8 Psychoacoustics

Musical notes are both sound events and auditory events – at least from the point of view of the perceptional psychologist. It is not denied here that these musical sounds may be – in the holistic-philosophical sense – even more than that. The **sound event**: that is the musical sound from the physical perspective. It is characterized by its physical parameters such as e.g. frequency, level, spectrum, or envelope. The investigation of individual physical parameters in isolation will, however, not give any information about the auditory perception: a tone of 40 dB level is audible under normal conditions if its frequency is 1 kHz, but at 50 kHz is will be inaudible. The perceived sound volume (psychoacousticians us the term *loudness*) therefore is not equivalent to the sound level. In the present context, the second syllable –ness in the term *loudness* is intended to indicate that not a physical quantity is meant but one that is connected to the **auditory event**. The science of psychoacoustics describes the functional connections between the parameters of the sound event and those of the auditory event. In other words, this science seeks to e.g. find out how the loudness may be calculated from the physical parameters. Besides the loudness, there are many more parameters (features, attributes) of the auditory event, examples being timbre, subjective duration, or **pitch**. The latter term in particular is often erroneously seen as equal to the frequency. A closer look shows that this is not tenable: the tone of 50 kHz can be precisely defined in terms of frequency, but it cannot be assigned any pitch at all because it is inaudible. The 1-kHz-tone, on the other hand, does generate a pitch, but the latter will be – despite constant frequency – dependent on the sound level, i.e. it is not constant.

Not every audible sound may be assigned a pitch: a voiceless *f*, for example, is perceived as a broadband noise lacking any pitch. Guitar sounds, on the other hand, are characterized by strongly pronounced **pitch** (although there are exceptions here, too). In a simplified consideration, the pitch of the guitar tone is matched to the fundamental of the string. This may lead to the following definition: *the pitch of the A-string amounts to 110 Hz.* However, the unit Hz is for frequencies and not for pitch. So how can we quantify the (subjectively) perceived pitch with sufficient accuracy? A frequently used method would be the comparison with a pure sine tone. The test person (in such experiments usually termed *subject*) alternately listens to the tone to be assessed and to a sine tone, and adjusts the frequency of the sine tone such that both sounds generate the same pitch. The frequency of the sine tone, given in Hz, may now be used as measure for the pitch of the tone to be evaluated. This methodology is sufficiently accurate for the following observations; in scientific explorations, more elaborate procedures are applied, as well. If we have the pitch of the A-string evaluated with the above method, we indeed obtain a frequency of the comparison tone of 110 Hz – but there are small yet significant differences. One reason for the deviations is the dispersion appearing in string oscillations, and the resulting inharmonicity of the partials. Moreover, the interaction of the partials in the perception process also plays a role.

While the first chapters in this book were dedicated to the physical principles of the sound generation, we will now focus on the auditory event. Actually, the guitar is not just a sound generator, but indeed a musical instrument. For extensive presentation and derivation of the fundamentals of psychoacoustics, reference is made to the literature cited in the annex.

8.1 Tonal systems

The strings of the guitar are tuned to E-A-D-G-B-E (for standard tuning). These notes are (inter alia) elements of a musical scale that itself is an element of a tonal system (other expressions used are pitch space, or system of tonality). The latter is understood as the (theoretically unlimited) set of all ordered notes, and includes defining the individual distances between pitches. In Western music, the tonal system with 12 steps is predominant, with the musical scale formed of 12 notes. Deviating from this system are, for example, the pentatonic system (based on merely 5 notes), and the diatonic system. The distance between the notes (the frequency relationships) can be derived from the rules of the tonal system, and from this we obtain the design rules of the guitar, and the tuning rules of the individual strings. In this context, basic knowledge of vibration engineering proves to be helpful.

It is mainly **transversal waves** that propagate on the guitar string; they are reflected at the termination of the free string (nut, bridge). A single-frequency excitation of the string leads to particularly strong vibration patterns at specific frequencies (Eigen-modes at the Eigen-frequencies i.e. natural modes at the natural frequencies). The lowest frequency at which such an **Eigen-mode** occurs is the fundamental frequency of the string. In a simplified view, all higher Eigen-frequencies are integer multiples of this fundamental frequency; a more detailed analysis shows a slight spreading of the frequencies (see Chapter 1).

Fig. 8.1 shows the first three Eigen-modes of an ideal string vibrating in a single-frequency fashion. If the excitation of the string is not with a single frequency but with a plurality of frequencies (e.g. via an impulse), the superposition of many of these Eigen-modes may lead to the formation of a complex vibration-pattern. Each one of the **Eigen-modes** (in theory there is a an infinite number of them) is characterized by four individual parameters: its **Eigen-frequency** that for the n-th Eigen-mode corresponds (in the dispersion-free string) to the n-fold fundamental frequency (n being an integer number); its **amplitude** and **phase**, and its **direction of vibration**. Of these 4 mode-specific quantities, only the frequency shall be considered in the following. Arbitrarily choosing 100 Hz as the fundamental frequency, the frequencies of the higher-order partials ($n > 1$) are 200 Hz, 300 Hz, 400 Hz, etc. Halving the length of the string while maintaining an equal tension-force yields twice the fundamental frequency, with the frequencies of the partials now 200 Hz, 400 Hz, 600 Hz, 800 Hz, etc.

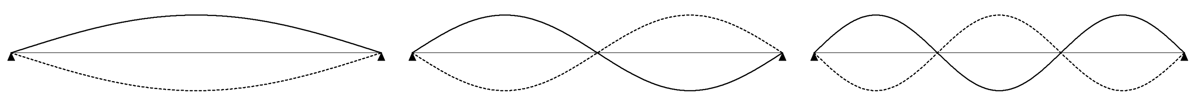


Fig. 8.1: The first three Eigen-modes of an ideal string.
Left: fundamental (1st partial), center: first overtone (2nd harmonic); right: 2nd overtone (3rd partial).

The individual partials do generate individual auditory perceptions in the sense that a multitude of tones becomes audible as a *single* string is plucked. Rather, the pitches of the partials (perceived on a largely subconscious processing plane) blend to form a single **pitch of the string**, with only this pitch being perceived consciously – given favorable conditions. The pitch of the plucked string corresponds *approximately* to the pitch generated by the fundamental vibration, but it is not identical♣. There are small deviations between the two – but for our first basic considerations the deviations shall not be regarded.

♣ The higher-order partials (overtones, $n > 1$) change the pitch of the string only to a minor degree, but they do contribute substantially to the timbre – which is not considered here.

The string mentioned in the example having a fundamental frequency of 100 Hz, and the string shortened by half (fundamental frequency 200 Hz) each generate a tone designated T100 and T200, respectively. Played one after the other in direct comparison, T100 and T200 sound very similar – this is not actually surprising since the frequencies of the partials contained in T200 represent a subset of those contained in T100. This example may be extended by subjecting the halved string (T200) to another halving (T400). The resulting frequencies of the partials (400 Hz, 800 Hz, 1200 Hz, etc.) are again a subset of the frequencies of the partials contained in T100 and T200. Further halving of the string length gives corresponding results. All notes generated by such halving (or doubling) sound very similar, although their pitches differ markedly. Since the frequency relation generated by halving and doubling of the string lengths (2:1 and 1:2, respectively) are designated **octaves** in the musical context, the resulting notes are called **octave-related**. The high degree of auditory relationship between two notes distanced by an octave has led to designating such notes with the same letter. For example, the **reference note** used for tuning to standard ("concert") pitch is internationally as a rule designated A_4, with the note one octave above being designated A_5. However, depending on the national context there are also variations to this system of designations, e.g. a^1 (or a'), and a^2 (or a''), respectively.

8.1.1 The Pythagorean tonal system

Continued halving of the string-length is a first step towards generating related notes of differing fundamental frequency. Following this approach, we find notes with corresponding frequencies of partials also when **reducing the string-length to one third**. The partials of the resulting note (designated T300) are located at 300 Hz, 600 Hz, 900 Hz, 1200Hz, etc. However, compared to T200 now only the frequencies of every other (even-numbered) partial is in correspondence, namely 600 Hz, 1200 Hz, etc. (**Fig. 8.2**). The fundamental frequency of the string reduced to 1/3rd in length relates to the fundamental frequency of the halved string, as would 3:2; this frequency relation (frequency interval) is called, in musical terms, a **fifth**. For the associated notes, the concept of **fifth-relationship** is derived from this. Compared to the octave-relationship, the fifth-relationship is less pronounced.

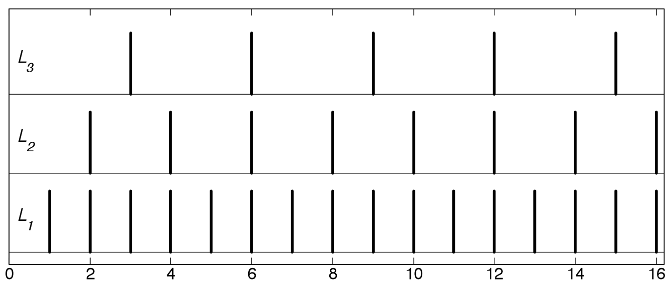


Fig. 8.2: Spectra of partials of strings with the relative lengths: L_1 = 1, L_2 = 1/2, L_3 = 1/3. Abscissa: normalized frequency; ordinate: amplitudes (arbitrary)

Applying jumps of fifths and octaves in combination allows for the generation of a multitude of notes that all are more or less related. Already in the ancient world a tonal system (among many others) was constructed from octave- and fifth- intervals; after its protagonist Pythagoras (ca. 530 B.C.), it is named the **Pythagorean tonal system**. In theory, an infinite number of different notes could be generated with it. However, in practice we arrive at a prominent end point after 12 jumps of one fifth each: after 12 subsequent intervals of one fifth each, the resulting frequency relationship is $1{,}5^{12} = 129{,}74_6$. This brings it close to the 7th octave, the frequency relationship of which amounts to $2^7 = 128$.

The small difference between these two values of 129,74$_6$ / 128 = 1,013$_6$ is called the **Pythagorean comma** in music theory. From the sequencing of fifths, and from octave shifts, all notes of Western music can be generated. In this approach, the frequencies of the notes positioned at a distance of a fifth are shifted by a number of octaves until all frequencies are located within one base octave. Starting from the arbitrarily chosen initial frequency 100 %, the following rounded (!) frequencies result (in order to be able to more easily interpret the frequencies, they are given in % to begin with; the corresponding frequencies are listed in Chapter 8.1.3):

100 –	150 –	225 –	338 –	506 –	759 –	1139 –	1709 –	2563 –	3844 –	5767 –	8650 –	12975	%.
100	150	113	169	127	190	142	107	160	120	180	135	203	%.
C	G	D	A	E	H	F#	C#	G#	D#	A#	E#	B#	

The first line in this table holds the ascending frequencies of fifths, the second line includes the corresponding frequencies in the base octave. The designation of the notes is given in the third line (# stands for 'sharp'). For example, 2563 % needs to be shifted (towards lower frequencies) by four octaves in order to arrive at 160%: $2563 / 2^4 = 160$. Rearranging the frequencies in the second line in monotonously ascending order, the sequence of frequencies of a scale results (values rounded off):

100 –	107 –	113 –	120 –	127 –	135 –	142 –	150 –	160 –	169 –	180 –	190 –	203	frequency / %
C	C#	D	D#	E	E#	F#	G	G#	A	A#	B	B#	note-designation

Besides the ascending sequence of fifths, the descending sequence of fifths may also be generated: again neighboring notes are fifth-related. In correspondence to the example above, the initial frequency 100% would have to be repeatedly divided by 3/2: 67 %, 44 %, etc. With suitable octave shifts (towards higher frequencies), again a scale results – with calculated frequencies that slightly differ from the ones given above, though.

In the classical **Pythagorean tonal system**, not all of the notes calculated above were employed. Starting from the keynote C, users made do with 5 ascending fifths (C-G-D-A-E-B) and one descending fifth (F). They were able to form a **scale** that way:

1	$Q^2/2$	$Q^4/4$	$Q^{-1}\cdot 2$	Q	$Q^3/2$	$Q^5/4$	2
C	**D**	**E**	**F**	**G**	**A**	**B**	**C'**
1\1	8\9	64\81	3\4	2\3	16\27	128\243	1\2

In this table, Q represents the interval of the fifth♣ (frequency ratio 2\3); the corresponding exponent indicates the number of the required jumps of a fifth each. From the denominator, we can take the number of the additionally required octave shifts. $Q^5/4$ indicates 5 fifth-jumps towards higher frequencies, and subsequently 2 octave-shifts ($2^2 = 4$) towards lower frequencies. The third line yields, referenced to the keynote, the frequency relation as a fraction. The notes of the scale given above, and their frequency relation (interval), is designated according to their place number:
C = prime, D = second, E = third, F = fourth, G = fifth, A = sixth, H = seventh, C' = octave.

♣ To specify the frequency relations in an **interval-designation**, two different styles are customary: for the fifth e.g. 2:3 but also 3:2. Both relations are self-explanatory, while the letter-designation (C-G) does not unambiguously identify which one of the two is the lower note. In the following, the lower note is always positioned first (to the left) as is usual for axis-scaling. However, following through with this train of thought would result in fractions that are smaller than 1, such as e.g. $f_{C1} : f_{G1} = 2:3 = 0{,}666...$ While this representation is in itself correct, it is in contradiction with the practice of indicating intervals with number that are larger than 1. This contradiction is resolved in the following via using the back-slash (as used in Matlab): $f_{C1} \backslash f_{G1} = 2\backslash 3 = 1{,}5$.

The terms are related to numeration in Latin: primus, sekundus, tertius, quartus, etc. In its precise meaning according to the theory of harmony, these expressions designate the *distances* between two notes (inter-vallum = space between palisade beams), but in everyday use they also represent the names of notes: *the fourth on the C-scale is an F*. Distance in the above sense means to indicate the distance to the root note i.e. the ratio of the frequency of the note in question (e.g. an F) to the frequency of the keynote; in this example it is 3\4, corresponding to a fourth. It is also possible to form the ratio of two notes directly neighboring on the scale; this yields:

$f_C \backslash f_D = 8\backslash 9$; $f_D \backslash f_E = 8\backslash 9$; $f_E \backslash f_F = \mathrm{HT}$; $f_F \backslash f_G = 8\backslash 9$; $f_G \backslash f_A = 8\backslash 9$; $f_A \backslash f_B = 8\backslash 9$; $f_B \backslash f_{C'} = \mathrm{HT}$;

Of these 7 frequency ratios, 5 correspond to a so-called **whole-step** (‘whole note’, ‘whole tone’), specifically C-D, D-E, F-G, G-A, A-H. The remaining two intervals of neighboring notes are **half-steps** (HT, ‘half notes’, ‘semi-tones’, ‘half-tones’). In Pythagorean tuning, the frequency ratio in a whole step amounts to 8\9 = 1,125, and the one in a half step (E-F, B-C) **HT** = 243\256 = 1,053$_5$. The resulting scale is called **diatonic scale** because it is comprised of two different steps (namely whole-step and half-step). As supplemental information, ‘Pythagorean tuning’ should be indicated – there are many different tunings, after all.

N.B.: with respect to the **note that is internationally designated B**, there is a particular idiosyncrasy when the German language is used: there, this note is designated **H**. Originally (in fact: obviously), letters (starting with A) formed the names of the notes in the scale: A-B-C-D-E-F-G. However, medieval hexachord theory required (on top of the B as mentioned above) a second note half a step lower. In order to distinguish between the two, the designations *B-quadratum (B-durum)* and *B-rotundum (B-molle)* were introduced – derived from the angular (hard, *durum*) and round (soft, *molle*) writing styles of the letter *b*. The angular b mutated to an h … and now musicians in Germany, Austria, and the German speaking part of Switzerland found themselves with a peculiarity that continues to lead to (sometimes serious) complications when communicating internationally.

The diatonic scale as introduced above consists of 5 whole-steps and 2 half-steps. Each one of the whole-steps can pythagoreically be subdivided into two half-steps – however this may be done in two different ways. In the international note designations, half a step upwards is indicated with adding the syllable “sharp” to the note, and half a step downwards by adding the syllable “flat”. The diminished D is called **D-flat** (D*b*, with the *b* standing for ‘diminished’), the augmented C is **C-sharp** (C#). It has already been shown that all notes can be generated by using upwards-fifths and downwards-octaves in the Pythagorean sense:
C–G–D–A–E–B–F#–C#–G#–D#–A#–E#–B#.
However, all notes may just as well be generated via downward-fifths and upward-octaves:
C–F–B*b*–Eb–A*b*–D*b*–G*b*–C*b*–Fb–B*bb*–E*bb*–A*bb*–D*bb*.
The notes B*bb*, E*bb*, A*bb* and D*bb* result from diminishing B, E, A, D by *two* half-steps, respectively.

Fig. 8.3 shows the keynote frequencies of these two Pythagorean-chromatic scales. Due to the Pythagorean comma, no frequencies in a pair in the sequence of upwards-fifths and downwards-fifths are the same (except for the starting pair). If we limit ourselves to diminishing by a *single* half-step, a scale of 21 steps results: each of the 7 diatonic steps C-D-E-F-G-A-B is allocated a lower and a higher half-step. This 21-note tonal system was actually the basis for keyboard instruments – however it was deemed too complex.

Many musicians therefore simplified the scale by enharmonically equating similar notes. The resulting **12-step Pythagorean-chromatic scale** is indicated on the top of Fig. 8.3 via squares. Only a single half-step is introduced each between all whole-steps, but the half-tone distances are of different size, as is clearly visible (□—□).

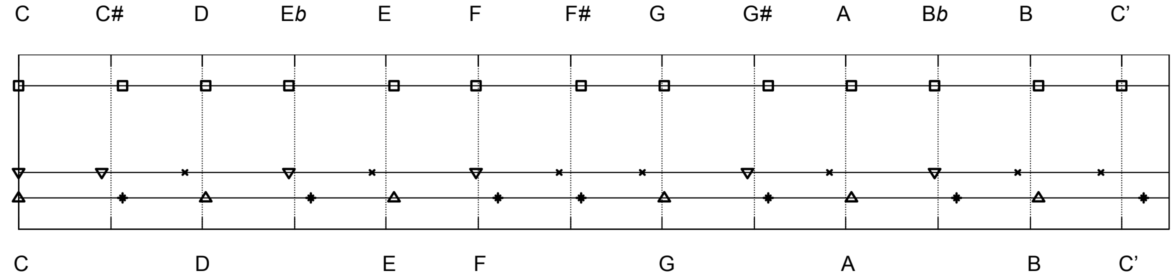


Fig. 8.3: Fundamental frequencies of the Pythagorean-chromatic scale, shown on a logarithmic frequency axis.
∆ = deduced from the first 6 upwards-fifths jumps; ∇ = deduced from the first 6 downwards-fifths jumps;
×,∗ = remaining 7 fifths jumps; □ = used in medieval times as chromatic scale.
The scale with equal temperament developed around 1700 is indicated with dashed vertical lines (8.1.3).

The different half-step distances complicate changing keys: the second (C-D) based on the keynote C has a larger frequency difference that the one based on C# (C#-E*b*), and other intervals (e.g. C-E, G#-C) meet a similar fate. Depending on the specific case, the flawed consonance when two notes are played simultaneously may be another problem. The fundamental thought behind the Pythagorean tuning was the note-relationship based on fifths and derived from the sequence of partials. Well meant that is – but you know how things are with relatives: as the distance grows, the similarities wane. **Fig. 8.4** schematically shows the frequencies of the partials for the prime (C) and the third (E). If, in simultaneous playing of the two notes, individual partials get to lie (frequency-wise) in immediate vicinity, **beats** may become audible. An example would be the 5th partial of the prime (C) and the 4th partial of the Pythagorean third (E_p).

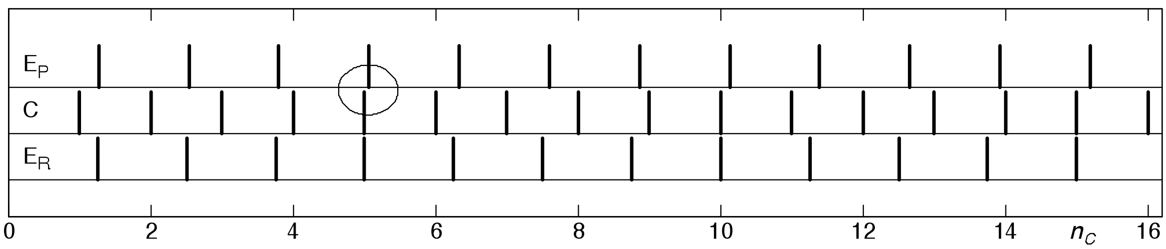


Fig. 8.4: Spectrum of partials of the notes C (prime) and E (third). Beats are generated between the 5th partial of the prime and the 4th partial of the Pythagorean third (E_P), due to the small frequency difference. For the pure third (E_R) the corresponding frequencies of the partial are identical. Abscissa: normalized frequency of the partials of the prime.

Beating happens when two mono-frequent notes of equal amplitude and similar frequency are played at the same time (i.e. they are added). Every note from a guitar consists of a multitude of (mono-frequent) partials, each of which is, individually considered, sine-shaped (a cosine-oscillation has the *shape* of a sine, as well). The 5th partial (= 4th overtone) of an ideal string vibrating at 100 Hz has the frequency of 100 Hz x 5 = 500 Hz, the 4th partial of the third according to Pythagorean tuning is at 126 Hz x 4 = 504 Hz. The frequency difference of the two partials is 4 Hz.

If we now regard merely the oscillation of the sum of the two partials, a figure similar to **Fig. 8.5** (center) results. The phase difference of the two partials fluctuates with the rhythm of the difference frequency, and amplification and cancellation alternate with the same rhythm. Given sufficient levels, *one single* partial with rhythmically fluctuating (i.e. beating) loudness is heard rather than two partials of almost equal pitch.

Interpreting the summation-curve (middle section of Fig. 8.5) is facilitated by reformulation towards a multiplicative operation:

$$\cos(2\pi f_1 t) + \cos(2\pi f_2 t) = 2 \cdot \cos(2\pi f_\Delta t) \cdot \cos(2\pi f_\Sigma t); \quad f_\Delta = \frac{f_2 - f_1}{2}; \quad f_\Sigma = \frac{f_2 + f_1}{2}$$

In this product-representation, f_Σ stands for the frequency of a cosine-oscillation with its amplitude changing "with the rhythm of the difference frequency f_Δ". The above example has f_Σ = 502 Hz, thus it lies exactly in between the primary frequencies f_1 und f_2. The term "difference frequency" should be used with care: it is calculated as f_Δ = 2 Hz, this is *half* the frequency distance between f_1 und f_2. However, the maximum of the beat-envelope appears (amount!) with double this frequency i.e. twice per f_Δ-period. The above beating with 500 Hz and 504 Hz as primary frequencies may therefore be seen as a tone at 502 Hz featuring 4 envelope maxima and 4 envelope minima per second. It therefore becomes louder and softer 4 times per second. The auditory effect of a beating of partials is difficult to predict – it may even be inaudible (despite its physical presence) due to masking by neighboring frequency components. If it indeed is audible, it may sound pleasant or displeasing. During many centuries the opinion was held that any beating of partials is undesirable, resulting in the beat-free **just intonation** (Chapter 8.1.2).

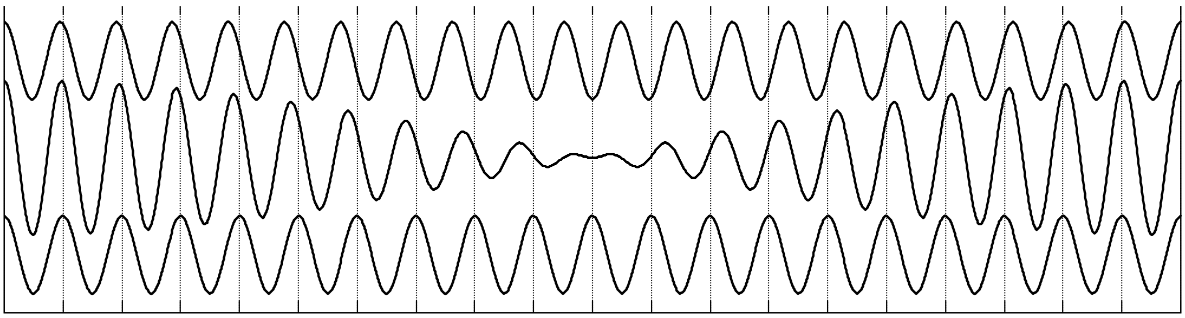


Fig. 8.5: Two cosine oscillations (top, bottom) slightly different (5%) in frequency, and their sum (middle). The curves are of equal phase at the left and right boundaries of the figure, and in opposite phase in the middle. Same-phase addition results in doubling of the amplitude (constructive interference), opposite-phase addition leads to cancellation (destructive interference). Abscissa: time.

8.1.2 Just intonation

In this context, *harmonic* and *natural* also stand as synonyms for *just* – the rationale being that nature herself allegedly had shown the way in the form of integer frequency ratios of the partials. The term *divine tuning* therefore is not far off, creating work for philosophers and esoterics, but mainly for mathematicians ... who not necessarily were musicians.

Just intonation – the term rings of *teachings of justice & purity*, and expressions such as *fairness*, *correctness*, or *well justified* come to mind – the opposite of *unjust*, *wrong*, or *unjustified*, and thus anything not conforming to the *just intonation* could in any case only be heresy. It is easily imaginable how hordes of mathematicians have deduced justifications for this or for that intonation … generating tables with an accuracy up to 12 figures! Or, rather: tables with 12 decimals, since the actual accuracy may have been a bit of an issue [Barbour]. Irrespective of any (not infrequently occurring) calculation- and rounding-errors: given a 1-m-long monochord string, 12-decimal-accuracy implies a length-tolerance of no more than 0,001 nm. Just to compare: the wavelength of visible light amounts to around 600 nm. Specifying pitch deviations with an "accuracy" of 1/10000000th of a cent is similarly nonsensical.

The **just intonation** may be traced back to ancient times. Two doctrines of thought emerged from the Pythagorean school (that originated around 530 BC): the **canons regular** (canon = rule, law) advocated the conservative opinion, while the **harmonists** gave priority to euphony, even if that required modification of mathematical laws of nature. The canonical doctrine regarded the frequency ratio 6 : 8 : 9 : 12 as "holy matrimony" between the forth and the fifth (**Fig. 8.6**) with the major second (full step F-G) being the result. Simbriger/Zehelein give an astounding assessment for this approach: *we have already met this grouping of notes in primitive music; with the Pythagoreans, we find that same basic occurrence substantiated and sanctioned with the background of advanced civilization.* There you have it: if – as a musician or listener – you recognize certain intervals as harmonic/consonant, then that's primitive ado. However, if you smudge some divine-cosmic-mystical mumbo-jumbo around that finding, it takes its place in high culture.

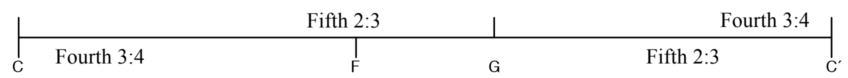


Fig. 8.6: The "holy matrimony"

Still: despite some massive mystical sanctioning it was not possible to hide that the use of Pythagorean intonation made some chords sound less than pleasant. Young J.-apprentice: "oh honorable master Y.: them chords, they will not sound – try as I might! Those fifths and thirds, they fail to soothe us." Y.: "Do or do not: there is not try … but quiet now be, young one; in a special realm here taken we are. Let be it, for divine this is – of The Force" ☺. Many will have conformed to this sage advice from a long time ago and a galaxy far, far away … but some went public. In the olden days, on this planet, that could well lead to premature termination under artificially elevated ambient temperature – or it could open the door to eternal fame and glory. Or both. **Didymos** (Didymus) and **Ptolemy**, Alexandrinian savants by trade (and, to begin with, both by all means proponents of the Pythagorean third), evidently found the silver bullet (at the time probably the silver arrow). They replaced the Pythagorean third (based on the divine fifth) by an at-least-as-divine relation of whole steps: the major third – in Pythagorean intonation the frequency interval 64\81 = 1,2656 – was shifted to 4\5 = 1,2500 in the so-called Alexandrinian system. Didymos borrowed the minor third (27\32 = 1,1852) from the Pythagorean system, and Ptolemy modified it to 5\6 = 1,2000. In principle, anyway. Looking closer, we find [e.g. in Barbour] two didymian intonations, and no less than 7 ptolemyan intonations. Nevertheless, the foundation block for the just intonation was laid.

Studying literature, it is easy to come to the impression that (as mentioned above) something divine is connected to the just intonation. However, as confusion grows, the realization does manifest itself that it must in fact be a kind of polytheism. Barbour defines *just intonation* as: based on octave (1\2), fifth (2\3) and major third (4\5); the intervals themselves are designated *just* (or *pure*), as well.

Elsewhere, however, Barbour extends the term *just intonation* to: based on octave (1\2), fifth (2\3), fourth (3\4), major third (4\5), and minor third (5\6). Other authors even designate as *pure intervals* all intervals the frequency ratios of which correspond to the whole-numbered ratios of the frequencies of the first 16 partials. All intervals? Well, almost … those ratios that fit to some degree, anyway. But not the 7th , 11th , 13th , and 14th partials! Of course not. Valentin substantiates: *the miraculous, natural, and therefore not worked-out order of the whole system stems from the sequence of the composition of just intervals contained in these notes that – with a suitable octave transposition – yield our whole scale system.* The 7th, 11th, 13th, and 14th partials are the "black sheep"; nature finds space for something like that, too. Only for C-F# (or C-G*b*) no fitting frequency ratio at all could be found in the natural order. Therefore, the devil had to be called in as the usual suspect – only he/she could have smuggled in such an inconvenient, devilish interval (Tritonus, Diabolus in Musica). The question: "how could God allow this …" again created many workplaces for philosophers (compare Theodizee), but this would go beyond the scope of scientific considerations.

The just intonation derives its rationale from the whole-numbered frequency ratios of the first 16 partials. But why exactly 16 partials? That's because the 16th partial is exactly 4 octaves above the fundamental. But why then not just 1 or 2 or 3 octaves? That would be because that way you could not yet generate a chromatic scale. Moreover, wind instruments can just about reproduce the 16 "natural tones" (Eigen-tones, partials). The peculiarity of the tritone with its 45\64-ratio was justified on the basis of this fact that about 16 but not those 64 Eigen-tones could be generated. **Fig. 8.7** shows the frequency ratios of a just-intoned scale. Besides the devil's interval, there indeed is nothing fishy in there: numerators and denominators are integers between 1 and 16. The **major third** C-E that would with the Pythagorean intonation carry beats – it now is beat-free (compare to Fig. 8.4).

1 15\16 8\9 5\6 4\5 3\4 45\64 2\3 5\8 3\5 5\9 8\15 1\2
C – D*b* – D – E*b* – E – F – G*b* – G – A*b* – A – B*b* – B – C'

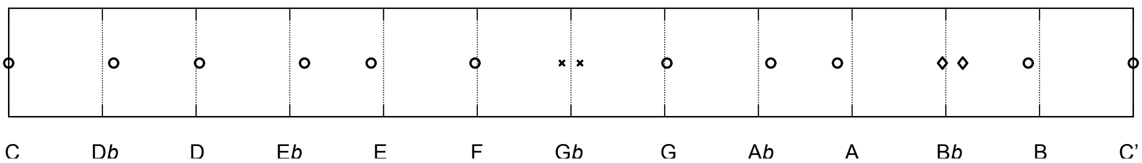


Fig. 8.7: Just intonation (Mersenne's lute tuning Nr. 2). The tritone was given also as F# with 32\45, for the Bb also 9\16 are found instead of 5\9.

Besides C-E, the combinations F-A and G-B (with 4\5) also make for a beat-free major-third interval. For the **minor thirds**, however, differences already appear now: E-G, A-C, and B-D yield 5\6, but D-F yields 27\32. Looking at the **fifth**-intervals: C-G, E-B, F-C, G-D and A-E yield 2\3, but D-A → 27\40. The **whole-step** intervals are at 8\9 or 9\10; the **half-step** intervals on the C-major scale are at 15\16, with the remaining (chromatic) half-step intervals at 24\25, 25\27 or 128\135. Despite the legitimization by nature herself, this gave opportunities for mockers: are you still learning, or do you play with a special intonation system?

It wasn't that these dissonances remained hidden to the working musicians. The latter knew about them, limited their music-making to a few keys, and tried to give a wide berth to the *howling-wolf intervals.* Alternatively, instruments could be built that divided every octave into 21 in-between notes. And if that didn't suffice: J. M. Barbour lists a plethora of other divisions, for example: the 31-division (Fibonacci-sequence), the 53-division (Bosanquet-harmonium), and don't you forget that *the 118-division has both fifths and thirds that are superlative (0,5 cent flat and 0,2 cent sharp, respectively)*.

There we are! … and there we go. On top of all that, other just (!) intonations were developed for the twelve-section octave, as well – which makes Barbour infer: **<u>the</u> just intonation does in fact not exist; rather, there are many different just intonations, with the best being the one that comes closest to the Pythagorean intonation**.

As desirable as “just” (or “pure”) intervals may be in polychoral play: for intervals succeeding each other errors do cumulate. Take, for example (and see **Fig. 8.8**), Jimi Hendrix’ “Hey Joe” (to solidly arrive back again in more modern times): the accompaniment first climbs down a minor third from E to C (or that could be interpreted as climbing up a minor sixth) and then runs through 4 jumps of fourths: E → C – G – D – A – E. Given just intervals, one revolution gets us this:

$$\frac{4 \cdot 3 \cdot 3 \cdot 3 \cdot 3}{5 \cdot (2 \cdot 2) \cdot 2 \cdot (2 \cdot 2) \cdot 2} = \frac{324}{320} = 1{,}0125 \ \hat{=}\ +21{,}5\,\text{cent}\,; \quad \text{C} \downarrow \text{E} = 5\backslash 4,\ \text{fifth} = 2\backslash 3,\ \text{octave jump} = 2\backslash 1.$$

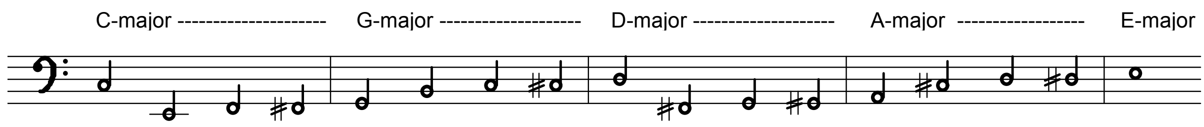


Fig. 8.8: Jimi Hendrix / Noel Redding: bass chromatic in "Hey Joe".

On the basis of just intervals, the full revolution of a cadence (lasting about 12 seconds in the original tempo) would lead to a frequency increase of 1,25% – after one minute, that would already make for no less that half a step. To execute every revolution at exactly equal pitch, e.g. the step from D to A (fifth) would have to be performed with the deviating ratio of 27\40. That, however, would mean a conflict with the pure (just) school of the first 16 natural notes.

Another “law of nature” (one that chalked up some success in architecture) is the **golden section** (or **ratio**). However, for Barbour the “golden tonal system of theoretical acoustics” is worth only a few lines. His bottom line: a jack-o’-lantern (ignis fatuus).

8.1.3 Tempered tunings

In music, the term temperament is used synonymously with the term tuning. **Tempered tuning** is no pleonasm, though. It is the technical term for tunings that, on a small scale and in a targeted manner, deviate from global tuning rules. Early versions of the tempered tuning may be traced back to Giovanni Maria Lanfranco (1533); starting from just intonation, he proposed to slightly down-tune the fifths, and to up-tune the thirds just as much as was tolerable (in terms of the perceived sound). During subsequent eras, there were countless attempts to define this advice more precisely. Starting from empirical results (the fifth should cause one beat per second), via graphical designs, nomograms, scary formula, and versatile tables, the path led to the equal-temperament tuning that dominates today: the octave is divided into 12 equidistant half-note-steps – and that’s it. That this seemingly simple rule has not been in practice much longer – that is probably due to its demanding a readiness to compromise. It does require, after all, detuning those just and highly consonant intervals (such as the fifth). Not all musicians show a corresponding capacity for suffering: the cellist Pablo Casals speaks of the *brainwashing of the tempered tuning*, and the violinist Carl Flesch allegedly was unable to play together with a piano (in tempered tuning) subsequent to a rehearsal with a string quartet.

Well, with direct access to the string and thus the pitch as a continuum, a violinist has the advantage of a completely free-wheeling intonation. The piano does not offer this possibility. If a growth of the number of keys into the infinite is to be avoided, the only remaining solutions are a highly key-specific temperament, or a universal equal-beating (equal tempered) tuning.

Intervals between notes are characterized by the corresponding frequency ratios. Using the equal temperament, 12 similar half steps succeed one another within an octave, with a geometric frequency sequence resulting: 1, HS, $HS{\cdot}HS$, $HS{\cdot}HS{\cdot}HS$, etc. Here, HS indicates the half-step interval, the 12-fold repetition of which yields the just (pure) octave: $HS^{12} = 2$. With this, the frequency relation of directly neighboring notes (at half a step distance each) calculates as:

$$HS = \sqrt[12]{2} = 1{,}059463...$$ Half-step interval in the equal temperament

The 12th root of 2 … that is an irrational number. In the actual sense of the word it is a number opposing reason. That may also be why a queasy feeling crept up on many a music-theorist. 3\4 for the just-intonated fourth is specified by nature herself; the counterpart in equal temperament, on the other hand, defies – with HS^5 – all sanity. And yet the numerical differences are not all that big: 3\4 = 1,33333... , HT^5 = 1,33484... , that's a gap of no more than merely 0,1%. However when principles are at stake, the gods themselves fight in vain. And sure: the differences may indeed be larger for other intervals. The following table lists all notes and frequency ratios in the equal-temperament scale. Other notes are not defined i.e. there is no distinction between C#/D*b*, E#/F, A*bb*/G, B/C*b*, and so on.

C	C#=D*b*	D	D#=E*b*	E	F	F#=G*b*	G	G#=A*b*	A	B*b*	B	C'
0	1	2	3	4	5	6	7	8	9	10	11	12
1	1,0595	1,1225	1,1892	1,2599	1,3348	1,4142	1,4983	1,5874	1,6818	1,7818	1,8877	2

Table: Notes and frequency ratios in equal-tempered tuning. The second line yields the half-note steps, the third yields the frequency ratios rounded to 4 decimal places. Reference = C.

In German-speaking lands, the term *gleichschwebend* (= with equal beating) could be misinterpreted such that all intervals would cause similar beating. This is not the case. The English designation EQUAL TEMPERAMENT is not self-explanatory, either. It is the half-note steps that are equal (in terms of the frequency ratios), and not the beats. Also equal (in the sense of relatively equal) is the distribution of the Pythagorean comma into all 12 jumps of fifths. Occasionally, *well-tempered* is found as a synonym for *equal tempered*; this can probably be traced back to **J. S. Bach's** preludes and fugues that he published under the title "The Well-Tempered Clavier". However, presumably Bach's instruments were not intonated with equal temperament (equal beats), but according to Werckmeister. Andreas **Werckmeister** (*Musikalische Temperatur*, 1691) had developed a tuning that comes close to the equal-temperament tuning but is not identical. Already one century earlier (around 1596), Simon **Stevin** had built a monochord the half-step frequency ratio of which corresponded to the 12th root of 2 (i.e. 1,059…). Presumably this was the first such instrument in Europe [Barbour]. Almost at the same time (around 1636), Marin **Mersenne**♣ carried out comprehensive theoretical groundwork.

♣ 1492 Franchinus Gafurius: Theorica musicae
1511 Arnolt Schlick: Book on organ-building
1533 Giovanni Lanfranco: Scintille di Musica
1544 Michael Stifel: Arithmetica integra, z.B. log
1596 Simon Stevin: Monochord mit HT = $2^{1/12}$
1636 Marin Mersenne: Harmonie universelle
1691 Andreas Werckmeister, Musicalische Temperatur
1706 Johann Neidhardt: Gleichschweb. Temp.

In his chapter *Equal Temperament*, Barbour lists no less than overall 41 different tempered tunings: eventual success had many parents that presumably had to fight vehemently for recognition. Even today, bitter adversaries turn up who are bothered by beating, "unnatural" intervals, while proponents of equal temperament revel in unlimited modulations. **Guitar players** should better make sure they run with the latter group because their instrument is manufactured using equal-temperament tuning.

In order to unambiguously define the whole relational range, it is also necessary to specify an **absolute value** besides just the frequency *relations* of the notes on a scale. As the long-standing reference (concert pitch), a^1 – the so-called middle A (also designated a' or A_4) – is in service. Today, the standard tuning frequency is **440 Hz** while in past centuries there were significant deviations in the range between 337 Hz and 567 Hz. In Germany, the reference was fixed to 422 Hz in Berlin in 1752. The year 1858 saw a proposal for international standardization on the conference on concert pitch in Paris, followed – on the corresponding conference in Vienna in 1885 – by the adoption of 435 Hz. On the ISA-conference in London in 1939, this value was increased to 440 Hz, and confirmed in 1971 by an ISO-resolution (ISO = International Standard Organization). In conjunction with the standardization, it was suggested to use the reference pitch for interval signals in radio and television, and as dial tone for the telephone. This was not a successful marketing idea: for the telephone, check measurements in 2004 showed a 6% deviation. The following table gives some fundamental frequencies for notes tuned to equal temperament; reference for A_4 is 440 Hz.

C	C#=D*b*	D	D#=E*b*	E	F	F#=G*b*	G	G#=A*b*	A	B*b*	B
523,25	554,37	587,33	622,25	659,26	698,46	739,99	783,99	830,61	880	932,33	987,77
261,63	277,18	293,66	311,13	**329,63**	349,23	369,99	392,00	415,30	440	466,16	493,88
130,81	138,59	**146,83**	155,56	164,81	174,61	185,00	**196,00**	207,65	220	233,08	**246,94**
-	-	-	-	**82,41**	87,31	92,50	98,00	103,83	**110**	116,54	123,47

Table: Frequencies of tones tuned to the equal-temperament scale, referenced to A_4 = 440 Hz; rounded to two decimal places. The open strings on the guitar E_2, A_2, D_3, G_3, B_3, E_4 are in bold.

In order to obtain convenient specifications of small deviations from correct tuning, Alexander John Ellis defined (in 1885) the **cent** as the (supposed) pitch-atom:

$$1\,\text{cent} = 2^{1/1200} = 1{,}0005778 \qquad\qquad \text{Interval} = 3986 \cdot \lg(f_2/f_1)\ \text{cent}$$

1 cent amounts to 1/100th of a half-step, or to the 1200th part of an octave. The frequencies 2000 Hz and 2001,155 Hz differ by 0,058% i.e. by 1 cent. Simbriger/Zehelein cite Preyer with the insight (questionable from a present-day perspective) that the hearing system was able to distinguish 1200 pitch steps between 500 Hz and 1000 Hz. Presumably, many a teacher scared away their pupils by demanding that the latter should be able to discern intonation errors of a 100th of a half-step. Chapter 8.2.2 has more on this topic.

8.1.4 Intervals in the equal temperament

The interval (inter vallum = space in between) is the distance of two notes; expressed numerically by the relation (ratio) of the frequencies of the corresponding tones. The names of the intervals are derived from the place numbers within the scale – for the C-major-scale, this implies: C = prime, D = second, E = third, F = fourth, G = fifth, A = sixth, B = seventh, C' = octave. Between the 3rd and 4th notes, and between the 7th and 8th notes, we find a half-step, all other notes are a whole-step apart each. In the equal-temperament tuning, a **whole-step** consists of two equal-size **half-step** (**HS**). All intervals can be represented by multiples of a HS:

Distance between notes (intervals) in the diatonic scale, represented by half-steps:
C-C = 0, C-D = 2, C-E = 4, C-F = 5, C-G = 7, C-A = 9, C-B = 11, C-C' = 12.

Intervals are not just definable as HS-multiples in their relation to the root note C of the C-scale, but also between all notes: e.g. D-E = 2 HS, G-H = 4 HS, F-A = 4 HS.

By the subdivision of the whole-step into two half-steps, new notes are obtained; they are designated by the chromatic sign relative to their neighbors: C# = C-augmented-by-one-HS, and (in the equal-temperament tuning) identical to the D*b* = D-diminished-by-one-HS. Corresponding: D# = E*b*, F# = G*b*, G# = A*b*, A# = B*b*. Equating the diminished notes and the augmented notes (e.g. C# = D*b*) is called the **enharmonic equivalent** (or enharmonic ambiguity). Out of experience, it appears that guitar players are more familiar with the augment-sign (#) than with the diminish-sign (*b*), and therefore we will give the former priority in the following. From the 7-step diatonic scale (C-D-E-F-G-A-B), a 12-step chromatic scale emerged:

C – C# – D – D# – E – F – F# – G – G# – A – A# – B chromatic scale

Each hyphen in this sequence represents a HS; the size of an interval can therefore be easily accounted for as HS-multiples. The regular numerals (second, third, fourth, fifth, etc.) are, however, already used (up) for the 7-step major (diatonic) scale, and this led to a somewhat confusing nomenclature: unison (0 HS, also called keynote or root), fourth (5 HS), fifth (7 HS) and octave (12 HS) are designated as **"perfect" intervals**, even if their tuning is not "pure" and free of beats! Caution is advised: C-G, for example, is designated a "perfect fifth" even in equal-temperament tuning. All other intervals within the major scale are "**major**" and thus: C-C = (perfect) unison, C-D= major second, C-E = major third, C-F = perfect fourth, C-G = perfect fifth, C-A = major sixth, C-H = major seventh, C-C' = perfect octave.

Reducing a large (major) interval by a HS results in a small (**minor**) interval. To get there, two possibilities exist: either the higher note is pushed down by a HS, or the lower note is pushed up by a HS: C-D*b* = C#-D = minor second, C-E*b* = C#-E = minor third, C-A*b* = C#-A = minor sixth, C-B = C#-H = minor seventh. If a perfect (or major) interval is enlarged by a HS we have an **augmented** interval; if a perfect (or major) interval is reduced by a HS we have a **diminished** interval. This results in two schemes:

diminished – minor – major – augmented (second, third, sixth, seventh)
diminished – perfect – augmented (unison, fourth, fifth, octave)

C-D# therefore represents an augmented second; in the sense of the enharmonic equivalent within the equal-temperament tuning, however, it also corresponds to the minor third C-E*b*. Purists turn away in horror, but the pragmatist just deals with it in everyday life: "C-D# is a minor third." Indeed, it is without purpose to ponder the differences between C# and D*b* when working with equal-temperament tuning. Of course, singers or violinists (as an example) will tend to intonate the augmented notes (#) slightly higher and the diminished notes (*b*) slightly lower, but that is then outside of equal-temperament tuning. When playing chords, the guitar player (and we are concerned with the associated instrument here, after all) has hardly any possibility to modify individual notes within the chord in their pitch. When playing single-note melody, higher-order knowledge of harmony could be put to use – unless the keyboard player in the band with his/her equal-temperament tuning shoots that down.

The following list gives an overview for **all intervals**, in this case referenced to C; with these representations: p = perfect, d = diminished, mi = minor, ma = major, a = augmented:

d-octave: C-C'*b*	p-octave: C-C'		a-octave: C-C'#
d-seventh: C-B*bb*	mi-seventh: C-B*b*	ma-seventh: C-B	a-seventh: C-B#
d-sixth: C-A*bb*	mi-sixth: C-A*b*	ma-sixth: C-A	a-sixth: C-A#
d-fifth: C-G*b*	p-fifth: C-G		a-fifth: C-G#
d-fourth: C-F*b*	p-fourth: C-F		a-fourth: C-F#
d-third: C-E*bb*	mi-third: C-E*b*	ma-third: C-E	a-third: C-E#
d-second: C-D*bb*	mi-second: C-D*b*	ma-second: C-D	a-second: C-D#
d-unison: C-C*b*	p-unison: C-C		a-unison: C-C#

This way, and given the enharmonic equivalent, every tone of the chromatic scale may exist in two different interval relationships to the keynote (in this case C):

C	perfect octave	12	augmented seventh	octave
B	major seventh	11	diminished octave	major-7th
B*b*	minor seventh	10	augmented sixth	seventh (mixo)
A	major sixth	9	diminished seventh	sixth (dorian)
G#	minor sixth	8	augmented fifth	#5
G	perfect fifth	7	diminished sixth	fifth
F#	augmented fourth	6	diminished fifth	tritone (lydian)
F	perfect fourth	5	augmented third	fourth
E	major third	4	diminished fourth	major 3rd
D#	minor third	3	augmented second	minor 3rd
D	major second	2	diminished third	whole step
C#	minor second	1	augmented unison	half step (phrygian)
C	perfect unison	0	diminished second	root

The *first* column in this table holds the designations of the note, the *second* column the preferred interval designations. The *third* column represents the half-note intervals relative to the keynote, and the *fourth* column represents the alternate designations. In the *fifth* column, some abbreviations customarily used by musicians are found (there may be others, of course). Again: this is based on equal-temperament tuning including enharmonic equivalents. Classical harmony theory finds reasons for a further differentiation; however, this is beyond the aim of the present elaborations [see secondary literature].

The below table indicates the numerical differences between just tuning and equal-temperament tuning. The deviation is just tuning vs. equal temperament tuning.

Interval name	no. of HS	notes	frequency relation	cents	deviation
Perfect octave	12	C-C'	1\2	1200,00	0,00
Major seventh	11	C-B	8\15	1088,27	-11,73
Minor seventh	10	C-B*b*	9\16	996,09	-3,91
Major sixth	9	C-A	3\5	884,36	-15,64
Minor sixth	8	C-G#	5\8	813,69	+13,69
Perfect fifth	7	C-G	2\3	701,96	+1,96
Tritone	6	C-#	32\45	590,22	-9,78
Perfect fourth	5	C-F	3\4	498,05	-1,95
Major third	4	C-E	4\5	386,31	-13,69
Minor third	3	C-D#	5\6	315,64	+15,64
Major second	2	C-D	8\9	203,91	+3,91
Minor second	1	C-C#	15\16	111,73	+11,73
Perfect unison	0	C-C	1\1	0,00	0,00
					cent

Table: Frequency relations of octave-internal intervals for just tuning. The deviations refer to the corresponding interval in equal-temperament tuning. Specifications in 1/100th cents should be in practice rounded off to whole cent-values. Compared to the major third in equal-temperament tuning, the major third in just tuning is to low by 14 cents. The other way round: compared to the just-intonated major third, the major third in equal-temperament tuning is too high by 14 cents. A deviation of 1 cent corresponds to a frequency difference of 0,058%.

We can see the frequency relations for different tunings in the following **Fig. 8.9**. The abscissa is a logarithmically divided frequency axis.

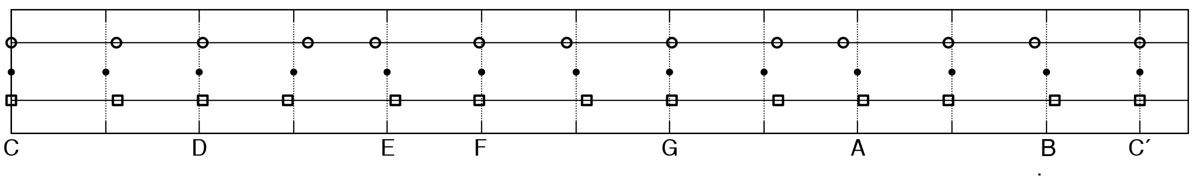


Fig. 8.9: Pythagorean (□), just-intonated (•), und perfect (o) intervals.

Since the half-step intervals are all equal in equal-temperament tuning, changing key (i.e. moving to a scale with a different keynote) does not represent a problem. For example, referencing to E results in the following intervals: E-F# = 2 HS = major second, E-A = 5 HS = perfect fourth. The reference to a particular key may now be omitted, because every interval is unambiguously defined by the number of its half-steps (HS).

Further interval designations exist **beyond the octave space**, as well: minor ninth (13 HS), major ninth (14 HS), minor tenth (15 HS), major tenth (16 HS), perfect eleventh (17 HS), augmented eleventh = diminished twelfth (18 HS), perfect twelfth (19 HS), minor thirteenth (20 HS), major thirteenth (21 HS); the half-step distances are given in brackets.

8.1.5 Typical detuning in guitars

Every guitarist will have experienced days when his/her guitar would just not tune properly. It typically gets <u>really</u> bad when we try to re-tune individual notes within *chords*. Even with a perfectly fretted neck and premium strings, this problem may occur – the most likely reason for which is the difference between just intonation and equal-temperament intonation. While the **fifth** tuned according to the latter is, with a deviation of 2 cents, really close to the perfect fifth tuned with just intonation, we find a much larger deviation for the **third**: that would be +13,7 cents for the major third, and as much as -15,6 cents for the minor third! Such detuning is already well audible, and the guitarist simply has to live with it. Trying to chord-specifically retune individual strings (towards just intonation) may easily generate deviations of as much as 29 cents for notes in other chords – and with that it now gets really shoddy. As an example:

An A-major chord (played without barré) consists of the notes [e-a-e-a-c#-e]. Given that all notes are tuned to equal temperament, it is in particular the C# played on the B-string that creates problems: it is sharp by 14 cents compared to a justly intonated C#. If we now retune the B-string by -14 cents (7,9 ‰), this A-chord will sound perfect. However: if, with the same retuning, e.g. an E-chord [e-h-e-g#-b-e] is played, the resulting sound is atrociously off. What happens is that the down-tuned B-string sounds a flat fifth – while the major third in that E-chord (the G# played on the neighboring G-string) is sharp. The interval between these two strings (3 half-steps G#-B) is too small by 29 cents! Changing from that re-adjusted A-major chord to a D-major chord creates a similar disaster: the down-tuned B-string now sounds too flat a D. The major third (F#) played on the neighboring E-string is already anyway too sharp by 13,7 cents and now sounds doubly out-of-tune relative to the tonic (that is lowered by 13,7 cents).

There may always be special cases when – given a limited selection of chords – a special detuning creates advantages. For example, it does not sound bad at all to slightly lower the tuning of the G-string for E and A7. E-major has [e-b-e-g#-b-e], and A7 has [e-a-e-g-c#-e]. In the E-major chord, the third profits, and in the A7 chord the diminished seventh – both are sharp in equal temperament relative to just tuning so that this detuning makes sense. For the same reason, the same detuning works well with the B7-chord [f#-b-d#-a-b-f#]. But don't you dare now changing to C or G … Thus, for universal deployment it is the equal-temperament tuning (executed as perfectly as possible) that remains a workable solution.

8.1.6 Stretched tuning

Piano tuners are known to tune not exactly according to equal temperament but in a slightly stretched-out fashion. In particular, in the very high and very low ranges, deviations of up to 30 cent can result. A spreading-out of partials, and in addition a narrowing of the pitch perception, are given as justification. In the guitar-relevant pitch range, however, the effect (merely 2 cents per octave) is rather weak, and the (compared to guitar strings) much heavier piano strings are no adequate equivalent. "Buzz" Feiten has obtained a US-patent for the stretched tuning – see Chapter 7.2.3). Fender, on the other hand, recommends adjusting the octave at the 12th fret with no more than 1 cent error – no spreading. To each his/her own …

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.

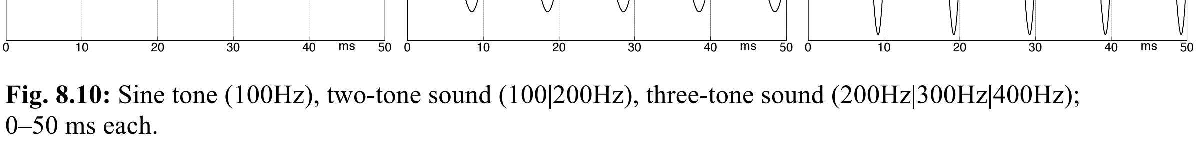


Fig. 8.10: Sine tone (100Hz), two-tone sound (100|200Hz), three-tone sound (200Hz|300Hz|400Hz); 0–50 ms each.

♣ 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 100^{th}, 201^{st}, 302^{nd}, and 404^{th} 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.

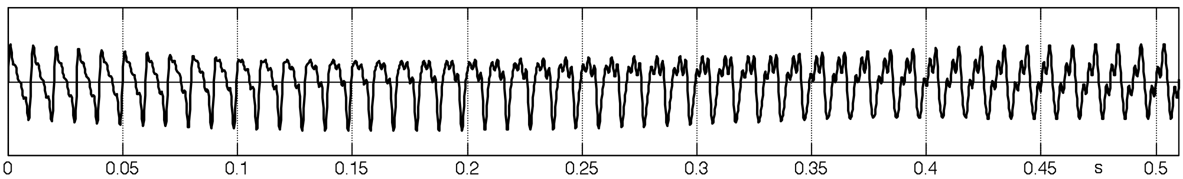


Fig. 8.11: 4-partials sound, f_1 = 100Hz, f_2 = 201Hz, f_3 = 302Hz, f_4 = 404Hz. $1/f$–envelope; t = 0 – 0,5 s.

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 $3/4^{th}$ 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 12^{th} 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 $(997{,}7\,\text{Hz})$ 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.

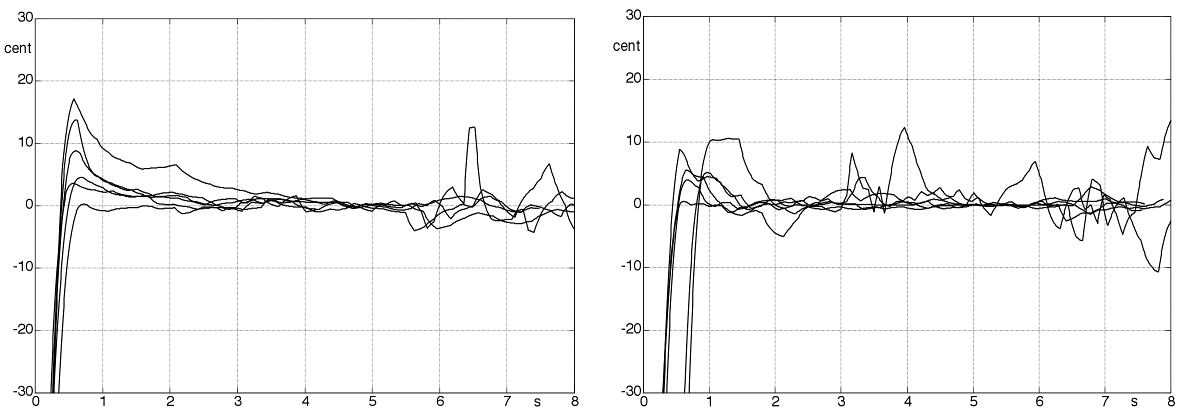


Fig. 8.12: Pitch measurement with the electronic tuner Korg GT-2. Tennessean (left), Ovation SMT (right).

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.

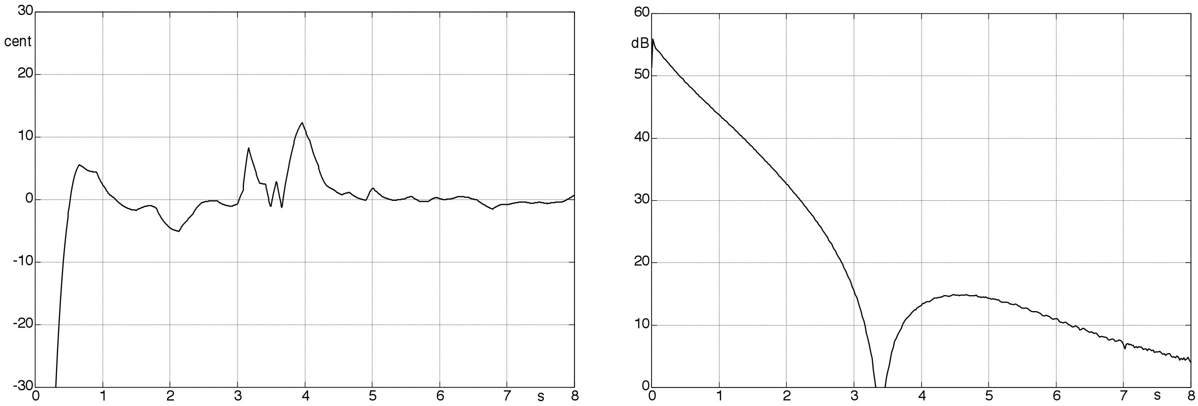


Fig. 8.13: Measured pitch-deviation, level of fundamental; Ovation SMT, B-string plucked at $t = 0$.

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 $16x10^{-6}$ 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?

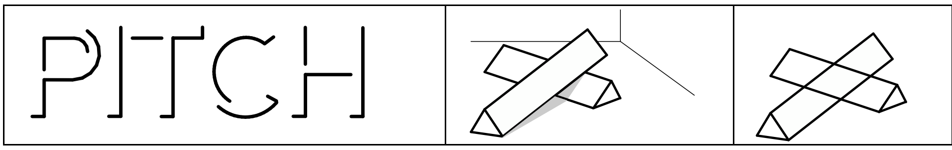


Fig. 8.14: Examples for visual grouping.

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.

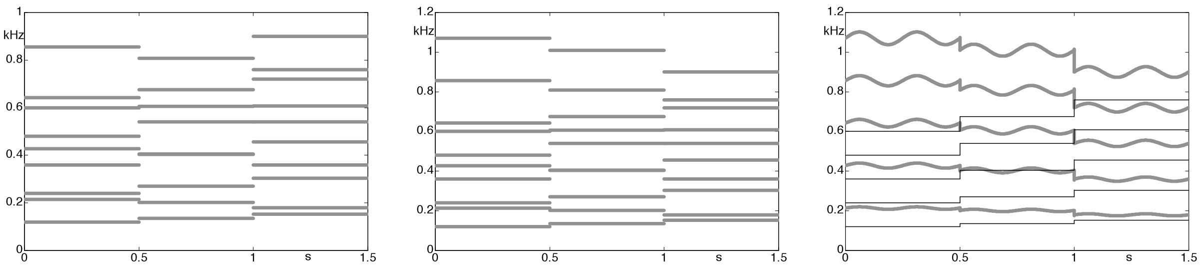


Fig. 8.15: Spectrograms of two tone-sequences: on the right, the descending sequence is frequency-modulated.

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 superimposed 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 1st,2nd, 3rd, 4th, 6th, 7th, 8th, and 9th 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 3rd, 6th and 9th harmonics are slightly modulated – the latter now form in a new grouping the 1st, 2nd, and 3rd harmonic of the additional 300-Hz-tone.

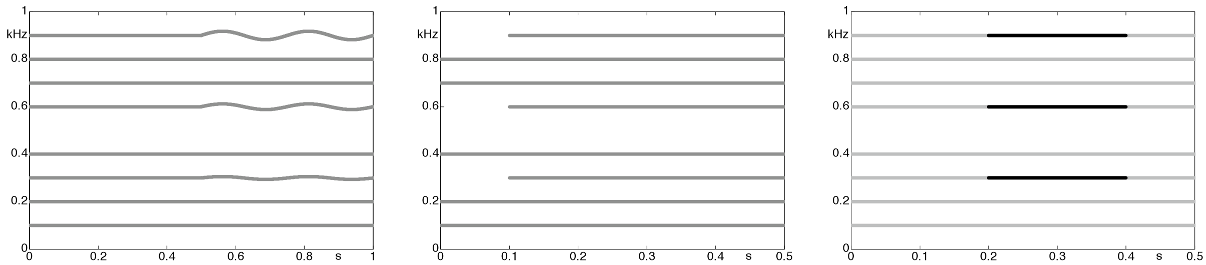


Fig. 8.16: Partial with/of a common fate are grouped to objects.

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 3rd, 6th, and 9th 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 3rd, 6th, and 9th 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 f_i of the partials are calculated as:

$$f_i = i \cdot f_G \cdot \sqrt{1 + b \cdot i^2} \qquad f_i = \text{frequency of the partial; } f_G = \text{frequency of the fundamental.}$$

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.

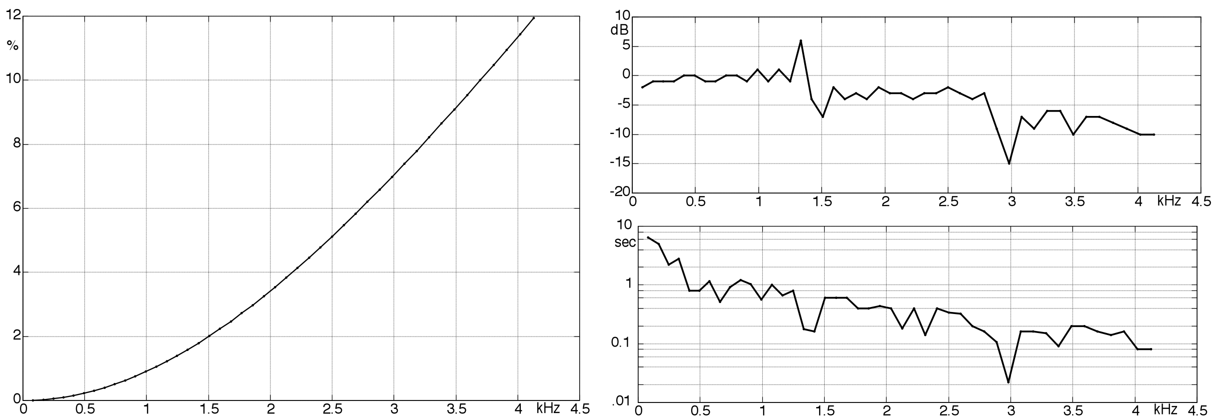


Fig. 8.17: Percentage of spreading-out of partials (left); levels and decay-constants of partials (right).

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

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

♣ 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.

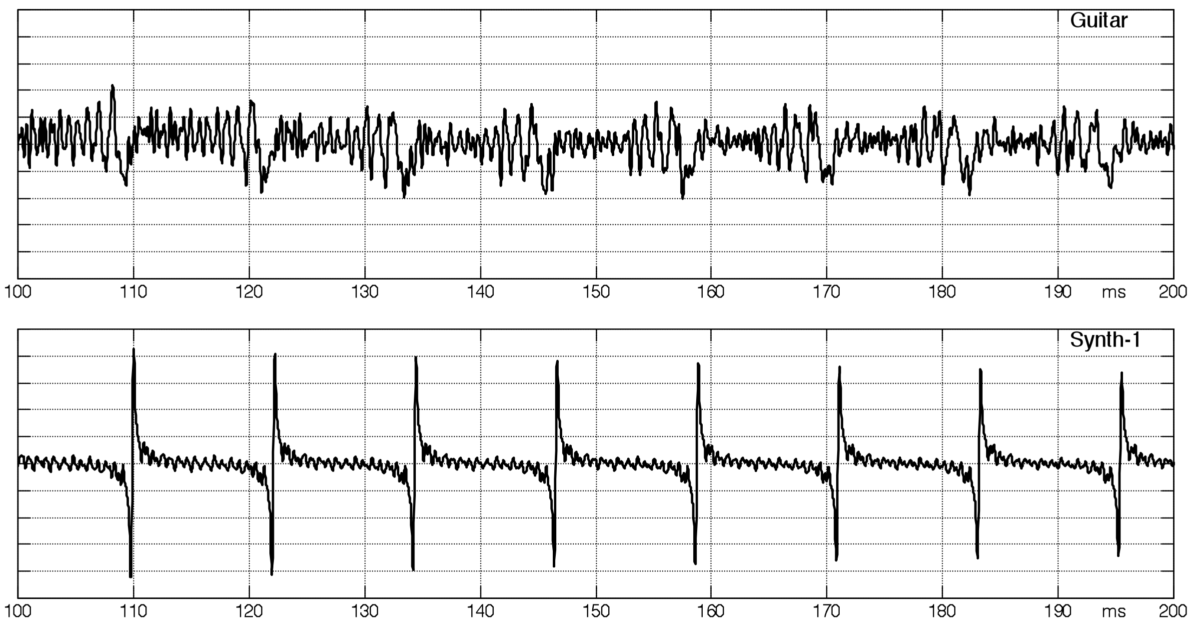


Fig. 8.18: Time-functions of the real signal and of the synthetic signals; E_2-string.

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.

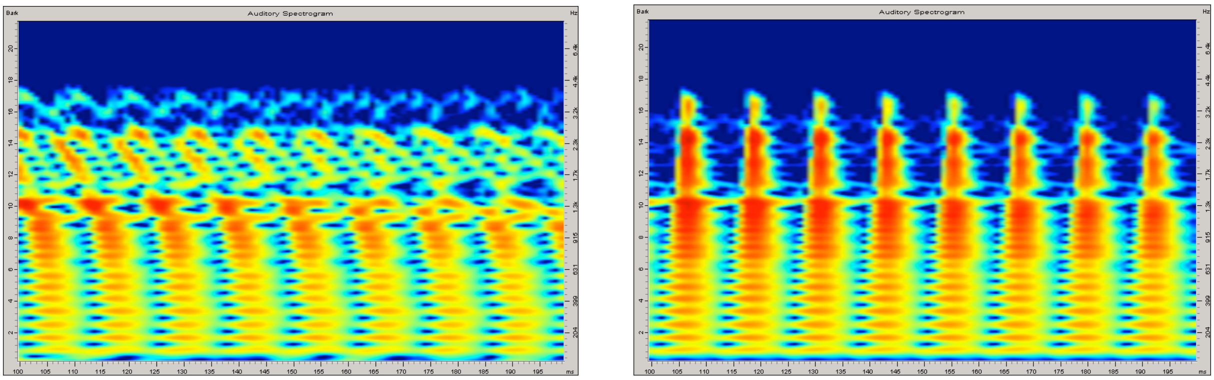


Fig. 8.19: Auditory spectrogram (CORTEX-VIPER), real signal (left), Synth-1 (right).

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.

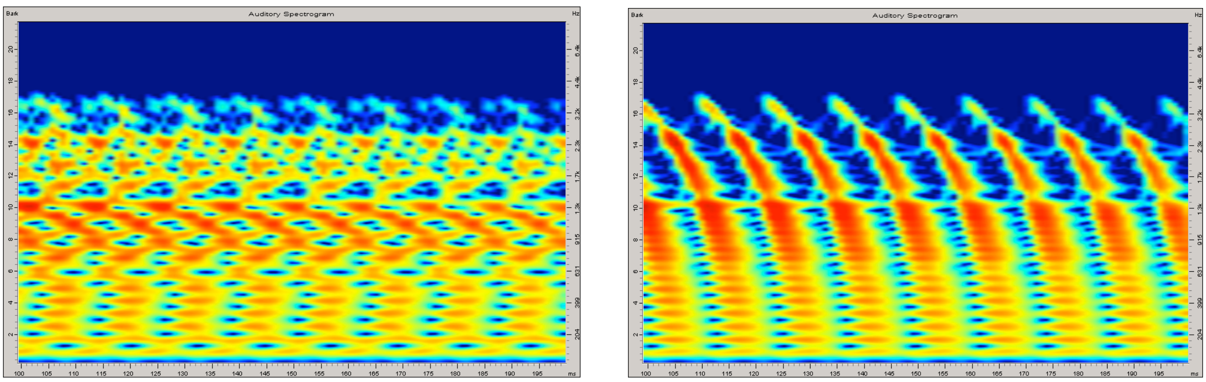


Fig. 8.20: Auditory spectrogram (CORTEX-VIPER), Synth-2 (left), Synth-3 (right).

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.

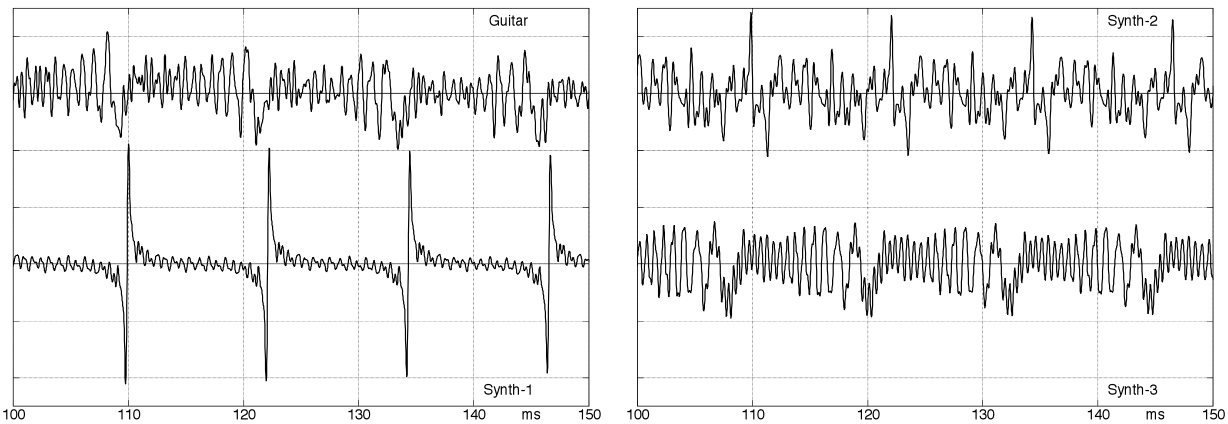


Fig. 8.21: Time functions of the real signal and of the three synthetic signals.

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.

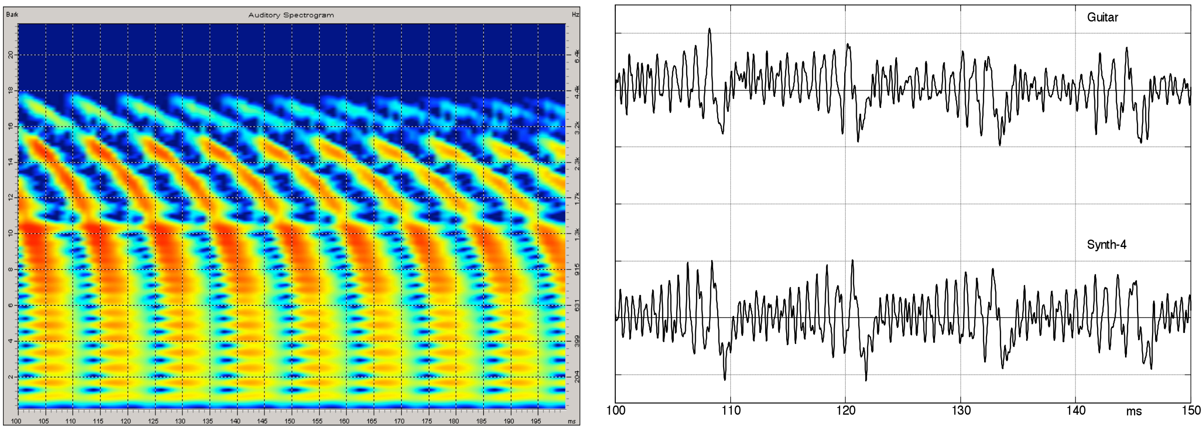


Fig. 8.22: Synthetic signal with spread-out spectrum (Synth-4).

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.

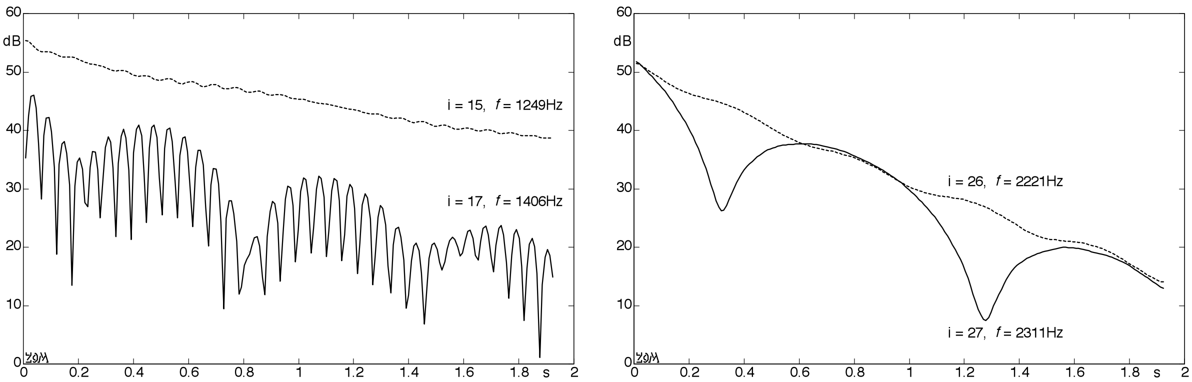


Fig. 8.23: Decay curves of individual levels of partials; Ovation-guitar, piezo pickup.

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)

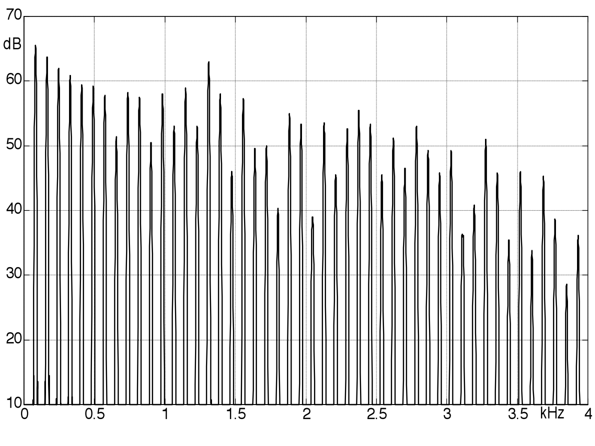
$$x(t) = \hat{x} \cdot \left[\cos(2\pi \cdot f_1 \cdot t) + \cos(2\pi \cdot f_2 \cdot t)\right]$$

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

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

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.



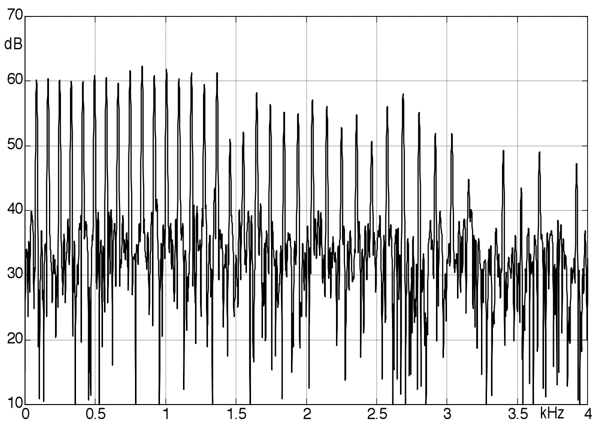


Fig. 8.24: Spectra of signals subjected to non-linear distortion. Left: harmonic primary signal; right: spread-out primary signal. Cf. Chapter 10.8.5.

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.

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

8.3 The character of the musical keys

"And indeed it appeared that for Beethoven, certain keys had certain characters that made them useful for corresponding moods and content[♣]". This statement and similar ones have led to attribute to a musical key an absolute character, in the sense of Eb-major = heroic, C-major = impersonal, E-major = solemn. This may have different reasons:

a) Musical keys do have a character. Brilliant musicians (such as e.g. Beethoven) have recognized this, and have oriented their compositions accordingly.
b) For whatever reason, brilliant musicians have believed in a character of the keys. Admirers of their music have internalized this, learned from it, imitated it, passed it on … and thus a self-fulfilling prophecy came into being: because Eb-major sounds heroic, heroic music is composed using Eb-major … and so Eb-major sounds heroic.
c) The whole shebang is nothing but coincidence.

Mies reports from an experiment that points to the existence of absolute character. He played Schubert’s Impromptu to about 20 pupils: once in G-major and Gb-major each. He asked which of the two was the original key. 3 pupils voted for G-major and the rest for Gb-major. They reasoned that the key they chose fitted better to the mood of the piece. Mies knew about the limited validity of such a single experiment and started systematic investigations with a multitude of piano pieces. At first glance, his results are contradictory: on the one hand he does arrive at a character correlation (see table below), but on the other hand he summarizes: “and here the investigations are clear proof that there is no general character of the keys across ages and composers, meters and time, rhythms and melodies. **A general character of a key that would be independent of composer, time, listener, etc., does not exist.**” In fact, this summary does not actually contradict the table because the latter answers the question of which matching the investigated composers have preferred. If D-minor feels passionate to Brahms, this is of no more significance than the statement that Eb-major was Beethoven's favorite key. Who would deny any great composer a subjective preference? A pars-pro-toto principle is, however, not justified by this.

C-major:	*Objective, superficial, impersonal. Key of truth. For thanks and salute.*
C#-major:	*Glimmering, sparkling, lively, virtuosic.*
Db-major:	*Soft, gentle, emotional.*
D-major:	*Key for marches, fanfare, cheerfulness, joy, festive splendor, scenes of revenge.*
Eb-major:	*Serious, grave, deep love, tormenting lovesickness.*
E-major:	*Solemn, serious to gloomy, belongs to exalted and otherworldly moments.*
F-major:	*Friendly, natural, moderate.*
F#-major:	*Passionate, ardent love.*
G-major:	*Simple, uncomplicated, cheerful.*
Ab-major:	*Quiet, emotional, longing. Sinister scenes.*
A-major:	*Manifold, lovely, serenade-like. Key of happy people. Expression of splendor.*
Bb-major:	*Cheerful, playful, gently. Cordial sentiment.*
B-major:	*No general character.*

Table: The character of the keys. Paul Mies, 1948. Strongly abbreviated representation.

[♣] P. Mies, Der Charakter der Tonarten, Staufen, Köln 1948

In his account, Mies explicitly points to the enharmonic identities and refers to the piano tuned to equal temperament. If the keys had an absolute character, Db-major could not appear soft and gentle while C#-major is sparkling and lively. That C#-major appears virtuosic – that is conceivable to the musician given its no less than 7 accidentals! If we look long enough, we do find contradictions: Gb-major = sad; F#-major = passionate. On the other hand, a lot can indeed be made to fit: the CD tries to make us believe that ***"Roll Over Beethoven"*** (with apologies, dear & highly esteemed Ludwig van) is played in Eb-major. Tormenting lovesickness? Maybe not … presumably heroic – after all, Chuck B. is the true R'n'R hero to many. However, what does the songbook tell us (*the authentic transcriptions with notes and tablature*)? D-major!! That would be "joy", then! That's gotta be it: the tape machines ran a few percent slow back in the 1950's, and Chuck must have certainly played in D-major. Cheerfulness, joy, festive splendor – that's really like him! However, Mies also lists "scenes of revenge" related to D-major, and lo and behold: "*Don't you step on my blue suede shoes*" – this little sideswipe is put in proper perspective, too. Or, how about something from the Stones' songbook: "***Let's Spend The Night Together***" of course is in Ab-major. That fits him to a T: longing and emotional chap that he is, our Mick. Did somebody say "sinister scenes"? Rather, a certain "sensitivity and delicacy of feeling", as Riemann elaborates. "Of practice-character", Mies complements, and is bang on target (*you need some guiding, baby*). Typical for Ab-major are also "the mid-tempos taking up the most extensive space" (*I'm in no hurry I can take my time*) and "medium and slow tempi with frequently on-going movement": an excellent match for what the Stones' front-man stands for. Not to forget: "sweet, romantic melancholy and longing" (*now I need you more than ever*), as well as Stephani's "soft-solemn seriousness" (*Oh my, da da da da da da da da*). And finally: " the movement often perceivable in the tempered pieces is also felt in the accompaniment" (*around and around, oh my, my, yeah*) – perceivable movement in the background vocals, indeed. Much could be added here, for example the *19th Nervous Breakdown* (E-major, otherworldly moments) or *Street Fighting Man* (F-major, friendly, natural, moderate). And many more …

Still, we do see criticism, as well: "in view of all these statements, an absolute character of Eb-major certainly cannot be observed." Or: "indeed, literature does not agree about Ab-major". Or: "would it not be possible that Beethoven's quotation (not actually from his own notes) was not correctly handed down in its relation to the keys; no support can be found in his own works." These are all Mies' citations. It would also be possible that everything is one big misunderstanding.

In the course of the last centuries, highly diverging opinions can be found about the absolute tuning of an instrument: the chamber pitch (concert pitch, standard pitch), i.e. the middle a (a', A_4), varied in its frequency by as much as 337 – 567 Hz! Even going back only to the 18th century (checking in with Beethoven or Brahms), we still find a scatter of just shy of a semitone). That could turn A-major ("key of happy people") into Ab-major and thus call for "sinister scenes". Nightmarish: the A-major scherzo played too low by a hair becoming the hotbed for sinister Hans-in-Lucks and happy gloom-o-philes – a cut set of joy and sorrow? "Die then, die now, die! Haha! Hahaaaa! Hahahaha! Die! Die!" "Welcome oh blissful woe – continue, go on." Without doubt: A-major, a quarter-step too low?

Nay, psychoacoustics does not know of a "tonal character" based solely on the frequency position. It is still conceivable that Schubert's Impromptu sounds more authentic in Gb than in G. There are no known recordings of Mies' experiment – we can therefore only speculate: Mies had presumably rehearsed the piece in Gb, with the transposition to G requiring different finger movements and possibly resulting in a different sound character just because of that. The experienced subject can detect timing-differences as small as 5 ms (Chapter 8.5)!

Also possible: the special tuning of the piano used resulted in characteristic beats that of course are key-dependent. Specific resonances of the individual piano may have played a role, changing individual notes/passages/chords in a key-specific manner. And finally it cannot be excluded that Mies (who knew when he was playing in the original key) was himself not convinced of the use of G-major, and therefore played with inferior expression in that key. For a double-blind test this was NOT.

Could we repeat the Mies-experiment as a double-blind test? We would require a very good pianist who practices both the G- and the Gb-version with the same dedication. That would seem doable. Presumably, however, this pianist would (wittingly or unwittingly) prefer one of the two versions, and thus would not be able to play both with the same expression. In this case the listeners would assess the way of playing and not primarily the key. We would therefore have to directly ask the pianist (or several pianists) but this would put the general validity of the experiment. Given modern options, a purely electronic transposition would be feasible: the piece is recorded e.g. in Gb-major and reproduced with a 6% higher speed (or sampling frequency). But then not just the key changes but also the timing: the G-major version is faster by 6% compared to the Gb-version. That's not optimal, either. Using harmonizers or pitch-shifters (special equipment used in recording studios) that change the pitch without influencing the reproduction speed calls for skepticism, as well, because with them the subject may judge the quality of the signal-processing algorithms and not just the character of the key.

Conceivable would be the following approach: the pianist plays the piece in the original key, and the key movements are electronically recorded (via MIDI or something better). From the stored data, artificial piano sounds can be created – both in the original key and in a transposed version. This 'electronic music' may now be judged with respect to the character of the key♣. Today, psychoacoustics assumes that such music has no inherent key-specific character, i.e. that aggressiveness, passion, or sorrow need to be expressed by means of harmony and rhythm.

However, this does not imply that the character of a piece of music accompanied by the guitar cannot change if the piece is transposed form G-major to A-major. If the guitarist plays a G-major chord without barring strings (g-b-d-g-b-g), and changes to the 'open' A-major chord (e-a-e-a-c#-e), the color of the sound will change significantly. However, this is not due to the different key, but results from the different chord composition. In the G-major chord the fifth appears only once, but three times in the A-major chord. Conversely, an A-major chord played in the 5th position (barré on the 5th fret) has only two fifths. Thus the simple conclusion is: when changing keys, the character of the sound can change – however this is not according to a generally applicable scale but specific to the respective interpretation and instrument.

♣ Similar experiments had already been carried out by Terhardt and Seewann – however with sounds that differed from those of the acoustic piano. The objective of these tests related to perfect pitch (absolute pitch) and not to the character of the musical key [Aural key identification and its relationship to absolute pitch. Music Percept. 1, 1983].

8.4 Consonance und dissonance

Although every musician thinks he/she knows what a **dissonant chord** is, a scientific description proves to be difficult. The Pythagoreans considered the octave, the fifth, and the fourth to be consonant (symphonical), and all other intervals as dissonant (diophonic). There is one peculiarity: *to the astonishment of the westerner, now and then the major second is also designated as being consonant. This has its basis in the totally different concept of melodic consonance: consonant is that which is easy to pitch* (*Simbriger/Zehelein*). Another explanation would be: the major second must not stand too far apart since it is the fruit of the "holy matrimony" (Fig. 8.6), and thus sanctioned via the insights of a "advanced civilization".

Apparently there is more than one type of consonance. As a synonym, we often find euphony, coalescence, serenity, relaxation. Playing in the middle range on the piano two notes at a distance of a fifth, both melt into one harmonic sound. The two notes "like each other", they sound well together, and that is exactly what con-sono means. Very different are two notes at a distance of a half-step: the esthete downright hears the fight they slug out, while the signal-theoretician detects beats, the psychoacoustician notices roughness – and the musician perceives dissonance.

Already early on, the nominal attribute became a ordinal attribute: for dyads, not only a statement was sought that they harmonize well, but also an assertion about how well they harmonize (concord) – in the sense of a ranking. **Franco von Köln** put together a five-step scale in the 13th century (C. = consonance, D. = dissonance):

Complete C.	Medium C.	Incomplete C.	Incomplete D.	Complete D.
Prime Octave	Fifth Fourth	Major third Minor sixth	Minor third Major sixth	Second Seventh

The high consonance of the fifth is already evident from Fig. 8.2: in the spectrum of the partials there is a close relationship. The 3rd, 6th, 9th, etc. partials of the lower note have the same frequency as the 2nd, 4th, 6th, etc. partial of the higher note – given perfect tuning and dispersion-free wave propagation. What could be closer than to derive, from the similarity of two notes, rules for the generation of consonance and dissonance? For example:

- The more shared partials, the higher the consonance. Or:
- The simpler the frequency ration, the higher the consonance.

However, there were also cautious rearguard actions: "essential are only the odd-numbered partials". Or: "the 7th, 11th, 13th, 14th, and 17th partials are excluded". Or: "the fourth is a perceptional dissonance". Or: "there are dissonant chords of highly consonant sound". Or "In context, a consonant chord very often is bestowed a dissonant purpose". And rather recent from Haunschild's 'New Theory of Harmony' (1998): "In general we can note that the human understanding of consonance and dissonance more and more shifts away from consonance, in favor of dissonance. This means that more and more intervals and chords that were surely classified as dissonant back in the day, are today rated as consonant. It is only the intervals with a so-called semi-tone-friction (minor second and augmented seventh) that are truly assessed as dissonant."

Let us give the philosophers some space, as well: "corresponding to the relations of the natural degrees of consonance it is possible to say that every entity, every form of being is the more complete, and thus the more in harmony with its physical and social environment, the closer it is positioned to its origin. The principle of consonance is the connecting within the differing – it therefore corresponds to the harmony, the organic integrated-ness in higher unity, in other words: love" (found in Simbriger/Zehelein). So then **Schönberg** possibly was a love-less person? He opined: “today we have already gone so far as to not make a difference anymore between consonance and dissonance.” Rossi similarly (but not quite as radically) says: "consonance and dissonance greatly depend on each individual's musical experience, and, more broadly speaking, musical culture".

It shall not be disputed that beats, roughness, fluctuations, frictions, or anything else you would want to call the **envelope variations** of the partials, represent a cause for the perception of dissonance or consonance. However, perception psychology increasingly distances itself form the so-called *unbiased scaling*, i.e. an absolute, purely signal-dependent scaling. At the 8th Oldenburg Symposium, Viktor Sarris elaborates: "Whereas classical sensory psychophysics relies mainly on the (illusory) assumption of absolute, i.e. invariant stimulus-response laws, the relation-theory in psychophysics is based on the general premise that, on principle, one and the same stimulus may be perceived and judged very differently as a function of the variables implied by the total 'contextual' situation at hand. … Contextual effects in psychophysics are of major importance since virtually all kinds of sensory-perceptual-cognitive judgments, whether in direct or indirect scaling resp. in discrimination and postdiscrimination-testing, are **contextual**." The insight that evaluations happen in relation to the given situation also concerns judgments of consonance – in particular if these are delivered by persons with musical experience or education.

We may use as an **example** a dyad with the two tones forming a **major sixth** – i.e. for example B-G#. Let us imagine two guitar players: one of them continuously plays an E-major chord, the other frets the B on the G-string, and alternatingly (e.g. with a 6/8th rhythm) the G# on the high E string. Both B and G# are included in the E-major triad; the two guitars play in harmony and the result is a tension-free sound. Now the “man of the 6th“ shifts his fretting hand upwards by 3 semi-tones, i.e. he frets the D on the G-string and the B on the E-string. After one bar he shifts upwards by another 3 semi-tones and plays F-D (**Fig. 8.25**). All the while the accompanying guitarist continues playing the E-major chord. The second sixth is – with D-B – still close to E-major; the D (representing the minor seventh) does already build some slight tension, though (E^7-chord). However, only the third sixth brings some serious dramatics to the game: the D can again be taken as the minor seventh, but the F – representing the minor ninth – is dissonant to a high degree ($E^{7/b9}$-chord). Every player of the electric guitar with some classical education (i.e. Beatles-Beck-Blackmore) knows this skewed chord from Lennon/McCartney’s *I want you*. The interesting thing in this example is: even if no accompanying guitar is playing along, the experienced payer of sixths still hears these mounting dramatics! The latter may be relaxed (resolved) e.g. via a concluding augmented sixth to E-C#. Again: a guitarist plays (now without accompaniment) the augmented sixths: B-G#, D-B, F-D, E-C#, and he/she hears an arc of suspense – although always an equal (not one and the same!) interval is being played.

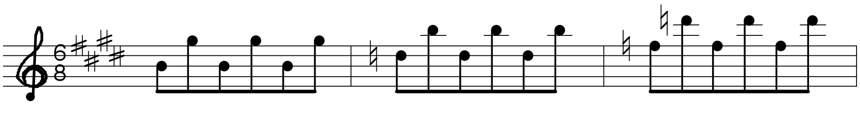


Fig. 8.25: Augmented Sixths.

Requirement for the changing musical tension is a **reference** carried along in memory – which needs to be available to every musician. Otherwise there would be no way to play an improvisation that is guided by accompanying chords. Now, the well-versed musician will (in contrast to the beginner) not need any audible accompaniment at all – he/she will generate it "within", using the "internal ear". The whole thing is less esoteric that one might fear. The reader could, as an example, begin to speak but stop at the last moment – intending to say "*a*" but keep the vocal chords shut. Dutifully, the tongue will already have moved into position, and a well-formed notion of how the vowel will sound (had only been allowed to do so) has emerged. The "internal ear" has already heard the "*a*" although the latter has physically not manifested itself at all. A vocalist could in addition also already set the vocal chords to the appropriate tension in order to produce a targeted pitch; however, already this will not work as well anymore without vocal training. The reason is that the internal ear requires connections between the motor-control areas and the sensory areas in the brain. Strangely, when it comes to hearing, the sensors not only comprise the 8th brain-nerve (N. acusticus). If a layperson-singer (i.e. in this case a person that wants to sing but lacks any skill) is played a note and then asked to sing it, a more or less horrible control process[♣] starts: the vocal chords generate a tone but only as the latter is physically existent can the hearing recognize the pitch and make the vocal chords change their tension. An expert singer, however, is expected to immediately produce the correct pitch without any interfering control processes. This he/she can do, too, because he/she has learned to pre-tension the vocal chords correctly already before the tone sounds ("muscular tone-memory").

Magnetic resonance imaging has enabled us to "watch the brain thinking", and we have started to understand how the individual brain regions cooperate. Or rather: we have a certain conjecture, because an actual comprehensive grasp has yet to be established. Some interesting connections have been observed in pianists: if a **pianist** listens to piano music, regions in his/her brain that are assigned to the fingers become also active. Presumably, the brain already practices how the fingers would have to be moved in order to play what is heard – even though the pianist merely listens and does not actually play. This works the other way round, as well: playing on a keyboard that does not sound any audible notes still activates brain regions related to auditory perception – that is the "internal ear". With beginners of the piano, these senso-motoric connections have, by the way, not been observed. Rookies need to first configure the hardware by practicing.

But back to our topic of **consonance**: at least the well-versed musician supplements the sounds aurally recorded by fundamental and accompanying notes that exist only in his/her imagination. The supplement may be more or less consonant, and therefore consonance is describable by physical signal parameters alone. The major sixths mentioned in the above example will generate an increasing tension only if the E, or the E-major chord, are retained. If the listener thinks of a concurrently changing fundamental note (in the example i.e. E – G – Bb), then the tension is not changed. Setting the respective current reference is an individual process that will follow some roughly predefined rules, but it will not run a predetermined course in the individual case. Rather, musical training as a general criterion, and musical context in particular, are significant. It is easily imaginable that probabilities related to the given choice of the fundamental are set up and evaluated, and that relations within the partials, as well as chord relationships, play an important role. After all, the brain is most powerful in supplementing missing sections in visual impressions – there should be similarities in the auditory system.

[♣] We are familiar with this from casting shows that have spawned frog-like superstars (Kermit on dope), the skillfulness of which with regard to intonation have called for critical voices to speak up even within Lower Bavaria (!).

The third sixth-dyad (F-D) described in the above example may be seen as part of a tetrad in a third relationship. Two tetrads are **third-related** if three of the four notes in the two chords correspond, and if moreover the root notes are located a third apart. In the example, the first two sixth-dyads form – with B-G# and D-B – the basis for an E^7-chord. The third sixth-dyad is part of a diminished seventh-chord (G#^{07}). Third-related to E^7, it forms an $E^{7/b9}$-chord with the latter. This rule of formation is not compulsory; alternative reference systems may be imagined. Indeed it is specifically the possibility of multiple reference systems that renders the degree of consonance not unambiguously definable.

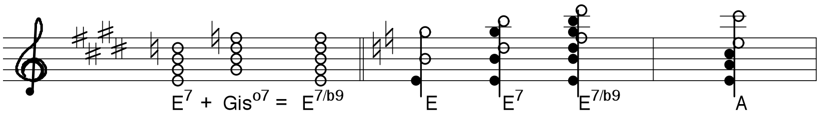


Fig. 8.26: Left: third-related tetrads; right: arc of suspense with resolution in A-major. “Gis” = G#

Fig. 8.26 clearly shows the third-relation mentioned above: E-G#-B-D forms an E^7-chord that has three notes in common with G#-B-D-F (G#^{o7}). The latter supplement the E^7 to an $E^{7/b9}$-chord. The right-hand section of the figure depicts the first sixth-dyad (open note-symbols), and the mentally supplied root note E (filled symbol). This pattern is stored in memory, and the next sixth-dyad is added, resulting in the E^7-chord. The latter is memorized as well (filled symbols) and supplemented by the third sixth-dyad … and there we have our dissonance. Actually played are merely the notes given by the open symbols; all other notes exist only in memory. In case the guitarist plays, in conclusion, also the major sixth E-C#, a nice resolution (relaxation) in A happens; this works in particular if he/she imagines E-A-C# in addition.

The above example was intended to show how the consonance of a major sixth can turn dissonant – if the imagination (the internal ear) plays along. Of course, not only the imagined, but also the notes existing in reality influence the perceived dissonance. In general, the **major seventh** (e.g. E-D#) is considered to be dissonant. However, if it is generated using two sine-tones, “actual” beats do not happen (in contrast to the minor second E-F), but octave beating (so-called 2nd-order beats) results. Experiments tapping the electrical potentials of the cortical nerve give rise to the assumption that our hearing system performs some sort of half-wave rectification within the analysis of the vibration of the basilar membrane♣. The patterns seen in the action-potentials on the nerve fibers change their shape in the same rhythm as the difference frequency (in this example defined by $T = 1 / (f_2 - 2f_1)$). **Fig. 8.27** depicts such a signal; the shown section corresponds to just this beating-periodicity. To compare: in Fig. 8.5, a 1st-order beating was shown. 2nd-order beats act in a more subdued fashion compared to 1st-order beats [Plomp, JASA 1967].

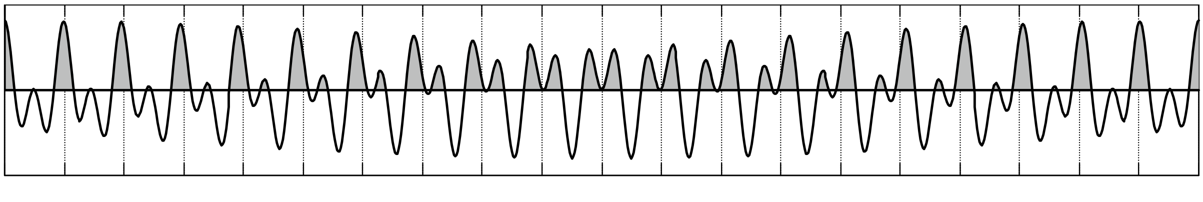
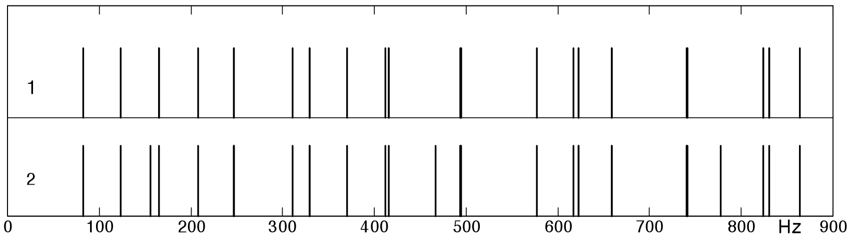


Fig. 8.27: Octave beating. Sum of two primary tones of the same level. The frequency of the higher tone is larger by 2,5% than twice the frequency of the lower tone: $f_2 = 2{,}05 \cdot f_1$.

♣ At least in the frequency range below 1,5 kHz.

Tones from instruments are almost never composed of only one single partial, though. In the guitar, we will normally have to deal with several partials – and in this case 1st-order beats do determine the sound, as the following example will show. Playing the just mentioned major seventh on the guitar (e.g. the E on the D-string and the D# on the B-string) indeed yields a sound that most would call dissonant. However, as soon as we supplement additional tones to these two tones to form a complete **E^{maj7}-chord** E-B-E-G#-D#-G#, the dissonance is gone♣. Causes may be found in the many consonant intervals that this chord features, or in the destruction of the strong envelope fluctuations by the additional partials. What is interesting in this context: in the chord sheets of e.g. the book "Rock Gitarre" (Bechtermünz publishers), a different E^{maj7}-chord appears: E-B-D#-G#-B-E. These are the same note-designations as above, but the root position has changed. The chord rumbles a bit and does not ring with the same beautiful melancholy as the chord mentioned first above. But again this is a subjective assessment. In fact, there can be no wrong chords – only wrong expectations.

Fig. 8.28 shows both E^{maj7}-chords in comparison. The spectra are based on equal-temperament tuning; all partials have (arbitrarily) the same amplitude. In the second chord, two partials with only 9 Hz distance appear at 160 Hz – they generate a fast beating that sounds rather unpleasant. The neighboring partials at 415 Hz have a distance of 3 Hz: they beat, as well, but slowly and more in the sense of a vibrato i.e. less annoying. What's happening at 311 Hz / 330 Hz? Here we have the intended dissonance of the major seventh that showed up already in the first chord – given by the E- and D#-partials.



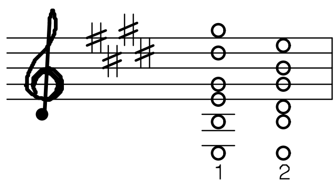


Fig. 8.28: Amplitude spectra and musical score of the E^{maj7}-chords elaborated in the text.

The closer two partials are spectrally located, the slower the resulting beats. Very small distances of partials (e.g. 1 Hz) happen in single notes, as well – due to slight detuning of the circular string-polarization, due to progressive spreading of partials, or because the instrument is polychoral (e.g. the piano). Somewhat faster beats (e.g. 4 Hz) may also appear for single notes, for example if the tone is generated using vibrato or tremolo. Even faster beats that are in part perceived as **fluctuation strength** [12] and in part as **roughness**, are typically only generated as several tones are played simultaneously. The borderline between fluctuation strength and roughness lies at a modulation frequency of about 20 Hz. Tones modulated that way – whether rough or fluctuating – can diminish the euphony and sound dissonant. As the modulation frequency further grows, the impression of dissonance decreases again – otherwise already the (harmonically complex) 100-Hz-tone would be dissonant (which it isn't). It may be deemed rough, but not dissonant. It cannot be specified by a single number at which distance between the partials a maximum dissonance occurs; the terms consonance and dissonance are too complex, and the sounds are too diverse.

♣ Again, this is naturally a matter of the approach taken, and may be subjectively judged differently in the individual case.

Psychoacousticians like the sensory to separate consonance and the musical consonance, or similar (often historically established) terms. Sensory consonance is represented in the absolute scaling, the "unbiased Scaling" that psychologists will readily put into question. So: put on the headphones, don't think of anything bad (and of course not of anything good, either), and evaluate the consonance of the two sine-tones presented. Just to avoid any misunderstanding: that is not pointless – from this we obtain elementary basic knowledge that may at some point form the fundamentals for a comprehensive theory on dissonance. However, it is still a long way from the dissonance of two sine-tones to the dissonance of a E^{maj7}-chord. That is true not only because here musical context, musical experience, and culture need to be involved (all elements of the musical consonance, also termed tonality), but also because already the purely psychoacoustical analytic poses considerable problems. Issues easily dealt with given an AM- or FM-tone turn voyage-into-the-unknown for a chord. As nice as the formulas about frequency- and level-dependencies of roughness and fluctuation strength are – they are of no help when dealing with signals containing complicated, time-variant partials. That E^{maj7}-chord has neither a modulation frequency nor a modulation index – just like the car engine the roughness of which has kept generations of acousticians busy. The synthesis of specific roughnesses proposed by Aures [Acustica, 58, 1985] shows a way but also reveals problems: we need to know not only the level of every partial (this could be measured) but also determine the phases of the partials because a cross-correlation is required across the specific roughnesses of neighboring frequency bands. That implies the time-functions, and therefore the phase is of importance. Your customary analyzer will, however, model only the damping function of the hearing-specific critical-band filters with reasonable accuracy. Not much is known yet about the (level-dependent!) phase response of these filters; and even if we would have that information, we would still have only captured one single dimension. Because: *One and the same stimulus may be perceived and judged very differently as a function of the variables implied by the total 'contextual' situation at hand* [Sarris].

If we don't pitch (sic!) our expectations that high and content ourselves with qualitative rules – then we can actually explain quite a few phenomena. Such as: if on a guitar the low E (open E-string) and the D# on the B-string (4th fret) are plucked with the fingernail, a dissonant dyad is sounded. The dissonance is significantly diminished if the fingertip is used for plucking. Explanation: dissonant beats may occur between the fundamental of the D# and the 4th harmonic of the low E. This will only happen, though, if this 4th harmonic is present with a sufficient level. Plucking with the fingernail or the plectrum will emphasize harmonics and generate a sufficiently strong 4th harmonic if the strings are not too old. Plucking with the fingertip, however, makes for a weaker excitation of the 4th harmonic – the dissonance thus is less pronounced. The markedness of the dissonance in this example is influenced by the playing technique (the interpretation) and cannot be determined merely on the basis of the interval. Of course, it is highly important how well the guitar radiates these neighboring partials, and how quickly they decay – and how the room transmits them … and whether further strings are plucked such that individual partials are masked. Roederer♣ describes a supplementary example: *if e.g. a clarinet and a violin play a major third with the clarinet playing the lower note, this interval sounds "smooth". If the clarinet plays the upper note, though, the interval sounds "rough".* The reasoning again lies in the instrument-specific structure of the harmonics which may not only be influenced by the mechanisms in the generator itself but also by the musician, the room and the setup in it, and of course by all other sources that may concurrently sound. In the end, a subjective assessment happens on the basis of the knowledge of the listener in relation to the musical context. The result is a highly subjective degree of dissonance the may certainly not merely be calculated just based on an interval relationship.

A tremendous range is covered from the older books on harmony[1] that attested to the major sixth a general dissonance, to more modern books[2] seeking to attribute this feature only to the minor second, from divine perfection and imperfect devil's notes via Helmholtz-ian tone-relations, all the way to multiple regressions[3]. They all share the search for rules, because: music is played according to rules … rather complicated ones at that, though. Auditory processing of acoustical signals also follows rules – and again the latter are all but simple … and they are subject to inter-individual as well as intra-individual scatter.

There are good reasons to assume that auditory perceptions emerge on the basis of audible partials. Audible are partials only if they surmount both the hearing threshold in quiet and masking thresholds caused by other tones. 'Audible' in this sense does not mean, though, that the partial would necessarily be *audible as individual tone*. To that effect, a partial is audible (i.e. it contributes to the overall hearing sensation) if the aural perception changes when the partial is filtered out. If the perception does not change, the corresponding partial is not audible. If we regard the interaction of individual partials as the source of the perception of dissonance, the (so defined) audibility of these partials is prerequisite. With this, however, dissonance becomes dependent on the individual sound spectrum and can by no means be calculated "from the score". If, conversely, the basis is the sound spectrum arriving at the ear, then orientating calculations are possible – albeit right now only with considerably reduced general validity. Daniel's[3] conclusion may serve to obtain three insights: roughness and sensory euphony are (negatively) correlated; roughness and unpleasantness are (positively) correlated, but: sensory euphony and unpleasantness are <u>not</u> correlated. Daniel moreover states: "this points to a significant difference between the opposite pairs *pleasant – unpleasant* and *euphonious – dissonant*". Daniel does not further delve into the subjects of pleasant dissonances or unpleasant consonances. It is now difficult, though, to repress the question of: what actually do subjects judge when asked about the consonance of a musical chord? Is it the pleasantness … or the euphony?

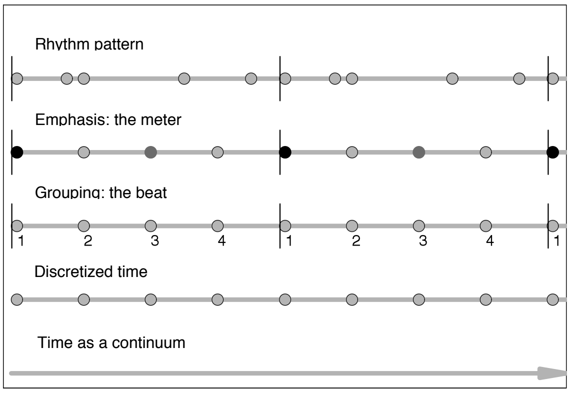
It is not a wonder that already 50 years ago Michael Dachs[4] arrived at this rationale: *in context, a consonating interval often gains a dissonant meaning*. Around the same time, Simbriger/Zehelein opine: *There is barely a second problem that would be as controversial in modern acoustics as that of consonance and dissonance*. While 50 years of supplemental research have considerably widened the body of knowledge available back then, an algorithm for calculating consonance that is at the same time manageable to the musician could still not be made available. Which is not necessarily a disadvantage: if you can feel it, you can play it. Oh yeah: those musicians, always having a solution at hand. And if you can't make it, fake it.

♣ Roederer J.: Physikalische und psychoakustische Grundlagen der Musik, Springer 1999.
1 Z.B. H. Grabner, Handbuch der funktionellen Harmonielehre, Max Hesses, Berlin 1950.
2 Z.B. F. Haunschild, Die neue Harmonielehre, AMA, Brühl 1998.
3 Z.B. P. Daniel: Berechnung und kategoriale Beurteilung der Rauhigkeit und Unangenehmheit von synthetischen und natürlichen Schallen, Universität Oldenburg, 1995.
4 M. Dachs: Harmonielehre, Kösel 1948.

8.5 Timing and rhythm

It's not only (apparently) world-weary John Rowles who lamented "If I only had time" – this is also *the* cri de coeur of every bad drummer. To stay in time is difficult, and even just to define time is not an easy feat. Usually, one refers to Augustine (a monk, not a drummer); his contemplations about time are widespread, and they are readily abbreviated to: "when nobody asks me what time it is, I do know it; as I seek to explain it to an asker, I do not." It is something like that when it comes to rhythm, groove, and timing. Even modern books on rhythmics♣ make do entirely without defining rhythm – and, yes, it is a challenge. If indeed someone tries, it reads something like this: *rhythm is the regular (periodic) repetition of accents that are pooled together.*

A pattern arises out of grouped accents – with an accent being a distinctive feature, i.e. for example the beats on a bass drum. Already now we can think of examples where this does not fit … in any case: if everyone can criticize this definition because he/she knows anyway what rhythm is, then an extensive definition is indeed superfluous. At least it is in the present context where the focus is on auditory perception, and not on teaching rhythmics. **Fig. 8.29** gives a brief outline on the hierarchical processing of continuous time, its discretization into basic beats, and the latter's grouping and accentuation. Based on this ordering scheme is the individual pattern that repeats within one bar in this example. In a two-bar pattern, two different patterns would alternate (Bossa) – but, again, the emphasis here is on the hearing system and not on the music.



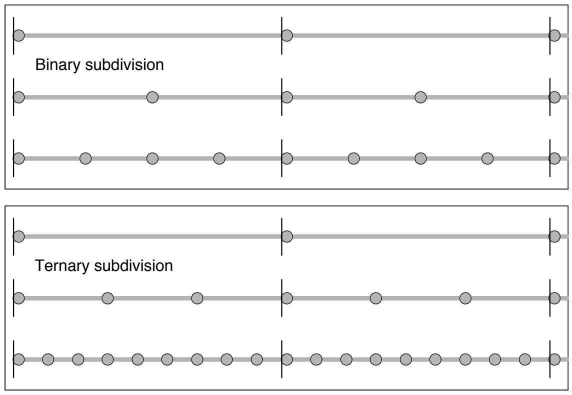


Fig. 8.29: Beat, meter, rhythm (left), binary and ternary subdivision of the beat (right).

In Fig. 8.29, the dots mark the start of individual notes – and the layperson believes that a musician with good "timing" needs to reproduce these starting points as precisely as possibly in order to receive the "playing like a machine"-distinction. Checking whether this is actually true shall be postponed to the next page; first, the focus is on the analysis: what is the accuracy that the hearing system can muster to analyze fluctuations in rhythm? Practical recording-studio experience provides the barely contested orientation value of **10 – 20 ms**. Timing errors of less than that quickly become meaningless. So, after all: the pro should hit his/her notes with a precision of about 1/100th of a second, and hard- and software in music needs to react very quickly in order not to make the ever-present signal-processing delays subjectively noticeable. Also: it must not be overlooked that an effects device with a basic delay of 7 ms is uncritical but 4 of them in series are not tolerable. For this reason some processors offer two settings: little time-lag (low latency) for playing live, and more time-lag (high latency) – but also much effect – for off-line processing.

♣ Marron E.: Die Rhythmik-Lehre. AMA 1991.

But now on to the actual topic of this chapter: **depending on circumstances, a good musician may need to do an objectively inaccurate performance so that it sounds correct subjectively**. The listener does the subjective assessment; the analyzer delivers the objective data. A good musician will not generally play his notes at the time as written but with a slight offset that is in part deliberate (determined) and in part unintentional (stochastic). Both (!) offsets are desirable and as such generic. This is why the sentence in bold above is not good as an excuse for the beginner – objectively wrong playing may indeed simply sound very wrong. But then: what is would be "correctly wrong"? Not in the sense of an aggravation or emphasis of "wrong", but rather: which deviation from the objective click of the metronome leads to a rhythm subjectively perceived as "good"?

Let's look at the determined (deliberate) deviations first. We have three main aspects here: the tone generation (transient, onset, attack), the tone perception (the auditory event), and the interpretation. Regarding the tone generation: 20 ms may have passed by the end of the tone onset (attack phase) of a wind instrument – or as much as 100 ms for low-volume-notes. Of course, it is not the moment when the player's lips open that counts as the onset of the tone, but a somewhat later point in time that can only be defined via the perception. Therefore the wind-instrument player needs to the start blowing before the tone is supposed to sound. When exactly the played note is considered to be existent – that is a decision made by the hearing system i.e. it is an act of the **tone perception**. In his book on psychoacoustics [12], Fastl lists eight examples for tones with different time-envelopes (TE), and determines the corresponding subjective start of the note. Only if a note has an abrupt onset of tone and immediately decays again (decaying TE) do objective and subjective starting points practically coincide. Given an increasing TE, the subjective start of the tone is up to 60 ms later than the objective one. For these special sounds, such numbers are of course dependent on the special experiment. However, even if all notes have a steep attack, we find astonishing differences in terms of seemingly equally long pauses: for the hearing system to assess a tone-duration as equally long as a duration of a pause, the objective duration of the tone needs to be considerably shorter relative to the objective duration of the pause! In Fastl's example, first an allegretto eighth-note and then an eighth-pause (both of 240 ms duration) are to be played. In order for tone and pause to sound equally long subjectively, the tone must not be played for 240 ms but for a shorter 100 ms, while the pause needs to be lengthened to 380 ms! For quarter-note and quarter-pause (nominal length 480 ms), the ratio is not quite as dramatic: tone duration = 260 ms, pause duration = 700 ms. The explanation of these indeed considerable discrepancies is found in the auditory transient processes (attack and decay) that extend the subjective length of a note (relative to the objective length) and thus shorten the subjective length of a pause. On top of these discrepancies (caused by the processing), the individual interpretation of the musician also needs to be considered. For example, given special stylistics, the "one" (1st quarter) will deliberately be played a bit early, or the "three" might arrive a tad later that it nominally should. This is not for a lack of exercise but to demonstrate individual virtuosity. If that weren't the case, all those 1000s of "Elises" celebrated on the pianoforte would have to sound identical.

It is in fact exactly this deviation from absolute rules that marks the virtuoso, the person "in the know" who understands not only where to deviate from the strict formula, but also how much. This knowledge often exists only implicitly i.e. without explicit awareness of it. If a virtuoso is asked to play the same passages, we will recognize always (almost) the same deviation. It is an expression of the personal style and not random at all. However, if we ask at which points he or she has shifted the "one", the artist will have difficulties supplying a complete list, and if we inquire about the degree of shifts, the answer is likely to be: *just as I feel it, I don't check the clock.*

In their work on swing-rhythmic, Lindsay/Nordquist♣ investigate an example regarding maintaining rhythm: **"Fever"** by Ray Charles (2004). The piece is dominated by a finger-snapping that puts the emphasis on the elsewhere often less accentuated even-numbered quarter notes (backbeat, more or less: iambus rather than trochee). These snaps are incredibly precise, as shown in **Fig. 8.30**: here, the envelope over the time reference is depicted (894 ms, corresponding to 67 bpm), and reference is hit with merely ±2 ms deviation. There is no further info but we can only surmise that some kind of assistance system was involved – it is hardly imaginable that a freely playing musician can after 3 minutes still be within 2 ms of the original time. That "Fever" – despite this machine-like precision – still never gets boring is to the credit of the continuously changing pattern played by other instruments. The bass basically plays half-notes, but already the first 4 bars reveal some of the bass notes locking into a ternary grid (splitting the fourths into thirds). The congas, as well, cannot do without the ternary splitting in their "da-dub-da". If merely accents on the four basic beats were allowed, the whole charm of the piece would be gone; it would be life-less and without that "swing".

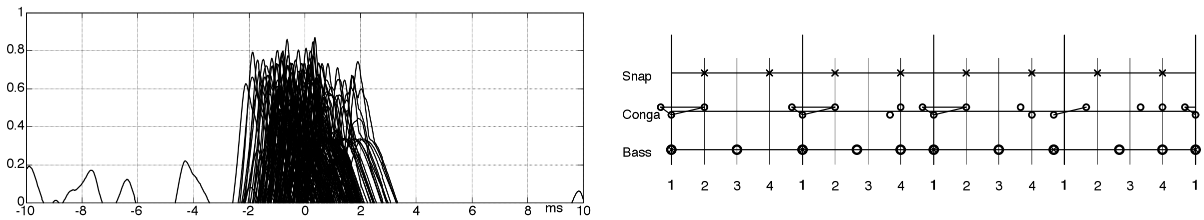


Fig. 8.30: R. Charles / N. Cole: *Fever*. Left: timing-analysis; right: four bars at the beginning of the piece.

"Hit the Road Jack" is another piece of Ray Charles' that provides aid to answering the question how precise the pro keeps time. In this version, Ray C. plays the first 1,5 minutes without accompaniment, and presumably also without click-track. He cranks up the tempo from 91 bpm to more than 95 bpm (beats per minute) – in a dance contest with rigid tempo-specifications (often as little as ±2%) that would be quite borderline. But hey, this is Brother Ray, and it ain't no dance contest, either: that piece needs to be exactly how he plays it.

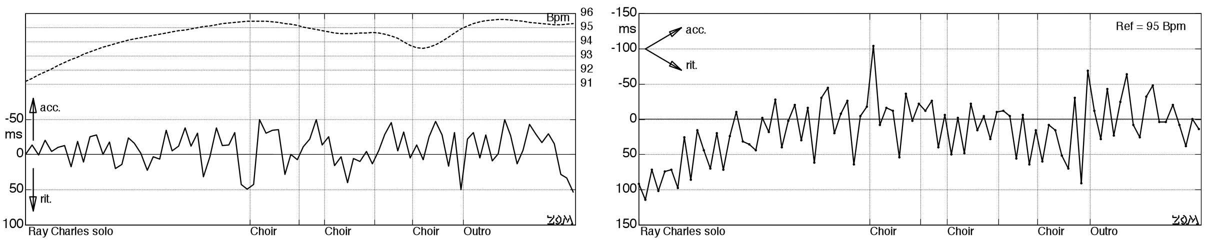


Fig. 8.31: Ray Charles 1981, *Hit the Road Jack*. Absolute (left) and differential-tempo (right) deviations. On the left, the (absolute) deviations are referenced to the smoothed tempo-model (---).

If we subtract out the slow tempo changes (as presented on the right of the figure), short-term fluctuations of a maximum of ±50 ms remain (there is some arbitrariness in that – of course other deviations will result when choosing another bpm-curve). A maximum of 50 ms in a bar of 2,5 seconds – that's all right … especially considering that the larger deviations are a good match to the form of the tune. The first entry of the choir is preceded by a minimal delay that in no way sounds off, but rather expresses the individual interpretation. Two hands on the piano and the voice generate a very lively rhythm that subjectively is perceived as correct – irrespective of what any timepiece says.

♣ Lindsay K.A., Nordquist P.R.: A technical look at swing rhythm in music; http://www.tlafx.com/jasa06_1g.pdf

Fig. 8.32 indicates that "Hit the Road Jack" may also be interpreted in a different manner. Same tune and same singer, but recorded 18 years earlier. Relative to the reference defined at 90,8 bpm, the tempo first minimally lags (-0,4 bpm), then catches up, and drops again towards the end. The deviations rarely cross the 15-ms-mark (it needs to be considered here that in contrast to Fig. 8.31 now quarter-notes are the reference). This is possible because the drums play along from the beginning, and their accents are easier to more precisely measure compared to the start of a piano chord. Neither video footage nor sound documentation reveals whether a metronome was put into service during the recording. Thus hypothetically: a metronome clicks along, the drummer realizes after about 18 s (towards the end of the second chorus) a slight time lag, counteracts and achieves perfect time again at the end of the 2nd chorus. We find larger deviations in the middle of each chorus – which fits the structure because each chorus consists of two halves: this would be a justification for that little swerve in their middle. It's a wrap – next take.

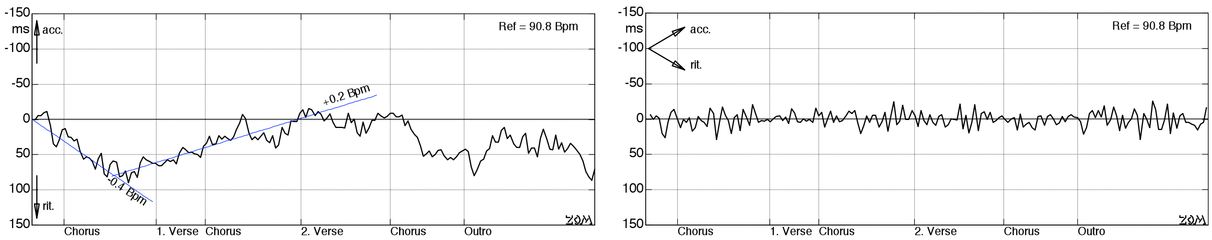


Abb. 8.32: Ray Charles 1963, *Hit the Road Jack*. Absolute (left) and differential (right) quarter-deviations.

Now on to the Eagles, the members of which are taken to be musicians that have a lot experience in the studio. At the start, their **"Heartache tonight"** exhibits a pronounced backbeat, i.e. an emphasis on the even-numbered quarter notes, generated by handclaps. In this tune, the handclap is not always present, therefore the snare drum was analyzed: **Fig. 8.33** shows only very small deviations in the quarter notes following each other at a distance of one second, with the extremes correlating with the structure. Here it can be assumed that a click track was used: both to achieve high precision but also because by now this is usual studio practice (post-processing simply becomes that much easier). This assumption does not at all seek to deny that Mr. Henley does precision work. Indeed, the presence of a click is no guarantee for a precise rhythm.

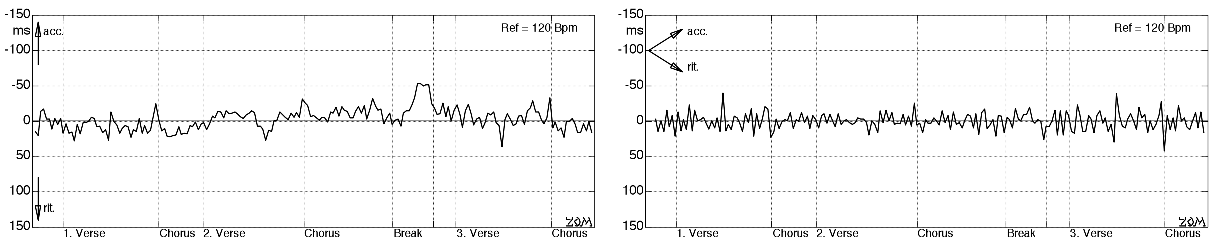


Fig. 8.33: Eagles, *Heartache Tonight*. Absolute (left) and differential quarter-deviations.

"Heartache Tonight" was analyzed as a 4/4-beat: the quarter notes are emphasized, and the counter-beats of the snare drum (offbeats) are on the even-numbered quarter notes. The subdivision of the quarter notes is ternary – it's a shuffle. Each 1st and 3rd quarter is given a grace note that sounds not on the eighth note before but is (due to the shuffle-partitioning) slightly delayed. In triplets notation, the second quarter would be divided in three parts; the grace note would then be located on the last third (eighth triplet). The same would be done ahead of the start of the bar.

Moving to the Rolling Stones, the "biggest Live-Band ever". Active since about 1962, and still rocking the scene♣. Somehow, anyway. When they recorded the live version of their **"Honky Tonk Women"** in 1969, they did have studio experience as well – but somehow in a different way compared to the Eagles[1]. The piece starts at about 103 bpm, speeds up mightily, and ends with about 120 bpm. Not that the guitar had accidentally started off too slowly – no, it's all supposed to be that way. Taking off with the heavy, earthy guitar riff, it ramps up and reaches operating temperature with the first chorus. Towards the second verse, the tempo eases off a bit, and then it's pedal-to-the-metal to the conclusion. Is that wrong? No way – it does groove. The deviations that, after all, are much larger compared to Fig. 8.33 do not stand out much, because the live recording offers a charming "togetherness" in particular in the chorus – it sounds like live recordings from that time simply were. Lively – but not wrong.

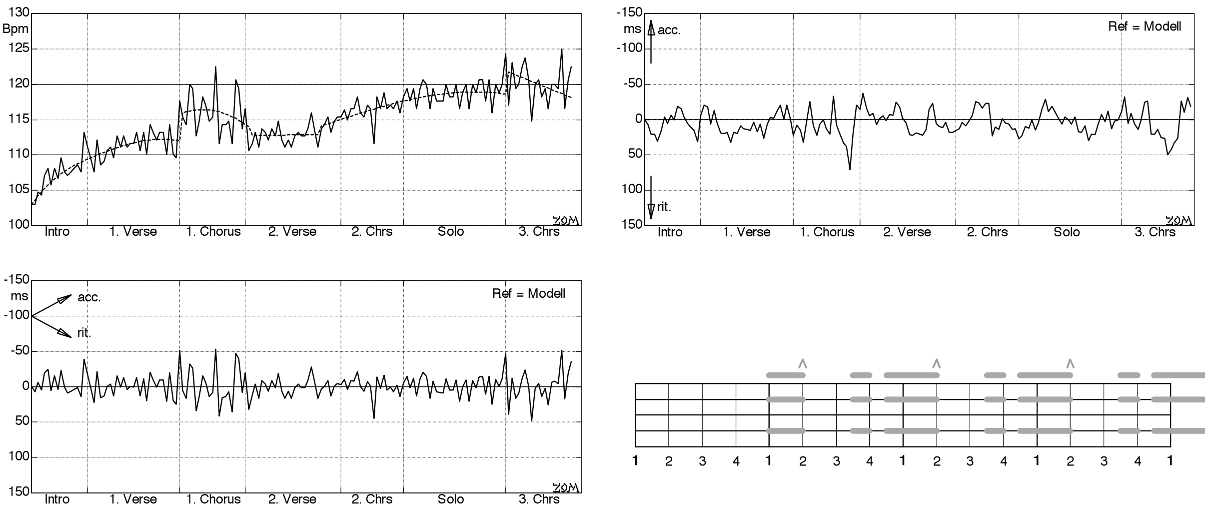


Fig. 8.34: Rolling Stones, *Honky Tonk Women* ("Get Yer Ya-Ya's Out!"-LP, 1969). Bpm-tempo (upper left), and absolute half-note deviations (upper right), referenced to the smoothed tempo model (---). The differential half-note deviations (lower left) also relate to the tempo model. Lower right: rhythm pattern (intro).

The world famous intro-riff is a nice example for the ability of the listener to detect the basic beat even if that is not even played at all. Only at the very beginning is the "one" emphasized in the intro, from then on the power chord is tied to the "four-and" and across the measure line to the following quarter note. The "three", elsewhere the other accent in the standard 4/4-beat, is not emphasized, either. Due to the accent on the "two" that immediately is interpreted as offbeat, the beat is found without any effort. In this example, the *quarter notes* with their time-distance of half a second are interpreted as basic beat; it is this rhythm that the listener's head synchronizes to as he/she "grooves along". To nod with the head two times per second is a very atypical movement. It would be possible to perceive the *eighth note* as basic beat but the corresponding head movement would already be too fast. However, the eighth-note tempo is a great match for drumming along with your fingers. You probably would not want to shake your whole human body in this tempo - but then that's a quite subjective decision. Customarily, 120 bpm is the best "groove along" tempo, which is why it is found often in dance music (moderato – allegretto). The dimensions and masses of the members of grown-up people specify – in conjunction with the spring stiffnesses – the natural frequencies of this "body"-system. And again we find: the tendency to oscillate is especially strong at resonance.

♣ Two musicians talk: „I've read that cockroaches presumably could survive an atomic war" – „Maybe – but Keith Richards would, in any case."

[1] *(Translator's note: about "live": the Eagles' live-version of "Heartache Tonight" is amazingly similar to the studio version – the point where one might ask how live the live-recording actually was ...)*

Not everything was entirely in time with the Rolling Stones, as shown by an early recording that (embarrassment-city!!) actually is called **"Time is on my side"**. The tempo remains at a rather constant 67,4 bpm during the first half of the song, and then suddenly drops to 65.4 at the start of the guitar solo (**Fig. 8.35**). That would not be a tragedy because we can see a motivation for this change. The short-term fluctuations that were analyzed using the beat of the snare drum are not really problematic, either (graph on the upper right). What is really off is the tambourine♣ that does get out of (time-) line again and again.

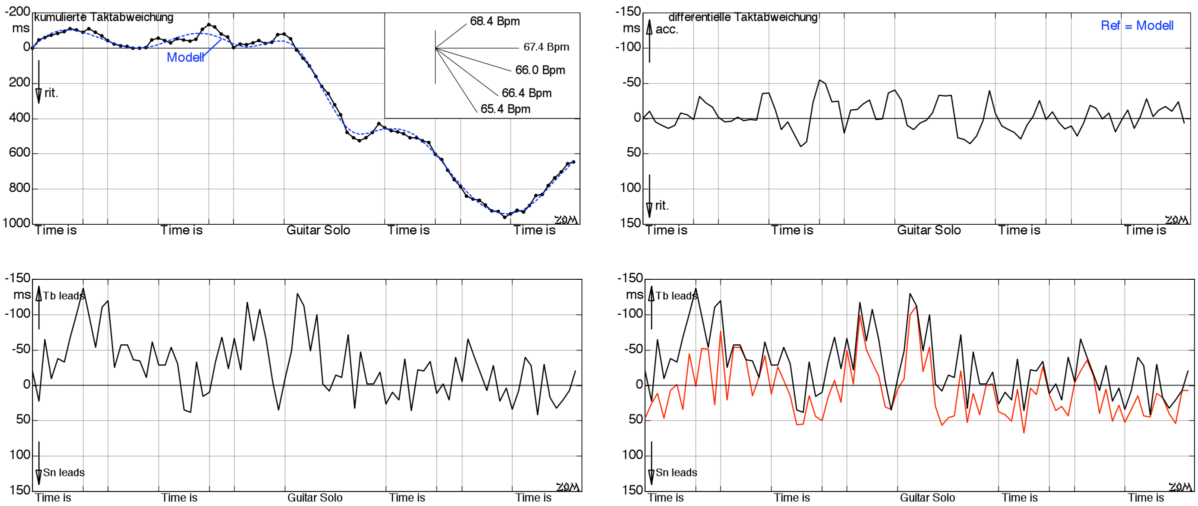


Fig. 8.35: Rolling Stones, *Time is on my side,* 6/8-beat; (Rolling Stones No. 2, 1964).

In the graph on the lower left, the time lag between tambourine and snare drum is depicted; it often surpasses 50 ms, and several times 100 ms – and that is audibly wrong. If the (still) tambourine is not hit *with the (one) hand* but the (other) hand is used to hit the (one) hand *with the tambourine*, then two sounds are created: one when accelerating the tambourine, and the other when it impacts/decelerates. Apparently, the latter playing mode was employed because often two tambourine hits are heard quickly following each other. The first one (shown in the graph on the left) is almost always early while the second one floats around the reference time (shown in the graph on the right as a bold line). So: is time on my (or rather the Stones') side? Not really, that was "out of time" even as early as 1964. Occasionally the two tambourine hits follow each other so quickly that they "melt together", and at other instances it will in fact have been only one single hit. Frequently, the time distance is more than 50 ms, and that's where (as an orientation value) the limit of the appearance of echoes lies. From a time distance of about 50 ms, single echoes become audible; repetitions with a smaller delay are pooled together by our hearing system into *one single* event. The haphazard change from single- to double-impact playing and the strongly fluctuating timing in this example – it is not perceived as stimulation anymore, but only and simply as timing errors.

A slightly delayed tambourine might be still acceptable because it is at least partially masked by the snare drum (accessory masking). However, the early tambourine beat (in the range above about 4 kHz mostly the only event) attracts much attention. This all the more because its percussive character marks a *point* in time – contrary to the slightly open hi-hat, the sizzle of which marks a *range* in time. Having the hi-hat ahead by a sixteenth (150 ms) – yes: that might also have worked. But then it should have been consistently ahead, and not: now on the "three", then on the "four" (together with the snare), and then again on the "three-and-three-quarters". Charley would surely have known all that – someone else must have Jaggered this.

♣ A head-less tambourine is what is meant – the sound of the jingles is discussed here.

As a résumé, let us jot down a few numbers: the limit of perception regarding group-delay distortion is around **2 ms**; for your usual music performance, this value is of no significance. From about **5 ms**, timing errors may be noticeable in individual cases – but only from about **10 ms**, the delay-range relevant for work in the studio starts. To avoid any misunderstanding: we are talking tone/note-onset here, and not a time-shifted superposition! Phaser, flanger, and similar effects devices (Chapter 10.8.3) do generate audible effects already at very short delay times, but these effects are noticeable because of the variations in the spectral envelope, and not as time-related in the first place. Given short repetition-periodicity, the perception threshold for timing drift may be as small as 10 ms (depending on the circumstances), but for rhythms that are not as fast it will extend to about **20 ms**. And that's it, then – even greater delay may indeed become problematic and sound wrong. Still: it may – but doesn't have to.

If the rules of timing could be summarized in a few lines, there would be great-sounding drum-machines only. Exceptions start already in a simple **¾-beat**: playing all three quarter-notes exactly on the regular time will sound wrong. Playing the second quarter note just a tad early creates a fit, making the listeners feel that all quarter notes have the same time-distance. A corresponding shift of as little as about $1/6^{th}$ of a quarter note will suffice, depending on the tempo. However, even with this shift of the second quarter note, the rhythm will start to become monotonous – for example if it's always the same drum pattern that is repeated. A drummer would indeed never repeat exactly the same beat; the drumstick will hit the skin at different positions leading to similar beats that still differ in the details. He/she would also introduce small variations: on top of the sound-color, he/she would vary the volume, and – yes – the timing, as well, in order to optimally support the musical piece. All this is alien territory for a drum-machine on its simplest programming level, and so it sounds just a primitive apparatus. Which it in fact is.

The **shuffle**, that galloping groove onomatopoetically described with "dumm-da-dumm-da-dumm", is an example for large time-shifts. Here, every eighth-note (if a 4/4-beat can be taken as a basis) is played later than the binary notation would require it. How much later – that is left to the artistic interpretation, and it determines whether the musicians in the ensemble play with or against each other. The musician ostentatiously playing a shuffle written "evenly" (i.e. binarily) just as evenly will kill the whole number although he considers him/herself as doing the right thing. Switching to ternary notation (triplets) helps only to some degree because this does not express, either, how strong the "shuffeling" actually is. That is decided by the "feeling", the experience, and expression of the musicians, and by their ability to empathize. That's empathizing with the interpretation of the co-musicians, and with the given piece of music. As a band "grows together" by frequent rehearsals, each musician acquires experience about how the others interpret the music, and in the end everyone "shuffles" so that they all match. That does not necessarily mean that they all play with exactly the same deviations.

Table: tempi (bpm)

Largo	40 – 60	Larghetto	60 – 66	Adagio	66 – 76
Andante	76 – 108	Andantino	100 – 108	Moderato	108 – 120
Allegretto	120 – 132	Allegro	120 – 168		
Presto	168 – 200	Prestissimo	from 200		

Table: meter

Iambus:	– o – o – o – o	Trochee:	o – o – o – o –
Daktyl:	o – – o – – o – – o – –	Anapaest:	– – o – – o – – o – – o

8.6 Loudness & timbre

Compared to an acoustic guitar, the electric guitar sounds different even if both play the same note. Pitch and tone duration may be identical, but given “species-appropriate” play, sound (timbre) and loudness will differ. This is due to the different way the partials evolve during the respectively sounding note, i.e. due to the individual, instrument-typical attack- and decay-behavior of the particular partials.

In simple models, each partial is assigned a frequency that is more or less an integer multiple relative to the fundamental (dispersion, Chapter 1.3). Per partial an initial level is specified, and also a decay time-constant that has the effect of an exponential decay of the amplitude over time, or of a linear decay of the level. However, more exact analyses show that most levels of the partials decrease according to more complicated functions; thus we have the basis for developing more complex models that define every (primary) partial as a sum of secondary partials. Other approaches are also possible – let’s remind ourselves here that the quantity of mathematically equivalent models is in fact not limited.

As we reduce the sound pressure level (SPL) of all partials by the same dB-amount, the volume drops; as we turn down the level of the higher partials by turning the treble control counter-clockwise, the sound gets dull – that’s well-known. It is much more difficult to answer the question *how* volume and sound depend on physical sound-parameters, and what would be the characteristics of a good or a bad sound to begin with. The volume of a tone (termed **loudness** in the following) is a monotonous function of the SPL-level (termed merely **level** in the following). Since the level is dependent on the power, “more power = more loudness” holds. Of course, we need to define this simple dependency in much more detail, because otherwise there’s the danger that the result of the consideration would quickly read: a 100-W-amp is louder than a 50-W-amp … however this cannot be the general statement.

First, we need to distinguish between amplifier power and sound power (or acoustic power). The amplifier power (that would strictly speaking have to be sub-classified into effective – or active, or wattful – power, reactive – or wattless – power, and apparent power) is the power that the amplifier delivers to the loudspeaker: e.g. 10 Watts (10 W). The largest part of this power is converted into heat by the loudspeaker (Chapter 11); only about 1 – 10% are converted into sound. For example, a highly efficient guitar speaker would convert 9 W of the 10 W electrical power fed to it into heat, and radiate 1 W as sound. In the immediate vicinity of the speaker, this acoustic power is concentrated onto a small spherical surface and generates a high intensity (the intensity is the power per area [3]). Assuming that the loudspeaker generates a short sound impulse, this results in an imagined spherical wave that propagates around the speaker and increases its radius (and thus its surface) with increasing time. Since the surface grows with the square of the radius, the intensity drops with the square of the distance. This is in the free, unperturbed sound field that we now focus on first. Because the intensity is in a square-relationship with the sound pressure, the simple $1/r$-law (**one-over-*r*-law**) is applicable: doubling the distance to the loudspeaker reduces the sound pressure by half, or as equivalent: the SPL drops by 6dB (more details in [3]). As an example: an efficient guitar loudspeaker generates an SPL of 110 dB at 1 m distance given an input of 10 W amplifier power. At a distance of 2 m the SPL is therefore 104 dB, and at 10 m distance it is 90 dB. If the objective is to generate not 90 dB at 10 m distance but 100 dB, the amplifier power needs to be upped to 100 W, and for 110 db it would have to be 1000 W. So, already here we notice the limits of this model that may remain linear only with regard to the sound wave – for the loudspeaker, load-limits need to be respected, the efficiency is of course power-dependent, and the speaker will die on us when overloaded.

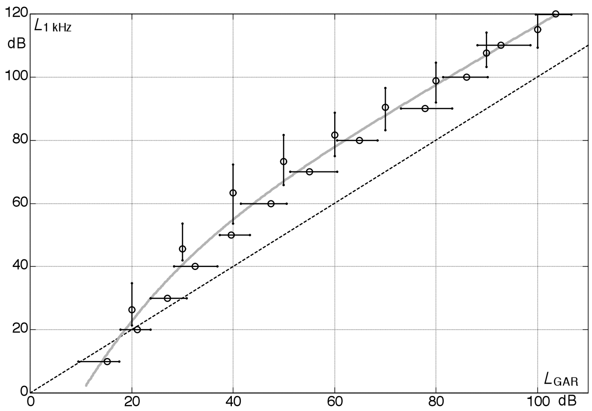
In the open, given unperturbed sound propagation, the level decreases by 6 dB per doubling of the distance. This fact is usually noticed with horror by the guitarist playing an open-air concert for the first time: that amp that was way too loud every time back in the club now is hopelessly drowned out all of a sudden. In the open, the reflections from the walls and the ceiling are missing – they lead to the sound reaching the listener not just once but (as echo) repeatedly. In the room, a superposition of free sound field and diffuse sound field is generated, with the free sound field dominating close to the loudspeaker, and the diffuse sound field dominating further away. The border between the two sound fields is represented by the diffuse-field-distance (also called reverberation radius). It amounts to a few meters in regular rooms (more precise information is found in [3]). Beyond the reverberation radius, the SPL stays independent of the location **within the room**; or so says simple theory. For the above example, this would imply: if the reverberation radius were 5 m, we would get (for 10 W input, and calculated starting from the speaker) at this distance a decrease in SPL down to 96 dB. In the remaining room (r > 5 m) the SPL would be 96 dB independent of the location. Of course, additional factors such as beaming effects, the actual geometry of the room, and the distribution of reflectors and absorbers would have to be considered – but this would go beyond the scope intended here. This example is to show that – before we start thinking about sound volume – sound source and room need to be looked into: which electrical power do we have, what is the efficiency of the loudspeaker, into what kind of room does the speaker radiate, and at last: where is the listener located? The SPL developing at the ear of the listener is the result of all these parameters, and from it – not just from the power of the amplifier – we can obtain indications for the generated loudness.

Psychoacoustics investigates the connection between SPL and **loudness**. Nowadays there is a standard for that – which is not undisputed. How loud you perceive a sound to be is a highly personal matter that is still interesting to science. And so we inquire with test persons (subjects) about their impression of loudness, we have them give categorical assessments (soft, loud, very loud), we make them perform magnitude estimates (double as loud as the reference sound), and let them determine thresholds (now the sound becomes audible). It is to be expected that not all human beings hear exactly the same thing, and neither that one and the same person will give the exact same response when asked again. This insight, however, will not be of much help – the psychoacoustician will want to know by how many dBs the level needs to be increased in order to make the subject perceive double the loudness. It is right here where the problems start: in fact, there is a multitude of experiments targeted to find out exactly that – but unfortunately there is also a multitude of answers or resulting models, not all of which generally correspond. Estimating the doubling or halving of loudness is a frequently practiced experiment from which the whole scale from *inaudible* up to *too loud* is assembled. Hellbrück [1993] has addressed this topic extensively and describes both the pros and the cons of the standardized loudness model of Stevens/Zwicker: power law, or exponential function? Stevens and his sidekicks had the subjects judge loudness relationships, and therefrom derived the **loudness power law** – it teaches that loudness depends on SPL according to a power law. In order to **double** the loudness of a 1-kHz-tone (in the level range > 40 dB), the **level needs to be increased by 10 dB** according to this law. Accordingly, upping the level by 20 dB corresponds to quadrupling the loudness, and +30 dB will match eight-fold the loudness. Recalculating this in terms of amplifier power: to double the loudness (and given linearity), the amplifier power needs to be increased by factor of 10 ten! Thus, compared to a 10-W-amp, only a 100-W-amp will be double as loud, and not a 20-W-amp. Still, a lot needs to be added here. To start with, the above law is applicable a priori only to a 1-kHz-tone. Then we find in Hellbrück's book the lovely but unsettling citation: *the possibility should be considered that the whole of the sone-scale is a pure artifact from psychometric methods that have been applied inappropriately and mindlessly.*

Sone, that's the unit for loudness. Mindlessly investigated? Let's not go there – psychologists and engineers will probably continue to bandy that ball for further decades. If we don't want to abort everything with the quite unsatisfactory insight that, due to the individual scatter, establishing an exact functional correspondence will not be possible, then what remains is forming statistical **mean values**. The difficulty is shown by an example from the beginnings of calculating loudness: during some auditory experiments it was noticed that broadband noise is much louder than a 1-kHz-tone although both have the same SPL value. Apparently, the SPL-value is unsuitable as a measure for the perceived loudness, leading to this question: by how many dB the two sounds will be different if both are adjusted to the same loudness? For the experiment described in [12], a special noise is used, the so-called *uniform exciting noise* (UEN) that may be imagined approximately as pink noise (kind of similar to a spoken long, slightly dark "sh"). One possibility to estimate the loudness is to present the 1-kHz-sinetone (e.g. at 80 dB) and ask the subject to adjust the level of the noise such that both sounds (presented alternately, not concurrently) are perceived equally loud. The reverse approach would also be possible; the noise is presented and the 1-kHz-tone is adjusted to the same loudness. Surprisingly, different values result from the two approaches even if the unavoidable small scatter is averaged out. There clearly is a **systematic deviation** (on top of the stochastic one): the adjustable magnitude is adjusted too high. For a presented 79-dB-noise, an adjustable tone is set to 90 dB, but for a presented 90-dB-tone, the noise is adjusted to 78 dB to be equally loud.

The measurements shown in **Fig. 8.36** give three results:
- For the two sounds to be subjectively of the same loudness, the level of the 1-kHz-tone needs to be in part more than 20 dB above the level of the noise.
- The results are dependent on the measurement procedures.
- The scatter is considerable.

In Fig. 8.36, the scatter is indicated as **interquartile ranges**; these represent 50% of the measurement values, with the values "above" and "below" discarded. As an example: 50 % of the subjects (the "middle" half) adjust the level of a 1-kHz-tone to an SPL of 83 … 97 dB for equal loudness with a 70-dB-noise, 25% of the subjects set the level to smaller than 83 dB, and the remaining 25% adjust the level to more than 97 dB. Additionally, the median value is given as a dot. We can unequivocally take from this experiment that noise is perceived louder that a 1-kHz-tone of the same level; however, the quantitative evaluation is subject to considerable scatter, and the latter moreover is dependent on the adjustment method. Psychoacoustics factors this in by defining two different loudnesses: a standard loudness level, and an object loudness level (that of the test sound). N.B.: the loudness comparison with the 1-kHz-tone historically was the first method to determine the loudness of any sounds, i.e. objects, via using a standard, i.e. the 1-kHz-tone).



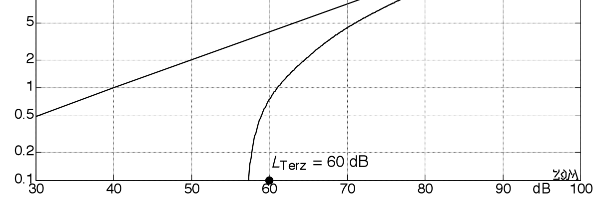


Fig. 8.36: Loudness comparison noise/sinetone (left), loudness with/without background noise (above), [12]. "GAR" = UEN, "Terz" = third octave band.

Keeping constant the level of the standard and making variable the level of the object (i.e. in this case the noise-level) yields the **object loudness**. Conversely, making variable the level of the 1-kHz-tone we obtain the **standard loudness**. The value interpolated between the two curves (the grey line in the figure) is called **interpolated loudness level** in older literature. The tem loudness level was introduced in order not to always have to talk about the "level of the equally-loud 1-kHz-tone" – rather, the **loudness level** with the unit **phon** is specified; the numeric value is that of the level of the 1-kHz-tone of the same loudness. Thus, if noise is perceived as equally loud compared to a 90-dB-tone (of 1 kHz), then this noise has a loudness level of 90 phon. This now makes the loudness sensation quantifiable – with numeric values that are difficult to interpret, though: 80 phon are not double as loud as 40 phon but 16 times that loudness. For this reason, additionally the **loudness** (measured in **sone**) was introduced. A 1-kHz-tone of 40 dB level serves as a reference point delivering the loudness of N = 1 sone. Since a level increase of 10 dB has the effect of doubling the loudness, 2 sone match 50 dB, 4 sone match 60 dB, 8 sone match 70 dB, and so on. Below the level of 40 dB, this correspondence is not valid anymore: in this range already smaller level changes have the effect to doubling the loudness.

The upper line in the right-hand section of Fig. 8.36 shows the relation between the level of the 1-kHz-tone (abscissa) and loudness (ordinate). Again: this is only for 1-kHz-tones – other spectral compositions necessitate other curves. Another prerequisite is that the 1-kHz-tone is presented by itself i.e. without other sounds being present. If the latter are presented concurrently, the loudness of the 1-kHz-tone may be **partially masked** i.e. reduced. The lower line in the right hand graph of Fig. 8.36 shows such a scenario: besides the 1-kHz-tone, a pink noise with the third-octave level of 60 dB is presented at the same time. If the 1-kHz tone has a high level (e.g. 90 dB), the two curves barely differ – the noise has little influence on the loudness of the tone. However, as the level of the tone is reduced (e.g. to levels below 57 dB), the tone becomes altogether inaudible because it is "masked" by the noise. Thus, when there is a masking sound present, the loudness grows more strongly with the level compared to the situation without masking noise.

For the practical musical performance situations we can learn from these relations that small variations in the sound power (e.g. +10%) are insignificant for the loudness perception. If the power of an amplifier is increased from 40 W to 44 W (and given a proportional change in sound power), we will – as a rule – not perceive a change in loudness. According to common practice, the just noticeable difference for amplifier power is estimated at about +50%. The difference between a 40-W-amp and a 60-W-amp is just about noticed – while doubling the power is clearly perceivable. Any musician deliberating whether to buy a 50-W-amp, or "for good measure" rather a 60-W-amp should be particularly weary of the efficiency of the loudspeaker. That is because, for example, a Celestion G-12-M is rated in the datasheet at 100 dB/1m while the G-12-M Greenback is rated at 97 dB/1m. Purely in terms of figures, the greenback requires double the power in order to generate the same SPL as the G-12-H. How these datasheets were established, is of course an entirely different story, and that (besides the loudness) the color of the sound (the timbre) plays a pivotal role – well, that opens yet another can of worms. It would go too far here to elaborate on all parameters that weigh in when determining loudness and timbre; those interested are recommended to read up in Fastl's book "Psychoacoustics" [12] – on 462 pages, it represents a comprehensive overview of the most important basics and models. The literature list in the appendix gives further info on related books.

The **color of sound** (timbre, sound- or tone-color) is the last sound parameter that we visit here. For many readers, it will be the most important one – but unfortunately it is also the most complex one. The sound-color – "the sound" – is being evaluated according to highly individual criteria, and trying to establish a model to calculate it always leads to failure. Of course, the sound-color depends on the sound spectrum, but already the metrological determination of the latter will be unsuccessful unless very simple sounds are analyzed. Harmonically complex tones are one thing, but a guitar solo played against a full accompaniment is another. Seeking to attribute roughness or fluctuation strength (based on modulation-indices and -frequencies) to a sound is futile because this cannot be determined in the guitar solo. Every spectral analysis may optionally be interpreted as a spectral weighing with the complex transmission function of a bank of band filters, or a convolution in time with the impulse responses of these filters. Bandwidth and impulse response cannot both be limited to a rectangular range, though, and thus every spectral analysis will lead to spectral and time-related **leakage**. The term spectral leakage intends to express that even the spectrum of a sine-tone is not measured discretely at on point of the frequency scale but as a continuously distributed spectral density. A Fourier series expansion is only possible in special cases (e.g. when the signal period is known), but this is meaningless in practice. Because the spectrum of the pure tone is presented in a broadened ('smeared') fashion, it is difficult to separate closely adjacent notes. Since spectral and time-related blur are reciprocal to each other, it would be possible to extend the duration of the analysis and thus to decrease the spectral leakage – but then the time-related leakage (describing the broadening – 'smearing' – along the time axis) increases. In concrete terms: if 1 Hz separation is desired in the frequency domain, the blur in the time domain is 1 s. The exact relation between the two quantities does not need to be deduced here[♣], for orientation $\Delta t \cdot \Delta f = 1$ suffices. If the analysis-blur along the time-axis is to be reduced to 10 ms, the spectral blur increases to 100 Hz. If we seek to, for example, extract from a musical piece the partials of the lead guitar, and therefore subject the wav-file to a DFT-analysis, it will be very difficult to decide which of the lines belong to the guitar, and which should be traced to other instruments. It may be possible in some cases, but fail in others.

Particular significance needs to be assigned to the "**attack**" (the onset of the tone). Many instruments can correctly be identified only via the structure of their attack; suppressing the first 100 ms tampers greatly with the sound. A good time- and frequency-resolution is desirable in this time range if the structure of the partials is to be meaningfully detected. The spectral and time-related leakage effects cannot be seen as errors per se; rather, they are kind of analysis-immanent artifacts. A Blackman-Harris-window is not more wrong or more right than a Kaiser-Bessel-window – it is just different. That, however, also means that one window modifies the structure of the partials differently compared to another window. If guitar tones were composed of harmonic partials of infinite duration, the analysis would be relatively simple. But they're not: the frequency relations of the partials are not integer multiples but they are spread out, and in addition they are slightly shifted (due to the frequency-dependent bearing impedances, Chapter 2). The amplitudes of the partials are not constant over time, and they do not decay according to simple functions, either. Moreover, the almost always present other instruments weigh in, as well, because pure solo-playing of any length of time does not occur much. Spectral analyses can certainly help to establish orienting impressions: are only odd-numbered partials dominant, how strong is the fundamental, do strong partials already stop at 1 kHz or do they extend up to 5 kHz? However, already with the evolution with time, with the fluctuations of the partials, it does get complicated, and the results of the analyses become dependent on the parameters of the analysis filters to a large extent.

♣ See e.g.: Zollner, M., *Frequenzanalyse*, Hochschule Regensburg, 2009;
or: Zollner M., *Signalverarbeitung*, Hochschule Regensburg, 2009.

Starting not with the spectral analysis of a whole ensemble, but recording and analyzing the sound of a single instrument played in the anechoic chamber, will usually result in spectra like those depicted in **Fig. 8.37**. They give the insight that e.g. a clarinet generates predominantly odd-numbered partials – this even being in good agreement with the wave mechanics of this aero-acoustic resonator (open on one side, i.e. “gedackt” pipe). The graphs on the left and in the middle stem from different books – both are supposed to show the spectrum of a clarinet. The graph on the right shows the spectrum of a cello. That the two clarinet spectra differ so much is not necessarily the result of grave measurement errors but easily due to the variability of this sound. Indeed, there is not “the” tone of a clarinet, and just as little is there “the” spectrum of a clarinet. We may be able to recognize characteristic differences in the cello-spectrum in Fig. 8.37 compared to the clarinet-spectrum, but these become meaningless in view of the spectral differences between the clarinets. Conclusion: **single spectra** hold little validity.

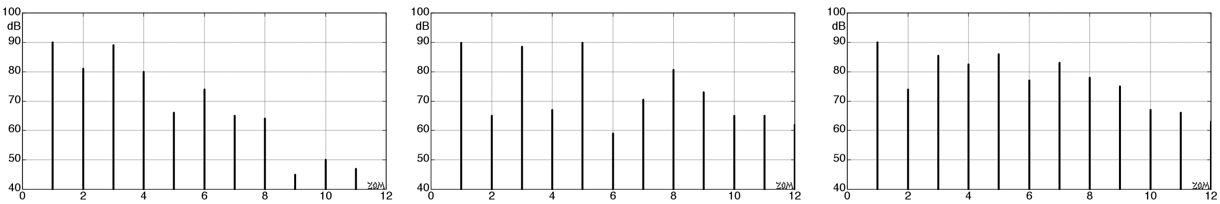


Fig. 8.37: Sound-spectra of some instruments: clarinet, clarinet, cello.

Regarding the sound of the violin, Dickreiter♣ elucidates: *the build of the partials of the violin is relatively irregular, i.e. it changes from note to note. The reason is found in the complicated resonance properties of the resonance body that strongly influence the material characteristics and the construction.* Thus, the spectrum of a d^1 may look entirely different from that of a g^1, and of course the relative position of violin and microphone plays a role since the radiation happens with a frequency dependent directionality. The first “electronic organs” sought to imitate the sound of specific instruments by generating periodic tones with a spectrum that had a supposedly instrument-typical envelope – such as e.g. the cello-spectrum from Fig. 8.37. It was more or less accepted that the resulting sound was only very remotely reminiscent of a cello; to sound “kind-of-electronic” was probably o.k. The main criticism was: the sound of simple organs is too “sterile”; it does not live – the instrument-typical beats are missing. The latter then were subsequently included via amplitude- and frequency-modulators (vibrato, tremolo), but again the result sounded artificial again because the effect was not relative to the partials but global. Only with the emergence of the sampling keyboards and the availability of huge solid-state memories could instrument sounds with an acceptable degree of naturalness be synthesized.

It’s not that the spectral representation would be entirely unsuitable to visualize instrument-typical characteristics – spectra can fully describe signals. It’s just that the information included in a single spectrum is too limited to already extract the instrument-typical from it. Typical is e.g. the presence of accompanying sounds that nevertheless contribute to the recognition of an instrument. The hammer-noise of the piano (“plock”), the blowing-noise of the flute, the scraping noise of the violin bow, the “squeak” in the horn attack, the impact of the strings of a bass on the fingerboard (drastically emphasized in the slap-bass style) – these are examples for such additional sounds, and there are many more. Typical are spectral maxima (**formants**) that are at a fixed frequency, or move along dependent on the fundamental; typical are time-related fluctuations of partials.

♣ Dickreiter M.: Handbuch der Tonstudiotechnik, Saur 1979.

All these characteristics aid the hearing system to categorize sound-colors, and to eventually allocate them to specific instruments. This then is done on the basis of learned knowledge – those who never have consciously heard an oboe will not recognize it, and only hear a strange nasal tone. Even those who in fact know how an oboe sounds will find recognizing the instrument difficult if one period is cut out from the oboe-tone and periodically repeated (*looped*). An oboe-typical spectrum is created – but it's out of typical context. In the auditory signal analysis (i.e. when we listen) the arriving sounds are automatically compared with known patterns stored in our memory. If the presently heard sound and the memorized one more or less match, the decision is made: sounds like an oboe, and/or like a musical instrument, and/or nasal, and/or dangerous, or whatever else could be found in the match. We can imagine the sound-color identification as a multi-stage process: in a first hierarchical stage, the inner ear determines the time-variant spectrum of the non-masked partials, i.e. the momentary sound-color – customarily described by *one single* spectrum. However, since (as taught be signal theory) a spectrum cannot be ascertained for a point in time but only for a time-range, the term *momentary* must not be taken too narrow a view on. The speech analytic evaluates sections of about 10 – 30 ms length, and it indeed is a powerful tool; as it is applied, it is often underlined that for the evolution of the hearing system, analyzing speech was even more important than analyzing music. That does sound convincing – but it does not mean that each and every musical analysis has to comply. For percussive sound, shorter durations of analysis may be purposeful, and for very low bass-notes longer ones, as well (because it allows for a finer frequency resolution). Still, an analysis-duration of 20 ms is quite workable as an orientation value; this means 50 spectra per second. These of course are not all identical but time-variant. On the basis of this spectral ensemble, the next-higher sound-color determination can happen which already yields more than just a "sound kinda like aaa". It could e.g. yield "sounds like a trumpet". In order for this already rather complex analysis to be successful, typical patterns about tone-onset, fluctuations, duration and decay need to be memorized. If the deviations are too big, the recognition algorithm fails. Cutting off the first 100 ms of a note will substantially lower the recognition rate; apparently already this short section includes important instrument-specific information that is not available in the later parts of the evolution of the note. Alternatively (and this is something we must not overlook), the cut sound will not be matched to the correct instrument because nothing about it has been learned yet (i.e. no corresponding patterns have been memorized).

In the processing stage still higher up, the evaluation steps can start that lead to the verdict: "sounds like Josh Redman", or "That be Hendrix on the Strat". Such judgments are, however, not part of the present reflections … so let us return to the color of sound, the timbre, and its signal-theoretical basis. We have already known for some time what the color of sound is NOT, and from this the following exclusion-definition originated: color of sound is that which remains if loudness and pitch are abstracted from. Alternatively, according to an old Acoustical-Society-of-America-definition: *color of sound is the perception attribute that still distinguishes two sounds although loudness and pitch are equal*. Somehow that feels like a trash-can-esque definition into which we can throw everything that cannot be defined precisely. Borrowing from optics helps to move along a bit: like we can objectively define visually perceived colors on the basis of spectral intensity distributions, the color of sound in auditory perception can be ascribed to the envelope of the sound spectrum. Like a picture consists of strung-together locally distributed color spots, the tone of an instrument consists of momentary timbres strung-together sequentially in time. We need to allow for the fact that this comparison will arrive rather quickly at its maximum load and hit a wall – the two sensory channels do, after all, exhibit strong differences besides some similarities.

In order to explain the possibilities and limits of the spectrum-based analysis of tone color, a dyad shall serve: two added-up sine-tones (300 Hz, 312 Hz) of equal level that are abruptly switched on at t = 100 ms (**Fig. 8.39**). The time-function would therefore be:

$$x(t) = \sin(\omega_1 t) + \sin(\omega_2 t) \;=\; 2\cdot\sin\left(\frac{\omega_1+\omega_2}{2}\cdot t\right)\cdot\cos\left(\frac{\omega_1-\omega_2}{2}\cdot t\right) \qquad \text{Beating}$$

Already this simple example exemplifies that there is more than one possibility of representation for every signal: the **dyad** may either be seen as the **sum** of two tones, or as the **product** of two other (!) tones. Instead of adding a 300-Hz-tone and a 312-Hz-tone, it is also possible to multiply a 306-Hz-tone by a 6-Hz-tone. A spectral analysis merely and always disassembles the signal into its additive components, and not into its multiplicative components, showing one 300-Hz-line and one 312-Hz-line in the spectrum. The 6-Hz-envelope that is so nicely revealed in the time function (Fig. 8.39, upper left) remains hidden in the spectral analysis. Even the 300/312-Hz-pair-of-lines will only be represented as two separate lines for suitable analysis parameters – and since there is an infinite number of parameter-variants, there will be an infinite number of spectra.

The long-term spectrum identified for $-\infty < t < \infty$ is pointless; rather, the **spectrogram** obtained by shifting a short window-section is required (**Fig. 8.38**). In the left-hand graph, a rectangular evaluation-window is shown; it is slid across the signal as a multiplicative weighing (over time). From the signal weighed this way (shown at ***b***), the DFT-spectrum is calculated as a function of the time-shift. Since undesirable jumps occur at the window-borders for this type of window, the rectangular window is not applied in practice; windows with a rounded-off shape are customary.

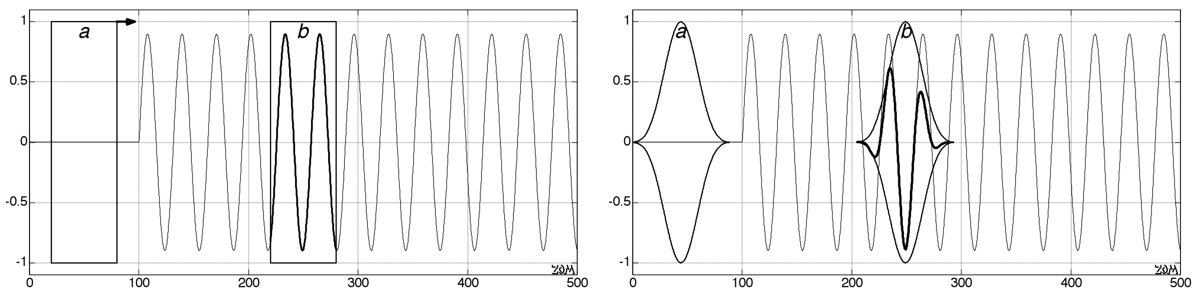
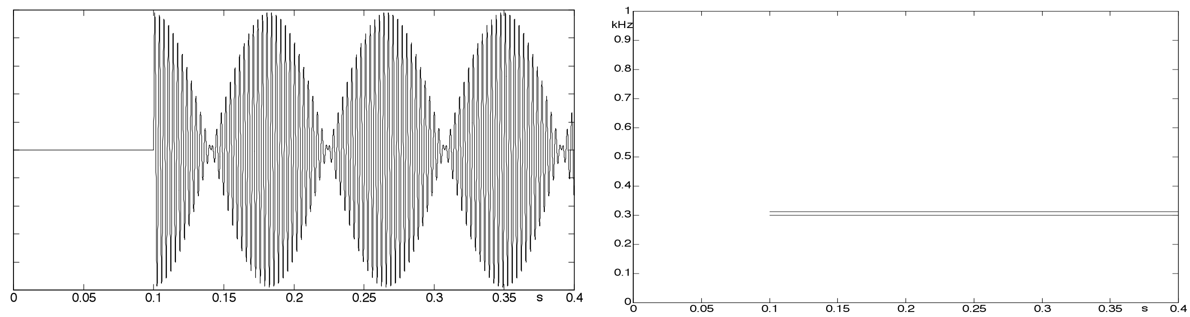
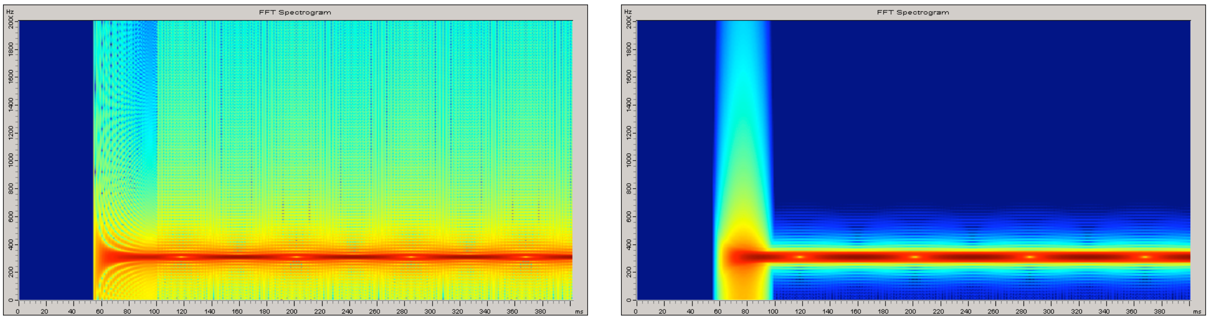


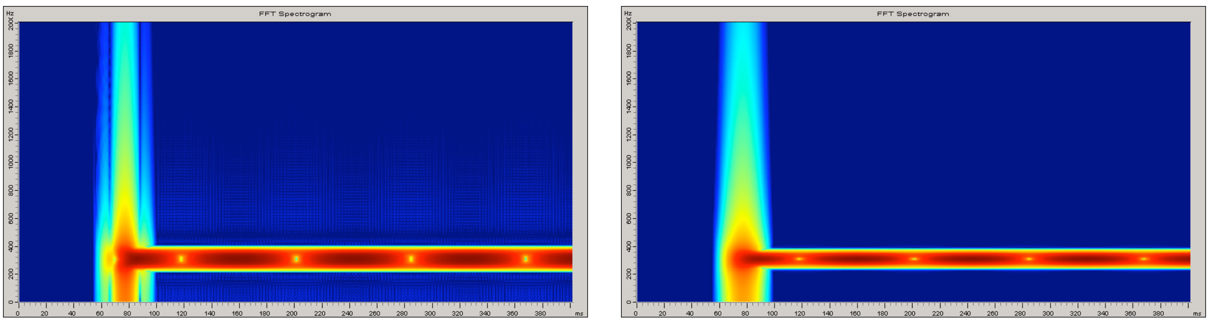
Fig. 8.38: Time-function of a sine-tone; with two different weighing windows.

However, a fundamental problem still remains with the window-weighing as shown on the right (here: Kaiser-Bessel window): the spectrum is determined based on the windowed (i.e. modified) signal. **Fig. 8.39** shows – for the two-tone signal mentioned above – spectrograms derived with different windows. The signal was identical for each spectrum; the differences stem exclusively from the different analysis-parameters. The window-length is specified by the point-number N, a frame-length of 46 ms belongs to N = 2048. The time specified as abscissa in the color-spectrum marks the beginning of the window. Since the width of the latter is not 0 but e.g. 46 ms, we understand why the analysis pushes the start of the dyad ahead e.g. to the 54-ms-point – although both sine-tones are switched on only at the 100-ms-point! At exactly this time shift, the start of the signal falls into the rectangular window, and therefore the corresponding spectrum also starts from 54 ms. Increasing the number of points to 4096, the window-length grows to 92 ms, and the spectrogram (linked to the rectangular window) starts at 8 ms.

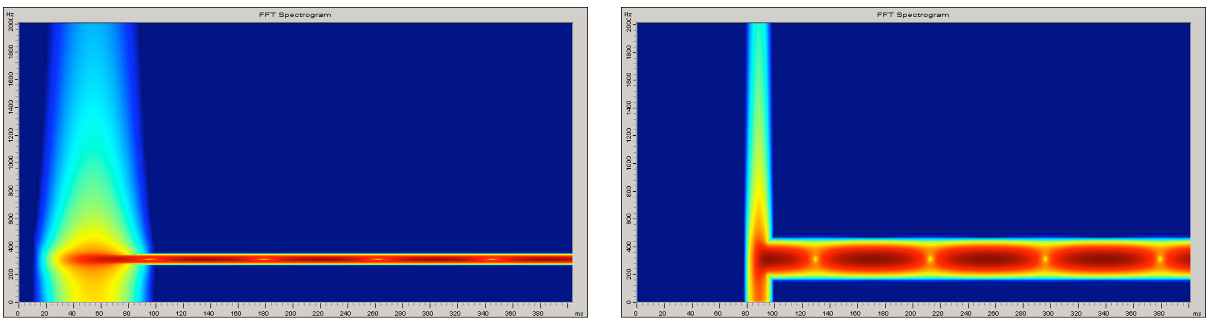




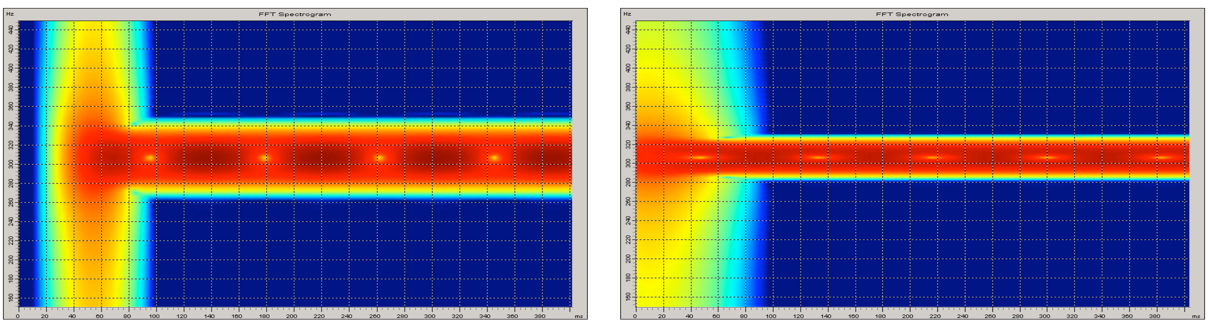
Rectangular window, $N = 2048$ Hanning-window, $N = 2048$



Flat-Top-window, $N = 2048$ Kaiser-Bessel-window, $N = 2048$

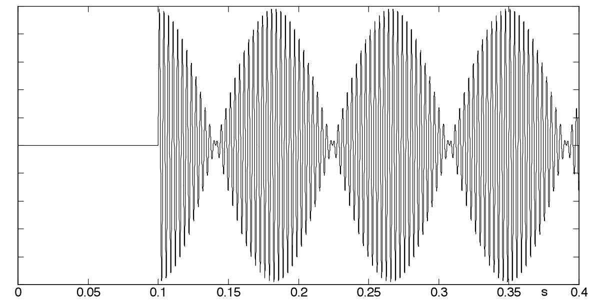
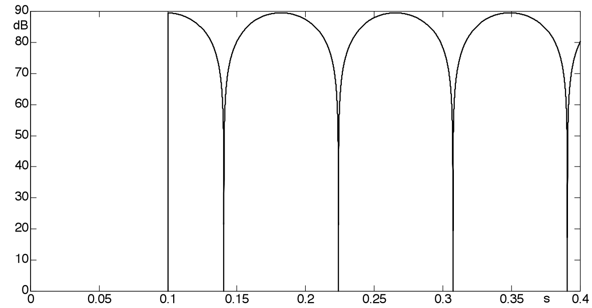


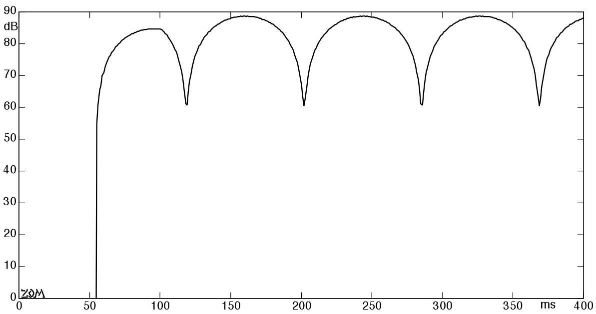
Kaiser-Bessel-window, $N = 4096$ Kaiser-Bessel-window, $N = 1024$

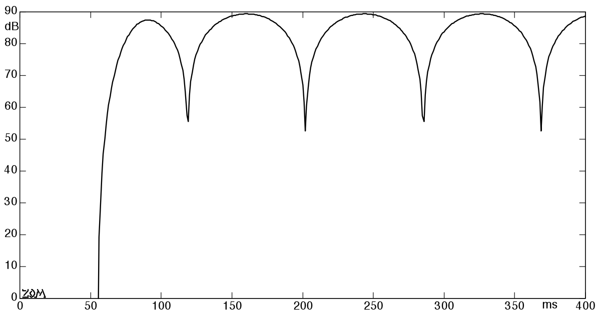


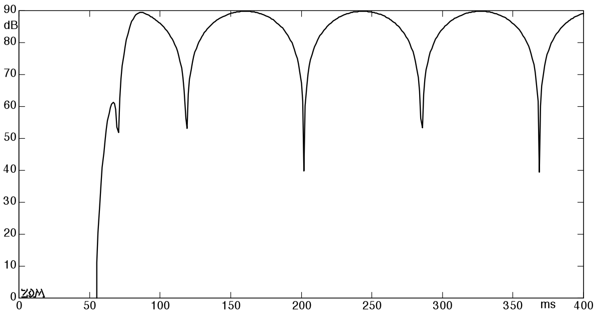
Kaiser-Bessel-window, $N = 4096$ Kaiser-Bessel-window, $N = 8192$

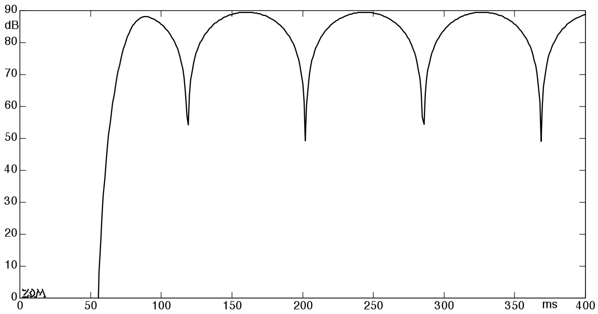
Fig. 8.39a: DFT-Spectrograms of an abruptly switched-on beating (300 Hz / 312 Hz), ΔL = 90 dB.

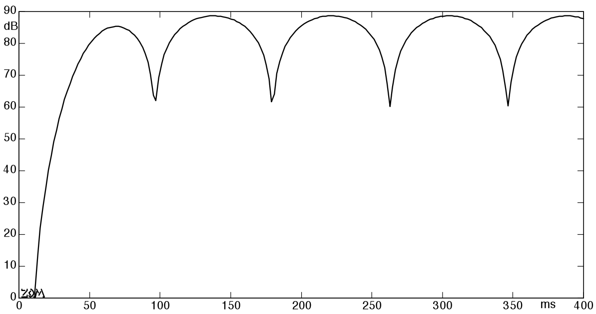



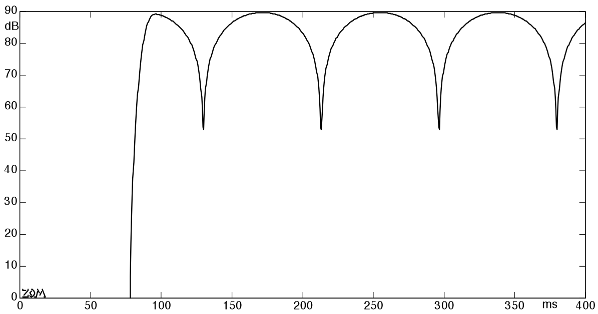

Rectangular window, $N = 2048$

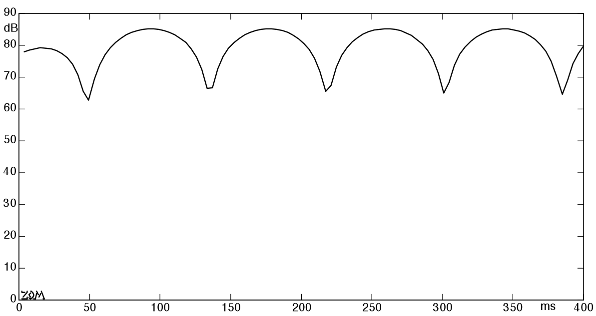

Hanning-window, $N = 2048$


Flat-Top-window, $N = 2048$


Kaiser-Bessel-window, $N = 2048$


Kaiser-Bessel-window, $N = 4096$


Kaiser-Bessel-window, $N = 1024$


Kaiser-Bessel-window, $N = 8192$

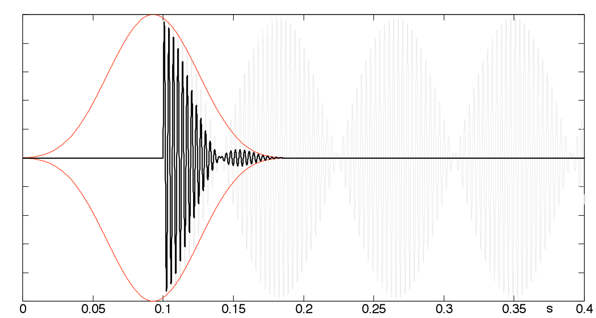

Kaiser-Bessel-window, $N = 8192$, time-domain

Fig. 8.39b: DFT-level graphs (at 306 Hz) of an abruptly switched-on beating (300 Hz / 312 Hz).

It would be possible to scale the time axis such that $t = 0$ specifies the *end* of the window; in that case the corresponding shifts would show up at the end of the signal. Just to be clear: it again needs to be emphasized that this is not a software error of the analysis program, but a system-immanent artifact of all spectral analyses. Depending on the window-length (= on the impulse response of the filter), the analyzed signal becomes longer. Moreover, changes result in the direction of the ordinate, as well: the switching-on click as vertical streak, and the spectral leakage as vertical broadening of the spectral lines. In fact, from 100 ms there should be two lines running in parallel towards the right, as shown in the top-right graph; instead *one single* streak is shown. The simple reason: for $N = 2048$, the analysis bandwidth is too small, and the two lines cannot be represented separately. If we take the bandwidth as the reciprocal of the window-width, we obtain the bandwidth of $\Delta f = 22$ Hz – that is too broad for a line distance of only 12 Hz. For the Kaiser-Bessel-window (**Fig. 8.40**) used in the following, we moreover need to consider that the effective duration is only about 1/4$^{\text{th}}$ of the frame length; and that the effective bandwidth therefore will be about four times that of the rectangular window♣.

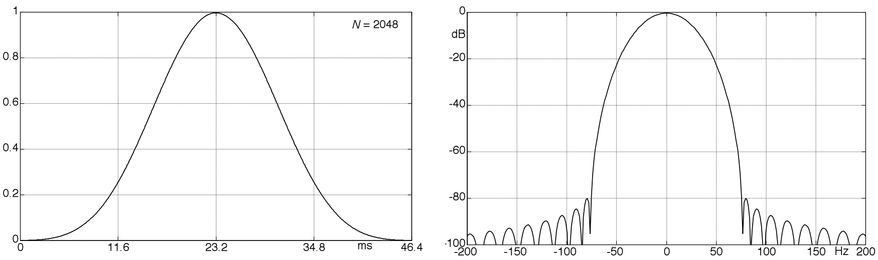


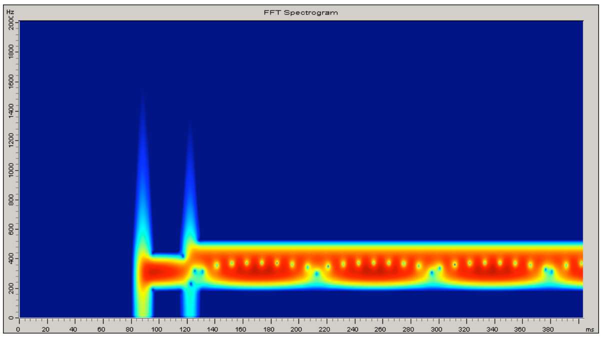
Fig. 8.40: Time-function (left) and spectral function of the Kaiser-Bessel-window, $N = 2048$, sampling frequency: fa = 44.1 kHz.

If indeed time-related and spectral leakage have effects in every spectral analysis, it stands to reason to ask whether the like would not appear also within the hearing process – after all, the signal is broken up into its spectral components there, as well. And sure, leakage will of course be present, too. However, because the **auditory filters** are adaptive and non-linear, we cannot specify *one* bandwidth and *one* attack time – things are more complicated. Too complicated for the present explanations that are merely intended as an overview, and therefore reference is made to specialist literature, e.g. Fastl's "Psychoakustik" [12, available also in English language]. The hearing system processes two tones of large frequency distance in separate channels, while tones close in frequency are jointly processed. The two-tone signal mentioned above cannot be separated into its two components by the auditory system, and one tone of quickly fluctuating loudness is heard – i.e. as product, not as sum. We hear something that does not actually exist in the spectrum: a 306-Hz-tone! Already this simple example proves how difficult it can be to extrapolate from a spectrum to the auditory perception. It is not entirely impossible; the parameters of the analysis can be adapted, after all. Therefore **Fig. 8.39** includes different analyses, with varying window-types and -lengths. All show the **switching-click**, to start with. The longer the window, the longer the switching click. It has to be that way: if, during the shifting of the window, the signal-start just about falls into the window, it is only an impulse of very short duration that is analyzed – the spectrum of which is necessarily broad-band. The more the window is shifted beyond the signal-start, the longer the signal to be analyzed (windowed), and the more narrow-band the spectrum. Is the switching click audible? No! In any case not as the figures would let us assume. It therefore is purposeful not to show the color-spectrum with a dynamic of 90 dB (as is the case in Fig. 8,39) but with only 40 dB: visual and auditory impressions are a better match that way.

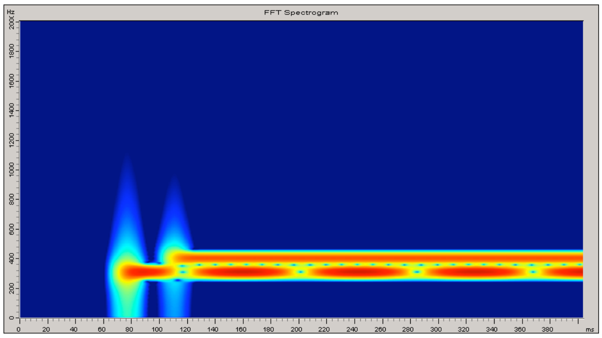
♣ We will not investigate in detail here what is to be understood by the term „effective".
More details may be obtained from: M. Zollner, *Signalverarbeitung*, Hochschule Regensburg, 2009,
as well as from: M. Zollner, *Frequenzanalyse*, Hochschule Regensburg, 2009.

We now take a look at the fast fluctuations that can be clearly seen in the time-function. They also appear in a time-section of the spectrum, in the so-called **slice** (level over time with fixed frequency, Fig. 8,39b). Forming the logarithm of the envelope yields the curve shown in the graph at the upper right, and the evaluation of the DT-analysis yields the graphs below. Again it is clear that the time-related leakage has the effect of very differently shaped level curves – depending on the window-type and -duration. Thus we retain: the **DFT-analysis** delivers a multitude of different spectra that – to begin with – allow for only few conclusions regarding the perception of the sound. Supplementary algorithms enable modeling of hearing-typical assessments (auditory critical-band filters, contouring algorithms, spectral and time-related masking), but the scientific investigations have yet to arrive at a true breakthrough.

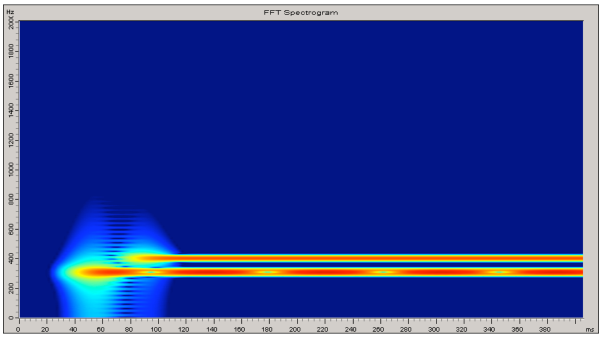
The two-tone signal analyzed in Fig. 8.39 already revealed the fundamental issues found in any spectral analysis. Yet, it is a very simple signal – instrument tones are of considerably more complex build, not to mention chords or tutti-sections. Compared to the latter, the **triad** analyzed in **Fig. 8.41** is still rather simple: three added-up sine-tones of equal level but switched on at different times. The 300-Hz-tone and the 312-Hz-tone are switched on at t = 100 ms, and the 400-Hz-tone comes in at t = 134 ms. Analysis is again done using the Kaiser-Bessel-window, the level dynamic in the figure is, however, reduced from 90 dB to 50 dB (compared to Fig. 8.39).



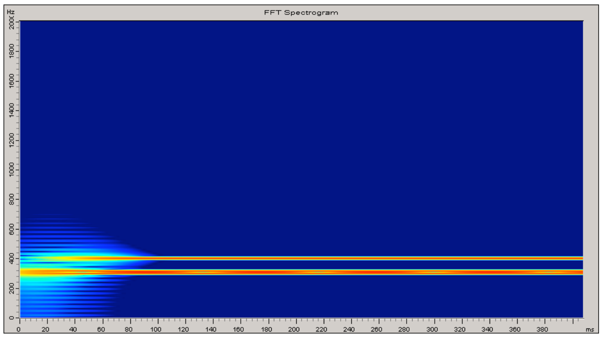
Kaiser-Bessel-window, N = 1024



Kaiser-Bessel-window, N = 2048



Kaiser-Bessel-window, N = 4096



Kaiser-Bessel-window, N = 8192

Fig. 8.41: DFT-spectrograms of a triad (300 Hz / 312 Hz / 400 Hz). ΔL = 50 dB.

The tone onset blurs as the window length increases, but the spectral separation improves in turn. The latter does not need be all that great, though – because with this triad, again only <u>one single</u> tone is heard. Not a sine-tone but rather a lively bubbling tone-mixture – with *one single* pitch. Only when listening repeatedly, one could also tend to hear an oscillation between two pitches … but certainly not anything like what the analysis done with N = 8192 would suggest.

As powerful PCs became available, the desire developed for sound analyses to – at last – depict “the correct” spectrum, meaning not just 512 lines but 4096, or even 16384, for good measure. The latter number implies, however, that the sampling window (at a sampling frequency of 44.1 kHz) has a length of 372 ms, which is too long compared to the hearing system even when applying the (shorter) effective length. For sound analysis, N = 4096 represents a tried and tested compromise that offers the basis for supplementary DFT-analyses and post-processing. The latter is urgently required: the 2k-analysis shown in Fig. 8.41 gives the impression of two sound-parts starting at different times. Objectively seen, this is indeed correct: a beat from 100 ms, and a sine-tone from 134 ms. Our hearing system, however, does not care: it perceives one single tone-onset and not two. Even when the two partial sounds start with a delay of 70 ms between them, they are not heard separately in time. The simple reason is that the beating in the dyad impedes the recognition of the time-structure. Only from an offset of about 100 ms, the additional tone coming in with the delay in this example (!) is recognized as such (compare to Chapter 8.5, though).

Not to stick exclusively to synthetic tones, let us now turn to a real guitar tone: **Clapton’s intro to “Stepping Out”**. The guitar plays by itself a number of times – this facilitates the spectral analysis a lot. **Fig. 8.42** shows spectra and time-functions: in the upper two lines of graphs those of a G_3, and below for a C_4. That’s four times that “same” G_3, but with considerable differences! Clapton’s sound may not be described with one single spectrum, after all – and that is the same for J.H., R.B., G.M. and all the other big names: virtuosity implies change, and that holds for the spectra, as well.

Still, we of course can wring a few commonalities from the G_3-spectra: they all feature a gap between 1 and 1.5 kHz, and a spectral maximum between 1.5 und 2 kHz. This is the range where the (second) formants of the vowels “ø” and “y” (using the definitions of the international phonetic alphabet, IPA) reside, so these tones can be attested an ø- and y-like timbre. Moreover, the strength of the low partials is notable: there are neither exclusively even-numbered, nor exclusively odd-numbered partials. And finally: the brilliance of a single-coil-guitar (which would feature a resonance of 3 – 4 kHz) is not achieved; rather we have a strong, mid-range-y, trumpet-y sound … or a saxophone sound, or a cello-sound with flute-like harmonics? Journal-literature – (rightfully) praising this phase of Clapton’s as pure genius – has found, and still finds, many comparisons. It seems strange that to describe a guitar sound, one would have to borrow from the realm of wind instruments, or strings – but maybe in the far distant future, a trumpet instructor will shout at his pupil: *blow with more emphasis on the mids; more like Clapton’s guitar sound*!

Irrespective of whether trumpet- or cello-like, what does determine that sound and its variance that appears even for the same notes? First, let’s look at the second part of that question which is easier to answer: even when fretting the same string at the same fret, the sound depends on the location of picking, and on the movement of the plectrum. And on the plectrum itself – although that was certainly not swapped during one take of the recording. The angle of the plectrum (parallel or slanted relative to the string), the basic movement (up- or down-stroke), the angle of the movement (relative to the fretboard), place of picking (closer to or further from the bridge – these are all sound-determining parameters. Then there is how the left hand is at work: even slight bends can make partials vanish into interference-gaps. That is why the four analyzed G_3’s are not identical, and that is why there is no “one” G_3-spectrum, and not “the” C_4-spectrum, either, and least of all “the” Clapton-spectrum. Not to forget: guitar, cable, amp, room, and recording technique of course also influence the sound – but these would be time-invariant per recording … presumably, EC will not have jumped back and forth between amp and mike. But then, come to think of … one could surmise that some musicians ……

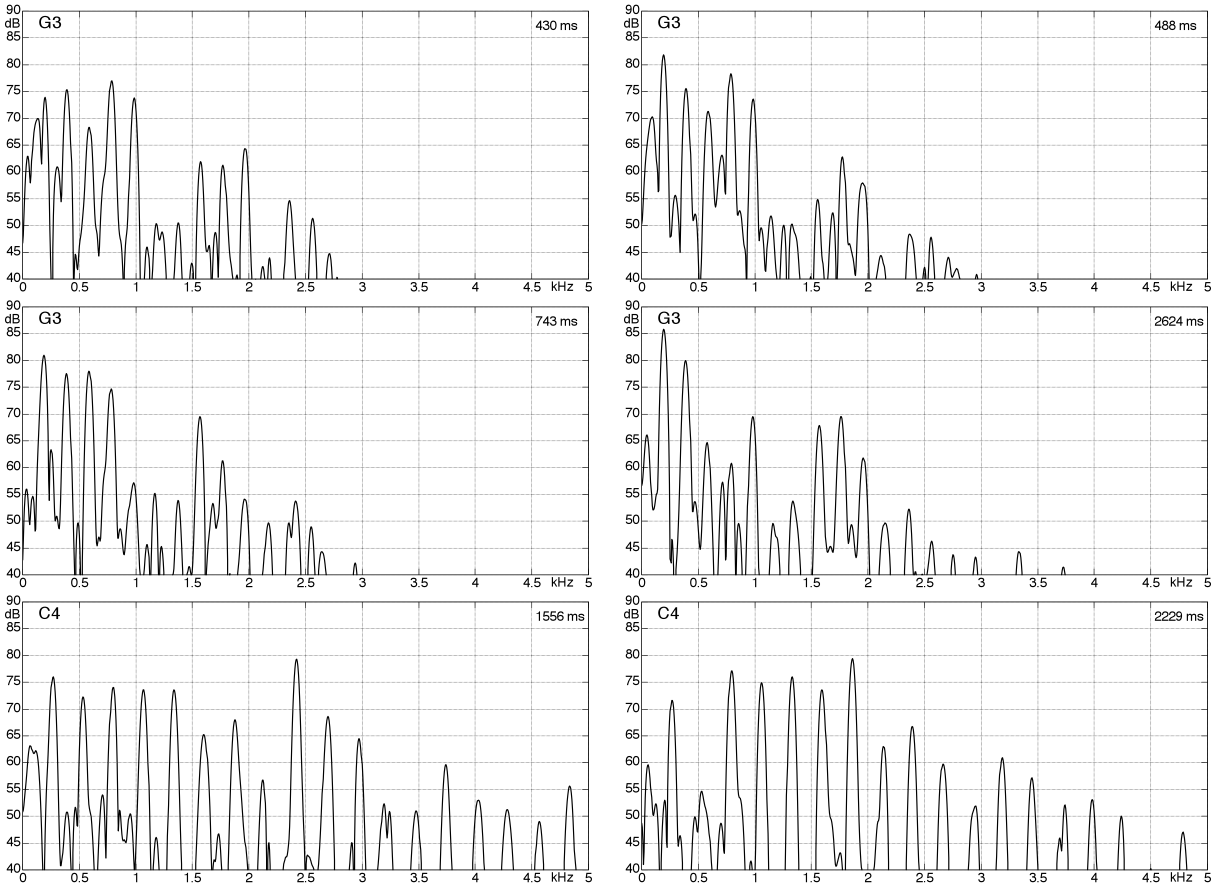


Fig. 8.42a: Individual spectra for the spectrograms in Fig. 8.43a. Kaiser-Bessel-window, N = 2048.

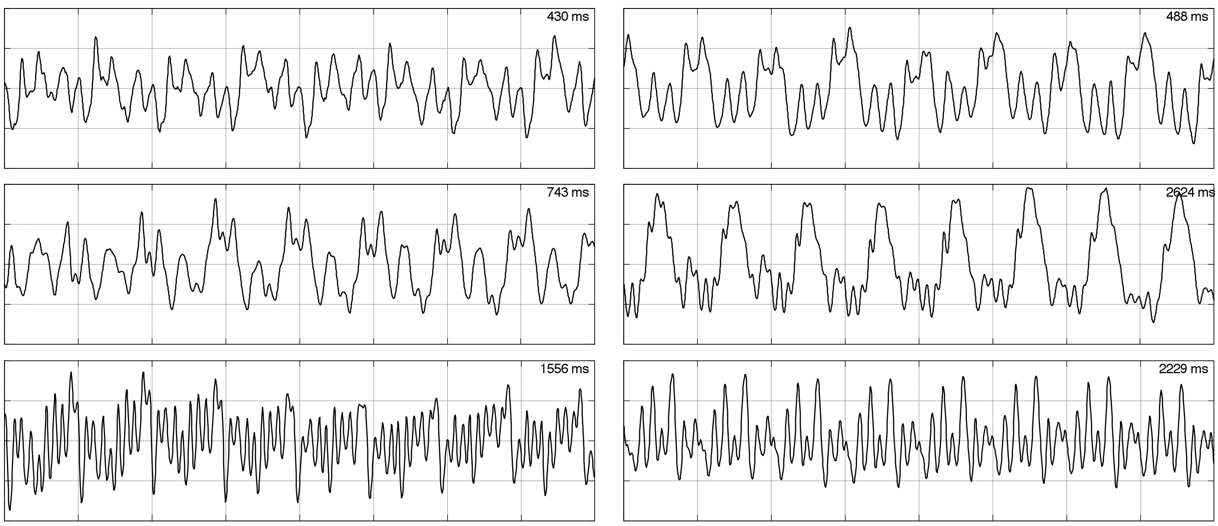


Fig. 8.42b: Time-functions for the spectra in Fig. 8.41a. Time grid = period of fundamental.

But where exactly do we now have the analytical proof for Clapton's "unique" or at least "groundbreaking" Bluesbreaker-sound; what is so special about these notes and their spectra? Ultimately: nothing at all! Listening to them in isolation, cut out of the intro, they sound plenty unspectacular. Maybe like a trumpet, or like a cello, or even synthetic. It gets interesting only as a group of notes is sounded, as soon as every note is presented with its attack- and decay-processes only existing in full context. However, it is exactly those processes that elude any spectral analysis that could reasonably be interpreted.

Somewhat easy to detect is the “lingering” of individual (in fact already terminated) notes which is due to the strong amplification. The **color-spectrogram** in **Fig. 8.43a** shows this, and it aurally creates the impression of a mighty, fat, powerful sound that can be reigned in only with difficulty. However, the short attack-noises limited to 20 – 40 ms duration yield all those problems that have already been described in the context of the dyad- and triad-signals. Of course it is possible to calculate corresponding spectra, but they will be highly parameter-dependent. Dear PhD-students who are just now in the process trying to cut another facet into the diamond that is psychoacoustics: don’t let yourselves be discouraged by that! EC needed 21 years to produce these sounds – you don’t have to have them analyzed within 2 days. Sure, it is not impossible, but simply applying a bank of Gammatone filters with contouring-algorithm – that ain’t enough. Here’s a hot tip: do *synthesize* the sound using the supposed partials, and listen. This approach very quickly reveals, which formation-rules are verifiable but not relevant to the auditory system, and what might constitute a “groundbreaking” sound. And speaking to the gear-heads: you won’t do anything wrong bringing out that original ’58 (or was it a ’59, after all?), but absolutely necessary it is not. Required are the right fingers, the right micro-timing, the right bends. “Clapton is God” was the writing, not “Paula is goddess”. This is illustrated by many EC-epigones appearing on Youtube, covering *Stepping Out* with at times remarkable equipment (but at times showing dismal timing, too). It becomes quite clear that the finger-work is much more essential than the question of “R8 or R9?”.

It is time to come back to the starting point of this chapter: to the timbre (or tone color). The latter may without doubt be determined on the basis of a spectrogram – but in infinite variations, because there are infinite possibilities to parametrization of spectrograms. If we do not want to test all of then, then an overlapping 4k-DFT with Kaiser-Bessel-window for the steady-state part of the guitar-tone will deliver some first orientation values. The onset of tone (attack) is more difficult to analyze because here the spectrum can change as much as 20 dB within 10 ms – a typical case of conflict between time-domain-resolution and frequency-domain-resolution. If several instruments sound at the same time, the analysis becomes particularly difficult. For the graph in **Fig. 8.43a**, only a single guitar plays, and the behavior of individual partials can clearly be observed. This behavior is, however, difficult to measure since these partials rarely maintain their frequency, not even in the seemingly steady-state part of a note. We find subtle up-bends (at around 1000 ms), down-bends (also called pre-bends, around 1900 ms), and half-step bends (around 1600 ms). Thus, it is not sufficient to set the cursor on the 180th DFT-line and to analyze how its level evolves. This would again be merely the behavior of the level of this DFT-line but not that of a special partial – the frequency of the latter is changing, after all (e.g. from the 180th line to the 191st DFT-line). Contouring- and pitch-follower-algorithms (which one indeed is that “closest neighbor”?) are applied to assist in this scenario, which is another reason for the multitude of parameters. Once these problems have been solved (it is, after all, not impossible to track partials), new challenges present themselves: the partials not only change their frequency but also their level! And not too little or too slowly, at that: we see e.g. 6 dB / 10 ms. Mind you, the attack- and decay processes of the DFT-analysis may run with the same speeds. Thus, if we change the DFT-parameters, the level fluctuations also may change. This multi-variant analysis (or optimization) would go far beyond the scope intended here, and so what can remain is merely the qualitative statement: the partials change their amplitude and frequency even within one single played note. At least the frequency shift is a global one (all partials change their frequency by the same percentage), but the amplitude shifts are partial-specific. Not all frequency- and amplitude changes are audible; there are absolute thresholds, masked thresholds depending on neighboring tones, and pre- and post-masking in time. Only that which is above threshold is fed to the final post-processor that then forms – among other things – the timbre, the tone color.

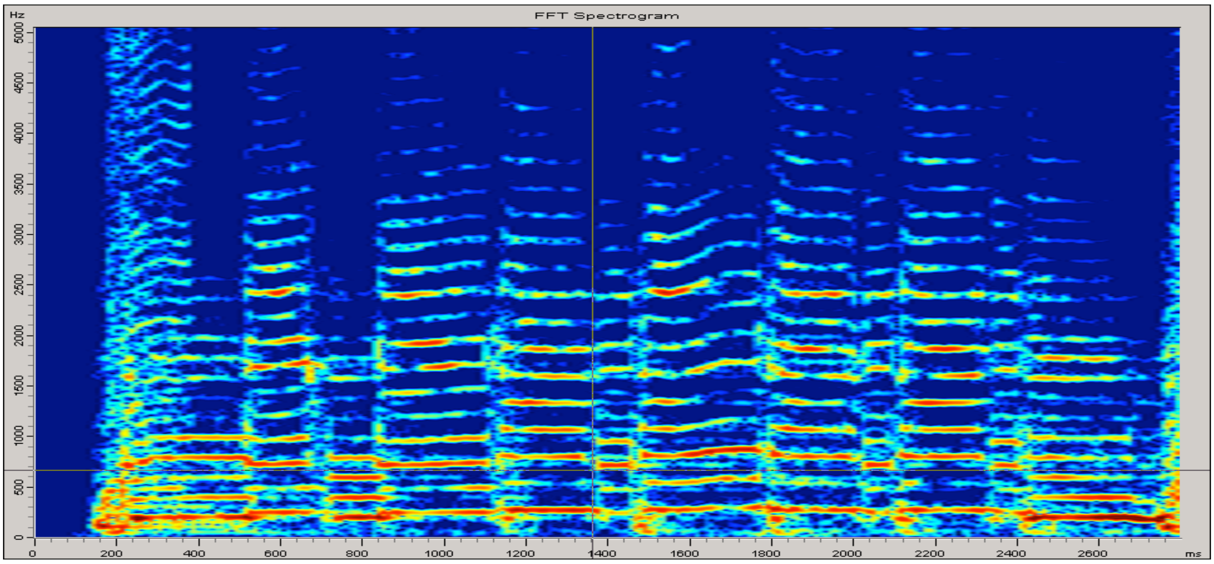


Fig. 8.43a: Excerpt from *Stepping Out* (Mayall / Clapton), guitar-intro. ΔL = 40dB.

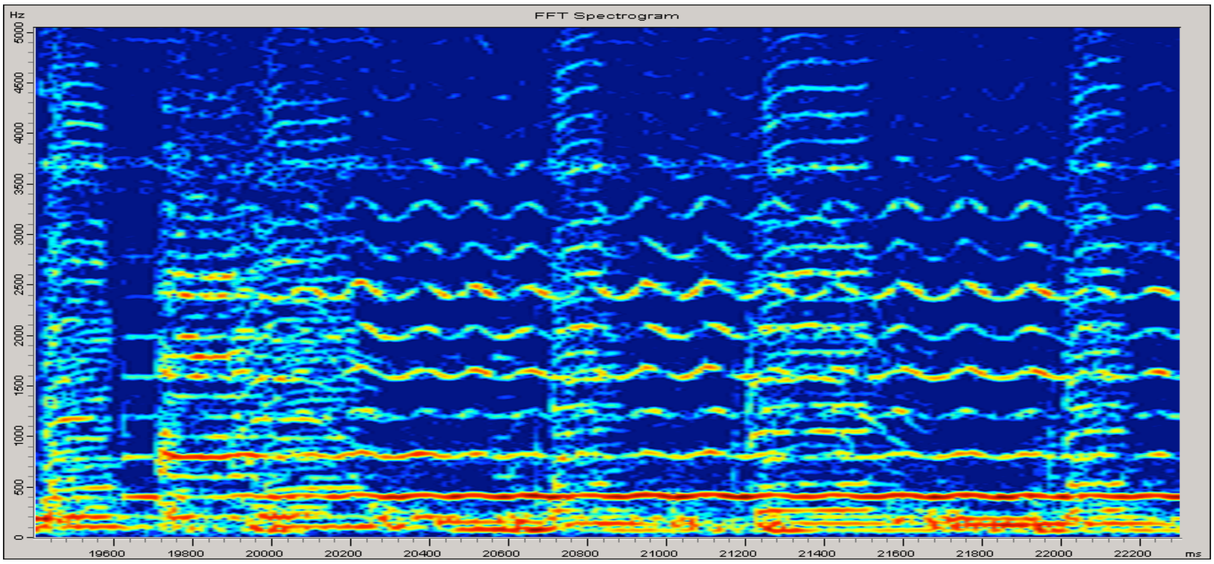


Fig. 8.43b: Excerpt from *Stepping Out* (Mayall / Clapton), guitar note with finger vibrato (7 Hz). ΔL = 40dB.

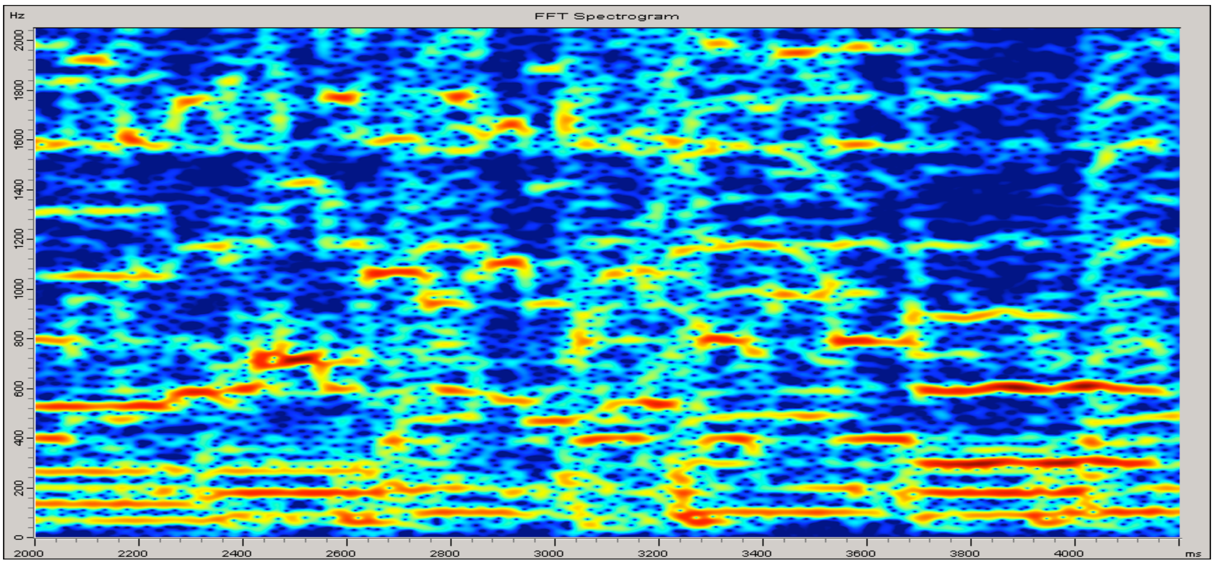


Fig. 8.43c: Excerpt from *Stepping Out* (Mayall / Clapton), fast eight-note triplets. ΔL = 40dB.

All in all: not a trivial analysis. This is not supposed to sound too discouragingly, therefore let's quickly look at **Fig. 8.43b**. In this graph, the visual analysis is facilitated by a gestalt-law that helps in auditory tone-recognition, as well: the low of common fate (see also Chapter 8.2.4). All lines that move back and forth in synchrony are partials of a guitar note, in between the horns provide (vertical) accents, and the electric bass lays the foundation below. It may be added for the Strat-purists: no, you do not need a whammy bar for that; this is done with a left-hand finger. To bend a note by ± ¼-step with a modulation frequency of 7 Hz – that is Clapton at his best. In **Fig. 8.43c**, things get more hairy again. This is one of the passages with faster playing, and vibrato is not really possible with note-durations of as small as 100 ms. In this section, already the pitch-tracking is a true challenge, not to mention an automatic timbre-analysis.

(Translator's note: the following paragraph only makes sense and works for German speech sounds <u>and</u> words. It was impossible to find suitable correspondences in English without a complete re-write/re-draw. I have tried to make sense nonetheless, using again the International Phonetic Alphabet – IPA – where necessary …)

If we do not want to wait until research offers reliable algorithm, we can only resort to onomatopoeia as it has been practice for centuries. This is an effort of pattern matching between the spectral maxima of the guitar tone to be described, and those of a speech sound (formant = frequency of a spectral envelope-maximum). From this, it suddenly becomes understandable that a "flute-like" *("flöten-artig" in German)* guitar sound does not need to unconditionally sound like a flute. Maybe that guitar sounds merely like a spoken "ø" *(as in the German "flöte")*, it "fløøøøøtes" without being that instrument. The corresponding (second) ø-formant is at 1500 Hz. It may be a bit higher up, if a female speaker is assumed (N.B.: it's *she the Paula*, after all). It wouldn't be counterproductive, either, that the famed blue Cøløstiøn-speakers have a maximum in their transmission curve around that frequency.

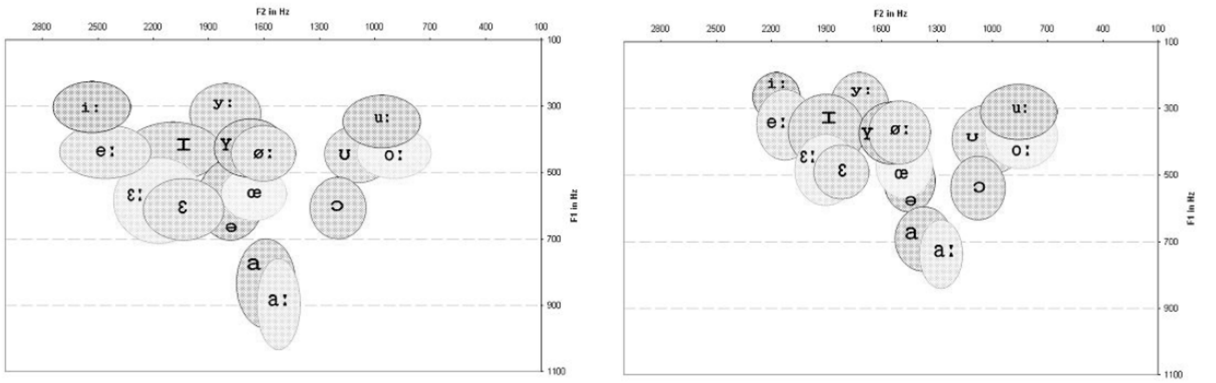


Fig. 8.44: Formant-frequencies of the German language, f/m speaker; (Sendlmeier/Seebode, TU Berlin).

In short: timbre (the tone color) depends on everything involved in tone generation. Not only that: the subjective assessment criteria of the listener play a role. To objectively visualize the sound that generates a timbre, the SPL-time-function is a complete but rather unsuitable and abstract quantity. The hearing system does not directly process the time-function, but a short-term spectrum determined according to complex rules. The phase is of secondary importance in this short-term spectrum; the behavior over time of the spectral amplitudes yields the primary hearing-relevant data-set. From the latter, and with suppression of masked (inaudible) ranges, a secondary above-threshold data-set is derived. Contouring-algorithms (maximum-detection), curve-following- and grouping-algorithms join what belongs together, and enable – on the basis of memorized knowledge – recognition of instrument-typical characteristics: timbre, pitch, and loudness, among others. There is a good deal of arbitrariness involved here: whether strongly modulated tones are attested a fixed pitch with a special modulation timbre, or a variable pitch with a fixed timbre: that is under the sovereignty of the listening "subject" .

8.7 Auditory experiments

The predominant part of this book discusses the function of the electric guitar by way of physical laws, documented via formulae and measurement protocols. This enables us to explain e.g. wave propagation, induction and signal filtering – but not the actual effect on the listener. The verdict of the latter is only made available in auditory experiments. Therefore, the following seeks to give a short summary of methods towards controlled sound appraisal.

8.7.1 Psychometrics

Psychophysics forms an interdisciplinary scientific area bridging **psychology** (= the science of sensory perception, among others), and **physics** (= the science of natural processes); it researches and describes the connection between physical stimuli on the one hand, and the sensations and perceptions caused by these stimuli on the other hand. **Psychoacoustics** narrows the wide area of physics down to sound phenomena, and connects the "science of sound" with the "science of hearing". **Psychometrics** is a sub-area of psychology that has specialized in the (in particular quantitative) measurement of sensations. Electrical voltage is measured with a voltmeter, temperature is measured with a thermometer – but how can we measure the sensation of sound resulting from listening to a guitar? This can work only if the human being is both **measurement object** and **measurement device**, with all connected problems. The human being is the measurement object because his/her sound-perceptions are to be determined; and he/she is the measurement device because he/she needs to describe these perceptions. Since measurement object and measurement device cannot be separated, errors are possible. The statement "I do not hear any tone" can mean that the measurement object (the "subject") indeed does not hear anything and responds truthfully. However, it could also mean that the subject lies and does actually hear something. It could also indicate, though, that the subject thinks that what he/she hears is not a tone but e.g. a noise – in this case the response "not … any <u>tone</u>" would be truthful from his/her perspective. In order to avoid such misunderstandings, and to obtain the subject's assessment in the most unaffected and most reproducible manner, psychometrics has elaborated guidelines for the execution of experiments and their evaluation.

Reproducibility of the sound-presentation constitutes a particularly essential aspect. The reason that a guitar sounds – compared to the studio – different on stage is found more in the (physical) room acoustics, and not primarily in perception psychology, although the assessment *criteria* (measurement device!) can be situation-dependent, as well. In order to guarantee reproducibility in the presentation, many experimenters used specially equalized **headphones**. While this is an improvement over exposing the test person to a totally undefined sound field, it does not warrant an exact sound exposure, either. The position of the headphone (relative to the external ear), and the individual shape of the earlobe and the ear canal do influence the sound level.[♣] Another problem is the fact that an entirely unnatural sound field is created that turns with the head. Using precise instructions, mechanical fixation, probe microphones, and figurative presentations, these uncertainties can be reduced to the point that they are seen as "bearable" in daily research routine – this is then simply is as good as it gets. Sound presentation via one or two **loudspeakers** would be the alternative – not small PC-monitors, though, but calibrated premium studio monitors. Indispensable is again documentation: room acoustics, transfer functions, impulse responses, best supplemented by dummy-head recordings. The more is documented, the easier the decision after an experiment series whether an effect is due to the hearing system or due to experimental methodology.

[♣] Zollner M.: Interindividuelle und intraindividuelle Unterschiede bei Kopfhörerdarbietungen, Cortex 1994.

 Translation by Tilmann Zwicker

It may be that not a stored (or artificially generated) sound is to be assessed, but a **sound source**, i.e. an acoustic guitar or a guitar loudspeaker. In this case the question should be considered whether a recording via microphone or dummy head is made (and the recording then is listened to as mentioned above) or whether live-presentation is preferred (incl. documentation like the one involving loudspeaker presentation). A curtain hung in front of a picture will change the visual perception, and similarly the room between sound source and listener will influence the auditory perception. If the filtering by the room is ignored, the assessment is unusable. Only after the sound presentation is fully optimized and documented, the assessment of the sound may be started.

Auditory experiments may be of simple or complex, of fundamental or special character. **Threshold measurements** are easy to do for the subjects. Psychometrics distinguishes between (absolute) thresholds of stimulation, and just noticeable differences. The **threshold of stimulation** tackles merely the issue of whether something is heard. Not every tone with a sound power different from zero is audible – to be heard, the tone-level needs to be above the threshold-level. The threshold of stimulation that is determined for tones that are presented *in quiet* is called the **threshold in quiet**. If another (interfering) sound is present besides the sound to be assessed, the term used is **masked threshold**, e.g. “threshold masked by pink noise”. When determining **just noticeable differences**, the question is from which degree of signal change a subjective difference is noticeable. For example: which change in frequency is necessary so that a change in pitch is perceived? The subject’s task becomes more difficult if the question is not just whether a change is heard but also how big this change is. This **magnitude estimation** targeting the numerical assessment of perceived difference can lead to significant scatter up to the point that it is actually impossible for some experiments. We can “force” assessments, but it is hardly measurable whether something sounds better by a factor of two or three. Psychoacoustics states that it is measurable whether a sound has double the loudness of another sound. Yeah, kind of – but with a scatter of ±6 dB, gripe the critics. Scatter of measurement results is not at all limited to psychometric experiments – all measurements will include variance. It’s just that in psychometrics, the variances are particularly pronounced and therefore need to be looked into with particular scrutiny.

No subject will increase the level always by exactly 10 dB when asked to adjust to double the loudness. That is why the experimenter will average the intra-individually varying values, deriving a subject-specific mean value. *One* subject would represent an unsuitably small sample, and thus e.g. 24 further subjects need to do this adjustment-experiment, leading to 25 different **mean values** that show inter-individual differences. Again, an average is taken, and finally we get the result that will e.g. express that “on average” the subjects will increase the level by 10 dB to achieve double the loudness. That this mean is not valid for each and every human being – that is often pushed to the back of our minds. So let’s play devil’s advocate: literature reports scatter between 5 – 17 dB, and even 4 – 30 dB is found [Hellbrück 1993]. Even so: *here the center of the distribution was in the class of 8,6 – 9,8 dB*. Well then … that is almost 10 dB. To conclude from the variance that the whole shebang is one giant hokum – that would show some uncalled-for ignorance, after all. Insofar as experimenter and subject are aware of what they evaluate, averaging methods offer the only possibility to reduce clusters of dots to functions. Whether the assessments of fluctuation strength include a scatter of factor 4 or 8 – they still clearly feature a band-pass characteristic with a maximum at a modulation frequency of 4 Hz. We simply have to avoid the mistake to declare such results – with a three-digit precision – as universally valid; average values do have a limited accuracy, too.

Of course, experiment and averaging become questionable if experimenter and subject have different attributes in mind. A strongly exaggerated example would be the following: the experimenter distributes written instructions regarding the scaling of the sonority of drums. Questions are not allowed so as not to influence the subject. And off we go – judging away on a scale from 0 to 10. Not wanting – as a spoor student – to forgo those hourly €15.-, one tags along. Either according to the best of one's knowledge (or rather: perception), or according to the Monte-Carlo-method: everything's coming 'round again, and even this hour will pass. The PC generates some averages, and we have a result. The concept what "sonority" is supposed to be – that should be shared by experimenter and subject … otherwise it all really is one big hokum. And nobody say that a good result proves that this term "sonority" is self-explanatory.

A less construed example from the Süddeutsche Zeitung (an internationally read German newspaper) published on 24.09.2009: positioned within an MRI scanner, a subject is shown various photographs. Depending on the motif, the MRI scanner establishes different brain activities. Exceptional here: the subject is a fish. And even much more exceptional: the fish is dead. In spite of this, the evaluating computer manages to arrive at a significant mean result. In this case, the experimenter is not a charlatan but an honorable scientist seeking to show *how much nonsense is often practiced in modern experimental brain research*. N.B.: having many subjects at hand and using modern ("Russian") averaging algorithms won't guarantee solid data … or, in other words: garbage in – garbage out.

Modern psychology, and in particular psychometrics, increasingly employs statistical evaluation methods; that may be pesky, but it's unavoidable. The most wonderful experiment is no good if the results are erroneously evaluated. Just as nonsensical is to continue to (without experimental experience) process mindless data until a convenient result is obtained. Consider that, in a source-recognition experiment, all guitars are given the numeral 1, all trombones the number 2, and all basses the numeral 3. If the subject has now recognized four times the 1, twice the 2, and four times the 3, then we may not average arithmetically and state that as a mean value a trombone is recognized. These assessments or nominal judgments, after all, and there is no mean value. It would be similarly absurd to calculate a "mean postal area code". That would be possible, yes, but not interpretable.

A **nominal judgment** groups according to names and thus congregates elements of equal attributes into groups. Only with an **ordinal judgment**, a ranking is created – however without any metric. In metrology, class-0 is more precise than class-1, and the latter is again more precise than class-2. Class-0, however, does not necessarily feature double the precision of class-1, and if that were the case, class-1 could well be 3 times as precise as class-2. More mathematically: an ordinal scale is determined via inequations but not via intervals of equal size. The latter comes into play only with **interval scales**, they allow for additivity based on equidistance. What is not required is that the property of the element with the value "0" disappears. 0°C does not imply "no temperature" but rather is an arbitrarily fixed neutral point, and that is also why 20°C is not double as warm as 10°C. At the end of this list we have the relational scale in which the relations of the numbers mirror the relation of the degree of manifestation of the assessed characteristics. The sone-scale is such a relational scale: if two loudnesses have the relation 2:1, the same ratio is also found in the corresponding sone-numbers (8 sone is double the loudness of 4 sone). Conversely, the phon-scale is not a relational scale: 60 phon is not double the loudness of 30 phon.

The following table summarizes scales, properties and operations. Nominal scaling only offers *equal* or *unequal*, ordinal scaling adds in *larger than* and *smaller than*, additivity comes in with the interval scale, and product/division is only there from the relational scale.

The median (numerical value) of a nominally scaled set cannot be determined because for this all elements need to be brought into a ranking – which does not exist in nominal scaling. Only the modus, the maximum rate of occurrence, may be identified. “Most letters were transported for postal code 93057” makes sense, but “the median is postal code 93057” does not. As a rule, to use ratios of levels is pointless – although there may be exceptions here and there, insofar as “0 dB” indeed is mean to imply “nothing”. In terms of the SPL, level ratios are usually without meaning – using an equalizer, however, a boost of 8 dB may be double the boost of 4 dB.

Scale	**Nominal**	**Ordinal**	**Interval**	**Relational**
Synonyms	Topologic scale		Metric scale, cardinal scale	
Allowable statistical measures	Absolute and relative rate of occurrence, modus	Cumulative rate of occurrence, median, percentile	Arithm. mean value, variance, standard deviation	Geometric mean value
Operations	$= \;\neq$	$= \;\neq , < \;>$	$= \;\neq , < \;>, + \;-$	$= \;\neq , < >, + -, \;\times \div$
Features	Nominal feature, categorical or qualitative feature	Ordinal feature, ranking feature, comparative feature	Cardinal feature, quantitative or metric feature	

Table: Scales, features, allowable operations. In addition to the statistical measures in each column, all measures on the left of these are, correspondingly, also allowed.

Once we now have perfected the sound to be presented, and once the feature-scale to be found is determined, the subjects (test persons) may arrive. From now on it’s: no influencing, and reproducible instructions. With a statement given right at the start of the sort that EC’s “Brownie” is to be assessed, an opinion like “sounds a bit thin” is not likely to be voiced – that guitar will simply sound “killer”. In order to prevent such bias, the desired objective is the **blind test**, although that is not always doable. It would be possible to assess two guitar amps without prejudice if the amps are hidden behind an opaque curtain (a rotary table takes care of positioning problems); however, the immediate difference between a Gibson Les Paul and an ES-335 may only be hidden from the guitarist if rather elaborate precautions are taken. The differences between different scale lengths (e.g. 24" vs. 25,5") are always recognized – blind tests are impossible here. **Written** instructions for all subjects ensure that everyone is told the same, and they also facilitate checking the instructions a year later. If we realize in the course of an investigation that the subject have difficulties doing an assessment, we must not change the instructions until the ”correct” result turns up and average subsequently over all experiments. Out of the question is also something like averaging only over the last five subjects (because only they have heard the difference). Difficult question: should one single out unsuitable subjects? To assess drumsticks, you would not ask harpists to give a verdict; the sound of a guitar amplifier can, however, certainly be judged by a non-musician, as well. Because there are no set rules here, documentation is particularly essential (questionnaires handed out to all subjects). If we want to do a true service to science, we measure the hearing threshold in quiet (audiogram) of the subjects ahead of the start of the experiments. This is because many a musician (and other people spending any length of time in noisy environments) have generated (and have been subject to) so much sound energy in the course of their lives that their auditory system has experienced considerable damage. Corresponding judgments may therefore not be typical for those of normal hearing. Wouldn’t you concur with that, dear Mr. Townshend? Mr. Townshend, sir? Mr. Peter Townshend – HELLO there?? **MR. TOWNSHEND!!!**

Last, we have to consider according to which method the subject is going to deliver the judgment. That is, "last" in the framework of this short overview, because the rules of professional psychometrics♣ are more extensive and go beyond the presently set scope. **Methods of acquiring judgments** differ (among other aspects) according to the degree of involvement of the subject. Is the latter merely supposed to give a verbal assessment ("I don't hear anything"), or does he/she need to twist a knob such that a tone becomes just audible (or inaudible)? Is a scale of the assessment presented, or can the subject make one him/herself? Is the verdict "no difference" allowed, or is a preference forced (forced choice)? Is the response of the subject considered when new test sounds are selected? May the subject compare test sounds as long as he/she likes, or is a decision called for after two repetitions? For decades, psychologists have never grown tired of preaching that all these details in the experiments are vital to the results, and so we engineers cannot but believe it, and promote it. All the while hoping that – vice versa – the advantages of correct level-measurements find a similarly strong lobby in the psychologist-camp.

Scientific auditory experiments are more than just calling in three pals to in order to verify the hypothesis that the new Fender is another milestone in rock history. The last trap is found in the formulation of the results. The statement "the Makkashitta VR-6 has some mighty sustain" is o.k.; however, declaring "due to its maple neck, the Makkashitta VR-6 has some mighty sustain" is, most probably, rubbish. Unfortunately, it is everyday practice in test reports: the tester hears something (which is his god-given right), and connects without any prove what he has heard to some kind of material characteristic (which is stultification of the reader). Often, e*vid*ent associations (i.e. from visual domain) are dragged into the arena in order to substantiate "ear-sounding" connections (i.e. in the auditory domain). Does a silver trumpet ring more "silvery" than a "warm-sounding golden trumpet? Science says: no, it's all but imagination, or influencing the player. If the latter has to play under yellow lights and cannot distinguish the metals, he/she plays the same, and then the sound is the same, too – despite different metals (and given equal geometry). Does that big loudspeaker have less treble because its heavy membrane is set in motion more slowly? Mechanics say: no, you are mistaking cutoff-frequency with efficiency. Are the sound pressures arriving at the two ear canals indeed the only excitation quantities for the auditory sense? Well, with the answer "of course not", the examinee would have most likely failed the psychoacoustics exam in 1979. But since then, much has progressed; we do learn all the time. The visual impressions play an important role in the auditory perceptions, and thus the perceived loudness is dependent on the distance at which we see the sound source. It's also why the red express train is perceived to be louder than the green one, despite equal SPL [Fastl]; and it is the reason why we may hear "behind us" although the sound source is in front. It's a wide field, and – for the most part – still an only sketchily examined one.

♣ e.g.: Kompendium Hörversuche in Wissenschaft und industrieller Praxis, www.dega-akustik.de

8.7.2 The sound of the un-amplified guitar

How does the expert test an electric guitar? He first listens to it without amplification (i.e. “dry”). “It is certain that – contrary to common opinion – the desired sound of electric guitars and electric basses is not mainly dependent on the pickups. Rather, the wood forms the basis. If a customer travels to see me in the ‘Guitar Garage’ in Bremen and seeks to discuss pickups, I first listen to the instrument without amplifier.” (Jimmy Koerting, Fachblatt Musikmagazin). Or: “For a first evaluation of the sound quality we do not need amplifier towers nor distortion boxes – a small combo entirely suffices. It would of course be even better to test the sonic behavior in a quiet corner, dry, purely acoustically, regarding response, balance, and sustain.” (G&B 3/97). But why then are two guitars that sound differently dry, not able to feature these differences anymore when played through an amp? “Surprisingly, the differences in sound show up – compared to the dry-test – much less when connected to an amp”. G&B 7/06, comparison: Gibson New Century X-Plorer vs. V-Factor. Or, from a different comparison: "The Platinum Beast sounds dry powerful, warm and balanced, with velvety brilliance and tender harmonics, while the Evil Edge Mockingbird is sonically somehow feeble, poor in the mids, with somewhat more pronounced bass, but also clearly more brilliant and harmonically richer. Connected to an amp, and thanks to the hot humbuckers, everything is different though: hard to believe, but the two instruments now sound almost identical.” G&B 8/06.

Extreme examples will not serve to help in this case. Plywood (or even rubber!) is used as material for the (solid) guitar body in order to justify significance and necessity of high-grade body-woods. That is one extreme: using a totally unsuitable (absorbing) body, a good guitar cannot be built; ergo-1: the wood is more important than the pickups. The other extreme: a brilliant (“under-wound”) Strat-pickup is swapped for a muffled, treble-eating Tele-neck-pickup with a cover made of thick brass, and the result is the statement ergo-2: the pickup is more important than the wood. Both approaches are too lopsided.

From the point of view of system theory, the vibrating string is a generator that on the one hand excites guitar body and neck to vibrate and thus to radiate airborne sound. On the other hand, the relative movement between string and pickup induces a voltage. Airborne sound and voltage are therefore correlated because of the excitation from the same source. If the string-vibration dies down already after a few seconds, the pickup cannot generate a gigantic sustain. Or can it? Within certain limits, indeed, it could – in cooperation with a suitable amplifier (+ loudspeaker). The decay behavior is changed if the signal experiences limiting via the amplifier (overdrive, crunch, distortion). This is the decay behavior audible *via the loudspeaker*, because the decay of the string vibration is not changed. Or is it? Things begin to become unfathomable, and exactly for this reasons we find such contradictory opinions in guitar literature. If guitar and loudspeaker are positioned closely together, feedback may certainly influence the string vibration, as well. Maybe this is where the expert advice comes from: first listen to the guitar without amp. However: hardly any guitar player will buy an electric guitar to play it un-plugged forever. Sooner of later he will plug it in, and then the forecasts from the dry-test are supposed to prove to be true. The likelihood of a fortunate result of that experiment is indeed not entirely zero: electrical sound and acoustical sound are somehow related (correlated) – but in which way exactly is unclear to begin with.

Let's imagine a simple **experiment**: the pickups of a Stratocaster are screwed directly into the wood so that they have a clearly defined position. Will already that change the sound? Anyway, let us assume this special sound to be the reference. Guitar, pickups, and now on to the peculiarity: once with pickguard, and once without. That's a pickguard made purely from plastic so that no metal layer may generate any eddy-current damping. Now, do we hear a difference in sound if that guitar is played with pickguard compared to being played without? In the acoustic sound: definitely yes – in the electric sound: definitely no. Via the body, the pickguard – if present – is made to vibrate. It has weakly damped natural modes (eigenmodes) and is able to radiate audible sound in several frequency ranges. Do these pickguard-vibrations act back to the string? Theoretically: yes, because "all things are connected" (as already reportedly pointed out to the US Government by Chief Seattle as early as 1854/5). Practically no, because between string and pickguard we find the body that weighs in with many times over of the mass of the pickguard. The string vibrations are changed by the pickguard only in such an insignificant degree that the electrical sound does not change audibly. However, the radiated airborne sound does. Or another **example**: singers perform in a concert hall, and listener A listens in that hall while listener B listens from an adjoining room via the open door. Now, the door is closed – what changes? For listener B, a lot – but for listener A, almost nothing. Very theoretically, we can again call in Chief Seattle and demand a correction factor for the wall absorption that the closed door has modified, but in practice not all of such lemmas have been rewarding, as the in the chief's case rather unfortunate history has shown.

What is the connection between the singer and the above electric guitar? In both cases there are two different transmission paths that modify the sound they carry in different ways. Knowing about one transmission path does not allow – in the general case – for any conclusions on the other transmission path. The listener in the concert hall cannot even be certain that the other listener (The Man Outside …) hears anything. This implies for guitars: what use is the great acoustical sound if the pickup winding is broken. Caution, though: we are again entering territory of extreme positions. Thus, not assuming a complete sound insulation for listener B, the latter will be able to make some statements: when singing is going on, when it is paused. Maybe, listener B can even recognize which one of the three sound sources is trying to get to that high C: the little one, the pretty one, or *Fat Lucy* (also called the stage-panzer). Any problems with intonation are perceived through the closed door, as well, as long as the latter in not totally soundproof – and if such problems are present within the expectations of the listener in the first place.

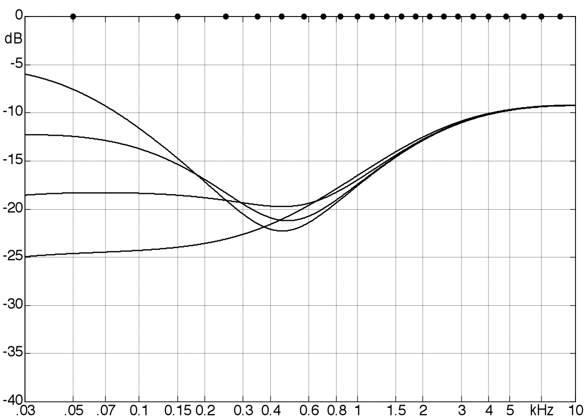
The thing with the expectations can be observed with guitars, also: it is astonishing how some guitar tester become victims of their own convictions. Irrefutable **credo**: "Of course, the original Les-Paul-mix of rosewood fretboard and mahogany body fitted with a thick maple cap – that gives us the unique Les-Paul-sound". That's just how it needs to be written – in this case in a comparison test (G&B 7/02). And then a copy with an alder body (stigmatized with a "!" in the test) dares to sound good. It even commands the tester's respect. "*... it can, in any case – be it alder or mahogany – convince with a first-class clean sound...*" Well, well – let's not exaggerate here! Don't forget, its alder!! And lo and behold: "*... overall somewhat subdued and a bit shy.*" There we are: typical alder. However, oh great Polfuss, what happens only a column further, with the Fame LP-IV also included in the test? *"Those who go for a typical forceful Les-Paul-sound without frills should check out the Fame LP-IV. Indeed, it sounds the most authentic. In all areas, its sound is very similar to that of the original*". **Question**: according to the test, the Fame LP-IV sports a maple neck, an oak fretboard, an alder body, and a mahogany cap – did I get anything wrong here?

Let us postpone the discussion on materials to later, though, and return to the question of how far the conclusion from the “dry” test to the electrical sound is admissible. Apparently there are “**robust**” signal parameters that win through on every transmission path, and “**fragile**” parameters that change on their way through the transmission medium. The pitch is quite robust: whether a guitar is in tune is audible both “dry” and amplified. Not to the last cent, as psychoacousticians know, but with a precision adequate for some first considerations. The sonic balance between treble and bass, however, depends on the tone settings of the connected amp – that much is as uncontested as it is trivial. The “dry” sound can make every effort: it can never hold its own against a fully turned-up bass control. “Anyway, that’s not what we mean”, the expert will object, “in the dry-test I can hear the fundamentals of the sound, and the soul of the wood.” Please, dear scientists and dear psychologists – no malice now … it’s o.k. to state something like that here, as a guitar tester who does neither have to understand much about physics nor of psychology. However, the **soul of the wood** does reveal itself to the seeker not a prima vista; it does require many séances in which the spirit penetrates the matter; much knocking on wood needs to happen, and a tuning fork must to be pressed against the solid body of a Stratocaster (in the Fender ads, anyway), and many years of ear training are necessary. At least for this last point we should be able to reach a consensus, shouldn’t we? This is not supposed to be about the guitar-o-phobe agnostic with progressive dysacusis, but about the more or less pronounced aficionado of the instrument. Those who – with their more or less extensive listening experience – indeed hear details in the sound not accessible to the layperson.

Problem: how do you describe such sound-details? This is the classic conceptual formulation and task of **psychophysics** and psychometrics that frequently leads to similarly classical misunderstandings. A verbal description (dead, woolly sound) is rejected at the physical docking-port as much too ambiguous and imprecise, just like the exact physical description (8,43% degree of amplitude-modulation at 944 Hz with f_{mod} = 6,33 Hz) is objected to by the artistic/mystical faction as pipe-dream-y and too abstract. Logically, any proposals of compromise trying to bridge the two realms are dismissed by both sides. Well then: rather than the wood’s soul, often a dead or a lively sound is mentioned. What distinguishes live from dead matter? The matter that is alive – it moves! And already we have the first objections, because that would define the pen dropping from the table as alive? O.k., so we turn to a fundamental philosophical contemplation of life in particular, and of the universe and everything in general … NOT! No, really not. **What is alive does move**. Period. Conferred to the guitar sound: an artificial tone with its strictly harmonic partials all decaying with the same time constant, sounds dead. However, if the partials decay with different speeds and with different beats, the impression is one of movement and life. In this, the term “movement” may indeed be seen in its original meaning as change in location: when a sound source changes its position in a (sound-reflecting) room, time-variant comb-filters vary the signal spectrum – the movement in space has the effect of a “movement” in the sound. Way back in prehistoric times it was presumably in support of survival if moving sound sources were given a higher priority than static sources; at the same time early researchers in communication discovered that speech sounds will only convey information if they include variations. Without pushing too far into foreign territory: there would be enough reasons why the human auditory system continuously hunts for spectral *changes*. Even though electric guitars are younger that roaring tigers and vandals screaming “arrrghh!”, our hearing has its capability to analyze, and it takes advantage of it. A lively tone rich in beats sounds more interesting than a dead sound – at least as long as instrument-typical parameters are maintained.

Similarly to the pitch of the string, the beats between partials can be rather **robust** relative to the transmission parameters, and therefore it is imaginable that the expert may be able to deduce criteria for the electrical sound from the "dry" test. On what does the robustness of the signal parameters depend? Frequency-dependent signal parameters – such as the spectrum – loose their individuality if the corresponding frequency-dependent system parameter (the transfer function) has a similar shape. Three examples follow:

1) Psychoacoustics [12] describes the balance between high and low spectral components as "**sharpness**": treble-emphasizing sounds have a high sharpness; turning down the treble control reduces the sharpness. Spectral details are not as essential for the calculation of sharpness as the basic (smoothed) run of the spectral envelope. To be more precise: the sharpness is taken from the weighted loudness/critical-band-rate diagram which has a mere 20 sampling points in the frequency range important for electric guitars. (Transmission-) frequency-responses of guitar amplifiers can be represented with the same increments (**Fig. 8.45**), and from the kinship of the two data-sets we can conclude that the sharpness of the "dry" guitar sound in general does not correspond to the sharpness of the amplified sound. In other words: changing the tone controls on the amplifier allows for changing the sharpness – from this point of view, sharpness is not a robust signal parameter.



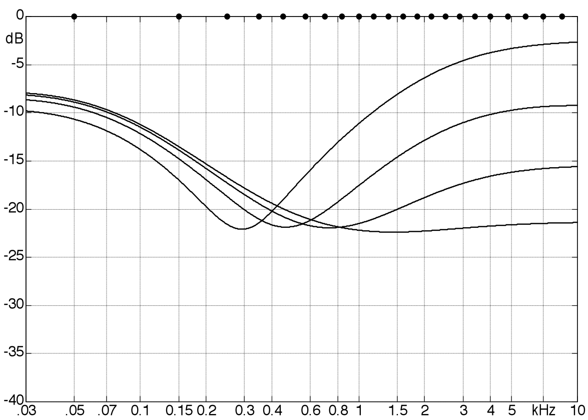


Fig. 8.45: Tone control of a Fender amplifier (transmission factor). The dots on the top mark the critical-band grid (discretization of the abscissa in order to calculate sharpness).

2) Beats between partials can in the time domain be described as amplitude fluctuations, and in the frequency domain as the sum of closely neighboring partials. For example, two partials of equal level but slightly differing frequencies (e.g. 997 Hz, and 1003 Hz) lead to the auditory perception of one single 1000-Hz-tone fluctuating in loudness with 6 Hz [3]. In order to change this beating, a highly frequency selective operation is necessary. Such an operation is untypical for tone controls in amplifiers. From this point of view beats in partials are robust relative to simple tone-control networks.

3) The spectrum of a quickly **decaying** sine-tone (**Fig. 8.46**) is largely limited to a narrow frequency range. Any changes in the decay behavior need to be done using highly frequency-selective methods, too. In other words, a linear, guitar-amp-typical tone control network will practically not change the decay-behavior of individual partials – the decay behavior is robust in this respect.

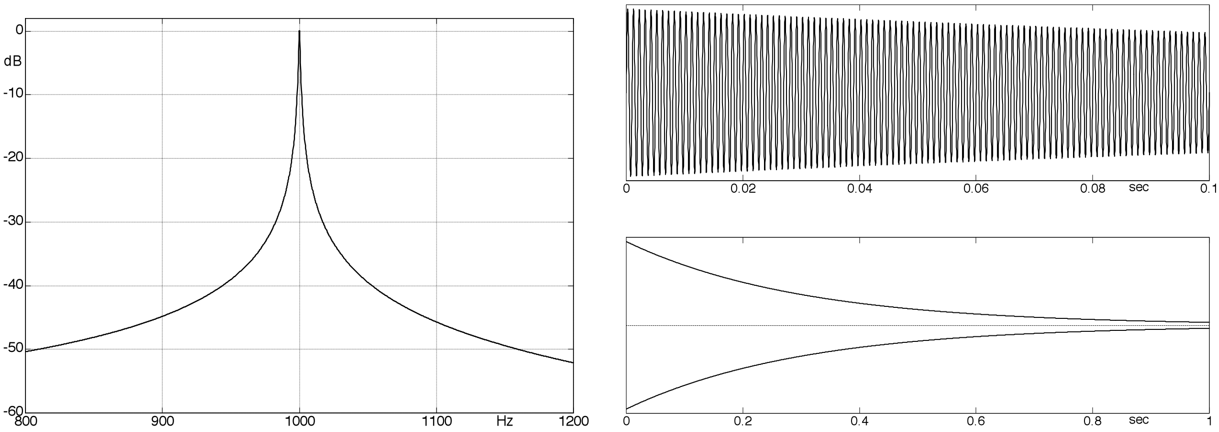


Fig. 8.46: Decaying sine-oscillation, $f = 1000$ Hz, time-constant $\tau = 0{,}3$ s.

These simplified representations do, however, require supplements in some points. It is not only the transfer factor of the guitar amplifier that changes the spectrum of the string oscillation. The loudspeaker (including its enclosure) acts as filter, as well; in the detail, its transmission curve is of stronger frequency dependence than the tone-control network is. Still, a loudspeaker membrane does not reach the high (resonance) Q-factors of decaying guitar partials – it would have to generate clearly audible natural tones for that, and this it does exactly NOT do. The last filter in the transmission path is the room with its reflective borders. Its effect cannot be neglected even in the "dry" test; when playing connected to the amp/speaker, the distance to the speaker needs to be added in as variable, as well. As long as one remains within the near-field of the loudspeaker, the effect of the room can be regarded as equal for both playing situation in a first-order approximation.

Special consideration is required for effects that achieve more than what simple tone control does. Adding artificial reverb can extend decay processes and feign life that is not included in the original in that form. Chorus/phaser/flanger are time-variant filters with high (resonance) Q-factors – their use always aims at changing the fine structure of the partials. Single band, and in particular multi-band, compressors change the decay time constant of individual groups of partials. Overdrive has similar effects but adds in additional partials. It is therefore very well possible to also influence the signal parameters designated as robust above. However, even without deploying radical effects one may – within certain limits – extrapolate from the sound of the unamplified guitar to the sound of the amplified guitar. Which of the many beat- and decay-parameters are crucial to the 'good' sound, though, is at the most implicitly appraisable … and we have not even touched the wide field of frequency- and time-related masking [12]. Therefore, only this principle can hold: **the unamplified sound of an electric guitar should basically not be evaluated**. Only for the expert, and in consideration of his/her special knowledge and listening-experience accumulated over decades the exception, this rule allows the exception that the "dry" test reveals "everything", after all, in the individual case. Experts who may claim this exception for themselves are: testers of all guitar magazines, all guitar sales-personnel, all guitar players who have had of who have wanted to have a guitar for more than a year, and all listeners to CD's who still have the sound of Jeff Beck's signature guitar ringing in their ears (see Chapter 7). And please, dear experts who have now received so much legitimization for your obviously indispensable "dry" tests: that the assessment of tactile vibrations is nonsensical – that should by now be o.k. for a consensus, should it not?

Concluding the topic of guitar tests, a few citations in the following:

Yamaha Pacifica-guitars (maple neck, alder body) in a comparison test: "The acoustically quite comparable basic characteristics of the Pacificas differentiate themselves rather clearly according to their pickups, after all. (G&B 6/04)."

Gibson Les Paul Faded Double-Cutaway: "Already with the very first striking of the string it becomes clear that the economy-varnishing curbs the resonance properties of the woods to a lesser degree. The guitar vibrates from the feet (strap-knob) to the tips of the hair (machine heads) so intensely that you can feel this even in your own body; (G&B 6/04)."

Ibanez IC400BK: "The slight underexposure of the E_6-string as it appears in the dry test has suddenly disappeared with the support of the pickups; (G&B 6/04)."

Squier-Stratocaster, comparison: **mahogany**-body vs. **basswood**-body: using the neck- or the middle-pickup, both guitars sound almost identical (G&B 5/06).

"Picking up the **Pensa-Suhr**-guitar and playing it un-amplified, the reasonably learned ear immediately hears where this is at. … both standing up and sitting down, you feel already in your belly the fantastic vibration-behavior of the outstandingly matched woods." (Fachblatt, 6/88).

"Despite using humbuckers, a Strat will never turn into a Les Paul"; G&B 2/00. **Ozzy Osbourne** on Joe Holmes: "In fact, I normally don't like Fender guitars. But Joe gets this fulminant Gibson sound out of them"; (G&B 2/02). "**Jimmy Page** recoded the whole of the first Led-zeppelin album using a **Telecaster**; the guitar sound on this album is exactly that of a Les Paul"; (G&B Fender-special-issue). **Mark Knopfler**: "If I look for a thicker sound, I use my Les Paul; it simply is more dynamic. That doesn't mean that I couldn't do the same thing with a Stratocaster"; (G&B Fender-special-issue). **Gary Moore:** "Some people think that a Fender Stratocaster is heard on 'Ain't nobody'; actually, that is my own Gibson Signature Les Paul"; G&B 7/06 p.91.

Big mass of wood (**3,9 kg**): Due to the big mass of wood, the response seems a bit ponderous and the notes don't get off the starting blocks that fast; (G&B 7/06).
Even heavier (**4,15 kg**): the guitar vibrates intensely, responds directly and dynamically, each chord or note unfolds crisply and lively; (G&B 8/06).
Despite the enormous mass of wood (**3,85 kg**), almost every note responds crisply and dynamically, and unfolds very swiftly; (G&B 7/06).
The lower **mass** can be more easily made to vibrate; (Thomas Kortmann, gitarrist.net).
A slender guitar **body** also creates a slender sound; (G&B 7/02).
Thinner **body** = less bass; (G&B 4/04).

Thick neck = sonic advantages; (G&B 8/02). **Thin neck** = round, fat tone; (G&B 10/05). **Thin neck**: the lower the mass that needs to be moved, the more direct and quickly response and tone-unfolding get off the starting blocks; (G&B 3/05). **Crisp** and direct in the response, every note gets quickly and lively away from the starting blocks, **despite the immense mass of the neck** (that needs to be first set into motion, after all); (G&B 9/05). **A thin neck** has no acceptable vibration behavior whatsoever; (G&B 3/97). Sonically advantageous is that the **neck** weighs in with **a lot of mass**; (G&B Fender-special-issue). The **Ibanez JEM 777** sports an extremely thin neck design: the basic tonal character is powerful and earthy; (Fachblatt, 6/88). Of course, the **shape of the neck** contributes to the tonal character of the guitar, as well; (G&B, 12/06). What is absolutely not true is that **thick necks** will sound better than thin ones. I have already built the same guitar with the thick and a thin neck and could not find any difference; (Luthier Thomas Kortmann, gitarrist.net)

8.7.3 Tactile vibration perception

There is scarcely any guitar test-report that does not praise the exorbitant vibration-happiness of the electric guitar under scrutiny: "The design shows considerable resonance properties, after each picking of a string it vibrates intensively and clearly noticeable.” G&B 9/06. Or: “ From a vibration-engineering point of view, the MTM1 ranks at the highest level because the whole structure resonates intensively into the last wood fiber after each picking of a string; this results in a slowly and continuously decaying sustain.” G&B 8/06. Or: “Combined with the given open freedom of vibration (sic), we achieve a beaming sound color.” G&B 8/06. Or: “Less mass can more easily be made to vibrate.” Luthier Thomas Kortmann, Gitarrist.net. Or: “At Fender they even proceeded to build bodies from several wood-parts … Of course, the ability of the wood to resonate is restricted by such a number of differently sized pieces.” And loc. cit.: “That Ash moreover has almost optimum resonance properties was thankfully acknowledged at the time. It does not bear contemplating that Leo Fender might have opted for mahogany back in the day.” Day et al. Or: “Clearly noticeable right into the outermost wood fibers, both Strat and Tele show very good resonance properties.” G&B 4/06.

Mind you: we are discussing electric solid-body guitars here, and not acoustic guitars. The clearly noticeable **vibrating of the guitar** is taken as a criterion for quality. Why don’t we let one of the fathers of the solid guitar, Lester William Polfuss, speak: *"I figured out that when you've got the top vibrating and a string vibrating, you've got a conflict. One of them has got to stop and it can't be the string, because that's making the sound."* Mr. Polfuss sought to let only the string vibrate, and not the guitar top. O.k., one could object that the man was a musician, not an engineer. Still, he was a musician that replied to the question of who had designed the Gibson Les Paul with *"I designed it all by myself"*. The string is intended to vibrate, and the body should just shut up and be quiet. Only the very nit-picking ones will throw in at this point that only the relative movement counts, i.e. if the strings remain at rest, and instead the body would … no, enough with the theories of relativity, it does work better the other way ‘round. However: what does that mean – better? What characterizes a better sounding guitar? In his dissertation [16], Ulrich May cites D. Brosnac with the insight that a guitar made of **rubber** would absorb the vibration energy of the string within a short time and therefore would not sound right. That is understandable but does not prove that ash (or maple, etc.) is better suited. Evidently, there are unsuitable body-materials that will withdraw an unbecomingly big amount of vibration energy from the strings. Rubber is among these materials – but who would want to build a guitar out of rubber? Presumably, damp towels♣ also rank among the unsuitable materials. Or, fresh from the sleep-lab: since a bed of a length of 1,45 m (about 5 feet) is uncomfortable for most grown-ups, a 2,12-m-long bed has to be more comfortable than a bed of 2,05 m length. Or, more guitar-specifically: what the luthiers have learned for the acoustic guitar cannot be wrong for the electric guitar. A guitar has to resonate. Right into the outermost wood-fibers. Intensively and clearly noticeable.

So, what can we feel – as human being in general, and as a guitar-tester in particular? That depends, of course, on the stimulus and on the receptor. However, in terms of vibrations, the subcutaneous Pacini-corpuscles react most sensitively to stimulus frequencies of 200 – 300 Hz; they sense vibration amplitudes as low as 0,1 µm. That also implies that the sense of vibration becomes increasingly less sensitive above frequencies of about 250 Hz. Sound-shaping harmonics remain largely hidden from the tactile sense.

♣ because of the high „damp“-ing ...

Fig. 8.47 shows the frequency dependence of the **vibration threshold**, i.e. the vibration amplitude that needs to be reached in order to generate any vibration perception in the first place. Besides the dependency on frequency and amplitude, the exact shape of the curve depends also on the area of the vibrating surface, and on the location that is stimulated. The given graph can be seen as typical for the thenar. Thus, if a guitarist feels a vibration in the neck or the body of the guitar upon striking the strings, it will be a case of low-frequency vibrations. To **check via a calculation**: if we take 10 N as force at the bridge, a mass of 4 kg, and 250 Hz as stimulation frequency, we get a displacement of 1 µm. It is therefore no wonder that vibrations result that are felt, even without any resonance-amplification.

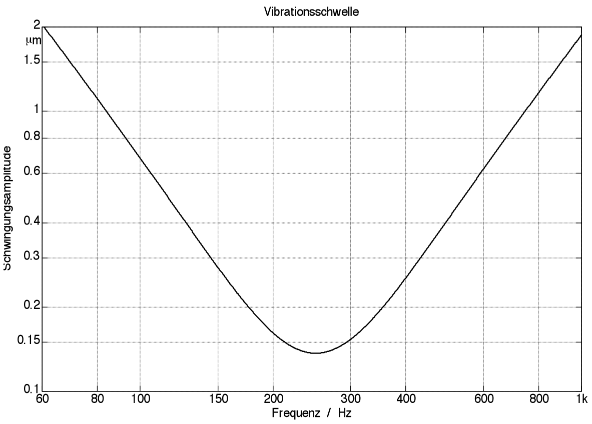


Fig. 8.47: Vibration threshold. Only values above the threshold lead to a perception of vibration. According to this graph, a vibration of an amplitude of 0,4 µm can be felt at 300 Hz; at 800 Hz it would not be felt anymore. "Schwingungsamplitude" = vibration amplitude; "Vibrationsschwelle" = vibration threshold

Therefore it is less a question of whether noticeable vibrations can emerge, but more how these should be assessed. If we take up again Les Paul's idea, any body-vibration to speak of would be counterproductive. With a lot of mass (ten-pounder Paula), this ideal can be approached at the cost of wearing comfort, and disregarding natural modes (eigenmodes) that amplify the vibration. The guitar neck in particular must not be too heavy; it will resonate to a noticeable degree in any guitar. What would in fact happen if guitar body and guitar neck could be manufactured to be vibration-free? On every guitar of this kind, comparable strings would vibrate in an identical manner given comparable picking! **Individuality is imperfection**, and it would fall by the wayside. In the acoustic guitar, the luthier seeks to form the transfer function frequency-dependently, and thus let some frequency ranges be radiated better, but conversely let other frequency ranges be radiated worse. An individual sound does result that way. The same principle could be applied for electric guitars, as well, and neck and body could be made to vibrate more at certain frequencies, i.e. the vibration energy would be more strongly dissipated. Whether this is indeed desired can only be judged in an overall consideration of all sound-shaping elements. Still, it would be a remarkable coincidence if exactly those frequency ranges for which the tactile sense is particularly sensitive would require the strongest damping. For one thing is certain beyond all doubt: the vibration energy that is felt, it is sourced from the string. The more intensive "the whole structure resonates", the less the string vibrates. One may agree or disagree with Les Paul's ideas – the law of conservation of energy should rather not be objected to.

Whether we would like to contradict Day et al., however, is again left up to us: "the vibrato-system itself was given a knife-edge-type shape at the six holes foreseen for the screws retaining it. The whole system was therefore mounted optimally in a very low-friction manner but still could transfer the vibrations of the strings optimally to the body." Indeed, this path is known: "**because the tawdry goes down to the corpus unsung**" … Schiller, Nänie. Or something like that.

The almost-empty page

Hi there old-timers,

Hey by ‘n’ large I can unnerstand and confir yer notions.

Old gittars just sound diff’rent than new ones, and it’s in the ear and the fingers of the beholder to decide if a gittar has some upwards potential or not. Course, the bottem supstanse has to be right butthat is then a matter of espirience so you can tax that. An axe who wont resone at all when its played won’t be impressd by that after 1000ds of plain hours, eithe.

So I pay atentsion to vibro-and reso-behavor in the newones.

Fact is too that plaing lots impoves yer own s kill and that change the sound agin (hope to the better).

Have read the article by U.P. He cam to the conlusion that such attack-apparatuss really change the gittar (cause stuff is done to it fysicaly like things happen in that cryo-tuning – you can see on youtube what Joe does at G-Cener):

What the two metods make clear, it is not really possible to esimate how the sound’ll differ. To show that they would have to record the gittars before and after the treatement with a fixed recording setup to make adifference somewhat possible to hear objectively.

If in the studio someone has tried to record one and same song on two days, he knows that really there can be differences in the sound if you jus drop yer axe after a cool session and leave everthing like that. The next morningthat super-sound sounds suddenly not as super even though you have changed nothin. Hereit’s again the subjetive hearin. As such even such a differece can not be felly objectively spread. We will have to try out such stuff ourself sometime if we really want to do that.

I kinda more think it better to play ones instrument so that one gets better instead all the time to run around in search for perfect soiund. Course too a new instrument, ol or new, needs to be won over by much plaing!

Play It, hear It and Sound like yourself !

Jack J.

<u>“Play-in” the guitar and find the desired sound</u> (loosely translated from the G&B-Forum of 11.01.2014)