decay of the first note extended by duplicating that one period.

One disadvantage of this technique (and the technique discussed next) is that the isolated period may differ slightly in pitch from the periods surrounding it; the jump in pitch at the begin-

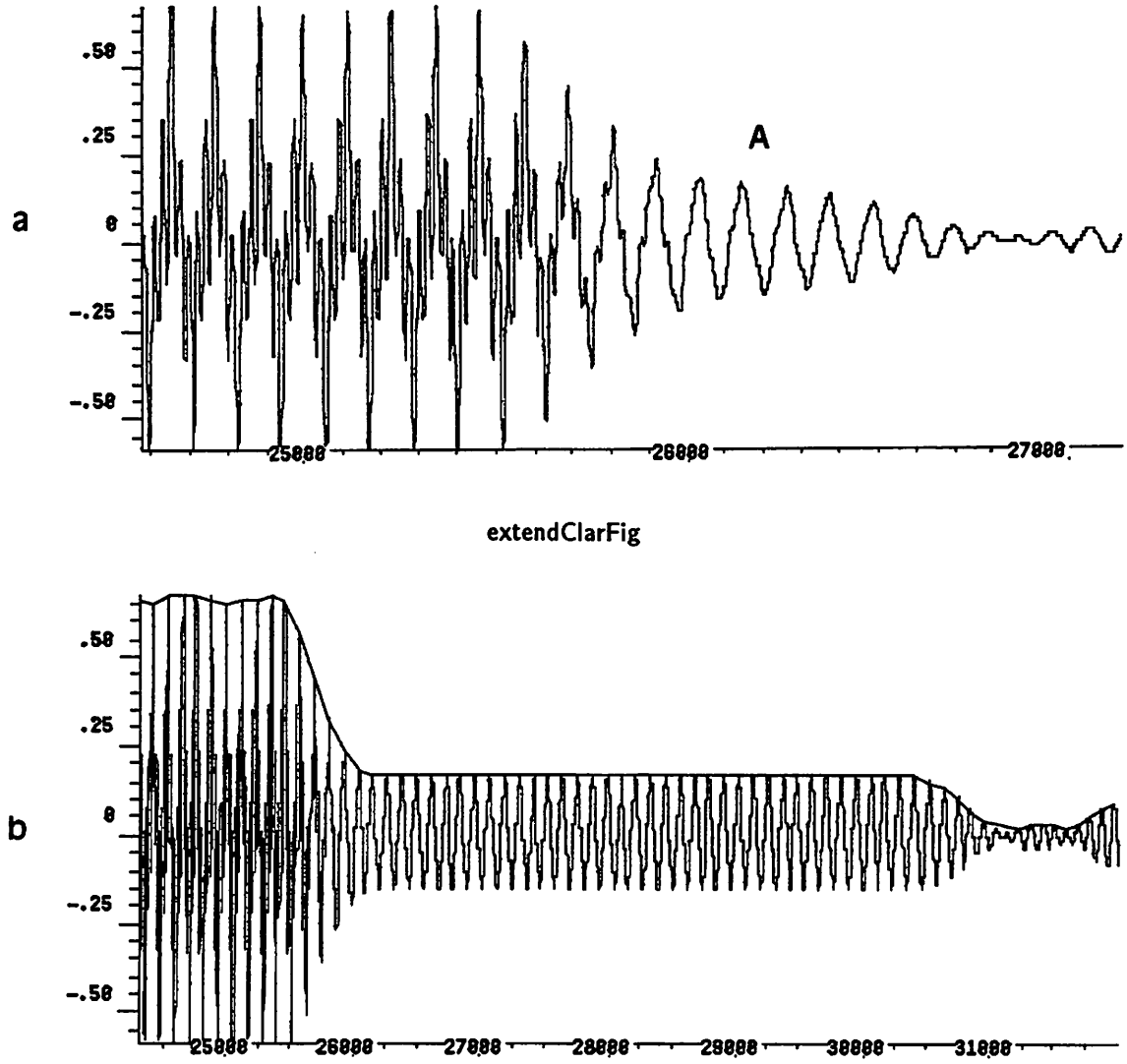


Figure A2.2. The period shown at A can be isolated and duplicated to extended the decay of the note, taken from a tongued clarinet transition.

ning and ending of the artificially extended signal can be quite noticeable. Changing the length of the period to be extended *by as little as one sample* can cause the pitch of the extended section to move up or down in pitch—too far. This is another artifact of the problem already mentioned; that is, isolating a truly steady period. This problem also caused difficulties in attempting to extend the *transition* between notes. Thus, it was impossible to use a period from E in Figure 5.3, because the period peaks (or zero crossings) that could be isolated all implied a periodicity which corresponded to a pitch vastly different from the pitch of the first note. It was necessary to back up to A–B–C in the figure in order to find a useable "period."

Another disadvantage of this technique (and the next one to be discussed) is that if the extended signal is very long, the ear hears that the extension is "artificial" and "electronic." For fairly short extensions, perhaps on the order of a 100 msec or so, I had good luck with multiplying every sample of the extension by a random number on a sample-by-sample basis, with the random numbers limited to about −60 dB from the amplitude of the extended signal. These random variations "fool" the ear into thinking that it is hearing a lively signal. For longer extensions, this simple solution does not work.

Method 3: Moorer's Overlap-Add Method

James A. Moorer was kind enough to suggest what I call an overlap-add method. This is not the same as the "overlap-add method" for resynthesis from the short-time Fourier transform; see (Allen and Rabiner 1977). The method, which has a vague historical similarity with "pitch-synchronous reverberation" (Miller 1973, pp. 46 ff), can be summarized as follows:

1. Remove two periods from the steady-state portion of the original recording. The period boundaries are defined to be the waveform peaks, as discussed earlier.
2. Across the two periods apply a window derived by inverting a cosine waveform scaled to the bounds $[0, +1]$.
3. Overlap and add the two periods to produce one period which may then be duplicated.

This is the technique, then, used to create the extensions shown in Figures 5.4 from the signal between points A and C in the original of Figure 5.3.

This method has the advantage that the endpoints of the individual period being extended match nicely. On the other hand, the problems mentioned under Method 3 still occur here: The change in pitch between the original and the extension, and the lack of time-varying spectral changes in the extension, can be quite noticeable.

Method 4: Fourier Resynthesis

Moorer and I also discussed a method which is presented here briefly for the sake of completeness, but which I did not implement. To create an isolated period, perform the Fourier analysis of a properly windowed but still small number of periods. Fourier resynthesis will produce a waveform in which the endpoints match as needed as long as both the real and imaginary parts of the analysis are used; in other words, phase information may not be discarded.

Method 5: The Phase Vocoder

The development of the phase vocoder has been highlighted if not accelerated by extensive research into time-scale modification of musical and speech signals (Portnoff 1978; Holtzman 1980; Dolson 1983). Rather than take the time to implement the clever phase-unwrapping which this method involves, I used a brute-force method—simply scaling the amplitude and frequency functions in time before resynthesis. This can be accomplished quickly and easily with a few modifications to the code given in (Gordon and Strawn 1985, pp. 254–57): basically, the quantities `ROverQ` and `nOSamps` in that code must be multiplied by the time scalar. It is possible to start and stop the time scaling without "phase pops", since phase is accumulated for each channel independently. This method has the further advantage that, except for extremely large time scaling factors, the extended signal sounds quite natural, as all the harmonics are still evolving independently. Ultimately, this proved to be the only method that would work for Experiment 3; Figure 4.6 gives one example of the results of this technique. It is computationally much more expensive than the other methods given here, but is likely to perform as needed when the other techniques fail.

APPENDIX 3

EXPERIMENTAL PROCEDURE

The Subjects

Ten volunteers took all of the experiments. Although I also took all of the experiments, my results are not included here; by the time I had worked on creating the test tones for two years, I was able to identify the process used to create the tones, which of course distorted my responses.

Table A3.1 summarizes the background of subjects, listed in the order in which they took the experiments. All subjects were males; only one (5) had any hearing problems—a slight tinnitus, audible only in the quietest of settings. Some were from the CCRMA community; others were writers and computer programmers. All were musicians with professional training; and all were familiar with electronic and computer music. In alphabetical order, they were: Jim Aiken, Thom Blum, Chris Chafe, Doug Fulton, David Jaffe, Kyle Kashima, Douglas Keisler, Brian Schober, Xavier Serra, and Amnon Wolman. The volunteer participation of these 10 subjects is gratefully acknowledged.

Recording the Test Tones

As was explained in Chapter 2, the original recordings were transferred to computer disk. All of the processing necessary to create the test stimuli for the various experiments was conducted in the digital domain; no analog processing was used.

The control and test tones were reproduced using the D/A converter of the "Samson box" at CCRMA (Samson 1980, 1985). This analog output was fed to the line input of a Sony F1 digital

Table A3.1. Background of Test Subjects

Subject	Age	Formal Musical Training	Instrumental Experience*	Previously Taken Psychoacoustic Experiments?
1	26	B.Mus., composition M.A., composition D.M.A., composition	piano (20) clarinet (3) guitar electric bass	no
2	33	B.A., music Ph.D., music, in progress	piano (25) violin (5) pipe organ (1) various electronic keyboards (10)	no
3	24	B.Mus., voice	piano (7) guitar (5) voice (4) saxophone (2)	no
4	30	B.A., computer music two years graduate work in composition	classical guitar (10) electric bass (4) clarinet (<1)	no
5	36	some undergraduate theory and composition	bass guitar (10) cello (10) piano (5) electronic synthesizer	no
6	33	B.Mus. D.M.A.	piano (25)	no
7	29	B.A., Music M.A., composition D.M.A., composition	guitar (17) violin (16) cello (2–3) oboe (3)	yes
8	29	B.Mus. M.Mus. D.M.A., in progress	piano (9) voice (4) recorder (4) guitar (2)	no
9	32	M.A., composition D.M.A., composition	cello (22) contrabass (17)	yes
10	25	M.Mus. Ph.D., music, in progress	guitar (15) cello (6)	no

*The number in parentheses gives the number of years studied and/or performed.

Table A3.2. Timing of Experiments

Experiment Number	Silence Between Stimuli	Silence Between Cases
1	0.20	0.50
2	0.20	0.65
3		0.50
4	0.75	1.50
5	0.25	0.75
7		0.75
8		0.40

Note: All times in sec. In some experiments, a case consisted of only one stimulus.

Table A3.3. Order of Presentation of Experiments

Order On Tape	Experiment Number	Duration (min)	Experiment
1	4	5	Amplitude dip without spectral cues
2	5	25	Variations in amplitude dip
3	3	5	Overlapped tones
4	7	3.5	Swapping amplitude envelopes
5	8	5	Categorical perception
6	1	1	Phase vocoder
7	2	7	Line segment approximations

tape recorder, a process already used for making record masters at CCRMA. The digital tape(s) recorded in this manner were played to the test subjects using a Sony F1. It was impossible to digitally transfer the digital samples from computer disk to Sony F1 at CCRMA at the time when these tapes were made.

Stimulus Timing

The average duration of each two-tone stimulus was 0.75 sec. In a given experiment, each case was separated from the next by a short amount of silence. When a case consisted of two stimuli, the stimuli were also separated by silence. The durations of these silences are given in Table A3.2.

Table A3.3 summarizes some other information about the experiments, including the duration of each. The total duration of the tape was approximately one hour.

Randomizing the Order of Trials

For each experiment, the trials were presented in a quasi-random order, which can best be explained with an example. In Experiment 5, there were 19 cases (see Table 5.5). Some cases were played three times, some twice. These 19 cases were created for five transitions: two on the clarinet, two on the violin, one on the trumpet.

For all of the experiments, the test tapes were arranged so that two stimuli from the same instrument were never played in succession. A further constraint existed for Experiment 5. Each case consisted of two recordings; each recording was in a file. Each of the 35 files appeared in more than one case. For example, the original recording for a given instrument appeared in cases 1–12 and 19; the case with 0 dB dip appeared in cases 1, 7, and 13; and so on. Only 31 files could be queued up for playing at one time on the CCRMA system. Therefore, the 250 trials for this experiment were broken down into groups of 50; cases were randomly selected within each group of 50 until the maximum number of files was reached. For the rest of the group of 50 trials, only cases requiring the currently available set of files could be selected.

At the beginning of each experiment, several "training examples" were presented. These were scored by the test subject, but were not included in the numerical analyses presented here.

Three different tapes were made. On each tape, the order of trials within a given experiment was different. Also, a given experiment started with a different instrument on each tape. Subjects 1, 4, and 7 used one tape; subjects 2, 5, and 8 used the second tape; and the other four subjects used the third tape.

Order of the Experiments on the Tape

Table A3.3 shows the order of presentation of the experiments, which remained fixed for all three tapes. This order was arbitrarily chosen for its convenience in making the tapes. Perhaps the order of experiments should have been changed on the tapes; possible fatigue effects were discussed in the conclusion to Chapter 6.

Collecting the Responses

Playback Setup

The subjects heard the tapes in the room used for recording the original tones (described in Chapter 2). They sat at a desk, behind which a loudspeaker was mounted such that it directly faced the seated subject. Playback level was adjusted to be comfortable, and stayed constant for all subjects. The digital tape machine was located in a side room, as its motor created a small amount of noise. The tape machine was thus controlled by the subject using a hand-held remote-control line-of-sight device.

Directions to Subjects

For each experiment the subjects were told to play the tape from start to finish without stopping. The digital tape machine used for playback could not be stopped and started at arbitrary places, as with an analog tape machine; readout of the video tape required a certain startup time. If the subject stopped the tape in the middle of an experiment, it might be possible to restart the tape later; but in order for this to work, markings from the "footage counter" on the tape machine would have to be included in the response sheets, which was deemed too heathen to be acceptable. Furthermore, it was difficult for the subject to see the "footage" counter on the digital tape machine used for playback, since it was set so far away. In general, the experiments were short enough so that the requirement of playing an experiment without stopping posed no problems. One subject did not follow these directions and as a consequence missed a few trials; this will be discussed shortly. The one exception to this paradigm was experiment 5, which lasted about 25 minutes. As was already mentioned, this experiment was broken up into 5 50-trial groups; the subjects were allowed and encouraged to stop the tape and rest every 50 trials. The subjects were also required to stop the tape after every experiment, and were allowed to rest between experiments as long as they liked.

The subjects completed answer sheets printed on normal 8.5×11" paper. Each experiment was introduced with a short written explanation; before I left a subject to his own devices, he read through each explanation and I answered any questions. The subject re-read the directions before each experiment ran; this was another reason for stopping the tape between experiments. To prevent confusion, a voice announced the end of each experiment, giving the experiment's number in each case.

The trials were numbered on the answer sheets, one trial per line, double-spaced. Each trial was given a number (starting with 1); the name of the instrument playing was included, to prevent confusion. Each line also included all possible responses (for example, "acceptable" and "unacceptable", or "A" and "B" in the preference tests). The subject marked his response with a pen.

After all of the experiments were run, the answers were entered manually into the computer. Earlier that year, Mr. Robert Currie of CCRMA had helped me proofread the page proofs for the volume (Roads and Strawn 1985). Following a suggestion from the publisher, Mr. Currie read the manuscript aloud while I checked the page proofs for errors. We followed a similar method for entering the data here—he read the responses, I typed at the terminal. Having proofread a 700-page book with him, I was convinced of his accuracy. I do not believe that errors were introduced in this process. Reading in the entire data for all subjects took a few hours. Mr. Currie's assistance is gratefully acknowledged.

It might have been possible to collect the responses directly on computer. This method was used at CCRMA by Gordon (1984), for example. However, there were many factors which spoke against doing so. The first was the possibility of computer failure with the subject sitting incommunicado at the other end of the building. (Obviously, there were no telephones in the room where the subjects heard the tapes). If the computer failed, then the subject would have to relocate a trial on the tape; but the difficulty in doing so has already been mentioned. Also, a large amount of software would have to be written for collecting the responses in some useable fashion. Since some of the subjects were not users of the CCRMA system, such software would have to be written for the novice. Initial estimates showed that the time involved in collecting the responses by computer would greatly exceed the time needed for collecting responses on paper; and this turned out to be the case.

Subjects' Written Responses

At the end of each experiment, the subjects were encouraged to write their own comments on the answer sheet. These have been cited, where appropriate, in the discussions of the experiments.

Missing Responses

With seven experiments on the tape, there were 667 "real" responses plus 64 training responses, for a total (across 10 subjects) of 7310 responses. Of these, 7 were missing: 1 in Experiment 7, 2

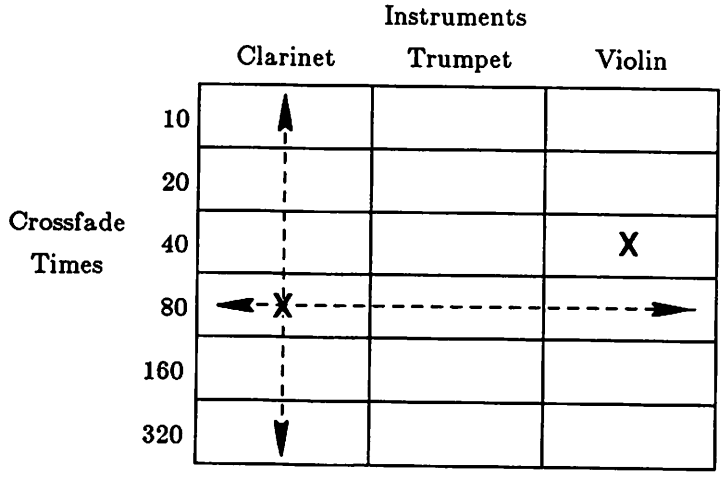


Figure A3.1. Arrangement of responses for analysis of variance in Experiment 3. Responses were missing in the bins marked with "X".

in Experiment 3, and 4 in Experiment 8. All of these missing responses were due to one subject, who stopped the digital tape in spite of instructions not to do so.

Calculating mean values with a few missing responses presented no problems. The means given in Chapters 3–6 were calculated by dividing by the reduced number of responses.

The situation was more complicated for analysis of variance. To give one example, Figure A3.1 shows how the data were arranged for Experiment 3. There were three columns, one for each instrument; the overlap times formed six rows. Each box in the figure contained 50 entries (5 trials per subject times 10 subjects). One response was missing for the 80-msec overlap time for the clarinet, and one for the 40-msec overlap in the violin, as shown by X's in the figure. To perform analysis of variance, the missing response had to be replaced with something besides 0.0.

A "dummy" answer was created by the usual method of summing all of the responses in all of the bins above, below and to the sides of the bin containing the missing entry; the arrows in the figure show the corresponding set of bins for one of the missing responses. This sum is divided by the number of entries in the bins in question; the result is then used as a dummy response for calculating analysis of variance.

APPENDIX 4

A LEXICON OF ANALYZED TRANSITIONS, PART 1:

POWER ANALYSES

Dedicated to John Snell, who originated the
Lexicon of Analyzed Tones in
Computer Music Journal

Preface

Analyses of time-varying power and spectrum have been presented in the main part of this document for some instruments. In this appendix (and the next), more analyses will be presented to complete the set.

Table A4.1 gives an overview of what this lexicon contains, and where the plots may be found. The figures included in this appendix follow the order given in the left-hand column of the table.

I have attempted to select a set of analyses that give a representative sample of what power and spectral plots look like; reproducing plots for all of the analyses would be prohibitively cumbersome. In particular, plots have been selected to allow the reader to selectively examine:

1. both power and spectral analyses for the same interval. This is the reason for listing the spectral plots of Appendix 5 here in Table A4.1.
2. both ascending and descending intervals
3. a range of different sizes of intervals
4. the contrast between playing styles
5. the effects of instrument size. The reader should compare oboe with bassoon, flute with piccolo and bass flute, violin with cello.
6. the repeatability of performed transitions. Multiple analyses are presented for the flute (power and spectrum) and the bassoon (power only).

The interpretation of time-varying power plots is discussed in Chapter 2. All power plots are shown on a decibel scale spanning 60 dB. Time is in seconds; the duration shown varies from recording to recording.

Table A4.1. A Lexicon of Analyzed Transitions.

Instrument		Amplitude and Power							
Direction		Tongued (With Bow Change)				Untongued (No Bow Change)			
		2	3	5	7	2	3	5	7
Flute	↑		2.30R				2.31R		
	↓								
Piccolo	↑		A4.1		A4.1		A4.2		A4.2
	↓		A4.1		A4.1		A4.2		A4.2
Bass Flute	↑		A4.3		A4.3		A4.4		A4.4
	↓		A4.3		A4.3		A4.4		A4.4
Clarinet	↑		2.3	2.3	2.3		2.4	2.4	2.4
			2.9	2.9	2.9		2.10	2.10	2.10
	↓								
Oboe	↑		A4.5		A4.5		A4.5		A4.5
	↓								
Bassoon	↑	A4.7	A4.7	A4.7	A4.7*	A4.8	A4.8	A4.8	A4.8
							A4.6R		
	↓								
Trumpet	↑	2.5	2.5	2.5	2.5	2.6	2.6	2.6	2.6
		2.11	2.11	2.11	2.11	2.12	2.12	2.12	2.12
	↓								
Violin	↑	2.7	2.7			2.8	2.8		
		2.13	2.13			2.14	2.14		
	↓	2.7	2.7			2.8	2.8		
		2.13	2.13			2.14	2.14		
Cello	↑	2.38	2.38			2.39	2.39		
	↓	2.38	2.38			2.39	2.39		

Note: Each entry is a figure number. The column headings 2, 3, 5, and 7 show the size of the interval played. ↑ is ascending, and ↓ is descending. R ("repeated") after the number of a power analysis figure means that more than one example of the given interval is presented.

*Four examples of the tongued ascending seventh on the bassoon are given in (Strawn 1985b)

Table A4.1 (continued)

Instrument		Spectrum							
Direction		Tongued (With Bow Change)				Untongued (No Bow Change)			
		2	3	5	7	2	3	5	7
Flute	↑		2.32		A5.2		2.33		
			2.34				2.35		
			2.36				2.37		
	↓		A5.1		A5.3				
Piccolo	↑		A5.4				A5.5		
	↓								
Bass Flute	↑								
	↓		A5.6				A5.7		
Clarinet	↑		2.18					2.19	
	↓								
Oboe	↑								
	↓		A5.8		A5.10		A5.9		A5.11
Bassoon	↑	A5.12	A5.19F			A5.13	A5.20F		
	↓			A5.14				A5.15	
Trumpet	↑		2.20				2.21		
	↓				A5.16				A5.17
Violin	↑								
	↓		2.22				2.23		
Cello	↑	A5.18F							
	↓		2.40				2.41		

Note: F after a figure number means that a three-dimensional frequency plot is shown.

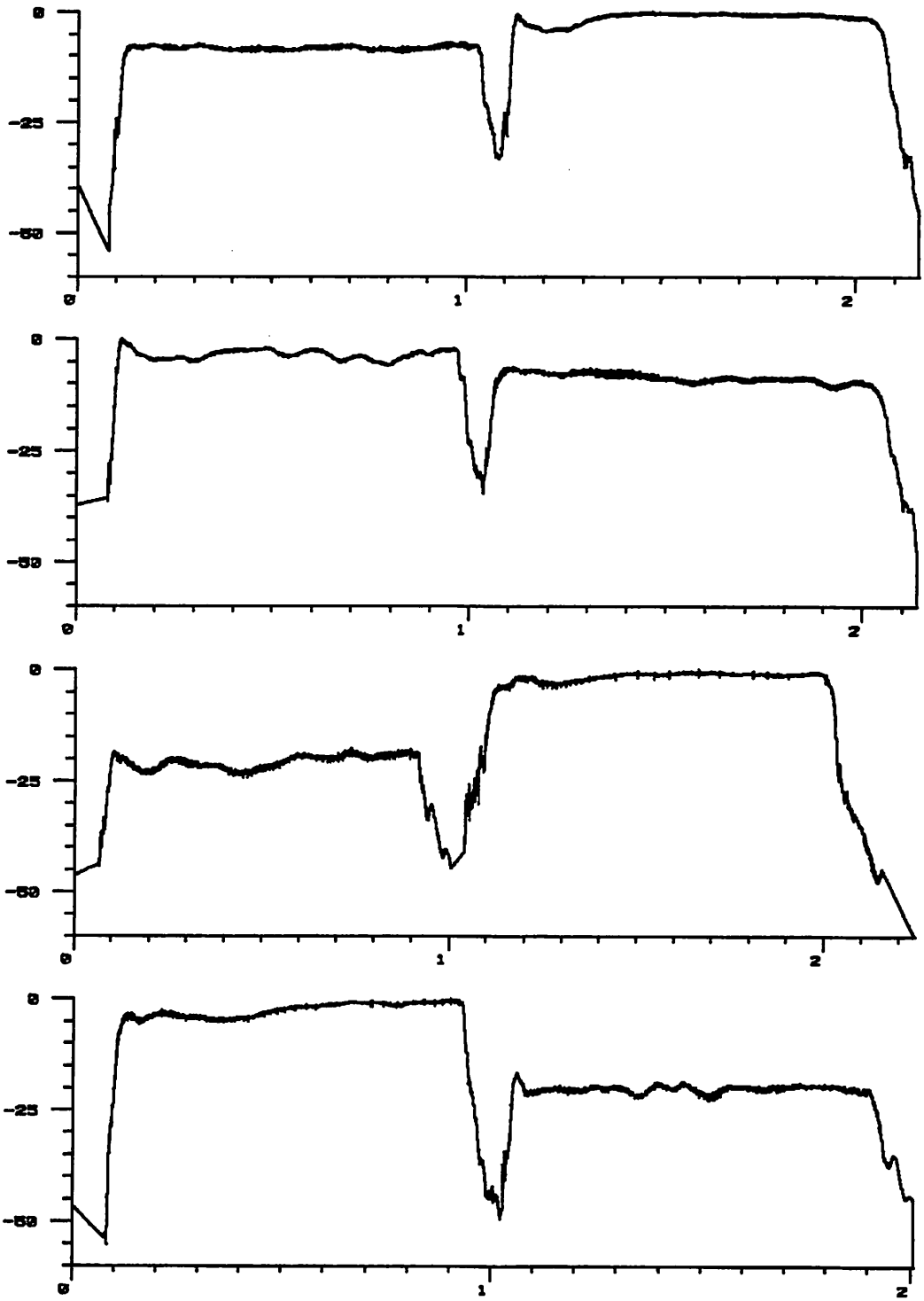


Figure A4.1. Time-varying analysis of four tongued piccolo intervals. From the top: major third ascending, major third descending, minor seventh ascending, minor seventh descending. The lower note in all four plots is A1760.

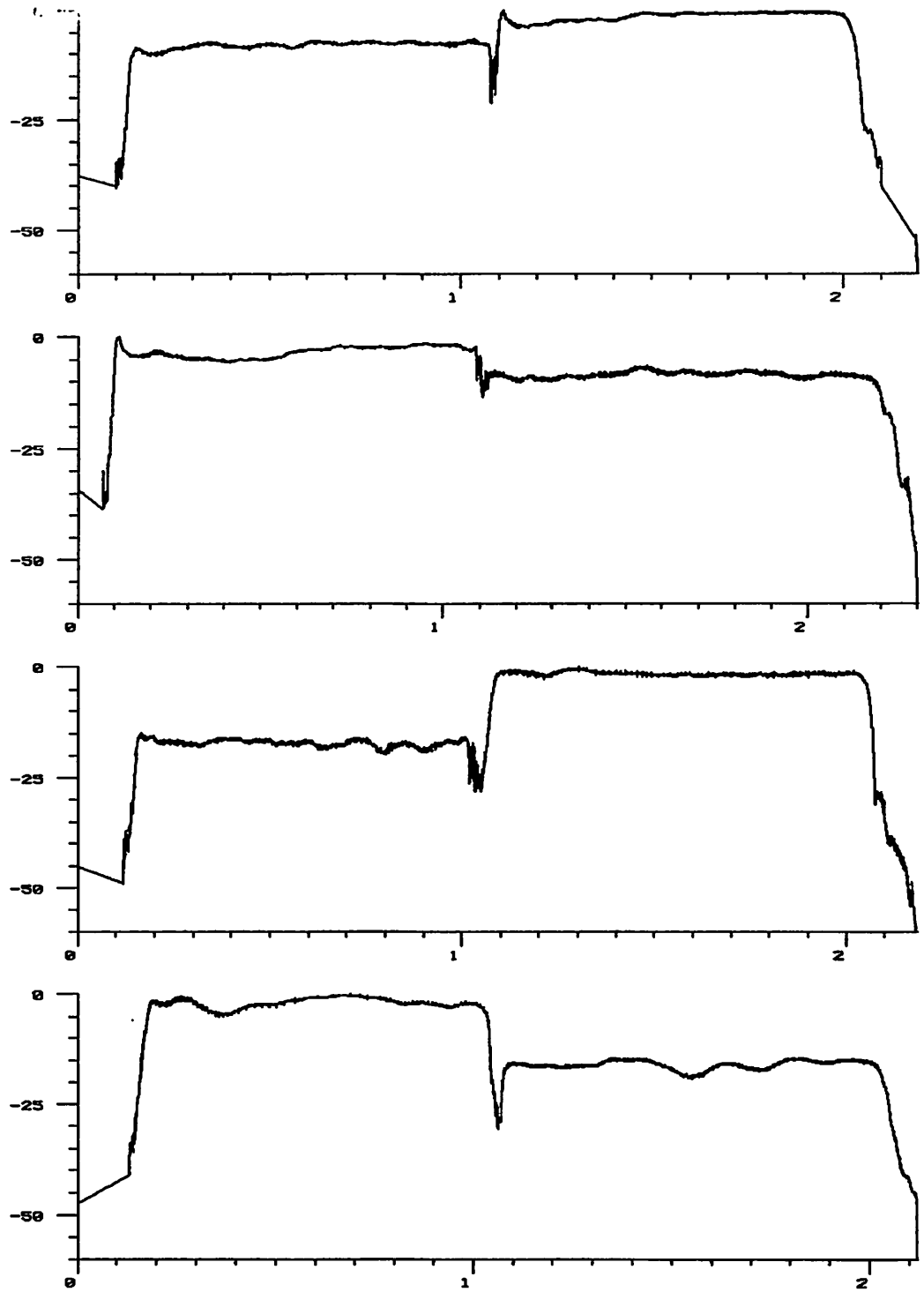


Figure A4.2. Time-varying analysis of four untongued piccolo intervals. From the top: major third ascending, major third descending, minor seventh ascending, minor seventh descending. The lower note in all four plots is A1760.

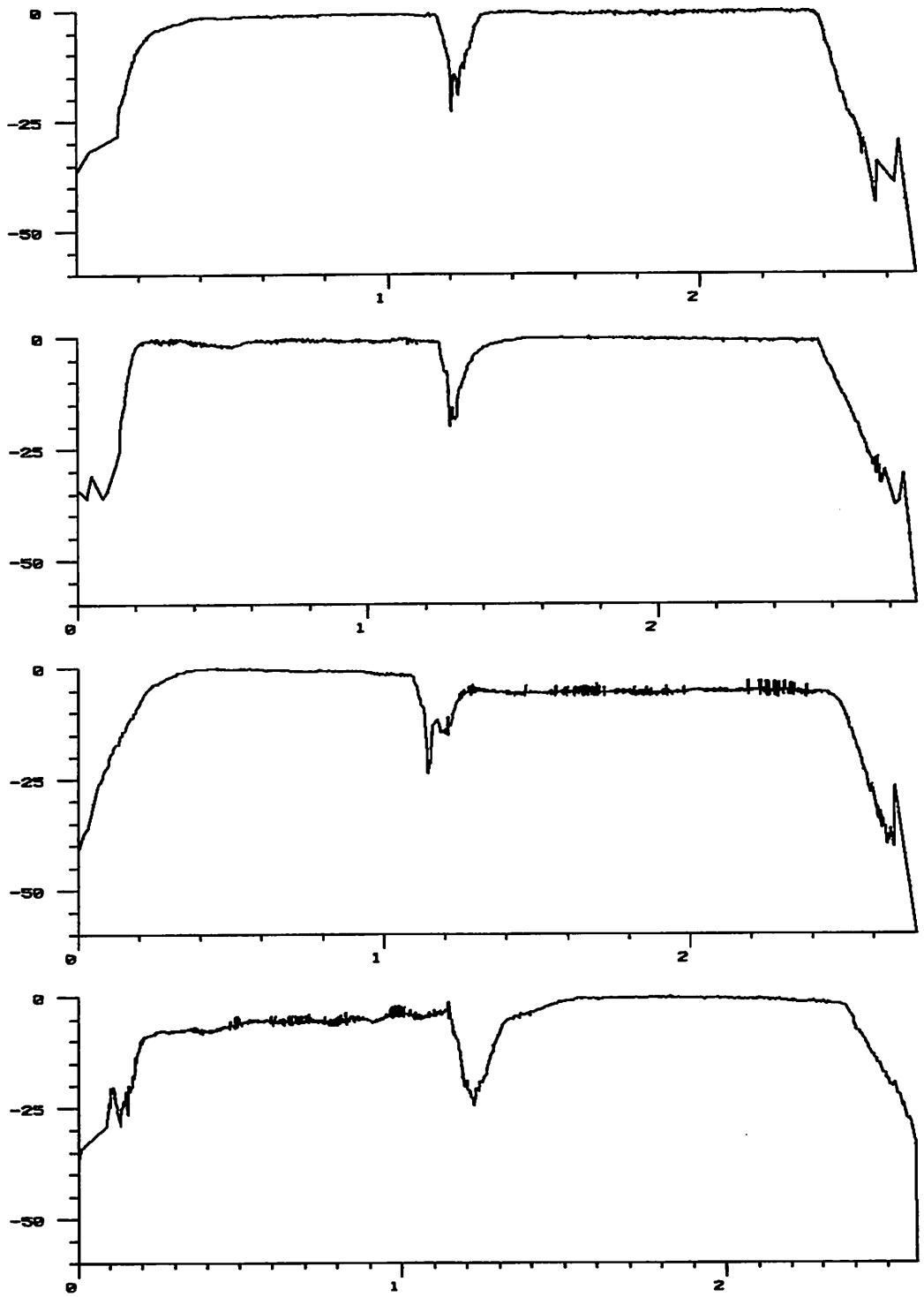


Figure A4.3. Time-varying analysis of four tongued bass flute intervals. From the top: major third ascending, major third descending, minor seventh ascending, minor seventh descending. The lower note in all four plots is A220.

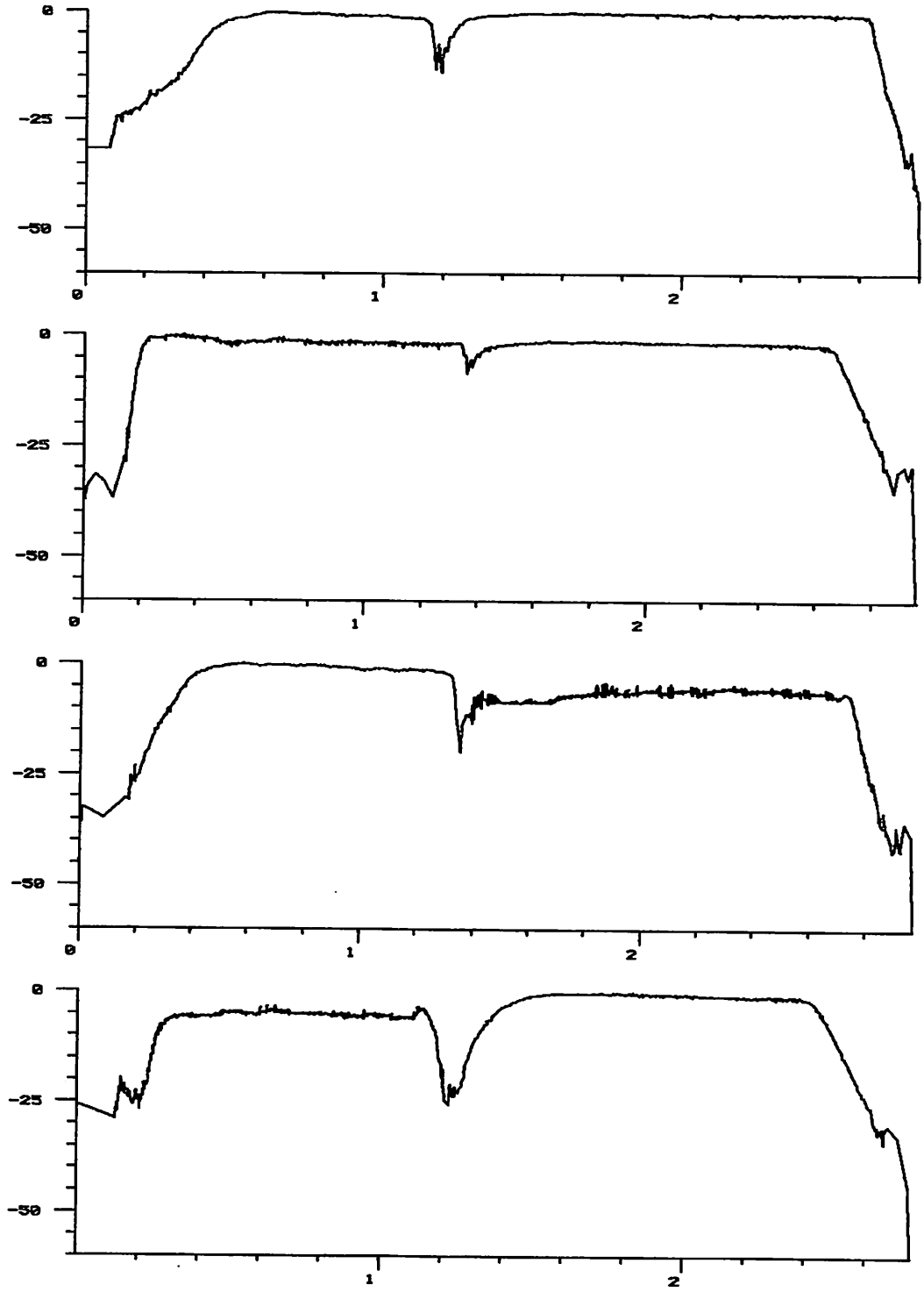


Figure A4.4. Time-varying analysis of four untongued bass flute intervals. From the top: major third ascending, major third descending, minor seventh ascending, minor seventh descending. The lower note in all four plots is A220.

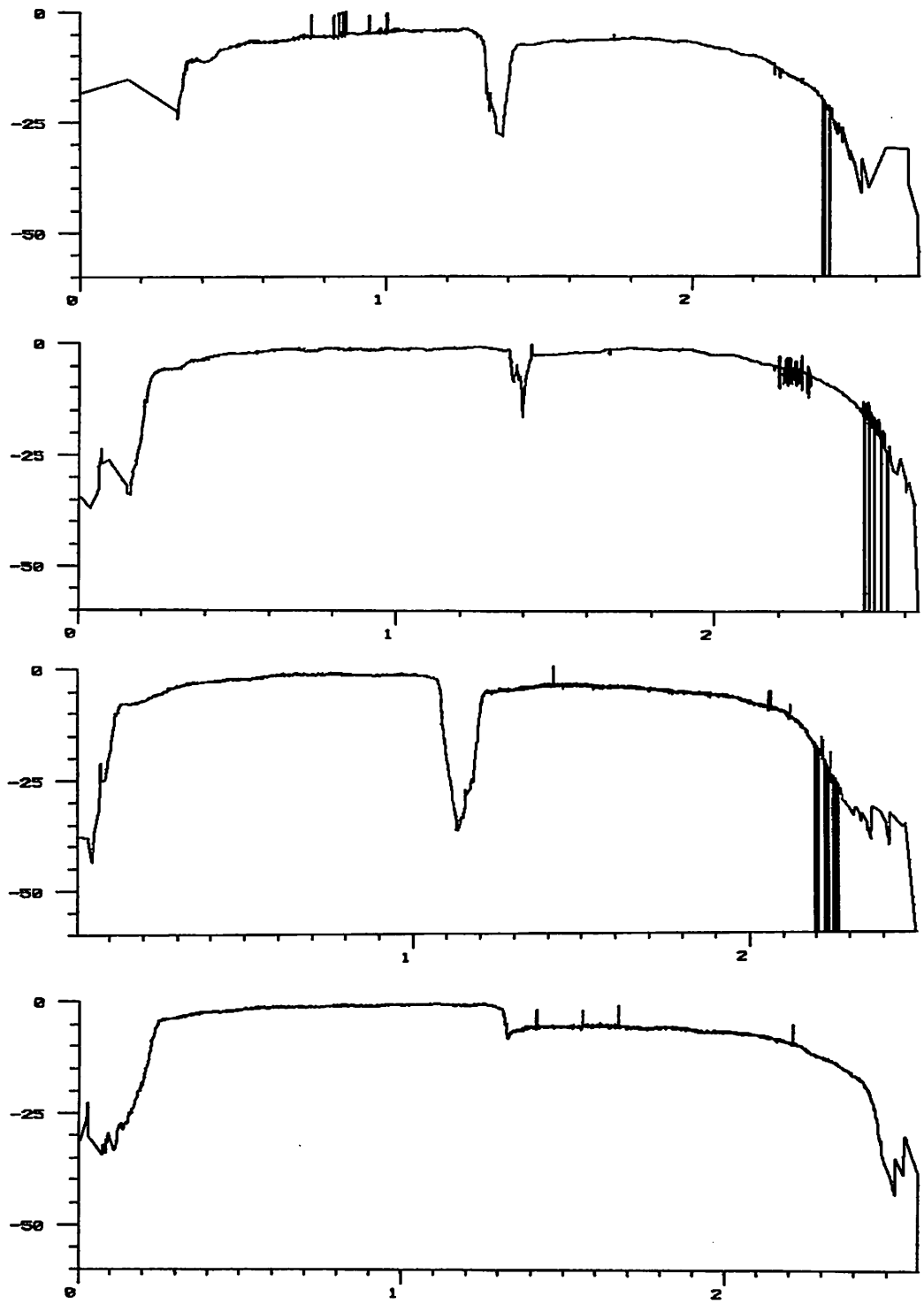


Figure A4.5. Time-varying analysis of four oboe intervals. From the top: tongued major third ascending, untongued major third ascending, tongued minor seventh ascending, untongued minor seventh ascending. The lower note in all four plots is A440.

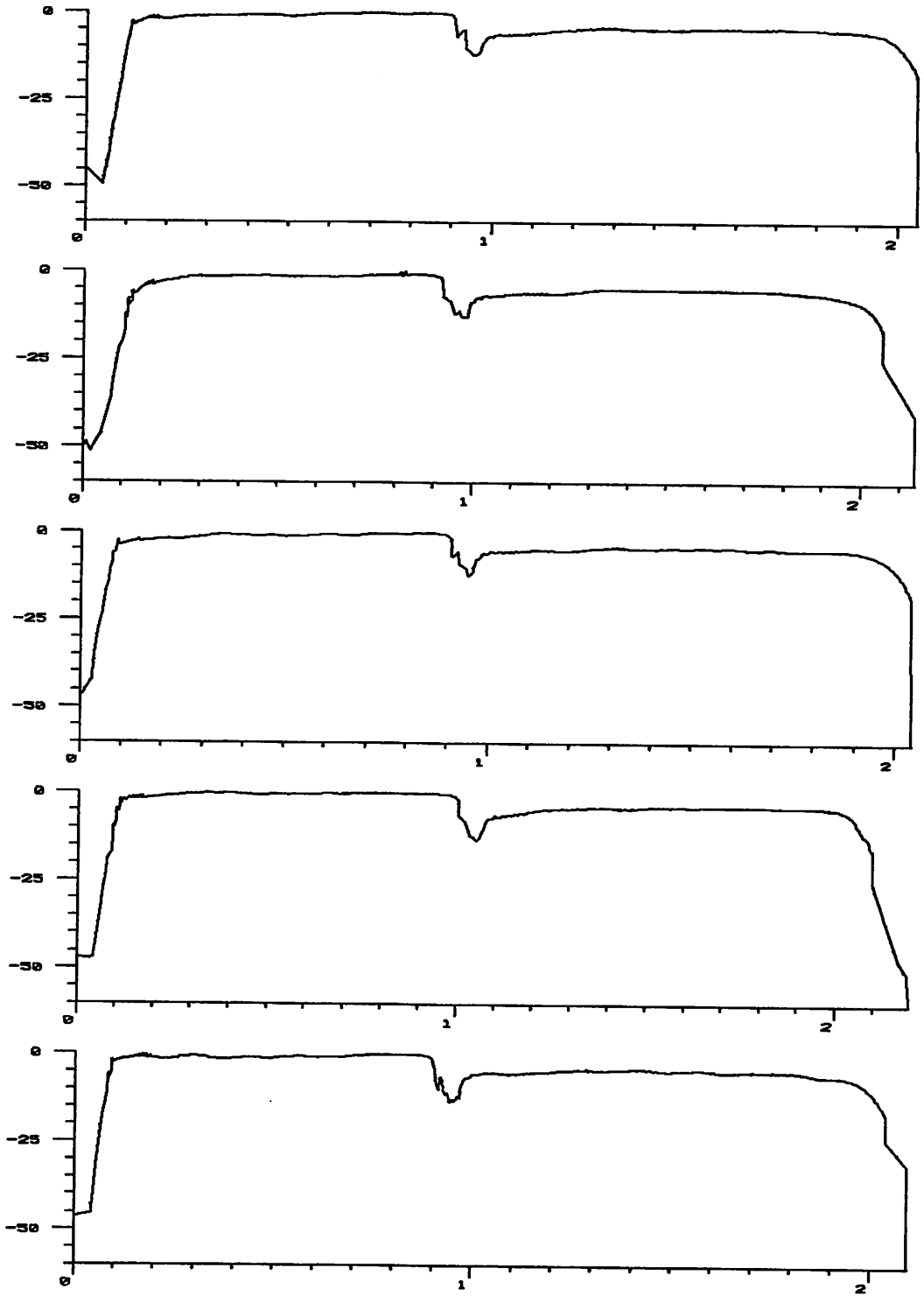


Figure A4.6. Five time-varying analyses of an untongued ascending major third played on the bassoon. The lower note in all five plots is A220.

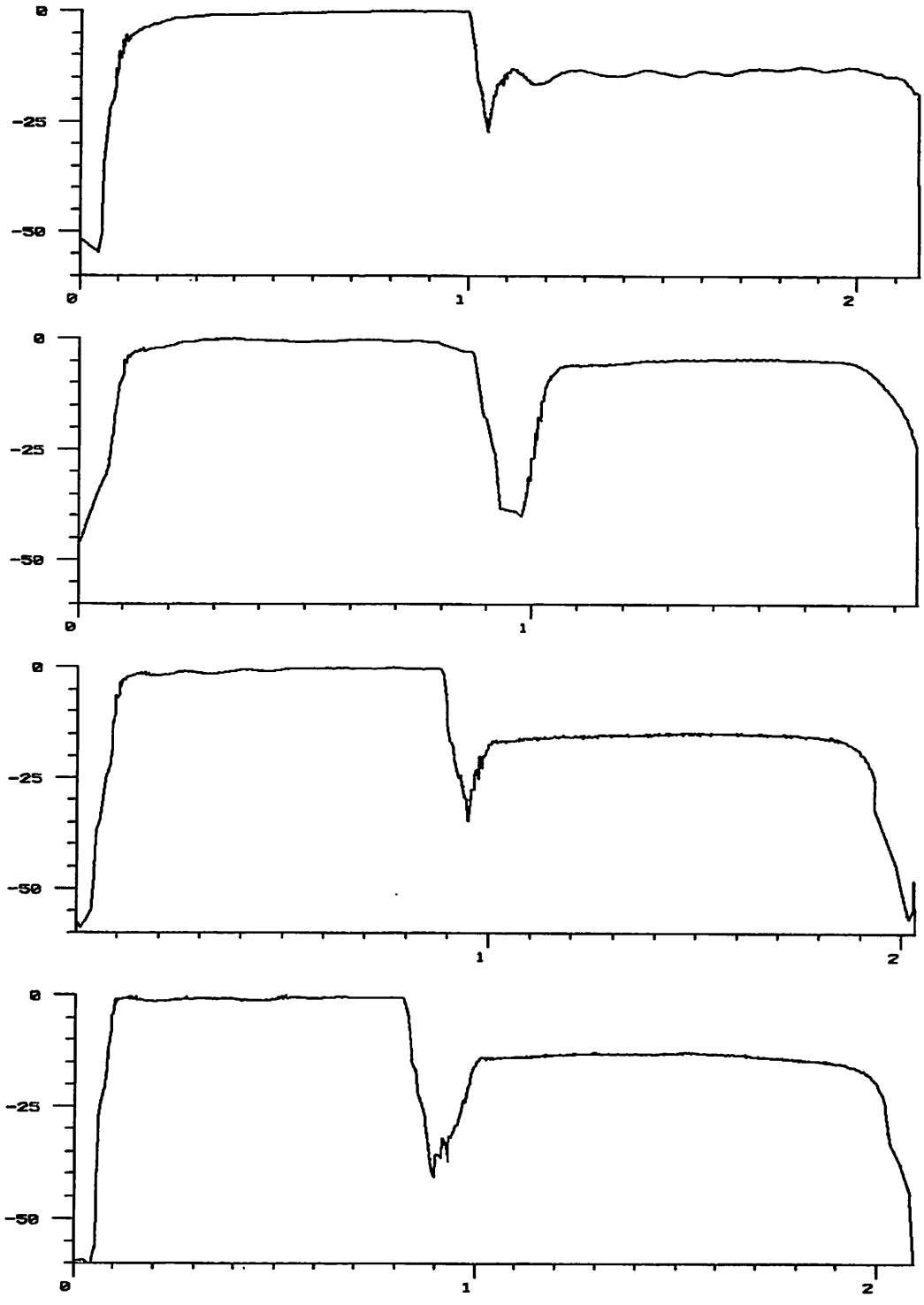


Figure A4.7. Time-varying analysis of four tongued bassoon intervals. From the top: major second ascending, major third ascending, perfect fifth ascending, minor seventh ascending. The lower note in all four plots is A220.

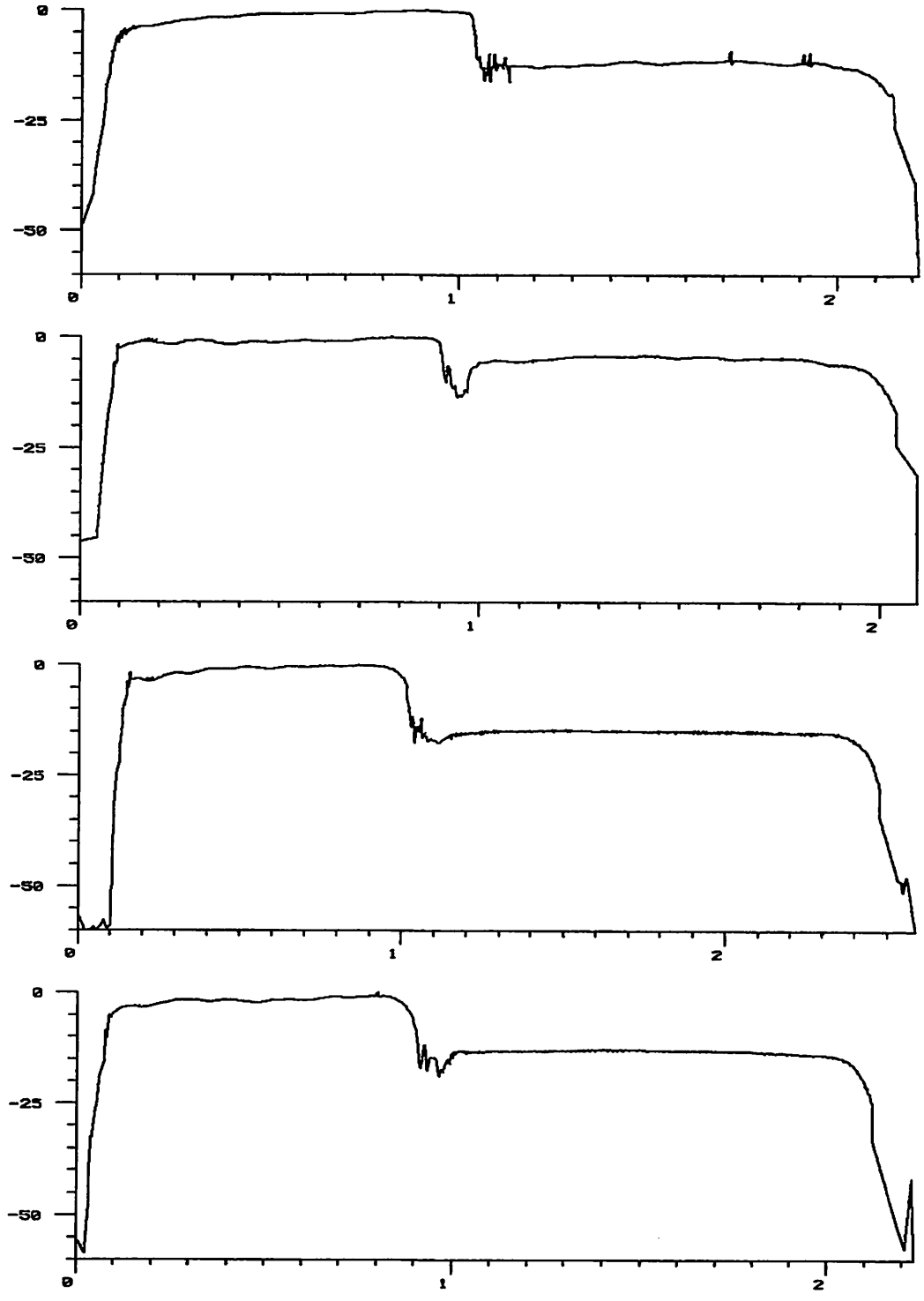


Figure A4.8. Time-varying analysis of four untongued bassoon intervals. From the top: major second ascending, major third ascending, perfect fifth ascending, minor seventh ascending. The lower note in all four plots is A220. The ascending third here is a duplicate of the bottom plot in Figure A4.6.

APPENDIX 5

A LEXICON OF ANALYZED TRANSITIONS, PART 2:

TIME-VARYING SPECTRAL PLOTS

Amplitude Plots:

Frequency plots:

Preface

Appendix 4 gives an introduction to the nature and purpose of this appendix; the remarks there apply here as well.

Three-dimensional analyses of time-varying spectra are discussed in Chapter 2. The amplitude scale covers a range of 60 dB for the harmonics in each plot. All of the spectral plots cover a time range of 300 msec. The number of harmonics varies from note to note; the cutoff point was the harmonic whose maximum amplitude in the steady-state never exceeded −60 dB. The number of harmonics was rounded off to a happy medium for all of the recordings of an instrument playing the same interval. The caption for each figure lists the number of harmonics.

Three plots of frequency traces during the transition are included at the end of the appendix. (Some problems with analyzing such plots were already discussed in Chapter 2.) In general, the frequency trace at the end of the first note anticipates the jump in frequency well before the pitch shifts. The amount of anticipation is puzzling, as the analysis window was only about 20 msec long. In some of the cases, the trace for the second harmonic gives a better idea of what the frequency is doing than the trace for the fundamental.

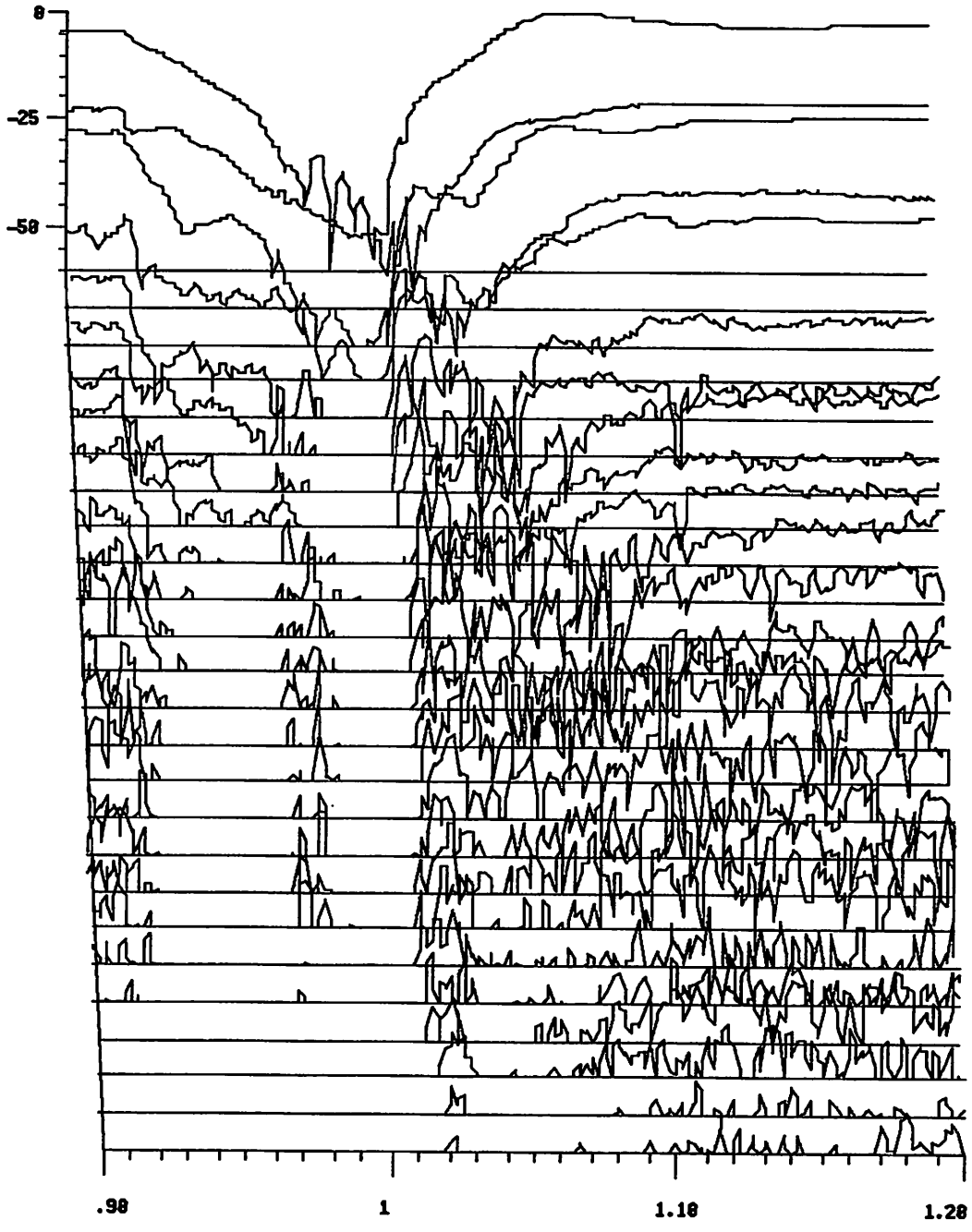


Figure A5.1. Time-varying spectral analysis of a tongued descending major third played on the flute. The lower note is A220; 25 harmonics are shown.

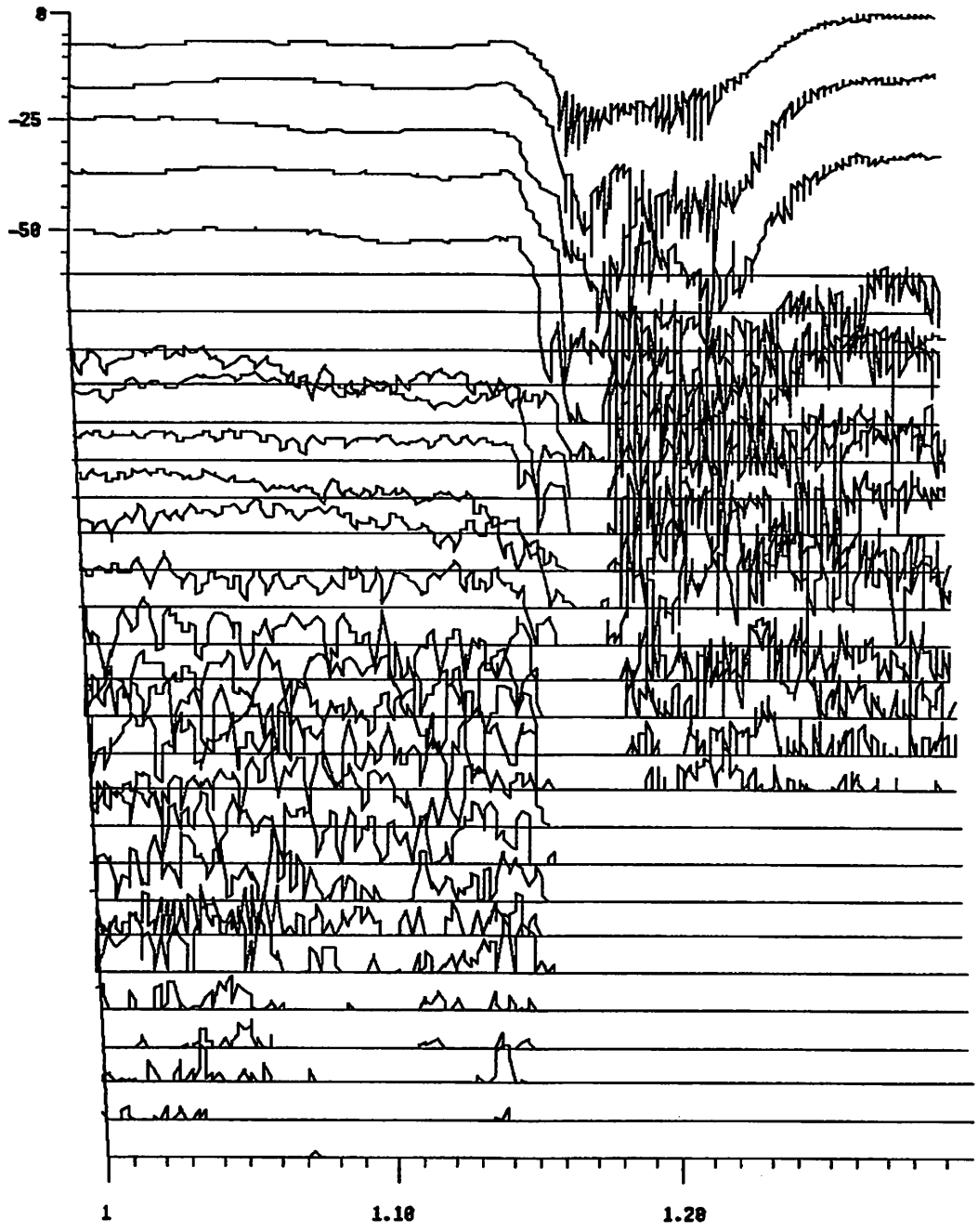


Figure A5.2. Time-varying spectral analysis of a tongued ascending minor seventh played on the flute. The lower note is A220; 25 harmonics are shown.

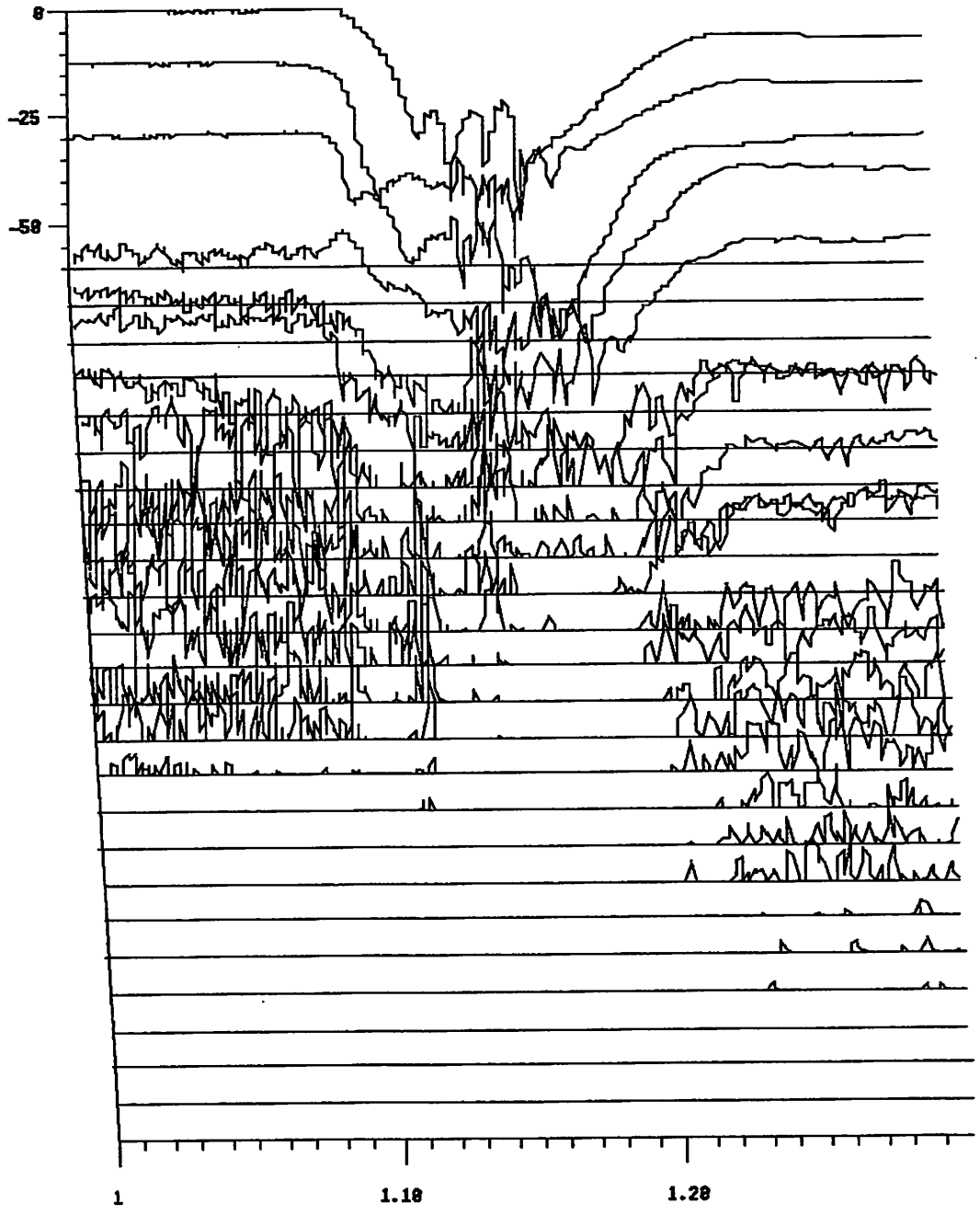


Figure A5.3. Time-varying spectral analysis of a tongued descending minor seventh played on the flute. The lower note is A220; 25 harmonics are shown.

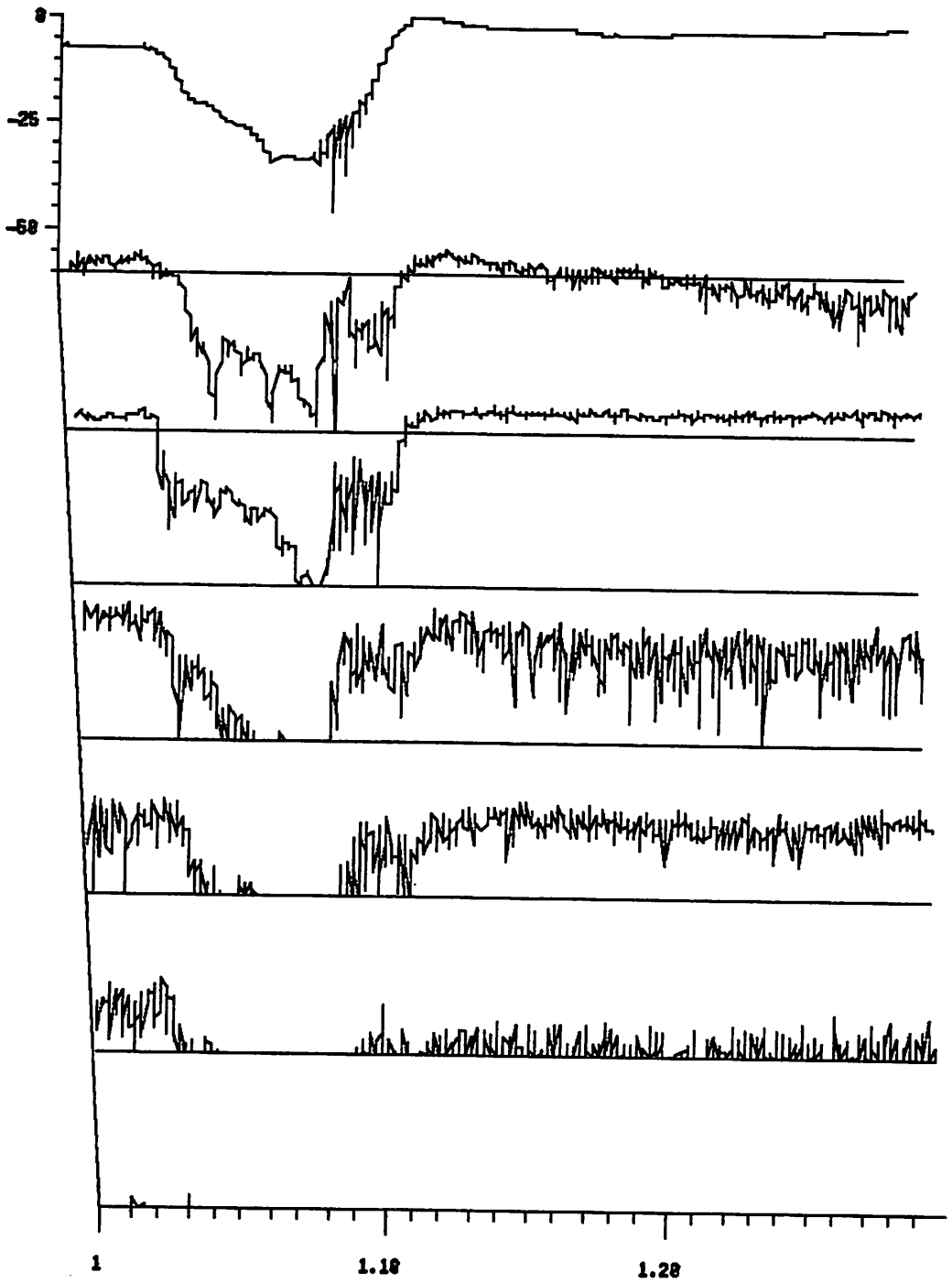


Figure A5.4. Time-varying spectral analysis of a tongued ascending major third played on the piccolo. The lower note is A1760; 7 harmonics are shown.

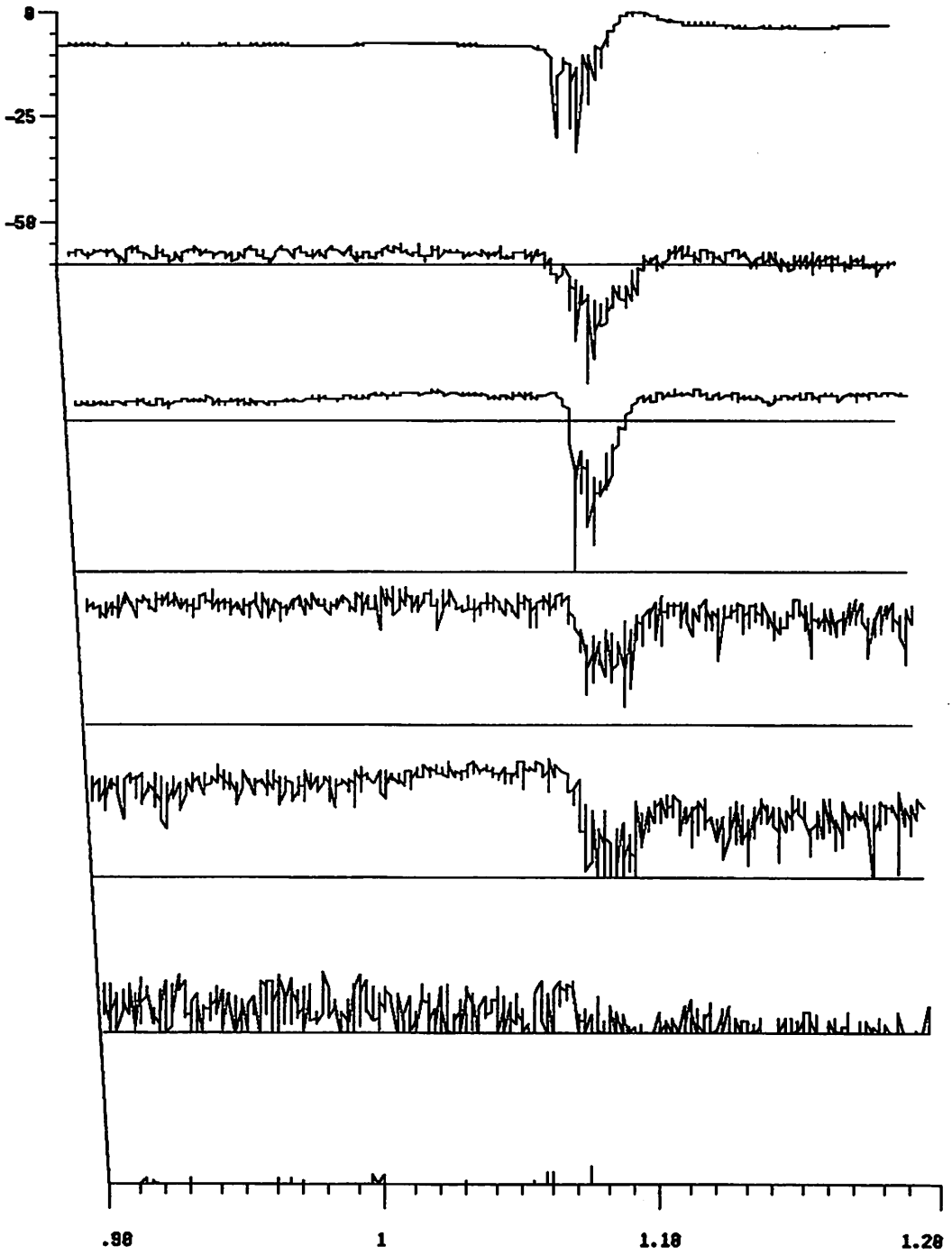


Figure A5.5. Time-varying spectral analysis of an untongued ascending major third played on the piccolo. The lower note is A1760; 7 harmonics are shown.

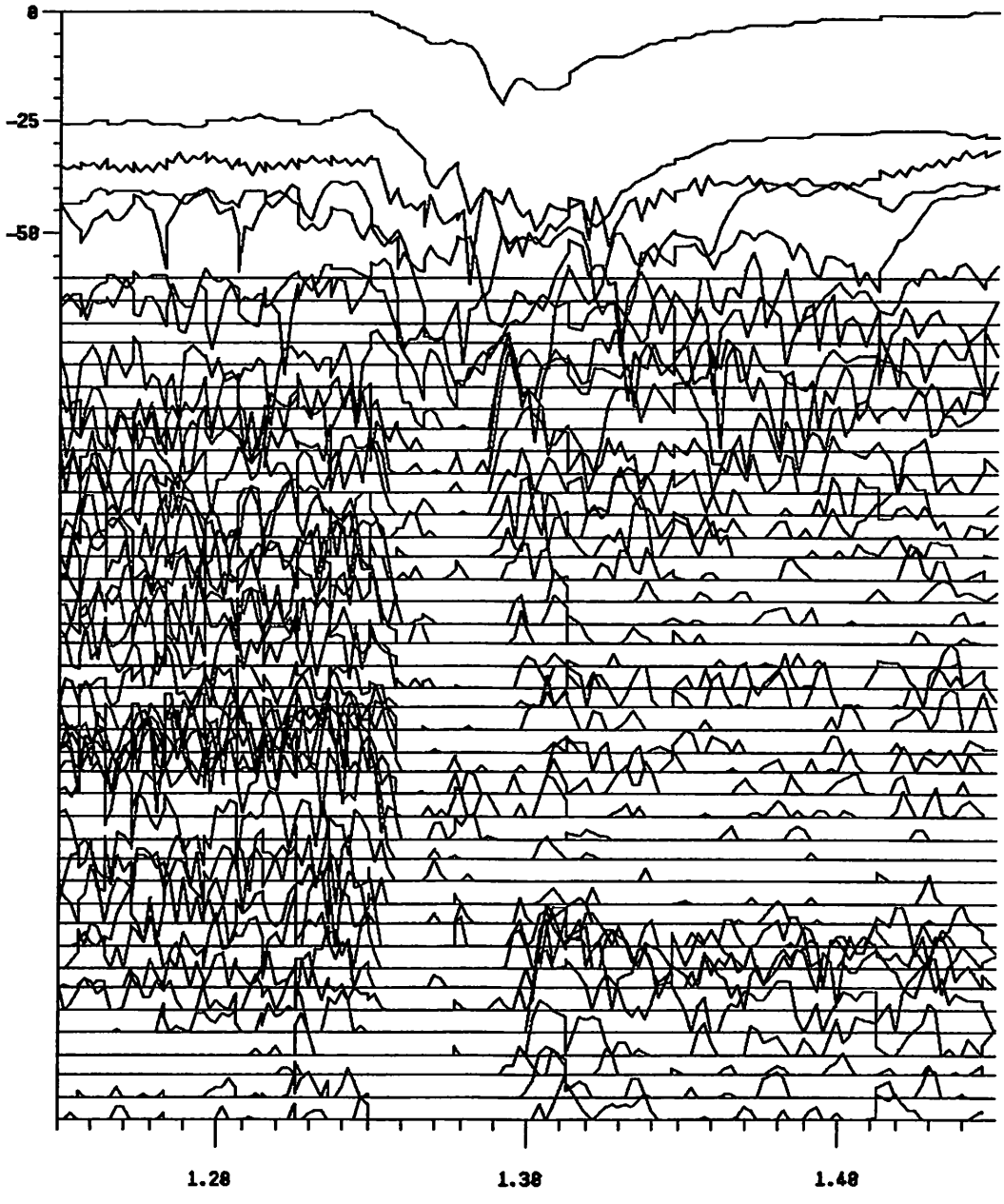


Figure A5.6. Time-varying spectral analysis of a tongued descending major third played on the bass flute. The lower note is A220; 40 harmonics are shown.

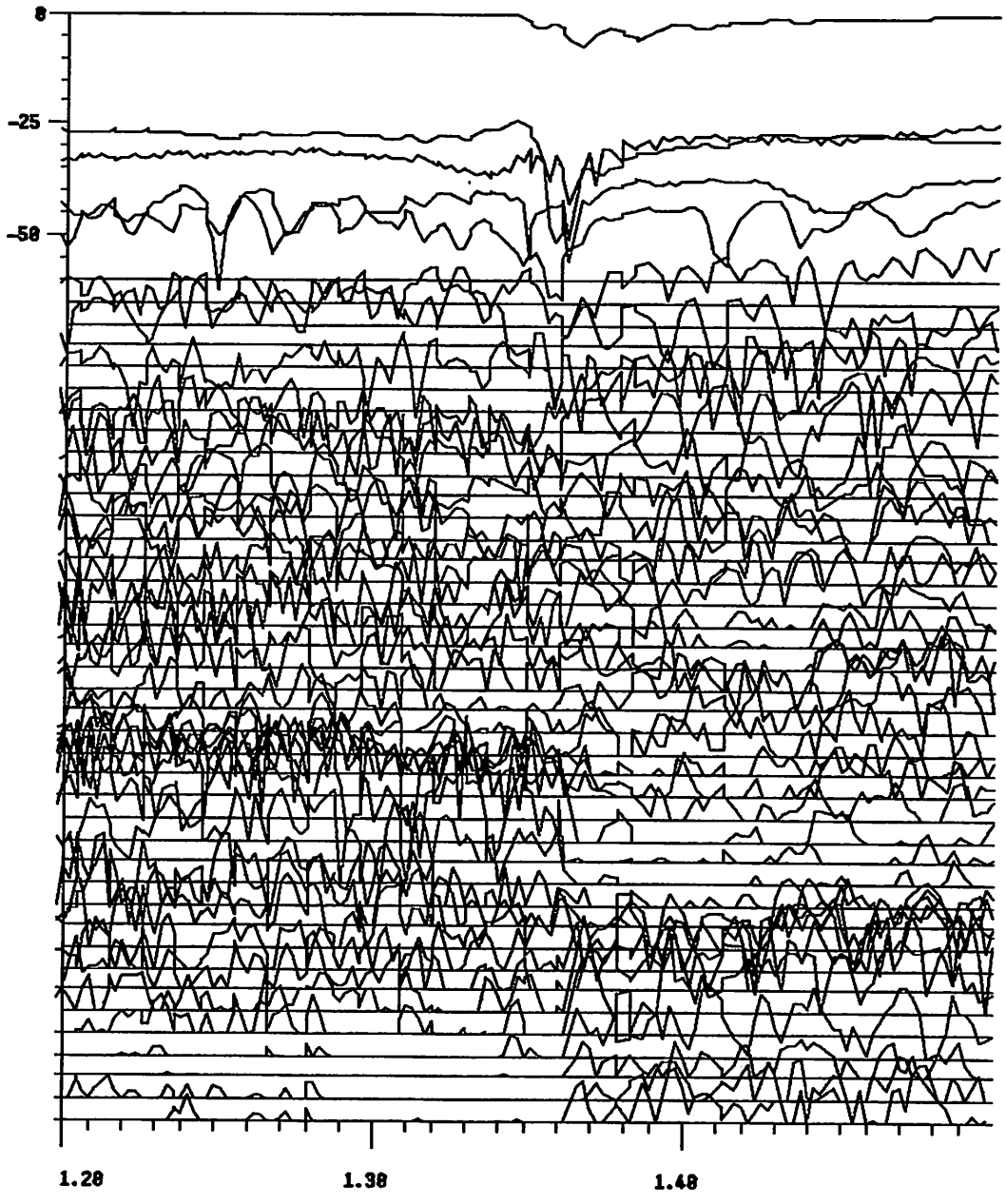


Figure A5.7. Time-varying spectral analysis of an untongued descending major third played on the bass flute. The lower note is A220; 40 harmonics are shown.

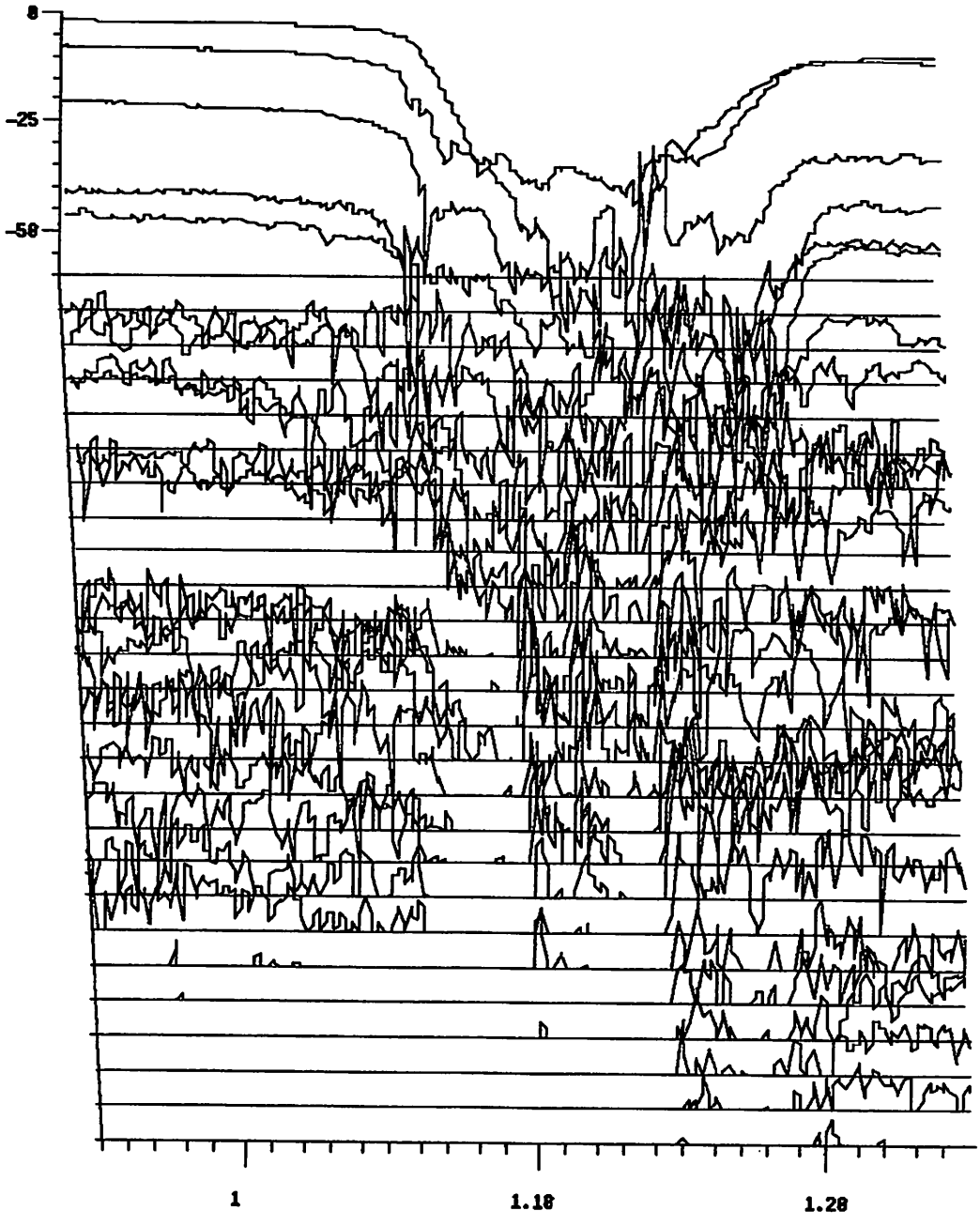


Figure A5.8. Time-varying spectral analysis of a tongued descending major third played on the oboe. The lower note is A440; 26 harmonics are shown.

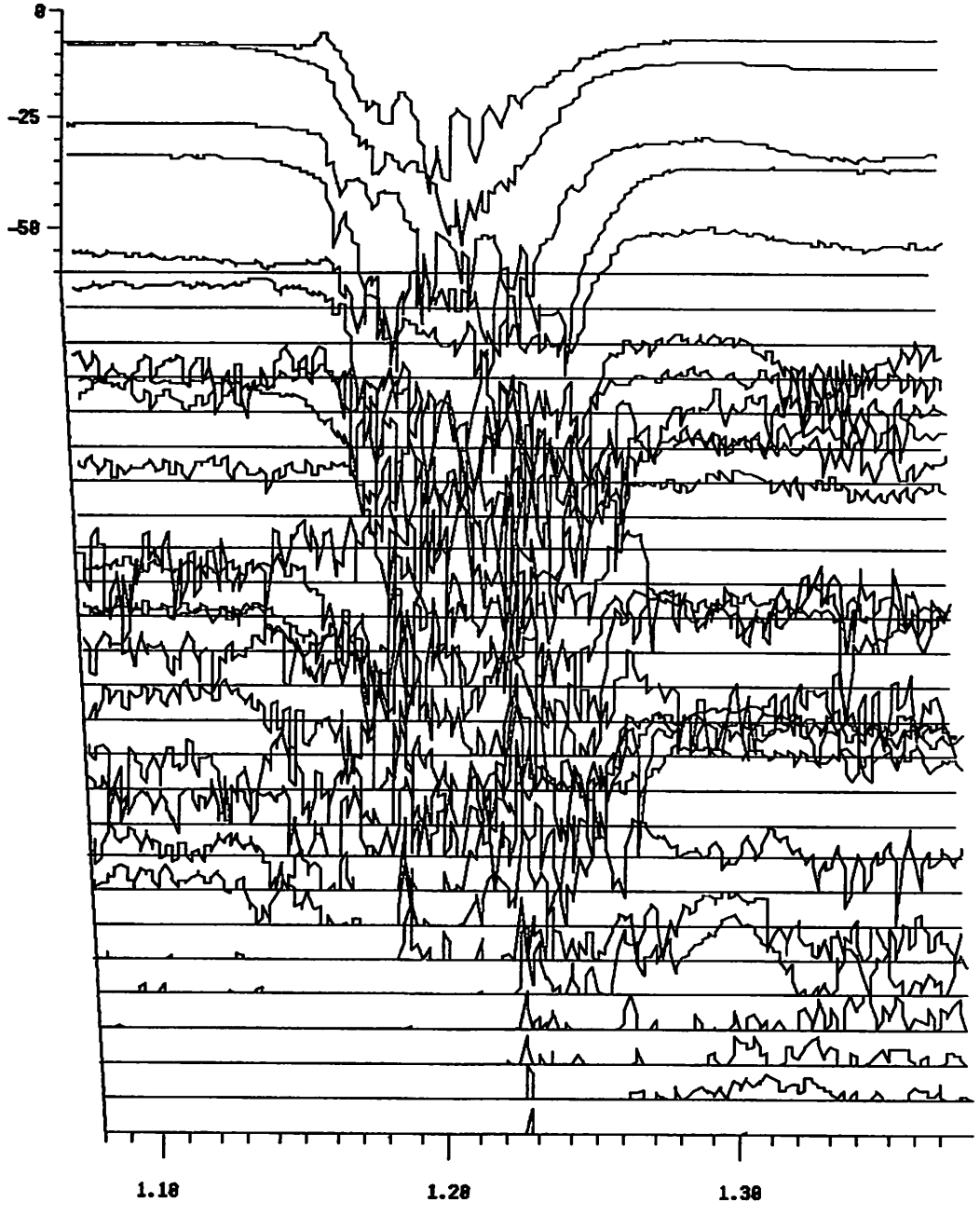


Figure A5.9. Time-varying spectral analysis of an untongued descending major third played on the oboe. The lower note is A440; 26 harmonics are shown.

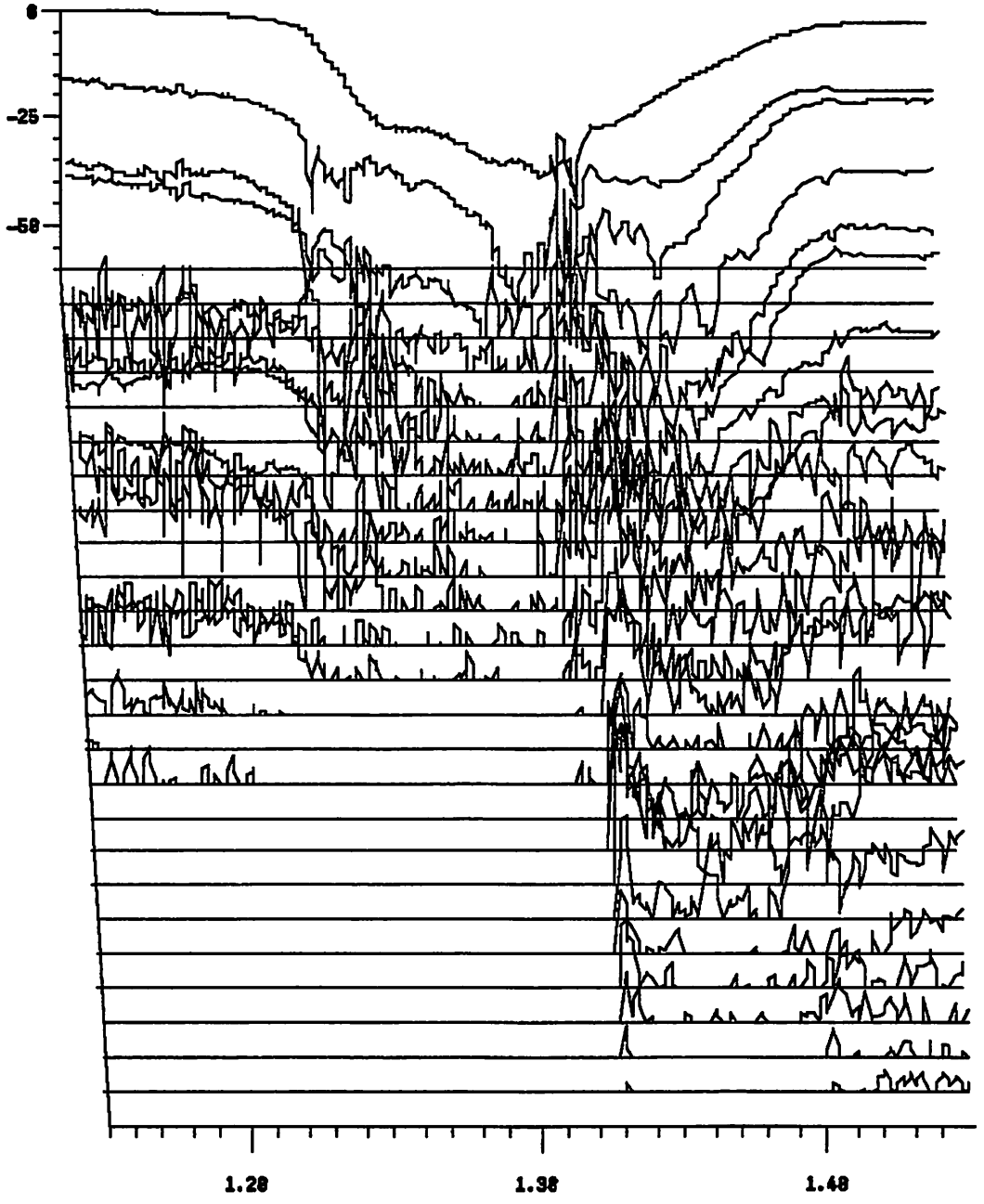


Figure A5.10. Time-varying spectral analysis of a tongued descending minor seventh played on the oboe. The lower note is A440; 26 harmonics are shown.

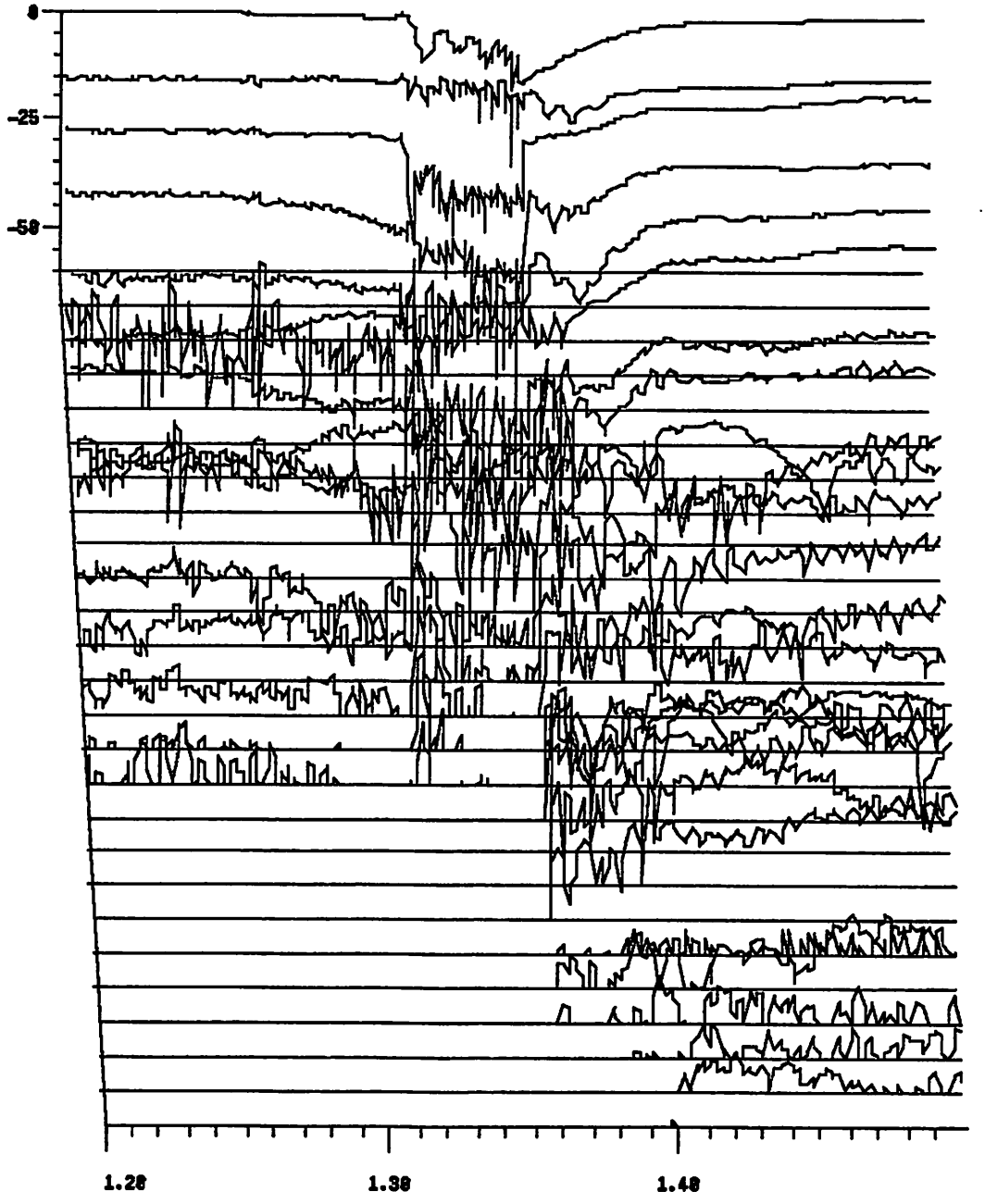


Figure A5.11. Time-varying spectral analysis of an untongued descending minor seventh played on the oboe. The lower note is A440; 26 harmonics are shown.

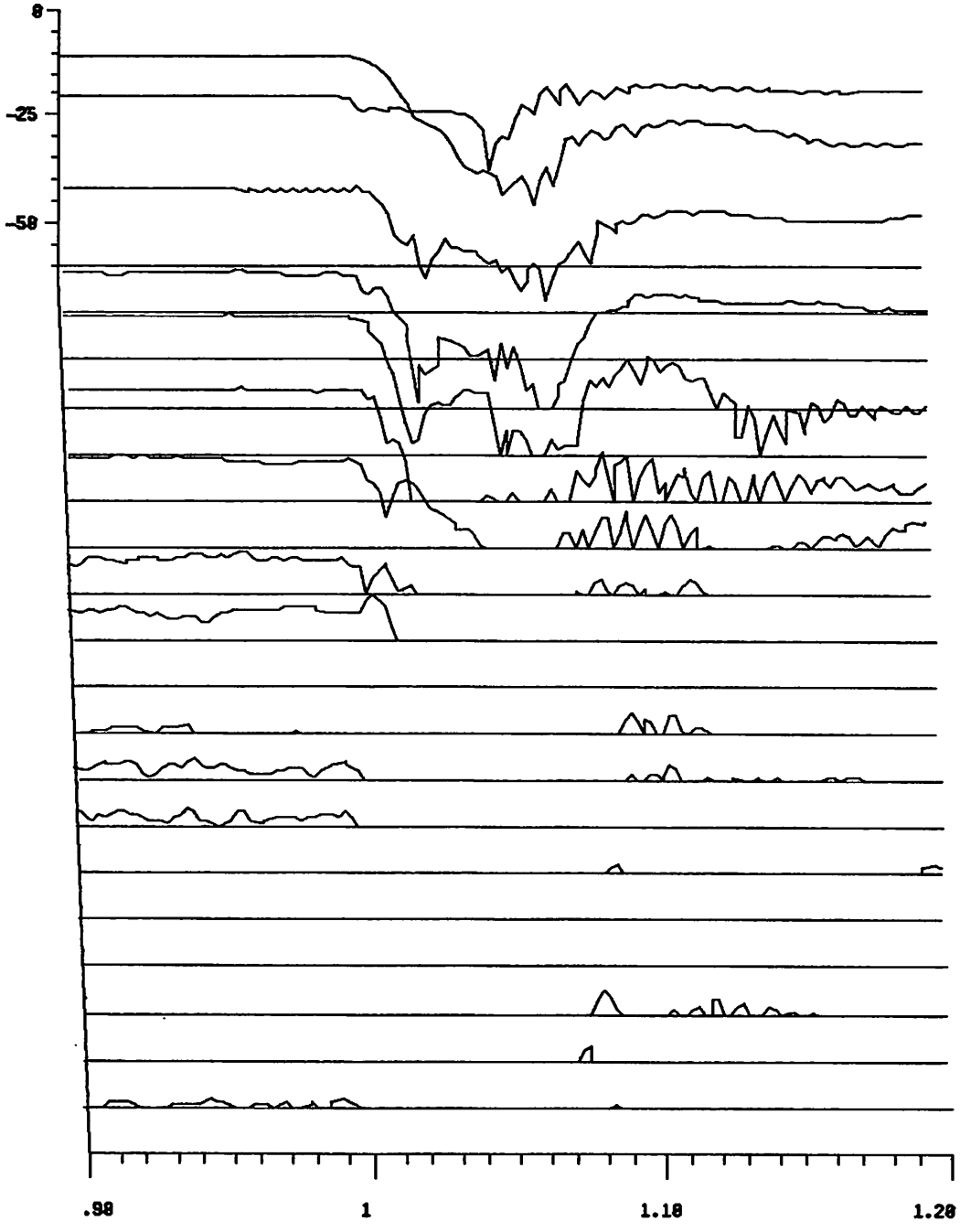


Figure A5.12. Time-varying spectral analysis of a tongued ascending major second played on the bassoon. The lower note is A220; 20 harmonics are shown.

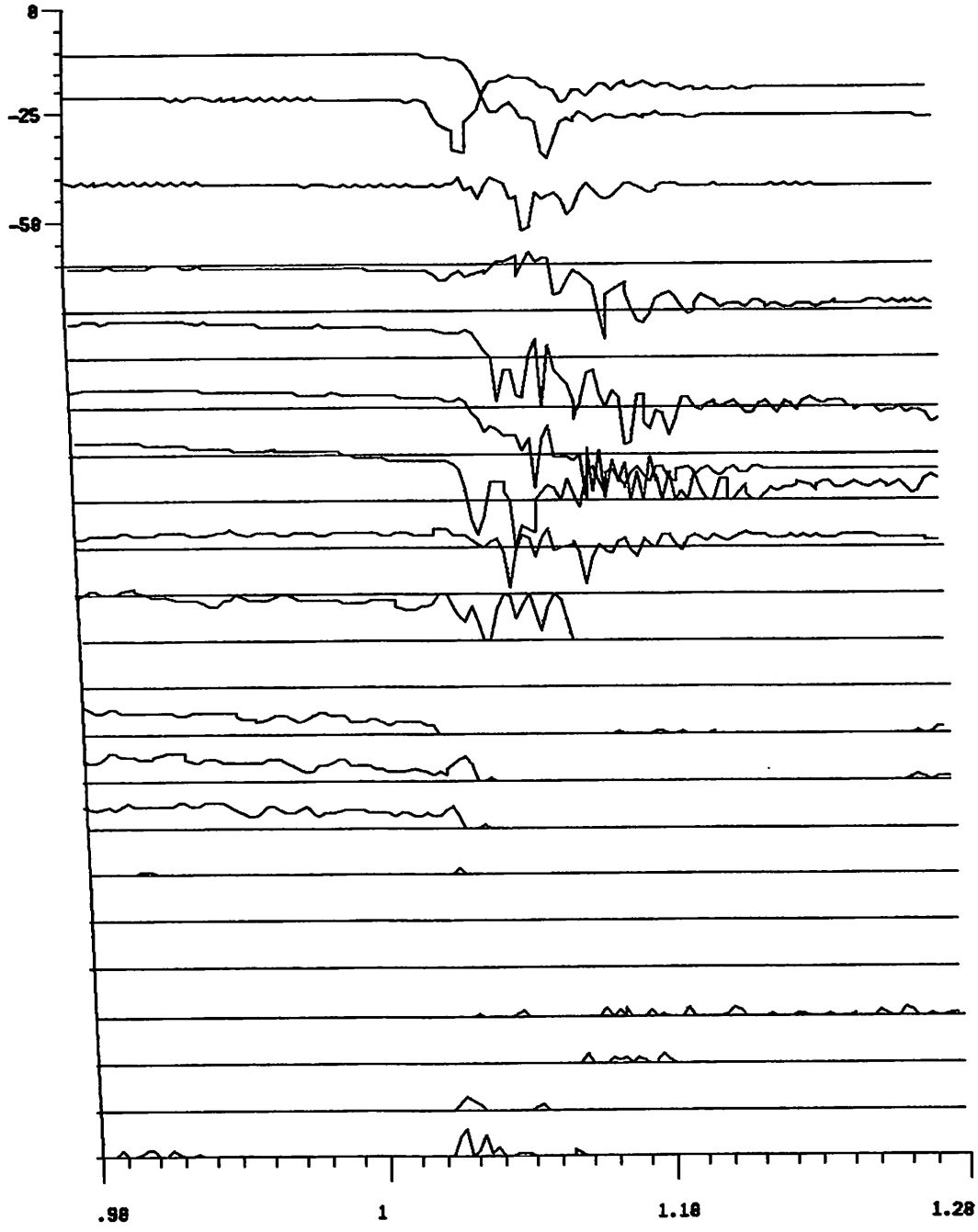


Figure A5.13. Time-varying spectral analysis of an untongued ascending major second played on the bassoon. The lower note is A; 20 harmonics are shown.

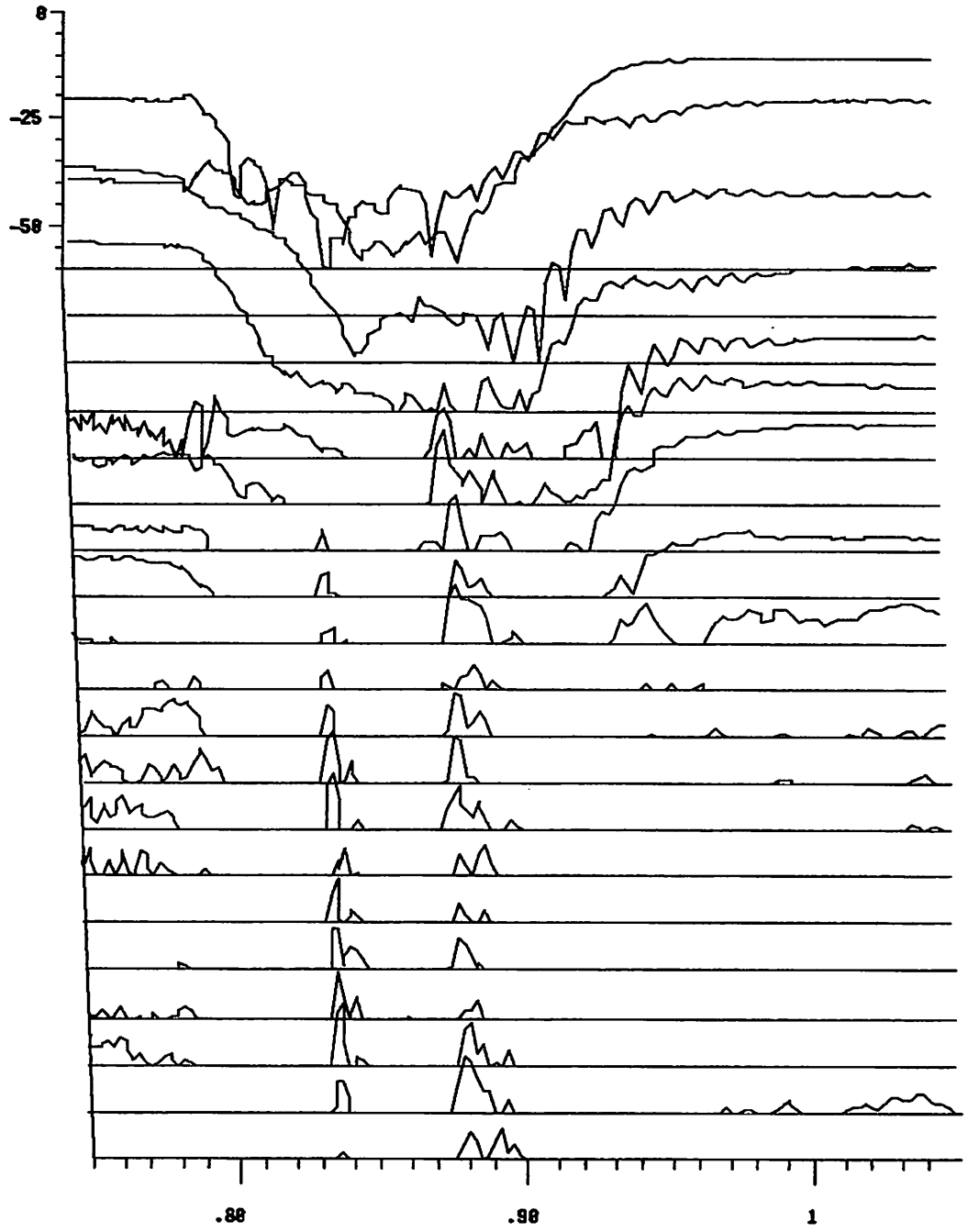


Figure A5.14. Time-varying spectral analysis of a tongued descending perfect fifth played on the bassoon. The lower note is A220; 20 harmonics are shown.

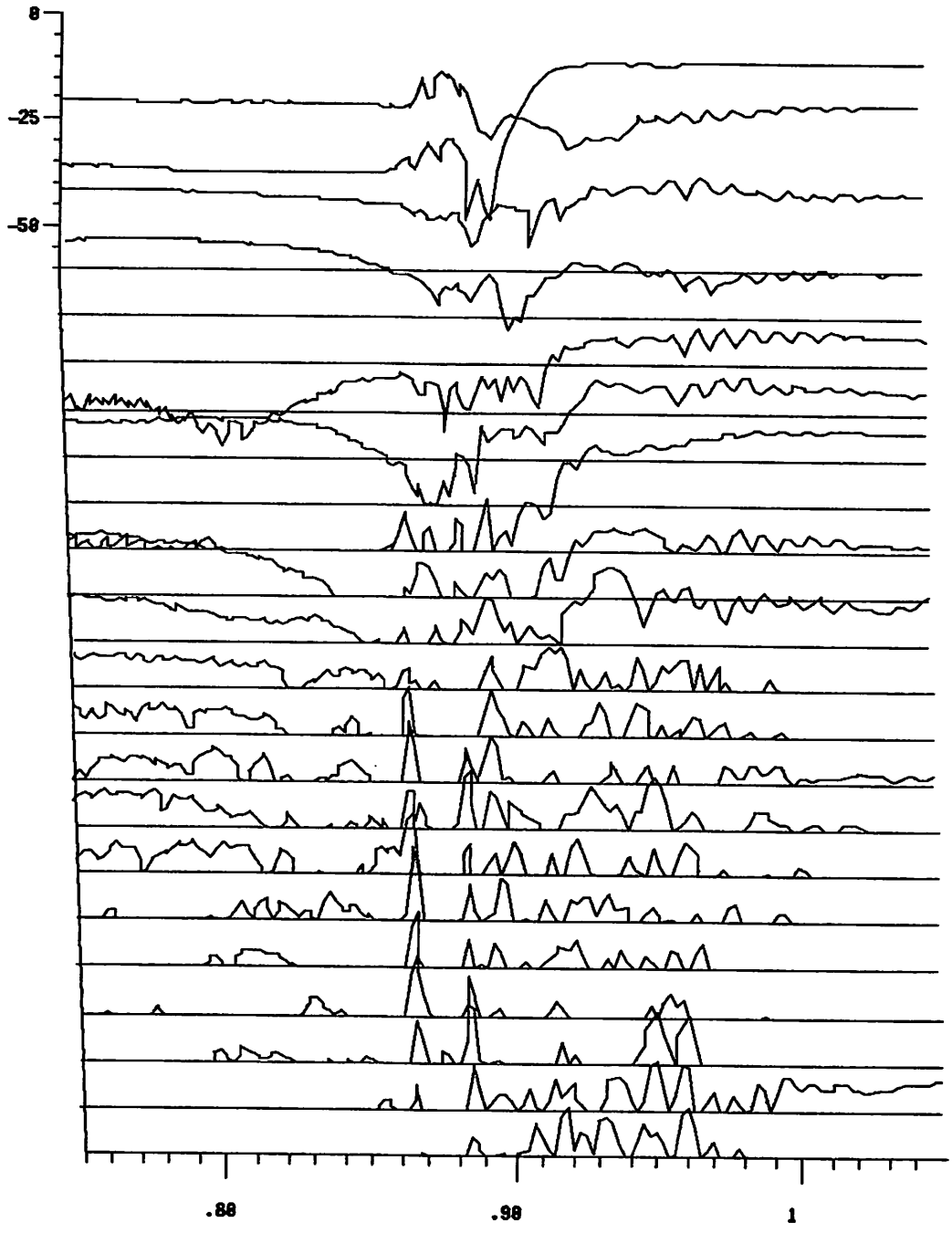


Figure A5.15. Time-varying spectral analysis of an untongued descending perfect fifth played on the bassoon. The lower note is A220; 20 harmonics are shown.

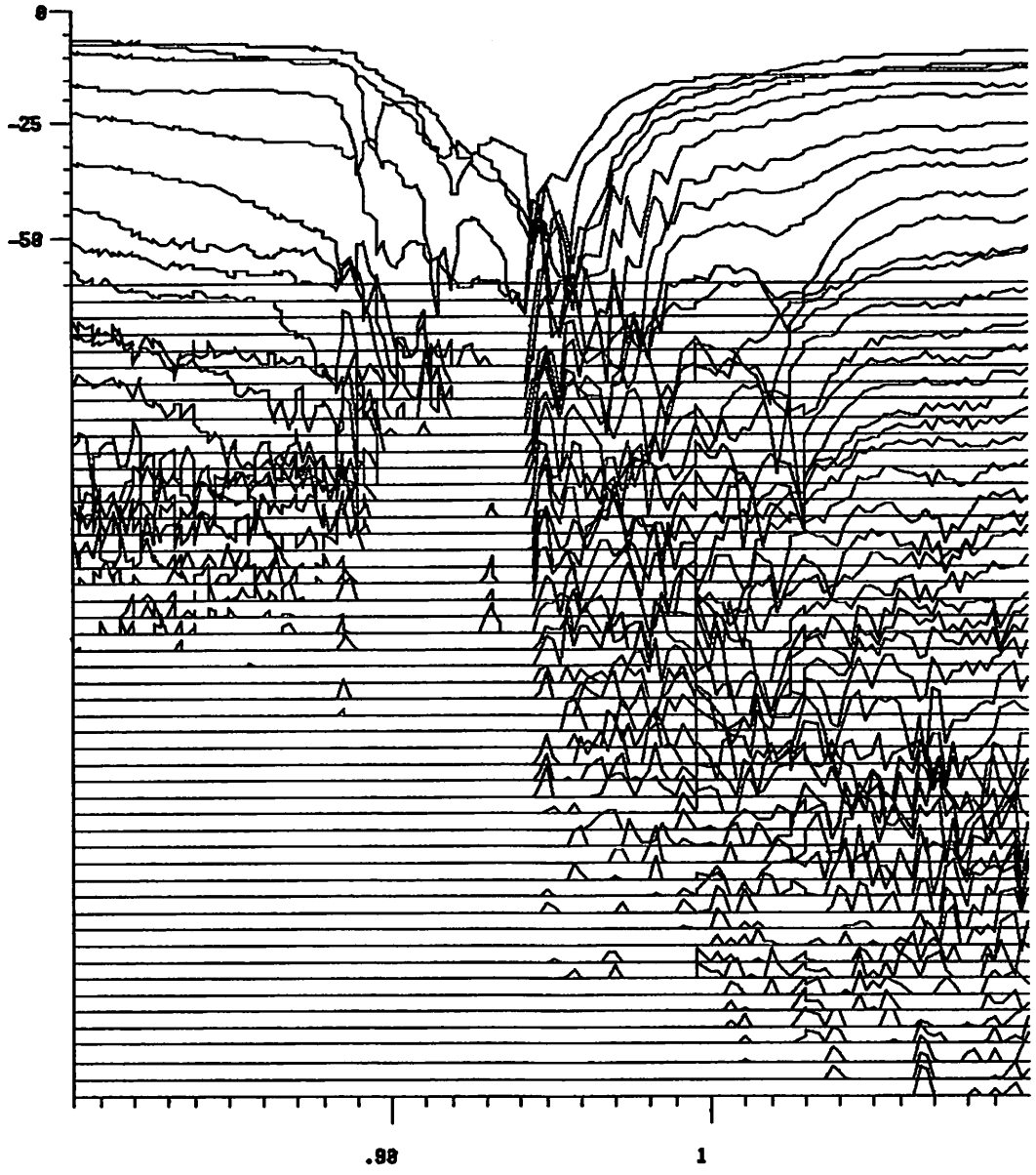


Figure A5.16. Time-varying spectral analysis of a tongued descending minor seventh played on the trumpet. The lower note is A220; 50 (!) harmonics are shown.

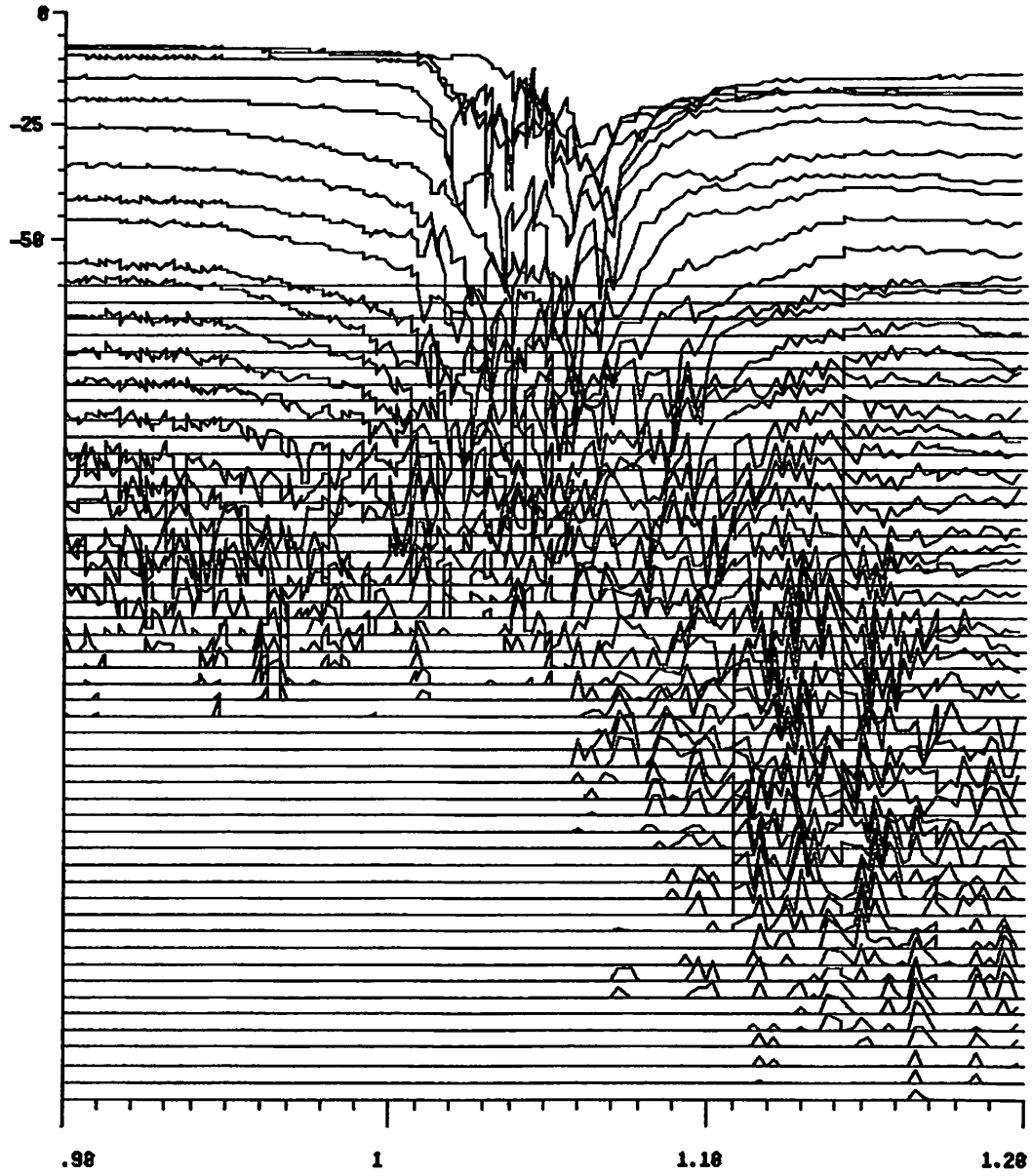


Figure A5.17. Time-varying spectral analysis of an untongued descending minor seventh played on the trumpet. The lower note is A220; 50 harmonics are shown.

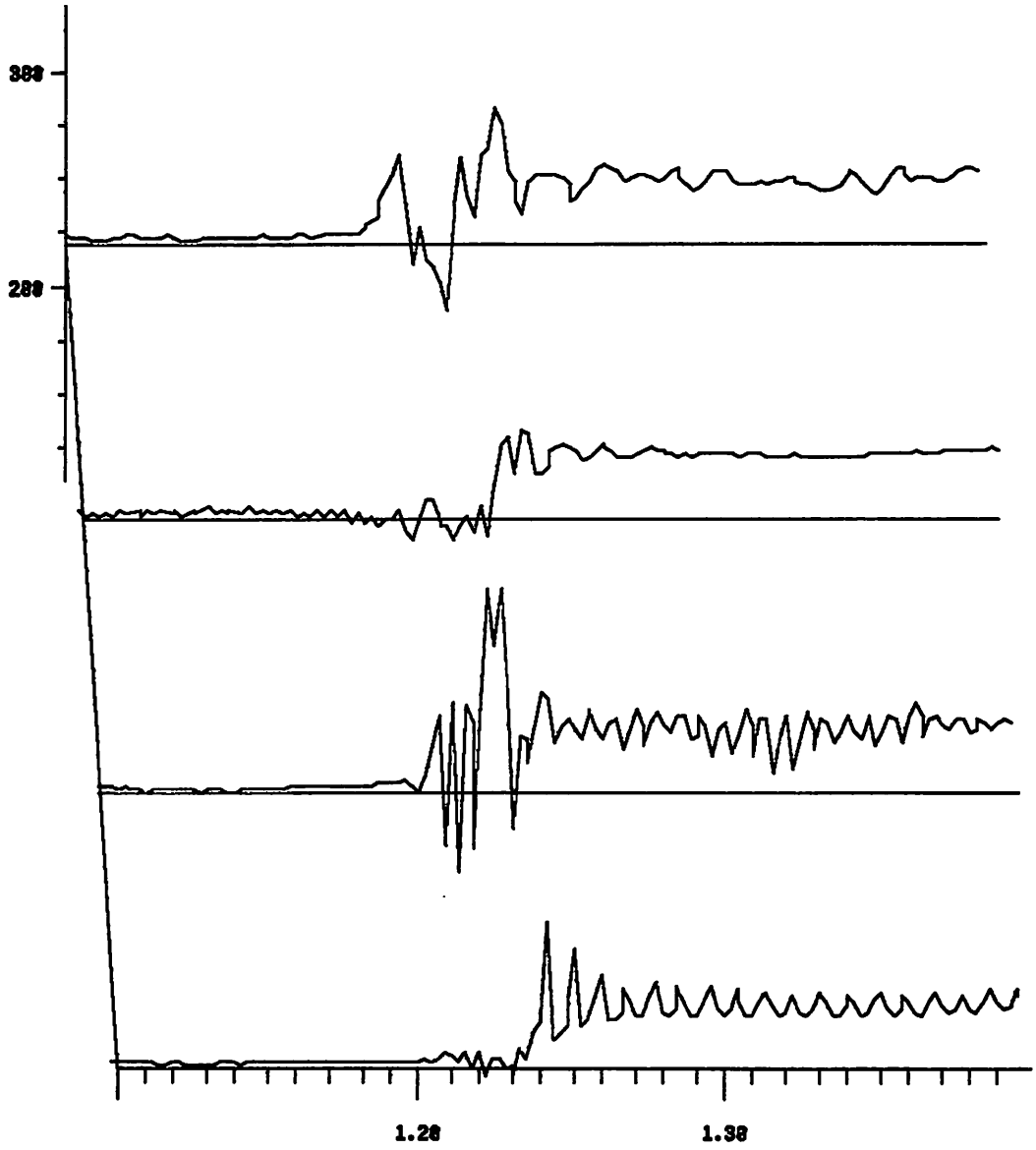


Figure A5.18. Time-varying frequencies of the first four harmonics of an ascending major second played with no bow change on the cello. The lower note is A220.

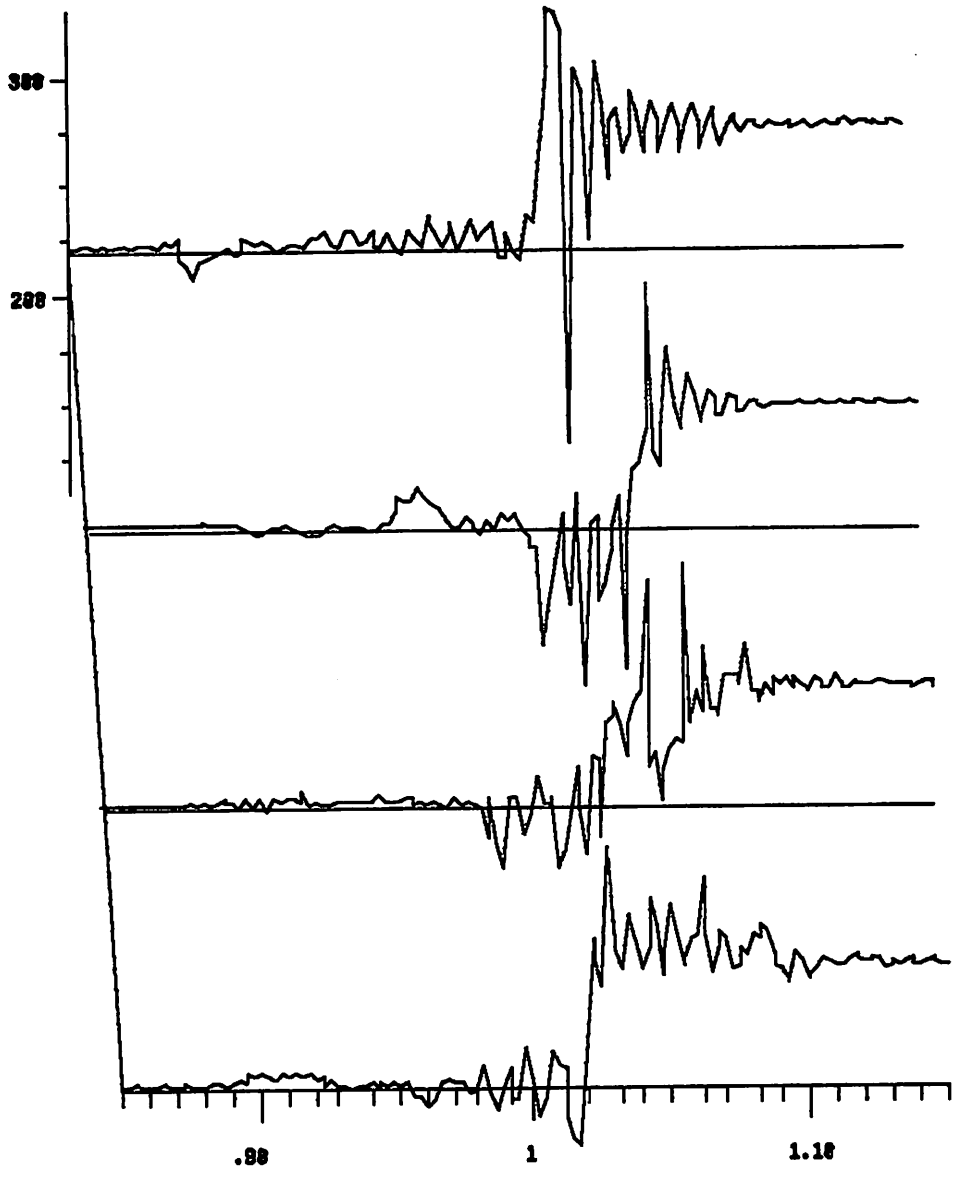


Figure A5.19. Time-varying frequencies of the first four harmonics of a tongued ascending major third played on the bassoon. The lower note is A220.

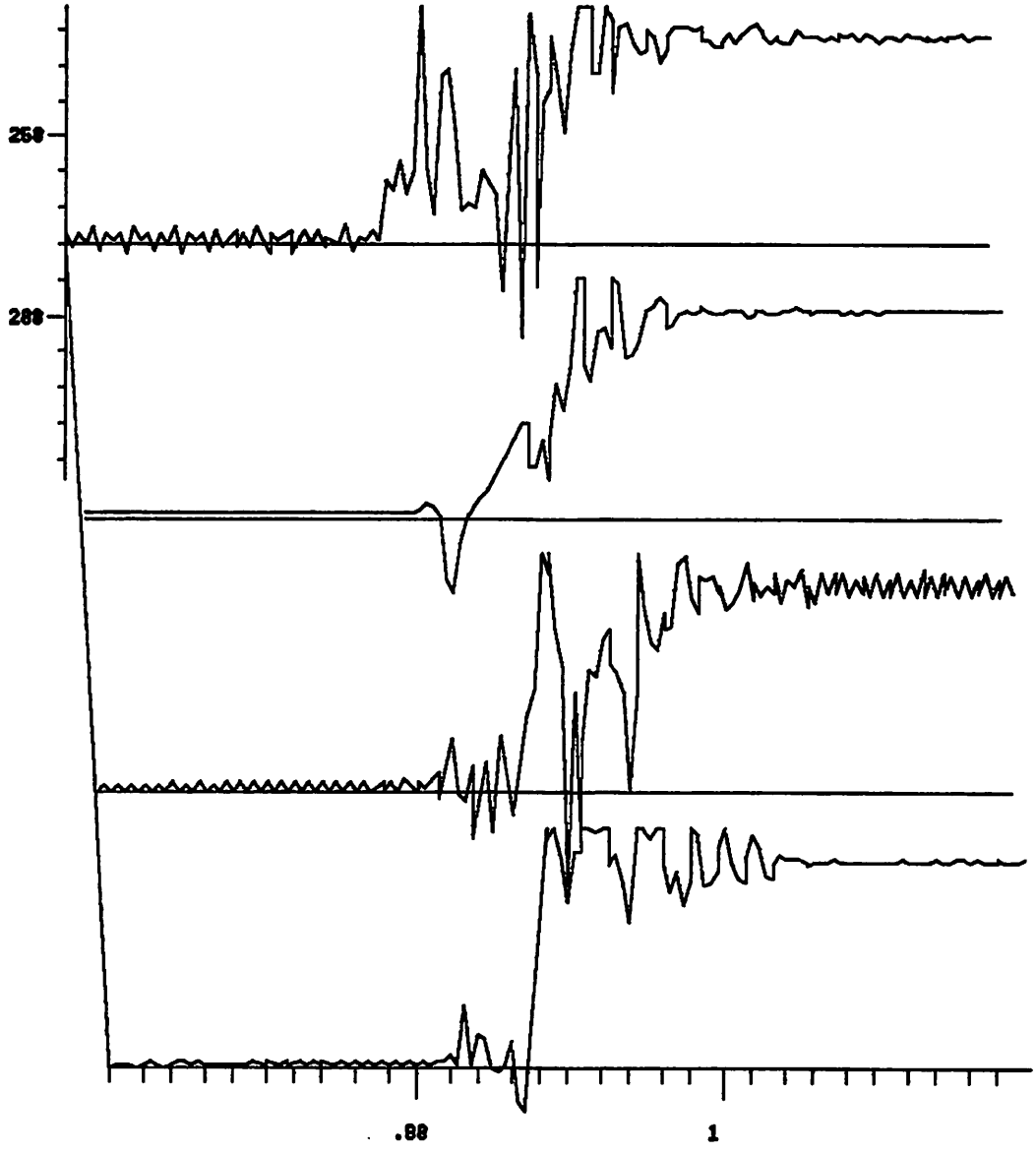


Figure A5.20. Time-varying frequencies of the first four harmonics of an untongued ascending major third played on the bassoon. The lower note is A220.

REFERENCES

Adler, Samuel. 1982. *The study of orchestration.* New York: Norton.

Allen, Jont B. 1977a. "Short term spectral analysis, synthesis, and modification by discrete Fourier transform." *IEEE Proceedings on Acoustics, Speech, and Signal Processing* ASSP-25(3):235–238.

Allen, Jont B. 1977b. "A unified approach to short-time Fourier analysis and synthesis." *Proceedings of the IEEE* 65(11):1558–64.

Allen, Jont B, and L. R. Rabiner. 1977. "A unified approach to short-time Fourier analysis and synthesis." *Proceedings of the IEEE* 65(11):1558–64.

Andersen, Arthur Olaf, 1929. *Practical orchestration.* Boston: Birchard, 1929.

Arfib, Daniel. 1979. "Digital synthesis of complex spectra by means of multiplication of nonlinear distorted sine waves." *Journal of the Audio Engineering Society* 27(10):757–768.

Backhaus, H. 1932. "Über die Bedeutung der Ausgleichsvorgänge in der Akustik." *Zeitschrift für technische Physik* 13(1):31–46. Translated as "On the Importance of Transients in Acoustics" by John Strawn, CCRMA, Stanford University, August 1982. Manuscript.

Bateman, Wayne. 1980. *Introduction to computer music.* New York: Wiley, 1980.

Beauchamp, James W. 1969. "A Computer System for Time-Variant Harmonic Analysis and Synthesis of Musical Tones." In Heinz von Foerster and James W. Beauchamp, eds. *Music by Computers.* New York: Wiley, pp. 19–62.

Beauchamp, James W. 1974. "Time-variant spectra of violin tones." *Journal of the Acoustical Society of America* 56:995–1004.

Beauchamp, James W. 1981. "Data reduction and resynthesis of connected solo passages using frequency, amplitude, and 'brightness' detection and the nonlinear synthesis technique." In Larry Austin and Thomas Clark, eds. *Proceedings of the 1981 International Computer Music Conference.* Denton, Texas: North Texas State University, pp. 316–323.

Benade, A. H. 1976. *Fundamentals of musical acoustics.* New York: Oxford.

Benade, A. H. 1980. "Wind instruments and music acoustics." In Johan Sundberg, ed. *Sound generation in winds, strings, computers.* Stockholm: Royal Swedish Academy of Music, pp. 15–101.

Berlioz, Hector. 1948. *Treatise on orchestration.* Enlarged and Revised by Richard Strauss. Tr. Theodore Front. New York: Kalmus, 1948.

von Bismarck, G. 1974a. "Sharpness as an attribute of the timbre of steady sounds." *Acustica* 30:159–172.

von Bismarck, G. 1974b. "Timbre of steady sounds: A factorial investigation of its verbal attributes." *Acustica* 30:146–159.

Blatter, Alfred. 1980. *Instrumentation/Orchestration.* New York: Longman.

Borish, Jeffrey. 1984. *Electronic simulation of auditorium acoustics.* Ph.D. Dissertation, School of Engineering, Stanford. Department of Music Report No. STAN–M–18.

Bracewell, Ron. 1965. *The Fourier transform and its applications.* New York: McGraw-Hill.

Burton, Stephen D. 1982. *Orchestration.* Englewood Cliffs, New Jersey: Prentice-Hall.

Bussler, Ludwig. 1879. *Instrumentation und Orchestersatz.* Berlin: Habel, 1879.

Callahan, Michael Wayne. 1976. *Acoustic signal processing based on the short-time spectrum.* Ph.D. Dissertation, Department of Computer Science, University of Utah.

Campbell, Warren C., and Jack J. Heller. 1978. "The contribution of the legato transient to instrument identification." Paper presented at the Research Symposium on the Psychology and Acoustics of Music, University of Kansas, Lawrence. Typewritten mss.

Carse, A. 1925. *The History of Orchestration.* New York: Dutton. Reprinted by Dover, New York, 1964.

Casella, A., and V. Mortari. 1950. *La Tecnica dell'orchestra contemporanea.* Second revised edition. Milan: Ricordi.

Charbonneau, Gerard. 1981. "Timbre and the perceptual effects of three types of data reduction." *Computer Music Journal* 5(2):10–19.

Chowning, J. 1973. "The synthesis of complex audio spectra by means of frequency modulation." *Journal of the Audio Engineering Society* 21(7):526–534. Reprinted in Curtis Roads and John Strawn, eds. 1985. *Foundations of computer music.* Cambridge, Massachusetts: MIT Press, pp. 6–29.

Chowning, J. 1980. "Computer synthesis of the singing voice." In Johan Sundberg, ed. *Sound generation in winds, strings, computers.* Stockholm: Royal Swedish Academy of Music, pp. 4–13.

Claasen, T. A. C. M., and W. F. G. Mecklenbrãuker. 1980. "The Wigner distribution—A tool for time-frequency analysis." Part 1: Continuous-time signals. *Philips Journal of Research* 35:217–250. Part II: Discrete-time signals. *Philips Journal of Research* 35:276–300. Part III: Relations with other time-frequency signal transformations. *Philips Journal of Research* 35:372–389.

Clinch, P. G., G. J. Troup, and L. Harris. 1982. "The importance of vocal tract resonance in clarinet and saxophone performance, a preliminary account." *Acustica* 50:280–284.

Cogan, Robert. 1984. *New images of musical sound.* Cambridge, Massachusetts: Harvard University Press.

Crochiere, Ronald E., and L. R. Rabiner. 1983. *Multirate digital signal processing.* Englewood Cliffs, New Jersey: Prentice-Hall.

Cutting, James E., and B. S. Rosner. 1974. "Categories and boundaries in speech and music." *Perception and Psychophysics* 16(3):564–570.

Cutting, James E., B. S. Rosner, and C. F. Foard. 1976. "Perceptual categories for musiclike sounds: Implications for theories of speech perception." *Quarterly Journal of Experiment Psychology* 28:361–378.

Del Mar, Norman. 1981. *Anatomy of the orchestra.* London: Faber and Faber.

Deutsch, D. 1982. *The psychology of music.* New York: Academic.

Dolson, Mark B. 1983. *A tracking phase vocoder and its use in the analysis of ensemble sounds.* Ph.D. Dissertation, California Institute of Technology.

Donnington, R. 1963. *The interpretation of early music.* London: Faber and Faber.

Dudley, H. 1939. "The vocoder." *Bell Labs Record* 17:122–126.

Flanagan, J. L., and R. M. Golden. 1966. "Phase vocoder." *Bell System Technical Journal* 45:1493–1509.

Forsyth, Cecil. 1936. *Orchestration.* New York: Macmillan, 1936.

Fourier, J.-B. 1888. *Œuvres de Fourier.* M. G. Darboux, ed. Paris: Gauthier-Villars.

Freedman, M. D. 1965. "A technique for analysis of musical instrument tones." Ph.D. Dissertation, University of Illinois, Urbana.

Freedman, M. D. 1967. "Analysis of musical instrument tones." *Journal of the Acoustical Society of America* 41:793–806.

Freedman, M. D. 1968. "A method for analyzing musical tones." *Journal of the Audio Engineering Society* 16(4):419–425.

Galler, Bernard A., and M. Piszczalski. 1978. "Automatic music notation translation from sound via 3-dimensional harmonic analysis." Final report, NEH grant no. RO–25315–76–656. Typewritten ms.

Gilson, Paul. *Le tutti orchestrale.* Brussels: Schott, 1922.

Gish, Walter C. 1978. "Analysis and synthesis of musical instrument tones." Audio Engineering Society, 61st Convention, New York, Preprint No. 1410(J–3), 1978.

Gordon, J. W. 1984. *Perception of attack transients in musical tones.* Ph.D. Dissertation, Department of Music, Stanford. Department of Music Report No. STAN–M–17.

Gordon, J. W., and John Strawn. 1985. "An introduction to the phase vocoder." In John Strawn, Ed. *Digital Audio Signal Processing: An Anthology.* Los Altos, California: William Kaufman, pp. 221–270.

Grey, John M. 1975. "An Exploration of Musical Timbre." Ph.D. Dissertation, Department of Psychology, Stanford University. Department of Music Report STAN–M–2.

Grey, John M. 1978. "Timbre discrimination in musical patterns." *Journal of the Acoustical Society of America* 64(2):467–472.

Hays, William L. 1963. *Statistics for psychologists.* New York: Holt, Rinehard and Winston.

Hiller, Lejaren A., Jr., and P. Ruiz. 1971. "Synthesizing musical sounds by solving the wave equation for vibrating objects." Part 1. *Journal of the Audio Engineering Society* 19(6):462–470. Part 2. *Journal of the Audio Engineering Society* 19(7):542–551.

von Helmholtz, Hermann Ludwig Ferdinand. 1913. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* Sixth Edition. Braunschweig: Vieweg. Translated as *On the sensations of tone as a physiological basis for the theory of music* by Alexander J. Ellis. New York: Dover, 1954.

Holtzman, Samuel. 1980. "Non-uniform time-scale modification of speech." M.Sc. and E.E. Dissertation, Department of Electrical Engineering and Computer Science, MIT.

Howe, Hubert S., Jr. 1975. *Electronic music synthesis: Concepts, facilities, techniques.* New York: Norton.

Humperdinck, Engelbert. 1892. *Instrumentationslehre.* Edited by H.-J. Irmen. Cologne: Verlag der Arbeitsgemeinschaft für rheinische Musikgeschichte, 1971.

Hunt, Frederick Vinton. 1978. *Origins in acoustics.* New Haven, Connecticut: Yale University Press.

Hutchins, B., ed. 1975. *Musical engineer's handbook.* Ithaca, New York: Electronotes.

Jacob, G. 1962 *Elements of orchestration.* London: Jenkins.

Jusczyk, Peter W., B. S. Rosner, J. E. Cutting, C. F. Foard, and L. B. Smith. 1977. "Categorical perception of nonspeech sounds by 2-month-old infants." *Perception and Psychophysics* 21(1):50–54.

Justice, James H. 1979. "Analytic signal processing in music computation." *IEEE Proceedings on Acoustics, Speech, and Signal Processing* ASSP-27:670–84.

Kaegi, W., and S. Tempelaars. 1978. "VOSIM—A new sound synthesis system." *Journal of the Audio Engineering Society* 26(6):418–425.

Kajiya, James T. 1979. *Toward a mathematical theory of perception.* Ph.D. Dissertation, University of Utah.

Kaplan, S. J. 1981. "Developing a commercial digital sound synthesizer." *Computer Music Journal* 5(3):62–73.

Kay, S. M., and S. L. Marple, Jr. 1981. "Spectrum analysis—A modern perspective." *Proceedings of the IEEE* 69(11):1380–1419.

Kennan, Kent W. 1970. *The technique of orchestration.* Second Edition. Englewood Cliffs, New Jersey: Prentice-Hall.

Kewley-Port, Diane, and D. B. Pisoni. 1984. "Identification and discrimination of rise time: Is it categorical or noncategorical?" *Journal of the Acoustical Society of America* 75(4):1168–76.

Kling, H. 1905. *Modern orchestration and instrumentation.* Translated by Gustav Saenger. New York: C. Fischer.

Kunitz, Hans. 1961. *Die Instrumentation.* Leipzig: VEB Breitkopf und Härtel. 13 volumes.

LeBrun, Marc. 1977. "A derivation of the spectrum of FM with a complex modulating wave." *Computer Music Journal* 1(4):51–52. Reprinted in Curtis Roads and John Strawn, eds. 1985. *Foundations of Computer Music*. Cambridge, Massachusetts: MIT Press, pp. 65–67.

LeBrun, Marc. 1979. "Digital waveshaping synthesis." *Journal of the Audio Engineering Society* 27(4):250–266, 1979.

LeCaine, Hugh. 1956. "Electronic music." *Proceedings of the IRE* 457–478.

Limacher, J. 1979. "A spectral analysis of eight clarinet tones." B.A. Honors Thesis, Department of Music, Stanford.

Locke, S., and L. Kellar. 1973. "Categorical perception in a nonlinguistic mode." *Cortex* 9:355–369.

Luce, David A. 1963. *Physical correlates of nonpercussive musical instrument tones*. Ph.D. Dissertation, Department of Physics, Massachusetts Institute of Technology.

Luce, David A., and M. Clark, Jr. 1965. "Duration of attack transients of nonpercussive orchestral instruments." *Journal of the Audio Engineering Society* 13:194–99.

Luce, David A., and M. Clark, Jr. 1967. "Physical correlates of brass-instrument tones." *Journal of the Acoustical Society of America* 42(6):1232–1243.

McAdams, Stephen, and Albert Bregman. 1979. "Hearing Musical Streams." *Computer Music Journal* 3(4):26–43, 60. Reprinted in Curtis Roads and John Strawn, eds. 1985. *Foundations of computer music*. Cambridge, Massachusetts: MIT Press, pp. 658–98.

Macmillan, N. A., H. L. Kaplan, and C. D. Creelman. 1977. "The psychophysics of categorical perception." *Psychological Review* 84(5):452–471.

Mancini, Henry. 1962. *Sounds and Scores*. N.p.: Northridge.

Mathews, Max V., and Joan E. Miller. 1982. "How to Make a Slur." Murray Hill, New Jersey: Bell Laboratories. Typewritten mss., no date. Lecture, Center for Computer Research in Music and Acoustics, Stanford University, Stanford, California, 7 July 1982.

Mattingly, Ignatius G., A. M. Liberman, A. K. Syrdal, anbd T. Halwes. 1971. "Discrimination in Speech and Nonspeech Modes." *Cognitive Psychology* 2:131–157.

Meyer, Erwin, and Gerhard Buchmann. 1931. "Die Klangspektren der Musikinstrumente." *Sitzungsberichte der Preußischen Akademie der Wissenschaften* (Physikalisch-Mathematische Klasse) 1931:735–778. Berlin: Verlag der Akademie der Wissenschaften/Walter de Gruyter.

Meyer, Jürgen. 1972. *Akustik und musikalische Aufführungspraxis*. Frankfurt am Main: Verlag das Musikinstrument.

Miller, D. C. 1935. *Anecdotal history of the science of sound*. New York: MacMillan.

Miller, Neil J. 1973. *Filtering of singing voice signal from noise by synthesis*. Ph.D. Dissertation, Department of Electrical Engineering, University of Utah.

Moorer, James A. 1973. *The heterodyne filter as a tool for analysis of transient waveforms*. Report No. STAN–CS–73–379. Stanford University: Computer Science Department.

Moorer, James A. 1975. "On the segmentation and analysis of continuous musical sound by digital computer." Ph.D. Dissertation, Department of Computer Science, Stanford University. Department of Music Report STAN–M–3.

Moorer, James A. 1976. "The synthesis of complex audio spectra by means of discrete summation formulas." *Journal of the Audio Engineering Society* 24:717–727.

Moorer, James A. 1977. "Signal processing aspects of computer music—A survey." *Proceedings of the IEEE* 65(8):1108–1137. Revised and updated version in John Strawn, Ed. 1985. *Digital Audio Signal Processing: An Anthology*. Los Altos, California: William Kaufman, Inc., pp. 149–220.

Moorer, James A. 1978. "The use of the phase vocoder in computer music applications." *Journal of the Audio Engineering Society* 26(1/2):42–45.

Moorer, James A., John M. Grey, and John Strawn. 1977. "Lexicon of analyzed tones. Part 2: Clarinet and oboe tones." *Computer Music Journal* 1(3):12–29.

Moorer, James A., John M. Grey, and John Strawn. 1978. "Lexicon of analyzed tones. Part 3: The trumpet." *Computer Music Journal* 2(2):23–31.

Morrill, Dexter. 1980. "The dynamic aspects of trumpet phrases." (French version: "Aspects dynamiques du phrase de la trompette," translated by Emmanuel Gresset). Paris: IRCAM.

Morrill, Dexter. 1982. Letter, dated 26 March.

Ohm, Georg S. 1843. Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen." *Annalen der Physik und Chemie* 54(8):513–65.

Petersen, Tracy L. 1980. "Acoustic signal processing in the context of a perceptual model." Ph.D. Dissertation, Computer Science Department, University of Utah.

Petersen, Tracy L., and S. F. Boll. 1983. "Critical band analysis-synthesis." *IEEE Proceedings on Acoustics, Speech, and Signal Processing* ASSP-31(3):656–63.

Piston, W. 1955. *Orchestration*. New York: Norton.

Plomp, Reinier. 1970. "Timbre as a multidimensional attribute of complex tones." In R. Plomp and G. F. Smoorenburg, Eds. *Frequency analysis and periodicity detection in hearing*. Leiden: Sijthoff, pp. 397–414.

Plomp, Reinier. 1976. *Aspects of tone sensation:* A psychophysical study. New York: Academic Press.

Portnoff, Michael R. 1976. "Implementation of the digital phase vocoder using the fast Fourier transform." *IEEE Proceedings on Acoustics, Speech, and Signal Processing* 24:243–48.

Portnoff, Michael R. 1978. "Time-Scale modification of speech based on short-time Fourier analysis." Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, MIT.

Portnoff, Michael R. 1980. "Time-frequency representation of digital signals and systems based on short-time Fourier analysis." *IEEE Proceedings on Acoustics, Speech, and Signal Processing* ASSP-28(1):55–102.

Portnoff, Michael R. 1983. Lecture, Stanford, May 25.

Potter, Charles E., and D. Teaney. 1981. "Sonic transliteration applied to descriptive music notation." In Hubert S. Howe, Jr., ed. *Proceedings of the 1980 International Computer Music Conference*. San Francisco, California: Computer Music Association, pp. 138–144.

Rabiner, L. R., and B. Gold. 1975. *Theory and application of digital signal processing*. Englewood Cliffs, New Jersey: Prentice-Hall.

Raman, C. V. 1918. "Mechanical theory of bowed strings." *Bull. Ind. Assoc.* 15:1. (not seen; Rösing's reference no. 170).

Rauscher, Donald J. 1963. *Orchestration: Scores and scoring*. New York: Free Press of Glencoe.

Reinecke, H.-P. 1953. *Über den doppelten Sinn des Lautheitsbegriffs beim musikalischen Hören*. Ph.D. Dissertation, University of Hamburg. [not seen]

Remez, R. E., J. E. Cutting, and M. Studdert-Kennedy. 1980. "Cross-series adaption using song and string." *Perception and Psychophysics* 27(6):524–30.

Rimsky-Korsakov, Nicolas. 1922. *Principles of orchestration*. Translated by Edward Agate. New York: Kalmus.

Risset, Jean-Claude. 1966. "Computer study of trumpet tones." Murray Hill, New Jersey: Bell Laboratories. Typewritten mss. *Journal of the Acoustical Society of America* 38:912, 1965 (abstract only).

Risset, Jean-Claude, and Max V. Mathews. 1969. "Analysis of musical instrument tones." *Physics Today* 22(2):23–40.

Roads, Curtis. 1978. "Automated granular synthesis of sound." *Computer Music Journal* 2(2):61–62, 1978. Revised and updated version in Curtis Roads and John Strawn, eds. 1985. *Foundations of Computer Music*. Cambridge, Massachusetts: MIT Press, pp. 145–59.

Roads, Curtis, and John Strawn, eds. 1985. *Foundations of Computer Music*. Cambridge, Massachusetts: MIT Press.

Rodet, Xavier. 1984. "Time-Domain Formant-Wave-Function Synthesis." *Computer Music Journal* 8(3):9–14.

Rodet, Xavier, Y. Potard, and J.-B. Barrière. 1984. "The CHANT project: From synthesis of the singing voice to synthesis in general." *Computer Music Journal* 8(3):15–31.

Rösing, Helmut. 1967. *Probleme und neue Wege der Analyse von Instrumenten- und Orchesterklängen*. Ph.D. Dissertation, University of Vienna, 1967. Vienna: Verlag Notring der wissenschaftlichen Verbände Österreichs, 1970.

Rosen, S. M., and P. Howell. 1981. "Plucks and bows are not categorically perceived." *Perception and Psychophysics* 30(2):156–68.

Rush, Loren. 1982. "The Tuning of Performed Music." Presented at the 1982 International Computer Music Conference, Venice, Italy.

Saldanha, E. L., and John F. Corso. 1964. "Timbre cues and the identification of musical instruments." *Journal of the Acoustical Society of America* 36:2021–26.

Samson, Peter R. 1980. "A general-purpose digital synthesizer." *Journal of the Audio Engineering Society* 28(3):106–13.

Samson, Peter R. 1985. "Architectural issues in the design of the Systems Concepts Digital Synthesizer." In John Strawn, Ed. *Digital Audio Signal Engineering: An Anthology.* Los Altos, California: William Kaufman, pp. 61–93.

Schafer, R. W., and L. R. Rabiner. 1973. "Design and simulation of a speech analysis-synthesis system based on short-time Fourier analysis." *IEEE Transactions on Audio and Electroacoustics* AU-21:165–74.

Schottstaedt, Bill. 1977. "The simulation of natural instrument tones using frequency modulation with a complex modulating wave." *Computer Music Journal* 1(4):46–50. Reprinted in Curtis Roads and John Strawn, eds. 1985. *Foundations of Computer Music.* Cambridge, Massachusetts: MIT Press, pp. 54–64.

Schouten, M. E. H. 1980. "The case against a speech mode of perception." *Acta Psychologica* 44:71–98.

Schroeder, Manfred R. 1966. "Vocoders: Analysis and synthesis of speech." *Proceedings of the IEEE* 54(5):720–34.

Schwede, Gary W. 1983. "An algorithm and architecture for constant-*Q* spectrum analysis." *Proceedings of the IEEE Internationl Conference on Acoustics, Speech, and Signal Processing,* April, pp. 1384–87.

Seagrave, B. G., and J. Berman. 1976. *Dictionary of bowing terms for stringed instruments.* Second Edition. American String Teachers Association.

Seashore, Carl E. 1938. *Psychology of music.* New York: McGraw-Hill. (Reprinted by Dover, New York, 1967).

Smith, Julius O. 1983. *Techniques for Digital Filter Design and System Identification with Application to the Violin.* Ph.D. Dissertation, School of Engineering, Stanford.

Smith, Julius O. 1984. Private communication, 4 March.

Stautner, John Paul. 1983. *Analysis and synthesis of music using the auditory transform.* M.Sc. Thesis, Department of Electrical Engineering and Computer Science, MIT.

Steiger, H., and A. S. Bregman. 1981. "Capturing frequency components of glided tones: Frequency separation, orientation, and alignment." *Perception and Psychophysics* 30(5):425–35.

Strawn, John. 1980. "Approximation and syntactic analysis of amplitude and frequency functions for digital sound synthesis." *Computer Music Journal* 4(3):3–24.

Strawn, John. 1982. "Research on timbre and musical contexts at CCRMA." In T. Blum and J. Strawn, eds. *Proceedings of the 1982 International Computer Music Conference.* San Francisco: Computer Music Association, pp. 437–65.

Strawn, John. 1983. "Spectra and timbre." Presented at the 1983 International Computer Music Conference.

Strawn, John. 1985a. "Editing time-varying spectra." Presented at the 78th AES Convention, Anaheim. Preprint 2228 (A-16).

Strawn, John. 1985b. "Orchestral instruments: Analysis of performed transitions." Presented at the 78th AES Convention, Anaheim. Preprint 2229 (B-10).

Strong, W., and M. Clark, Jr. 1967a. "Synthesis of wind-instrument tones." *Journal of the Acoustical Society of America* 41(1):39–52.

Strong, W., and M. Clark, Jr. 1967b. "Perturbations of synthetic orchestral wind-instrument tones." *Journal of the Acoustical Society of America* 41(2):277–85.

Studdert-Kennedy, M., A. M. Liberman, K. S. Harris, and F. S. Cooper. 1970. "Motor theory of speech perception: A reply to Lane's critical review." *Psychophysical Review* 77:234–49.

Sundberg, Johann, A. Askenfelt, and L. Frydén. 1983. "Musical performance: A synthesis-by-rule approach." *Computer Music Journal* 7(1):37–43.

Teuchert, Emil, and E. W. Haupt. 1924. *Musik-Instrumentenkunde in Wort und Bild.* Leipzig: Breitkopf und Härtel.

Wedin, Lage, and Gunnar Goude. 1972. "Dimension analysis of the perception of instrumental timbre." *Scandanavian Journal of Psychology* 13:228–40.

Winckel, Fritz. 1960. *Phänomene des musikalischen Hörens.* Berlin: Max Hesses Verlag, 1960. Translated by Thomas Binkley as *Music, Sound and Sensation: A Modern Exposition.* New York: Dover, 1967. The English version is an abominable translation.

Wyss, Niklaus. 1984. Personal communication, Stanford, April 19.

Youngberg, James E. 1979. *A constant percentage bandwidth transform for acoustic signal processing.* Ph.D. Dissertation, Department of Computer Science, University of Utah.

Zwicker, Eberhard, and Richard Feldtkeller. 1967. *Das Ohr als Nachrichtenempfänger.* Second edition. Stuttgart: Hirzel.