# Orchestral Instruments: Analysis of Performed Transitions\*

#### JOHN STRAWN\*\*

The Droid Works, San Rafael, CA 94912, USA

The region between performed notes was examined in nine instruments (flute, bass flute, piccolo, clarinet, oboe, bassoon, trumpet, violin, cello). On each instrument, eight intervals (2nd, 3rd, 5th, 7th, ascending and descending) and two playing styles (tongued/untongued, bow change/no bow change) were digitally recorded. Using a computer, the recordings were analyzed for time-varying power and time-varying spectrum. Analysis of as many as five recordings of a given interval and playing style showed that the performer could easily replicate a given transition, so the recordings were judged representative. In the tongued (bow change) case, the notes were farther apart, the amplitude dip between the notes (except in the cello) was greater, and spectral changes were more extensive than in the nontongued (no bow change) case. This pattern was not significantly influenced by variation in the size of the interval, the direction of the interval, or the instrument performing.

#### **0 INTRODUCTION**

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 [Fig. 1(a)]. This model still dominates thinking about the overall shape of a musical note. It is thus not surprising that most of the research on instrumental sounds to date has concentrated on the analysis of just one note at a time [1].

But much of an instrument's tell-tale "sound" lies in how the notes are connected, and thus in how musical phrases help create the instrument's "signature." It is thus important to examine more than one note at a time if the nature of musical sound is to be fully understood. The advent of digital audio and digital signal processing have now made analysis along these lines possible. As a first step in this direction, this paper concentrates on the transition between notes. In this paper a transition will be understood to include the ending part of the decay of one note, the beginning and possibly all of the attack of the next note, and whatever connects the two notes. It will become clear that a transition typically lasts on the order of tens to hundreds of milliseconds.

Some studies have been done on musical structures longer than one note [2]-[8], but none of these has concentrated on analyzing the transition between notes. After this work was completed we discovered that Miller [9] had been able to photograph the time-varying waveform as a violin bow changes directions. His photograph [9, p. 197] and his observations agree with the material presented here.

There are several possible models [10] for the juncture between two notes: 1) there might be a gap between the decay of one note and the beginning of the next [Fig. 1(a)]; 2) the attack of the second note might start right when the decay of the first note finishes [Fig. 1(b)]; or 3) the decay of the first note might overlap the attack of the second [Fig. 1(c)]. The first step was to find out which of these models occurs in nature.

#### **1 GATHERING THE DATA**

When this study began, no corpus of suitable audio recordings was available. Thus from 1979 through 1983 we recorded transitions played on nine instruments at the Center for Computer Research in Music and Acoustics (CCRMA), Stanford University: flute, piccolo, bass flute, clarinet, oboe, bassoon, trumpet, violin, and cello. It seemed wise to select at least one instrument from each of the traditional families of orchestral in-

<sup>\*</sup> Presented at the 78th Convention of the Audio Engineering Society, Anaheim, CA, 1985 May 3-6; revised 1986 March 26 and August 8.

<sup>\*\*</sup> Now with S Systems, P.O. Box 623, San Rafael, CA 94915-0623.

struments (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.

#### 1.1 Making the Digital Recordings

Each player was recorded in a room measuring approximately 20 by 24 ft (6 by 7 m) at CCRMA. This was the room used for measurements by Borish [11]; many of the experiments by Grey [1], Cohen [12], and Gordon [13] were also conducted in this room. Its isolated location made it adequate for undisturbed recording and playback. The walls were treated with absorbent material to reduce reverberation in the room.

The 14-bit analog-to-digital converters installed on the PDP-10 of the Artificial Intelligence Laboratory, which occupied the building at the time, were used to make initial digital recordings directly to computer disk. (This recording setup is discussed in the earlier version of [14].) The remaining (16-bit) digital recordings were made using a Sony F1 recorder; these recordings were digitally transferred into the CCRMA Foonly computer using a specially constructed interface and likewise stored on disk. For some recordings we used a B&K 2619 microphone, and a Crown PZM for others. All of the recordings were resampled [15] to 25.6 kHz, selected because it was low enough for practical work but still high enough to ensure adequate fidelity. There were some low-frequency artifacts in the recording, such as a dc component, which were removed by highpass filtering the recordings, typically using a digital eighth-order Butterworth filter with -3-dB point at 50 Hz or so. The recordings were spliced apart into in-



Fig. 1. Possible methods for creating a transition between notes. (a) No interaction occurs between notes. (b) Notes are abutted. (c) Notes overlap.

dividual (monaural) files, each containing a two-note segment.

The noise level of the recordings turned out to be around -60 dB. Given the maximum theoretical limits of around 84 dB (for the 14-bit DAC) or 96 dB (for the Sony 16-bit DAC), this may seem surprisingly poor. However, monitoring amplitude in the direct-to-computer-disk method (discussed in the earlier version of [14]) was clumsy at best, so recording at close to full amplitude was ill-advised. Experience with the trumpet likewise showed that the "overload" light on the Sony F1 recorder was not as reliable as one needs when working with very tight headroom (there was no console between the microphone and the Sony F1, so no other monitoring was available). Thus the trumpet recordings were clipped in a few places in the steady states of the tones (not in the transitions). Fortunately the clipping lasted for only two or three samples each period. We were able to remove the clipping by low-pass filtering the recordings before doing the downsampling. Still, this warned us against trying again to exploit the full dynamic range of the F1; for the other recordings, more headroom was allotted, with a resulting reduction in dynamic range. The available dynamic range posed no problems in any of the work presented here. In particular, the transitions themselves all lay well above the noise floor [10].

### 1.2 Choice of Intervals

"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. 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 [1], 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, as their normal playing ranges do not overlap. The next best solution was to have many of the instruments play the same pitches, and to have other instruments play an octave above or below. Another constraint was to avoid open strings

Table 1. Instruments recorded.

Family	Instrument	Base pitch
Air reed	Flute	A220 A1760
Single reed	Bass flute Clarinet	A220 A220
Double reed	Oboe Bassoon	A440 A220
String	Violin Cello	A220 A220
Lip reed	Trumpet	A220

on the string instruments; all pitches were to be played on stopped strings. Thus only seconds and thirds were formally analyzed on the strings, although other intervals were recorded and examined. In the end, the "base" pitches given in Table 1 were used for each instrument; that is, these were the lower notes for both ascending and descending intervals. Fig. 2 shows the complete set of intervals (based on A220).

## **1.3 Playing Methods versus Perceived** Articulations

At this point an aside on instrumental performance techniques is in order. 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 Fig. 1(b) 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.

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 wide range of 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. 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.

In a comprehensive study of the transition between performed notes, it is necessary to examine more than one kind of performed articulation. At the same time, it would be impractical to analyze all of the possible nuances of perceived articulation. The situation is complicated because the perceived articulation is not necessarily a clue to the playing technique for both the wind and the string instruments [10]. The purpose of



Fig. 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.)

this study was not to learn more about the physical correlates of any given playing techniques per se; rather, it was necessary to choose a small set of playing techniques which are well known to both performer and listener and which one may readily assume may be perceptually distinguished from each other. Thus the winds and brass played the second note with or without tonguing; and the strings performed the second note with or without bow change. (In this paper the term "tongued" will include the transition with bow change on the strings, and "untongued" will also include "without bow change.") Subsequent research [10], which will not be discussed here, showed that these recordings provided a reasonable data base.

In order to check whether the recordings were representative, as many as five separate recordings of each transition were made.

### 1.4 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.

No attempt was made to tune all of the instruments to one frequency reference. The musicians were instructed to play each note for a duration of 1 s; a metronome was available in the recording sessions. Still, the durations of the notes surrounding the transitions varied considerably.

Grey [1] equalized his test tones for loudness, duration, and pitch. It proved impossible to equalize our two-note recordings along these lines. In order to equalize the recordings for duration or 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, it was also impossible to find a way to equalize, say, the loudnesses of the tongued transitions, without again distorting the transition as recorded.

## **2 RECORDED TRANSITIONS**

A sample set of transitions, plotted using Rush's EdSnd program [16], is shown in Figs. 3-8. Each pair of figures contains amplitude plots for one instrument: the clarinet, trumpet, and violin are shown. Each set of plots contains four tongued intervals (three for the clarinet, as the ascending seconds were lost), followed by the same intervals played without tonguing. For the clarinet in Figs. 3 and 4 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; the clarinet plots show 300 ms. The "waves" in the clarinet plots are artifacts of the display process and should be ignored; they disappear

when a smaller time range is displayed. The trumpet in Figs. 5 and 6 has the same intervals, plus the major second at the top. A smaller time range (130 ms) is shown for the trumpet in order to avoid certain artifacts in the display. Only ascending transitions are shown for these two instruments. The larger two intervals were not analyzed in the strings, because the recordings were limited to intervals that could be played on one string. So the violin transitions shown here (Figs. 7 and 8) include the major seconds and thirds, both ascending and descending. The time range shown for the violin plots is 200 ms.

## 2.1 General Nature of Musical Transitions

The conclusions in this section are based on the plots shown in Figs. 3-8 as well as on similar plots, not reproduced here, for the other instruments recorded (some are given in [10]).

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.

In no case did the amplitude between the 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 Fig. 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, 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 in Fig. 3), or the decay in the ascending fifth in the untongued trumpet (second from bottom in Fig. 6). Some instruments showed a swelling on some notes, such as on the ascending third played with bow change on the violin (second from bottom in Fig. 7). Some transitions, such as the ascending second at the top of the same figure, show "shoulders" in the decay (or the attack). Only some of the decays follow the exponential path which one would expect; the tongued clarinet tones in Fig. 3 are perhaps the "closest to theory." This "nonexponential" behavior occurs because the performer is still introducing energy into the instrument, even during the decay of the note.

Although these amplitude plots are unclear in this matter, the change in pitch can be shown [10] to occur at the earliest in the middle of the transition, and usually right at the attack of the second note. In none of our recordings did the change in pitch occur during the decay of the first note. Hereafter we will use the term "point of pitch change" to refer to the change in frequency, as opposed to "transition" as defined above, meaning the whole region between two notes.

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 those for the clarinet. 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



Fig. 3. Tongued transitions (300 ms) between two notes on the clarinet. The first note is A220 in each case. From the top: ascending third, ascending fifth, ascending seventh. In each plot, the first note ends on the left, the transition is in the middle, and the second note begins on the right.



Fig. 4. Untongued transitions (300 ms) between two notes on the clarinet. The first note is A220 in each case. From the top: ascending third, ascending fifth, ascending seventh.



Fig. 5. Tongued transitions (130 ms) between notes on the trumpet. The lower note is A220 in each case. From the top: ascending second, ascending third, ascending fifth, ascending seventh.



Fig. 6. Untongued transitions (130 ms) between notes on the trumpet. The lower note is A220 in each case. From the top: ascending second, ascending third, ascending fifth, ascending seventh.



Fig. 7. The transition (200 ms) 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.



Fig. 8. The transition (200 ms) 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.

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 Fig. 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, we found at least a little bit of noise in all the attacks; this is contrary to what Rösing [7] observed in his extensive but rarely cited study. Probably the frequency resolution and dynamic range of the Kay Sonograph that he used were not adequate to make this noise visible in his plots. Also, we often found 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 makes it difficult to pinpoint exactly where the point of pitch change lies. At the same time, this observed behavior matches well Benade's remarks on how the instrument can quickly reach a stable state of oscillation: "[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" [17, p. 425].

The situation is not so simple, of course, when the player purposefully includes noise, as can be seen in some of the violin plots. In smooth, more or less noisefree 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.

Most important, the tongued case often exhibits a wider gap between the two notes, and a greater dip in amplitude, than does the nontongued case. The next two sections will examine this finding in more detail.

## 3 ANALYSIS OF TIME-VARYING POWER IN TRANSITIONS

It is misleading to base a detailed analysis of timevarying musical signals solely on amplitude plots such as those just given. The wave shapes 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 Fig. 8, second from bottom). Furthermore, the positive-going peak excursion is significantly different from the negative-going one in some recordings (see Figs. 7 and 8). The goal here is to find some acoustical behavior common to all these instruments. But the raw waveforms are inadequate for this purpose. Another measure of the signal's amplitude is therefore needed. Average power is widely used; but here a time-varying measure is called for.

To this end, we adopted an algorithm developed by Smith [18], which identifies the peak of each period in the waveform (see also [19]). It was possible to tune this algorithm so that it worked automatically 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 [2] 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 [20].) 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); Risset [21, 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

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

where N(n) is the length of a period (peak to peak) beginning at sample number n, and y(i) 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. Figs. 9-14 show time-varying power plots for the two-note pairs given in Figs. 3-8; these 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 [6].) To facilitate comparisons, the order of these power plots matches exactly the order of the earlier amplitude plots. Power analyses for some of the other instruments recorded are given in [10].

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 Fig. 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 Fig. 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 Fig. 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; Fig. 14 is a case in point. These are areas where the peak-tracking algorithm was confused, often by a largescale shift in phase due to time-varying spectral changes. It is possible to remove these "burrs" by correcting the locations of the peaks by hand and then recalculating the power curves; but we 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 area 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 Fig. 9 on the one hand, and the differences among the three plots in Fig. 10 on the other, are not nearly as large as the differences between the two figures. In Fig. 9 the amplitude dip is deeper, and the time gap between the two notes wider, than in Fig. 10. The same generalizations can be made about the trumpet power plots (Figs. 11 and 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 for the other analyses [10] shows that this pattern holds for all but one 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 Fig. 13 are quite similar, as are the two top traces in Fig. 14; the largest difference occurs between the two figures. 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 Fig. 14.

We thus concluded that time-varying power varies more with the playing method than with the 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.

## 4 ANALYSIS OF TIME-VARYING SPECTRUM IN TRANSITIONS

These recordings were also analyzed using the discrete short-time Fourier transform (DSTFT). Fig. 15 gives an overview of this process (detailed references are given in [22], [23]). In effect, a signal x(t) is passed through a set of bandpass filters whose center frequen-

cies are equally spaced from dc to one half the sample rate. In analyzing tones from musical instruments, one usually arranges the filters so that one harmonic falls into each passband. The real and imaginary outputs of the filters shown in Fig. 15 give a time-varying spectral representation of the signal. If the analysis outputs are fed directly to the synthesis part of the technique, the output y(t) is virtually identical to the input x(t). The analysis data may be modified in various ways to produce tones which are more or less close to the original. Also, the real and imaginary outputs may be converted into time-varying amplitude and frequency terms. The DSTFT and related techniques have been used for several decades to analyze musical instruments, yielding results useful for psychoacoustic researchers [1], [5], [10], [21], [24]-[28] as well as synthesizer manufacturers and the recording industry in general.

The problem with using the DSTFT 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. Also, since the amplitude of the signal drops several tens of decibels during many transitions, the frequency traces are difficult to interpret, because the frequency trace becomes unstable at very low amplitudes. Furthermore, if the frequency traces are moving too quickly, it can be shown that the DSTFT does not track them accurately, which might lead to distortion. Analysis of data from psychoacoustic experiments showed that any such distortion could be ignored. These issues are discussed further in [10] and [23].

#### 4.1 Creating Plots of the Amplitude Traces

We solved the problem of the displaced analysis channels by running the analysis twice, once with the filters set for the first note, and again with the filters set for the second note. The end of the first note analyzed in this manner looked reasonable; and so did the beginning of the next. Psychoacoustic experiments [10], [23] confirmed that this was the case.

The only way we could find to make useful spectral plots was to splice together these two analyses in a three-dimensional representation [10], [23]. It was necessary to modify spectral editor [20] to handle these two analyses properly. Figs. 16 (the tongued clarinet ascending major third) and 17 (the same interval, played untongued) show a sample of the result. These are the same ascending thirds already presented in Figs. 3, 4, 9, and 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 DSTFT analysis for the first note

to that of the second; this is approximately the point of pitch change.

## 4.2 Analysis of Amplitude Plots

Clearly, there is a spectral rolloff at the end of the first clarinet note in the tongued clarinet transition of Fig. 16; of the 30 harmonics shown here, perhaps the top 20 drop out. Note that the pattern with which the harmonics drop out and reenter is not entirely regular. However, in general the higher order harmonics leave sooner and reenter later than their lower frequency counterparts. This corresponds to the "wedgelike funnel, open toward the higher frequencies," found by Rösing (our translation). Comparison of Fig. 16 with Fig. 17 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 shown 300 ms). 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 [5, p. 107].

To cite some more examples, Figs. 18 and 19 show the tongued and untongued transitions, respectively, for the ascending major third on the trumpet. These transitions were already shown in Figs. 5, 6, 11, and 12. As the attack of the second note begins, there is some noise visible in the higher order harmonics. The irregularity in the pattern according to which the harmonics drop out and reenter is striking. Another set, this time for the violin, is given in Figs. 20 and 21; here we have the ascending third, already met in Figs. 7, 8, 13, and 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 lowpass filtered version of the spectrum at the end of the steady state of the first note.

More plots of this kind are given in [10]. The following conclusion is based on an analysis of threedimensional time-varying spectral plots for all 212 (!) 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. We found this to be a general principle, regardless of 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.

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. Recall that as many as five separate recordings were made (by the same performer) for a given instrument, interval size, interval direction, and playing style. Analyses of these multiple recordings are given in [10]. Thorough examination of these data quickly led to the conclusion that the performers could reliably reproduce a given transition, and that the transitions illustrated in the figures given here were representative performances.

#### 4.3 Plots of the Frequency Traces

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-di-



Fig. 9. Period-synchronous power of two entire notes on the clarinet, with the second note tongued. The first note is A220 in each case. From the top: ascending third, ascending fifth, ascending seventh.



Fig. 10. Period-synchronous power of two entire notes on the clarinet, with the second note untongued. The first note is A220 in each case. From the top: ascending third, ascending fifth, ascending seventh.



Fig. 11. Period-synchronous power of two entire 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.



Fig. 13. Period-synchronous power of two entire 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.



Fig. 12. Period-synchronous power of two entire 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.

J. Audio Eng. Soc., Vol. 34, No. 11, 1986 November



Fig. 14. Period-synchronous power of two entire 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.

ORCHESTRAL INSTRUMENTS

mensional 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. A few plots of this kind are given in [10].

#### **5 ON THE EFFECTS OF INSTRUMENT SIZE**

Analyzing power plots for the cello, given in Figs. 22 (with bow change) and 23 (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 Figs. 13 and 14, respectively.

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. Figs. 24 and 25 show the time-varying spectrum of the ascending thirds with and without bow change, respectively. (These plots should be compared with Figs. 20 and 21 for the violin.) Compared with Fig. 25, Fig. 24 shows the dip in the spectrum found



Fig. 15. The DSTFT analysis stage (left) works like a set of band-pass filters spaced equally from 0 Hz to one half of the sampling rate. If the data analyzed from x(t) are used for resynthesis (right), a signal y(t) results which is nearly identical to the original. (Reprinted with permission from [22].)

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



Fig. 16. Time-varying spectral analysis (300 ms, 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 s.



Fig. 17. Time-varying spectral analysis (300 ms, 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 s.

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



Fig. 18. Time-varying spectral analysis (300 ms, 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 s.



Fig. 19. Time-varying spectral analysis (300 ms, 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 s.

J. Audio Eng. Soc., Vol. 34, No. 11, 1986 November

regular flute. Examples of power and spectral plots for these instruments are given in [10]. 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.



Fig. 20. Time-varying spectral analysis (300 ms, 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 s.



Fig. 21. Time-varying spectral analysis (300 ms, 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 s.



Fig. 22. Period-synchronous power of two entire 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, descending third.



Fig. 23. Period-synchronous power of two entire 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.

### **6 SUMMARY AND CONCLUSION**

We were not able to develop empirical methods for analyzing and comparing the plots resulting from these two kinds of analysis. However, examination of these plots suggests the following conclusions, some of which were borne out in later research [10], [23].

1) Notes in musical performance are in fact connected; they do not occur in isolation. The Helmholtzian model of Fig. 1(a) must be modified to include the transitional material. The decay of the first note is affected more than the attack of the second.

2) Except for the cello (see below), there is a characteristic drop in overall amplitude between two notes,



Fig. 24. Time-varying spectral analysis (300 ms, 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 s.



Fig. 25. Time-varying spectral analysis (300 ms, 50 harmonics) 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 s.

as one would expect. Likewise, at the end of the first note there is a gradual spectral rolloff. The transition region consists in general of a low-passed version of the tail of the first note. During the attack of the second note, the higher frequency components reenter.

3) The transition from one pitch to the next occurs very quickly. A good player can make the transition between notes in the time required for just a few periods of the waveform.

4) The skilled player can easily replicate a given articulation. Power and spectral analyses of as many as five repetitions of a given size-direction-playing style combination showed no significant differences. The durations of the notes surrounding the transition may be slightly different in each case. But the shape of the transition (such as the slope of the decay of the first note, the slope of the attack of the second note, the very short trough between the two notes) is surprisingly consistent.

5) Looking more closely at the general behavior given under 2), we have concluded that the tongued case often exhibits a wider gap between the two notes, and a greater dip in amplitude, than the nontongued case. In no case did the amplitude between notes for the tongued case reach or stay at 0.0, as one might expect. Likewise, the spectral rolloff is in general deeper for the tongued case than for the untongued.

6) There is no systematic difference between ascending and descending intervals for a given instrument, nor for the intervals of varying sizes. This finding runs counter to what one might expect. Of course, some intervals on some instruments are harder to play than others, but no systematic differences were found.

7) 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. In the strings, one might expect "bow change" to produce a more abrupt attack on the second note. However, the "no bow change" produces its own abrupt attack on ascending notes, because the finger "thwacks" the string to make it shorter.

8) In spite of instructions to these players to play the notes at the same loudness, the amplitudes of the two notes are often quite different. As one might expect, the decay time of the first note is often different from the attack time of the second. The attacks and decays often follow a two- or even three-tiered pattern, which was unexpected. Some instruments showed characteristic swellings on some notes.

9) Only in the string instruments does the size of the instrument seem to have any effect on the overall behavior of the transition, and then only in the power analyses. In the cello it is difficult to distinguish the amplitude dips of the bow-change and no-bow-change cases.

## 7 ACKNOWLEDGMENT

This work was performed at the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University and was supported in part by the System Development Foundation. The author gratefully acknowledges the contributions of: 1) John Chowning, Earl Schubert, James A. Moorer, John Pierce, and John Grey for their support and guidance; 2) Mark Dolson and John Gordon for their assistance with the DSTFT; 3) Julius O. Smith for the peak-following algorithm; 4) Phil Gossett for the modification to the Sony F1; and 5) Emily Bernstein, David Burkhardt, Angela Sohn, Dexter Morrill, and David Jaffe for recording the test tones. The support of the Stanford Artificial Intelligence Laboratory, Lucasfilm, and The Droid Works is also gratefully acknowledged.

#### **8 REFERENCES**

[1] J. M. Grey, "An Exploration of Musical Timbre," Ph.D. dissertation, Dept. of Psychology, Stanford University; Dept. of Music Rep. STAN-M-2, 1975.

[2] J. W. Beauchamp, "Data Reduction and Resynthesis of Connected Solo Passages Using Frequency, Amplitude, and 'Brightness' Detection and the Nonlinear Synthesis Technique," in *Proc. 1981 International Computer Music Conf.*, L. Austin and T. Clark, Eds. (North Texas State University, Denton, 1981), pp. 316-323.

[3] W. C. Campbell and J. J. Heller, "The Contribution of the Legato Transient to Instrument Identification," presented at the Research Symp. on the Psychology and Acoustics of Music, University of Kansas, Lawrence, 1978.

[4] R. Cogan, New Images of Musical Sound (Harvard University Press, Cambridge, MA, 1984).

[5] M. B. Dolson, "A Tracking Phase Vocoder and Its Use in the Analysis of Ensemble Sounds," Ph.D. dissertation, California Institute of Technology, Pasadena, 1983.

[6] D. Morrill, "The Dynamic Aspects of Trumpet Phrases," in French, E. Gresset, transl. (IRCAM, Paris, 1980).

[7] H. Rösing, "Probleme und neue Wege der Analyse von Instrumenten- und Orchesterklängen," Ph.D. dissertation, University of Vienna, 1967 (Verlag Notring der wissenschaftlichen Verbände Österreichs, Vienna, Austria, 1970).

[8] C. Seashore, *Psychology of Music* (McGraw-Hill, New York, 1938; reprinted by Dover, New York, 1967).

[9] D. C. Miller, *The Science of Musical Sounds* (Macmillan, New York, 1926).

[10] J. Strawn, "Modeling Musical Transitions," Ph.D. dissertation, CCRMA/Dept. of Music, Stanford University; Dept. of Music Rep. STAN-M-26, 1985.

[11] J. Borish, "Electronic Simulation of Auditorium Acoustics," Ph.D. dissertation, School of Engineering, Stanford University, Dept. Music Rep. STAN-M-18, 1984.

[12] E. A. Cohen, "The Influence of Nonharmonic Partials on Tone Perception," Ph.D. dissertation, School of Engineering, Stanford University, 1980.

[13] J. W. Gordon, "Perception of Attack Transients in Musical Tones," Ph.D. dissertation, Department of Music, Stanford University: Dept. of Music Rep. STAN-M-17, 1984.

STRAWN

[14] J. A. Moorer, "Signal Processing Aspects of Computer Music—A Survey," *Proc. IEEE*, vol. 65, pp. 1108–1137 (1977); reprinted in *Digital Audio Signal Processing: An Anthology*, J. Strawn, Ed. (William Kaufmann, Los Altos, CA, 1985), pp. 149–220.

[15] J. O. Smith and P. Gossett, "A Flexible Sampling-Rate Conversion Method," *Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing*, vol. 2 (San Diego, 1984 March), pp. 19.4.1-19.4.2.

[16] L. Rush, "The Tuning of Performed Music," presented at the 1982 International Computer Music Conf., Venice, Italy.

[17] A. H. Benade, Fundamentals of Musical Acoustics (Oxford, New York, 1976).

[18] J. O. Smith, personal communication, 1984 March 4.

[19] B. Hutchins, Ed., Musical Engineer's Handbook (Electronotes, Ithaca, NY, 1975).

[20] J. Strawn, "Editing Time-Varying Spectra," J. Audio Eng. Soc. (to be published).

[21] J. C. Risset, "Computer Study of Trumpet Tones," Bell Laboratories, Murray Hill, NJ; J. Acoust. Soc. Am., vol. 38, p. 912 (1965).

[22] J. W. Gordon and J. Strawn, "An Introduction

to the Phase Vocoder," in *Digital Audio Signal Processing: An Anthology*, J. Strawn, Ed. (William Kaufman, Los Altos, CA, 1985), pp. 221–270.

[23] J. Strawn, "Analysis and Synthesis of Musical Transitions Using the Discrete Short-Time Fourier Transform," J. Audio Eng. Soc. (to be published).

[24] H. Backhaus, "Über die Bedeutung der Ausgleichsvorgänge in der Akustik," Z. tech. Physik, vol. 13, pp. 31-46 (1932).

[25] J. Beauchamp, "A Computer System for Time-Variant Harmonic Analysis and Synthesis of Musical Tones," in *Music by Computers*, H. von Foerster and J. W. Beauchamp, Eds. (Wiley, New York, 1969), pp. 19-62.

[26] G. Charbonneau, "Timbre and the Perceptual Effects of Three Types of Data Reduction," *Computer Music J.*, vol. 5, no. 2, pp. 10–19 (1981).

[27] M. D. Freedman, "A Method for Analyzing Musical Tones," J. Audio Eng. Soc., vol. 16, pp. 419-25 (1967).

[28] D. A. Luce, "Physical Correlates of Nonpercussive Musical Instrument Tones," Ph.D. dissertation, Dept. of Physics, Massachusetts Institute of Technology, Cambridge, 1963.





John Strawn was born in 1950 in Ohio. He received a bachelor of Music degree (double major in organ and music theory) from Oberlin Conservatory in 1973. From 1973 through 1975 he studied music history and theory on a Fulbright in Berlin. In 1976 he travelled through the Middle East and Asia to Japan, where he conducted independent research on the electronic music scene with a Thomas Watson Fellowship.

Strawn studied with John Chowning at the Center for Computer Research in Music and Acoustics (CCRMA), Department of Music, Stanford University, graduating with a Ph.D. in 1985. During his years at Stanford he was an editor of The *Computer Music Journal* (1977–1982), and served as a consultant for a number of digital audio synthesizer manufacturers such as Mattel Electronics, Music Technology (GDS and Synergy synthesizers), and Kurzweil. In 1984 he established the Computer Music and Digital Audio Series (published by William Kaufmann, Inc.), for which he is series editor. Together with Curt Roads, he has published *Foundations of Computer Music* (MIT Press, 1985) and is the author of numerous papers and translations.

From 1985 January through 1986 September, Dr. Strawn was a member of the Digital Audio Research and Development Group of The Droid Works, an affiliate of Lucasfilm. There he was primarily involved in digital audio signal processing research (including microcode implementation) for the Droid Works' digital audio products such as the SoundDroid. He is now active as a consultant, working in digital signal processing and computer programming. A member of the AES, he is Chairman of the 1987 International Conference on Music and Digital Technology.