BEYOND THE SPECTRUM OF MUSIC:

AN EXPLORATION THROUGH SPECTRAL ANALYSIS OF SOUND COLOR IN THE ALBAN BERG VIOLIN CONCERTO

by

Diego Bañuelos

A written project submitted in partial fulfillment of

the requirements for the degree of

Doctor of Musical Arts

(Music)

at the

UNIVERSITY OF WISCONSIN-MADISON

2005

To Cathy Ann Elias

CONTENTS

ILLUSTRATIONS

SOUND EXAMPLES

PART I

INTRODUCTION

Among the arts, music is privileged to enjoy a particularly rich vocabulary that describes its various objects and phenomena. For centuries and in more than one culture, we have been fascinated by attempts to describe and explain—through language—the ways in which sounds come together to be heard as a meaningful whole. We tend to separate this complex web of events that happen over time into various phenomena that we call rhythm, melody, harmony, and the like. Our understanding of these elements and the ways in which they interact has changed constantly over time, as emphases shift from culture to culture, period to period, composer to composer, and even listener to listener. Much as it is often arbitrary (at least in detail) within the wholeness of that language we call music, this partitioning into discrete phenomena seems to help us understand the ways in which that very wholeness is achieved.

One such element that has received increased attention throughout the 20th century and into the 21st is what I call sound color: the inner life of each sound as a moment in time (however long a "moment" may be), perceived as a whole whether produced by a single source or a combination of sources, and whether it's composed of definite pitches or noises or both. "Sound color" as it is used here, is a term that can be compared to Schoenberg's *Klangfarbe*, as it is described by Alfred Cramer (2002). Cramer claims that Schoenberg's usage of the word *Klangfarbe* "refers to an idealized hearing of tones for the timbral contributions of [spectral] frequencies rather than of pitch values. The timbres of *Klangfarbenmelodie,* then, result from pitches heard alone or in harmonic combination; such *Klangfarben* are not attributes of discrete tones, and they are not distinct from pitch" (2).

Theodor Adorno describes a similar notion of this element in his *In search of Wagner* (1991). He calls it "sonority" and he points to its "two dimensions of harmony and colour [*sic*]." He adds: "Through sonority, time seems transfixed in space, and while as harmony it 'fills' space, the notion of colour, for which musical theory has no better name, is directly borrowed from the realm of visual space" (63). Adorno further explains his concept of "colour" by relating it directly to orchestration. "Orchestration in the precise sense," he writes, is the "productive share of colour in the musical process 'in such a way that colour itself becomes action'" (71). We can thus think of sound color as an amalgamation of harmony and timbre. Even though the two dimensions can be thought of separately in any given context, they often intertwine, especially in complex orchestral textures, creating an overall impression of how the sonic tissue is fashioned as a whole.[1] Independently of how a sound is produced or notated, one could simply describe its sound color as the configuration of its partial frequencies, represented in its Fourier spectrum.[2]

The search for new colors or new "sounds", whether through the creation of new sources or by recombining existing ones in ever-new ways, has been a recurring theme that can be traced back to the very origins of music, and about which much could be written if we looked at the last hundred years alone. The emphasis Debussy placed on sound color, for instance, often at the expense of harmonic progression and even melody, has almost come to

[1] For a more thorough discussion of the relationship between harmony and timbre and their specific spectral characteristics see pp. 27-29.
[2] For a definition of Fourier spectrum see p. 15.

define his style, and it has proven extremely influential to this day (Morgan 1991, 46). Composers such as Messiaen, Boulez, and Stockhausen have repeatedly honored this influence in their work, where time and again our attention is directed to the inherent sonic properties of each particular moment, as inseparable as they may be from the structural considerations that bring them about (Griffiths 1995, 31). Stockhausen remarked of the way in which he meant his "through-organized" pieces to be heard: "One stays in the music [and] needs nothing before or after in order to perceive the individual now (the individual sound)" (37).

It was the search for new sounds that drove composers like Henry Cowell, Harry Partch, and John Cage (and many others thereafter) to experiment with the vast sonic possibilities provided by unorthodox instrumental techniques, such as playing inside the piano, preparing it in various ways, or in the case of Partch, inventing wholly new instruments. It was also one of the driving motives (and ultimately one of the main results) that led to the development of electronic music. Once again, it was an effort to bring timbre to the status of importance and organization enjoyed by melody and rhythm that drove Stockhausen to make his first electronic compositions in the early 1950s (Griffiths 1995, 45). Scientific research suggesting that any sound could be decomposed into pure frequencies or sine tones, also suggested that, conversely, new sounds could be systematically synthesized from pure tones, therefore allowing the creation of a kind of organized timbral system. Even though the results might not have been the creation of an absolute "timbral organization," as he had hoped, his efforts did help to uncover an immense field of exploration (computer-generated music) which has fascinated composers for more than half a century and that continues to yield ever-new sonic worlds today.

A further step in the exploration of what I have defined as sound color can be seen in the so-called textural music of composers like Penderecki and Ligeti. In their case, a conscious step was taken not only to emphasize this parameter, but to actually obliterate all other elements including the notions of melody and rhythm, in an effort to shape music "through its larger sonic attributes rather than as an accumulation of individual details" (Morgan 1991, 386).

Ironically, sound color is one of the elements of music for which our vocabulary is most limited and least precise, perhaps because the property itself is one of the most elusive and hardest to define and pinpoint. Thus far it has been difficult to talk objectively about sound color due to the complexity of factors that affect its structure. And yet, its impact on the listener is usually immediate and powerful. A single texture played for as little as one bar before any melodic material in the proper sense is introduced can serve to set up the tone or mood of an entire movement, as is the case in the opening measures of the Mendelssohn or the Sibelius Violin Concertos. In cases such as these, the instrumentation and voicing of a single chord, as well as the figurations used to elaborate it, all come together to create a characteristic sound color that serves to set up the right atmosphere for the introduction of the first thematic idea.

One of the difficulties of attempting a comprehensive discussion of sound color stems from the fact that responses to it tend to be extremely personal and subjective, and are often inseparable from notions of affect, mood, and emotion. As a performer and a violin student, I can recall countless instances in which my attention was directed to color—both in my own tone and in the composite texture of a particular ensemble—in order to achieve a specific affect or to characterize a certain emotion. Violinist Charles Treger, for instance, often stressed that in any piece of music, notions of emotion and feeling had to be looked for in its vertical organization—that is in the change and evolution of harmonies, textures, and colors. For this reason, whenever I was to play a melodic line on the violin, he stressed the importance of knowing and understanding the vertical structure of the accompaniment that supported it. Furthermore, he often placed the relevance of color or chord quality, above that of "function" within a chord progression, as he believed quality played a stronger role in the creation of affect. Only if I understood the colors of my accompaniment, he would say, could I understand how they related to the melodic line, enabling me to shape it and color it accordingly, through my own tone.

I also heard Daniel Baremboim in rehearsal speak of a similar outlook in which he saw harmony as the heart and carrier of emotion in music. Violinist Midori, on the other hand, recently told me I should always put the full "character" of a passage in the tone of every note I play, so that if any note was isolated from the rest and listened to by itself, it would still transmit the character of the entire phrase. Even Schoenberg himself associated the world of sound color somehow with that of the emotions, while lamenting at the same time our lack of vocabulary to describe it and understand it. In his *Harmonielehre* (1978), he writes that "our attention to tone colors [*Klangfarben*] is becoming more and more active, is moving closer and closer to the possibility of describing and organizing them. At the same time, probably, to restrictive theories, as well. For the present we judge the artistic effect of these relationships only by feeling" (421).

But what are these "relations" of which Schoenberg spoke, and which lend themselves so well to interpretation? How can sonorities have such an immediate impact on us? Surely this question cannot be fully answered without plunging the depths of the human psyche and their relationship to its social context. But there must be something material and tangible in sound itself that sparks the flame of emotion, independently of what the specific emotion might be and whether it is the same in every listener. In fact, it is this ambiguity that makes music so wonderfully powerful, allowing it to speak in a slightly different way to each one of us, and ultimately, allowing us to communicate with one another.

Somewhere between a merely acoustical phenomenon and the mysterious depths of the human psyche, the miracle of music takes place; surely the relationship between the two poles must not be arbitrary. Perhaps the key lies in the word “relationship.” Both the world of sounds and the world of human emotion and experience are but a complex web of relationships, and for some reason, the relationships within sound as it exists in time and through time, lend themselves particularly well to interpretation and analogy to the relationships in that other realm, the one of human experience.

So far, the language we have used to describe the characteristics of sound color in music has been tightly linked with our interpretation of it. Given that the material relations that rule its nature have been almost entirely hidden to us until very recently, we have resorted to those relations which are closer to us, the ones in our own experience. Thus we often describe sounds using words like “somber”, “cheerful”, “sorrowful”, or even “loving”—as in bar 152 of the Alban Berg violin concerto, movement II, marked *molto espressivo e <u>amoroso</u>*. But this makes it very hard to speak objectively about sound color and to separate—at least in our speech—our interpretation, from those events that actually happen in the air before they step into our ears to ignite our imagination. Another approach has been to describe sound color by describing its means of production, mainly through instrumentation and performance techniques. But once again, this tells us little about the inner life of the sounds themselves. All this approach does is evoke familiar sounds and use our aural memory as a way of description.

It was only when spectral analysis was developed that we could peer into the actual structure of sound and have a glimpse of the myriad relations that govern its shape. The first person to point out the correlation between tone color and the harmonic structure of a sound was Hermann von Helmholtz in his ground-breaking *On the Sensations of Tone* (1954) first published in 1863. Helmholtz used resonators as his sole technology for analyzing sound. Since then, the science and technology of sound analysis have grown at exponential rates. Interestingly, though, most of this knowledge, has been applied to speech science, sound synthesis, and signal processing, while its application to music theory and to musical analysis in particular has received comparatively much less attention.

Even though the potential of these advances to further our understanding of sound color within musical contexts has intrigued many (among them Stockhausen, as noted above), the first attempts to exploit this potential within the framework of actual musical compositions were not made until the 1960s and 1970 when Robert Cogan and Pozzi Escot published their *Sonic Design* (1976). Their pioneering work, however, was significantly slowed down by the limitations of the technology available to them at the time, as Cogan himself remarked later in *New Images of Musical Sound* (1998). In this work, "spectrum photographs," photos of the spectral formation of musical contexts, allowed Cogan to discuss for the first time and in great detail what he called the "sonic design" of entire compositions. He, then, went on to attempt a comprehensive theory of tone color based on spectrum analysis and modeled on the phonological theory of oppositions.

The seminal importance of this theory can hardly be ignored. However, it is but the beginning of what could become an entirely new aspect of musical analysis. In the conclusion of *New Images of Musical Sound* Cogan writes that "music's sonic flesh and substance have been oddly unnamable. Not until quite late in its history has it become possible to approach, picture, and name its sonic essence. . . . Still, it is [now] possible to return form this terra incognita with initial pictures and names, which will undoubtedly be refined and enriched by later explorers. . . . Within its newly envisioned boundaries, and among its newly glimpsed details and forms, myriad discoveries lie ahead" (153).

The present study follows in the footsteps of Cogan's *Images* in that it is based on spectral analysis. However, it differs from Cogan's in one fundamental way: whereas Cogan works with "spectrum photographs," which he goes on to interpret by eye, I am working with digital sound, which can be analyzed as a series of data points with the use of a computer program. This opens an enormous range of possibilities that were simply not available to Cogan with the technology he possessed at the time he wrote his *Images*. First of all, spectrograms (like the ones found in appendix B) can be plotted with much more precision, and the ways to display their information can be controlled more precisely to achieve optimal clarity. But more importantly, once the spectrum has been found, it is possible to treat it, in turn, as a series of data points, which can be analyzed as well.

One of the great problems one experiences when looking at a spectrogram, especially when it represents a complex texture, is that it contains more information than we can fully process with the naked eye (one second of recorded sound consists usually of 44,100 data points). What we see in one page of a spectrogram is a visual representation of literally all the information that creates its corresponding sound and gives it the qualities we perceive though the ear—as well as many we don't.[1] Furthermore, the spectrogram allows us to visualize sound as it evolves through time, hence displaying at once all the information contained in long stretches of music. If we were to listen to the sound we would receive this information on a moment-by-moment basis, but in a spectrogram the time domain simply becomes another axis of the plot.

By looking at enough pictures one can begin to associate certain shapes with specific sonic phenomena, and even recognize certain families of timbres, provided they are presented in isolation and as single sounds. But this is not to say that we can truly analyze and understand the nature of sound simply by looking at its spectrogram, especially when dealing with complex textures and larger musical contexts, and not just with single, isolated sounds. Generally speaking, our eye will only be able to identify in the spectrogram the most general characteristics of a sound. Rarely can we say something that will truly enrich our understanding of a sonority that we can otherwise hear and see represented in a musical score. But the idea is to use technology to further the understanding we already have of musical sound. Therefore, if what the spectrogram tells us can also be gathered from a regular score, it is of little use to us.

Most of the properties of rhythm and pitch (melody) fall in this category. On the other hand, details of a spectrogram that can't actually be distinguished by the human ear, are once again utterly useless in terms of our better understanding of music. It is in the realm of sound color that spectrum analysis can provide unprecedented insight, perhaps even the kind

[1] For a complete explanation of how a spectrogram works and how it is derived from actual sound see Part II, pp. 14-19.

Schoenberg yearned for as noted in the quote above. The key lies in simplifying the amount of information we display in a way that it isolates certain elements of interest.

My goal is to be able to focus on specific qualitative properties of sound and trace their development through time within a musical context. It is not my intention to simply describe the elements of isolated sounds, but to explore the *relationships* among elements of sound in large musical contexts, both momentarily and through time. Given the complexity of the objects in question, the only way to do this is to systematically analyze the spectrum through strict mathematical methods; in other words, to find concrete relationships between the elements of the spectrum through mathematical algorithms, in order to provide ways to express these relationships more succinctly. Due to the sheer amount of data within each sound, this can only be done with the aid of a computer.

The genesis of this kind of analysis can be found in William Sethares's *Tuning, Timbre, Spectrum, Scale* (1998), where he develops a procedure that yields what he calls a "dissonance score" of a piece (a Scarlatti harpsichord sonata), which he goes on to interpret in conjunction with the actual musical score (189).[1] Sethares shows that this dissonance score can be used in conjunction with the musical score to discuss notions of dissonance. Following in his footsteps I have adapted this procedure and created a number of others, all with the same basic objective: to condense the amount of information in the spectrum, zeroing in on specific relationships between its elements, thus isolating, through mathematical methods, material and aurally relevant features of sound color. This yields new

[1] A comprehensive discussion of Sethares's dissonance scores the kind of results they may yield can be found on pp. 30-35.

visual representations that can be more easily interpreted and understood than the raw spectrogram.

The first step in this method of analysis is to simplify the spectrum in order to create an abstract model of it in which its elements can be easily separated and described, facilitating further analysis by a computer program. One way of doing this is to create an alternate spectrum in which harmonic elements (pitched components) and noise elements (non-pitched components) are separated into discrete lines.[1] The new graph is a greatly simplified version of the spectrum but it nonetheless still contains most of its salient attributes. A further step is to operate on these elements in such a way as to gather information about how they interact. For instance, we can determine how much energy is contained in the higher frequencies compared to that in the lower frequencies. This would give us a reading that could be called the "brightness" or the "acuteness" of a sound. If we analyze in this way an entire beat (or any other arbitrary measurement of a "moment"), we will get a single number assessing its "brightness." Naturally, this number itself is irrelevant as far as its magnitude is concerned, but we can perform the same operation on a number of consecutive beats within a musical context and plot them against time. With the new graph—a kind of "brightness score"—we can gather invaluable insight on the way this aspect of sound develops over the course of a passage.

The actual word we choose to describe this phenomenon is of less consequence to this study than the phenomenon itself. The reading, which we could more accurately call "registral balance" for example, will be no different than any other descriptive label in the usual vocabulary of music theory, such as "triplet" or "dominant seventh." In the same way

[1] This technique is discussed at length on pp. 23-27.

in which these labels are meaningless for their own sake, our studies won't be complete unless they can be interpreted and used to further our understanding of a piece and to facilitate our communication of that understanding. Furthermore, such an interpretation may (and probably ought to) vary from listener to listener, which takes nothing away from the validity of the readings. In the end, the point is to incorporate the concrete information we acquire about the particular sounds in a piece of music into a comprehensive and living-breathing interpretation of it.

It should be noted that even though the various readings obtained through the kind of method described above will always reflect some tangible aspect of the sound or a relation between its elements, there is always a question of whether such a property is actually relevant as far as its aural perception is concerned, and if so, in what measure. Surely one could divide each element in the spectrum by the next, for instance, and plot the results. The new graph certainly reflects material relationships between elements of the spectrum, but it would be hard to know how to interpret them in a way that could be mapped to our aural perception of the sound. Once we have established the significance of a particular reading, the next thing we can ask is just how well it is isolating and representing the sonic property it describes. And it won't be until substantially more research has been done in this field that we will be able to arrive at a definite answer to this question, if we ever do. Once again, with this question we approach the rickety bridge that separates the world of physical phenomena and that of human perception.

The Part II of this paper presents the various procedures I have developed or adapted (mostly from Sethares's book), and the way in which I have gone about dealing with the questions and problems presented above. Even though some issues of perception will be necessarily dealt with in this section—being as inseparable as they are from the creation of effective analytical routines—the question of interpretation is the focus of Part III. Here, the analysis and the synthesis are performed on the second movement of the Alban Berg Violin Concerto. Using the techniques developed in Part II, I create a number of graphs and alternate "scores" that enable me to discuss comprehensively the ways in which sound color is used in the concerto to dramatize its program and to provide unity and form. What I intend to present in this section is nothing but my own personal interpretation of the piece as enriched by the results of my spectral analysis. Rather than proposing this interpretation as an ultimate conclusion, I intend it as an invitation for the reader to examine and interpret the results of my sonic analysis in the same kind of personal way. It is my belief that only after many people have crossed that rickety bridge between acoustic fact and human interpretation time and again can we begin to probe that "miracle" that links sound with music; only by looking ever closer at the breach between the physical and the psychological can we further our understanding of the former and the latter.

PART II

ANALYTICAL PROCEDURES

Spectrum and Spectrogram

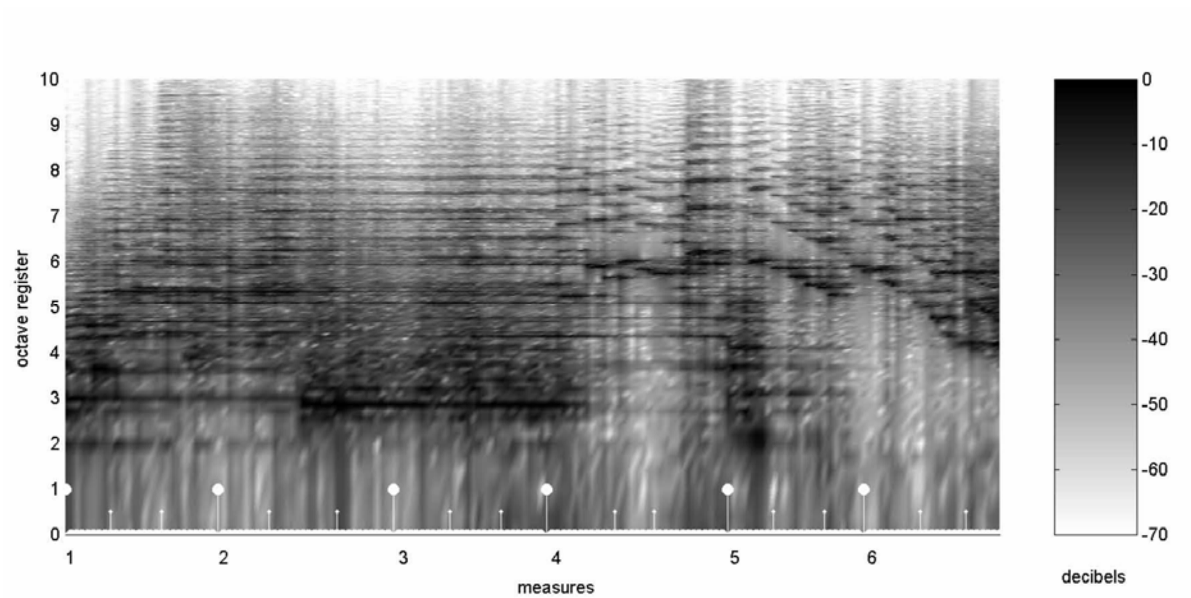


Figure 2.1. Spectrogram of the first six measures of Alban Berg's Violin Concerto, movement II performed by Ann Sophie Mutter and the Chicago Symphony Orchestra, conducted by James Levine (1992).

Figure 2.1 is a spectrogram of the first six measures of Alban Berg's Violin Concerto, movement II (1992). The horizontal axis represents time, which is marked in musical beats and measures, and the vertical axis shows the frequency components of sound (partials) measured in octave registers (4=C_4). Their respective sound levels, expressed in decibels, are indicated by the color map: a black line is the loudest sound and a white patch is anything less than 70 dB below that sound.

The spectrogram is a widely used visual representation of sound, with numerous applications in disciplines that include linguistics, voice pedagogy, sound synthesis, sound recording, and music theory. [1] This kind of representation is based on the principle that any complex periodic wave can be expressed as a sum of simple sine waves, each with a specific frequency, amplitude and phase. The procedure used to obtain these component waves is called the Fourier Transform—after Jean Baptiste Joseph Fourier, who first developed it in the early 19th century. This mathematical model is particularly useful in the study of sound because the human ear analyzes sound in a very similar way. Furthermore, the disposition of the Fourier spectrum is closely related to our perception of timbre, a fact first pointed out by Helmholtz (1863) and corroborated by psychoacoustical research thereafter (Sethares 1998,17).

When analyzing digitized sound, the Fourier spectrum can be obtained using an algorithm called the FFT (Fast Fourier Transform), which is usually built into standard audio processing software and numerical packages such as MATLAB and Mathematica. What this algorithm does is transform a series of data points from the time domain into the frequency domain. In other words, if we input one second of digitized sound consisting of 44,100 data points (the standard sampling rate for commercial recordings), the FFT returns 44,100 frequency points. Only the first half of these points is used to create the spectrum, that is, a series of 22,050 numbers corresponding to frequencies raging from 1 to 22,050Hz. Each one of these numbers has a magnitude that represents how much energy at that particular

[1]Robert Cogan's *New Images of Musical Sound* (1998) provides a good example of how spectrograms can be applied to music theory. Similar applications can be found in the budding discipline of spectromorphology, which is used in electroacoustic music to "describe and classify the way sound spectra are shaped in time" (John Young, "Denis Smalley ," in *Grove Music Online* ed. L. Macy, [accessed 20 February 2005], <http://www.grovemusic.com>).

frequency is present in the original signal.[1] But when using the FFT to analyze real-life recorded sounds a few problems arise. First of all, the FFT assumes that each sound is perfectly periodic and hence infinite in duration. However, real sounds aren't absolutely periodic, and certainly no sound lasts eternally. So we must pretend that the signal analyzed is periodic with period equal to the length of the analysis window. As a consequence, there will be a small distortion in the spectrum. This takes us to the next problem: How long should analysis windows be? How long is a "moment"? What is a "single" sound?

Invariably the answer to these questions will be arbitrary. No acoustically generated sound is absolutely steady in a way that it could be appropriately called a "single" sound. In fact, transients—rapid and minute variations in a vibrating pattern—play a determining role in the recognition of timbre (Roederer 1995, 123). The most common approach is to take the FFT repeatedly for several adjacent small windows (0.2 seconds or so). All these "momentary" Fourier spectra are then collected together to create a *spectrogram* like the one displayed in Figure 2.1.

This kind of display shows quite accurately how the overall spectrum changes with time, but the problem arises that small analysis windows result in poor resolution of the image. For example, using a window that is one tenth of a second long, the lowest frequency point in the spectrogram is at 10Hz, and all subsequent frequencies are rounded off to multiples of this "fundamental", smearing the image considerably. This is particularly problematic in the low register due to the fact that the ear perceives pitch level on a logarithmic scale—meaning it perceives *powers* of two as being octave equivalences (Roederer 1995, 26). Figure 2.2 illustrates this fact by comparing the linear representation of

[1] Phase is usually discarded since it plays only a minor role in the perception of timbre (Roederer 1995, 150).

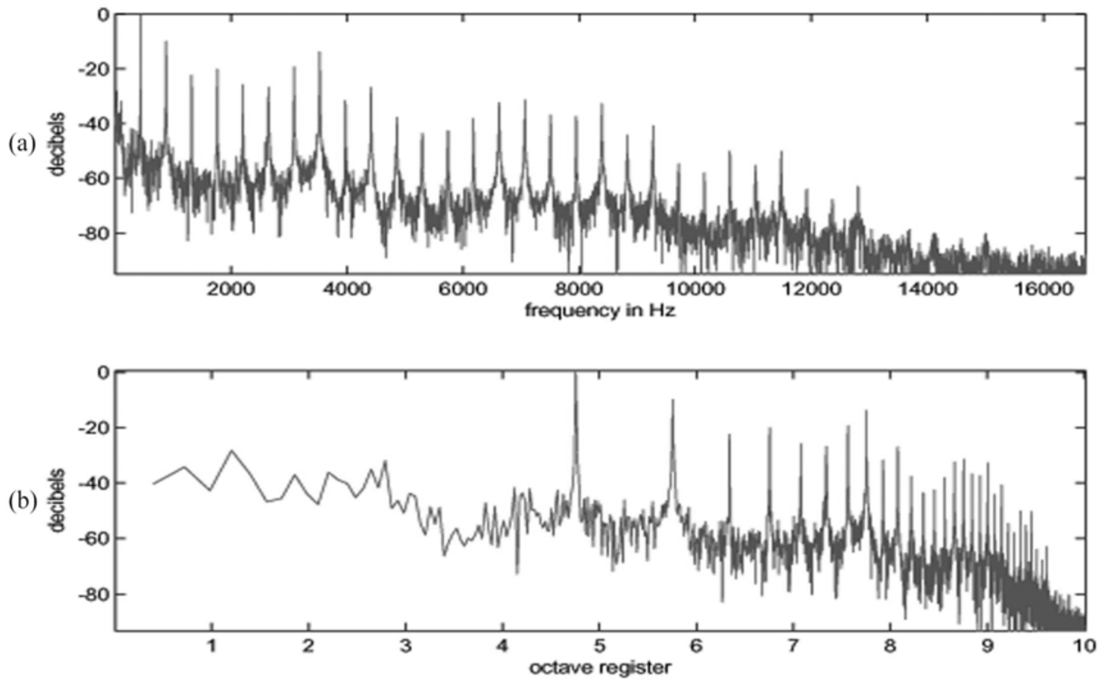


Figure 2.2. Spectrum of a violin open A-string analyzed over a period of 0.19 seconds. In (a) the horizontal axis represents linear frequency, whereas in (b) frequency is represented logarithmically in octave registers. In both cases the vertical axis represents relative intensity measured in decibels below the loudest partial.

a spectrum (a), with its counterpart in logarithmic, or octave register scale (b). In both cases the sound of a violin open A-string is being analyzed over a period of approximately two tenths of a second.[1] It is not hard to see from these pictures that as we move down the octave register there are fewer and fewer samples, and the resolution gets poorer and poorer.

In the case of an actual spectrogram, this distortion is not particularly alarming. After all, the spectrogram is only a visual representation, meant to be looked at with the naked eye and nothing more; whether the noise in the low register is shown to be 5 dB more or less than what it ought to be, or whether a particular partial appears a half step sharper or lower than it should will not change much the overall *visual* impression of a passage. Accurately

[1] The source signal can be found in audio track 1

representing the rapid changes that happen over time is more important, as these changes may constitute some of the most salient features of a spectrogram, especially in fast passages. For this reason, in all the spectrograms that appear in this paper I have used analysis windows of about a tenth of a second, choosing to represent time more accurately than frequency. For more rigorous kinds of analysis, like the ones found in the various algorithms developed below, I will rapidly turn to larger-sized windows, as will be discussed on p. 36.

One thing my spectrograms do *not* show is dynamic change or absolute sound intensity levels. I chose to ignore this parameter for a number of reasons, not least of which is our inability to calculate any kind of absolute sound intensity level out of a recording. Rather, I chose to show the spectral makeup of each "sound" as clearly as possible by scaling the spectrum of each "moment" to its loudest partial. Thus, using the decibel scale[1], I have placed the highest peak at zero for each window, measuring all other partials according to how soft they are in comparison to this peak. In this way, we can always see clearly what the most prominent frequencies are at any time, regardless of their actual intensity in comparison to other moments in the piece.

Figure 2.3 shows in three stages how the spectrogram is put together. Once more, the sound represented here is a violin open A-string. It is worth stressing that this is *only* a visual representation, an analogy of the sound, and not the sound itself. In other words, a spectrogram does not pretend to *replace* sound, but rather to provide added perspective and insight in order to illuminate the way we perceive it aurally. The first image (a) is an FFT of the first window (approx. 0.1 seconds). This is done about 30 times, and the resulting curves

[1] By definition the dB scale does not describe an amount of intensity but rather a ratio between intensities; for example, a sound that carries ten times more energy than another will be 10dB higher (Hall 1991, 74).

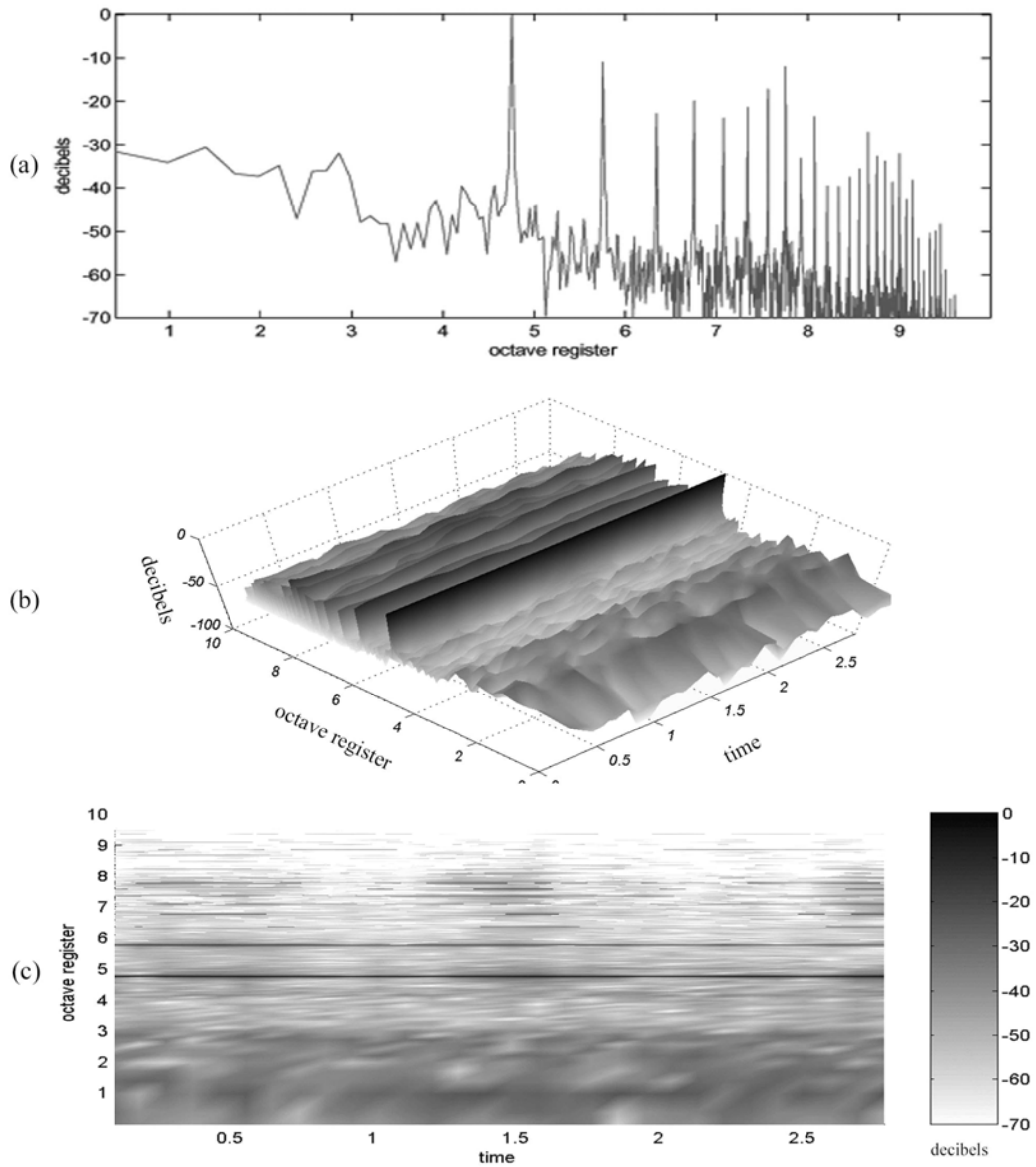


Figure 2.3. Stages in the construction of the spectrogram: (a) Fourier spectrum of the first window (0.09s). (b) Surface map made up of multiple spectra strung together. (c) Flat image in which intensity is represented by the color map alone.

are collected together to create the surface in (b). Each spectrum curve can be seen as “slice” of this surface. The last step (c) is to “flatten” this picture out (or simply look at it from above) so that intensity level is only shown by the color map, making the image more intelligible.

Noise and Peaks

As mentioned earlier, my goal is to be able to isolate aurally relevant elements of the spectrum and find relationships between them. But before doing that, it is necessary to first label such elements, both conceptually and electronically. What the FFT provides is nothing but a list of numbers, and even though we can *visually* pick out various objects, such as particular frequencies that stick out above the others, there is nothing in the actual FFT differentiating these objects from the frequencies around them. So we must create yet another representation—a kind of abstract model of the spectrum—in which various elements appear separately and can be recognized by a computer for further analysis.

The ear can separate and catalogue simultaneous sounds extremely well, to the point that it can isolate single instruments in a complex orchestral texture. Our technology could not come near such a level of sophistication in pattern-recognition. Even simulating and automatizing the pattern-recognition process we perform when *looking* at the original spectrum (itself a much more rudimentary process than the aural response) is far from trivial, as we will see next.

Consider figure 2.2 (a): right away we can see a number of equally spaced peaks, corresponding to the various overtones of the fundamental frequency of the violin (A_4). It is easy to visually separate these peaks from the “noise” around them, some of which is created by the numerical errors implicit in the procedure and some of which comes from the actual acoustic noise of the signal. Most of the information in this picture corresponds to noise of either kind. However, most of it lies more than 50dB below the fundamental frequency (the largest peak). A difference of 50dB is roughly the difference between the noise level in an empty auditorium and the noise level in a factory. This would mean that whereas the harmonic elements sound at a more or less *mezzo piano* level, most of the noise has the intensity of a falling pin[1]. So it is safe to assume that in this particular example the peaks are the most prominent features of the sound (they are the loudest frequencies). This is not to say that the noise is unimportant all together. However tenuous it is, the noise of the bow rubbing against the string can have an impact on our recognition of this particular timbre. Certainly, though, we don't need 40,000 or more data points to represent it.

Juan Roederer (1995) defines noise in the following way: "When the vibration patterns are randomized [...] or when their complexity exceeds a certain threshold, the neural processing mechanism simply gives up: no definite pitch and timbre sensations can be established. The ensuing sensation is called noise" (171). It makes sense, then, to represent noise not as a collection of frequencies, but rather as a single line that nonetheless can have some registral characteristics. This line can be drawn by passing the magnitude of the FFT through a median filter, as shown in figure 2.4.

The next step in creating the model is to pick the peaks that allegedly best represent the pitch content of the sound. In theory—for most western instruments—these peaks ought to form harmonic templates in the signal (integer multiples of a fundamental), corresponding to the perception of pitch. In practice, however, truly harmonic templates are rare, partly because of round-off errors, partly because many real-life sounds that we perceive as having a pitch are not entirely harmonic in all registers—like the piano, for instance, where upper partials tend to be slightly sharper than expected due to string distortion (Roederer 1995, 110). For this reason the peaks in the FFT graph can not be properly understood only as the "harmonic elements" of the sound (although harmonic elements may be included). Rather,

[1] All comparisons between absolute intensity levels in the dB scale are taken from Hall (1991, 76).

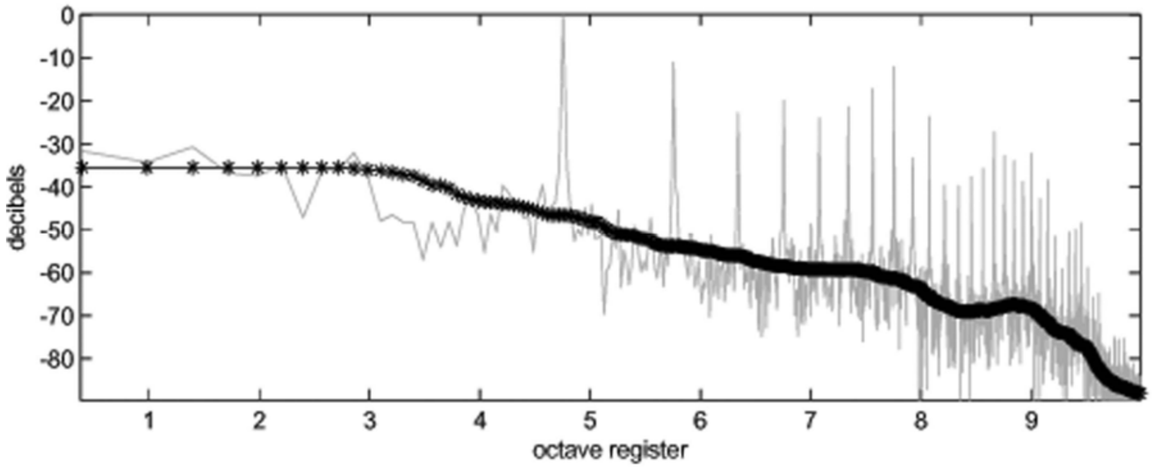


Figure 2.4. Spectrum of a violin A-string. The starred line represents the noise floor, which was calculated using a median filter.

they can be seen as "the most significant spectral frequencies (and their magnitudes)" to quote Sethares (1998, 190).

Needless to say this raises the question of how significant is "significant." And the answer to that question will invariably be subjective. This is why a pick peaking algorithm, such as the one Sethares (1998) uses for his dissonance scores (295) or the one I am about to present, is not grounded in an actual mathematical definition of "peakness." Rather than claiming any kind of absoluteness or definitiveness, this peak-picking algorithm is an attempt to emulate the kind of process one would follow when choosing peaks visually, and as such it is not devoid of personal choice. This would make it hard to justify conclusive statements about quantitative attributes of specific sounds drawn from this algorithm alone. However, the model proves useful within a setting in which various quasi-harmonic sounds are evaluated in the same way and the algorithm is used to elucidate *qualitative* relationships between them.

My peak-picking algorithm[1] takes Sethares's as a point of departure, adapting it to automate some of the human choices his involved. This is done taking under account the context in which the algorithm is meant to be applied: The Berg Violin Concerto (or other similar orchestral pieces). The original routine has five basic steps: (1) Finding all local maxima in the spectrum. (2) Filtering out the noise by ignoring all local peaks with amplitude less than a *specified* fraction of the highest peak—this would, for example, disregard anything below 40dB or 50dB in figure 2.4, depending on the input. (3) Discarding anything below a given fundamental and above a certain top frequency—this is useful if we actually know the fundamental of the sound to be analyzed, but dangerous otherwise. (4) Making sure adjacent peaks are at least a certain distance (in Hz) apart. This is done by choosing only the highest peak within bands of a specified width. The reason for this step is that often the peaks we see in a spectral image consist of several data points, which often create sub-peaks, so to speak. (5) Choosing only a specified number of peaks starting with the lowest in frequency.

Thus, in order to calculate the peak vector (a list of points in the frequency domain where the peaks happen) using this procedure, five different parameters need to be entered manually. This becomes impractical if the sounds we are trying to analyze are not as straightforward as the single string of the violin, and especially if the goal is to be able to analyze an entire piece continuously. We simply cannot go through the entire piece manually choosing different parameters every time the instrumentation or the register changes. What my algorithm does is automate this process using some musically-founded assumptions. These are some of the key features:

---

[1] The full routine in MATLAB code (*nsoctpeak*.m) can be found on the web at http://eceserv0.ece.wisc.edu/~sethares/banuelos/banuelos.html.

(1) Instead of taking the largest peak as a reference to filter noise, I first calculate the actual noise floor using a median filter as described above. The frequencies that get chosen as peaks in the second stage of the algorithm are those which lie more than 10dB *above* this line.

(2) Furthermore, the algorithm is sensitive to the overall intensity level of the passage.[1] When the intensity levels are high (e.g. full orchestra playing *ff*), the assumption is that the noise floor will be high as well, so the algorithm is more tolerant and allows for peaks that are only 10dB above the noise floor. It makes sense to assume that in the context of a loud orchestral texture we will consider "significant" the sound of a flute, for example, even if it is barely louder that the sound of a snare-drum playing at the same time. When the intensity is low however, most of the noise is going to be close to inaudible, so it is suitable to make the algorithm more intolerant and admit only peaks that lie 20dB above the noise floor. 20dB is approximately the difference between an empty auditorium and a *pp* sound. The function used for this purpose transitions smoothly between 10dB and 20dB according to the average power in the window. It is worth noting that the noise floor actually changes with register as can be appreciated in figure 2.4. This means that peaks will be chosen according to how "relevant" they are *within* their specific registral context.

(3) Rather than specifying particular starting and ending frequencies every time, I make the entire range of human hearing the domain. Thus peaks can be found anywhere between about 20 Hz and 20 KHz (octave registers 0-9). However, given the poor resolution of the FFT in the low registers (up to C_2), a peak there must be 30dB above the noise floor in

[1] The intensity scale (which in this case is simply the power of the signal) was drawn directly from the recording used in part III of this paper (Berg, 1992). Therefore, the scale goes from the softest to the loudest sound that can be found in this particular recording.

order to be selected. This is also consistent with the Fletcher-Munson loudness curves, which show that human hearing is less sensitive at very low (and to a lesser extent at very high) frequencies (Roederer 1995, 89).

(4) Instead of using a set frequency to be the minimum difference between peaks, I make the default minimum a quartertone, which once again is register-dependent. I chose the quartertone as the minimum difference somewhat arbitrarily, but not without musically motivated reasons: Anything smaller than a quartertone within an orchestral texture will blend as a single pitch and will only affect minimally the overall color of the sound—perhaps in the form of slow acoustical beats, that could potentially be measured in a different way. A half step, on the other hand seems too big a distance if we want to take under account upper partials and the potential role they may play in sensory roughness, which will be discussed at length on pp. 30-38. No set number of peaks is specified in my algorithm.

All that is needed to calculate the peak vector using this new algorithm is the original signal and the size of the sample. None of the parameters have to be entered manually, which is very convenient when analyzing large stretches of music. Once we have the peak vector and the noise floor we can easily create a new, simplified representation of the spectrum as shown in figure 2.5. The original FFT spectrum (once again the violin A-string) appears in (a). The beams show the locations of the selected peaks and the thick line represents the noise floor. In (b) we can see a reconstruction of the spectrum (called a *line spectrum*) using only the information in the peak vector and the noise vector. The creation of this abstract model allows us to reduce the information provided by the FFT from ten of thousands of undistinguished data points to (in this case) about 20 *labeled* objects; each peak can be seen

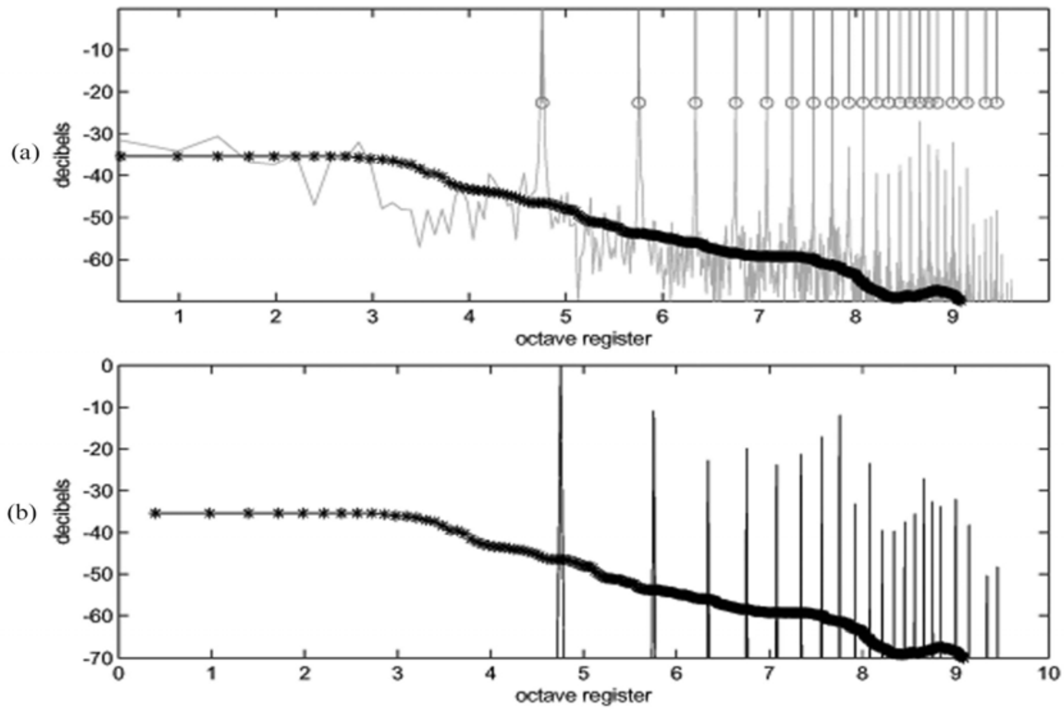


Figure 2.5. (a) Spectrum of a violin A-string where the starred line represents the noise floor and beams show the locations of selected peaks. (b) Line spectrum representation of the original spectrum.

as a separate object with a particular frequency and amplitude. Furthermore, each peak bears a specific relationship to the noise around it.[1]

It is important to keep in mind that most of the choices I made when creating this peak-picking algorithm were geared towards optimizing results in the specific context to which the procedure was intended to be applied (the Berg Violin Concerto). And even then, there are surely several instances where the results are not entirely satisfactory, that is, where the algorithm clearly chooses peaks that could not really be considered "the most significant elements." One must keep this in mind when interpreting results of the procedures outlined below. It is worth stressing that all these procedures are only meant to be used as *tools* to

[1] The relationship between noise and signal and how to plot it over time is discussed at length on pp. 44-47.

elucidate and further our already vast understanding of relationships in music, rather than to dictate definitive attributes of sound.

Plotting Single Attributes of Sound

Dimensions in Sound Color and the Purpose of Spectral Analysis

The simplified and labeled model of the Fourier spectrum can be used in a variety of ways in order to isolate aurally relevant attributes of sound color and observe them as they evolve through time in a musical context. What to isolate, then, becomes an important question. One of the main problems in understanding sound color—and even in defining it with precision—is the fact that, like its visual counterpart, it is *multidimensional*. In the case of timbre, Sethares (1998) remarks: "It is not possible to construct a single continuum in which all timbres can be simply ordered as it is done for loudness or for pitch. Timbre is thus a "multidimensional" attribute of sound, though exactly how many "dimensions" are required is a point of significant debate" (28).

A similar case can be made for harmony, even though in acoustics (and often in music theory) harmony is commonly thought-of *only* as the product of single pitches or melodic lines sounding simultaneously. When considered as an element of sound color—or of the overall "sonority" of sound, as Adorno puts it (see p. 2)—harmony becomes not just an incidental product of various melodic lines, but a living entity that morphs through time, "filling" it like color does a canvas. In this way, the *sonority* of harmony (as opposed to its functionality), which depends as much on pitch content as it does on voicing and instrumentation, creates a rich universe of shades and qualities as it moves through time in any given musical context. The trajectories and impressions created by this movement may or may not support the formal structure of a piece as dictated by its melodic content, and the degree to which they do varies greatly from composer to composer and from style to style. In certain contexts, however, this ebb and flow of sound colors or *Klangfarben* could reach levels of organization comparable to those found in tone melodies, yielding perhaps *Klangfarbenmelodien* (melodies of sound color) such as the ones envisioned by Schoenberg (Cramer 2002, 4).

The multiple qualities and colors of harmony cannot be simply put in a linear continuum. For instance, trying to arrange linearly the allegedly finite number of simultaneous pitch combinations that can be produced with the chromatic scale within the hearing range would seem to me less than musically relevant. This is because the nature of harmony—and of sound color in general—is intrinsically *qualitative*, and depends as much on absolute sonic properties as it does on context—and here I am including cultural context as well as musical context. For this reason, rather than trying to "catalogue" sounds based on their sonic attributes, I strive to use the spectral representation to better *describe* and *understand* interactions between otherwise aurally recognizable elements of real music. And rather than trying to look for these elements *in the spectrum*, my approach is to isolate and observe the larger attributes *of this representation* and explore how they can be used as a tool in understanding and tracing the development of various musically relevant aspects of sound color within a closed environment.

The most immediate and general characterization of spectral features has already been suggested; that is the distinction between "harmonic" versus "timbral" attributes. In the most general sense we can think of harmonic traits as depending on the disposition of peaks along the frequency axis, whereas timbral traits depend more on the size of these peaks relative to each other. In truth, however, the relationship between the two parameters is a bit more complicated, as will become more apparent in the course of this paper. For instance, we can define a certain harmony (chord) as having a particular pitch content; if we keep this pitch content intact but vary the instrumentation, allegedly the only thing that has changed is timbre. In terms of the spectrum of the sounds however, the disposition of peaks may not be exactly the same. It is entirely possible for one of the versions of the chord to have considerably fewer peaks than the other, for instance, and whereas the first dozen or so peaks may actually be in the same place in both chords, the upper-partial content of each may be rather different. Furthermore, it is entirely possible to perceive two chords with the exact same pitch content *and* instrumentation as considerably different, depending on how they are balanced, meaning how much emphasis is given to any particular voice. Here the only thing that changes is the size of peaks and yet we would say that timbre has stayed constant. While some times it is useful to make the distinction between harmony and timbre clear, it is important to always take it *cum grano salis* and to keep in mind the influence they will have on one another, given that both are part of the same basic phenomenon (whether it is called sound color, sonority, or sonic structure).

The five spectral attributes I have chosen to isolate (based on their musical relevance) can be described as sensory roughness (SR), registral brightness (RB), timbral brightness (TB), peak variance (PV), and noise-to-signal ratio (NSR). SR can be seen as the most "harmonic" of the five, while PV and NSR would qualify as the most "timbral;" RB and TB lie somewhere in between as they can be affected both by the disposition of peaks and by their size. The remainder of this chapter will be devoted to defining and describing these five spectral characteristics and the algorithms used to isolate them.[1]

Sensory Roughness (SR)

The idea of isolating a single aspect of sound and plotting it as a curve that changes over time has its origin in Sethares's (1998) dissonance curves, as noted earlier. His approach enables us to derive and plot "sensory dissonance" directly from the Fourier spectrum, generalizing to all sounds a theory of dissonance that since Helmholz's time had remained in the abstract realm of single and isolated intervals (Sethares 1998, 81).

Whatever the implications of the word "dissonance" in any particular circumstance, Sethares's graphs express a material and, furthermore, an aurally relevant attribute of sound that can be observed and described in any context, regardless of the interpretation or the aesthetic evaluation it is given. In other words, a sound does not have to be "dissonant" (whatever that may mean in the wake of the 21st century) in order to display a certain degree of "sensory dissonance." In fact, the only sound that would give a reading of "zero dissonance" using Sethares's method would be a single sine wave. Between the single sine wave and a saturation of frequencies so dense the ear can only perceive it as white noise lies the continuum of "sensory dissonance," which I prefer to call *sensory roughness*. This acoustic property, which can be sensed to various degrees in *any* sound, could be seen as a sophisticated measure of peak density. In other words, SR will be large when we have many peaks all close to each other, and small when we have fewer peaks further apart.

---

[1] The actual algorithms in MATLAB code can all be found on the web at http://eceserv0.ece.wisc.edu/~sethares/banuelos/banuelos.html.

But this is not all; the SR also tells us something about how these peaks are interacting with one another. Every time two sine waves are played simultaneously with different frequencies, they create *acoustic beats* at a rate equal to the difference between them.[1] If the two frequencies are really close to each other, the ear will not perceive them as separate pitches but as a single sound with slow fluctuations in volume (slow beats). As the frequencies draw apart, the beats get faster until "the beat sensation disappears giving way to a quite characteristic *roughness* [*sic*] or unpleasantness of the resulting tone sensation" (Roederer 1995, 32). Eventually the "roughness" begins to recede at the same time as the ear begins to differentiate the two waves into separate pitches, but it never disappears completely. This can be appreciated in audio track 2 which plays two sine waves, one of which is held at 220 Hz while the other sweeps slowly from 220 Hz to 470 Hz.

This phenomenon was tested extensively by Plomp and Levelt (1965), who asked a significant number of musically untrained volunteers to rate the "consonance" of pairs of sine waves (which, when it was needed, they defined as "euphoniousness" or "pleasantness"). The results of these experiments yielded a series of curves that "have since become widely accepted as describing the response of the auditory system to pairs of sine waves" (Sethares 1998, 45). Naturally, the responses in Plomp and Levelt's experiment varied considerably from subject to subject in terms of the actual numerical rating (limited to a scale from 1 to 7). But the overall shape of the curves, which can be seen succinctly in figure 2.6, was quite consistent with the *roughness* pattern described above. Sethares (1998) condensed Plomp and Levelt's results into a single equation[2] that, given a fluctuating frequency over any sustained

[1] For a comprehensive psychoacoustical description of this phenomenon see Roederer (1995, 28-36).
[2] A complete mathematical explanation of this equation and its application to the "sensory dissonance" algorithm can be found in (Sethares 1998, 299-302)

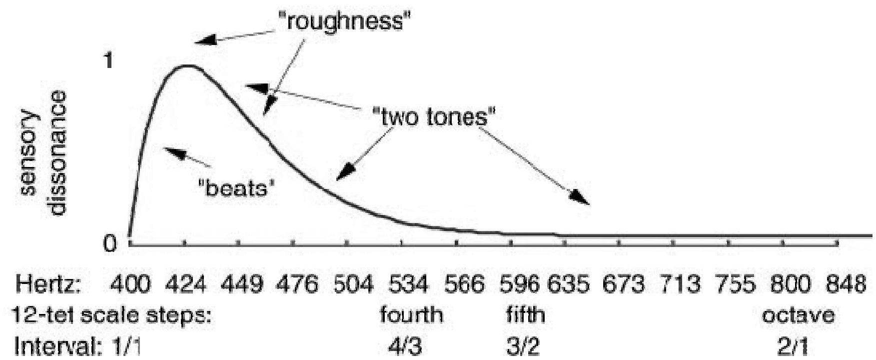


Figure 2.6. When two sine waves are played simultaneously, aural perceptions include pleasant beating (at small frequency ratios), roughness (at middle ratios), and separation into two tones (at first with roughness, and later without) for larger ratios. The horizontal axis represents the frequency interval between the two sine waves, and the vertical axis is a normalized measure of "sensory dissonance." The frequency of the lower sine wave is 400 Hz. Reproduced from Sethares (1998, 44).

bass, reproduces the essential shape of the corresponding Plomp-Levelt curve. Figure 2.7, shows a few examples of Sethares's curves, from which a number of interesting things can be observed:

(1) The curves do not display any distinction between intervals that are normally considered "dissonant" and intervals that are not. (2) The location of the apex (maximum roughness) varies depending on the lower frequency. (3) The apex seems to always occur somewhere between the unison and a major third.

It is indeed remarkable that the curves do not display any distinction between intervals. This is because the notion of dissonance, as it is usually understood in music, is a phenomenon that arises from the superposition of *complex* tones and the multiple interactions between their respective partials. What we see here is what Roederer calls "basic dissonance," and it applies only to pure sine tones, as opposed to "complex dissonance" which applies to richer sounds (1995, 168).

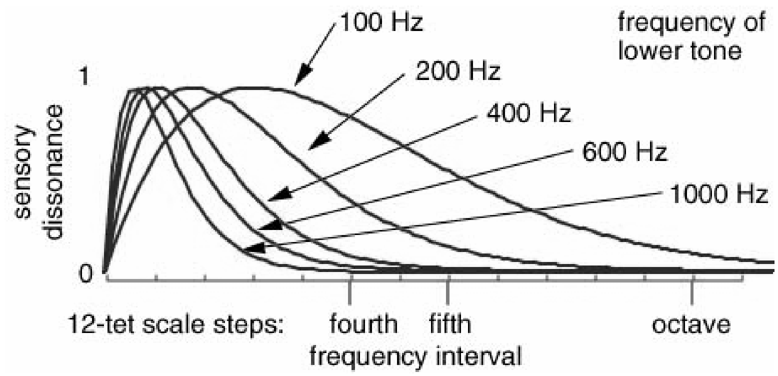


Figure 2.7. Various "sensory dissonance" curves for pairs of sine waves. The horizontal axis represents the frequency interval between the two sine waves, and the vertical axis is a normalized measure of "sensory dissonance." The plot shows how the sensory consonance and dissonance change depending on the frequency of the lower tone. Reproduced from Sethares (1998, 44).

The specific locations of roughness maxima are not fortuitous either. Plomp and Levelt showed that there is an intimate relationship between this distance and what is called the *critical bandwidth* of the ear. Two frequencies are said to be in the same critical band when they are close enough to each other that their responses along the basilar membrane overlap[1]. The basilar membrane is the natural frequency analyzer of the auditory system. It stretches along the snail-shaped cochlea and its elasticity varies along the way, making it responsive to different frequencies at different places. For a comprehensive description of the inner ear see Roederer (1995, 22-24). In general, the point of maximal roughness occurs consistently at about ¼ of the critical bandwidth (Sethares 1998, 86). However, the bandwidth itself varies according to register, which accounts for the registral dependency of the curves.

After investigating these responses to "basic dissonance" Plomp and Levelt went on to suggest that the dissonance of more complex tones could be calculated by adding up the

[1] The basilar membrane is the natural frequency analyzer of the auditory system. It stretches along the snail-shaped cochlea and its elasticity varies along the way, making it responsive to different frequencies at different places. For a comprehensive description of the inner ear see Roederer (1995, 22-24).

dissonances between all pairs of partials. This is exactly what Sethares's algorithm does, automatically, and in a way that can be applied continuously to any electronically recorded sound—provided the peak-picking algorithm can be applied successfully. Sethares's "dissonance curve" algorithm can be summarized in the following steps:

(1) The sound file is partitioned into a number of equal-sized windows (about 0.2 seconds in length).

(2) For each window it takes the FFT and calculates the peak vector as described on page 25. The peak vector and a vector of peak amplitudes are then fed to a new function called *dissmeasure.*[1]

(3a) The *dissmeasure* function goes through every possible pair of peaks and calculates its "basic dissonance" in the manner of Plomp and Levelt, using the above mentioned equation. The "curve," so to speak, corresponding to each pair of frequencies is scaled in such a way that takes amplitude into account, giving more weight to the roughness between louder pairs. This makes sense, even if the "basic dissonance" for a louder pair was the same as in a softer pair; clearly, the roughness in the louder partials will be more prominent in the composite sound. Within each pair, however, the amplitude of the smallest peak is the only one considered. This means that each pair is dissonant only inasmuch as the smaller peak is dissonant to the larger one; the bigger the amplitude difference between peaks, the smaller the dissonance will be.

(3b) All the pairs are then summed up to give a *single number* that represents the "sensory dissonance" in the window.

[1] The *dissmeasure* function, which is also used in my SR calculation, can be found at the above mentioned website.

(4) Single “sensory dissonance” readings for each window are plotted against time illustrating how it changes throughout the entire file.

A few things seem problematic in this procedure, not least of which is the peak-picking algorithm (my solution to this problem is discussed on pp. 24-25). Another problem, I believe, is the scaling of the peak-size vector. On the one hand it seems reasonable that louder pairs of frequencies within the same sound receive more weight. However, it does not seem appropriate for an entire window to receive a higher dissonance reading just because it is louder than another. Unfortunately the *dissmeasure* function considers the raw amplitude of each peak, and so it is entirely possible to obtain a drastically different reading for two intervals or chords that are exactly the same except in volume. Furthermore, the fact that the “size” of the peaks within each window is taken by measuring *amplitude* on a linear scale is problematic in itself, given that what the auditory system perceives is *intensity* level on a more or less logarithmic (decibel) scale.

What I have done in my SR algorithm is transform all the peak sizes from amplitudes to decibel levels. This eliminates any trace of absoluteness there might have been, making all the readings dimensionless and self-referent; every peak is measured only with respect to other peaks within the same spectrum (as the decibel scale does not express units of *quantity*, but *ratios* between quantities). At the same time, this is more akin to what the ear does, as we can identify specific timbres (or harmonies) regardless of their intensity. What we perceive as a certain color then, is the relationship between spectral elements—or to quote Roederer, “the activity distribution evoked along the basilar membrane” (1995, 150)—regardless of their absolute size. Furthermore, in the SR routine, each window is scaled in the same way, as if every sound had the exact same intensity level. The range is always 70dB, which is about the difference between an empty auditorium and the loudest a symphony orchestra can play. Thus the largest peak in each sample always gets a 70dB level and all other peaks are expressed as fractions of it. This information is what gets fed to the *dissmeasure* function as the "peak size" vector on step (3a) of the algorithm.

The last problem I detected is the size of the windows. We have already seen that the smaller the window the poorer the resolution of the FFT spectrum. In a situation in which we are measuring minute differences between adjacent peaks we can not afford to have such errors as the ones we encounter when dealing with windows of this size. Plus, the peak-picking algorithm is less likely to be successful when the resolution is poor, especially in the lower register. Unfortunately we cannot simply increase the size of the windows arbitrarily. One of the properties of the FFT is that when it analyzes a sound that changes within the timeframe of a window, the spectrum will display *all* the frequencies that appear in that timeframe, regardless of whether they were played simultaneously or in succession. This is called the *averaging* property of the FFT (Sethares 1998, 293). This means that if we are analyzing an actual piece of music and our windows are larger than the smallest unit of time, the FFT might turn adjacent notes into "chords." Furthermore, if the partitioning is arbitrary, this may happen even when the windows are the size of the smallest note values or smaller, but are not aligned with them.

My solution to this problem was to manually partition the wave files I analyzed into *musical beats*. This allowed for much larger windows that in turn provided better resolution. Needless to say though, rarely do these windows represent a "single sound." The division makes sense musically, however, and seems a natural and uniform choice for a "single moment." Rather than *measuring* the dissonance of *a sound*, my solution provides an

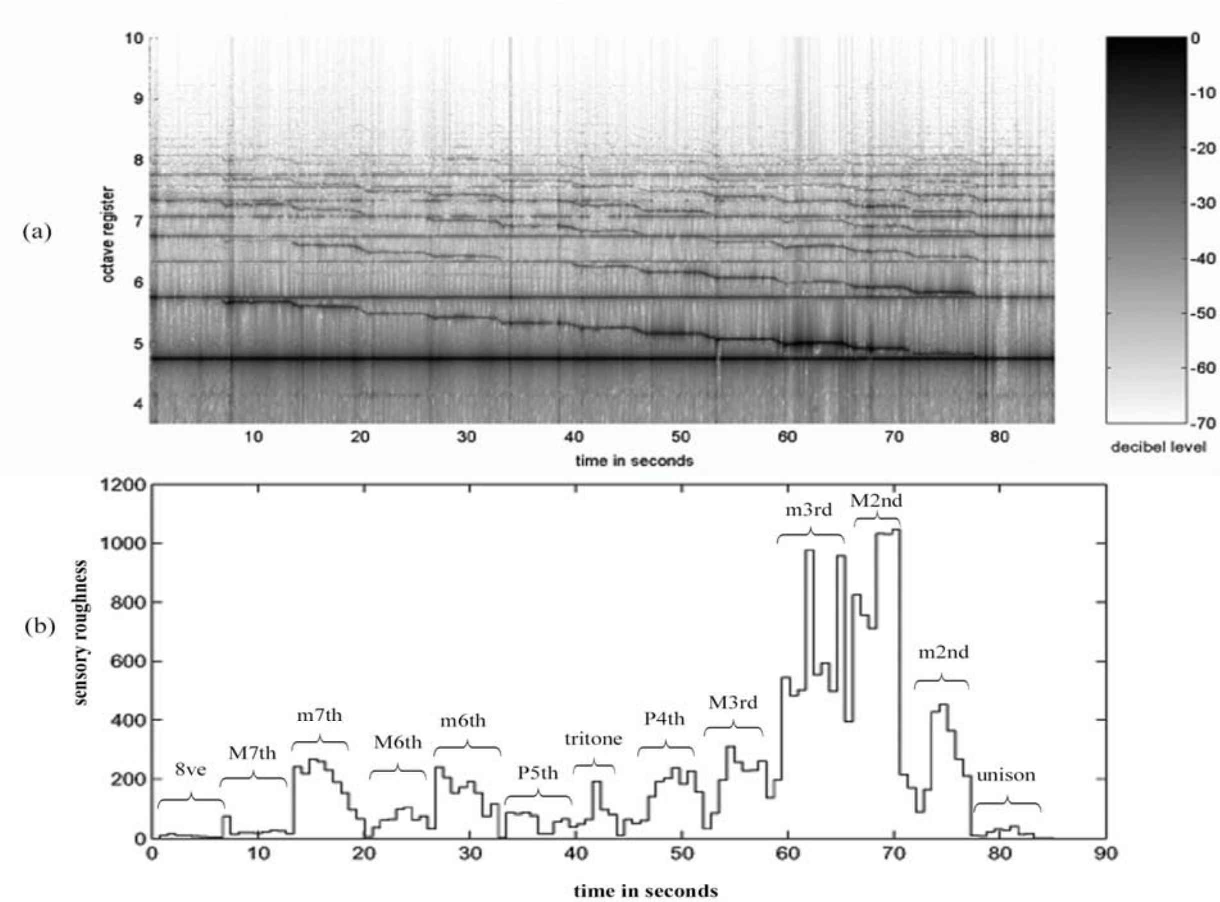


Figure 2.8. A violin playing double stops slides from an octave to a unison over a sustained A_4. (a) Shows a spectrogram of the file while (b) displays its SR as it changes over time. Brackets indicate the interval between the two tones at any particular time.

assessment of how much *spectral activity* there is in a particular beat, be it harmonic or melodic, and how all its main frequencies interact with one another. The fact that the SR routine cannot tell the difference between harmonic and melodic elements does not really matter because that is something *we* can do by looking at the score or listening to the sample. The SR algorithm then, summarizes the overall activity in a beat; much like a roman numeral would summarize its basic harmonic makeup regardless of whether the notes of the chord are all played simultaneously or not.

The precise nature of this activity will become more apparent in Part III of this paper, when we see the application of the SR algorithm in an actual musical context. Meanwhile, let us observe its results in more controlled conditions. Figure 2.8(a) (audio track 3) shows a violin playing double stops. While one voice sustains A4, the other slides chromatically from A_5 to A_4, thus sampling every interval between the octave and the unison. (b) Shows an SR reading of the file using analysis windows of about a second. Unlike in the "basic dissonance" curves of Plomp and Levelt, here it is easy to see differentiations between the various intervals—despite the natural fluctuations resulting from live performance. When comparing these SR levels with standard notions of dissonance for familiar intervals it is important to keep in mind that the curve is specific to this particular timbre and even to this particular performance. Therefore, the curve does not show the SR levels of intervals in general, but of the intervals played in this particular example.

Similarly, it is possible to compare the results of the SR algorithm for the somewhat more complex structure of single chords. The following table displays a few chord qualities and voicings, and respective single SR levels obtained over a window of one second. All these chords have A_3 as their bass and are played on a pipe organ. The corresponding sound files can be found in audio tracks 4-15.

Table 1. Chord types and their SR levels

Chord type	SR level	Chord type	SR level
Major triad (open position)	239.3	Wholly diminished	534.5
Major triad (closed position)	395.7	Quartal tetrachord	386.9
Minor triad (closed position)	424.3	Pentatonic cluster	770.4
Dominant seventh	616.5	Diatonic cluster	1382
Dominant seventh (fourth inversion)	626.3	Whole-tone cluster	1318
Minor-minor seventh	581.3	Twelve-tone cluster	3114

Registral and Timbral Brightness (RB and TB)

The concept of isolating particular inter-spectral relationships of perceptual relevance and plotting them through time can be advantageously applied to more timbral characteristics of sound. In a 1978 study, Gordon and Grey suggested that the spectral energy distribution of a signal—which they claim serves as the sole basis for interpreting perceptual relationships between various timbres in steady-state state tones (1978, 1494)—could be reduced to a single number. By creating a “peak” representation not unlike the one proposed above and calculating its *centroid* or “balance point” they were able to create a model of spectral energy distribution which adequately took into consideration “the many factors which may be important: overall bandwidth, balance of levels in the lower harmonics and the existence of strong upper formants” (1498). They also showed that this model could be directly correlated to *perceived* similarity relationships between instrumental tones, which they thoroughly examined in the same study.

Both RB and TB algorithms calculate the centroid of a peak spectrum, in the manner of Gordon and Grey, in order to reduce it to a single-number representation. The centroid can be described in the following way. Suppose we had a massless rod spanning the length of the frequency hearing range. Each peak in a given spectrum can be seen as a mass point along that rod, with mass equal to the height (in this case in dB) of that peak. The centroid is the place where, if suspended, the rod would “balance.” Figure 2.9 shows the peak spectrum of two violin sounds (using approximately 1 second windows); the first one is an open A-string while the second is a *muted* A-string. The centroid in both cases is shown with a triangle. Both sounds can be heard in audio track 16.

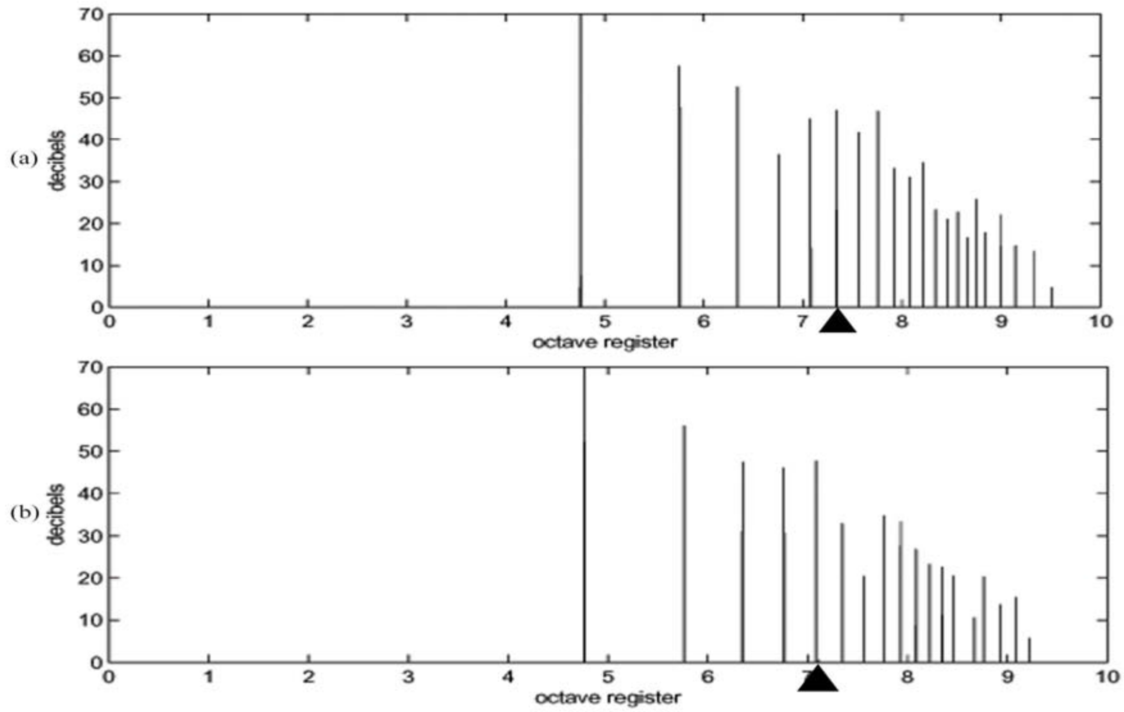


Figure 2.9. Peak spectra of two sounds and their respective centroids (indicated by triangles). In both cases the sound represented is a violin playing A_4. In (a) the A-string is played *ordinario*, while in (b) the same note is played on the D-string with mute. This change in timbre is reflected in the shifting of the centroid to the left.

The RB routine can be summarized as follows: (1) Parse the sound file into windows of specified length (in applicable cases musical beats). (2) Calculate the FFT of each window and reduce it to a peak vector. (3) From the peak vector create a peak spectrum in the manner of figure 2.8 and calculate its centroid. (4) *Take the distance from the centroid to the highest frequency in the range (C_{10}) and divide it by the distance from the centroid to the lowest frequency (C_0)*—distance is measured in octave registers. This gives a sense of how much energy there is in the upper registers compared to that in the lower registers. It also allows for a dimensionless value (once again a ratio). (5) Take the results from step (4) and string them together to represent the transformations of RB through time.

I would not claim that the results of this algorithm fully describe the energy distribution of spectra, nor that they map perfectly onto the perceptual basis for timbre discrimination of steady state tones. However, they do tell us something about a phenomenon that among musicians is generally referred to as "brightness" (hence my choice of name). It is a known fact in phonetics that even single sine waves are perceived as having different tone colors according to register. As Cogan (1998) describes it: "In the lowest registers a sine tone suggests the dull grave sounds [u] (oo) or [o] (oh); [and] in the higher registers its quality resembles [a] (ah) or the bright, acute sound [i] (ee)" (12). Hence the presence of higher partials in a complex sound will invariably make it sound "brighter" or more "acute," whereas their absence will create a "duller," "darker" effect. Even the same exact energy spectrum distribution will change timbre when transposed in register, as playing a tape at double speed would illustrate.

One can easily verify the results of the RB procedure in the example shown above (audio tracks 16 and 17). Similarly, I asked a soprano to sing the same pitch (A4) in two different ways, namely with a bright tone and with a dark tone (audio track 17). When analyzed in isolation over a window of about three seconds, the four sounds display the following RB values[1]:

Table 2. Examples of RB

Sound source	RB level
Unmuted violin A_4	2.75
Muted violin A_4	2.28
Soprano with bright tone	2.30
Soprano with dark tone	1.75

[1] An example of what an actual RB curve looks like in time can be found on p. 55.

The TB algorithm is quite similar to the RB one. The only difference can be found in step (4). Rather than measuring the distance from the centroid to the extremes of the register I consider only the bandwidth of the particular spectrum, that is, the distance from the lowest peak to the highest peak. In step (4) then, the algorithm takes the distance from the centroid to the highest peak and divides it by the distance from the centroid to the lowest peak. In a way this ignores the information pertaining to register and considers only the energy distribution within the particular bandwidth. This reading does not map onto our perception of brightness or acuteness as well as RB does. However, it can be useful to make the distinction between the kind of brightness that is achieved mainly through the overall register, and the kind that is created by emphasizing higher components of the spectrum, especially in complex musical contexts involving several independent voices. TB then, can isolate aspects of voicing and instrumental balance that may get lost in the all-encompassing RB reading. For this reason, both curves are meant to be considered together. Applications of TB and its relationship to RB can be found in Part III, particularly on p. 73.

Peak Variance (PV)

Consider the two sounds represented in Figure 2.10, which can be heard in audio track 18. Once again, the source of both sounds is a violin A-string, played at two different contact points with varying bow speed. At first glance the two spectra may seem very similar. Both contain a similar number of peaks and even the overall energy distribution is quite similar. Yet, it is not hard to recognize aurally that there is a marked timbral difference between the two sounds. How this difference reflects on the spectra becomes more apparent upon closer examination. Whereas in (a) the peak heights decrease more or less smoothly as

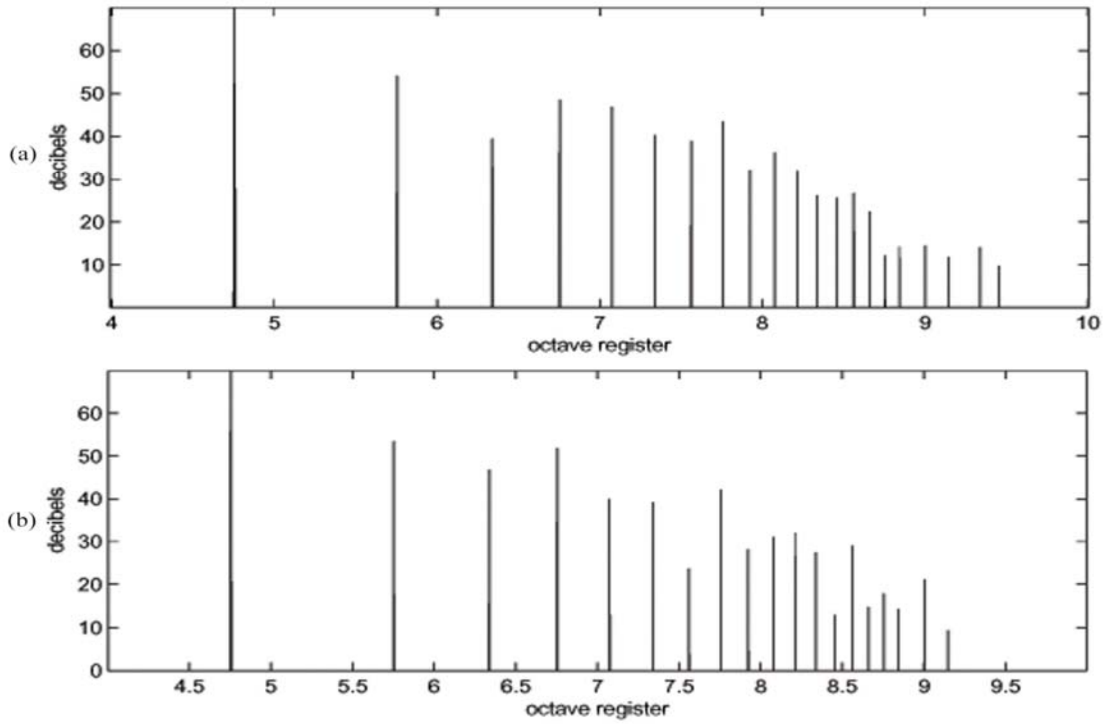


Figure 2.10. Peak spectra of two different timbres on a violin A-string. In (a) the string is bowed close to the bridge with slow bow, while in (b) it is bowed close to the fingerboard with fast bow.

we move up in register, in (b) the picture is more uneven; a number of peaks seem underemphasized and as a consequence there are a number of others that "stick out." In other words, the *envelope* of the spectrum fluctuates much more in (b). As a result, the timbre seems more saturated and "poignant" in (a), whereas in (b) it seems more "cottony" and "round."

The program I have written to analyze and trace this property is called simply peak variance and can be summarized in the following steps: (1) Parse the sound file into windows (possibly musical beats). (2) Calculate the FFT of each window and generate a peak vector. (3) Create a second vector of ordered peak magnitudes. (4) *Calculate the difference in magnitude between every pair of adjacent peaks and take the average value.* This single number models how much variance there is overall between adjacent peaks. (5) Take the results from step (4) for every window and string them together to represent the transformations of PV through time.

The following chart displays a few single PV readings for familiar sounds analyzed over a window of approximately three seconds. The first two are the violin sounds displayed in figure 2.10 (audio track 18). The next pair is once more a soprano singing A_4 with two distinctly different tone colors (audio track 19). This time she described the two sounds herself as "sweet" and "ugly" respectively. Lastly I consider a clarinet and a trumpet, both playing the same note.

Table 3. Examples of PV

Sound source	PV level	Sound source	PV level
Violin A-string (a)	6.95	Violin A-string (b)	9.27
Soprano with "sweet" tone	11.9	Soprano with "ugly" tone	7.41
Clarinet A_4	21.7	Trumpet A_4	5.49

It is not hard to see from this simple example that the clarinet and the trumpet lie at the extremes of the PV continuum. The violin and the voice on the other hand fall more towards the middle. One could see the trumpet as the sharpest and edgiest of the two, while the clarinet is the most ethereal and delicate. Also noteworthy is how similar the voice and the violin are in their PV ranges. When I first wrote the algorithm I was unaware of this fact, but as it turns out, these particular relationships can be seen carefully orchestrated in the Berg Violin Concerto, to the point of being almost motivic, as will be shown throughout Part III.

Noise to Signal Ratio (NSR)

So far, all the algorithms discussed have considered only the relationships between peaks of the spectrum. And since all the instruments we have studied have been pitched, this means that disproportionate attention has been placed on harmonic and quasi-harmonic elements, that is, those elements that the auditory system tends to clump together into definite pitches. But, what about inharmonic, non-pitched elements? It has already been suggested that we can not simply throw this information away as it is doubtlessly relevant to our perception of timbre.

The NSR algorithm provides a simple model that can be used to show the relationship between pitched and non-pitched elements of the spectrum, or perhaps more accurately, between single, recognizable peaks and masses of nonperiodic sound. The algorithm consists of the following steps: (1) Parse the sound file into windows. (2) Calculate the FFT of each window, generate a peak vector, and calculate the noise floor using a median filter as discussed on p. 21. (3) Calculate the "height" (amplitude) at which each peak is intersected by the noise floor and take the sum of all these values. (4) Take the sum of all the peak heights. (5) *Divide the result of (3) by the result of (4).* This number evaluates the relationship between the energy in the peaks and the energy of the noise floor, at the peaks. In other words, this is a kind of average of *how high the noise floor is in relation to the peaks*. (6) Take the NSR results for each window and string them together to represent its transformation through time.

Figure 2.11 shows yet two more timbres in a violin A-string (audio tracks 20 and 21). This time (a) is played normally while (b) is played *sul ponticello*. It is easy to see that the noise floor is higher in the second case. A NSR reading of both sounds (analyzed over a period of about three seconds) gives 0.0076 and 0.0248 respectively. For these particular sounds the NSR reading tell us all we really need to know about noise floor, namely that it is more prominent and hearable in the second example. We don't need to characterize this noise

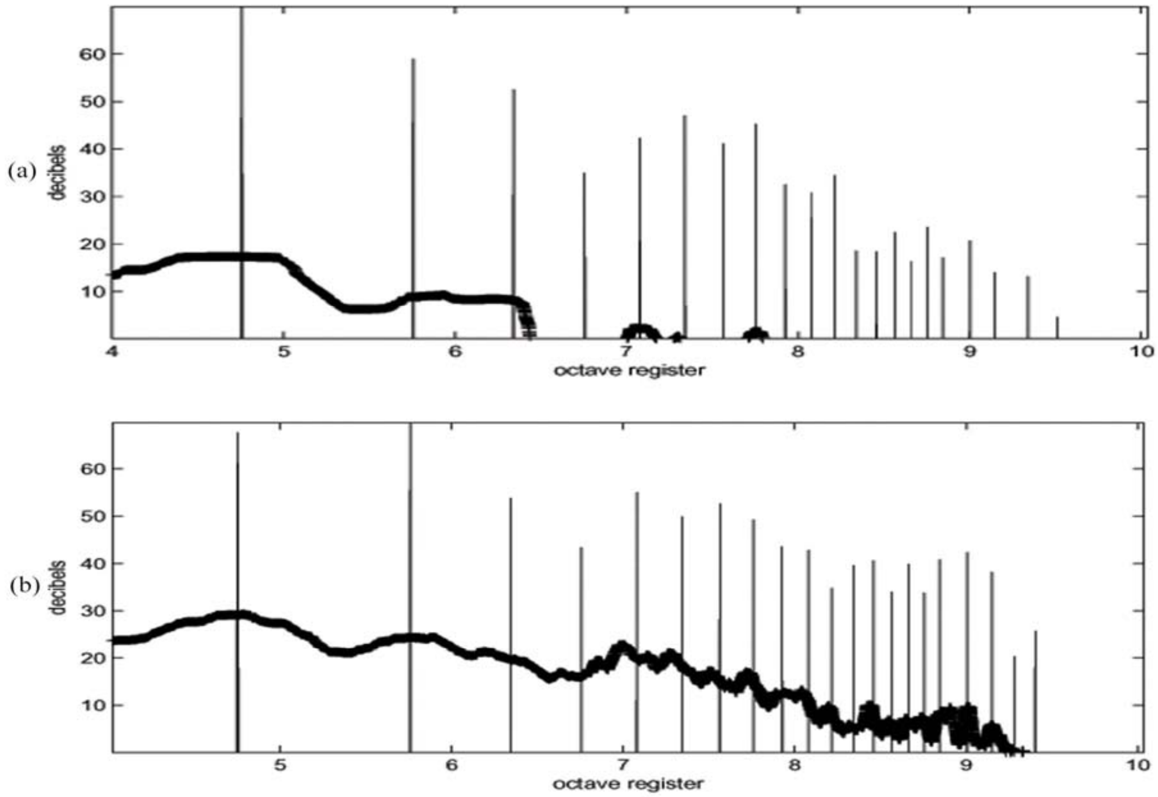


Figure 2.11. Line spectra of two different timbres on a violin A-string. In (a) the string is bowed *ordinario*, while in (b) it is bowed *sul ponticello*. One of the most prominent differences between the two spectra is the placement of the noise floor, which is considerably higher in (b).

neither registrally nor temporally, in order to convey the idea that (b) is "scratchier" than (a). There might be instances however, where a more precise characterization of "noise" might be needed. In some large orchestral textures, for example, the main sources of noise are actual percussion instruments and the nature of their sound is substantially more complex, both temporally (different kinds of attacks, rolls, etc) and timbrally (e.g. bass noise vs. treble noise). All the NSR curve can do in these cases is alert us about the presence of noise. The actual nature of this noise though, can then be better understood by looking at the raw spectrogram. There, we can see the energy distribution of the sound in question and also observe how it changes over time. Examples of this can be found in Part III, pp. 74 and 77.

Summary

The Fourier spectrum and the spectrogram allow us to create visual representations of sound and to recreate electronically the kind of frequency analysis that takes place in the human auditory system. Using the data in Fourier spectra of finite sounds, it is possible to derive simpler models that in turn can be further analyzed and used to explore the relationships between elements of sound in musical contexts. Such models can be created with the aid of a peak-picking algorithm. In an effort to investigate how the spectrum and its representations can be used to further our understanding of music, I propose five algorithms that isolate aurally relevant relationships between elements of the spectrum and plot them through time. The first one, which I call sensory roughness (SR), is closely derived from Sethares's (1998) "dissonance curves," and it plots the level of interaction between multiple partials of sound through time, or in other words, the amount of spectral "activity", as might be perceived by the human ear.

Registral brightness (RB) and timbral brightness (TB) assess the balance between higher and lower partials, which maps loosely onto the perception of how bright a sound is. While these parameters can be affected both by harmony and by timbre, peak variance (PV) reflects mostly timbral qualities. PV measures the fluctuations in the height of adjacent peaks, which relates to the perception of how poignant or saturated a sound is. Lastly noise-to-signal ratio (NSR) provides an assessment of how much noise there is comparison to the perceived pitch content.

PART III

SPECTRUM-BASED ANALYSIS OF ALBAN BERG'S VIOLIN CONCERTO (SECOND MOVEMENT)

In a 1958 review, John Hind wrote the following about a broadcast of Tibor Varga and the BBC Orchestra performing the Berg Violin concerto:

> During the broadcasts form the Festival Hall on 4 February . . . there was evidence of a large audience being held spellbound by one of the arch-dodecaphonists of the century. As Berg's Violin Concerto drew to its ethereal close, we had that rare thing, a momentary hush followed by tumultuous applause: an indication, more often than not, that the audience has been not only impressed by the significance of the subject matter (and its manner of presentation), but also deeply moved; moved, I think, by an almost embarrassing glimpse into the composer's spiritual self, and a part lifting [*sic*] of the veils which normally screen from us the opaque worlds of Truth and Beauty. Such experiences are rare, personal, and their significance not easily conveyed from one listener to another, least of all in analytical terms. Even so, just to describe this work as one of the masterpieces of our age is to beg further questions. By its very revolutionary and artificial nature, the atonal or twelve-note system of composition seems to preclude all possibilities of meeting us on any of our familiar grounds; and certainly we neither expect nor often receive, any striking emotional impact from some of the mathematically devised tone-rows of Schoenberg and his disciples. But the whole *raison d'être* of Berg's Concerto is in feeling and emotional response, as it follows the peregrinations of a human soul on earth. (1958, 207)

Opinions such as this one are not rare; another critic wrote in 1937 about the work's first performance in England:

> The astonishing thing is that, though the method itself is confessedly cerebral, when Berg applies it it becomes an eloquent vehicle of emotion. For the Concerto is beyond contradiction an expression of whole-hearted and sincere emotion, lyrical to a degree that many may have thought scarcely compatible with such esoteric calculations. (Review of London Concert 1937, 67)

It is apparent that even though trained listeners like these are aware of the highly sophisticated techniques involved in the craftsmanship of the work, they are more concerned with what they perceive as its "emotional" content, which some how for them resides elsewhere. Independently of whether there is or not a connection between the systematic pitch organization of the work and its expressive power, most listeners would agree (especially now, seven decades after its composition) that the piece is indeed quite powerful. However, the question of just what about this music makes it so, is more problematic, as Hind points out in the excerpt quoted above.

Anthony Pople suggests that what lends the concerto its communicative power is "recognition, rather than structuration" (1991, 88). He describes Berg's mature style as an amalgam of ideologies, some borrowed from the past, and some belonging to the most revolutionary musical tendencies of his time. Pople thus compares Bergs late works to a mosaic or a pointillist painting (7). According to Pople, in such works, and in the Concerto in particular, Berg was able to create the illusion of a true synthesis of old and new by distilling "tiny elements of style and technique from music of all kinds . . . to reconstitute them in a way which melded them together" (13). On the one hand, there is a plethora of tonal references in the work; in fact, the possibility of tonal reference is built right into the row that serves as the basis for the piece. But in spite of their structural importance and the occasional brief moments where they could be considered hearable and actually recognizable as such, these references to tonality lack the kind of continuity that would be necessary in order to be able to understand the work as a whole as "tonal." On the other hand, the piece is admittedly constructed through twelve-tone techniques. A more continuous account of how the Concerto is put together can indeed be achieved by looking at it from a serial point of view. But even this approach will only account for part of the events in the Concerto. There are many instances—especially in the second movement—where twelve-tone technique is abandoned altogether, such as the four-part canon (ms. 78-89), all the chorale quotations, and the beginning of the coda (m. 214), to name a few.

Independently of just how thoroughly serial the Concerto is, more often than not the presence of the row is so stealthily handled throughout, that most listeners (musically trained or not) will agree it does not play a truly crucial role in the work's expressive effect. Pople himself writes that the "general imperceptibility of the series denies it any central interpretative function" (88). This fact supports his original claim that the reason for the Concerto's coherence and communicative power does not lie in any kind of consistent system, but rather that it "makes sense" because at every point *something* in it is always recognizable through an active cognitive framework" (89). Personally, however, I wonder whether this recognition is necessarily *cognitive* at all times. In fact, I would claim that select places that lend themselves more readily to cognitive recognition are set in relieve by the very fact that they *are* recognizable (the found objects in this piece are good examples of this).

Searching the territory of sound color through spectral analysis may shed some light on those kinds of musical relationships that are not so easily or usually recognized cognitively, but are perceived at a more subconscious or intuitive level, advancing perhaps our understanding of what lends the Concerto some of its emotional impact. At the same time, a comprehensive discussion of sound color will provide further insight on the compositional coherence and interpretative function of sections of the concerto which may not be as easily or as thoroughly explained through other approaches like the ones mentioned above.

The sonic space of the Violin Concerto, in a way, provides complete unity for the work—if perhaps not systematically. The piece creates a world of sounds in which every moment bears some relation every other moment, by virtue of the very fact that they all belong to the same sonic micro cosmos. Continuing along the lines of Pople's "recognition" argument, we can say that one thing that is surely recognizable or at least hearable from moment to moment throughout the piece, whether cognitively or not, is the sonic relationship that every moment bears to those that came before. This is not to say that in every piece these recognitions necessarily amount to any kind of larger communicative impression. In the case of the Violin Concerto and many other pieces of the like, however, the overall sonic tissue is crafted in such a way as to articulate and reinforce the events carried out by more "cognitive" parameters such as thematic and motivic processes, and in many cases perhaps provides the kind of continuity that these parameters may lack. Furthermore, the Concerto, uses its sonic map, its "sound coloring", not only as a way of reinforcing, but what is more, as a way of *dramatizing* his musical narrative. It is no surprise that this is the case in the mature work of a composer who owed his success in great measure to his genius as an opera composer.

The fact that this piece is openly programmatic provides a very useful tool in verbalizing and rationalizing the sonic roadmap to the music, which is not to say that a similar kind of study could not be performed on a non-programmatic piece; it might be more challenging "making sense" of it, but it may be equally if not even more useful. More than anything, the openly programmatic aspects of the Concerto justify the spontaneous creation of a more meticulous and continuous narrative for the piece—an activity routinely employed by performers in general regardless of the repertoire, albeit in varying degrees of thoroughness.

Berg openly composed the piece as a requiem to Manon Gropius, the young and charismatic daughter of Alma Mahler, who died untimely of polio at age eighteen in April of 1935, when Berg was barely beginning to sketch his Violin Concerto. The piece thus bears the dedicatory “to the memory of an angel.” The published program for the piece, written by Berg’s pupil and friend Willi Reich and approved by the composer himself reads:

> Insofar as a transcription into words is possible at all, the tone—a favorite expression of Berg’s—of the whole work may be described as follows: delicate Andante melodies emerge from the rising and falling movement of the introduction. These crystallize into the Grazioso middle section and then dissolve back into the waves of the opening. The Allegretto Scherzo rises from the same background; this part captures the vision of the lovely girl in a graceful dance which alternates between a delicate and dreamy character and the rustic character of a folk tune. A wild orchestral cry introduces the second main part, which begins as a free and stormy cadenza. The demonic action moves irresistibly towards catastrophe, interrupted once—briefly—by a reserved point of rest. Groans and strident cries for help are heard in the orchestra, choked off by the suffocating rhythmic pressure of destruction. Finally, over a long pedal point—gradual collapse. At the moment of highest suspense and anxiety, the chorale enters, serious and solemn, in the solo violin. Like an organ, the woodwinds answer each verse with the original harmonization of the classical model. Ingenious variations follow, with the original Chorale melody always present as a *cantus firmus*, climbing ‘misterioso’ from the bass while the solo violin intones a ‘plaint’ [*Klagegesang*] that gradually struggles towards the light. The dirge grows continually in strength; the soloist, with a visible gesture, takes over the leadership of the whole body of violins and violas; gradually they all join in with his melody and rise to a mighty climax before separating back into their own parts. An indescribably melancholic reprise of the folk tune ‘as if in the distance (but much slower than the first time)’ reminds us once more of the lovely image of the girl; then the Chorale, with bitter harmonies, ends this sad farewell while the solo violin arches high over it with entry after entry of the plaint. (Cited in Pople 1991, 33)

It is also well known by now—thanks to the discovery by George Perle (1977) of an annotated score to the *Lyric Suite*—that in his late works Berg included autobiographical references in the form of secret ciphers and numbers. No secret score has been found for the Concerto, but as Pople (1991) points out, it is in deed possible to find a number of recurrent ciphers and autobiographical references in several passages. For instance, he shows that the

Carinthian folk tune that appears at the end of the first movement (m. 214) and later at the end of the second (m. 200) is a clear reference to an affair Berg had at the age of 17 with one of the servant-girls on his Carinthian estate, and which resulted in his fathering of an illegitimate daughter (33-34). Once again, rather than provide a script for how the piece should be interpreted, this information serves to ignite the imagination and create the possibility for not one, but an endless array of possible programs for the piece. Making some use of these programs, the following sections consider the use of sound color in a recorded performance of the second movement of the Concerto by Ann Sophie Mutter and the Chicago Symphony Orchestra, conducted by James Levine (Berg 1992). Appendix B (pp. 89-107) presents a continuous spectrogram of the entire piece.

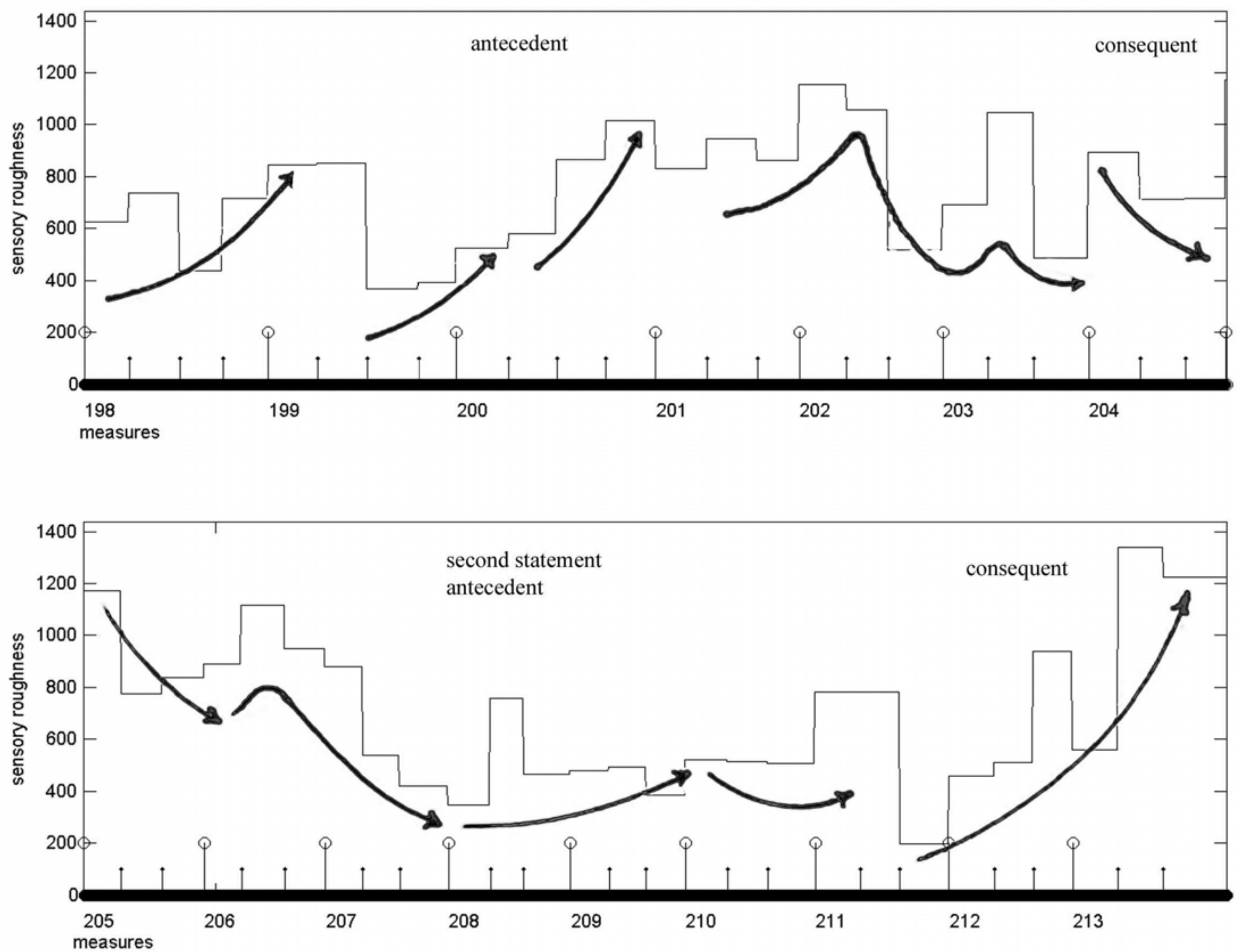
Micro-Structure

At the smallest levels—those of phrases and subphrases—the range of expression through color and texture is somewhat more limited than the kind of contrast that can be achieved from one large section to the next through drastically different orchestration. Nonetheless, even when the instrumentation remains constant throughout a passage, subtle yet significant differences in sound color occur in the harmonies and the inflections with which they are performed. In these cases, such fluctuations of sound color can generate varying degrees of spectral tension, which I will conceptualize in terms of sensory roughness (SR).

The return of the Carinthian folk song in ms. 200-213 is a good example of this, as the clarity of the texture and the straightforwardness of the phrasing allow the SR reading to be quite clean. The song itself as it appears the first time (in the first movement) is sixteen bars long and is divided into two antecedent/consequent-like periods that come together to form a double period. The song emerges right out of the last phrase of the chorale (CH_3), which is presented in inversion (i) by the cellos at m. 198. During ms. 198-201, the four-note figure gradually transforms into an ornamented version of what would have been the second measure in the original song. In this way, even though the beginning of the song is punctuated by the cymbal and the bass drum in m. 200, it is also somewhat obscured by the metamorphosis of CH_3i.

As we can see in figure 3.1, which shows a plot of SR for ms. 198-213, each statement of the cello's ascending figure starting in m.198 is accompanied by a small surge in SR that intensifies it and propels it forward as the melodic line ascends. Furthermore, all four statements together describe an ascending gesture that peaks on the downbeat of m. 202. At this point of maximum tension within the phrase, the cellos complete the transformation of CH_3i and arrive at the second two-measure subphrase as we first heard it in the first movement. After this arrival point, the SR drops quickly, as does the melodic line, and as we come to the realization that what we are hearing is actually a return of the folk song. The following measure is a kind of half cadence. The whole passage is permeated by tonal tendencies, but there are simultaneously two tonal centers scattered among the participant instruments: E^b major (the key of the song and its countermelodies) and A major (the key of the harp's chordal accompaniment). It is this conflict between the two tonal centers a tritone apart that creates the momentary surge in SR in the middle of m. 203, fleetingly intensifying the half-cadencial measure before the resolution that will close the antecedent phrase at m. 204.3. Whereas most of the gestures in the antecedent phrase pushed forward in intensity

Figure 3.1. Sensory roughness plot of ms. 198-213. The horizontal axis represents time, marked in musical measures and beats and the vertical axis represents the average SR level of each beat.



(ascending curves in the SR graph), gestures in the consequent (ms. 204-208) all descend, both melodically and in intensity. The 1+1+2 organization of the phrase is also supported by the SR continuity, which describes two descending motions, each one a measure long, followed by a larger arching gesture that reaching the lowest point in intensity thus far brings the phrase to a close in ms 208.

As we enter the second statement of the antecedent phrase beginning in m. 208, we perceive a marked change in texture/instrumentation. The violins pick up the folk song, with the soloist playing a countermelody mostly absent in the first statement, while the winds—except for the first clarinet—all drop out. The significance of this change and of this instrumental color in particular will only become apparent when we consider this passage in the context of the entire movement. For the time being, it is easy to hear that even though the harmonies remain fairly similar (we remain polytonally in E^b and A), SR levels are consistently lower in ms. 208-212 than they were in the previous phrase. It is also quite evident that both the composer and the performers go to great lengths to suspend the phrase and make it as floating as possible. This is reflected in the relatively smooth SR representation in these bars. Despite this stillness, we can see how the first two bars (ms. 208-209) again lean towards the third (m. 210), after which the intensity decreases in approaching the half cadence. The arrival of the cadence in m. 211 displays a sudden increase in SR, brought about by the clash of the dominant chords (B^b and an E) of our two simultaneous keys, and emphasized by the performers with a slight accent. In the last beat of this measure, a couple of instruments drop out and the remaining ones come down in dynamic level to near silence, leaving the soloist hovering almost alone before the arrival of the consequent. This marks the lowest point of SR in the entire passage, as shown in figure

3.1. The consequent phrase serves the dual purpose of bringing the song to an end and transitioning to the coda, which overlaps with it. The full woodwind section enters in the third measure, m. 214, to begin a new statement of the chorale. This leaves the cadence (both in E^b and in A) entirely to the soloist. As the trombone prepares the arrival of the new instrumental force throughout m. 213, the SR level also raises in anticipation of an overall harmonically rougher section.

So far we have looked only at the harmonic elements of color, and the perceptual measure that best isolates them, the sensory roughness. I would like to turn now to the other algorithms presented in Part II, which can provide significant insight in the sphere of timbre. The key to how timbre is used structurally in this piece can be found in the chorale (the only other quotation employed in the Concerto). The chorale melody "Es ist genug!" appears for the first time in its entirety at the beginning of the Adagio (m. 136). Interestingly, Berg includes the words in the score right below the melody, and in one of the most salient gestures of the piece, he assigns every other phrase to a four-part clarinet ensemble that adapts Bach's harmonization of the chorale as it appears in Cantata BWV 60 (1723). In this way, each melody is first presented by the soloist with a semi-serial accompaniment in the low strings and bassoon, and then answered by the organ-like clarinet choir playing Bach's version of the melody. The text by F.J. Burmeister, as it appears in the score (Berg 1996), is reproduced below. The labels in brackets denote the characteristic melodies used in each phrase.

Es ist genug! Herr, wenn es Dir gefällt, so spanne mich doch aus!	[CH_1]	It is enough! Lord, when it pleases Thee, Relieve me of my yoke!
Mein Jesus kommt: nun gute Nacht, o Welt! Ich fahr' in's Himmelshaus.	[CH_1]	My Jesus comes: So goodnight now, O world: I'm going to my Heavenly home.
Ich fahre sicher hin mit Frieden,	[CH_2]	I'll surely journey there in peace,
mein groβer Jammer bleibt darnieden.	[CH_2]	My great distress will stay below.
Es ist genug.	[CH_3]	It is enough.
Es ist genug.	[CH_3]	It is enough.

The three graphs below (figures 3.2, 3.3, and 3.4) represent sensory roughness (SR), peak variance (PV), and noise-to-signal ratio (NSR) respectively, for measures 136-157. A quick glance at these pictures may be enough to corroborate our aural intuition that at this point of "highest suspense and anxiety" (Willi Reich), Berg sets up a clear dichotomy between two very different elements. Whereas the first solo/serial statement in ms.136-141 emerges organically from the previous texture, the clarinet response in ms. 142-147 comes in as something curiously new. Not only is the difference in overall SR considerably lower, but there is a marked contrast in timbre as well. The organ-like sound of the clarinets is so clean (lowest NSR in the movement) and ethereal (highest PV) in comparison to its predecessor, that it sounds almost otherworldly—especially since the suddenly tonal harmony, is quite foreign to the mostly twelve-tone world we have established up to this point. "My Jesus comes," reads the text on the score as we hear the clarinet choir introduce this most pristine and otherworldly sound, thus far unprecedented in the piece.

But this new texture is not entirely disconnected from its context, neither harmonically nor timbrally. On closer examination of the PV graph for ms. 142-146, we can

Figure 3.2. Sensory roughness plot of ms. 136-157. All the SR values in each phrase are boxed together in order to show how the overall SR range changes form phrase to phrase.

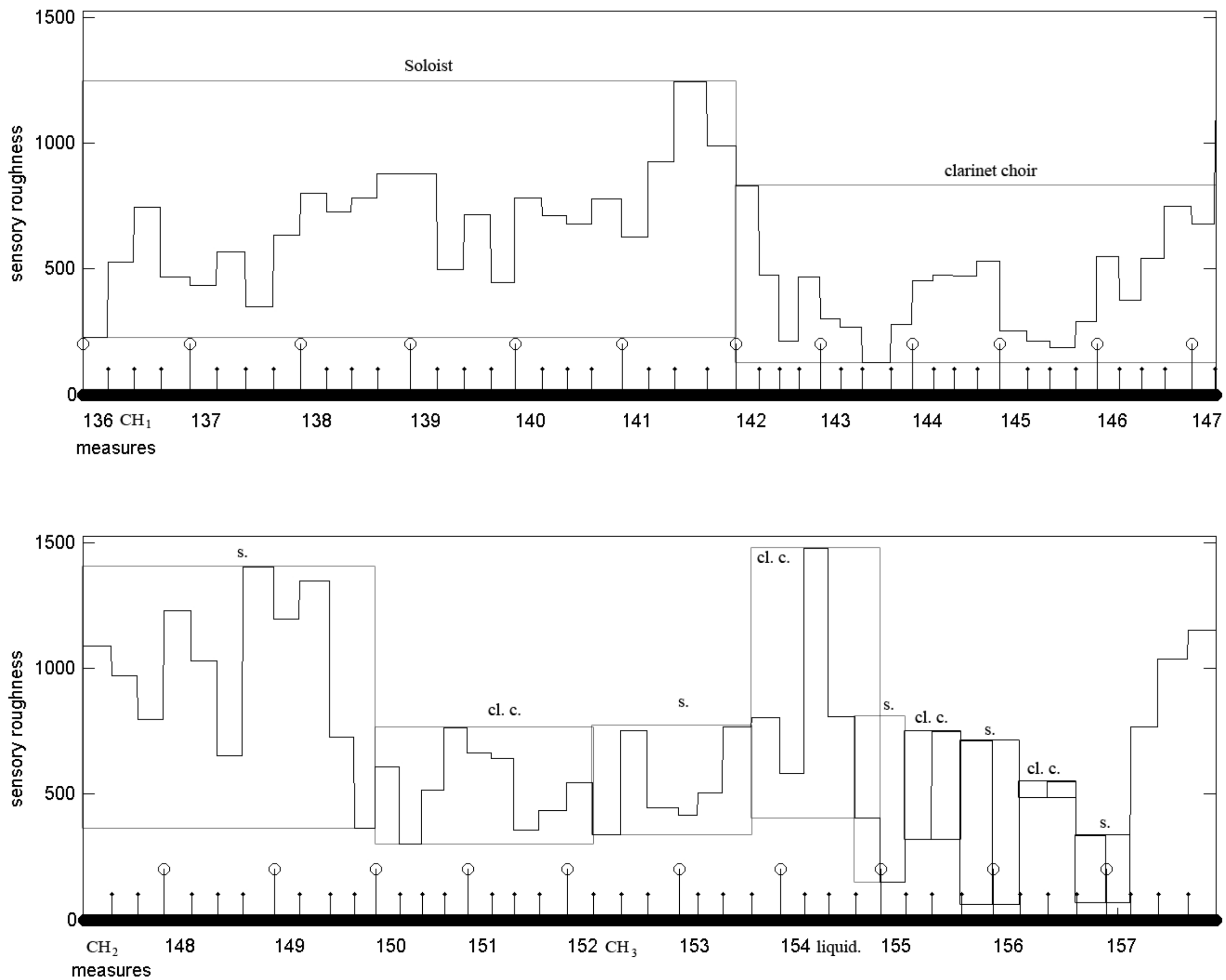


Figure 3.3. Peak variance plot of ms. 136-157. All the PV values in each phrase are boxed together in order to show how the overall PV range changes form phrase to phrase.

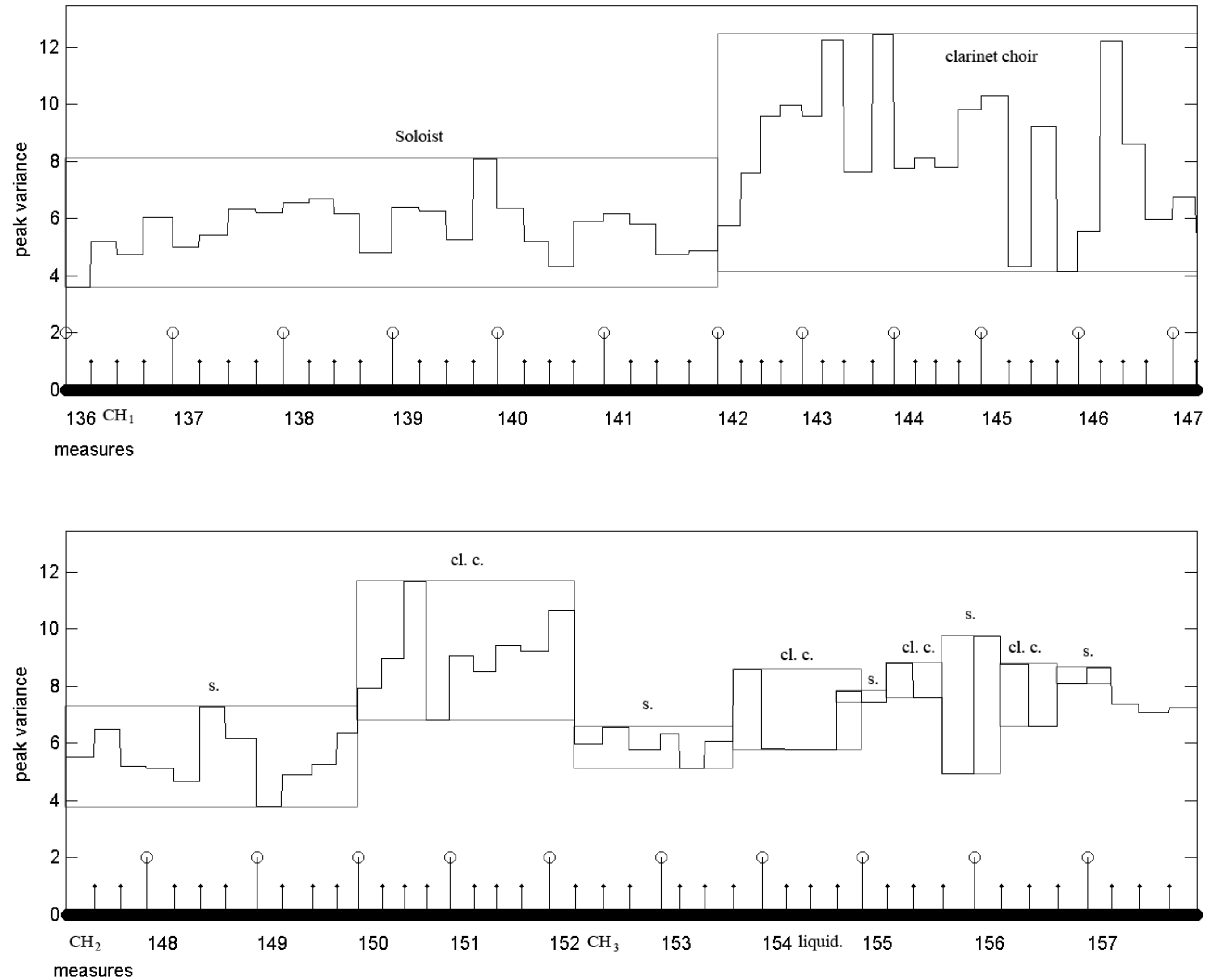
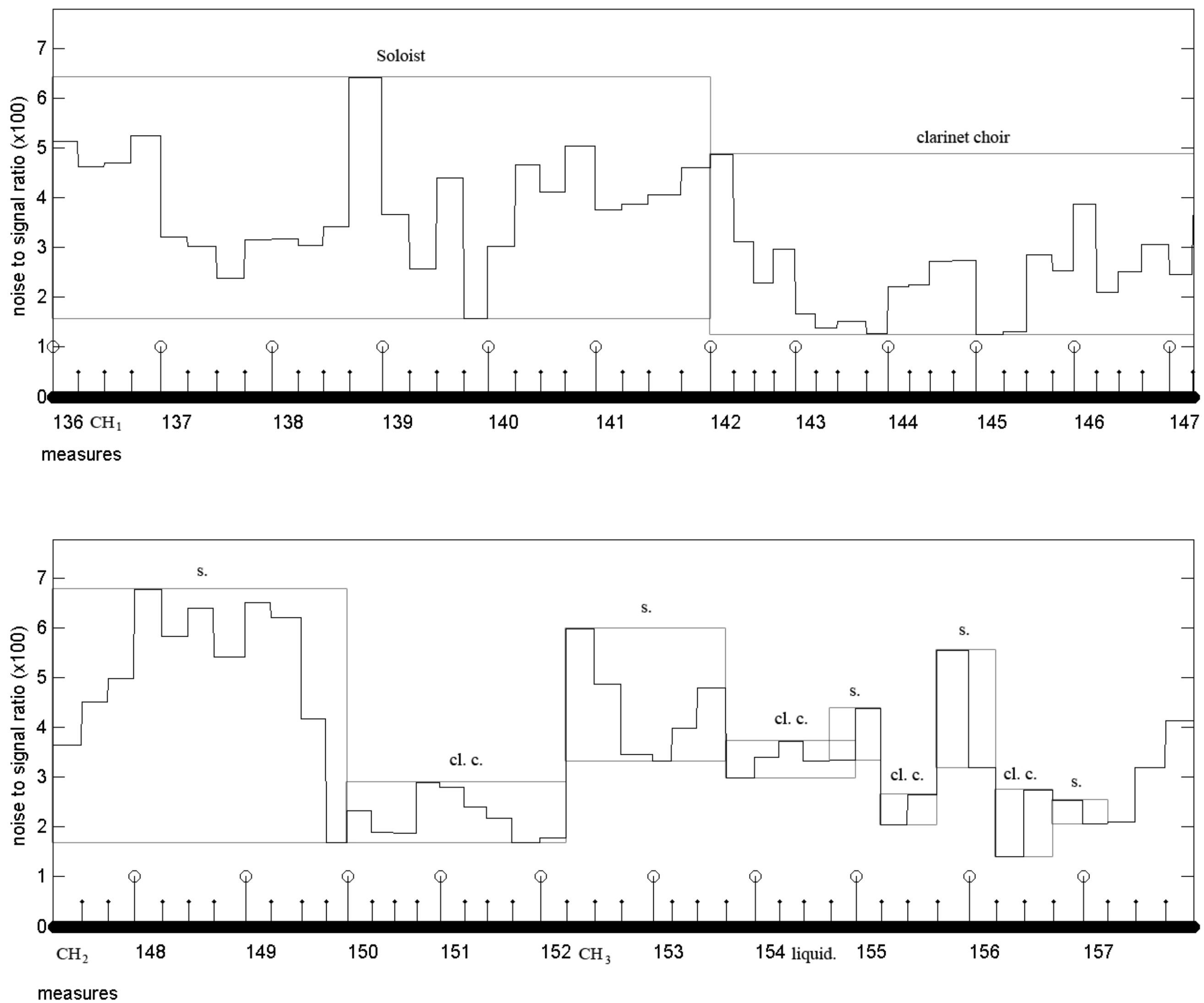


Figure 3.4. Noise to signal ratio plot of ms. 136-157. All the NSR values in each phrase are boxed together in order to show how the overall NSR range changes form phrase to phrase.

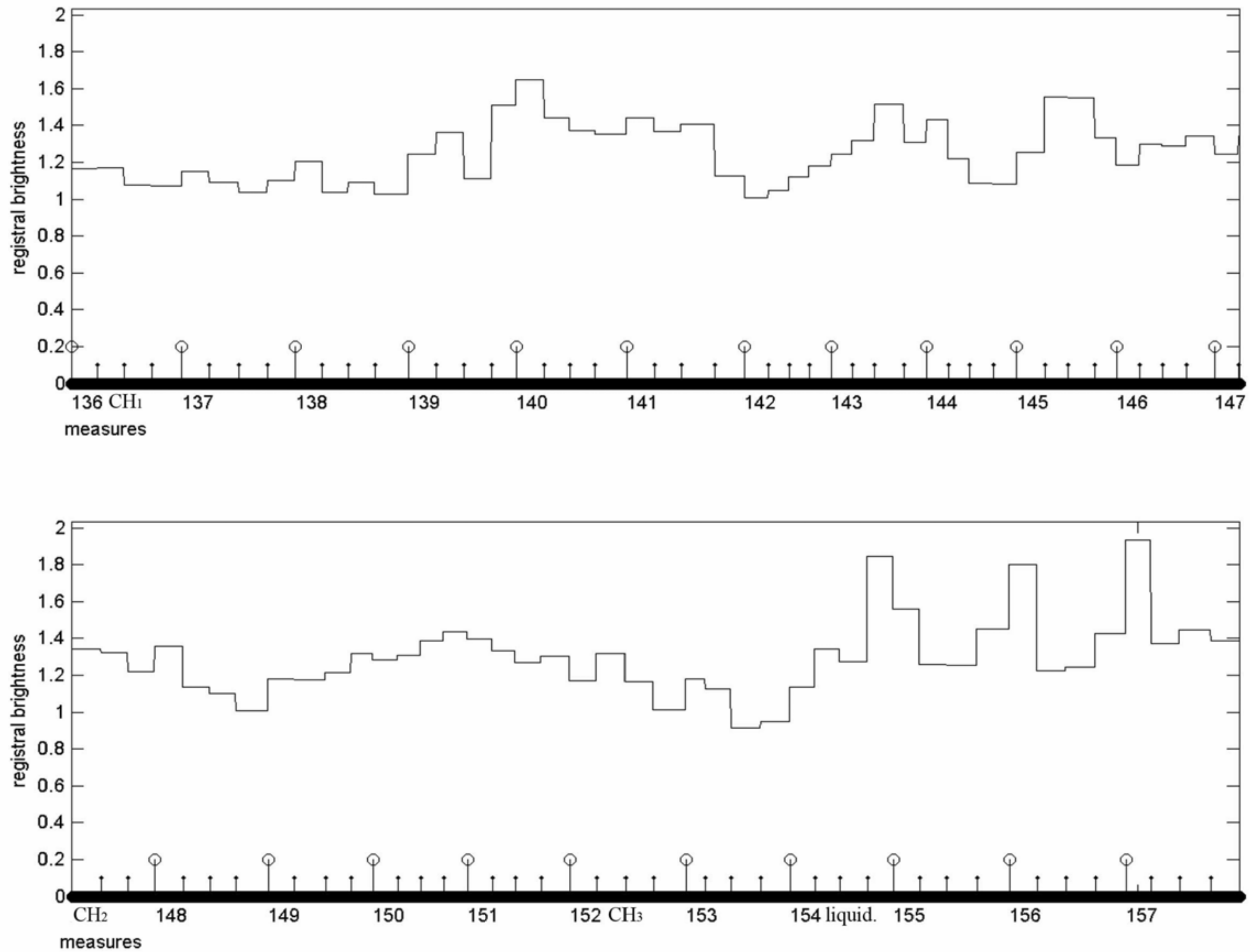


see that it displays a number of "dents," or sudden low points such as the ones in ms. 145.3 and 145.2. These fluctuations correspond to the interjections made by the first violins in every other bar. Given the indications to use a mute and to play with no vibrato, the timbre of this third group provides the perfect bridge between the hollowness of the clarinets and the fuller espressivo sound of the soloist, thus connecting timbrally the first and second statements of the chorale (ms. 136-141 and ms. 142-147) in spite of their intrinsic timbral dissimilarities. The violin interjections also serve to insert a tinge of tonal ambiguity into the otherwise functional harmony of the passage, which again provides a link between the two chorale statements.

Ms. 147-151, continue in the manner of ms. 142-147, this time using CH_2, which is a shorter phrase. This time, there are no muted violin commentaries, which perhaps attests to their transitional character the first time. What happens next is remarkably interesting and can be best understood by looking at the PV graph. As we move into CH_3 in ms. 152-154 and its subsequent liquidation in ms. 154.3-157, the two opposites begin to move toward each other to the point of almost reversing their roles. The performers contribute to this effect as much if not more than the composer. On the one hand the clarinets bring out their swells and change their tone according to the *molto espr. e amoroso* indication, while on the other hand, Mutter, now playing alone, employs her most ethereal and pristine tone. It is noteworthy that the whole process in ms. 136-157 is accompanied by an overall ascent in registral brightness (RB), displayed in figure 3.5 below.

I will save a more personal reading of the passage for a later time, when we can observe it in context. In the mean time, we are now in a position to return to our earlier

Figure 3.5. Registral brightness plot of ms. 136-157.



discussion of the folk song. Now, the texture of the second period can be understood as a communion of the two textures that were set against each other in the chorale, and which approached each other, yet never combined. It is not hard to see from the continuity PV graph, found in Appendix A, p. 85, that this texture created by strings and clarinet playing simultaneously, does indeed generate levels of PV that lie somewhat between the two original characters in the chorale. So, in summary, we have that at the beginning of the Adagio the soloist, with her warm and expressive tone, is set against the initially distant and empyrean sanctity of the clarinet choir. In the course of the passage, however, the soloist increasingly approximates the timbre of its counterpart, which in turn moves closer to her own sound color.

This is not the only place in the piece where the soloist is set against another force, though, nor is the unearthliness of the clarinets the only thing the soloist ever opposes. The section just before the chorale, labeled *Höhepunkt des Allegros* (climax of the Allegro) stages a dichotomy of a very different kind. Whereas in the chorale two foils seemed to be conversing in a way that brought them closer together, in the *Höhepunkt* the two characters presented seem to be battling to death. Figures 3.6, 3.7, and 3.8 once again display the SR, PV, and NSR for ms. 125-135. Looking at the SR plot, the difference between the two groups is huge. The two could not be more diametrically opposed nor could the conflict between them be more excruciating. The highest levels in SR in the entire piece are found in this passage, where the whole orchestra, including four percussion, play a nine-note cluster (the first nine notes of P_1); and the complexity of this texture is set against a three-note cell played by the soloist and a few others in unison! The effect is to set the single voice of the soloist against the massive force of the full orchestra, but this would be impractical balance-wise,

Figure 3.6. Sensory roughness plot of ms. 125-135.

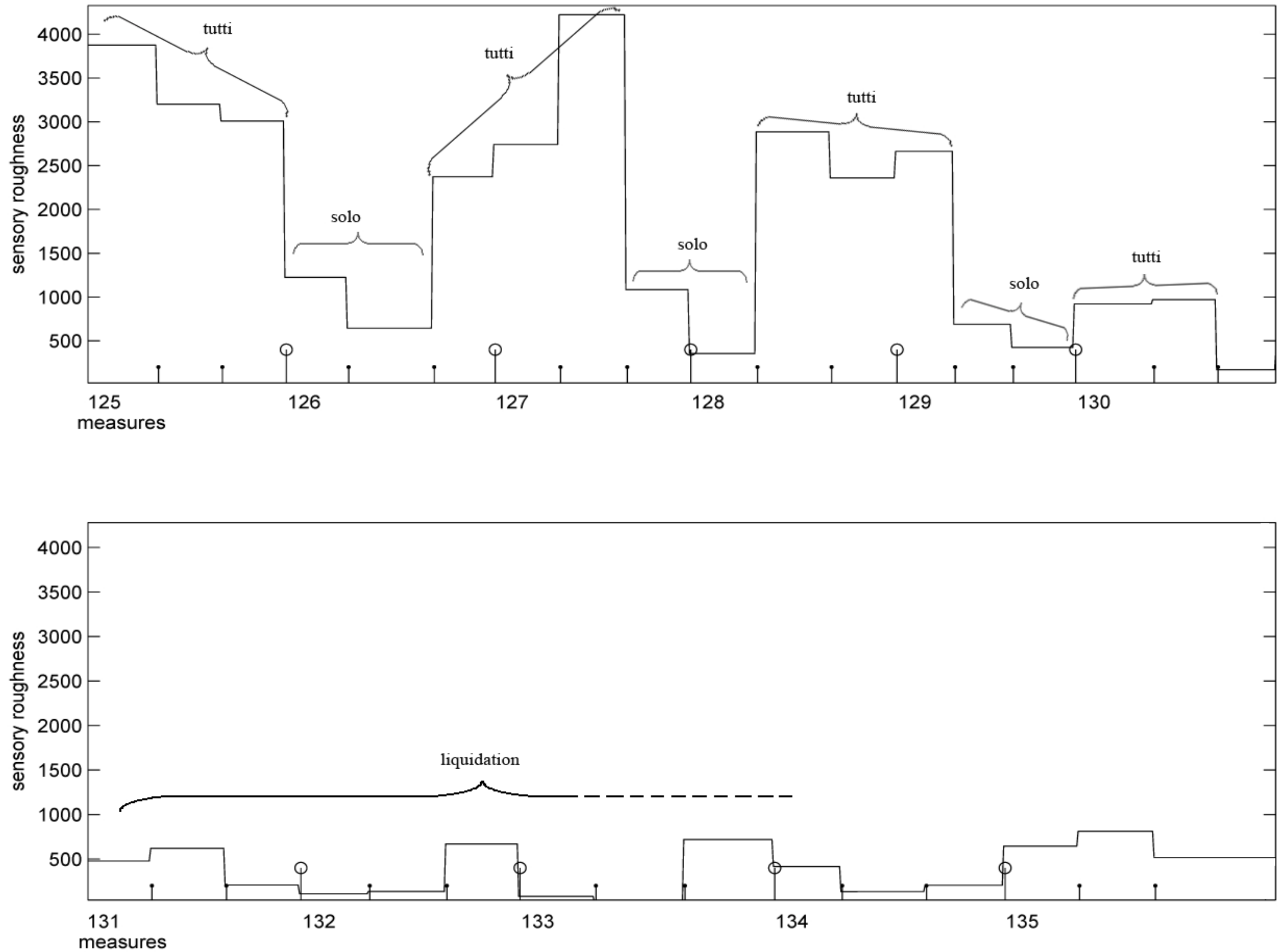


Figure 3.7. Peak variance plot of ms. 125-135.

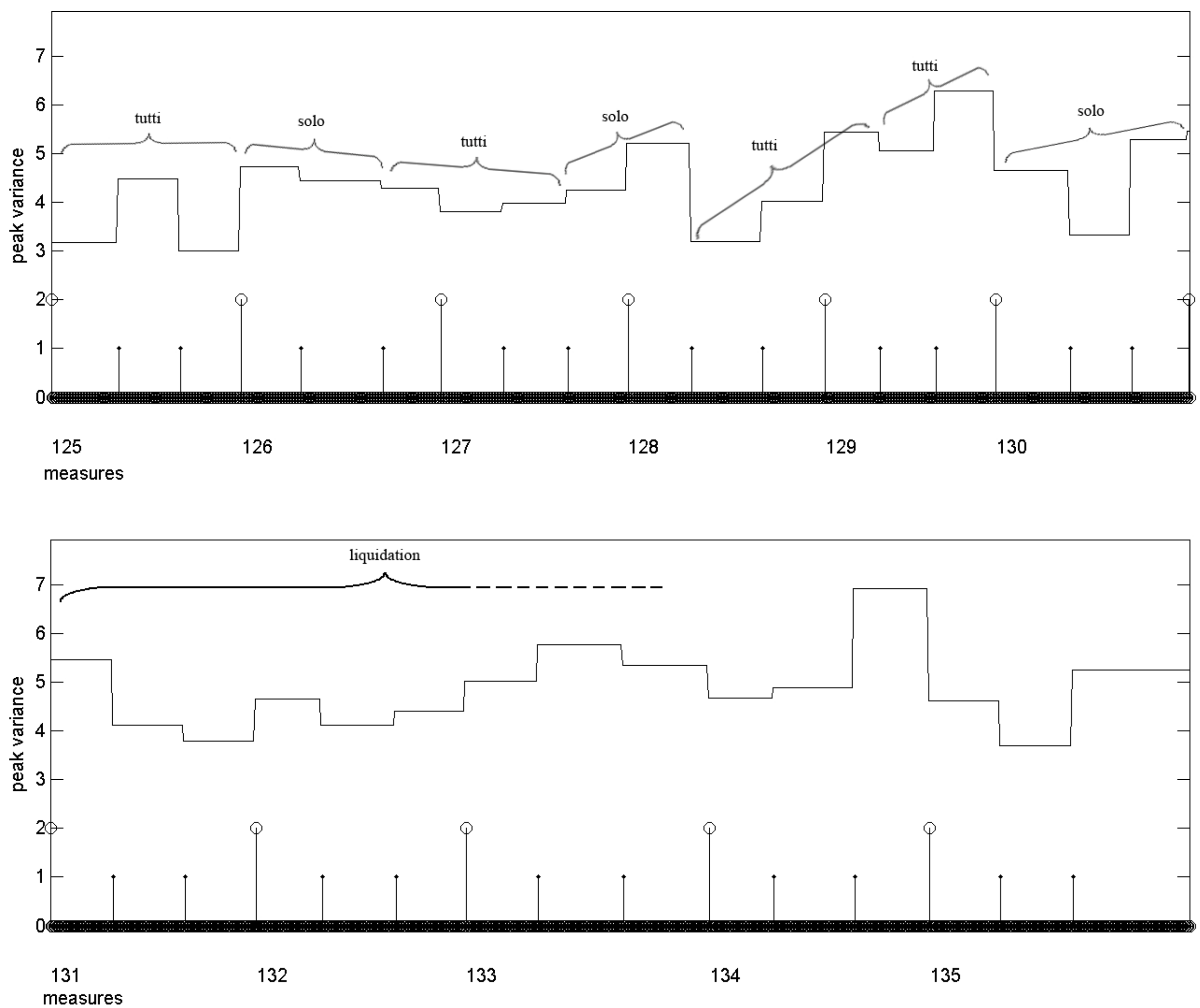
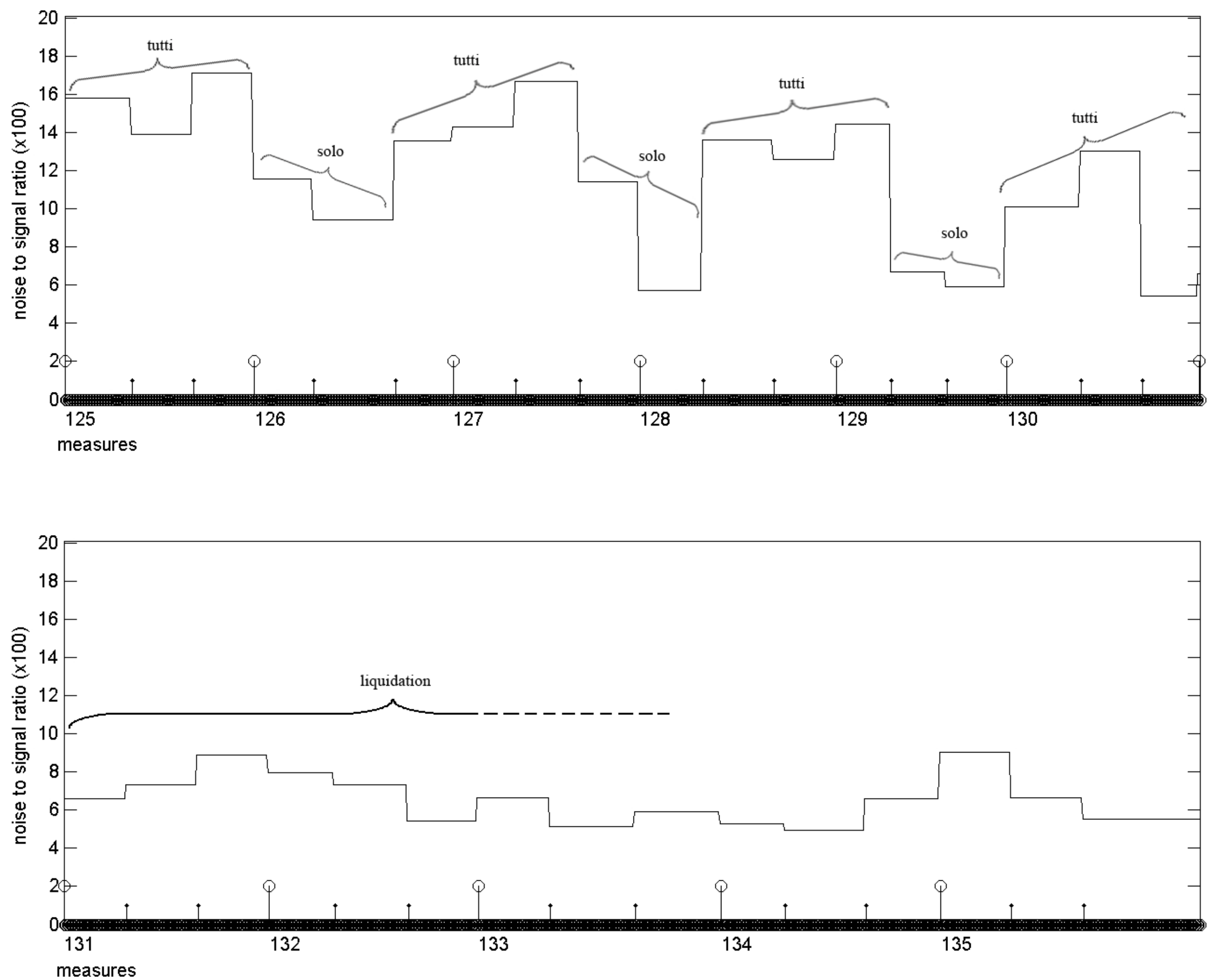


Figure 3.8. Noise to signal ratio plot of ms. 125-135.



considering that the two groups overlap this time. For this reason, a quartet of solo strings and a few winds, support the solo line, which is always one dynamic level higher.

Harmonic tension is not the only thing that differentiates these two groups however. Timbrally as well, these two groups are as different as it is compositionally possible. First of all, with the first group the noise floor is at its maximum in m. 125 (maximum NSR in the whole piece) with the furious utterances of the cymbals, timpani, bass drum, and tam-tam. This last one, marked triple forte, is so powerful and rich that it dominates the entire ensemble, saturating every register in the human hearing range with non-harmonic partials. Perhaps this can be best appreciated from the spectrogram itself (Appendix B, p. 98, m.125.3) where the darkest line represents the resonance of the tam-tam (and not the F pedal as one could infer). In terms of the timbre of the melodic tutti instruments, the texture is dominated by the brass—the instrument group with lowest PV in the orchestra. The remaining winds (oboes, piccolos, and clarinets) are all scored at the top of their register, ensuring their shrillest sounds. Winds and brass are supported by the whole string section playing pizzicato, adding to the poignancy of the *Hauptrhythmus* (main rhythm), which Reich describes as "the suffocating rhythmic pressure of destruction." The result is one of the lowest PV moments in the score.

Rather than dropping out completely for the soloist's howl, the winds and brass drop to mezzo piano, and sustain their chord while the soloist and her squad respond to the percussiveness of the *Hauptrhythmus* with three sustained, swelling notes. This accounts for the fact that even though we can see the contrast between the two groups in the PV graph, this contrast is not as marked as it is in the chorale. The effect resembles the crashing of huge waves in a storm while a desperate shipwrecked swimmer tries to keep her head above the water. It is noteworthy that Berg did not choose to pair the strings with the clarinets for the sake of a more drastic contrast in PV. He chose instead timbres that are closer to the violin in PV, namely the double reeds, the sax and the *bass* clarinet. A very similar timbre (solo violin, violas and bassoon), can then be set against a yet higher level of PV, when the clarinet choir comes in with the tonal statement of CH_1.

As we approach the chorale, what we perceive is not any kind of understanding between the two forces, but rather the kind of "gradual collapse" (to quote Reich's program again) experienced in a battle that has been drawn out for too long. Both groups begin transitioning color-wise to what will be the new texture, the noise receding, the roughness renouncing, and the PV slowly ascending. As the *Hauptrhythmus* begins fragmenting, note by note between each anguished cry, the soloist begins formulating her final words, which will begin the Adagio: "It is enough."[1]

Macro-Structure: Movement II in Continuity

In this section I will consider the entire movement as a continuous whole, placing the three sections already studied within their larger context and venturing a more comprehensive interpretation of their function within the piece. Figure 3.9 provides a synthetic map of movement II of the Concerto, represented by its most salient structural divisions. Labels for the various sections are my own and are congruent with the interpretation I will outline in the following pages. The pertinent SR (sensory roughness), PV (peak variance), RB (registral brightness), TB (timbral brightness), RS (registral Span), and

---

[1] This would be Berg's own final words before his death, five months after the completion of the work (Pople 1991, 43).

Figure 3.9. Structural Map of Movement II

Movement IIa: Allegro

Main parts	**A** *(Cadenza)*			**B** *(Fate Rhythm)*
Measures	1	7	19	23
Sections Thematic Material	**a**	**b**	**a'**	**HR**: Hauptrhythmus (main rhythm)

C *(Memories)*		**D** *(Dream)*			**C'** *(Searching)*	
44	54	61	68	73	78	91
$\mathbf{c_1}$ *(love)*	$\mathbf{c_2}$ *(scherz.)*	c_1 + RH+P-Row	CH_1 *(premonition)*	improvisatory *(waking)*	$\mathbf{c_1}$	$\mathbf{c_2}$ *(hope)*

A'	**B'** (interpolated)		**A'** (resumed)	**Höhepunkt** *(Climax)*
96	104	111	120	125
a	**HR** ⟶ (intensified)		**a'**	HR⟶CH_1 climax/transition

Movement IIb: Adagio

Chorale				**Variation I**		
136	147	152	154	158	164	176
$\mathbf{CH_1}$	$\mathbf{CH_2}$	$\mathbf{CH_3}$	Liquidation	**"Plaint"** (solo) ⟶		$\mathbf{CH_3}$
				CH_1 (cantus firmus) → CH_2 →		

Variation II		**Folk Song**	**Coda**	
178	186	200	214	223
Klage. (cont.)	***Höhepunkt***→ *Klage.* (cont.)	Folk Song	$CH_{1,2}$, *Klage.*	CH_3, P-Row *(final ascent)*
CH_1i ⟶	→CH_2i → CH_3i→			

NST (noise-to-signal ratio) graphs for the whole movement in continuity can be found in Appendix A (pp. 84-88).

Reich describes the opening of the Allegro as “a wild orchestral cry.”[1] Certainly the wild and brutal tone of this movement asserts it self immediately as the sound escalates, in the span of three beats, to extreme levels of SR and NSR. The texture is dominated by the brass and given that the strings only punctuate the wind and brass entrances with pizzicati, the overall texture is nothing short of a premonition of the “catastrophe” that will come in the *Höhepunkt*. This is reflected by low PV levels. Images of lovely girls and youthful dances have vanished and all that remains is struggle and pain. Once the tone is established and the full force of the orchestra has been heard playing the primary form of the row as a cluster, intensity transfers to the soloist, who emerges from the texture as she ascends in register. Whereas SR and NSR drop down immediately and permanently after m. 2, TB and RB continue to escalate till m. 5. Marked “always rubato, free like a cadenza,” section A as a whole (ms. 1-22), gives the impression of a fiery improvisation. Starting in m. 7, the soloist rambles furiously, while the orchestra comments impassively, save for an occasional off-beat pummel. Brightness quickly descends to a local minimum at m. 14, both timbrally and registrally. Much of the contrast in this section derives from these two parameters, which again escalate as the opening material returns in m.19. This is supported by the NSR, which also rises into m. 19, first through the thickening of texture and the choice of attack (pizzicati and accents) and then with the help of the tam-tam at m. 19.

M. 23 marks the beginning of section B, which is characterized by the ostinato *Hauptrhythmus*. We know from the secret score of the *Lyric Suite* that 23 was a number Berg

[1] See program on p. 52.

associated with himself, having had his first asthma attack at age 23 on the 23rd of July (Pople 1991, 61). Thus we can see the *Hauptrhythmus* as the agent of catastrophe, the symbol of inescapable fate. This section is structured in two parts, one dominated by the brass and winds with the soloist as obbligato (ms. 23-34), and one in which the roles are reversed (ms. 35-43). Both the SR and NSR graphs show each one of these segments starting at a relatively low level of intensity and then increasing to the end. Save for the parting commentary of the brass section in m. 42, the entire section furtively transitions in timbre to the dreamy world of the middle section, as shown clearly from the PV and NSR graphs. Whereas the first intensification was brass-heavy and made extensive use of the bass drum, the second combines solo violin, bassoon, and bass clarinet and has no percussion. This is where the two forces that will contend for life or death in the *Höhepunkt* appear for the first time; on the one hand, fate and worldly suffering as represented by the brass with percussion and pizzicato strings, on the other, the struggling heroine (or hero)—whether Manon Gropius or Berg himself—struggling on the verge of catastrophe, represented by the solo violin and low reeds.

The inner sections of this quasi mirror form (C, D, and C', ms. 44-95) deviate from the character that has prevailed so far in the movement. It is as though the reality experienced up to this point is too crude and painful to take in and we must take refuge in a world of memories and dreams. Approaching C, the soloist continues with a variant of the *Hauptrhythmus,* but the focus, which before had been on the crudeness of exterior agents (i.e. fate, worldly suffering, etc.) as represented by the strident cries and blows of the tutti orchestra, here turns inward. The *Hauptrhythmus* becomes merely an accompaniment to the memory in the flutes of what I call the *love* theme from m. 155 in movement I. The texture, consisting only of these three instruments with an occasional harp roll, is the thinnest and most delicate in the entire Allegro, which accounts for the extremely high PV levels, the highest in the entire Allegro. This delicate, dream-like timbral quality, contrasts with the earthbound, strident attributes of its predecessor, and becomes the main trait of the middle section, generating relatively high PV levels throughout.

As the three instruments continue their inward journey, the *Hauptrhythmus* is forgotten and the soloist becomes completely immersed in the world of the flutes, of her own memories, leading to an apex in RB at m. 50. It is not surprising that this section shows very low SR and NSR in addition to the PV attributes discussed above. Both of these parameters escalate with the approach of the scherzando theme, which in the first movement we were taught to expect after the *love* theme. It is remarkable that even though the harmonic intensity (SR) increases suddenly with the appearance of a 12-note cluster (I_0) in ms. 53-54, this is done within a pretty low dynamic level (mezzo forte at most), maintaining the overall sense of introspection. It is also worth noting that although the register expands downward with the enlargement of the ensemble, the texture remains top-heavy (as shown by the increase in TB) maintaining a sense of detachment from reality, of distance from the ground. In this case, rather than adding actual darkness to the color, the presence of lower frequencies provides a sense of perspective, which reinforces the impression of elevation and brightness in the solo line, while setting it in relieve against its background in a kind of chiaroscuro effect.

As we reach D memories assume the irrationality of dreams; random themes and ideas melt freely into a tranquil improvisation in ms. 61-67. Then something remarkable happens: the orchestra drops out entirely in m. 68, and there is a pristine premonition of the chorale melody (CH_1), a beam of sunshine in the eye of the storm, a vision of something greater than our worldly toils. The timing of this performance is such that m. 68 happens exactly halfway through the duration of the movement; the very heart of the Allegro. As the orchestra evaporates and the soloist ascends to this magical state, RB is brought to a local maximum. The only accompaniment to this vision is the soloist's own left hand pizzicato, which is why SR drops dramatically. It is also the reason why the NSR is laden with momentary fluctuations, which lend the texture its scintillating quality.[1]

The vision lasts only a few moments; soon, in ms. 73-74, gradual awakening becomes desperate pleading. This process is still carried out mainly by the soloist, with some help from the low strings. Hence, until we reach the return of the *love* theme in m. 78, the color itself is not much different from what it was before the vision. Mutter and Levine in this recording perform the ossia notated in the score, which splits the strict four-part canon based on the *love* theme between the soloist and a solo viola (instead of the soloist alone). This allows them to attain slightly higher PV values, a very clean sound (the lowest NSR level yet), and low dynamic levels, perhaps making the section timbrally more akin to the first version (the original C) than if the soloist played all four parts at once.

After the magical vision of the chorale, the return to the *scherzando* idea is barely recognizable by its sheer character. Nothing is left of the contained roughness of the twelve-note cluster; on the contrary, SR drops to remarkably low levels as the soloist brings the section to a close with a lighter and cheerful spirit—while accompanied only by two clarinets! Our vision has given us the hope we lacked in the first half of the movement; hope of a synthesis that will only be truly possible towards the end of the piece (as we saw in the

---

[1] Given that the pizzicati are performed at the sixteenth-note level and our readings are done on a beat-by beat basis, these fluctuations are not as clear in the graph as they would be if the readings were done with a smaller window. Perhaps they can be better appreciated in the raw spectrogram on p. 94.

discussion of the folk song), and promised to us by this brief and unique moment of cheer, where clarinets and violin sing together for the first time in the movement.

Painful reality returns forcefully and abruptly with the recapitulation of A in m. 96. In our memory, this music is the roughest thing encountered so far, and as such we experience it again, but curiously Berg chooses to voice and orchestrate the twelve-note cluster in such a way that it creates the least possible SR. Whereas at the beginning of the movement the voicing of the chord was rich in sevenths and seconds, here the voicing is such that the final chord is simply a vertical alignment of the primary row (P_5)—not unlike the chord at m. 54 in the first scherzando section, which displays a similar SR level. Levine heightens this effect by directing his brass section to drop down immediately after their entrances, leading to a texture dominated by the wood-winds in ms. 96.3-97. This is a way of rationing intensity, for there is a long way to go before the *Höhepunkt* and conserving energy will heighten its effect. Unlike the first time, the apex of this section—both in terms of SR and NSR—is not within the first couple of measures but five measures later in m.100, when the timpani announces the return of the *Hauptrhythmus*—the relentless beating of fate. From this moment on the *Hauptrhythmus* will become the driving force, eventually leading to the *Höhepunkt*. SR escalates dramatically yet momentarily in m. 100, mainly due to the non-harmonic overtones of the timpani playing *fortissimo*. It immediately drops again as the previously held C^b major heptachord crumbles down in preparation for the interpolation of B' beginning in m. 104.

In contrast to the exposition, here section B' is inserted within the two iterations of the opening idea; see structural map on p. 70. Once again B' comprises two intensifications of the *Hauptrhythmus.* This time, however, the roles are reversed, and instead of traversing from the extrospective inwards like they did the first time they do the opposite: the first intensification falls to the solo violin (ms. 104-110) and the second one (ms. 111-119) mostly to winds, brass, and percussion, with solo violin obbligato. The pacing of intensification leading to the *Höhepunkt* can be best appreciated from the NSR graph. Throughout the solo statement there is an ominous timpani roll, which nevertheless stays relatively quiet. When the winds and muted brass take over the *Hauptrhythmus* in m. 111, however, the timpani stops rolling and begins to support the rhythm, while gradually increasing in dynamic level. In m. 115 we take a step back as we approach the resumption of A' in m. 120, where the timpani transfers its duties to the snare drum (which lacks the low-frequency noise of its predecessor). It is only in m. 122, three measures before the *Höhepunkt* arrives, that the timpani picks up again, creating an enormous surge in NSR that would not have been possible had it not receded in the preceding measures. This final surge in noise is mirrored in the instrumentation, which thins out dramatically in m. 121 and then, in the span of four measures, goes from cellos and soloist alone to absolutely every instrument in the orchestra at the downbeat of m. 125. Because noise levels are so high throughout this section, we cannot rely on the SR algorithm to say much about the actual harmonic tension of the passage; the routine operates on harmonic elements only, and these are blurred by the noise saturation. On the other hand, we can clearly see the PV steadily descending form m. 111 to its lowest point at the *Höhepunkt*—the single most saturated and abrasive texture in the entire piece.

After an intensification of such gigantic proportions, the pressure of fate is unbearable. In the midst of a catastrophe where roughness and timbral saturation reach excruciating proportions, our heroine struggles for life. Time and again she is assaulted by the inclemency of a world too ruthless to spare her life; time and again she is engulfed by the colossal waves of her own destruction, and the only way to placate them is through one last prayer: “Es ist genug! Herr, wenn es Dir gefällt, so spanne mich doch aus!” Thus begins the Adagio. The extreme intensity levels of the *Höhepunkt* have all receded, and all that remains is the bleeding resonance of the tam-tam. We don’t hear its final *ppp* stroke, but its presence (as shown by the RB graph)[1] makes the passage one of the darkest yet. Gradually, this resonance clears like smoke after a battle, and the prayer rises from the ashes. There, earnest and solemn, stands our heroine between two worlds: the one she has just left and another that comes without delay to meet her, compassionately answering her prayer.

Just as on a micro-structural level we could see high and low levels of PV set against one another to symbolize the position of our soloist within a certain context, now we can see on a grand scale two different worlds (or perhaps one world viewed from two different angles), as symbolized by the Allegro on the one hand and the Adagio on the other. The PV graph shows neatly how the Allegro is bound at the beginning and end by extremely low levels, only escaping from them in the dreamy middle section, which tends toward the opposite side of the spectrum. On the other hand, the Adagio is framed by the ethereal character proper to high PV. It too provides contrast to these bounds toward the middle, where it descends, discreetly approaching the Allegro, yet never quite reaching its extreme levels. This time, however, the relationship between the low and high levels of PV (among other things) is handled in a much different way, assuming an entirely new meaning.

[1] The predominance of the tam-tam resonance in this passage can be seen most clearly in the raw spectrogram (Appendix B, p. 99).

The music between the chorale and the folk song (ms. 158-200) consists of two variations that use the chorale melody as a cantus firmus, first in its original shape, then in inverted form. Once again Berg indicates where the climax of the movement takes place: m. 186 (*Höhepunkt des Adagios*), about half way through Variation II. This moment is the culmination of a journey that leads from the darkest texture in the Adagio to the brightest one yet (as can seen in the RB graph): indeed a "struggle towards the light," as Reich describes it. This journey takes the shape of a double cycle, with an intensification towards a smaller climax (m. 176) preceding the massive surge towards the second.

The first cycle spans the length of variation I, which begins with a sudden drop in RB accompanied by an increase in overall SR. A chill runs down our spines (cymbal, m. 158.2) as we hear the double basses lugubriously intone CH_1, followed in canon by the low register of the harp. Six measures of somber uncertainty pass by before the soloist (now muted) begins her pious lament, her *Klagegesang*. SR begins to drop slowly, and as RB rises, something remarkable happens: in m. 170 the soloist is almost imperceptibly joined in unison by the concertmaster, then the second chair (m. 173), then half a section of first violins (m. 175). Together, they reach the first acme in m. 176, where the *Klagegesang* suddenly becomes the first statement of CH_3. With SR at a minimum and RB at its first apex, this is a moment of great clarity, but something strange has happened: as more and more violins join the soloist in her *Klagegesang* and the acoustic beating they create increases, her individual voice begins to dissolve in the collective song.

The entire first violin section then resumes the *Klagegesang* in m. 178, thus beginning Variation II. The new intensification leading to the formal *Höhepunkt* is compressed to half the length of its predecessor and, almost like a fractal image, it mirrors in small scale the overall double cycle of the variations. After a brief decrease in RB at the beginning of variation II, the *Klagegesang* begins to ascend again, bringing the texture to a first apex in m. 184. A cymbal roll accompanies this first crescendo, and both violin sections now intone the song with a single voice, meeting the brass chorale on an A dominant chord in m. 184 (very low SR). At this moment the thus far compact register begins expanding downwards while the *Klagegesang* continues to ascend. A tam-tam roll saturates the spectrum, adding depth to the already unprecedented fullness of the texture. Thus, the entire upper string section reaches together the *Höhepunkt* in m. 186. Registrally and timbrally, this is the brightest moment yet, even though all registers are active (including the extremely low). The final intensification brings the *Klagegesang* to its most exalted heights, but it also adds perspective and depth. At the same time, the individuality of our heroine is now completely absorbed by the immensity of her context. Yet her presence is felt and her voice is strong. She is no longer at odds with the world, as she feels herself an integral part of it. Together with the entire string section, she echoes in diminution the chorale in the brass, and the exalted figure resounds through the orchestra in a moment of utmost unity, timbral and otherwise. Then the vision begins to dissolve and gradually the strings begin to withdrawal in the same way they entered until the soloist is again alone in m. 197. As she hovers above the orchestra one last vision emerges before her. It is the folk song of her youth (m. 200), which she accompanies melancholically as she continues her transit towards evermore ethereal planes of existence (as symbolized by the change in timbre, discussed on p. 56).

The Coda (m. 214) reenacts the process observed in Variations I and II, bringing the piece to its conclusion in two more ascending gestures. First, the winds resume the chorale and the soloist intones the *Klagegesang* once more, ascending to the highest levels of RB and

TB in the movement. The fist apex is occurs in m. 222 as the *Klagegesang* once again becomes CH_3. The final melodic ascent in m. 223 covers all string sections, starting with a solo double bass and climbing all the way to the now un-muted soloist playing at the top of her register. It is accompanied by the chorale in the brass (recalling the synthesis of the *Höhepunkt des Adagios*). This ascent does not lead to extreme registral or timbral conditions. On the contrary, in the final measures the soloist reaches a state of complete serenity (the long, sustained, high G), and all parameters fluctuate around moderate levels while all instrumental groups make one final appearance. Thus the soloist floats away above a texture that resembles, peacefully, the one at the beginning of the movement.

Conclusion

In an effort to expand our understanding of sound color and its structural and interpretative role in music, this paper has presented an analytical approach that can be used to facilitate discussion of this most elusive of musical parameters on the basis of its concrete sonic attributes. The Fourier spectrum has served as a point of departure to examine specific elements of sound color and trace their development within a given musical context—in this case, a specific performance of the Alban Berg Violin Concerto. Using the data provided by Fourier spectra, the various algorithms I have outlined make it possible to derive visual representations of specific sonic elements directly from a digital recording. Whether sensory roughness, noise to signal ratio, brightness, or timbral poignancy (peak variance), these algorithms isolate a particular characteristic of the sonic tissue by consistently computing a set of relationships between spectral elements at various times. When the results are plotted against time we obtain a rough visual analogy to the aural perception of each sonic characteristic.

As seen in the discussion of the Berg Violin Concerto, these graphs allow one to step outside of time and contemplate the development of sound color through large sections of music. Long musical gestures, which may be perceived aurally as slow transformations of sound color, become shapes in a plane, and notions of contrast and change find their analogy in the slopes and turns of the various plots. This approach not only elucidates the development of sound color through a given piece of music, but it also provides a clear and concrete vocabulary to discuss it. Regardless of whether my conclusions about this music are shared by other listeners, the very possibility of verbalizing and discussing them on a concrete basis has allowed me to assimilate and rationalize aspects of the concerto that, otherwise, would have remained at the level of vague intuitions.

Having performed and studied the Concerto intensively prior to my spectral analyses, I saw reflected in them many of my old observations and intuitions, but the more I plunged into the sonic world of the piece, the more my attention was directed to details and aspects of the score I had never considered. These realizations and the interpretative conclusions I drew from them will undoubtedly have a significant impact on the way I will perform the work in the future. Now equipped with a new found understanding of how various sound colors interact and evolve throughout the work, I can stress their tendencies and italicize large and small-size gestures, contrasts, and shapes. This will, in turn, allow the listener to identify and interpret such gestures, while endowing the piece with an added degree of unity and cohesion.

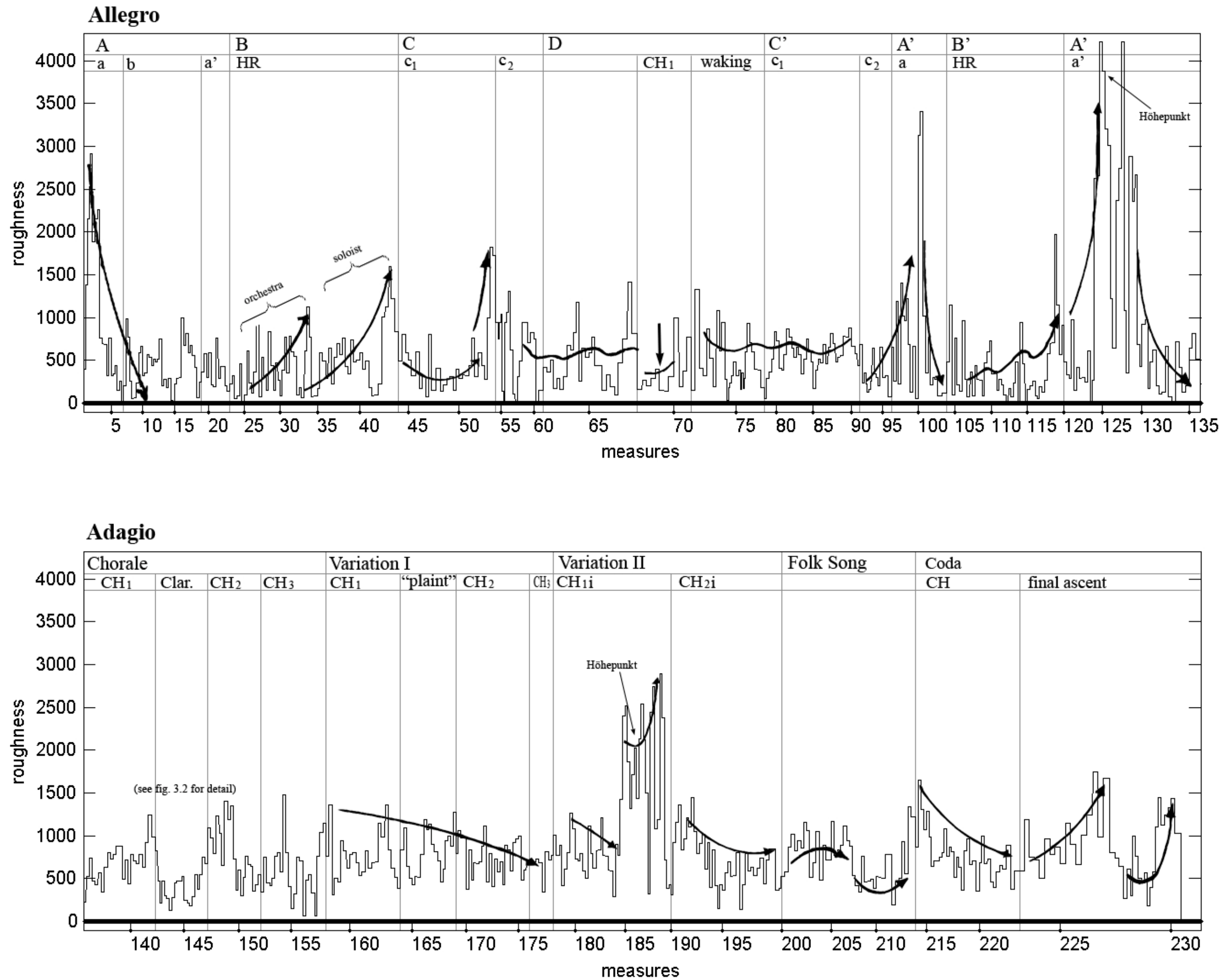
Naturally, the techniques and tools presented in this paper are nothing but a point of departure for future and more rigorous exploration. As mentioned above, the various algorithms and the measurements they yield are but rough approximations to perceptual processes, and a more precise and unambiguous simulation of their perceptual counterparts would be most beneficial. A few improvements in this direction would not be too difficult to implement on the existing algorithms. For instance, instead of using decibels as the standard scale of intensity level, one could use *sones*, (a measurement of perceived loudness that varies with frequency range). Other possible improvements would be significantly more involved. The peak-picking algorithm in particular could be improved in a number of ways: masking effects could be taken into account, as well as register-dependent features of frequency and pitch perception, such as critical bands. An effective integration of some of these notions into the kind of algorithms presented here, however, would require further psychoacoustical research.

Though this paper defines and explores five different parameters of sound color, this is not to say that there are not many more. An attribute of sound that according to Roederer (1995) is instrumental in the recognition of timbre and the identification of its source is envelope—which includes onset and decay, as well as the presence of transients. But Roederer himself acknowledges that this attribute is far more difficult to quantify (151). For this reason, envelope and its impact on sound color are hardly considered in this paper. However, it would be worth exploring this parameter and its interpretative implications, particularly in the case of acoustic beats and vibrato, which can have a tremendous impact on the interpretative affect of a given sound.

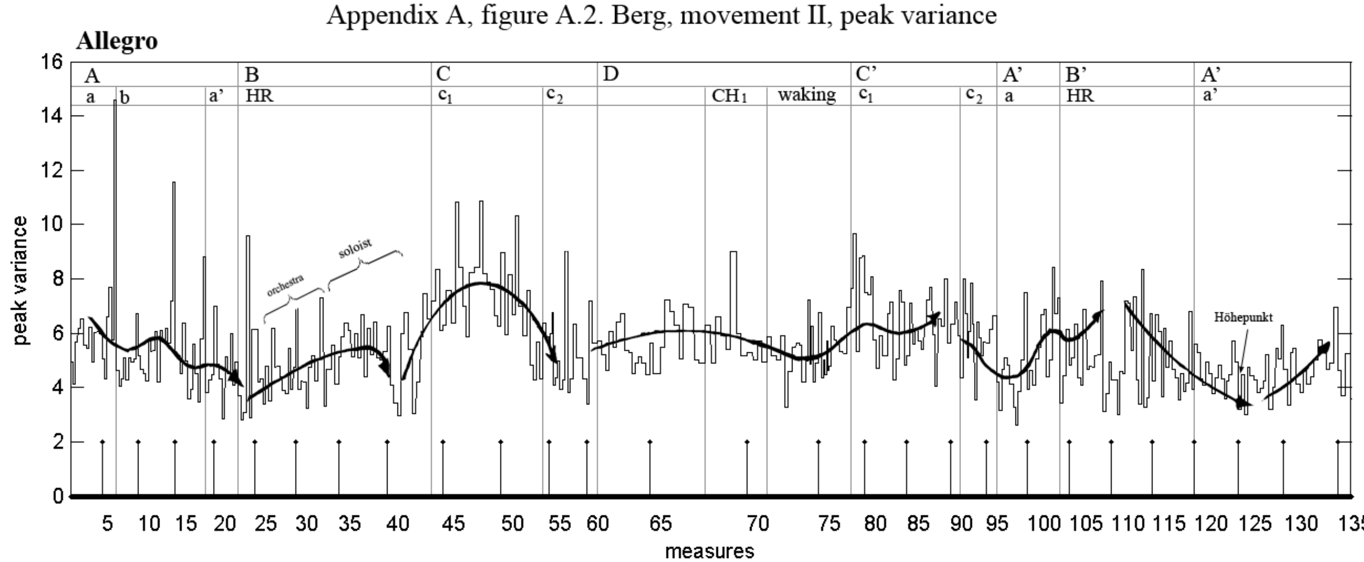
Notwithstanding their weaknesses and their potential for future refinement, the tools presented here have already proven helpful to elucidate sound color in the context of the Berg Violin Concerto. Now that the first analysis of its kind has been performed, many other pieces await to be explored in the same fashion; and not only that, but specific performances, as well. A valuable enterprise would be to compare several recordings of the Berg Violin Concerto—or any other piece for that matter—and contrast the specific musical choices of the various performers. This would not only further our understanding of the pieces at hand, but would allow us to investigate more concretely some of the elements that make certain performances more effective than others.

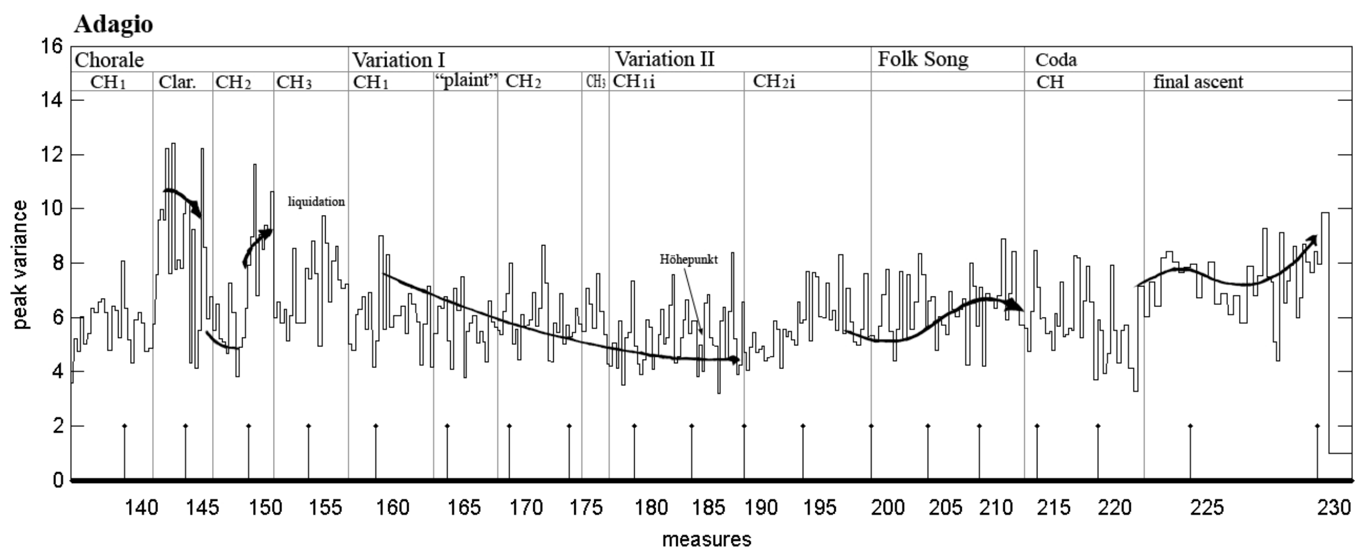
Given the remarkable degree of freedom and choice performers have over sound color—compared to its counterparts, pitch and duration—the benefits for performance of furthering our understanding of this parameter go without saying. Not only can we study how the great masters of the past and present chose to manipulate sound color, but we can apply this kind of analysis to our own playing, shedding unprecedented light, at the individual level, on the relationship between the mechanics of sound production and their actual sonic effect. Furthermore, this analytical approach could be used by composers and conductors alike to better understand the intricacies of orchestration. Finally, the more we can understand the connection between the physical and the psychological/emotional realms, the closer we will be to understanding ourselves, as well as our relationship to one another and to the world in which we live.

Appendix A, figure A.1. Berg, movement II, sensory roughness



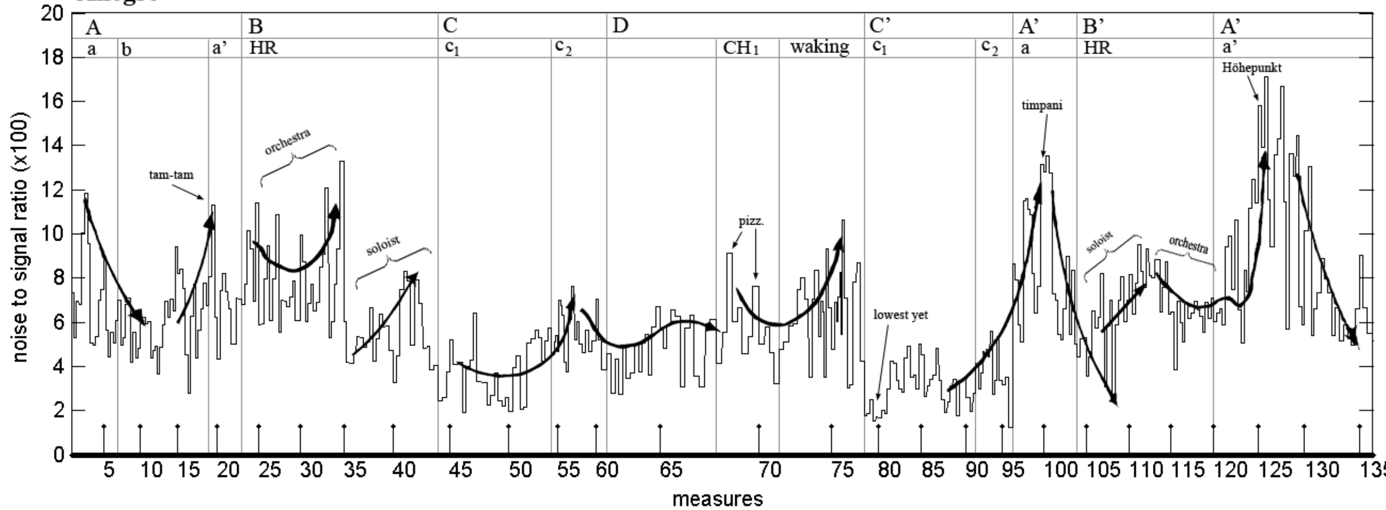
Appendix A, figure A.2. Berg, movement II, peak variance



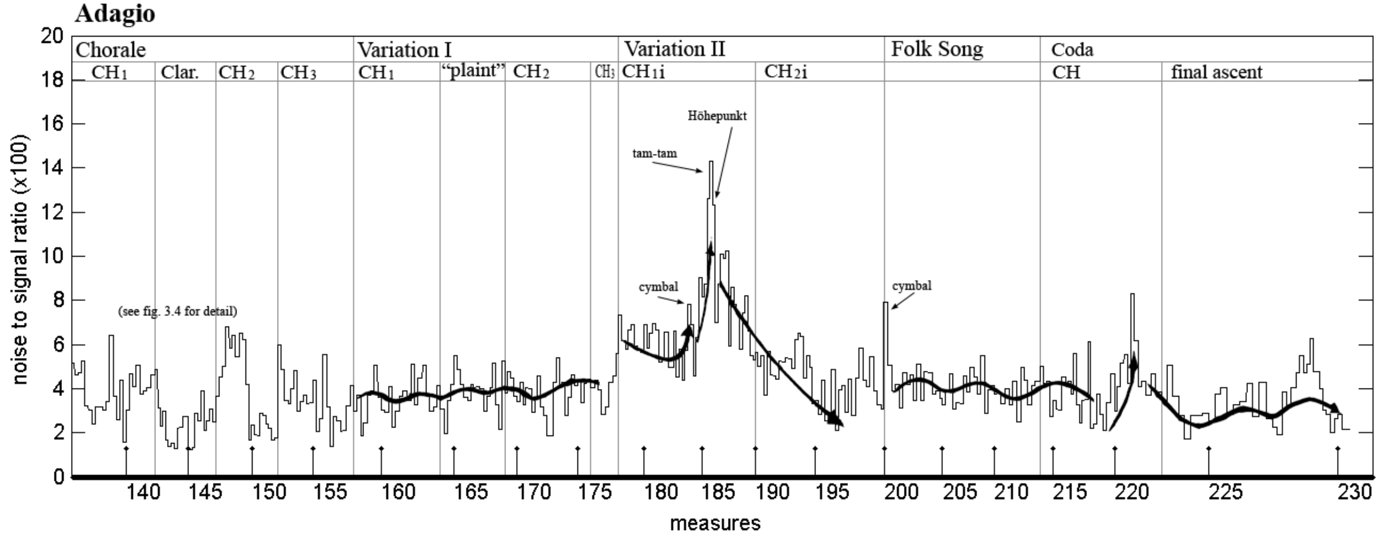


Appendix A, figure A.3. Berg, movement II, noise to signal ratio

Allegro

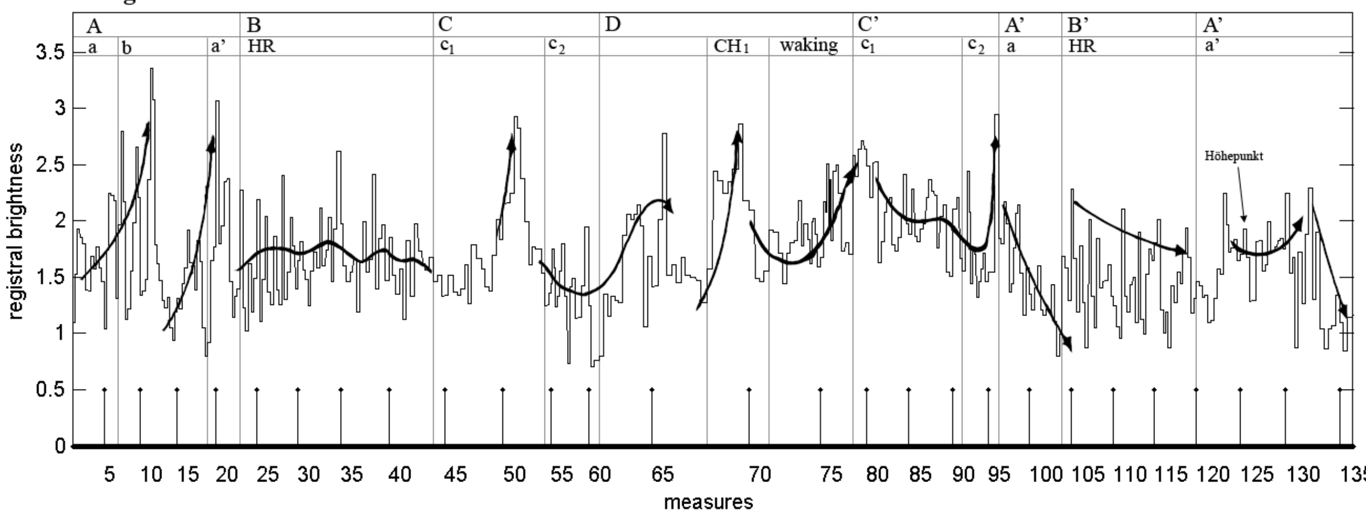


Adagio

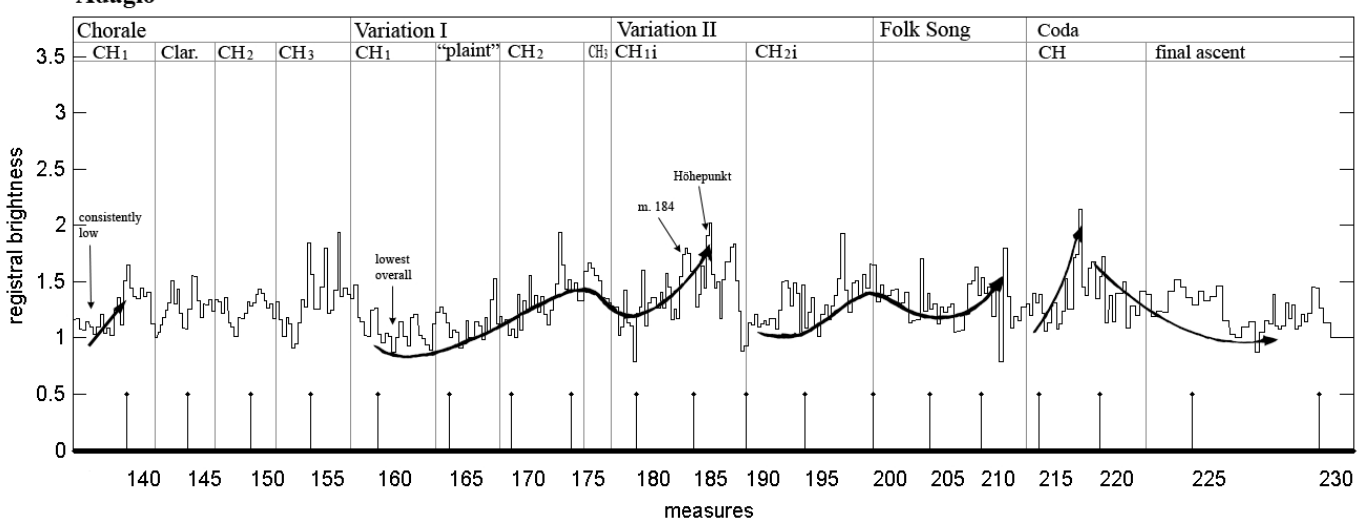


Appendix A, figure A.4. Berg, movement II, registral brightness

Allegro

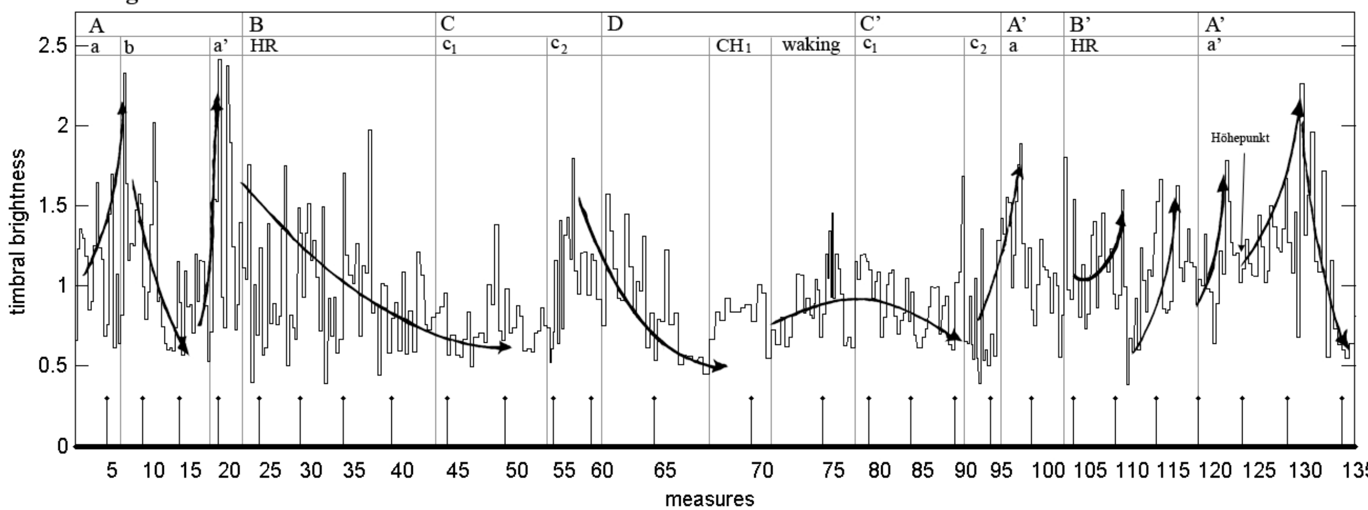


Adagio

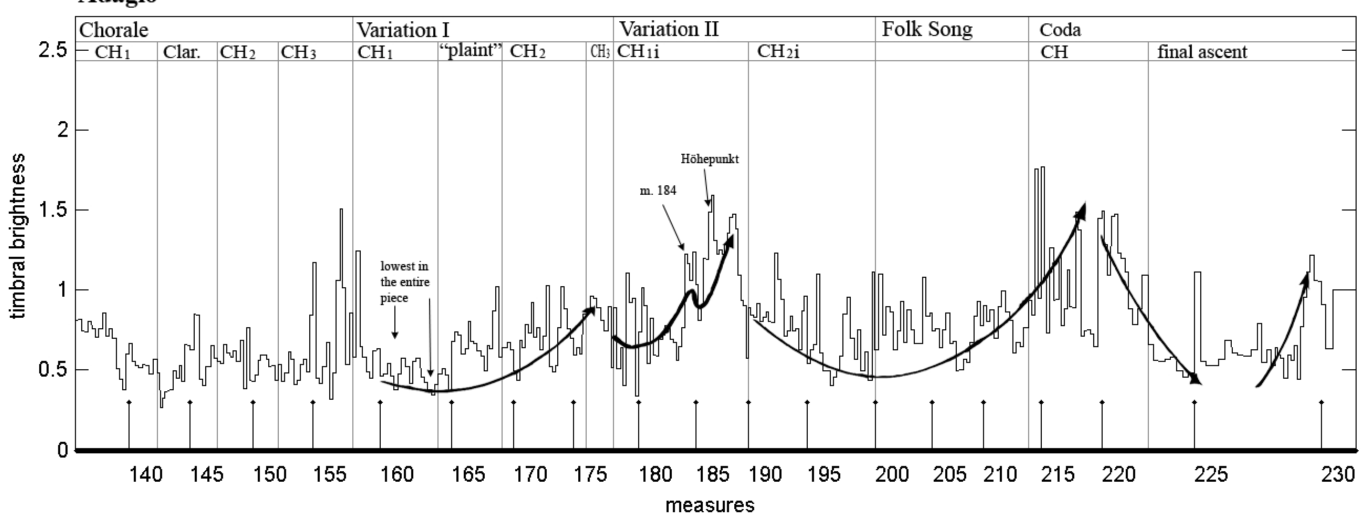


Appendix A, figure A.5. Berg, movement II, timbral brightness

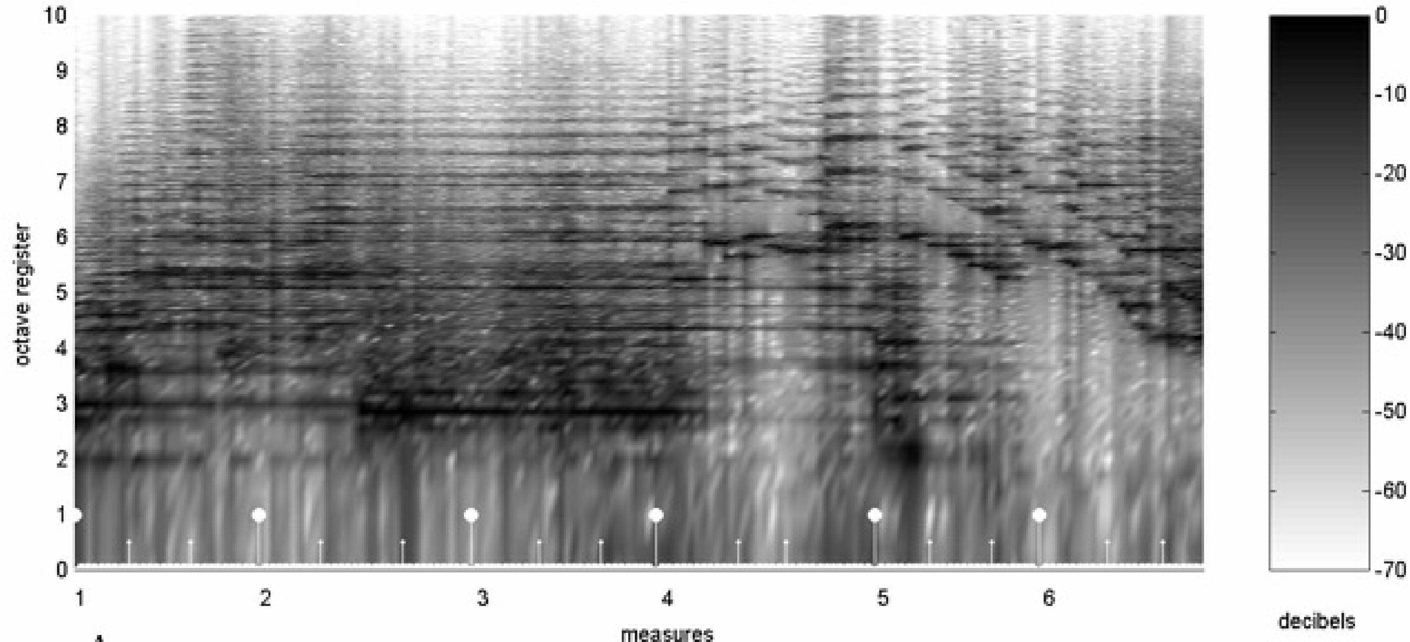
Allegro



Adagio

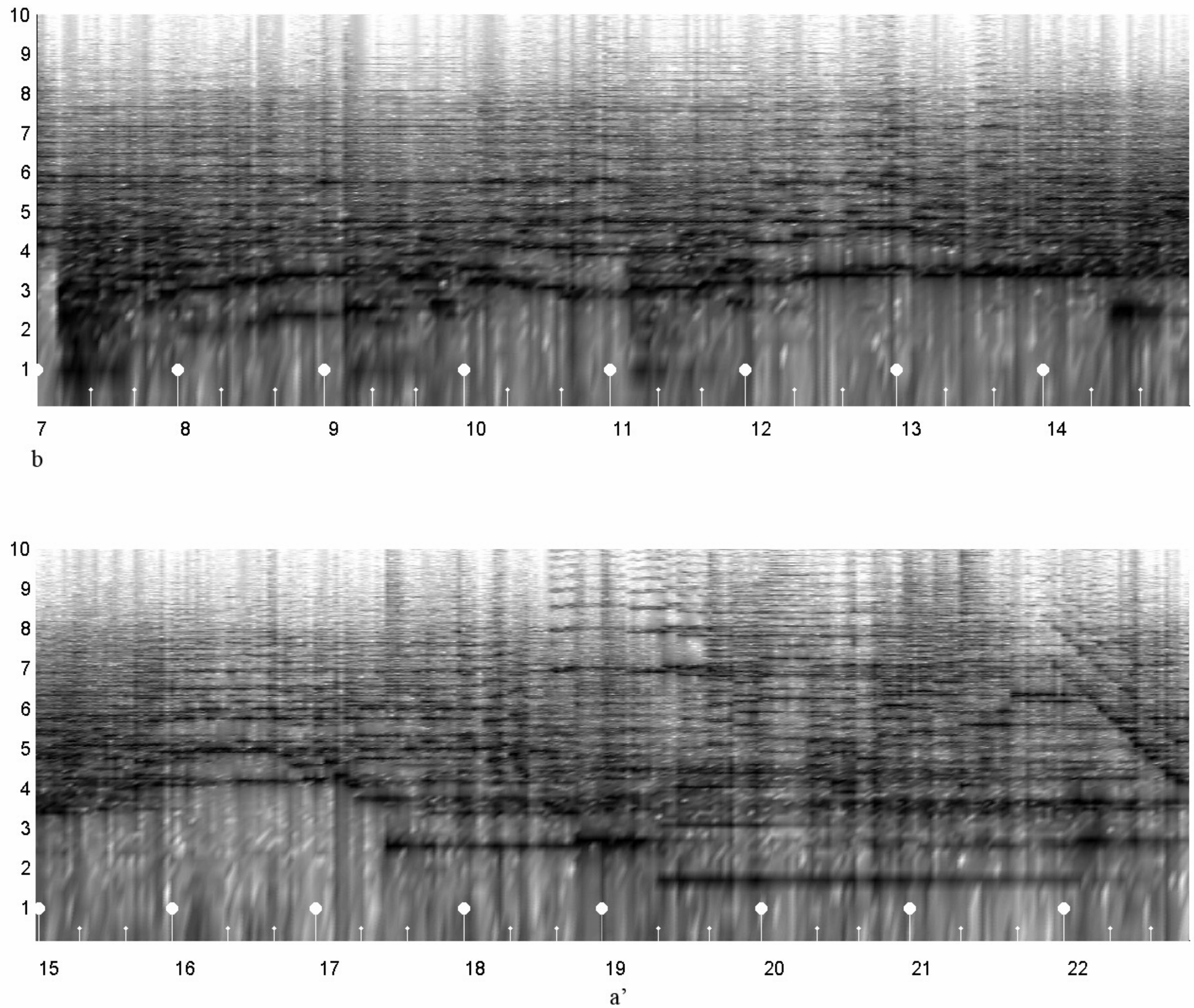


Appendix B, figure B.1. Spectrogram of ms. 1-6

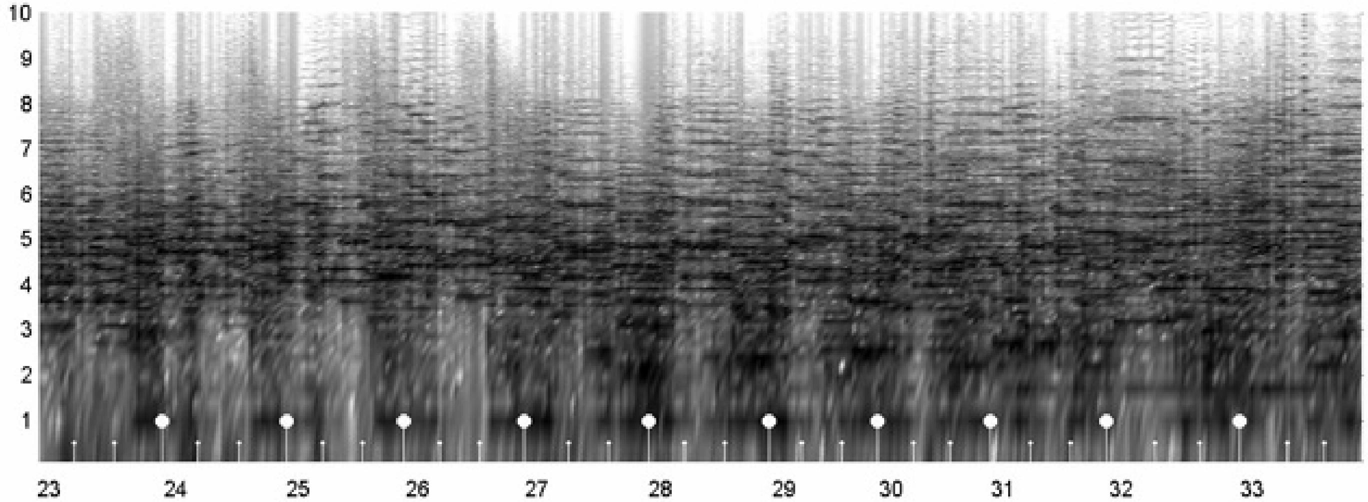


section/ A
thematic material a

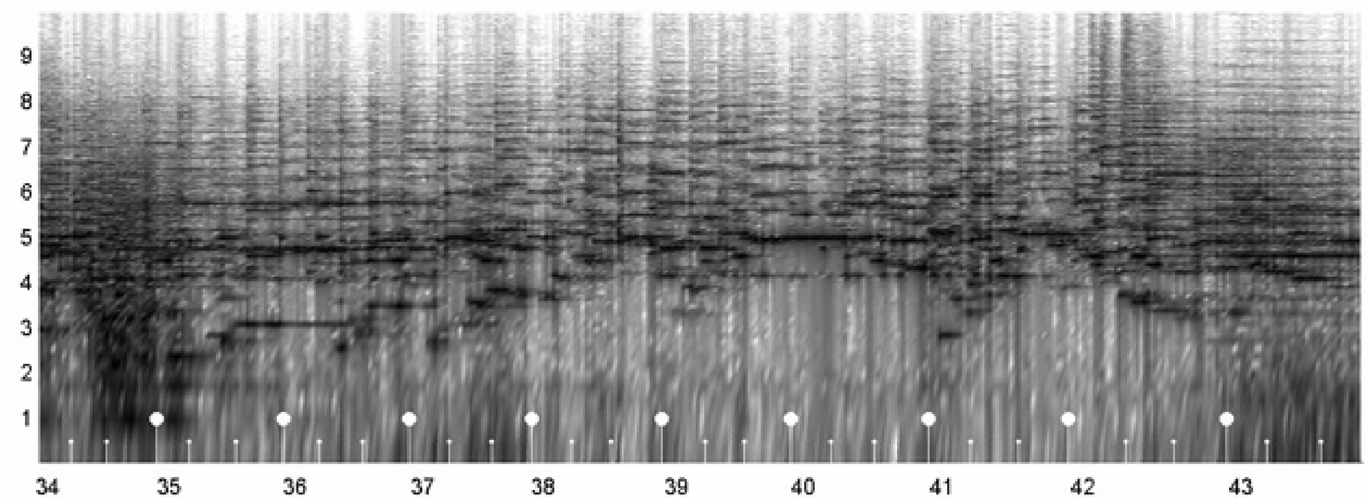
Appendix B, figure B.2. Spectrogram of ms. 7-22



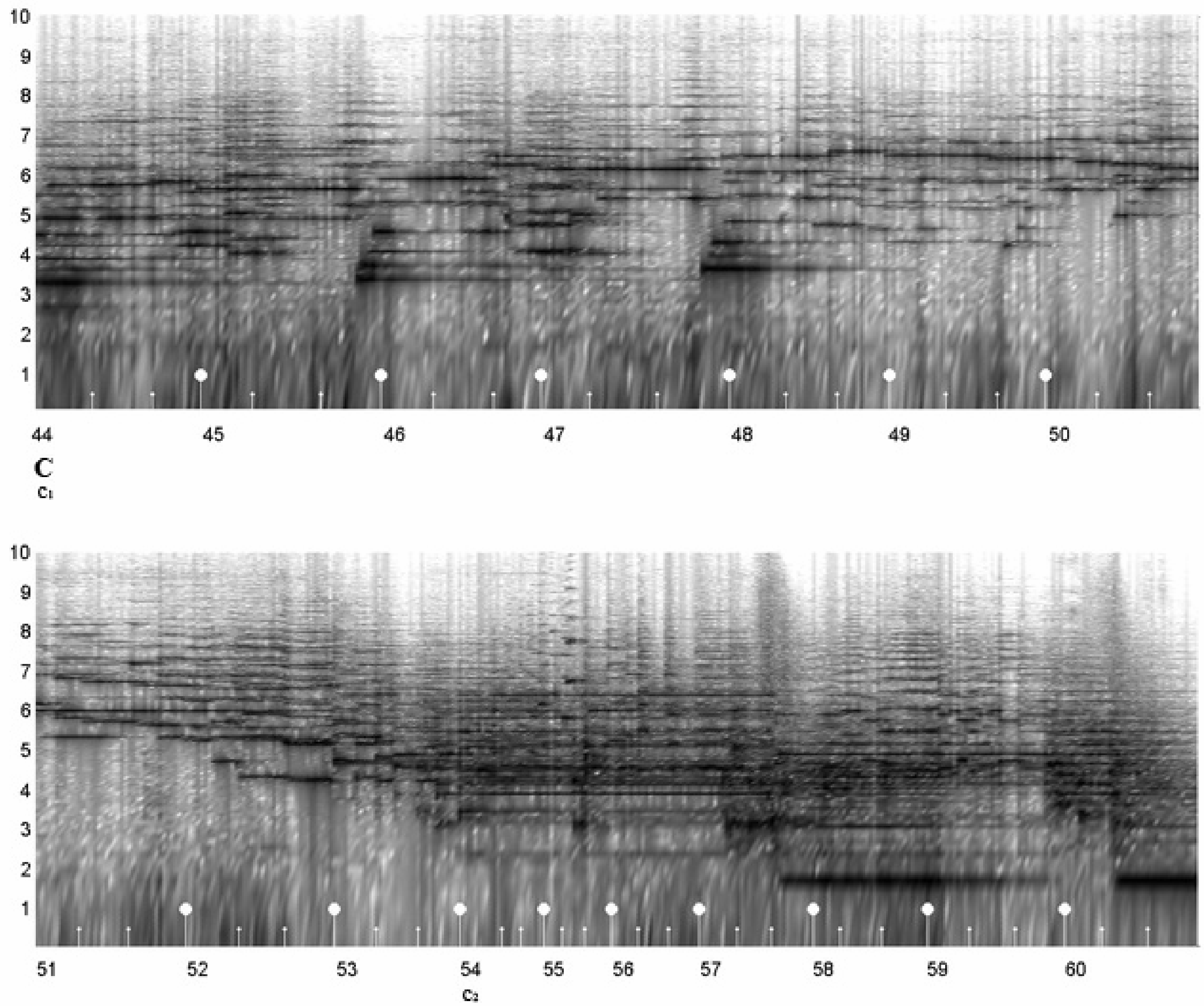
Appendix B, figure B.3. Spectrogram of ms. 23-43



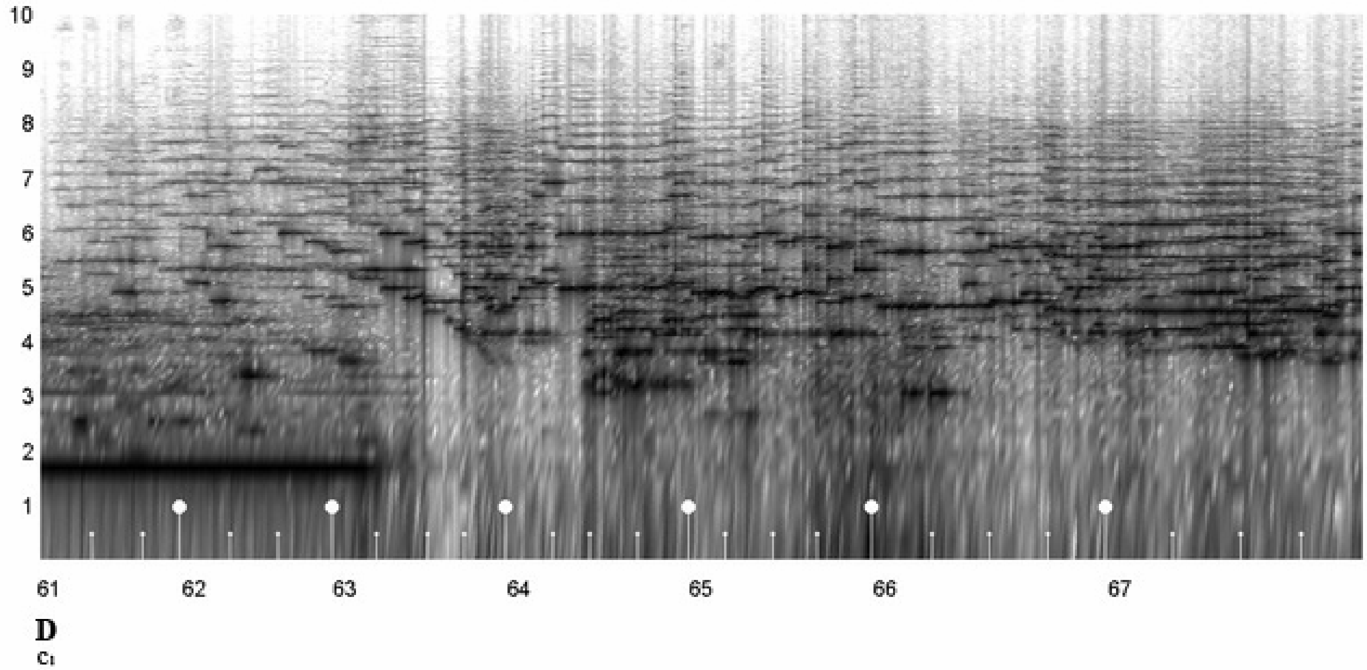
B
HR



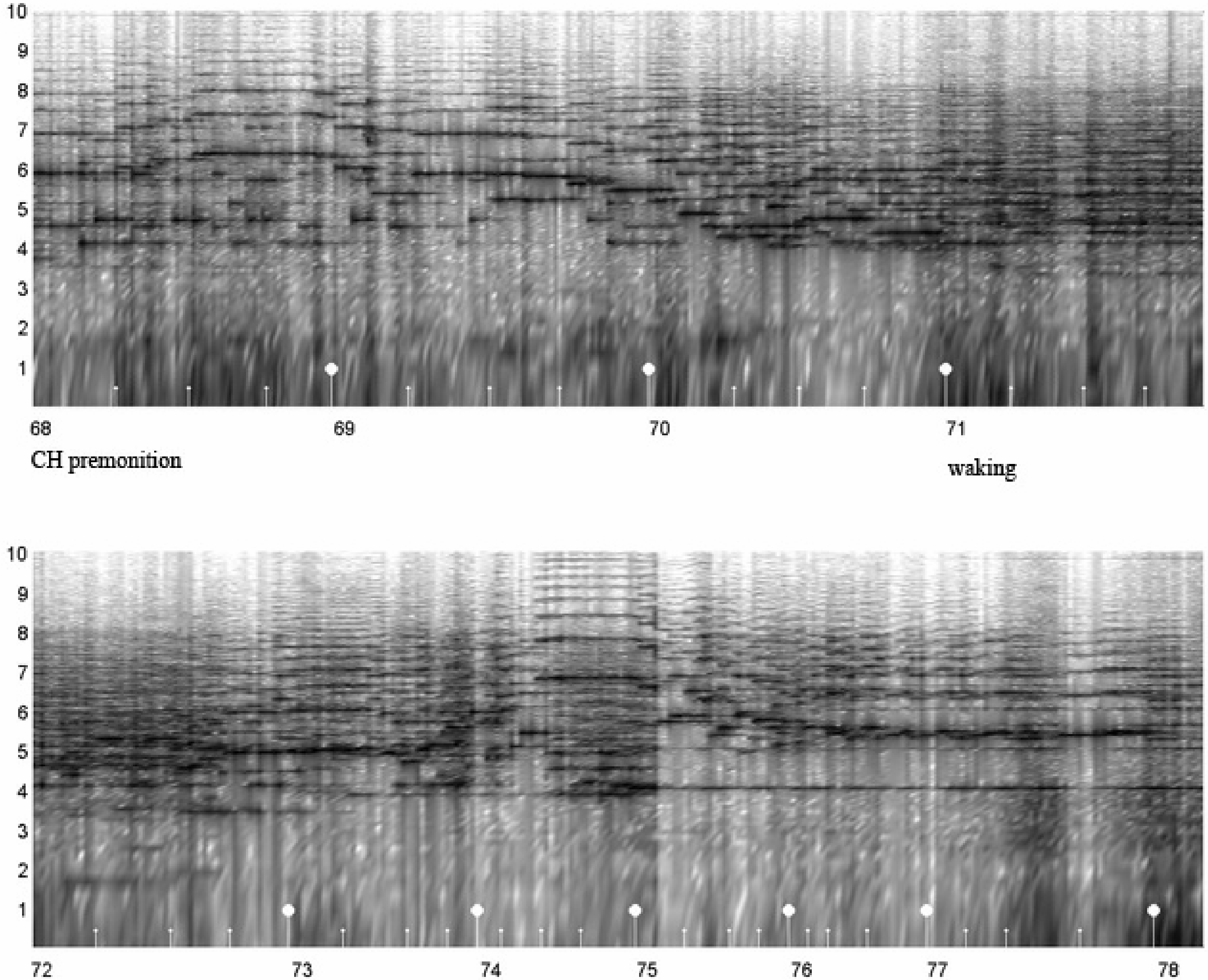
Appendix B, figure B.4. Spectrogram of ms. 44-60



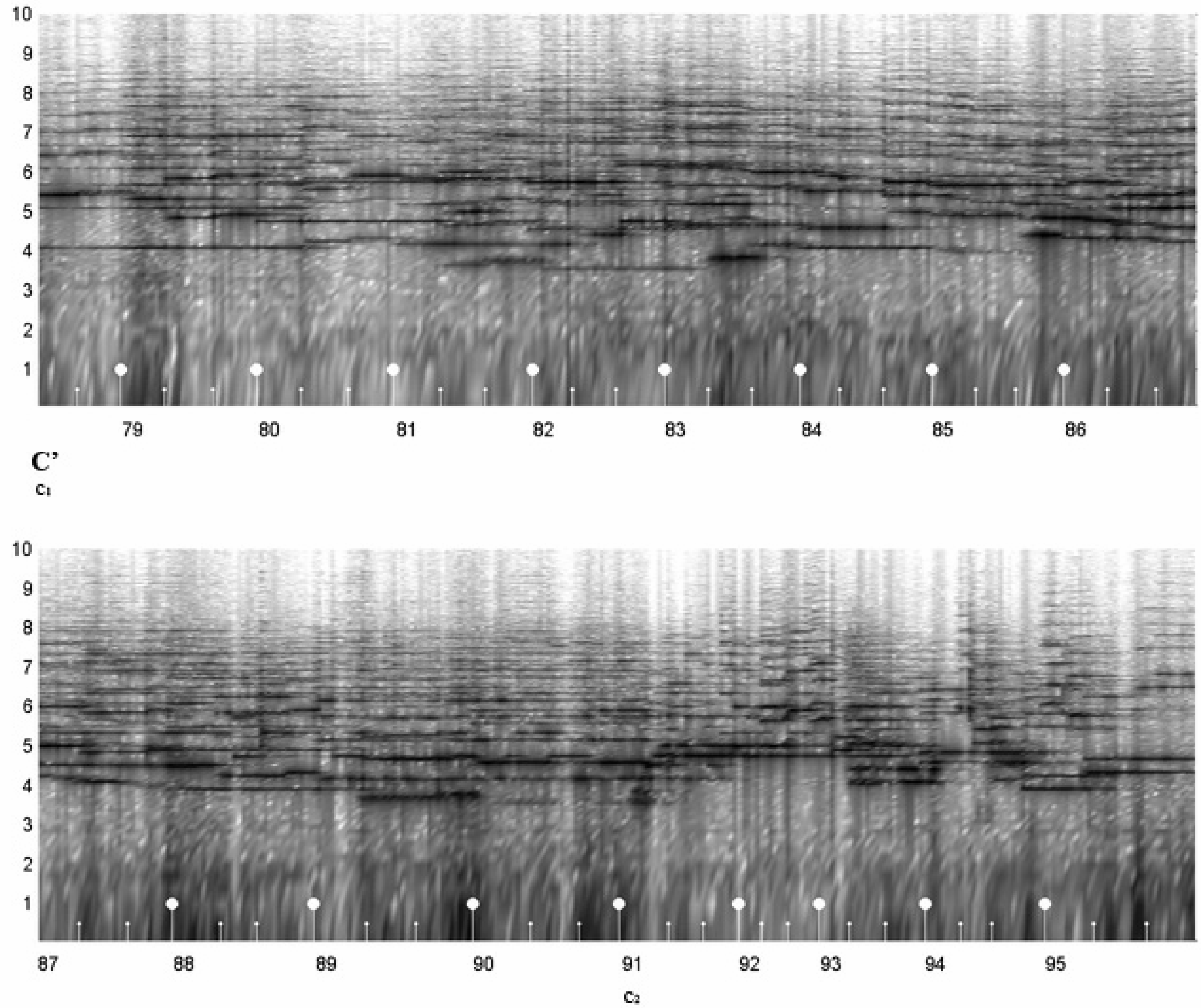
Appendix B, figure B.5. Spectrogram of ms. 61-67



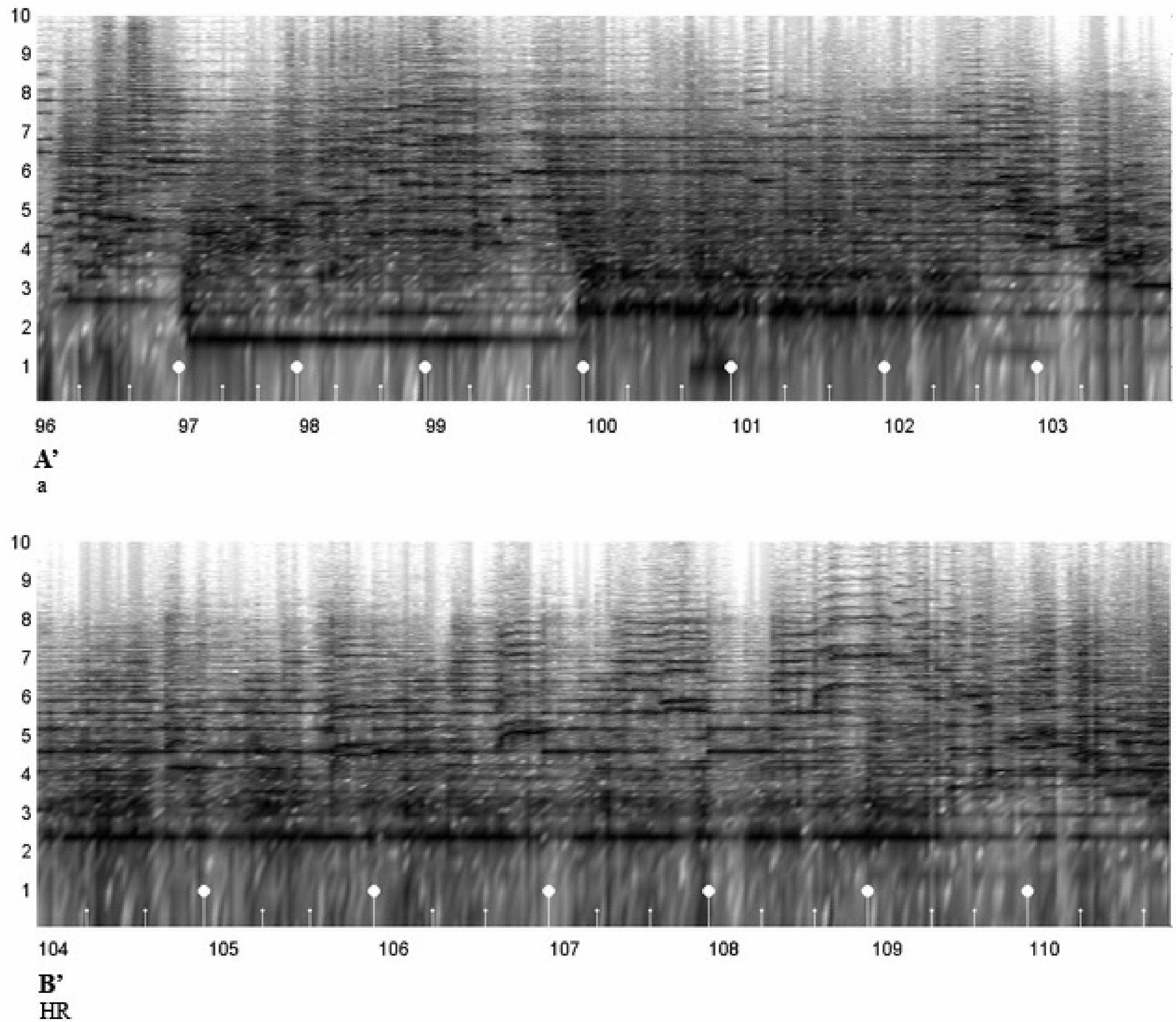
Appendix B, figure B.6. Spectrogram of ms. 68-78



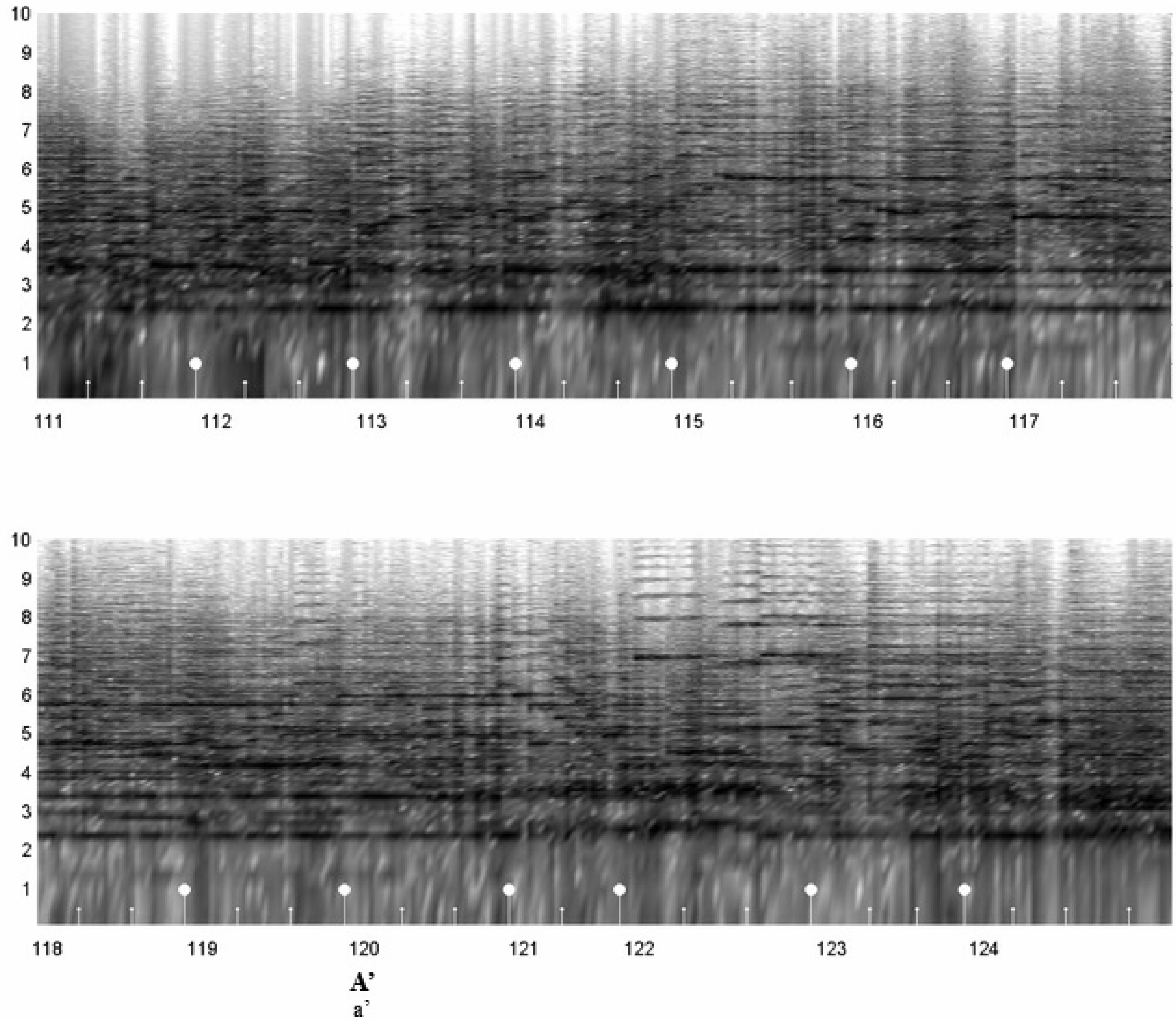
Appendix B, figure B.7. Spectrogram of ms. 79-95



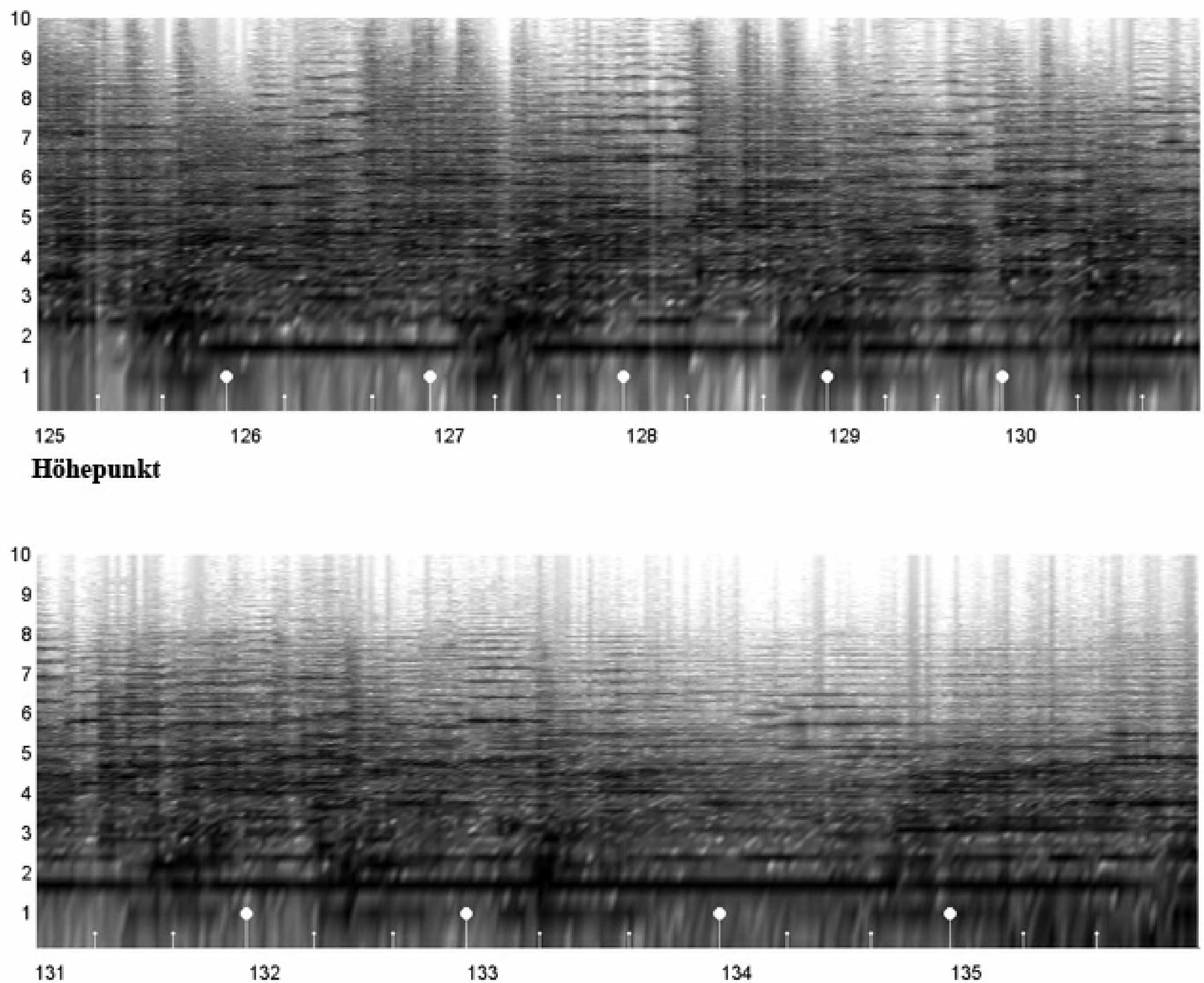
Appendix B, figure B.8. Spectrogram of ms. 96-110



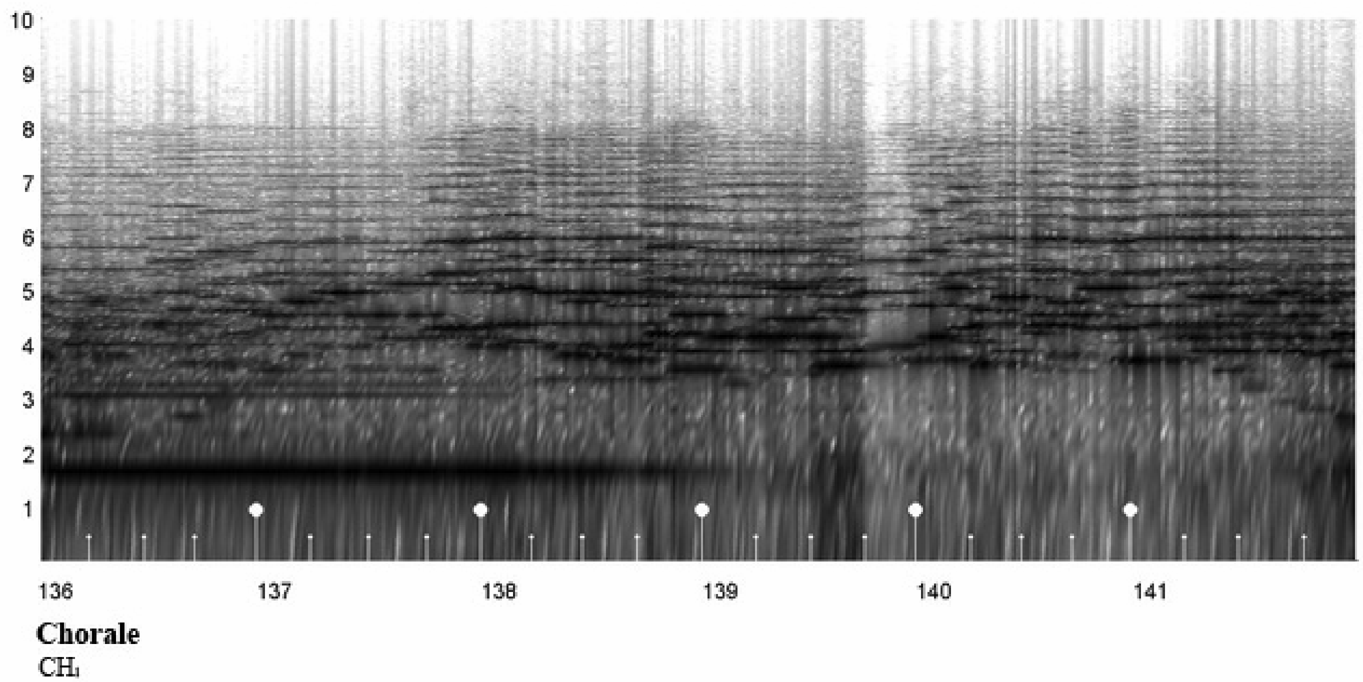
Appendix B, figure B.9. Spectrogram of ms. 111-124



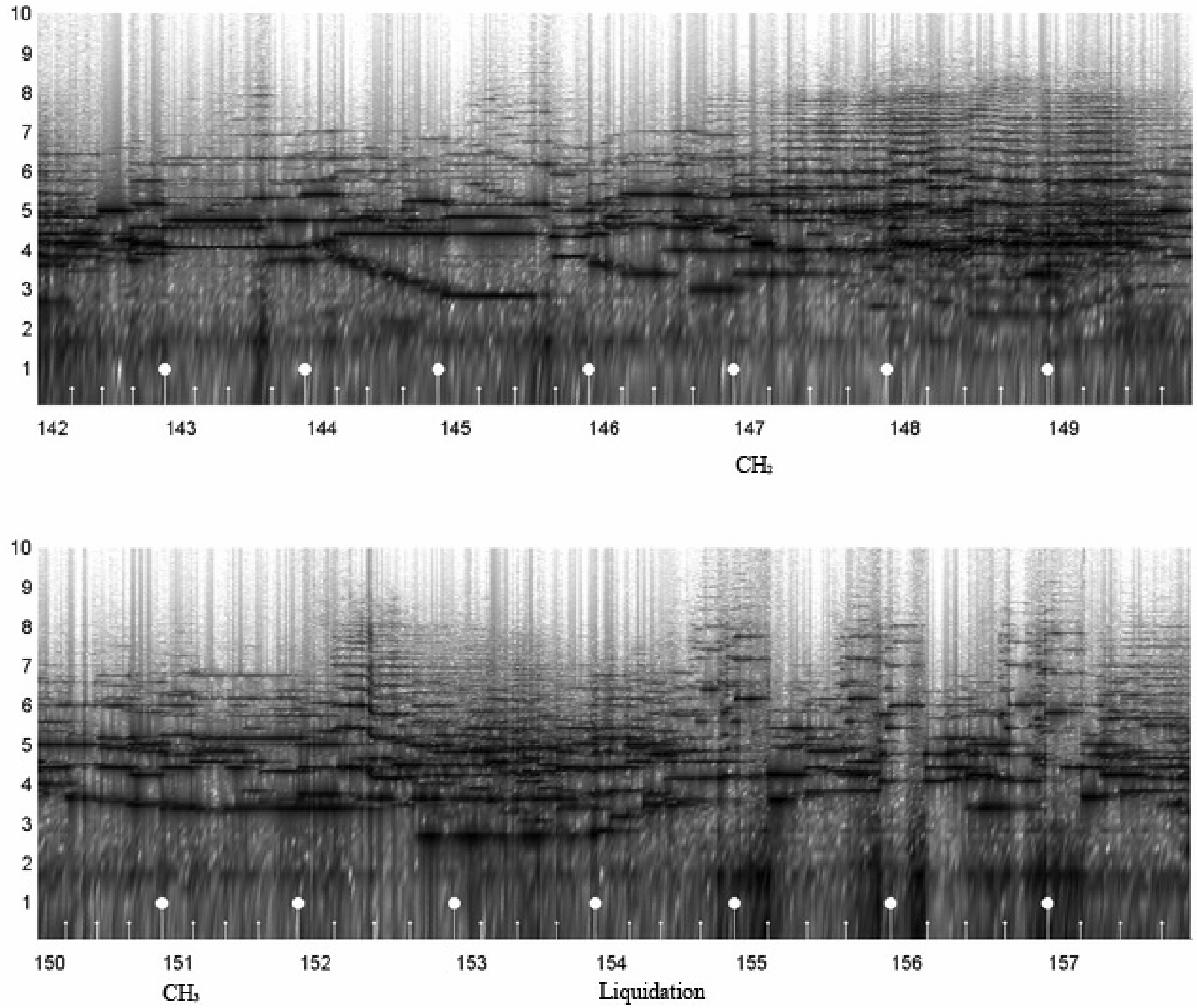
Appendix B, figure B.10. Spectrogram of ms. 125-135



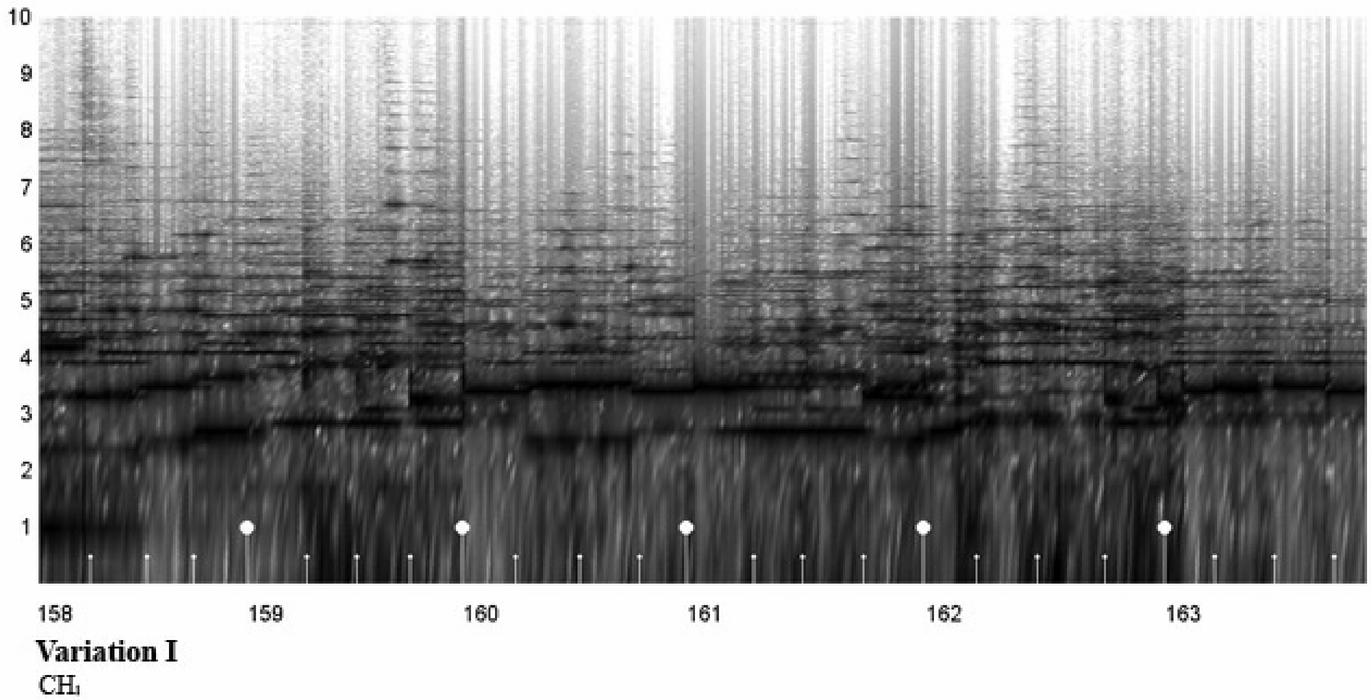
Appendix B, figure B.11. Spectrogram of ms. 136-141



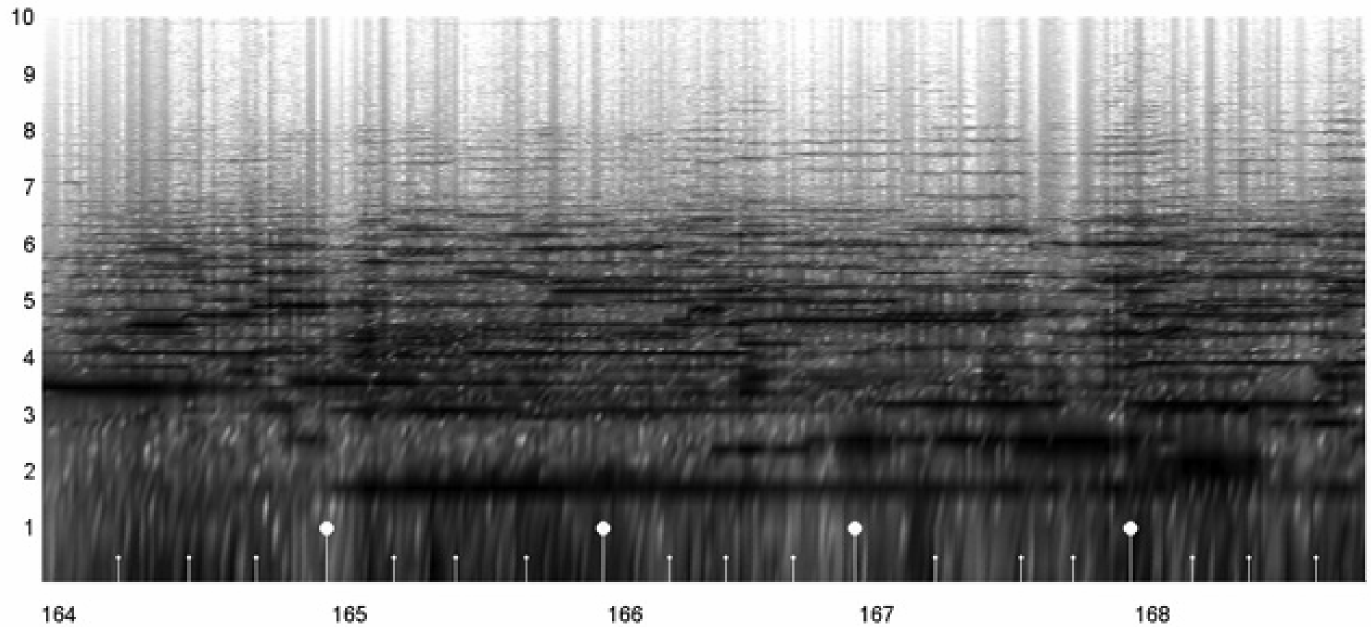
Appendix B, figure B.12. Spectrogram of ms. 142-157



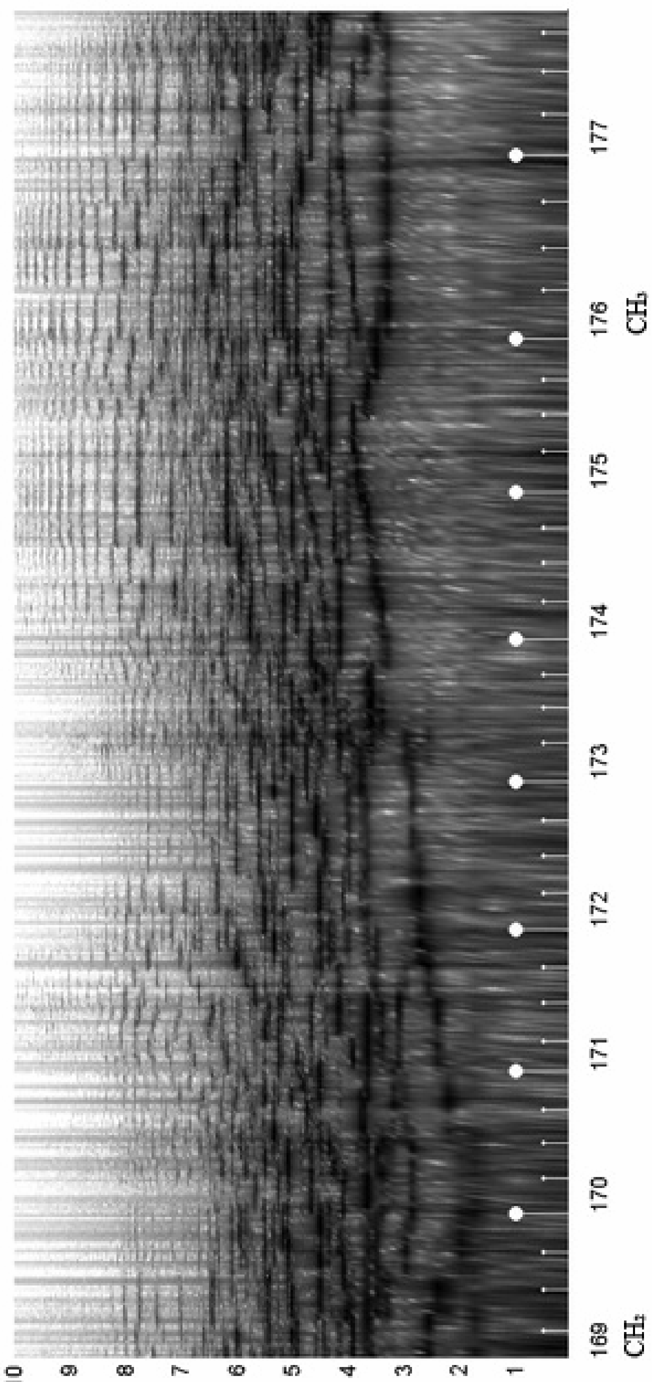
Appendix B, figure B.13. Spectrogram of ms. 158-163



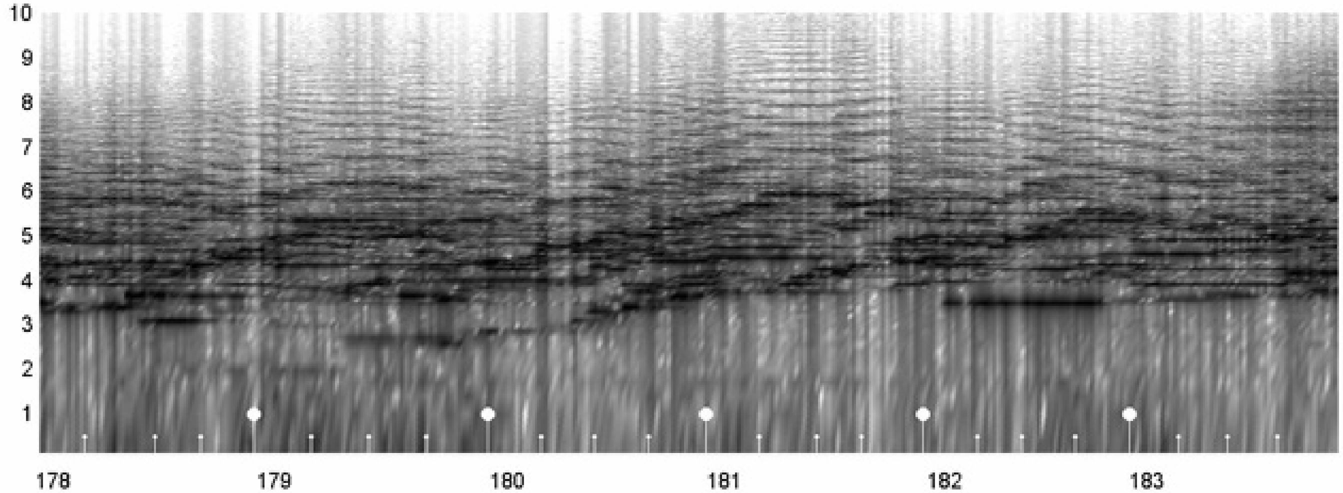
Appendix B, figure B.14. Spectrogram of ms. 164-168



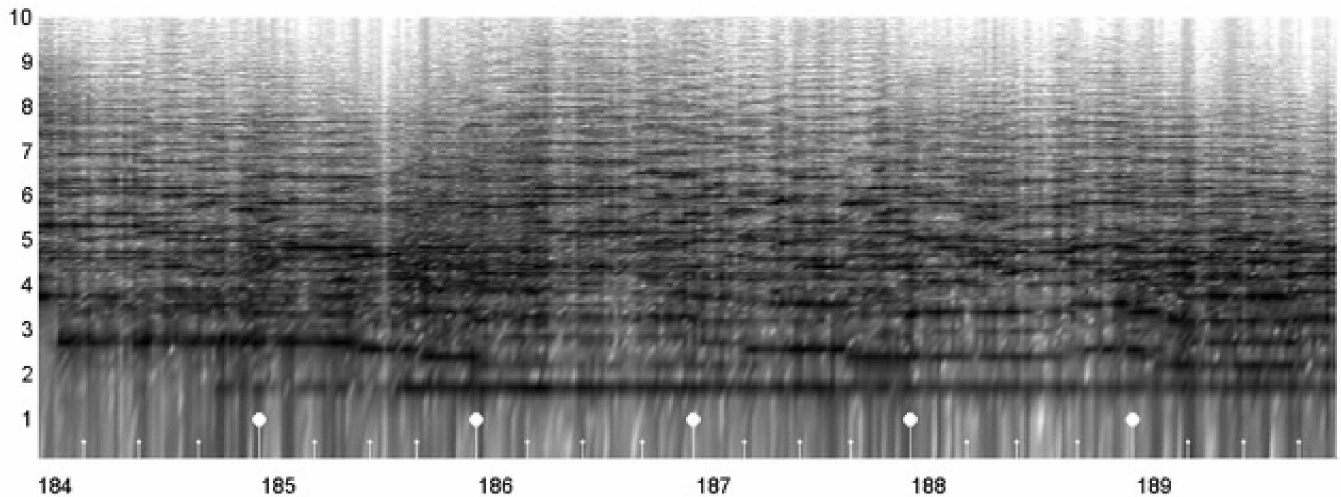
Appendix B, figure B.15. Spectrogram of ms. 169-177



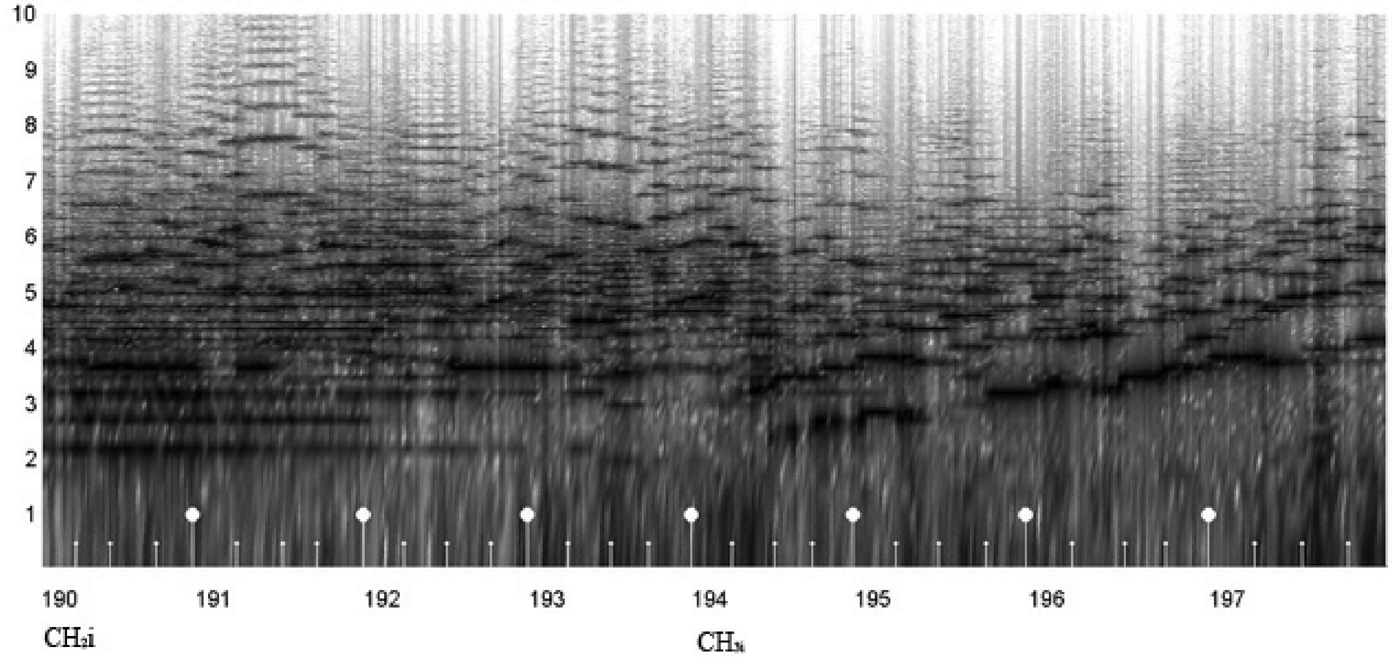
Appendix B, figure B.16. Spectrogram of ms. 178-189



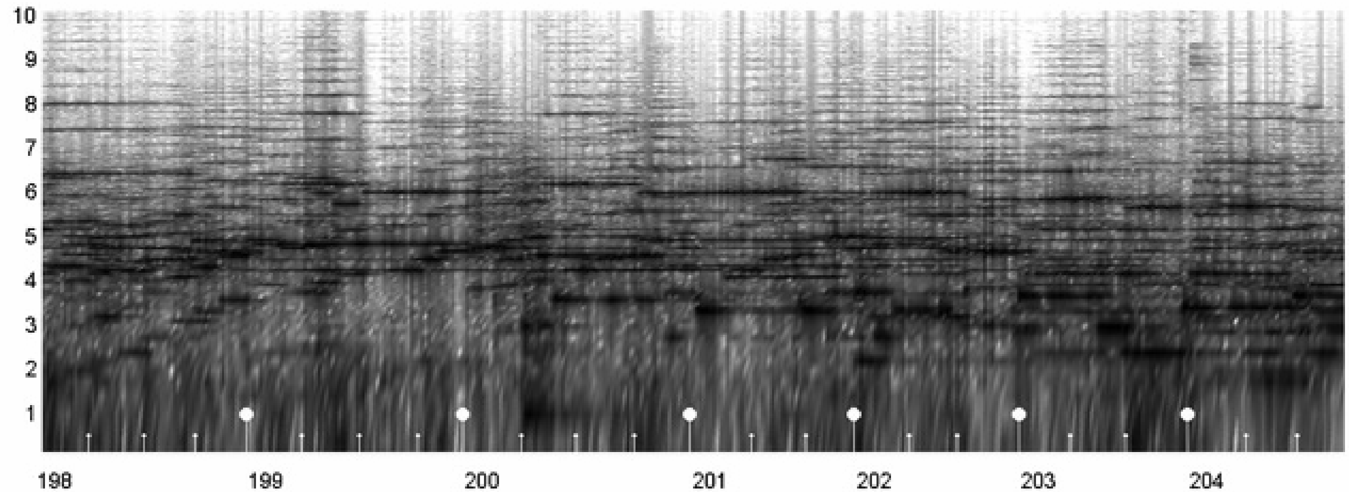
Variation II
CH$_{1}$i



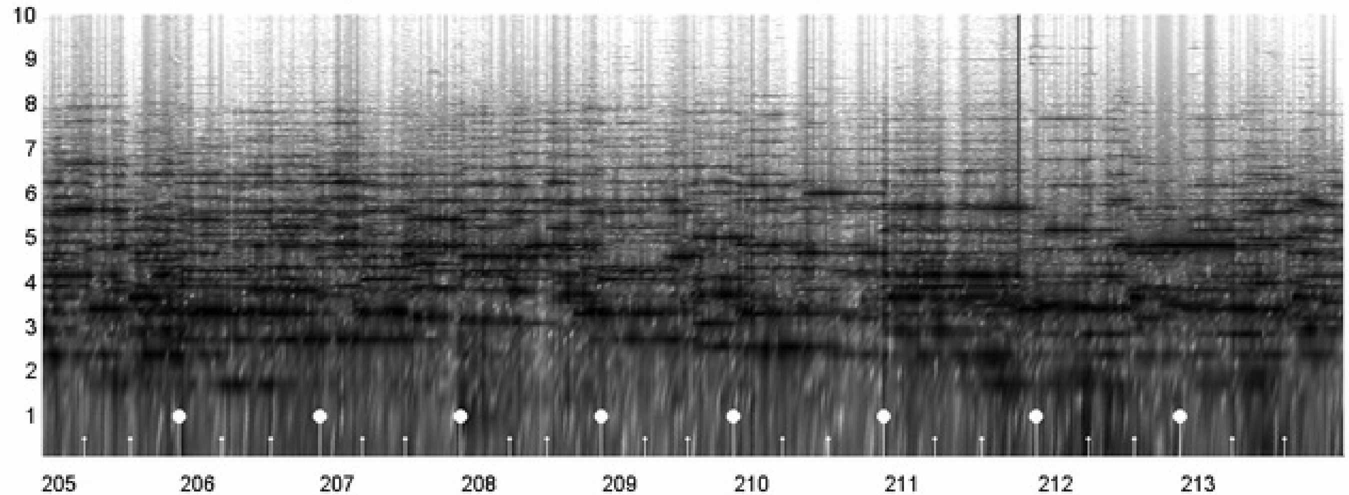
Appendix B, figure B.17. Spectrogram of ms. 190-197



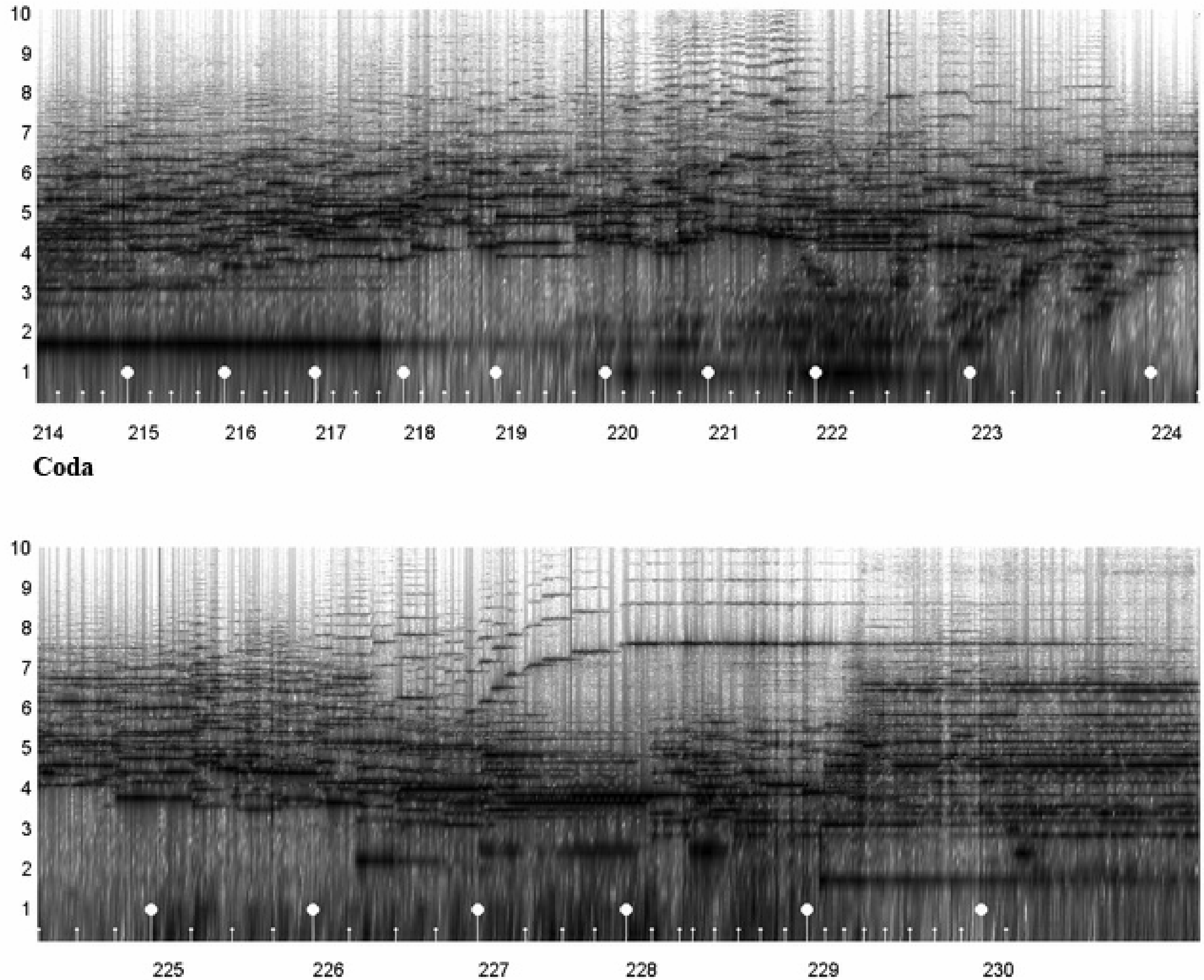
Appendix B, figure B.18. Spectrogram of ms. 198-213



Folk Song



Appendix B, figure B.19. Spectrogram of ms. 214-230



WORKS CITED

Adorno, Theodor W. 1991. *In search of Wagner*. Translated by Rodney Livingstone. London and New York: Verso.

Berg, Alban. 1992. *Violinkonzert.* Anne-Sophie Mutter, violin. Chicago Symphony Orchestra. Cond. James Levine. Hamburg: Deutsche Grammophon, 437 093-2.

________. 1996. *Violinkonzert.* Edited by Douglas Jarman. W.Ph.V. 537. Wien: Universal Edition.

Cogan, Robert, and Pozzi Escot. 1976. *Sonic design: the nature of sound and music.* Englewood Cliffs: Prentice-Hall.

Cogan, Robert. 1998. *New images of musical sound*. Cambridge: Publication Contact International. Original edition, Cambridge: Harvard University Press, 1984.

Cramer, Alfred. 2002. Schoenberg's Klangfarbenmelodie: A principle of early atonal harmony. *Music Theory Spectrum* 24, no. 1 (spring): 1-34.

Grey, John M., and John W. Gordon. 1978. Perceptual effects of spectral modifications on musical timbres. *J. Acoust. Soc. Am.* 63, no. 5 (May): 1493-1500.

Griffiths, Paul. 1995. *Modern music and after.* Oxford and New York: Oxford University Press.

Hall, Donald E. 1991. *Musical acoustics.* 2nd ed. Pacific Grove: Brooks/Cole Pub. Co.

Helmholtz, Hermann von. 1954. *On the sensations of tone as a physiological basis for the theory of music.* Edited and translated by Alexander John Ellis. New York: Dover Publications.

Hind, John. 1959. Review of Broadcast Music. *The Musical Times* 100, no. 1394 (April): 207.

Morgan, Robert P. 1991. *Twentieth-century music: a history of musical style in modern Europe and America.* New York: Norton.

Perle, George. 1977. The secret program of the Lyric suite. *International Alban Berg Society Newsletter,* no. 5: 4-12.

Pople, Anyhony. 1991. *Berg, Violin concerto***.** Cambridge and New York: Cambridge University Press.

Review of London Concerts. 1937. *The Musical Times* 78, no. 1127 (January): 67-71.

Roederer, Juan G. 1995. *The physics and psychophysics of music: an introduction.* 3rd ed. New York: Springer-Verlag.

Schoenberg, Arnold. 1978. *Theory of harmony.* Translated by Roy E. Carter. Berkeley: University of California Press.

Sethares, William A. 1998. *Tuning, timbre, spectrum, scale.* London and New York: Springer.