Chapter 2
Structure of Musical Sound

2.1 Introducing the Model

Every single tone which reaches our ear in the course of a musical work, contains a fullness of information. We perceive a pitch, loudness and tone color. We can also make statements about the steadiness of the pitch, or if the tone is enlivened with vibrato. Furthermore we notice changes and fluctuations in loudness as well as the nature of the tone entrance, be it attacked softly or sharply; similarly we can draw conclusions about the decay of the note. All these details give a characteristic tone picture from which we extract the musical content and also recognize what instrument generated the tone. In this, previously gathered listening experiences, play a not unessential role. Finally we can even draw conclusions about the nature and size of the room in which the music resounds.

This acoustic phenomenon can be described by a number of physical factors. In all this there is a certain difficulty posed by the circumstance that the individual characteristics perceived are not each determined by only a single physical quantity, but come about by the cooperation of several components. Nevertheless, the problem of finding an association between objective acoustical data and tonal impressions on the basis of practical experiences should not be overestimated.

A model, as schematically presented in Fig. 2.1, shall serve to clarify the complex sound processes which correspond to a single note. It proceeds from the concept that a tone is a vibrational process which changes in time, which is characterized by the number and strength of vibrations. Accordingly, the model contains a time axis (in the picture it runs from front to back), a frequency axis (from left to right) as well as an axis for sound pressure level (upwards), which is the quantity responsible for the loudness impression (Winckel, 1960).

Additionally the shape of the vibration is of great significance: Most often the sound does not consist of a simple sinusoidal oscillation, as known from the motion of a pendulum, but rather it exhibits a complicated time sequence. Thus the ear is placed in the position of needing to differentiate between tones of different timbre for the same pitch. However, such a complicated vibrational form can be considered as a superposition of a series of sinusoidal vibrations with different frequencies.

J. Meyer, *Acoustics and the Performance of Music*.
DOI: 10.1007/978-0-387-09517-2_2, © Springer Science + Business Media, LLC 2009

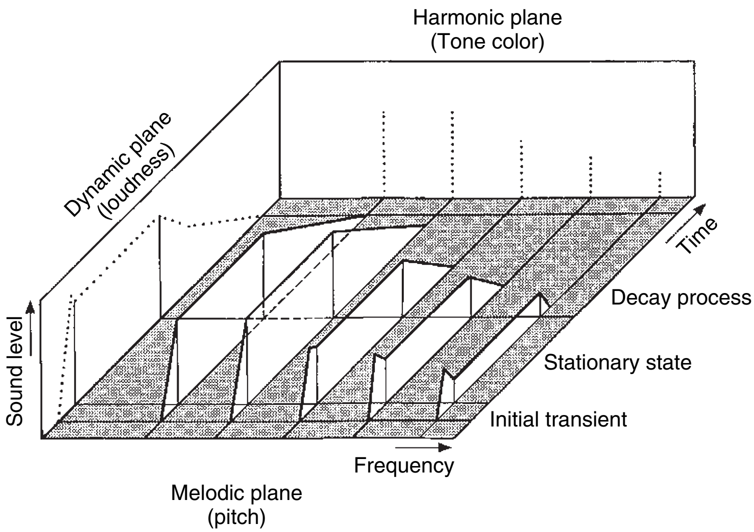

Fig. 2.1 Three dimensional representation of a single tone in schematic form

The entire vibration process, therefore, appears in our model as a sequence of several partial vibrations or overtones.

From a time standpoint, each tone can be structured into three sections:

1. The starting transient, i.e., that portion of time during which the tone is developed from complete rest to its final state.
2. The stationary condition, i.e., that portion of time during which the tone is practically not subjected to change.
3. The decay, i.e., that portion of time during which the tone, after completion of the excitation, dies out to complete silence.

The starting transient, in particular measure, contains characteristic features, which make it possible to distinguish between instruments. Often components are present which are no longer contained in the later tone picture.

A stationary condition is, strictly speaking, reached only for instruments with very uniform vibration excitation. These are above all those instruments which do not use the energy of the player but an external (constant in time) energy source, as for example the organ or many electronic instruments. Even the minimal air pressure variations of a wind instrument player, or the bow pressure changes of a string player lead to minute variations of the vibrational process. This is also the reason why one speaks of a quasistationary state in such cases. The most important features can, however, be describes as though it were a stationary state.

The decay process plays a special role for plucked- and percussion instruments, since in the absence of continuing excitation there is no stationary or quasistationary state. Yet, also for other instruments it has a certain significance for convincing connections between tones in flowing passages.

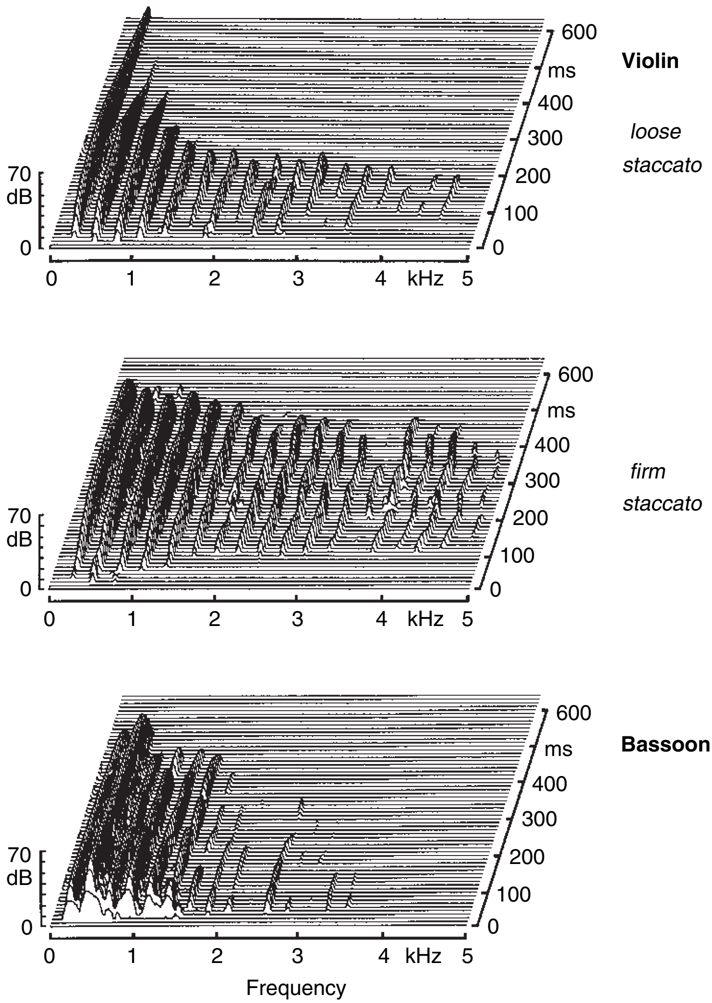

Fig. 2.2 Observed measurements of tone spectrum developments in time for three (C_4) *staccato* tones

Figure 2.2 shows the transition from abstract model presentation to real tones. It contains three dimensional sound spectra as measurement results of the analysis of three staccato tones with equal pitch, and approximately equal duration, however, of different tonal character. Differing numbers of partial vibrations are recognized. Furthermore, in the partial picture for the firm violin staccato, a vibrato becomes apparent in the form of a slight wave motion of the individual partials. The tone onset for example, occurs most rapidly in the bassoon, and is largely simultaneous for all partials, while the partials in the violin enter in part only with delay. Likewise a differential behavior is noted for the end of the tones, the long ringing of the loose violin staccato is particularly noticeable. Already these few suggestions make it appear rewarding to pursue a deeper consideration of such methods of analysis to gain detailed information about tonal characteristics of musical instruments. Aside from that, the examples shown in this picture show how difficult it is to compare the durations of tones of differing tonal characteristics.

For the sake of clarity it is often recommended to simplify the three dimensional representation of the tone by sacrificing one of the three quantities and projecting the model onto one of the planes represented in Fig. 2.1. Relinquishing the frequency dependence results in a time dependent sound pressure level graph, as indicated by the dotted line on the left sidewall of the model. Inasmuch as this concerns the time flow of the quantity most responsible for the loudness sensation, this plane is frequently referred to as the dynamic plane. If the tone model is projected onto the base plane with its frequency and time axis, one obtains the temporal pitch flow, including possible fluctuations by such things as vibrato; this plane is called the melodic plane.

A projection into the third plane with the frequency and sound level axes is equivalent to a cross section through the stationary part of the vibration process, more exactly a median value. For this viewpoint individual partials of the complex sound are represented with their respective frequency and strength, as the dotted lines on the back wall of the model show. Their height gives the sound pressure, or sound pressure level, their foot point on the horizontal axis indicates their frequency. Since this type of representation also makes it possible to visualize the frequency composition of cords, one speaks of the harmonic plane of the model. In analogy to the optical decomposition of light into individual portions of various primary colors, the representation of individual frequency components with their strengths is designated as a sound spectrum. Above all, this gives information about the tone color in the stationary part of the tone and in addition permits conclusions about the starting transient and decay behavior.

2.2 Frequency- and Level: Structures

2.2.1 The Harmonic Tone Structure of Sound Spectra

For periodic vibration processes, as they are present in the stationary and also quasistationary portion of sounds of almost all musical instruments, the individual vibration contributions form a so called harmonic series in reference to their frequencies, i.e., starting with the lowest frequency, partial tones are present whose frequencies are integral multiples of the fundamental frequency. As an example for such a partial tone sequence the note $C_{2,}$ as sounded on a contrabassoon, is broken down in Fig. 2.3 into its first 16 harmonic partials. The note picture shows (without octave transposition) the location of the individual partials within the musical scale, below each note the relevant frequencies are indicated, the corresponding harmonic index is shown above the note picture showing that they form the 2-, 3-, 4-, etc fold frequency of the fundamental.

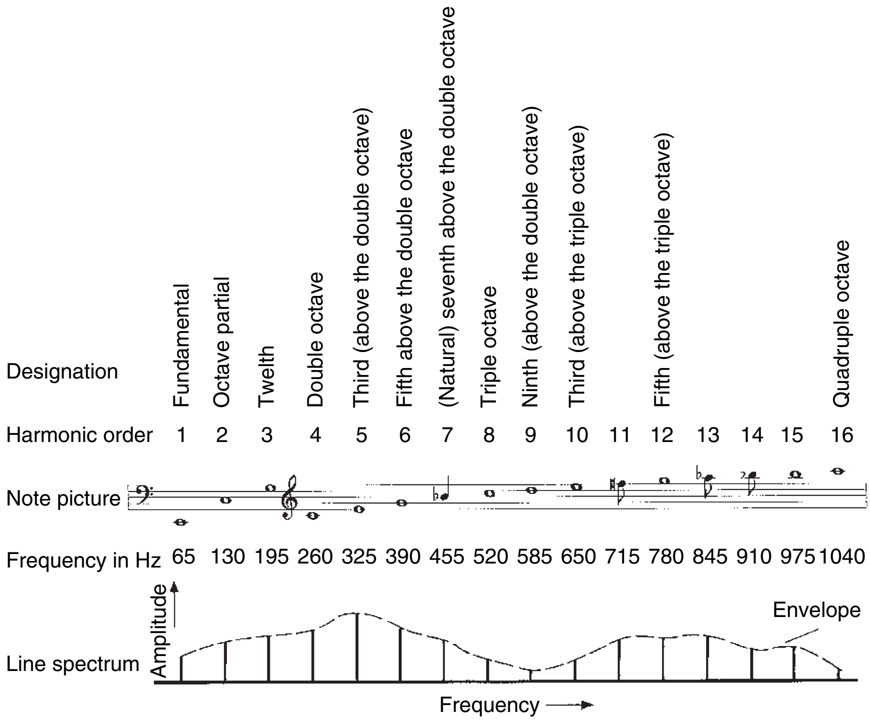

Fig. 2.3 Harmonic structure of the note C_2, represented schematically

This fundamental is principally responsible for the perceived pitch. In the example shown it is located at 65 Hz. A doubling of the number of vibrations corresponds to exactly an octave, so that the 2nd, 4th, 8th, and 16th partials again result in a C (or C_3, C_4, C_5, and C_6 respectively). Tripling the frequency leads to the fifth above the octave (G_3), correspondingly the 6th and 12th harmonics form fifths in higher positions. The 5th and the 10th partials are identified as notes E_3 and E_4 which form major thirds in relation to the corresponding C's, the 9th an 15th partials form the fifth and the major third respectively to the higher octaves of the 3rd partial, i.e., G_4–D_5 and G_5–B_5. All of these indicated intervals should be considered as "perfect" intervals.

Accommodating, however, the 7th, 11th, 13th, and 14th partials, which lie somewhere in between, proves more difficult, since their frequencies do not coincide with those of notes on a normal scale. Thus the 7th and 14th partials lie lower than B_4^b and B_5^b, they, along with the fundamental, form the basis for the interval of a so called natural seventh, which, however, is by no means sensed as dissonant in the context of an entire partial series since it is included in the full sound without beats. A similar situation exists for the 11th partial which lies below the $F_5^{\#}$, as well as for the 13th partial, which is just a little higher than A_5^b.

Clearly, sounds can consist of more than 16 partials, in the octave from the 16th to the 32nd partial, additional 15 intermediate values are found, which are increasingly crowded together within the scale. While a trained ear can still detect the lowest 6–8 partials individually from within a sound, since they fall within different critical bands (see Sect. 1.2.1), the higher harmonics melt together into a ton color impression, even for the experienced listener. In all this, the intensity relations between individual partials naturally play a significant role.

The amplitudes or sound pressure levels of the individual partials can be represented as a line spectrum in a schematic manner, as is already suggested in the harmonic plane of Fig. 2.1, and also in the lower part of Fig. 2.3. Every line by position and length denotes the frequency and strength of the relevant partial. Connecting the endpoints of the spectral lines results in the so-called spectral envelope. This curve gives a clear representation of the amplitude distribution as a function of frequency without considering the harmonic number of the individual partial. The spectral envelope is therefore especially suited for summary representations.

2.2.2 The Frequency Range of Sound Spectra

The lower limit of the spectrum of partials is always determined by the fundamental, that is, the frequency which corresponds to the written representation (possibly under consideration of transposed notation). Still lower, stationary, i.e., steadily vibrating tonal components, do not exist. Below the fundamental, only noise-like, or nonsteady tone components are found. Consequently, the lower limit of the spectrum of partials moves upward with increasing pitch.

In order to give a general view of the systematics of spectral composition for a musical instrument, an entire chromatic sequence for a horn is given in Fig. 2.4. In this type of analysis, each individual partial appears as a peak of certain width, which is connected with the sharpness of the instrumentation used. It is clearly recognizable how the first peaks of the spectra move toward the right with increasing pitch, i.e., toward higher frequencies. Similarly the spacing between the individual partials increases. For the low notes, the high frequency partials are so closely spaced that they assume a noise-like character ("metallic") for sufficiently high intensity, while partials of higher notes in the same frequency region are still sensed as "pure sounds" by reason of their larger spacing.

The upper boundary of the spectrum of partials is very different for individual instruments, it also depends in large measure on dynamics. Room acoustical conditions also play a more important role than for the low frequencies. For this reason it is most advantageous to be able to refer to sound power spectra when describing tonal characteristics typical of a particular instrument. Sound power spectra summarize the entire sound radiated by an instrument. They are independent of the surrounding room, as well as of the distance and direction of an observation- or measurement point. They are best recorded in a special room with high grade sound reflecting walls, a so called reverberation chamber.

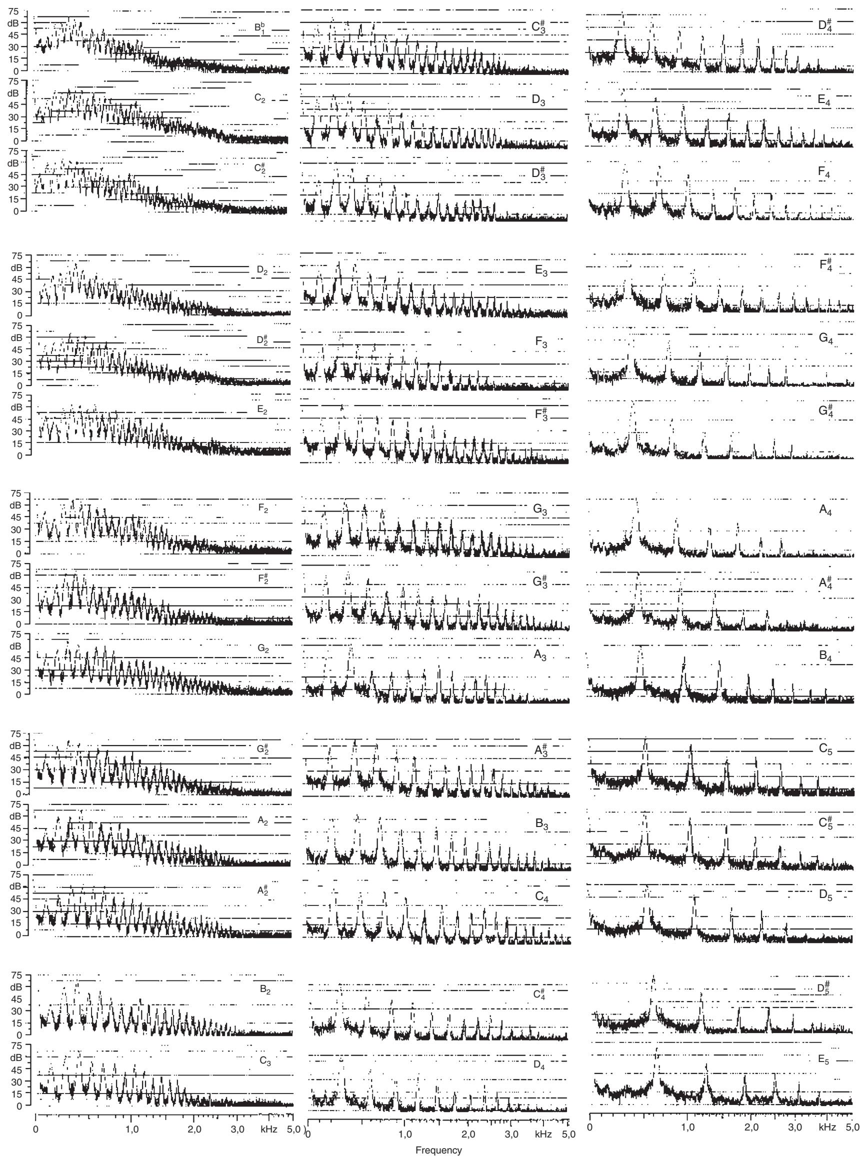

Fig. 2.4 Chromatic sequence of sound spectra for an F-Horn. Reproduction from experimental measurements

In order to be able to compare the behavior of different instruments at high and highest frequencies, it is desirable to establish a characteristic frequency as a reference point. For several reasons the frequency of 3,000 Hz appears particularly suited; one of the reasons is that sound power spectra – with only a very few exceptions – show a linear level decay above 3,000 Hz. Therefore, by specifying the slope of this decay in level above 3,000 Hz, on the one hand, and by specifying the level difference between the strongest partials and the 3,000 Hz component on the other, the course of sound power level spectra at high frequencies can be largely described.

Aside from the fact that tone color naturally changes over the tonal range of an instrument, as determined by the location of the fundamental, it can be said in general, that a tonal impression is brighter, and possibly sharper, as richness in overtones increases (in view of the frequency range and the intensity of the upper frequency components) (von Bismarck, 1974). For low tones, rich in overtones, the dense partial sequence in the upper frequency region leads to a rough character. This effect can occur for tones from the 3rd octave upward. While, for example, for G_3, overtones above 2,000 Hz effect a roughness, the corresponding limit for G_1 already lies at about 500 Hz (Terhardt, 1974). In contrast, overtone- poor sounds have a tendency for dark or soft timbre.

2.2.3 Formants

The fundamental certainly does not need to be the strongest partial in the sound spectrum. As the representation for the horn shows in Fig. 2.4, the fundamental dominates in this example in relation to other tone locations only in the upper register approximately from C_4 on upwards. For lower notes the intensity maximum is found at higher order partials. In this context it is remarkable that the location of the frequency maximum remains unchanged, and thus represents a particular characteristic for the sound of the instrument.

A similar phenomenon is also known from speech: When a singer sings a scale on a particular vowel, the region of strongest intensity remains fixed in its frequency location in spite of pitch changes. It is precisely for this region that all tones obtain the same (or at least similar) tone color. These amplitude maxima within a spectrum, which do not change their frequency with changing singing pitch, are called formants.

For the most important vowel colors the formant regions are assembled in Fig. 2.5 according to various authors. The individual maxima of the scheme mark the frequency region of strongest amplitude for the indicated sounds, between them transition colors are to be assumed, which can no longer be uniquely described by letters. They can be compared with the sound of the “same vowels” in various dialects. Generally for speech sounds, two or three more or less strong formants appear, for which, for the sake of clarity, only the most important ones are represented. Accordingly the dark vowels [u (oo), o (oh), a (ah), and å (aw). Translator’s note: the German vowel is given first, the English equivalent sound is given in parentheses] are each characterized by one maximum, and the bright vowels each by two. To clarify the frequency values, the peaks are also entered as notes.

Through the connection between formant regions and tone color of vowels, a second possibility (in addition to associating frequencies with the scale) presents

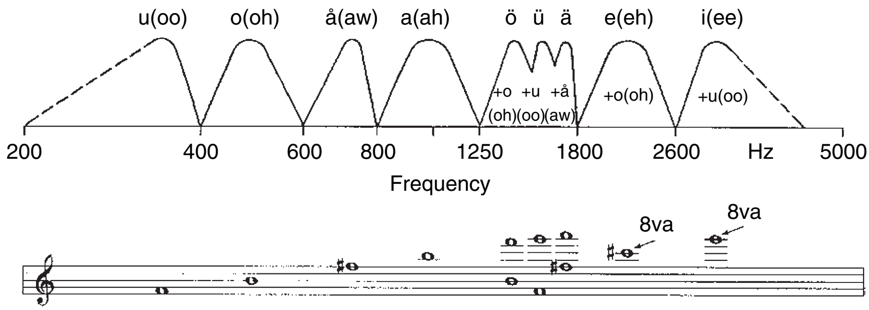

Fig. 2.5 Frequency location of formants for the vowels of the German language (after Thienhaus, 1934; Winckel, 1960; Trendelenburg, 1961)

itself to visualize the frequency regions of the sound of musical instruments. Strong components in the region of the u (oo)-formant (200–400 Hz) and above all the o (oh)-formant (400–600 Hz) are responsible for the fullness and sonority of the sound, while a pronounced a (ah)-formant (800–1,250 Hz) results in a forceful timbre. Especially the contributions between about 1,000 and 1,250 Hz prove very important for a striking tone impression. In contrast, excessively strong components in the region of modulated vowels (ö, ü, ä see translator's foreword) are sensed as uncomfortable, because they can lend a nasal character to the tone (Thienhaus, 1954), if the fundamental is too weak, and additionally insufficient intensity is present in the upper frequency region. Contributions in the region of the e (eh)-formant (1,800–2,600 Hz) and the i (ee)-formant (2,600–4,000 Hz) cause lightness and brilliance of tone. The typical regions for hissing sounds, which are perceived as voiced or voiceless, depending on whether they contain only noise components or also harmonic contributions, are located at frequencies from 4,000 Hz on upwards. Hissing sounds, however, possess pronounced formant character only in Slavic languages.

In order to achieve a vowel-like impression, a formant must be sufficiently clearly developed; i.e., the maximum of the frequency spectrum cannot be too broad. The so-called half-width can be used to characterize that. This is the difference between the two frequencies for which the sound intensity is just half the value at the maximum. Occasionally the width or clarity of a formant is described by the so-called logarithmic decrement which is calculated by f/f_m, where f is the half-width and f_m the mid-frequency. For example, the logarithmic decrement for the vowel "o (oh)" is 1.2, for the first and second formant of the vowel "e (eh)" it is 0.8 and 0.4. Raising the formants by 4–6 dB can intensify the character of a musical instrument, a further increase, however, is detrimental, making the sound rough and shrill (Mertens, 1975).

The esthetic effect of formants on the sound of musical instruments rests primarily on the similarity with singing, since here the human element itself, in a measure, becomes the standard. Furthermore, formants in the musical tone picture receive particular significance, since this characteristic of sound (in contrast to the upper frequency limit of the spectrum) is essentially independent of room acoustical influences. While it is true that the intensity of high frequency secondary formants can be weakened by room absorption, their frequency position, nevertheless, remains unchanged, so that the amplitude maximum continues to determine the tone picture.

2.2.4 The Effect of Individual Partials

The spectra of the horn in Fig. 2.4 show relatively smooth envelopes, which are largely characterized by the frequency range and the location of the formants. It is, however, entirely possible for individual partials to stand out within a spectrum, or on the other hand, have no intensity at all. In particular, sounds exist, for which odd numbered partials are more strongly developed than the even partials. A typical example of this is given by the gedackt organ pipes. For the clarinet in the low register, this type of spectrum also dominates. These lead to a covered and occasionally hollow tone, in this the absence of the octave components (2nd and 4th partials) also supports the dark timbre.

A similar hollow tone effect can also be achieved synthetically by appropriate instrumentation, as the score example 1 from the Bolero by M. Ravel shows. Naturally, because of the high pitched tone location of this passage, the tone color cannot be called dark, rather it resembles the Rohr-Flöte of an organ. The theme is played by the horn (in F) beginning from the C_5 in C-major. The two

Score example 1 M. Ravel, Bolero, measure 131 *ff*. Excerpts without strings and harp (in octaves) piccolo-flutes supplement the spectrum of the horn by a strengthening of the 3rd and 5th partials by playing the theme starting from G_6 and E_7 in G-major or E-major respectively. In spite of the complicated way of writing the score, a fully harmonic tone is produced with the horn part as the key note, with the celeste pointing to the octave at each note entrance.

While the dominance of odd partials leads to a covered tone, strong even harmonics and, above all, octave components lead to an open and bright timbre. This phenomenon is often used to advantage in instrumentation of orchestral works by using parallel octaves or parallel motion in double octaves. Even triple octave combinations are found (e.g., bassoon – oboe – flute in the third movement of the 9th Symphony by L. v. Beethoven, measure 65 sqq.). A sound with very equally strong partials, in contrast, can sound very hard, especially if no formants are present. Also the typical snarling sound of the reed pipes of an organ results from a spectrum very rich in overtones, with no dominant individual partials in the upper register, in contrast to flue pipes.

2.2.5 Frequency Width of Partials

The representation of partials, using lines, is a schematic simplification resting on the assumption that the tones do not change in their frequency, so that an exact location on the frequency scale can be assigned. In most cases minute amplitude variations are present, which, however, are not analyzed by the ear in their fine structure. Since the spectra represent a median value for a (quasistationary) steady state, these frequency variations can lead to a broadening of the spectral lines, i.e., the partials each fill a narrow frequency band. This effect becomes particularly apparent in a vibrato, when the sound gains "fullness" for a steady sound pressure level; time averaging due to the room also plays a role in this.

A similar broadening of the spectral lines occurs in cases where several sound sources radiate with nearly the same frequencies. Since this process is especially pronounced when in a group the intonation of each singer or player is slightly different, it is also called the chorus effect. This effect, among others, is also responsible for the difference in tone between a string orchestra and a string quartet. In a good choir the half width of a tone, i.e., the frequency width within which the sound pressure level drops by no more than 3 dB, lies in magnitude from 1/5 to 1/3 of a half step. In contrast, one characteristic of the Don-Cossaks is that the half width of their low notes is a full half tone. For instrumental ensembles the intonation width is generally relatively small and does not exceed a value of 1/5 of a half step (Lottermoser and Meyer, 1960).

2.2.6 Noise Contributions

Because of the nature of tone production, most musical instruments exhibit some noise background in addition to the spectrum of partials. This represents an important part of the total sound picture (Winckel, 1969; Meyer, 1964a). Thus, for example with string instruments, the bowing noise will always be somewhat perceptible, and for wood wind instruments the blowing noise can never be totally suppressed. An attempt will always be made to minimize such extraneous noise, however, such a minimal contribution is actually essential for the sound to retain its natural character. As experiments with electronic tone synthesis have shown, instruments cannot be imitated satisfactorily with spectra of partials alone.

These background noises come about through the fact that by irregularities in excitation all resonances of the instrument are slightly stimulated each time. Consequently the admixture of noise has a unique character for each instrument type. A comparison between the violin and the flute in Fig. 2.6 is intended to visualize this: While for the violin below the fundamental (which lies at around 1,400 Hz) and between the partials, noise components of significant frequency width are present with only small dips, in the flute, at a very low noise level, only a few small noise peaks are present at the frequencies of a fundamental, fingered as F_6, and overblown at the twelfth.

The superposition of the noise components over the harmonic spectrum, as well as the broadening of the partials, naturally presents a departure from the mathematically exact vibration form. Esthetically, however, these phenomena are extremely important, since they not only enliven the tone picture, but also prevent the appearance of fatigue symptoms of the ear. An analogy of these somewhat less than sharp tonal contours to painting comes to mind, where a sharpness scale, or resolution grades can be followed from an extremely photographic-like precision of a Canaletto, to the pointillist manner of impressionism. Similar cross-connections to other art forms can also be found.

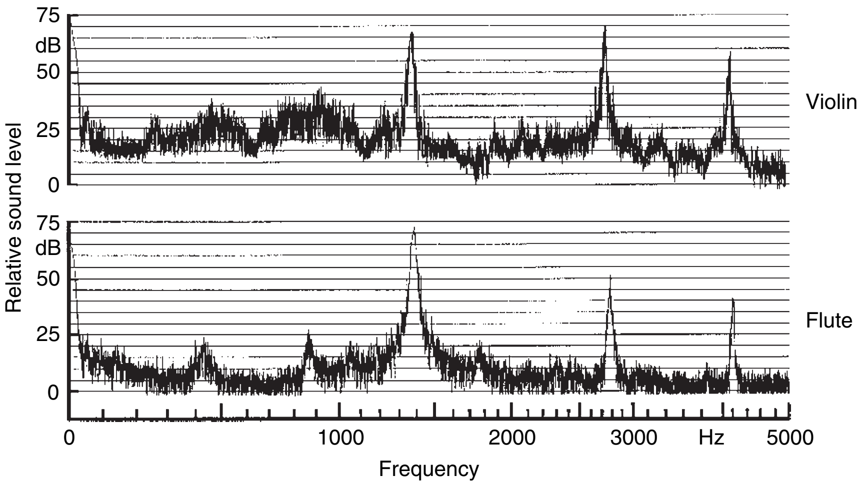

Fig. 2.6 Tone spectra with different frequency characteristics of the noise background. (Note played: F_6)

2.2.7 Dynamics and the Sound Spectrum

Playing volume is one of the most important factors influencing the sound picture. In this context the musical concept of dynamics refers principally to loudness, however, other factors such as relative motion or rest within a sound, or the nature of the attack can have a effect on the listener's impression of dynamics (Hadamowsky, 1958). The expressional value of dynamics thus by no means rests exclusively on differing sound pressure levels as they could be achieved, for example, by adjusting amplifier volume: Individual dynamic steps experience a particular characterization by the fact that for almost all musical instruments a change in playing volume varies not only the sound intensity in the stationary part of the sound, but also brings about clear changes in the spectral composition (Reinecke, 1953). By the way, it is this effect which has to be thanked for the fact that a *forte* is recognized as such for low volume sound reproduction by a loudspeaker even though it only has the intensity of a natural *piano*.

This phenomenon is illustrated in Fig. 2.7 for the upper register of a horn in the form of three spectra corresponding to the dynamic steps *fortissimo, mezzoforte* and *pianissimo* (Meyer, 1967b). The relative dB scale is set so that the strongest *ff* partial with 75 dB just reaches the upper limit of the registering level region. The very large overtone content of the upper spectrum is particularly obvious, here the partial contributions around 10,000 Hz still show amplitudes of about 45 dB, while in the *mf* spectrum only six partials and in the *pp* spectrum only four partials exceed the background noise. This already characterizes the tonal difference between the penetrating *ff* and the round and soft *mf* and *pp*. To that is added, that in the *ff*

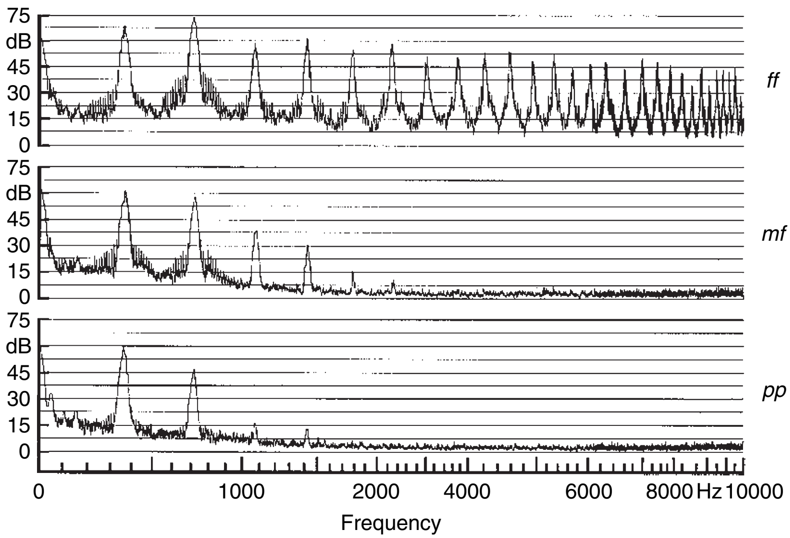

Fig. 2.7 Sound spectra of a French Horn played at different dynamic levels (F_4)

spectrum the maximal intensity has been shifted to the second harmonic (in contrast to the fundamental for *mf* and *pp*), so that the formant color becomes brighter. The importance of this spectral change for dynamics is recognized especially clearly through the fact that amplitude differences between *ff* and *pp* amount to 50 dB in the region of higher frequencies, while the difference between the corresponding strongest partials of the upper and lower spectrum lies at only around 15 dB. This results not only in a stepping of loudness (which can naturally still be subdivided), but also in a modification of the tone color.

When describing dynamic differences between instrument sounds, it is desirable, therefore, to take into consideration the level change of the 3,000 Hz component as well as possibly the change in slope of the level drop above 3,000 Hz, and also the change of the overall level, which is caused almost exclusively by the change in the strongest partials. All these data should be in relation to the sound power spectrum. A quantity, characteristic for each instrument group, is that level change of the 3,000 Hz component, which occurs when the strongest partials experience a change of just 1 dB. This quantity will be called the "dynamic tone color factor."

However, even when taking connections between loudness and tone color into account, individual dynamic steps such as *forte, mezzoforte, piano* etc. cannot be specified by a firm universally valid numerical value. The meaning of these performance specifications is as dependent on musical context as on the room environment. Finally the instruments themselves play an essential role, each instrument group has its own dynamic range between largest possible and minimum (still sounding) loudness. Furthermore, for many instruments this dynamic range is not the same over the entire tone range, where in addition even the quality of the individual instrument participates as a further factor.

2.2.8 Dynamic Range and Sound Power

Sound power level data are best suited to represent the dynamic performance range of an instrument in summary form, since they represent only the sound radiation from a particular instrument independently of the room, and can thus be recalculated for each room situation. Experience during measurements of the dynamic range accessible for performance have shown that there are noticeable differences in the accessible level limit, depending on whether the performers need to sound a quick tone sequence or can concentrate on each individual note. For a summary representation, both cases thus need to be considered.

In the sections on the dynamics of instruments (Chap. 3) the limiting values are given for the softest possible pp and the loudest possible ff, while playing fast scales covering two octaves, with the condition for the player that indeed each individual note is sounded. In addition, extreme values are given for individually played notes. Since the dynamic performance range of instruments often depends on the pitch – the low register of an instrument is often softer for pp, as well as for ff, than the high register – it is often the case that the softest playable tone is different than the loudest possible. These extreme values are especially of interest for microphone recordings. Under usual performance conditions it is, however, more important to be aware of a realizable dynamic range. Appropriate values should therefore be given as averages for fast and slow tone sequences.

Most of the sound power values presented in the following chapters are taken from sound power measurements performed using the reference sound source procedure in a reverberant chamber (Meyer and Angster, 1981; Meyer, 1990). They will be augmented by the earlier results of Burghauser and Spelda (1971), which were obtained in a medium size studio, as well as those of Clarke and Luce (1965), which were measured in a reflection poor room. All results by these authors were recalculated as sound power levels with consideration of all relevant boundary conditions.

For subsequent room acoustical considerations it is also necessary to know a characteristic value for the sound power of each instrument. For this purpose an "average *forte* sound pressure level" will be used, which is established on the basis scale, and single tone measurements in such a way that the playable dynamic range between *pp* and *ff* is divided by the steps of *p*, *mf* and *f* into equal segments. Starting from this *forte*- level the sound power of entire ensembles, or the sound pressure level to be expected in a room for arbitrary dynamics can be calculated (Meyer, 1990).

2.3 Time Structures

2.3.1 Deviations from a Steady Vibration Process

Decomposition of a sound into a sequence of harmonic partials is, strictly speaking, only possible for a stationary state. In contrast, a sudden change means an unsteadiness, which can no longer be described by a line spectrum. Such unique processes thus possess a spectrum which is not formed by discrete individual frequencies, but by a frequency continuum. This presence of arbitrarily closely spaced frequencies gives the ear a noise-like impression, which, during the onset, can assume the character of a crack.

Figure 2.8 shows the vibration processes and the spectra, next to each other, for a steady vibrating sinusoidal tone, and for the onset of a sinusoidal tone. The steady state can be represented by a line, while the onset of the tone exhibits a spectral broadening in the region of the tone, and beyond that, a decreasing amplitude with increasing frequency deviation. Also the end of a tone can be understood as such a process. The sound proceeds as though, in addition to the existing tone, a tone equal in amplitude and opposite in phase were suddenly turned on. The tones cancel each other, and the second crack remains. In practice, switching processes without some kind of a transition phase, occur only for some older types of electronic organs, in which the keys switch the already running tone generators to the amplifier and speaker. These instruments thus begin each tone with a noticeable crack, since the

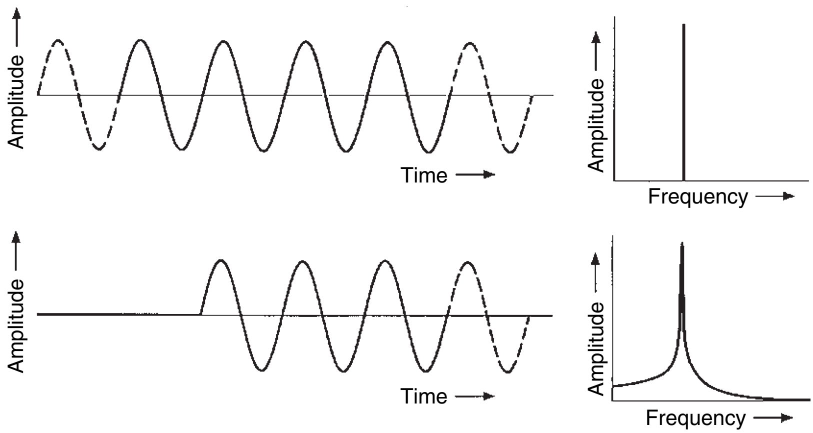

Fig. 2.8 Wave form and Spectrum for a steady tone and a tone entrance

eigenresonances of the speaker are excited. To some extent, similar cracks can also be heard at the end of the tone, which is perceived by the ear as an additional accent (Lottermoser and Meyer, 1962a).

Generally, such a switching process does not occur as suddenly as represented in Fig. 2.8, but is extended over a certain time period. Consequently the amplitudes of the noise components, which are present in addition to the actual vibration frequencies, become increasingly less as the duration of the switching process increases, and as the break in the vibration process at the onset is smoothed out. Naturally the final vibration amplitude in such a case is reached only after the full course of the onset process which is stretched in time. Since for such a soft transition from rest to the tone, the unsteadiness in the vibration process is reduced, the noise-like impression in the ear of the transient also decreases.

2.3.2 The Starting Transient

Musical instruments are complicated physical structures, for which in most cases, several coupled resonance systems influence the vibration process between excitation by the player and the radiated sound. A resonance system, however, cannot react suddenly to an excitation, rather, vibrations must slowly build up to their final strength. This is connected with the fact that a portion of the energy provided externally for the resonance system is in turn radiated, and a portion is absorbed by the instrument. As long as more energy is provided than used, the amplitude rises. Only when an equilibrium is reached between the input energy on the one hand and the absorbed and radiated energy on the other hand will the oscillation reach its ultimate strength. This starting transient is easily observed by placing a vibrating tuning fork on the corpus of a violin: the loudest volume is reached slowly after a gentle transition.

The more an instrument absorbs and the stronger it is damped by radiation, the sooner it will reach equilibrium, i.e., the shorter will be the duration of the starting transient. Thus damping plays an important role during the attack for an instrument. For most resonance systems, this damping is in large measure frequency dependent. This means on the one hand, that characteristic vibration processes proceed at different rates within the tonal range of the instrument, and on the other hand that individual partials grow at different rates in the process of developing the tone. The sequence of starting transients of the overtones belongs to those special characteristics which form the tone picture of the individual instrument groups. In all this, high tonal contributions with fast initial transients have the effect of suggesting a precise attack, similar to articulation in speech, while starting transients which are too slow, are perceived as a poor attack for the instrument.

Inasmuch as the amplitude during the starting transient rises very quickly initially, but then reaches its final value only with relatively small intensity increase, and since furthermore, this final value is difficult to determine precisely due to the appearance of fluctuations, the literature usually specifies the time by which the sound pressure level has reached a value 3 dB below the stationary state as the duration of the starting transient (Luce and Clark, 1965; Melka, 1970). This definition is generally applied to the total level without frequency weighting of individual sound portions, in practice this emphasizes the strongest partials.

Naturally, the duration of the starting transients can be influenced by the performer within certain limits. The harder a tone is attacked, the richer in overtones will be the unsteady contribution during the time process of the excitation; consequently the high tone portions will develop faster and more precisely. Nevertheless, natural limits are in place through the damping of the individual resonances. The more softly a tone is attacked, the slower the initial transient process of the instrument will be developed, and consequently the formation of the higher components will be weakened.

This difference is shown clearly in the juxtaposition of two starting transients for a G_4 on a flute in Fig. 2.9. As seen, the onset of the strongly attacked tone is marked by an attack noise of 40 ms duration. The fundamental and the third partial develop very quickly at the same time; the fundamental already reaches its full strength after 70 ms. The octave partial and the 3rd partial reach their final value after 90 and 100 ms respectively, the higher partials need 100–120 ms. In comparison to that, for the soft attack, the duration of the starting transient for the fundamental (120 ms) is not that much longer; the soft character primarily comes from the delay of the higher partials and the freedom from noise. Otherwise, the example also shows that the speed of transient development is a typical characteristic of performance style, which can be utilized as a means of expression. In contrast, possibly only the shortest possible duration of the starting transient of a staccato tone can be utilized to characterize an instrument.

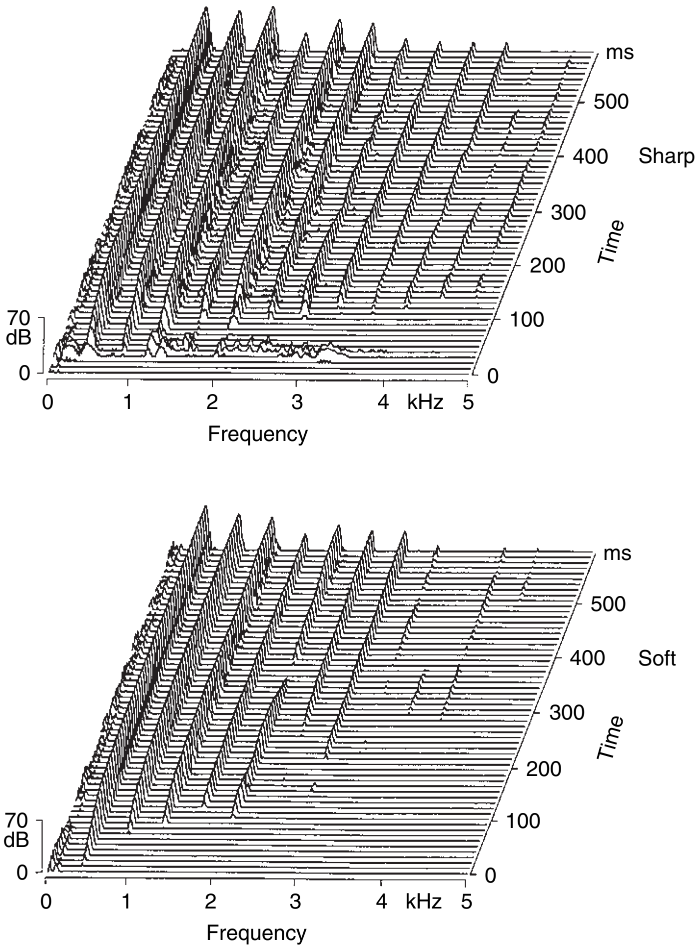

Fig. 2.9 Time development of a tone spectrum for different attacks (Flute, G_4)

2.3.3 Inharmonic Components

Since the sudden stimulation of an oscillating system represents an excitation of a frequency continuum because of unsteadiness in time, the starting transient includes not only the harmonic tone contributions, which will later be present in the stationary state, but the instrument will be caused to vibrate at all resonance frequencies. To the extent in which these frequencies do not coincide with later partials these oscillations form inharmonic tone contributions, which, however, essentially decay again during the initial transient (Lottermoser, 1958; Lottermoser and Meyer, 1966). In this way characteristic admixtures come about in the onset of the tone, which present properties typical of each instrument group by reason of the position and width of the resonances.

When resonances are broad and strongly damped – as for example in wide frequency ranges for string instruments – then corresponding broad band noise components arise during the initial transient, which cause an auditory impression of a tone color comparable to a whisper, however, without a definite sensation of pitch. Sharp resonances, which are not utilized for the formation of harmonics – as they are present, for example, in the overblown tones of woodwind instruments, are recognized by the ear at their pitch level, in spite of the noise character.

In the special case when the exciting frequency lies close to a resonance, beats are generated during the starting transient between the excitation frequency and the resonance frequency, which only disappear with the decay of the resonance frequency (Meyer, 1985). Depending on the separation of these two frequencies, the beats can enhance the tone picture or otherwise influence it through roughness. Such phenomena can occur particularly for the high partials of wind instruments when the upper resonances are not situated strictly harmonically; they are occasionally also found in string instruments when the corpus resonances are too sharp.

Though noise contributions in the initial transient should not be too obvious for tone esthetic reasons, they, nevertheless, contribute in an essential way to the precision of the tone entrance, as seen clearly in Fig. 2.9. In this respect they are comparable to the consonants in speech, which likewise form a noise-like introduction to the vowel vibrations, though in most cases with greater intensity. This articulation, emphasized by the nature of the tone entrance, can be further enhanced when tones in the frequency neighborhood are excited shortly prior to the actual tone. Some anticipatory grace notes serve in this way to accentuate the following main note, and are, so to speak, to be considered as "composed initial transients." In score example 2, there are even harmonically placed double grace notes which give the accent to the heavy beat.

While it is possible to use the ability to obtain an audibly precise, recognizable entrance, as a further criterion to evaluate the ability to address the instrument, the strength of the articulation as a means of artistic expression is subject to the playing technique of the musician. Yet it must not be overlooked, that pitch recognition by the ear is made more difficult through the noise components associated with very short notes, even if the rhythmic precision increases. The instrument groups with low noise level, but short initial transients are thus best suited for a sharp staccato.

Score example 2 Joh. Strauß, Polka "Leichtes Blut" 'Opus 319, 1st Violin, measure 5 sqq.

2.3.4 Decay of Resonating Systems

For no mechanical instrument does the termination of excitation occur as abruptly, as it begins for a staccato tone. Consequently, in contrast to initial transients, no new tone or noise components are excited. However, it is important for the decay process, that sound energy is still stored in the resonance systems of the instrument, which is radiated until it is used up. Of concern is a process opposite to the building of oscillations in the initial transient. Depending on their damping, each resonance decays more slowly or more rapidly. A tuning fork, for example, has a very sharp resonance, after the attack, its vibrations decay only slowly. In comparison, the resonances of wind instruments are strongly damped, so that they use the rest energy in a very short time, and practically no decay is audible.

Naturally, the decay can be influenced within certain limits by playing techniques. As was already recognized from Fig. 2.2, the nature of the decay for bowed string instruments depends on the bowing force, and is longest when the bow is lifted. It can also be shortened when the string is retuned by a change in fingering. Such a string frequency change can be compared to the opening or closing of keys or valves in wind instruments. These phenomena have particular significance for connecting two tones: by overlapping the decay of the first tone into the onset transient of the second, the melodic line gains continuity, while an all too sudden cut often leads to unwanted gaps.

Naturally, decay processes gain preferred interest in instruments for which, by reason of a very short excitation, no steady state can be developed. Percussion instruments fall into this category, as do struck or plucked string instruments. For these, the duration of the decay is particularly long, and it determines the tone picture more strongly by far than does the duration of decay processes in other instruments.

Since the amplitude decreases exponentially for damped resonances, a logarithmic representation of sound pressure level shows a nearly linear time dependence of decay, as is also shown in the model in Fig. 2.1. However, the slope of this drop is by no means equal for all partials, since the damping of the resonances is frequency dependent in most cases, where generally the high tone contributions drop faster than the low ones. Furthermore, tones, whose frequencies fall directly on a resonance, decay differently than tones which lie between two resonances.

2.3.5 Decay Time and Reverberation Time

The audible duration of the decay, on the one hand, naturally depends on the loudness of the tone, and on the other hand, on the ambient noise level in the room, since, as far as the ear is concerned, the end of the decay is determined by the point at which the tone drops below the ambient noise level. In analogy with room acoustics, the time span during which the decay is audible is referred to as the decay time. This measure is quite suitable for the description of the tonal

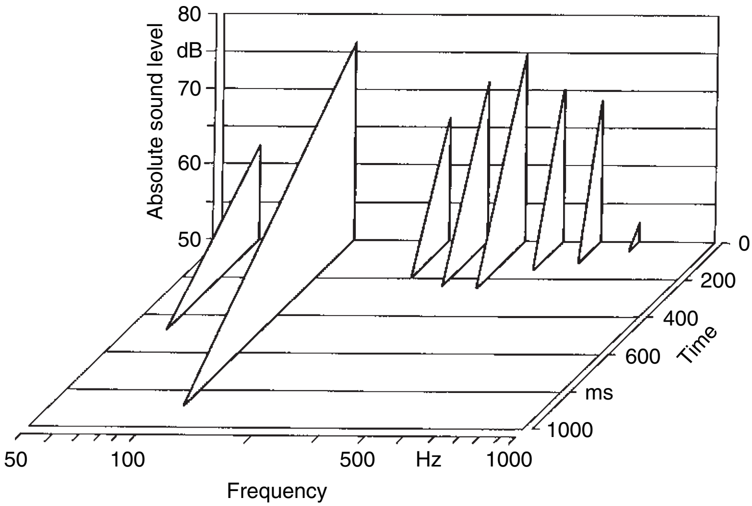

Fig. 2.10 Decay of a double bass tone (B_1, *pizz.* after Spelda, 1968)

impression, nevertheless, because of the dependencies mentioned above, it is not an objective measure of the slope of the level decrease.

The decay time for a pizzicato tone of a contrabass is reproduced in Fig. 2.10 as a typical example of a decay time. For this representation a noise level of 50 dB was assumed as a lower limit for the level drop, this level forms the base plane of the model. Time is plotted toward the front, the sound pressure level goes up, and the frequencies of the overtones increase toward the right. It is noted from this picture that the low tone contributions exhibit the shallowest decay curves, while the high components decay rather rapidly. The 2nd partial can be followed for the longest time, particularly since it also has the highest starting value. The overtone content accordingly changes in the decay in such a way that the tone color becomes increasingly darker and softer. As an alternative to this spectral representation, the decay time can also be given as a numerical value, which only gives the duration of audibility without frequency specification. The pizzicato tone of the pictured example would have a decay time of about 900 ms.

In order to obtain an objective measure for the slope of the level decrease, and thus for the tonal characteristic of the relevant instrument, often the time for which the level drops by 60 dB relative to its original value is determined, and this quantity is designated as the reverberation time -again in analogy with room acoustics- (Meyer and Lottermoser, 1961; Plenge and Schwarz, 1967). Since the level decrease is linear, this quantity can naturally also be calculated when the level record only covers a smaller region. For example, for the 2nd partial in Fig. 2.10, from a 900 ms time lapse and an associated level drop from 77 to 50 dB, i.e., of 27 dB, a reverberation time of 2 s is calculated (for a 60 dB drop).

Complicated resonance systems at times do not exhibit a linear level decay, but rather follow an initial steep portion with a later more shallow decay. This is most frequently explained by the fact that a strongly damped vibrating system radiates much energy, while at the same time a different section of the instrument, which absorbs less energy, radiates that energy only slowly. This causes a break in the level curve, and the decay process must be described by specifying two values for the reverberation time (for short and for long tones).

2.3.6 Fluctuations in the Quasistationary Part

Changes of the vibration excitation during a tone also need to be considered as unsteadiness in the vibration process, even if they are significantly less noticeable than during the starting transient and the decay. Persistent statistical fluctuations of the excitation are however, finally the reason for the noise admixture to the sound of the instrument, as discussed earlier. The measure of these variations transmits an impression of the stability or – in the negative case – the uncertainty of the tone. On occasion, however, such instabilities are deliberately utilized as tonal effects (*flautando*).

A particularly important structuring of the tone, introduced as an artistic means of expression is represented by the vibrato (Winckel, 1960; Gärtner, 1974). While it is used by singers, string-, and wind-players to very different degrees, yet, a good vibrato has something in common for all voices and instruments: The fluctuation frequency of the vibrato almost always lies in the region of 5–8 Hz, which is connected – as already mentioned – by the fact that the ear still senses a definite pitch for fluctuations at that frequency. In contrast, the width of the vibrato characteristic for singers, string- and wind-players is different. It can move by ± 5 cents near the limit of the audible, or for singers it can certainly exceed the range of ± 100 cents. The effect of the vibrato on the tone quality is also different: while the vibrato in all cases depends principally on a time modulation of the exciting frequency, still, depending on the resonance structure of the instrument, more or less pronounced modulations of the individual partial amplitudes result. If these occur in phase in a larger frequency region, a time dependent tone color modulation results. Occasionally this shapes the tonal impression more strongly than the original frequency modulation. The latter, for example, is the case for brass instruments and the flute (Meyer, 1991).