8.2 Frequency and pitch

Frequency is a quantity from the realm of physics, while **pitch** – as a sensory perception-quantity – belongs with the auditory event. Usually, frequency is measured in Hz, representing the numbers of oscillations per second. The unit Hertz (abbreviated Hz) is named after the physicist Heinrich Hertz. The inverse of the frequency is the period (short for duration of periodicity of a cycle). A period of T = 2 ms corresponds to a frequency f = 500 Hz. The **pitch** may either be determined via self-experiment (introspection), or indirectly via evaluation of the reaction of a test-person (a "subject"). Although the pitch is a subjectively rated quantity, it can be measured numerically. **Measuring** means in this context to allocate numbers to an object-set according to predetermined rules, with these numbers being reproducible within purposeful error margins. Now, what one considers purposeful – that again is a rather subjective decision♣. Most psychometric experiments yield intra- and inter-individual **scatter**: one and the same test person may give different evaluations when carrying out the same experiment a number of times (intra-individual scatter), and the assessments of different test persons may vary when an experiment is presented once for each person (inter-individual scatter).

8.2.1 Frequency measurement

Simple measurement devices for frequency count the number of oscillations occurring per time-interval: 5 oscillations per 0,1 s yields 50 Hz, for example. 'Oscillation' always implies a whole period. For a string, this means: swinging from the rest-position in one direction, reversal at the crest (apex), swinging (across the rest-position) fully to the other apex, reversal at the latter, and swinging back to the rest-position. Given an ideal oscillation, terms such as frequency or period are thus easily definable – real oscillations are, however, not ideal. Signal theory defines a **periodic** process as *infinitely repeated in identical form*. Thus, a sine tone is periodic and has one single frequency. A sound composed of a 100-Hz-tone and a 200-Hz-tone (in music this would be called a note) would be periodic, as well (**Fig. 8.10**). However, since more than one frequency appears here (i.e. 100 Hz and 200 Hz), we need to distinguish between **frequency of the partial** and the **fundamental frequency**. Now, the fundamental frequency is not necessarily that of the lowest partial, but the reciprocal of the period. The oscillation-pattern of a sound comprised from sine components at 200 Hz, 300 Hz, and 400 Hz repeats after 10 ms; the fundamental frequency therefore is 100 Hz although there is no actual partial found at 100 Hz within that sound. Generally speaking, the fundamental frequency is the largest common denominator of the frequencies of the partials, and the period is the least common multiple of all periods of the partials.

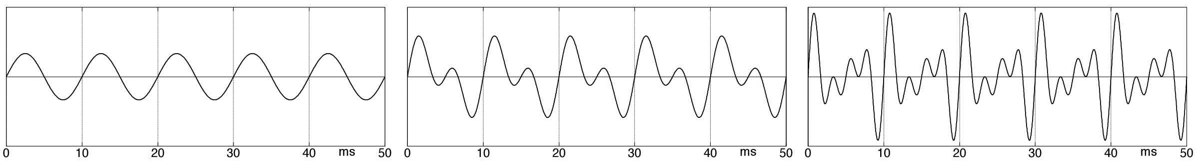

Fig. 8.10: Sine tone (100Hz), two-tone sound (100|200Hz), three-tone sound (200Hz|300Hz|400Hz); 0–50 ms each.

♣ A driver of a vehicle that has just reflected a high-frequency radar-beam may possibly demand a larger margin of error than what the municipal administration profiting from motoring fines would see as appropriate.

Evidently, a tone does not need to be of mono-frequent characteristic to feature *one* frequency (more exactly: one single fundamental frequency). In theory, there even may be an infinite number of partials (as is the case for an ideal square-wave sound). However, the partials have to be **harmonic** i.e. their frequencies need to be integer multiples of the fundamental frequency. This condition cannot be met e.g. for irrational numbers such as $\sqrt{2}$ und $\sqrt{3}$. In practice, though, frequencies can be specified only to a finite number of decimals, e.g. 1,414 Hz, or 1,732 Hz. If these examples would be rounded-off roots, a specification of "the fundamental frequency is 0,001 Hz" would be very arbitrary. Nor would it be within the meaning of the largest common denominator; 0,002 Hz, at least, would be a common denominator. It should be noted that the issue with the irrational numbers is of a less academic nature than one would think. This is because string vibrations are never of an exactly harmonic nature. The decay process gives every "period" different amplitudes, and the partials are not actually in a strictly harmonic relationship (i.e. they are **un-harmonic**), due to bending stiffness, and to the dispersive wave propagation connected to it. Let us assume that the decay process is so slow that its effects on the spectral purity may be disregarded. Let us further assume that the analysis of a guitar tone has yielded four components (partials) at the frequencies of 100 Hz, 201 Hz, 302 Hz, and 404 Hz. What would be the frequency of this tone? It makes no sense to specify 1 Hz as the fundamental frequency, and to call the partials the 100th, 201st, 302nd, and 404th harmonic, respectively. What remains is the sobering insight that **a guitar tone has no fundamental frequency**. It does, however, have a pitch! Determining that pitch shall be reserved for a later paragraph – first we still have to clarify what a **tuning device** is in fact doing given the above finding, and why a string may be tuned – despite all this.

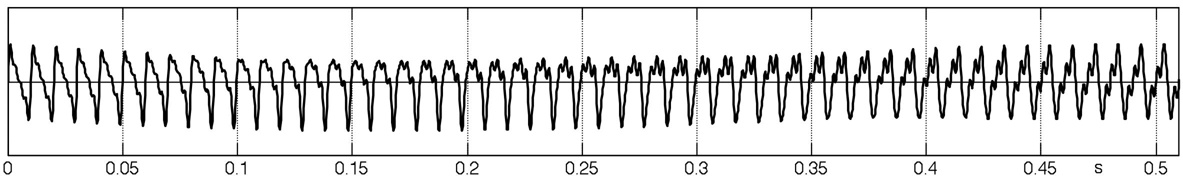

Fig. 8.11: 4-partials sound, f_1 = 100Hz, f_2 = 201Hz, f_3 = 302Hz, f_4 = 404Hz. $1/f$–envelope; t = 0 – 0,5 s.

Fig. 8.11 depicts the first 0,5 seconds of a sound comprised of the frequencies mentioned above. How many periods appear during that time interval? Trying to count the maxima, we get into a bit of trouble approaching the mid-section of the figure, but we can make it to the right-hand end with the finding that there will be about 49 and 3/4th periods. But what is that in this case: a "period"? To deliver a *visual* evaluation, our optical sense seeks to – as well as possible – perform visual smoothing (i.e. filtering!) and locally limited auto-correlations. What else could a visual system do in the first place. But will that be helpful in the context of an acoustical signal? What could an exact algorithm be? Simple measurement devices determine the upward (or downward) zero-crossing. Given the above signal, that will make for considerable problems between 0,15 and 0,2 s, and between 0,35 and 0,45 s. Of course, **smoothing** (i.e. low-pass filtering) is a solution, but with it the frequency of the *filtered signal* will be determined. In the extreme case, the filtering will pass on merely the 100-Hz-oscillation – with that, the frequency-measurement certainly is most straightforward. Presumably most tuning devices (electronic tuners) have a built-in low-pass filter that filters string-specifically, or at least instrument-specifically. Also, they will accept small deviations from the nominal value. It may still happen that the display sways back and forth between correct and incorrect. The well-versed guitar player will then turn down the tone control (low-pass filter) or relinquish any high demands on accuracy. Some may celebrate an act of the gripping drama: "Guitarists never stop tuning, guitars eternally refuse to be correctly tuned".

The frequencies on which Fig. 8.11 is based show the fundamental problem but do exaggerate the situation. The spreading of the partials found with electric guitars amounts to about 0,2% for the E_2-string at 500 Hz, and to about 1% at 1 kHz. Still: if the 12th partial of the low E-string is represented with substantial level in the overall signal, a possibly annoying discrepancy of about 9 Hz between ideal $(12 \cdot 82{,}4\,\text{Hz} = 988{,}8\,\text{Hz})$ and real (997,7 Hz) may result. Such an error would be inacceptable for precise tuning. However, the amplitudes of the higher partials usually decay much faster than those of the lower partials, and thus most electronic tuners achieve an acceptable reading, especially since the guitarist will pluck the string rather softly so as not to emphasize the harmonics too much. For the lower partials of the low E-string, the inharmonicity will then be rather unproblematic with 0,02% for the third harmonic, and 0,05% for the fourth. There will be even less of an issue for the higher guitar strings: due to the smaller string diameter, the bending stiffness plays not as big a role, and the number of the possibly interfering harmonics decreases due to the low-pass character of the pickup.

As a summary, we may therefore note: even though the string vibration is comprised of inharmonic partials and therefore in theory has no fundamental frequency, electronic tuners will in practice detect the frequency of a “practical” fundamental, or a value that is very close to it. Whether our hearing system arrives at the same conclusion is, however, an entirely different question (see Chapter 8.2.3).

8.2.2 Accuracy of frequency and pitch

Following a chapter on frequency measurements, it would seem natural to explain pitch determination in more detail. First, however, desired accuracy and measurement errors shall be looked into. This way it will be easier to assess the properties of the hearing system that will be the focus in the subsequent chapter.

The frequency of a strictly periodic tone can be measured with an accuracy that is more than adequate for musicians. Precision frequency counters feature relative measurement errors in the range of 10^{-5}, and 10^{-6} is not impossible, either. In a watch, for example, an error of 10^{-5} leads to an inaccuracy of 1 second / day. The problem does not lie in the underlying reference (oven-stabilized quartz generators are extremely accurate) but in the signal to be measured. Measuring does become tricky if this signal does not have exactly identical periods. Given a known shape of the signal, frequency measurement is simple and quick: three points on a sine curve (excluding a few special points such as the zero crossing) suffice to determine the three degrees of freedom: amplitude, frequency, and phase. In theory, the three points may succeed one another very quickly, and thus achieving both high measurement precision and a short measuring time is not a contradiction. These highly theoretical findings based on function analysis do not help for measuring the frequency, though. This is because the shape of the signal is not known, and with that the rule holds that the **duration of the measurement** and the **accuracy of the measurement** are reciprocal to each other. If the frequency measurement is based on counting periods of the signal, a measurement of a length of 10 s is required in order to achieve an accuracy of 0,1 Hz. Interactive tuning would be impossible given such long durations. Frequency-doubling or half-period-measurements could be advantageous, but requires that the duty-factor of the signal is known – which is not the case with sounds of musical instruments. What remains is to determine the frequency of individual partials. Presumably, most tuning devices will indentify the frequency of the fundamental, and – in the case of the guitar – will indicate that as the frequency of the string.

It is not only the measurement process that requires us to consider the measurement duration, but also the fact that the signal to be measured is **time-variant**. The amplitudes of the partials decay with different speed as a function of time, and moreover the **frequencies of the partials** will change. This is connected to the string being elongated and thus stretched more as it moves from its rest-position: the larger the vibration-amplitude, the higher the frequency. Further, it needs to be considered that real string oscillations are never limited to remain in exactly **one single plane**. During the decay process, the plane of oscillation rotates; this can be seen as the superposition of two orthogonal vibrations. Due to direction-dependent bearing-impedances, these two vibrations may differ slightly in their frequencies, and consequently there will be changes in amplitude and frequency over time.

A (non-representative) field experiment shall give some indications of how accurate the frequencies of strings can be measured despite all these issues. From the many digital **electronic tuners** on the market, three were selected and checked using a sine generator and a precision frequency counter. The ranges within which the devices registered a “correct tuning” measurement were ±1,6‰, ±2,0‰, and ±2,3‰, i.e. on average **±2‰**. This corresponds to **±3,5 cent**. To be clear: “correct tuning” in this context means that, for example, the device under test evaluated all frequencies between 439,4 Hz and 440,7 Hz as *correctly tuned to A*. The width of that tolerance interval is a compromise between high precision (possibly never achievable due to the aforementioned issues) on the one hand, and more easily achievable “kind-of-in-tune” state (that may not be accepted due to audible deviation from the ideal value) on the other hand.

Fig. 8.12 shows the progress over time of such a measurement. Using a tuning device (Korg GT-2), the tuning of two guitars was assessed; depicted are the deviations of the value indicated by the tuner from the reference value (during 8 seconds of a measuring time; for each string). The string was plucked with regular strength at t = 0; all non-involved strings were damped in order to avoid interferences. For the measurement with the Gretsch Tennessian, the stronger decrease of the pitch during the first seconds stands out. This effect was not further investigated; a cause may be found in the relatively thin strings: their average tensile stress is increased with strong vibration. Towards the end of the shown measuring time, the deviations increase; this is due to the decreasing signal level. The Ovation (with the signal of the piezo pickup measured) also caused some fluctuations during the measuring time; the causes for these were looked into in more detail.

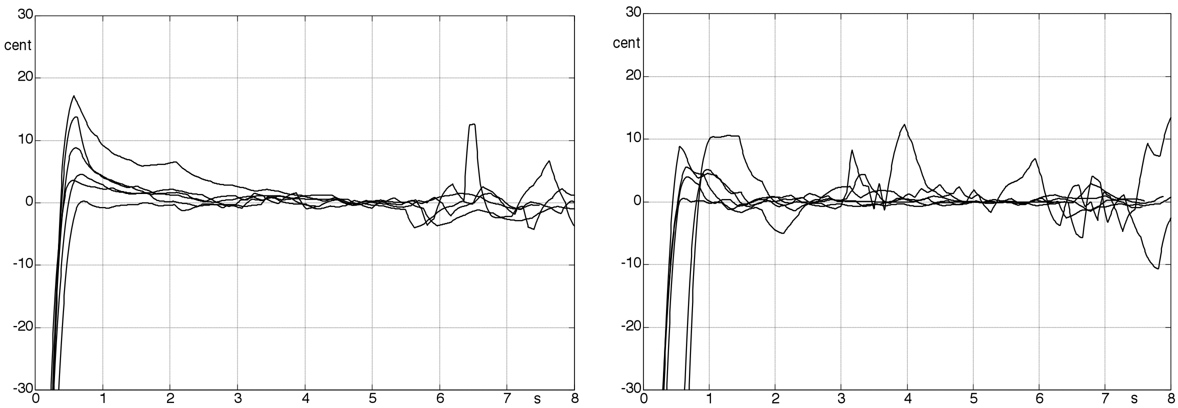

Fig. 8.12: Pitch measurement with the electronic tuner Korg GT-2. Tennessean (left), Ovation SMT (right).

In **Fig. 8.13**, the measured pitch is compared to the level of the fundamental over time. The signal generator is in both cases the plucked B-string of the Ovation SMT. At 3,5 s we see a minimum of the level of the fundamental. Assuming a time lag of around 0,5 s due to the processing, pitch-fluctuations at about 4 s can be explained; the other fluctuations cannot be attributed to anything specific with any certainty.

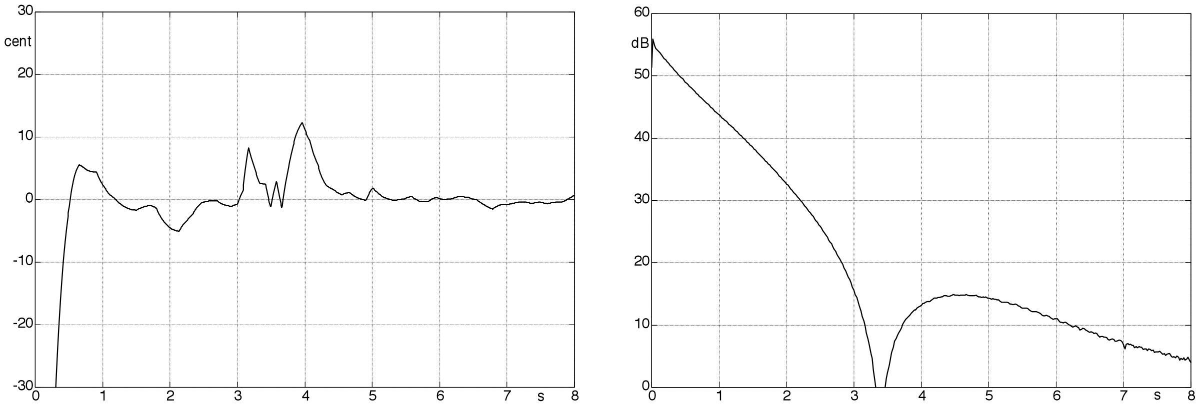

Fig. 8.13: Measured pitch-deviation, level of fundamental; Ovation SMT, B-string plucked at $t = 0$.

The measurements show that – despite alleged digital precision – considerable fluctuations in the display value are to be expected. Since the electronic tuning shows a highly accurate display without any noteworthy fluctuation when a precision generator serves as input, only the guitar tone itself can be the reason. The more "lively" this tone is, the larger the fluctuations in the measurement result will be, and the larger the variations in pitch.

At this point, a short digression into **thermodynamics** makes sense. The linear thermal expansion coefficient describes how dimensions change dependent on temperature. If the dimensions are "imprinted" (forced), the mechanical stress will vary as the temperature changes. This implies for steel strings: the un-tensioned string will experience an elongation by a factor of 16×10^{-6} for a temperature increase of +1°C. While this appears insignificant compared to the 2‰ mentioned above, we need to consider that for the change of the string frequency, the relative change in *stress* is the influential factor. Typically, an E_2-string needs to experience an elongation (strain) of about 1,5 mm for correct intonation. It is this 1,5-mm-strain that needs to be seen in connection with the change in length caused by the temperature change. The relative frequency change corresponds to half the relative change in strain (square-root in action here!). For our example, this means: **with 1°C temperature change, the string frequency changes by 5,3‰**. Here we assumed that the dimensions of the neck and body of the guitar remain constant; given the highly different thermal time constants over short time-periods, this is justified. Confirmation was provided by an experiment: taking a correctly tuned guitar (Gibson ES-335) from a room to the outside (cooler than in the room by a few degrees) raised the frequency of the E_2-string within a few seconds by 12‰. Conversely, it follows: if you seek to keep the tuning of a guitar constant within 1‰, you need to demand that short-term temperature fluctuations remain within 0,2°C ☺.

We have saved the most important question for last: **how precise actually is the hearing system?** In the terminology of psychoacoustics: how large is the threshold of pitch discrimination? You will find quite different answers – it depends on the experimental methodology. Fundamentally, we need to distinguish between a **successive pair** (2 tones follow each other in time) and the **dyad** (two-tone complex; two tones are sounded at the same time)

When concurrently presenting two tones, the smallest of differences between frequencies may be noticeable – depending of the circumstances. For example, if two 1-kHz-sine-tones are detuned by 0,1 Hz with respect to each other, a beating results: i.e. a tone is gradually getting louder and becoming softer again, with its amplitude reaching its maximum value every 10 seconds. The latter duration is short compared to the average life expectancy, and also small relative to the tolerance-span of the test persons (subjects) – therefore it is well observable. For the same reasons, a periodicity of 0,01 Hz would still be observable – but with 0,001 Hz the limited patience of the subject might become a problem. Relative to 1000 Hz, 0,001 Hz already represents is factor of 10^{-6}. However, to conclude that the frequency resolution of the auditory system would always be 0,001‰ – that would be nonsense. The result is only usable in the given experimental context.

Clearly, a large part of music consists of sounds comprising two or more tones – so: what gives? The answer will necessarily remain unsatisfactory because music is diverse, but there are rough guidelines. A first borderline is defined by the duration of sounds. If a sound consisting of two tones lasts only for a second, a frequency deviation between the two tones of 0,1 Hz will not be detected. Sounds of longer duration generally facilitate recognizing frequency differences. Still: long sustained notes are often played with **vibrato** (for the terminology see Chapter 10.8.2), and in this case a small detuning will be noticed less. Pitch vibrato, however, cannot be generated on every type of instrument – but then a polychoral design will make for audible modulations already in single notes. On the piano, for example, most notes are generated by two or three very closely tuned strings; beats will be inherent in the system here. Even when trying to tune all strings of one piano note to exactly equal pitch, the overcritical coupling of the strings will result in beating. Besides the beats audible in the single note, additional beating between different notes may be audible as a separate characteristic – but this will depend on too many factors to make an analysis with simple algorithms feasible. Looking at the distribution of how often musical notes of certain durations occur, and considering the auditory fluctuation assessment, we may cautiously estimate the following: upwards of an envelope-period of about 1 s, beats loose their sensory significance. This corresponds to a frequency resolution of about 1 Hz.

Given a **sequential presentation** of tones, beating is excluded. Or so many psychoacousticians believe. However, of significance is not which sounds are generated, but which sounds actually arrive at the ears of the subjects. Presentations of sounds in a room are always accompanied by **reflections** – if these occur in great numbers, they are called **reverb**. If the pause between sequentially presented tones is too short, there may still occur a short beating at the transition, and this beating may be perceived depending on the circumstances. Such experiments should therefore exclusively be carried out using headphones. A room as a transmission system has other issues, as well: due to the superposition to interleaved reflections, the impulse response is lengthened. The Fourier transform (the transmission function) obtains selective minima and maxima, and between these includes steep flanks. A frequency change of 1 Hz that is inaudible as such may now receive a change in level of several dB. This will be audible – however, although the original cause is a frequency change, it is the threshold of the hearing system for amplitude discrimination that is decisive for the detection.

For sine tones of a duration of no less that 0,2 s (sequentially presented via headphones), the **threshold for frequency discrimination** is about 1 Hz in the frequency range below 500 Hz. Above 500 Hz, this threshold is not constant anymore, but about ca. 2‰ of the given frequency. With shorter duration (< 0,2 s), the discrimination threshold deteriorates. These data are averages from a large number of psychoacoustical experiments.

For a sine tone, it is easy to assess whether it ties in with the 1-Hz-criterion, or with the 2‰-critrerion: the limit is at 500 Hz, with a transition from one limit value to the other♣. For sounds comprising several partials, this decision is not so simple anymore. Given an E_2-string, the first 6 partials are below 500 Hz, all further partials are above that limit. In such cases the following holds: frequency changes become audible if for at least one (audible) partial the threshold or frequency discrimination is surpassed. For the E_2-string it thus is not the 1 Hz / 82,4 Hz ≙ 12‰ criterion that forms the basis for the decision but the 2‰-harmonics-criterion. This is a good match to the tolerance range we found in electronic tuners. With the conversion into the unit cent that is customary among musicians, the tolerance range is **3 – 5 cent** (with 1 cent = 1/100th semi-tone interval ≙ 0,58‰). The 1-cent-accurracy that is sometimes demanded is exaggerated: on the guitar, the temperature of the strings would have to be kept constant within 0,1°C (which may be difficult when playing your hot grooves, as cool as they may feel). If the guitar can be tuned with an accuracy of ±2‰, we are on the safe side. This does not mean, though, that every larger deviation will immediately sound out-of-tune. Our hearing system can be quite forgiving and ready to generously compromise in certain individual situations.

8.2.3 Pitch perception

It has already been noted above that pitch and frequency are different quantities. Our auditory system determines the pitch according to complex algorithms – an associated comprehensive discussion would go beyond the scope of this book (specialist literature exists for this). A first important processing step is the frequency/place transformation in the inner ear (cochlea): a travelling wave runs within the helical cochlea, with the wave-maximum depending on amplitude and frequency of the sound wave. Tiny sensory cells react to the movement of this travelling wave; they transmit nerve impulses among various nerve fibers to the brain. The latter performs further advanced processing. A regularly plucked guitar sound consists of a multitude of almost harmonic partials. Round about the first 6 – 8 of these partials result in distinguishable local travelling-wave maxima, the higher partials are processed grouped together.

Normally, we cannot hear the individual partials when a string is plucked. Rather, we hear a complex tone with *one single* pitch. With a little effort, however, these individual partials may be heard, after all. To do this, we first suppress a given partial using a notch-filter, and then switch off the filter-effect so that the original signal is reproduced. From the moment the filter is switched off, the partial in question will be audible for a few seconds, and then merge again with its colleagues to form the integral sound experience that was originally audible. A sufficient level of the partial is a requirement; the partial may not be masked to such an extent by its spectral neighbors that it does not contribute at all to the aural impression. How the single elements are grouped and combined together – that has long been a topic of research for the Gestalt-psychologists. This topic resulted first of all in the **Gestalt laws** for the visual system (see Chapter 8.2.4). In particular, it is the “principle of common fate” that also plays a role in the auditory system if the issue is to group the individual partials of a complex sound event, attributed them to sound sources, and to assign to the latter characteristics such as e.g. a pitch.

♣ Both “1 Hz” and “2‰” are to be taken as approximate values that are subject to individual variations.

As a rule, the pitch recognition works rather well for complex tones with *exactly harmonic* partials – especially if there are lots of partials. However, just like in the visual system with its optical illusions, we know in the auditory area of special sounds that lead to seemingly paradox perceptions. If the partials are not harmonic – as it is the case e.g. for bells – the pitch algorithm develops estimates based on probabilities. Results can be that a subject (test person) cannot decide between two pitches, or that two subjects allocate entirely different pitches to the one and the same sound. Sounds of strings are, however, only mildly in-harmonic, and merely octave confusions are conceivable in the worst case. As a rule, for the **pitch of a string tone** a value is determined that is close to the fundamental but not identical to it. In a first step, the auditory system allocates to all non-masked partials their **spectral pitch**, and on that basis calculates a **spectral rating curve** that has a flat maximum at **around 700 Hz** ♣ – this is the **virtual pitch**. Higher-frequency and lower-frequency partials therefore contribute less to the pitch than middle-frequency components. Experiments carried out by Plomp♥ show that it is – in particular – not the fundamental that defines the perceived pitch. In a piano tone, the frequency of the fundamental was decreased by 3%, while all other partials were increased by 3 %; the result being that the perceived pitch went up by 3%. While the fundamental can have a big influence on the **tone color**, it is rather insignificant for the pitch as long as there are sufficient higher harmonics available.

Now, in the **guitar**, the harmonics are progressively shifted towards higher frequencies (at 1 kHz easily by + 15 cents). If we calculate back the pitch from this, we arrive at a value that is higher than the reading on an **electronic tuner** (measuring merely the fundamental). We should still not retune to make the tuner display 15 cent more – things are more complex. The perceived pitch of the fundamental (or its frequency) is not simply the *n*-th fraction of the frequency of the *n*-th partial: Fastl/Zwicker [12] report of hearing experiments with harmonically complex tones with a perceived pitch *lower* that the objective fundamental frequency. The error of the mentioned electronic tuner would thus tend in the same direction as processing in the hearing system. Moreover, it needs to be considered that the pitch (despite constant frequencies of the partials) is dependent on the **sound level**: as the level increases, the pitch decreases by as much as 5 cents per 10 dB. Even larger effects can be created by **additional sounds** that are superimposed on the guitar sound: literature [12] knows of pitch shifts that can be as large as a semi-tone in extreme cases! Such shifts may not be part of everyday guitar playing, but all in all there is a wide field leaving much space for fundamental research. What also transpires: **cent-exact tuning is not actually possible**. Even though frequencies of individual partials may be measured and adjusted with high precision – it's the hearing system that decides whether the tuning is "correct" … and it will use complicated, situation dependent and even individual criteria. That laboratory experiments indicate that pitch differences of **3 – 5 cent** are recognized does not imply that this accuracy needs to be always observed. It is impossible to specify a mandatory limit for tones hat would be audibly out-of-tune, because too many parameters determine the individual case – but in practice the following rule-of-thumb has proven itself: **a tuning error of no more than 5 cent is desirable**, with 10 cents often being acceptable. Those listeners who have privilege to experience sound through "golden ears" may happily halve these numbers.

♣ Terhardt E.: Pitch, Consonance, and Harmony. JASA 55 (1974), 1061–1069.
♥ Plomp R.: Pitch of complex tones. JASA 41 (1967), 1526–1533.

8.2.4 Grouping of partials

Customarily, string vibrations are described as a sum of differently decaying partials. This "expansion according to harmonic members of the series" is not imperative, but it is the standard tool of spectral analysis – and in fact it derives at least some of its justification from the hydromechanics in the cochlea♣. Even though it is, after all, a model: the tone of a guitar does "consist of" partials. Upon plucking of a string we do, however, not hear a multitude of tones but only *one* tone – so there are **grouping mechanisms** in auditory perception that form groups of connected partials from the spectral pitches (of the non-masked partials), the latter having been gathered on a low processing level. The brain (the human CPU) receives information from the sensory receptors and evaluates it, i.e. reduces this immense flood of data by categorization- and decision-processes. Just as an example: 1,4 million bits of information are contained in just one second of music from a CD! Whether it's 50 bits (per second) that reach our awareness or a few bits more or less: the major portion of the arriving information needs to be jettisoned. But which portion would that be?

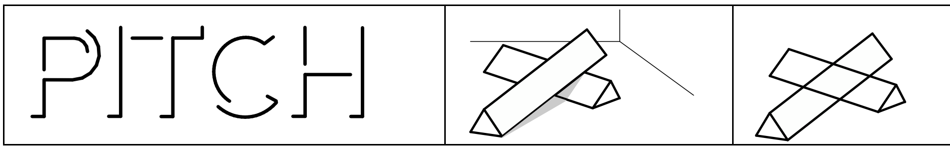

Fig. 8.14: Examples for visual grouping.

On the basis of experiments relating to visual perception, Gestalt-psychologists such as e.g. Max Wertheimer have formulated the **Gestalt laws** that are applicable also in auditory perception. Presumably, the recognition mechanism includes a reduction of the multitude of data delivered by the receptors according to grouping-processes and -patterns already stored in memory. The already-known-and-plausible is given a higher priority compared to the unknown and illogical. The arrangement of two logs of wood shown in the middle section of **Fig. 8.14** can be interpreted three-dimensionally at first glance, even though the drawing plane has merely two dimensions. Some small changes (graph on the right) make the spatial impression all but (or completely) go away. It would go too far to elucidate in detail the principles of closeness, similarity, smooth flow, coherence, and of common fate – the reference to literature in perception psychology must suffice here. As an example that circles back to acoustics, Fig 8.14 shows on the left the word "pitch" represented via an incomplete outline-font. Despite considerable deficits in the picture as such, our visual sensory system succeeds without problem in completing the given lines in a sensible manner, and in giving them a meaning. "Pitch" is captured as a word, and not as a bunch of lines. Perceiving the latter is also possible, though – our visual system is more flexible in this respect compared to our hearing. While it is visually possible to deliberately separate the lines or a grouped object, this is very difficult or even impossible in auditory perception: compared to "pitch" in the figure, it is not at all as simple to switch back and forth between the individual object (the partial) and the grouping (guitar tone). Plucking the string, we hear *one* (musical) tone but find it difficult to pick out individual partials. It may not be entirely impossible but we have serious difficulty doing it compared to separating a read word into its letters and their lines and curves. Insofar there exists a difference between the visual and the auditory processing, but there are also shared characteristics, such as the ability to group, or the hierarchical structure. According to the pitch model by Terhardt, spectral pitches are determined first (in the cochlea) and from these the virtual pitches (on a higher processing level).

♣ Frequency-place-transformation [12] chapter 3.

The processing step on the lowest (peripheral) level of this hierarchy is similar to a short-term Fourier analysis (although with very special parameters). Already on this processing level, partials are sorted out – those, the energy of which is so small that “you wouldn’t recognize if they were missing”. This is because not every partial contributes to the aural impression: if its level is too small compared to its spectral neighbors, it is suppressed (this effect being termed **masking** in psychoacoustics). The partials that are not or only partially masked are given a corresponding spectral pitch each. This pitch will be subject to weighing in the higher processing levels, and synthesized into a virtual pitch. It is no issue in this process if the fundamental of a harmonically complex tone is entirely missing. For example, the telephone – with its band-limitation to 300 – 3400 Hz – is not even able to transmit the first two partials of a male voice (f_G = 120 Hz), but the pitch of the fundamental can still be reconstructed when listening. The perception of a speaking child never appears.

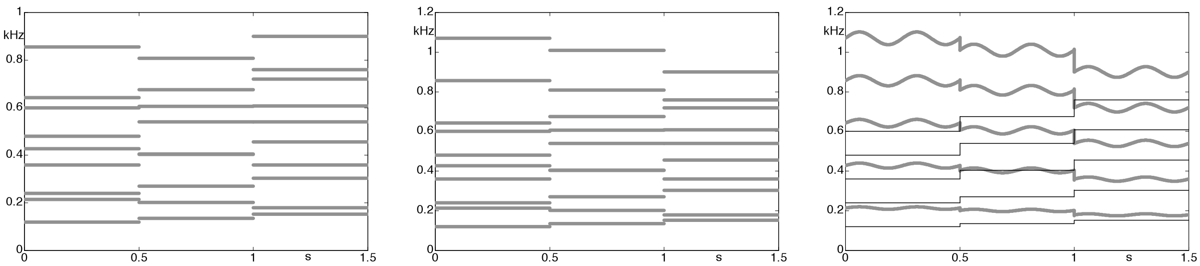

Fig. 8.15: Spectrograms of two tone-sequences: on the right, the descending sequence is frequency-modulated.

One grouping-rule (of several) says that *concurrently* starting sinusoidal tones with an *integer* frequency relation are likely to stem from the same sound source and should be grouped together into one object. Natural sound sources (and only those were available for training the ear during its evolution) almost never generate pure tones. Even if that would occur, it would be extremely improbable that at the same instant several of such sound sources would start to emit sound, and even less likely for them to have an integer frequency relation. If such a *harmonically complex sound* is heard, it can therefore only come from *one* source. Given this, it is purposeful in the sense of information reduction to combine the corresponding spectral lines, just as (optically) the two lines of the letters L, V or T (respectively) are seen as belonging together. The visual signal processing can separate two sup@rimposed letters, and the hearing system can follow *one* speaker – even in the presence of a second concurrent speaker. That does not function perfectly, but still astonishingly well: Chuck’s “long distance information” is clearly intelligible, despite the competing accompanying instruments, and similarly fare “O sole mio” or “We’re singin’ in the rain”. More or less, that is – depending on orchestra/band and singer. The latter may have to push himself quite a bit (or instruct/bribe the sound man conducively) to make sure that the audience (if they listen that closely at all) will not with surprise take cognizance of the fact that “there’s a wino down the road” ☺, rather than that Mssrs. Plant, Page, Jones & Bonham, jr. “wind on down the road” (if they ever play the tune in question again together). Indeed, the grouping of harmonics (and thus their decoding) does not always succeed flawlessly. **Fig. 8.15** gives an idea of difficulties that may occur: on the left we see the spectrogram of a little two-part melody: it is not easy to say which lines belong together. In the figure’s middle section with its larger frequency-span, a formation rule starts to emerge – but only on the right we get some clarity: given different line width and a frequency modulated top voice, the separation becomes easy. The hearing system (especially the musically schooled one) will separate the two voices already without **vibrato** into an ascending and a descending one; with vibrato it comes even more naturally. That would be one reason why singers and soloists often use vibrato: they can be identified more easily among the multitude of accompanying tones. Since the modulation in the soloist-generated sound will run similarly for all partials, the hearing gets help for grouping.

The perceptional psychologist uses the term **law of common fate** in this context: everything that starts concurrently and ends that way, too, “presumably” belongs together. In order to further facilitate the recognition (or the grouping), the soloist chooses a modulation frequency of about 4 – 7 Hz; this is because the hearing system is particularly sensitive for such modulations (fluctuation strength [12]). Accompanying musicians (in the choir or orchestra) also often use vibrato: in part because they just can’t help it anymore, but in particular because that way messy beatings can be avoided that would otherwise automatically arise from playing with several voices. From the "orchestra hacks", however, some restraint is required with respect to vibrato – unless some serious bedevilment is actually called for.

How vibrato will influence the grouping of partials is shown also on the left of **Fig. 8.16**: first, a 100-Hz-tone sounds that is comprised of its 1^{st},2^{nd}, 3^{rd}, 4^{th}, 6^{th}, 7^{th}, 8^{th}, and 9^{th} harmonics. From half the shown time interval, an **additional tone** comes into play in a fifth-relationship (strictly speaking it’s the twelfth) because the 3^{rd}, 6^{th} and 9^{th} harmonics are slightly modulated – the latter now form in a new grouping the 1^{st}, 2^{nd}, and 3^{rd} harmonic of the additional 300-Hz-tone.

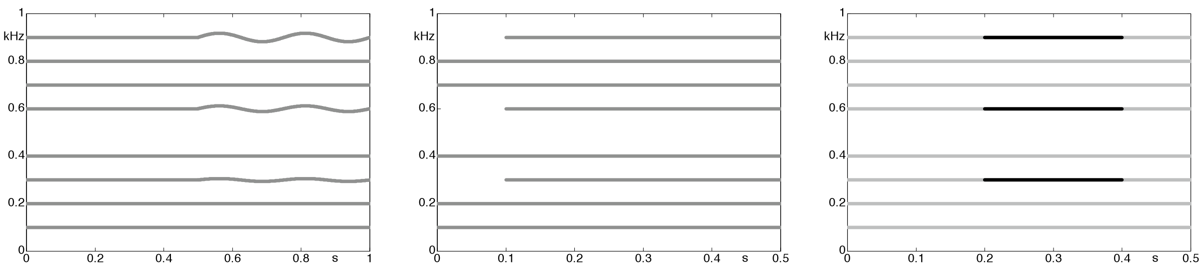

Fig. 8.16: Partial with/of a common fate are grouped to objects.

In the middle section of Fig. 8.16, some partials are started with a delay: first, a 100-Hz-tone sounds, followed by a 300-Hz-tone. However, this happens only if the delay is long enough (e.g. 100 ms). With a delay of about 30 – 50 ms, a sort of initial accent results, with the delayed partials only audible for a short time, as a sort of “livening-up” of the 100-Hz-tone. For an even shorter delay (e.g. 5 ms) this accent looses significance and we hear only *one single* tone. Despite the objective delay, a subjective commonality results that is assigned *one single* common cause.

In the right-hand section of Fig. 8.16 the level of the 3^{rd}, 6^{th}, and 9^{th} harmonic is abruptly changed – indicated by the darker lines. We hear a 100-Hz-tone, and an additional 300-Hz-tone in the time interval between 0,2 – 0,4 s. However, if the levels of the 3^{rd}, 6^{th}, and 9^{th} harmonics are changed continuously, we hear only *one single* tone with a changing tone color. Our experience teaches us that an abrupt change can only stem from a newly introduced object, while slow changes may be attributed to single objects, as well.

The discovery and understanding of the auditory grouping algorithms (here only outlined via a few examples) is not only of interest to musicians and psychoacousticians, but increasingly also to neuro-scientists. Those who seek to immerse themselves into cortical hard- and software find a profound supplement in Manfred Spitzer’s book "Musik im Kopf" [ISBN 3-7945-2427-6] *(translator’s note: this book is apparently only available in German, the translation of the title would be: "Music in the Head".)*

8.2.5 Inharmonicity of partials

Due to the dispersive transversal-wave-propagation, the partials of guitar tones are not strictly harmonic♣, but spread-out spectrally: the frequency of the i^{th} partial is not $i \cdot f_G$, but a bit higher. The analytical connection between bending stiffness and spreading-out of partials has been already discussed in detail in Chapter 1.3 – we will now look at the connected effects on the perceived sound.

In the following analyses, a real guitar signal will be juxtaposed to several synthetic signals. **The real signal** was picked up (without any sound filtering) from the piezo-pickup of an Ovation Viper EA-68 guitar; it was stored in computer memory. For these recordings, the open E_2-string (D'Addario EJ-26, 0.052") was plucked with a plectrum right next to the bridge in fretboard-normal fashion; the first second of decay was used for the psychoacoustic experiments (listening tests). Exponentially decaying sinusoidal oscillations were superimposed and saved as a WAV-file for **the synthetic signal**.

The DFT-analysis of the real signal yielded (with very good precision) the spreading-parameter of b = 1/8000; given this, the frequencies fi of the partials are calculated as:

$$f_i = i \cdot f_G \cdot \sqrt{1 + b \cdot i^2}$$

f_i = frequency of the partial; f_G = frequency of the fundamental.

Fig. 8.17 shows the percentage of frequency-spreading for the spread-out partials; f_i is the abscissa – and not $i \cdot f_G$. On the upper right, the levels of the partials are depicted; on the lower right, we see the time-constants of their decay. With many partials we find in good approximation exponential decay; some partials, however, show strong fluctuations in their envelopes. For the first experiments, these beats were ignored – they were approximated (replaced) via exponential decay.

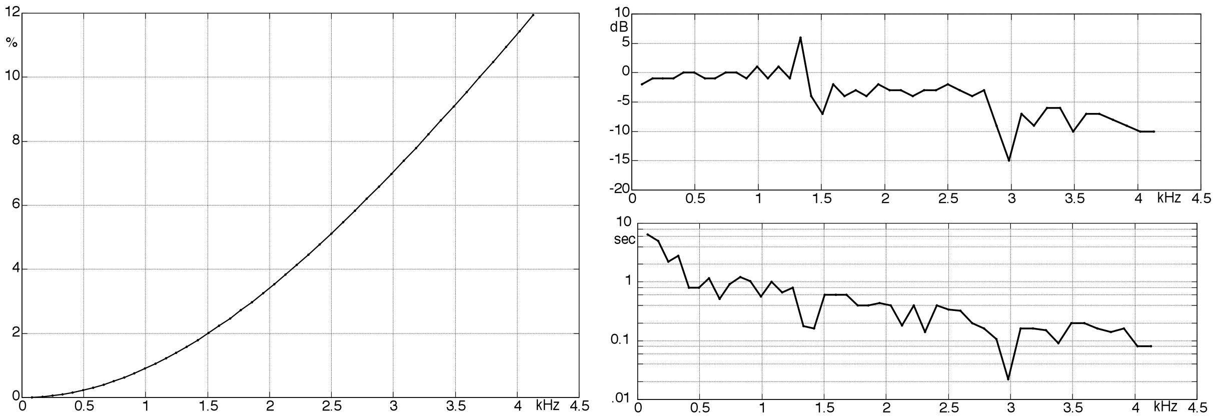

Fig. 8.17: Percentage of spreading-out of partials (left); levels and decay-constants of partials (right).

The data for levels and decay of the partials of the real signals formed the basis for generating the different synthetic signals.

$$u_{\text{synth}} = \sum_i A(i) \cdot e^{-t/\tau(i)} \cdot \sin\left[2\pi \cdot f_i(i) \cdot t + \varphi(i)\right];$$

synthetic signal

♣ Harmonic spectrum: the frequencies of the partials are all in integer ratios relative to each other.

In the formula, A is for the amplitude, τ is for the decay time-constant, f_i is for the spread-out frequency, and φ is for the phase; all these parameters are functions of the order i of the partials. The phases of the partials had not been measured – contrary to the level-spectra, phase-spectra require considerable post-processing in order to obtain graphs that can be reasonably well interpreted.

For a first listening experiment, a synthetic signal was generated that consisted of partials with amplitudes and decay time-constants corresponding to those of the real signal. All phases of the partials were set to zero, though, and the frequencies of the partials were integer multiples of the fundamental frequency (i.e. they were not spread-out). A signal synthesized that way sounds different compared to the real signal. In view of the frequency shifts shown in Fig. 8.17, one might spontaneously consider a difference in pitch – this was in fact indeed noticed during the first listening test. However, the "exact" fundamental frequency of the real signal can – at a signal-duration of 1 s – not be determined with sufficient accuracy; it moreover also changes during the decay (mechanics of the string). Therefore, the synthetic signal was tuned by ear to f_G = 81,9 Hz; the pitch was sufficiently well matched that way. Subsequently, the **essential difference in sound** could be determined via the listening experiment: the synthetic sound was described as "clearer, more buzzing, spatially smaller", while the real sound received the attributes of "more rusteling, more metallic, spatially larger". When presenting the sounds using loudspeakers (broadband speakers, normally reflecting room), an interesting effect with respect to distance could be observed: as the distance to the loudspeaker increased, real and synthetic signals became more and more similar.

The hearing system has no receptor that would analyze the sound pressure arriving in the ear canal with respect to time. Rather, the sound signal is first broken down into spectral bands (called critical bands in this specific context) with a hydro-mechanical filter [12], and is only subsequently recoded into the electrical nerve impulses (action potentials). It is nevertheless purposeful to take a look at the time-functions of the sound signals – at least as long as we do not loose sight of the band-pass-filtering included in the hearing system. **Fig. 8.18** depicts the time-functions of the real signal and of the synthetic signal – they differ considerably.

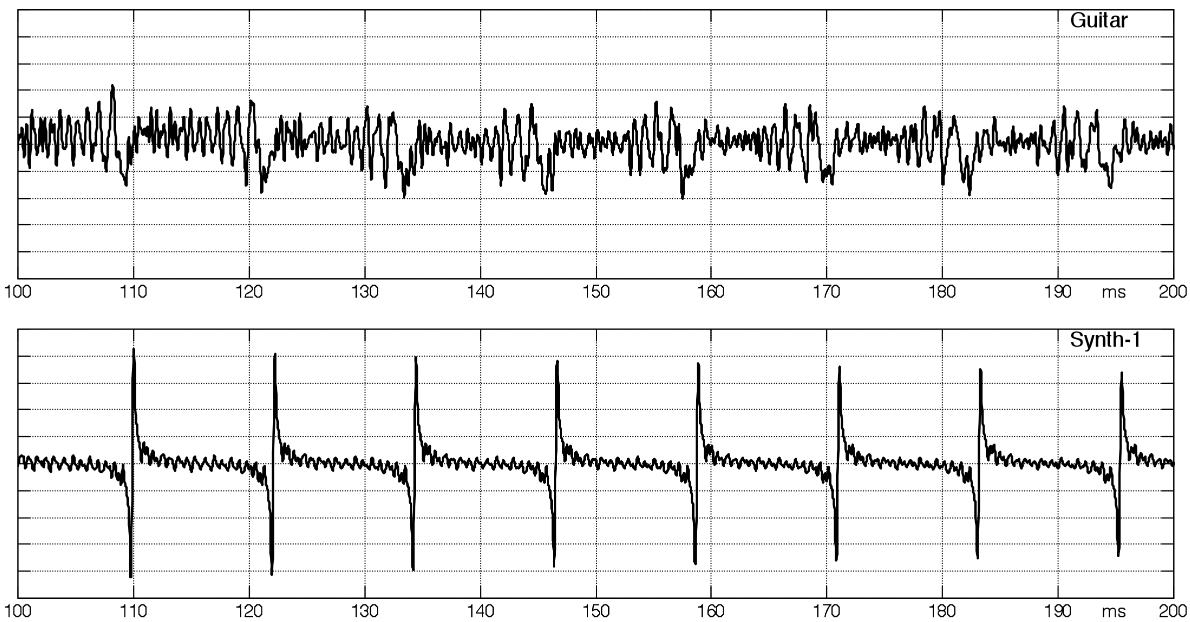

Fig. 8.18: Time-functions of the real signal and of the synthetic signals; E_2-string.

The synthetic signal shown in Fig. 18.8 is periodic while the real signal is not. However, the main difference between the two signals is not found in the periodicity but in the crest-factor (ratio of peak value to RMS value). The considerable content of impulses in the synthetic signal also shows up in a hearing-related spectral analysis (**Fig. 8.19**) as it is generated e.g. in the CORTEX-Software "VIPER": here, we see time represented along the abscissa, and along the ordinate the critical band rate (a non-linear mapping of the frequency as it occurs in the auditory system [12]), scaled in the unit Bark. Coded via the color is a specific excitation quantity derived from the signal filtering as it occurs in the inner ear (i.e. in the cochlea). While the synthetic signal excites the hearing system across the whole signal bandwidth, this synchronicity appears only in the low-frequency range for the real signal. Looking at the pictures it becomes clear why the synthetic signal would be designated “buzzing”, while the attribute “rusteling” is used for the real signal. We can also surmise why the distance between loudspeaker and listener has such a big influence on the sound: given a larger distance, the gaps between the impulses in the synthetic signal are filled with echoes, and it comes closer to the real signal. Evidently, it is not the inharmonicity per se that is so special about the real signal, but the lack of a strictly time-periodic structure featuring a high content of impulses.

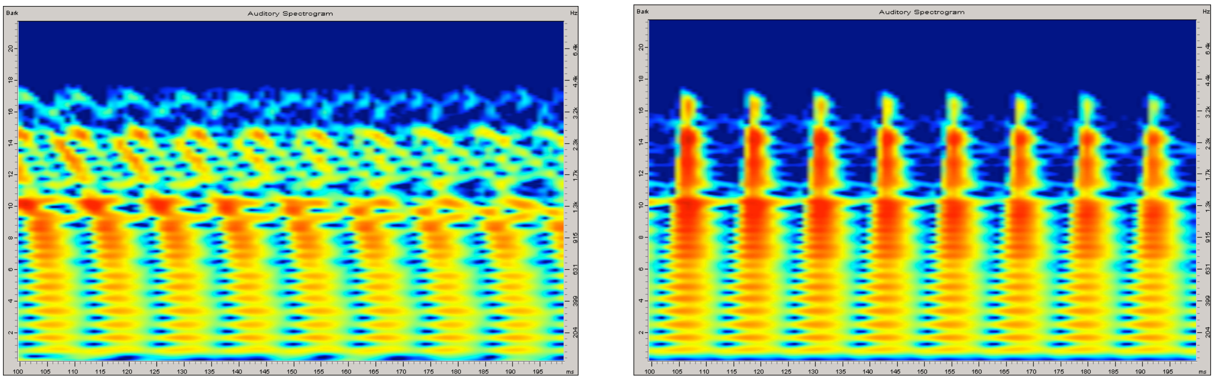

Fig. 8.19: Auditory spectrogram (CORTEX-VIPER), real signal (left), Synth-1 (right).

There is a simple way to check the hypothesis related to impulse-content (or hamonicity): not setting all phases of the partials to zero but having them statistically uniformly distributed yields a so-called pseudo-noise-signal. Due to the strictly harmonic structure of the partials, this signal is periodic, but the wave-shape within one period (in this case amounting to about 12 ms) is of random nature. **Fig. 8.20** shows the auditory spectrogram, and **Fig. 8.21** depicts the time-function. Although this signal (like the Synth-1-signal) does not include the frequency spreading of the real signal, it sounds almost exactly like it. Some test persons with a trained hearing will still detect small differences; in particular in the attack, the signal Synth-2 does not sound as precise.

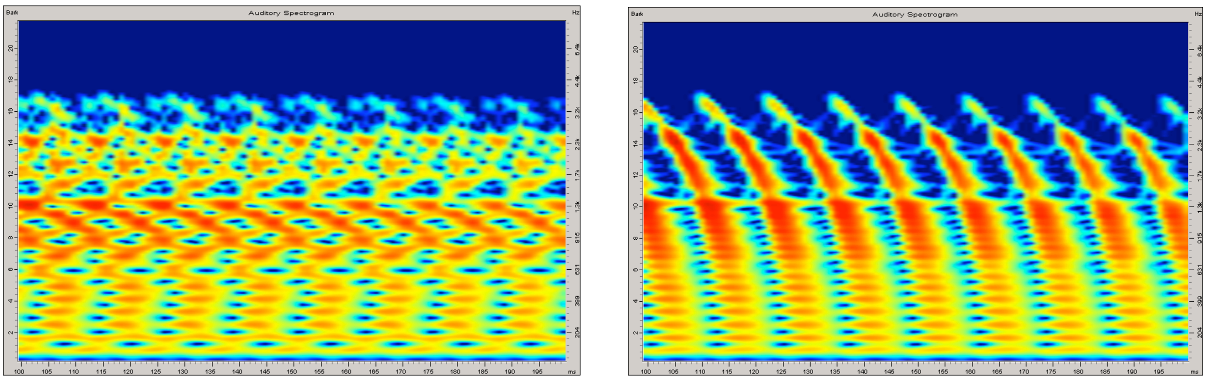

Fig. 8.20: Auditory spectrogram (CORTEX-VIPER), Synth-2 (left), Synth-3 (right).

Still, the difference in sound between the real signal and Synth-1 is much larger that the difference between the real signal and Synth-2. The *rusteling* heard in the real signal is present in Synth-2, as well, but the latter lacks the *buzzing* that is characteristic of Synth-1. Highly discriminating subjects may even hear “a tad too much rusteling” in Synth-2, but most test persons will perceive no difference at all compared to the real signal. An alternative to the equal-distribution phase would be a phase frequency-response suggested by M. R. Schröder♣ that will again guarantee a small crest-factor. The signal designated Synth-3 comprises a harmonic spectrum (i.e. non-spread-out), with the phases of the partials defined according to the following formula:

$$\varphi(i) = i^2 \cdot \pi \cdot 0{,}04\,; \qquad \text{Schröder-phase}$$

Hearing them for the first time, real signal, Synth-2, and Synth-3 differ little; Synth-1, however, sounds distinctly different. Given headphone presentation, a trained ear will notice differences between all four signals, but with presentation via loudspeaker at close distance only Synth-1 sounds different, and for bigger loudspeaker distances, all four signals sound practically the same.

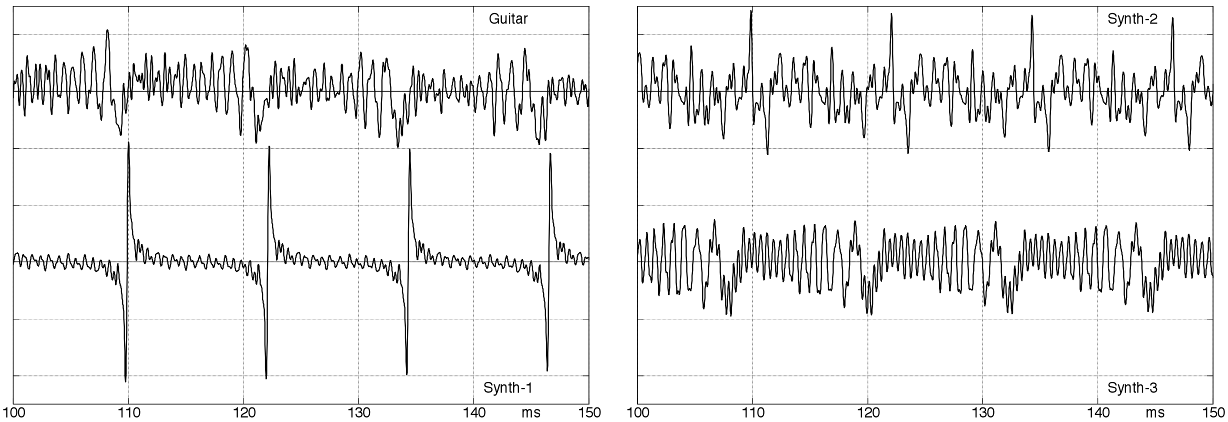

Fig. 8.21: Time functions of the real signal and of the three synthetic signals.

Since all three synthetic signals have identical amplitude spectra but still sound partly similar and partly different, the frequency resolution of the hearing system cannot be of significance in this respect. Exclusive basis for the differences in sound is the difference in the phases – it is only in this parameter that the formulas used for the synthesis distinguish themselves from each other. If one of the signals is transmitted via loudspeaker, the frequencies of the partials do not change, but the phases of the partials do. This bold statement may not be entirely correct from the point of view of signal theory (because a decaying partial is not described by a single frequency but via a continuous spectrum that may well be changed via loudspeaker and room), but it is quite usable as an approximation. The direct evaluation of frequency responses of the phase is, however, of no help: the auditory system does not include a receptor that would a priori determine the phase. Rather, small sensory (hair-) cells within the organ of Corti sense the frequency selective vibration of the basilar membrane. The vibration-envelope of the latter delivers the basis for the auditory sensations of sound-fluctuations and -roughness [12]. The attribute of *buzz* given to the signal Synth-1 is typical for a “rough” sound. Classical psychoacoustics defines **roughness** as the sensation belonging to a fast signal modulation. “Fast” modulations are those with a modulation frequency of between 20 and 200 Hz.

♣ M. R. Schroeder, IEEE Trans. Inf. Theory, 16 (1970), 85-89.

At 82,3 Hz, the frequency distance of the spectral lines of all three synthetic signals is very close to 70 Hz (i.e. the reference frequency for roughness-scaling). However, besides the modulation frequency we need to also evaluate the time-functions of the excitations on adjacent ranges of the basilar membrane: their cross-correlation functions are a kind of weighing-function for the **overall roughness**♣ that is generated from the sectional roughnesses. In Synth-1, all frequency bands are active concurrently – shown in Fig. 8.19 by the fact that the red ranges lie on top of each other (for the same *t*-values). Concurrence is a required condition for roughness. In Synth-2 (Fig. 8.20) the red ranges are dispersed; they appear in the individual frequency bands at different times. This is the reason why the resulting sound is not a buzzing one – but rather one of a rusteling character.

Besides assessing the roughness of the signals, the subjects also judged the perceived size of the sound source. This is a typical phenomenon in perception psychology: while the objective size of the sound event (the dimensions of the loudspeaker) remains unchanged, the **size of the auditory event** varies with the changes in (relative) phase. Synth-1 appears to arrive punctiformly from the middle of the loudspeaker membrane, while Synth-2 seems to be radiated from a range in space. The latter does not appear very big (maybe 10 cm by 10 cm) but is still not punctiform. And something else attracts attention: all sounds except Synth-1 seem to originate from behind the loudspeaker; they have more spatial depth. This impression is created in particular if first Synth-1 is listened to, and then one of the other synthetic signals. An explanation could be that the hearing system is not able to detect any echoes in Synth-1, and interprets the other two synthetic sounds as similar but containing very early echoes. Echoes do lend spaciousness and size, even when arriving from the same direction as the primary sound.

In summary: the frequencies of the partials of a real signal are spread out, but this spreading-out is merely of secondary influence on the pitch. If we compare the real signal with a synthetic one that carries the same partial levels as the real signal but has the partials set harmonically (i.e. not spread out), a very similar aural impression results as long as the phases of the partials are chosen such that the crest-factor does not become too high. If, however, all phases of the partials are set to zero, a different, more buzzing sound results that seems to originate form a *point* in space (for loudspeaker presentation), while all other sounds are perceived to originate from a *range* in space.

Next, the synthesis is modified such that the frequencies of the partials are defined via the spreading formula given above (b = 1/8000). **Synth-4** is a synthetic signal with the frequencies and the level-progressions of the partials corresponding to those of the real signal. Differences exist in the phase of the partials (in Synth-4 these are all at zero), and in the details of the progression of the levels of the partials. As already noted, the partials decaying with beats are replaced in all synthetic signals by exponentially decaying partials. Right off the bat, the inharmonic synthesis is convincing: Synth-4 is barely distinguishable from the real signal even given headphone presentation. And yet, the two time-functions and spectrograms show differences (**Fig. 8.21**) … but this was to be expected: the synthesis is limited to merely 45 partials (f < 4,1 kHz) that all decay with a precisely exponential characteristic.

♣ W. Aures: Ein Berechnungsverfahren der Rauhigkeit, Acustica 58 (1985), 268-281.

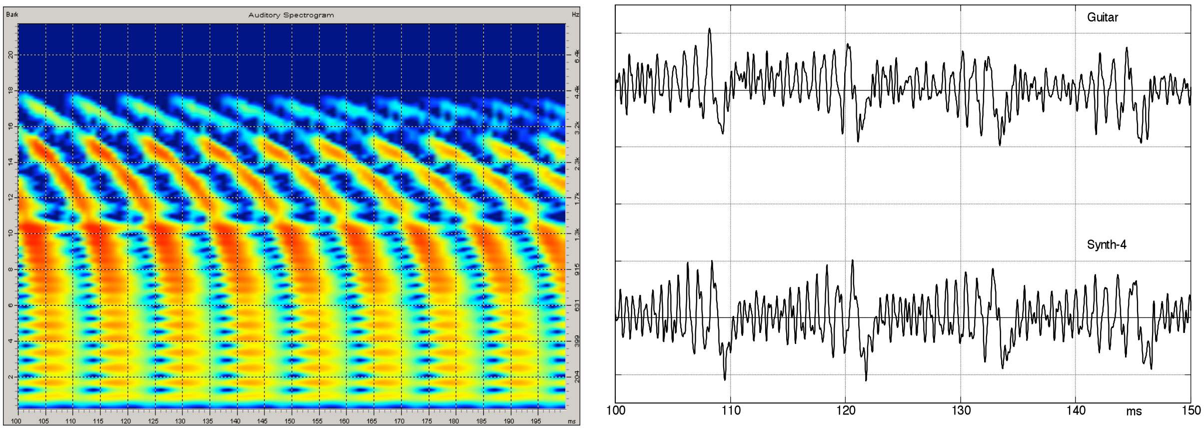

Fig. 8.22: Synthetic signal with spread-out spectrum (Synth-4).

The spreading of the partials leads to a progressing loss of synchronization in the time domain. At the instant of plucking (attack), all partials need to cooperate with equal phase in order to effect the abrupt change in signal. In Synth-1, the attack is repeated in identical shape: the maxima appear at the same times (pan-cochlear **synchronicity**); the tone buzzes. In Synth-2 and Synth-3, this pan-cochlear synchronicity is by and large destroyed, but the period of excitation remains constant in all critical bands. In Synth-4, the period of excitation decreases with increasing frequency, and the cross-correlation function (that the formation of roughness is based on) becomes time-variant. It is no issue that – due to the intra-cochlear time delay of 6 ms (max.) – a true pan-cochlear synchronicity does not actually appear: the hearing system is used to that. All impulses suffer the same fate … and still remain one object.

It is not a matter of course that changes in the **phase spectrum** become audible. If we would repeat the above experiment with a fundamental frequency of 500 Hz, the mentioned phase shifts would still change the time function, but they would not be perceived. It has proven to be purposeful to assume the time-resolution of the auditory system to be about 2 ms: at a fundamental frequency of 82 Hz, the hearing can still "listen into the gaps" but not anymore at 500 Hz. However, apparently a particular sensitivity towards how of the critical-band-specific loudness evolves over time does not exist: Synth-1 is clearly recognized as being different, while Synth-2 and Synth-3 sound very similar despite different cross-correlation functions. It should be noted that this similarity is subject to inter-individual scatter: it may happen that a special sound is perceived as tuned too low. Changing the fundamental frequency (e.g. from 81,9 Hz to 82,3 Hz) removes this discrepancy … now we are in tune. Perfectly, even. A few minutes later, however, the same tone is suddenly too high – and needs to be retuned down to e.g. 81,9 Hz. In the best case, our hearing may notice a frequency difference of 0,2% [12]. It may – doesn't have to. The listening experiments convey the impression as if the attention of the test-person works selectively: sometimes, more attention is paid to pitch, other times roughness is in focus – or other attributes that go beyond the scope of generally understandable adjectives for sound such as "steely", "wiry", "metallic", "rolling", or "sizzling", "swirly", "brown". We seek to describe the remaining difference in the color of the sound somehow, but semantics do let us down here. And then: lets hope that a translation into another language is never needed. Who would think that "**kinzokuseino**" means metallic? Or that "**hakuryokunoaru**" means strong? What does "**namerakadadenai**" sound like? Can you hear "roughness" in there? Or "**r-aow-hig-ka-it**" (to try – and fail – to represent the German word *Rauhigkeit* for this attribute)?

Most partials of the real guitar signal decay in good approximation with an exponential characteristic, but with some we observe a **beating**. The reasons for this shall not be investigated here – we are looking into auditory perception at this point. Already the second partial gives rise to the conjecture that a beating minimum would occur shortly after the end of the recording (duration 2 s), i.e. a beat-periodicity of about 5 s. Within the duration of the listening experiments (1 s), this can still be nicely approximated by an exponential decay, but in the 17th partial there are two beats in combination: a slower one with 1,6 Hz beat-frequency, and a faster one with 18,4 Hz (**Fig. 8.23**). This partial has, however, a low level (in particular relative to the 15th partial), resulting in this beating being practically unperceived – it is masked [12]. For the 27th partial, we find an again different scenario: if features a classical beating with a periodicity of 950 ms. At first glance there seems to be no strong masking: all neighboring partials have similar levels – but they all decay relatively smoothly such that the overall critical-band-level (that is formed from the levels of 4 partials) features almost no fluctuation. The levels of the partials obtained via narrow-band **DFT-analysis** deliver objective signal parameters but do not allow for any conclusion about the audibility of special sound attributes. Psychoacoustical calculation methods such as roughness- or fluctuation-analysis also are to be taken with a pinch of salt: our knowledge about the interaction in inharmonic sounds is still too limited. Listening experiments yield the best results about the audibility of beats in partials – no surprise there, of course. In the case of the above guitar tone, they lead to the clear statement: despite inharmonic partials, beating is practically inaudible.

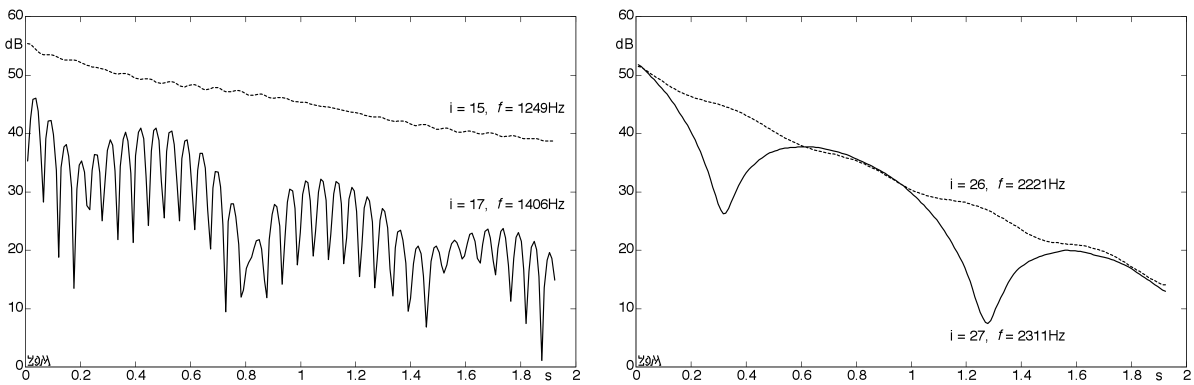

Fig. 8.23: Decay curves of individual levels of partials; Ovation-guitar, piezo pickup.

Still, we must not conclude from the fact that no beats are perceived in the guitar tone presented here that beats are inaudible in general. They are present, and they will be audible if the levels of their partials stand out sufficiently from their spectral neighborhood. Cause for the beats may be found in magnetic fields of pickups (Chapter 4.11), or coupling of modes within the string bearings (Chapter 1.6.2). The inharmonicities of partials, however, can (regarded by themselves) generate only minor fluctuations. Beats within octaves [Plomp, JASA 1967] or time-variant cross-correlations [Aures, Acustica 1985] explain only very subtle fluctuations – partials creating a clearly audible beating require two spectral lines that are in close vicinity, and of similar level. Such lines cannot be generated merely by inharmonicity, though. “In the LTI-system”, we are temped to add in order to have really thought of everything … and we suddenly realize that in particular this limitation is not fulfilled in many cases for guitar amplifiers. Spectral inharmonicity can certainly generate neighboring tones if **non-linearities** are allowed!

In guitar amplifiers, **non-linear distortions** appear to various degrees. While the acoustic guitar amplified via piezo-pickup will usually not be given audible distortion, the contrary might be the case for the electric guitar with its magnetic pickup (depending on musical style). A non-linearity – or, to put it simply, a curved transmission characteristic – enriches the spectrum by additional tones. A mixture of two primary tones (at the input of the nonlinearity)

$$x(t) = \hat{x} \cdot \left[\cos(2\pi \cdot f_1 \cdot t) + \cos(2\pi \cdot f_2 \cdot t) \right]$$

is mapped onto the output signal $y(t)$ via the nonlinear transfer function (in a series expansion)

$$y(t) = v_1 \cdot x(t) + v_2 \cdot x^2(t) + v_3 \cdot x^3(t) + \ldots \; .$$

For purely 2nd- or purely 3rd-order distortion, the spectrum belonging to $y(t)$ may be easily calculated [e.g. 3]. For distortion of any order, the above input signal will create a distortion spectrum that is harmonic relative to the new fundamental frequency $ggt(f_1, f_2)$. The operation $ggt(f_1, f_2)$ determines the *largest common denominator* of the two frequencies f_1 and f_2. Given e.g. f_1 = 500 Hz and f_2 = 600 Hz, a distortion spectrum with spectral lines at the integer multiples of 100 Hz results, while for e.g. f_1 = 510 Hz and f_2 = 610 Hz, a distortion spectrum at integer multiples of 10 Hz is created.

If we generalize the two-tone signal $x(t)$ to an n-tone signal, then the distortion spectrum of the latter will be harmonic relative to a fundamental frequency corresponding to the largest common denominator of all n frequencies of the participating primary tones. If $x(t)$ is a time-periodic signal with the periodicity of T, then its spectrum will be harmonic, i.e. all frequencies of the partials are an integer multiple of f_G = $1/T$. The largest common denominator of all frequencies of the partials is also f_G, and therefore a non-linearity does not change the harmonicity (or the time-periodicity). However, given a spread-out spectrum, a vast variety of new frequencies is created (the root-function is irrational), and these create a noise-like or crepitating additional sound. **Fig. 8.24** depicts the spectrum resulting from a time-periodic signal (Synth-1), and a synthetic signal (similar to Synth-1 but with b = 1/3000), both being fed to a point-symmetric distortion characteristic. In this conglomerate of superimposed primary tones and distortion tones, everything is possible – including beats.

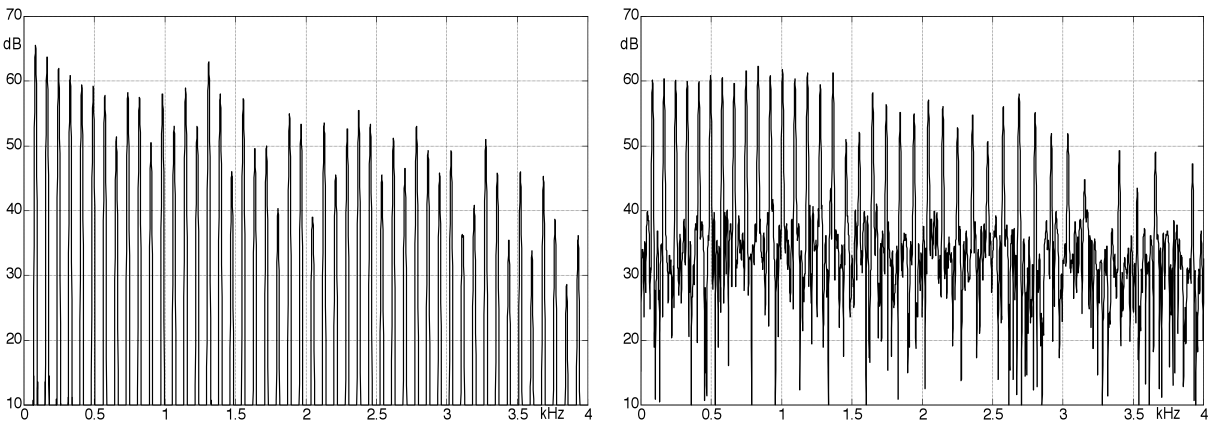

Fig. 8.24: Spectra of signals subjected to non-linear distortion. Left: harmonic primary signal; right: spread-out primary signal. Cf. Chapter 10.8.5.

Conclusion: due to their bending stiffness, strings do not have a harmonic spectrum but a spread-out spectrum; therefore the corresponding time-function is not periodic. If we compare the inharmonic sound with a harmonic sound (of the same fundamental frequency) that features levels of partials at least approximately corresponding to those of the inharmonic sound, we realize that the *phase of the partials* is significant. A harmonic sound carrying partials that all have a phase of zero (or π) sounds buzzing and clearly different from a real guitar sound. However, given a suitable phase function that creates a small crest-factor, harmonic tones that can be synthesized that differ only marginally from a real guitar sound. Using headphones, the trained ear may still recognize differences, but with loudspeaker presentation, the sounds are practically identical. The inharmonicity is clearly noticed only if the spreading parameter b is set significantly above about 1/5000 (this would not be typical for guitar strings). For example, at b = 1/500 a dark chime like that of a wall clock results, while with b = 1/100 synthesizer-like sounds are created. However, if a strongly non-linear system (such as a distortion box) is connected into the signal path, even weakly inharmonic signals may drastically change their spectrum (including additional frequencies) and thus their sound. In such a configuration, harmonic signals experience a change in amplitude and phase only – they remain harmonic.

These statements should be interpreted as results of a small series of experiments and not be generalized to every instrument sound. The aim of these investigations was not to find the absolute threshold for perception of inharmonicity but to demonstrate the rather small significance of guitar-typical inharmonicities. If the decay of higher-order partials is different, inharmonicities based on a much smaller inharmonicity parameter may well be noticed (Järveläinen, JASA 2001).

Compilation of formulas:

$$u_{\text{synth-1}} = \sum_i A(i)\cdot e^{-t/\tau(i)}\cdot\sin\left[2\pi\cdot i\cdot f_G\cdot t\right] \qquad \text{Synth-1}$$

The function of the angle was formulated as a sine in order to not make the crest-factor even larger.

$$u_{\text{synth-2}} = \sum_i A(i)\cdot e^{-t/\tau(i)}\cdot\sin\left[2\pi\cdot i\cdot f_G\cdot t+\varphi(i)\right] \qquad \text{Synth-2}$$

The phase angles $\varphi(i)$ are equally distributed within the interval [0...200°].

$$u_{\text{synth-3}} = \sum_i A(i)\cdot e^{-t/\tau(i)}\cdot\sin\left[2\pi\cdot i\cdot f_G\cdot t+\varphi(i)\right] \qquad \text{Synth-3}$$

The phase angles $\varphi(i)$ are calculated (according to Schröder) as $\varphi(i) = 0{,}04\cdot\pi\cdot i^2$. This corresponds to a group-delay linearly increasing with frequency.

$$u_{\text{synth-4}} = \sum_i A(i)\cdot e^{-t/\tau(i)}\cdot\cos\left[2\pi\cdot f_i(i)\cdot t\right] \qquad \text{Synth-4}$$

The frequencies of the partials are inharmonically spread out.

$$f_i = i\cdot f_G\cdot\sqrt{1+b\cdot i^2}\,; \qquad f_G = 81{,}9\text{ Hz};\quad b = 1/8000;\quad i = 1{:}45;$$
