### 8.2.4 Grouping of partials

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

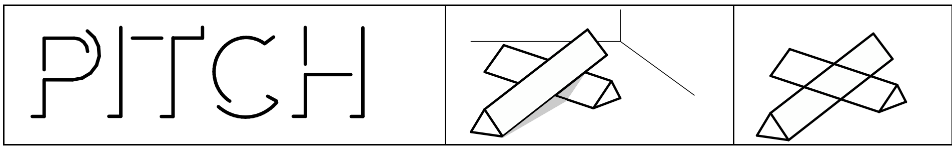


Fig. 8.14: Examples for visual grouping.

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

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

 

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

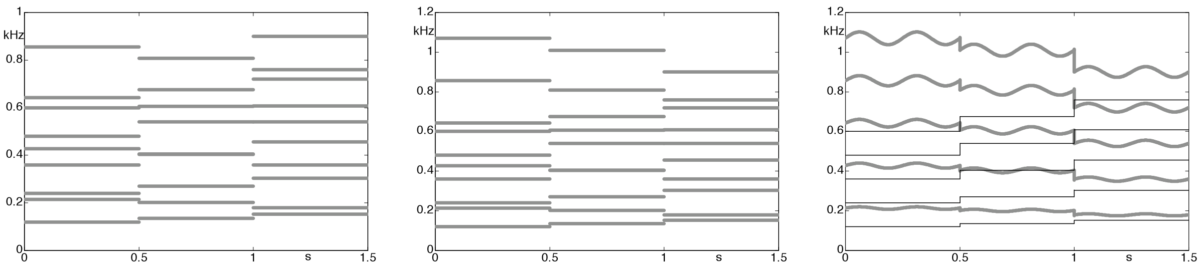

**Fig. 8.15:** Spectrograms of two tone-sequences: on the right, the descending sequence is frequency-modulated.

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

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

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

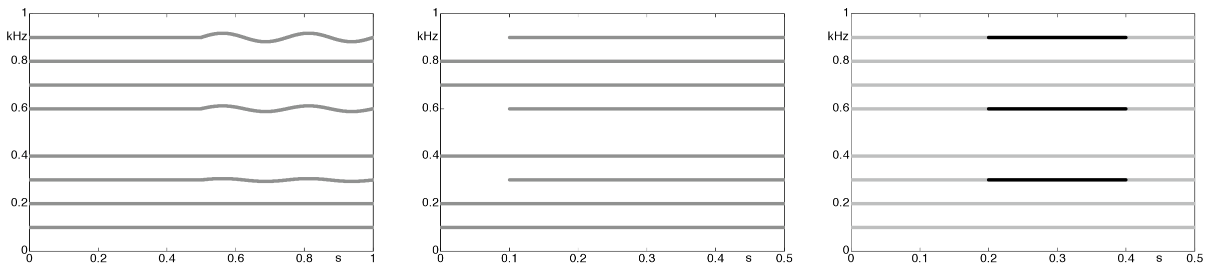


**Fig. 8.16:** Partial with/of a common fate are grouped to objects.

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

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

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