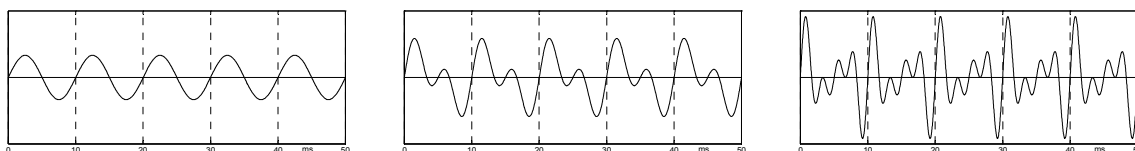


## 8.2 Frequenz und Tonhöhe

Die **Frequenz** ist eine Größe aus der Welt der Physik, während die **Tonhöhe** als sensorische Wahrnehmungsgröße zum Hörereignis gehört. Üblicherweise misst man die Frequenz in Hz, und gibt damit die Schwingungsanzahl pro Sekunde an. Die Einheit Hertz, abgekürzt Hz, ist nach dem Physiker Heinrich Hertz benannt. Kehrwert der Frequenz ist die Periodendauer, kurz **Periode** genannt. Zu einem Ton der Frequenz  $f = 500$  Hz gehört die Periode  $T = 2$  ms. Die **Tonhöhe** kann entweder im Selbstversuch bestimmt werden (Introspektion), oder indirekt im Fremdversuch durch Auswerten der Reaktion einer Versuchsperson. Obwohl die Tonhöhe eine subjektiv bewertete Größe ist, kann sie numerisch gemessen werden. **Messen** heißt, einer Objektmenge nach festgelegten Regeln Zahlen zuzuordnen, die innerhalb sinnvoller Fehlergrenzen reproduzierbar sein sollten. Was man als sinnvoll ansieht, ist allerdings wieder eine sehr subjektive Entscheidung\*. Die meisten psychometrischen Versuche ergeben intra- und interindividuelle **Streuungen**: Ein und dieselbe Versuchsperson kann beim mehrmaligen Durchführen eines Versuchs unterschiedliche Bewertungen abgeben (intraindividuelle Streuung), die Urteile verschiedener Versuchspersonen können sich bei je einmaliger Darbietung aber auch unterscheiden (interindividuelle Streuung).

### 8.2.1 Frequenzmessung

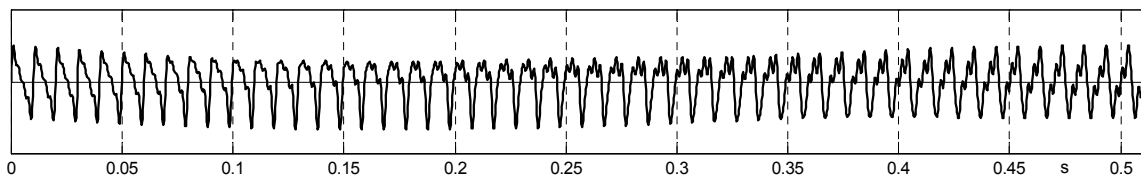
Einfache Frequenzmessgeräte zählen die Anzahl der pro Zeitintervall erfolgten Schwingungen. 5 Schwingungen pro 0,1 s ergibt 50 Hz. Unter Schwingung wird immer die komplette Periode verstanden, bei der Saite also von der Ruhelage ausgehend: Schwingen in eine Richtung, Umkehren am Scheitelpunkt, Durchschwingen durch die Ruhelage in Gegenrichtung, Umkehren am anderen Scheitelpunkt, Rückschwingen zur Ruhelage. Bei einer idealen Schwingung sind Begriffe wie Frequenz oder Periode somit leicht definierbar, aber reale Schwingungen sind nicht ideal. Die Signaltheorie definiert einen **periodischen** Vorgang als *in identischer Form unendlich oft wiederholt*. Ein Sinuston ist somit periodisch, und hat eine einzige Frequenz. Auch ein aus einem 100-Hz-Ton und einem 200-Hz-Ton zusammengesetzter Klang (den man in der Musik auch Ton nennen würde), ist periodisch (**Abb. 8.10**). Da hierbei aber mehr als eine Frequenz auftauchen, nämlich 100 Hz und 200 Hz, ist in **Teiltonfrequenz** (auch Partialtonfrequenz genannt) und **Grundfrequenz** zu unterscheiden. Die Grundfrequenz ist aber nicht zwangsläufig die des tiefsten Teiltons, sondern der Kehrwert der Periode. Das Schwingungsbild eines aus 200 Hz, 300 Hz und 400 Hz bestehenden Klangs wiederholt sich nach 10 ms; deshalb ist seine Grundfrequenz 100 Hz, auch wenn in diesem Klang gar kein Teilton mit 100 Hz vorkommt. Allgemein gesagt ist die Grundfrequenz der größte gemeinsame Teiler aller Teiltonfrequenzen, und die Periode ist das kleinste gemeinsame Vielfache aller Teiltonperioden.



**Abb. 8.10:** Sinuston (100Hz), Zweitonklang (100|200Hz), Dreitonklang (200Hz|300Hz|400Hz); je 0–50ms.

\* Der Autofahrer, dessen Fahrzeug gerade einen hochfrequenten Radarstrahl reflektiert hat, wird möglicherweise mehr Toleranz fordern als die von Bußgeldern profitierende Gemeindeverwaltung.

Offensichtlich muss ein Ton nicht monofrequent sein, um *eine* Frequenz (exakt: Eine Grundfrequenz) zu haben. Theoretisch können sogar – wie beim idealen Rechteckton – unendlich viele Teiltöne vorhanden sein. Die Teiltöne müssen aber **harmonisch** sein, d.h. ihre Frequenzen müssen ganzzahlige Vielfache der Grundfrequenz sein. Das ist z.B. bei idealen Irrationalzahlen wie  $\sqrt{2}$  und  $\sqrt{3}$  nicht möglich. In der Praxis sind Frequenzangaben aber immer nur mit endlicher Stellengenauigkeit möglich, z.B. 1,414 Hz und 1,732 Hz. Falls es sich hierbei nun um gerundete Wurzeln handelt, wäre die Angabe: "Die Grundfrequenz ist 0,001 Hz" sehr willkürlich. Und übrigens auch nicht im Sinne des ggT: Zumindest 0,002 Hz ist ein gemeinsamer Teiler. Das Problem mit den Irrationalzahlen ist weniger akademisch, als man zunächst glauben möchte: Saitenschwingungen sind nämlich niemals exakt periodisch. Ihr Abklingen gibt jeder "Periode" eine andere Amplitude, und ihre Teiltöne sind wegen der Biegesteifigkeit und der damit auftretenden dispersiven Wellenausbreitung **unharmonisch**. Nehmen wir an, das Abklingen erfolgt so langsam, dass seine Auswirkungen auf die spektrale Reinheit vernachlässigt werden können. Nehmen wir weiter an, die Analyse eines Gitarrentons hätte vier Teiltöne bei den Frequenzen 100 Hz, 201 Hz, 302 Hz und 404 Hz ergeben. Was ist die Frequenz dieses Tones? Es macht wenig Sinn, 1 Hz als Grundfrequenz zu spezifizieren, und die Teiltöne zur 100., 201., 302. und 404. Harmonischen zu ernennen. Es bleibt nur die ernüchternde Erkenntnis: **Ein Gitarrenton hat keine Grundfrequenz**. Aber er hat eine Tonhöhe! Die Ermittlung dieser Tonhöhe bleibt einem späteren Abschnitt vorbehalten, zunächst muss noch geklärt werden, was dann ein **Stimmgerät** macht, und warum eine Saite trotzdem gestimmt werden kann.



**Abb. 8.11:** Viertertonklang,  $f_1 = 100\text{Hz}$ ,  $f_2 = 201\text{Hz}$ ,  $f_3 = 302\text{Hz}$ ,  $f_4 = 404\text{Hz}$ .  $1/f$ -Hüllkurve;  $t = 0 - 0,5$  s.

**Abb. 8.11** zeigt die ersten 0,5 Sekunden eines aus den o.a. Frequenzen zusammengesetzten Viertertonklangs. Wieviele Perioden treten während dieser Zeit auf? Wenn man versucht, die Maxima zu zählen, kommt man in Bildmitte etwas in Bedrängnis, kann sich aber mit der Erkenntnis an den rechten Bildrand retten, dass es ca.  $49\frac{3}{4}$  Perioden sind. Aber was heißt hier "Periode"? Auge bzw. Gehirn versuchen halt, so gut sie können, visuelle Glättungen (Filterungen!) und lokal begrenzte Autokorrelationen vorzunehmen, und eine *visuelle* Bewertung abzuliefern. Was anders sollte ein visuelles System auch machen. Ob das aber für akustische Signale hilfreich ist? Wie könnte ein exakter Algorithmus aussehen? Einfache Messgeräte bestimmen den steigenden (oder fallenden) Nulldurchgang. Das führt zwischen 0,15 – 0,2 s und zwischen 0,35 – 0,45 s zu erheblichen Problemen. **Glättung**, d.h. Tiefpassfilterung, ist natürlich ein Weg – aber dann wird die Frequenz des *gefilterten* Signals bestimmt. Im Extremfall lässt der Glättungstiefpass nur die 100-Hz-Schwingung durch; dann ist Frequenzmessung natürlich einfach. Die meisten elektronischen Stimmgeräte haben vermutlich ein Tiefpassfilter eingebaut, das saitenpezifisch oder zumindest instrumentspezifisch filtert. Außerdem akzeptieren sie kleinere Abweichungen vom Sollwert. Trotzdem kann es passieren, dass die Anzeige zwischen korrekt und falsch hin- und herpendelt. Der versierte Gitarrist dreht dann den Klangregler zu (Tiefpassfilter), oder reduziert seine Genauigkeitsansprüche. Manche zelebrieren auch den Akt: Der Gitarrist stimmt immer, die Gitarre stimmt nie.

Die Abb. 8.11 zugrunde gelegten Frequenzen zeigen die prinzipielle Problematik, übertreiben aber den Sachverhalt. Die bei E-Gitarren typische Teiltonspreizung beträgt für die E<sub>2</sub>-Saite bei 500 Hz ca. 0,2%, bei 1 kHz ca. 1%. Aber immerhin: Ist der 12. Teilton der tiefen E-Saite mit wesentlichem Pegel im Gesamtsignal vertreten, so entsteht eine u.U. störende Diskrepanz von ungefähr 9 Hz zwischen der idealen Lage ( $12 \cdot 82,4 \text{ Hz} = 988,8 \text{ Hz}$ ) und der realen Lage (997,7 Hz). Für eine präzise Stimmung wäre ein derartiger Fehler inakzeptabel. Die Amplituden der höheren Teiltöne klingen aber viel schneller ab als die der tiefen Teiltöne, und deshalb bringen die meisten Stimmgeräte eine akzeptable Anzeige zuwege – zumal der Gitarrist die zu stimmende Saite weich anschlägt, um die Obertöne nicht zu sehr zu betonen. Bei den tiefen Teiltönen der tiefen E-Saite ist die Inharmonizität dann ziemlich unproblematisch: 0,02% beim zweiten und 0,05% beim dritten Teilton. Bei den höheren Gitarrensaiten ergeben sich bezüglich der Inharmonizität noch weniger Probleme: Wegen geringerer Saitendicke spielt die Biegesteifigkeit eine geringere Rolle, und die Anzahl der störenden Teiltöne sinkt wegen der Tiefpassbegrenzung des Tonabnehmers.

Als Fazit bleibt somit festzuhalten: Auch wenn die aus unharmonischen Teiltönen aufgebaute Saitenschwingung theoretisch keine Grundfrequenz hat, in der Praxis ermittelt ein Stimmgerät die Frequenz der Grundschwingung bzw. einen Wert, der diesem sehr nahe kommt. Ob das Gehör auch zu diesem Ergebnis kommt, ist hingegen eine ganz andere Frage (Kap. 8.2.3).

## 8.2.2 Frequenz- und Tonhöhengenaugigkeit

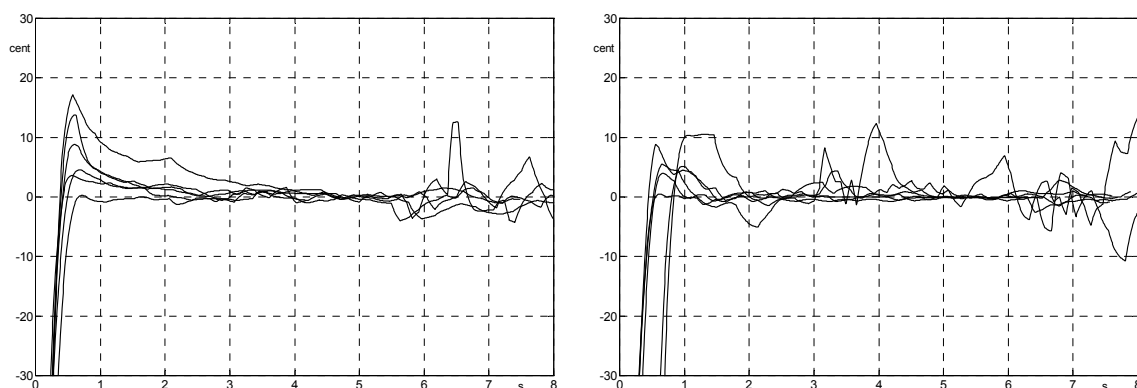
Es wäre naheliegend, nach einem Kapitel über Frequenzmessung die Tonhöhenbestimmung ausführlicher zu erläutern. Zunächst werden aber Fehler und wünschenswerte Genauigkeit vorgestellt, um die im dann folgenden Kapitel dargelegten Gehöreeigenschaften besser bewerten zu können.

Die Messung der Frequenz eines streng periodischen Tones ist mit für Musiker mehr als ausreichender Genauigkeit möglich. Präzisionsfrequenzzähler haben relative Messfehler im Bereich um  $10^{-5}$ ; auch  $10^{-6}$  ist nicht unmöglich. Beispielsweise führt bei einer Uhr ein Fehler von  $10^{-5}$  zu einer Ungenauigkeit von 1 Sekunde / Tag. Das Problem liegt nicht in der zugrunde liegenden Referenz (geheizte Quarzgeneratoren sind hochgenau), sondern im zu messenden Signal. Wenn dieses nicht exakt identische Perioden aufweist, wird die Messung schwierig. Bei genau bekannter Kurvenform ist Frequenzmessung einfach und schnell: Mit Ausnahme einiger weniger Punkte (z.B. Nulldurchgang) reichen drei Punkte einer Sinusschwingung, um die drei Freiheitsgrade: Amplitude, Phase und Frequenz zu ermitteln. Theoretisch dürfen diese drei Punkte in sehr kurzem Abstand aufeinander folgen, und somit sind hohe Messgenauigkeit und kurze Messdauer keine Gegensätze. Diese sehr theoretischen Erkenntnisse der Funktionsanalyse helfen aber bei der Frequenzmessung nicht weiter, denn hierbei ist der Kurvenverlauf nicht bekannt, und dann gilt: **Messdauer** und **Frequenzgenauigkeit** sind zueinander reziprok. Wenn Frequenzmessung auf Periodenzählen beruht, ist für 0,1 Hz Genauigkeit eine Messdauer von 10 s erforderlich. Interaktives Stimmen wird bei derart langen Zeiten natürlich unmöglich. Frequenzverdopplung bzw. Halbperiodenmessung könnten Vorteile bringen, setzen aber voraus, dass das Tastverhältnis bekannt ist – was bei Instrumententönen nicht der Fall ist. Somit bleibt nur, die Frequenz einzelner Teiltöne zu ermitteln. Vermutlich wird in den meisten Fällen die Frequenz des Grundtons (1. Teiltons) bestimmt, und als Saitenfrequenz angegeben.

Nicht nur mit Blick auf das Messverfahren muss der Messdauer Beachtung geschenkt werden; es ist auch zu berücksichtigen, dass das zu messende Signal **zeitvariant** ist. Die Amplituden der Teiltöne nehmen als Funktion der Zeit unterschiedlich schnell ab, es ändern sich aber auch die **Teiltonfrequenzen**. Dies hängt damit zusammen, dass die Saite beim Verlassen ihrer Ruhelage länger, und somit stärker gespannt wird. Je größer die Schwingungsamplitude, desto höher die Frequenz. Des weiteren ist zu berücksichtigen, dass eine reale Saitenschwingung niemals genau **in einer Ebene** erfolgt. Während des Ausschlagens dreht sich die Schwingungsebene, was auch als Überlagerung zweier orthogonaler Schwingungen aufgefasst werden kann. Diese beiden Schwingungen können sich wegen richtungsabhängiger Lagerimpedanzen in ihren Frequenzen geringfügig unterscheiden, und somit treten über der Zeit Amplituden- und Frequenzänderungen auf.

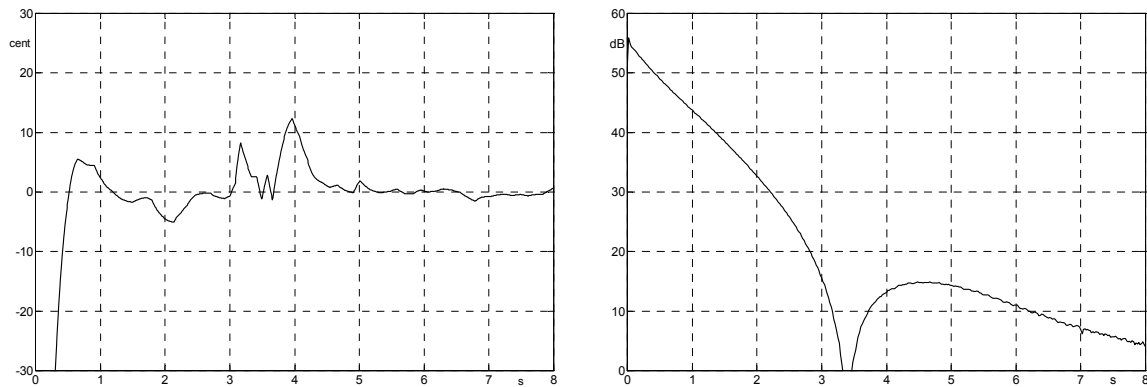
Ein (nicht repräsentativer) Feldversuch sollte Anhaltspunkte liefern, wie genau – trotz all dieser Schwierigkeiten – Saitenfrequenzen gemessen werden können. Aus der Vielzahl der am Markt erhältlichen digitalen **Stimmgeräte** wurden drei ausgewählt und mit Sinusgenerator und Präzisionszähler überprüft. Der Bereich, innerhalb dessen die Messung als "gut" angezeigt wurde, betrug  $\pm 1,6\%$ ,  $\pm 2,0\%$ ,  $\pm 2,3\%$ , im Mittel also  $\pm 2\%$ , entsprechend  $\pm 3,5$  cent. Um nicht missverstanden zu werden: Die drei untersuchten Stimmgeräte zeigten nicht "falsch" an, vielmehr wurden beispielsweise alle Frequenzen zwischen 439,4 und 440,7 Hz als *korrekt auf A gestimmt* bewertet. Die Breite dieses Toleranzintervalls ist ein Kompromiss zwischen hoher Genauigkeit einerseits, die aber wegen der o.a. Probleme u.U. nie erreicht werden kann, und leichter erreichbarer "Stimmung" andererseits, die aber wegen hörbarer Abweichung vom Idealwert nicht akzeptiert wird.

Den zeitlichen Ablauf einer Messung zeigt **Abb. 8.12**. Hierfür wurden mit einem Stimmgerät (Korg GT-2) die Stimmungen zweier Gitarren gemessen; die vom Messinstrument angezeigten Abweichungen vom Sollwert sind über einer Messzeit von 8 Sekunden für jede Saite dargestellt. Jeweils bei  $t = 0$  wurde die Saite normal angezupft; alle nicht beteiligten Saiten waren bedämpft, um Interferenzen zu vermeiden. Bei der Tennesseean fällt auf, dass die Tonhöhe während der ersten Sekunden deutlich abnimmt. Der Effekt wurde nicht näher untersucht, Ursache sind möglicherweise die relativ dünnen Saiten, deren mittlere Zugspannung beim starken Schwingen erhöht wird. Zum Ende der dargestellten Messdauer nehmen die Abweichungen zu, was am abnehmenden Signalpegel liegt. Bei der Ovation (gemessen wurde das Piezosignal) zeigen sich auch zwischendurch immer wieder Schwankungen; deren Ursache wurde genauer analysiert.



**Abb. 8.12:** Tonhöhenmessung mit dem Stimmgerät Korg GT-2. Tennesseean (links), Ovation SMT (rechts).

In **Abb. 8.13** wird die gemessene Tonhöhe mit dem Zeitverlauf des Grundtonpegels verglichen. Signalgeber ist in beiden Fällen die angezupfte h-Saite der Ovation SMT. Bei 3,4 s erkennt man ein Minimum im Verlauf des Grundtonpegels. Mit einer Verarbeitungs-Totzeit von ca. 0,5 s lassen sich die Tonhöhenschwankungen um 4 s erklären; die restlichen Schwankungen können nicht sicher zugeordnet werden.



**Abb. 8.13:** Gemessene Tonhöhenabweichung, Grundtonpegel; Ovation SMT, h-Saite bei  $t = 0$  angezupft.

Die Messungen zeigen, dass trotz vermeintlich digitaler Präzision mit erheblichen Schwankungen des Anzeigewertes zu rechnen ist. Da das Stimmgerät bei Betrieb an einem Präzisionsgenerator sehr genau und ohne nennenswerte Schwankungen anzeigt, kommt als Ursache nur der Gitarrenton selbst in Frage. Je "lebendiger" er ist, desto stärker fluktuiert das Mess-Ergebnis bzw. schwankt die Tonhöhe.

An diese Stelle passt ein kurzer Exkurs in die Grundlagen der **Thermodynamik**. Der lineare Ausdehnungskoeffizient beschreibt, wie sich Abmessungen in Abhängigkeit von der Temperatur verändern. Falls die Abmessungen eingepreßt (erzwungen) sind, ändert sich bei Temperaturänderung die mechanische Spannung. Für Stahlsaiten bedeutet dies: Die ungespannte Saite verlängert sich bei  $+1^{\circ}\text{C}$  Temperaturerhöhung um  $16 \cdot 10^{-6}$ . Dies scheint unbedeutend, verglichen mit den o.a. 2‰. Allerdings muss berücksichtigt werden, dass für die Änderung der Saitenfrequenz die relative *Spannungsänderung* zu berücksichtigen ist. Typischerweise muss eine  $E_2$ -Saite um ca. 1,5 mm gedehnt (gespannt) werden, um richtig gestimmt zu sein. Und in Relation zu diesen 1,5 mm ist die durch Temperaturänderung hervorgerufene Längenänderung zu sehen. Die relative Frequenzänderung entspricht (Wurzel!) der halben relativen Dehnungsänderung. Dies bedeutet für das Beispiel: **Bei  $1^{\circ}\text{C}$  Temperaturänderung ändert sich die Saitenfrequenz um 5,3‰**. Hierbei wurde angenommen, dass die Abmessungen von Hals/Korpus unverändert blieben, was wegen der sehr unterschiedlichen thermischen Zeitkonstanten im Kurzzeitbereich Berechtigung hat. Ein Versuch bestätigte: Als eine gestimmte Gitarre (Gibson ES335) aus dem Zimmer ins Freie gebracht wurde, erhöhte sich innerhalb weniger Sekunden die Frequenz der  $E_2$ -Saite um 12‰; im Freien war es einige Grad kühler als im Zimmer. Umgekehrt folgt daraus: Wer die Stimmung der Gitarre auf 1‰ konstant halten möchte, muss fordern, dass kurzfristige Temperaturschwankungen unter  $0,2^{\circ}\text{C}$  bleiben ☺.

Die wichtigste Frage wurde bis zuletzt aufgespart: **Wie präzise ist eigentlich das Gehör?** In der Terminologie der Psychoakustik: Wie groß ist die Tonhöhenunterschiedsschwelle? Hierzu kann man sehr unterschiedliche Antworten finden, je nach Versuchsmethodik. Grundsätzlich ist zwischen dem **Sukzessivpaar** (2 Töne folgen zeitlich nacheinander) und dem **Zweiklang** (2 Töne werden gleichzeitig gespielt) zu unterscheiden.

Bei der gleichzeitigen Darbietung zweier Töne können u.U. äußerst kleine Frequenzabweichungen bemerkt werden. Wenn man z.B. zwei 1-kHz-Sinustöne um 0,1 Hz gegeneinander verstimmt, so entsteht eine Schwebung, d.h. ein an- und abschwellender Ton, dessen Amplitude alle 10 Sekunden seinen Maximalwert erreicht. Diese Dauer ist gering, verglichen mit der durchschnittlichen Lebenserwartung, und auch klein gegenüber der Duldungsdauer üblicher Versuchspersonen, und somit gut beobachtbar. Auch 0,01 Hz sind aus gleichem Grund beobachtbar, ab 0,001 Hz wird dann vielleicht die endliche Geduld der Versuchsperson zum Problem. Bezogen auf 1000 Hz sind 0,001 Hz bereits  $10^{-6}$ . Daraus aber zu folgern, die Frequenzauflösung des Gehörs würde immer 0,001‰ betragen, ist Unsinn – das Ergebnis ist nur versuchsbezogen nutzbar.

Ein großer Teil der Musik besteht aber aus Zwei- und Mehrklängen; was gilt hier? Die Antwort muss unbefriedigend bleiben, weil Musik vielgestaltig ist, aber grobe Richtlinien gibt es. Eine erste Grenze wird durch die Klangdauer definiert. Wenn ein Zweiklang nur 1 Sekunde dauert, können 0,1 Hz Frequenzabweichung nicht bemerkt werden. Bei lang ausgehaltenen Klängen fällt das Erkennen von Frequenzunterschieden prinzipiell leichter, aber: Lang ausgehaltene Noten werden häufig mit **Vibrato** gespielt (zur Terminologie siehe Kap. 10.8.2), dann fallen kleine Verstimmungen weniger auf. Nun können zwar nicht alle Instrumente ein Tonhöhen-Vibrato erzeugen. Häufig sorgt dann aber mehrhörige Auslegung dafür, dass bereits bei Einzeltönen hörbare Modulationen auftreten. Beispielsweise werden beim Klavier die meisten Töne von zwei oder drei sehr ähnlich gestimmten Saiten erzeugt, da sind Schwebungen systemimmanent. Auch wenn man versucht, alle Saiten eines Klaviertones auf exakt gleiche Tonhöhe zu stimmen, entstehen wegen überkritischer Saitenkopplung Schwebungen. Ob nun neben den im Einzelton hörbaren Schwebungen die beim Zusammenklang zusätzlich entstehenden Schwebungen als eigenes Merkmal hörbar werden – das hängt von zu vielen Faktoren ab, als dass es mit einfachen Algorithmen fassbar wäre. Mit Blick auf die Häufigkeitsverteilung der Dauer musikalischer Töne, und unter Berücksichtigung der auditiven Schwankungsbewertung, kann eine vorsichtige Prognose erfolgen: Ab etwa 1 s Hüllkurvenperiode verlieren Schwebungen an sensorischer Bedeutung; dem entspricht eine diesbezügliche Frequenzauflösung von etwa 1 Hz.

Bei der **zeitlich aufeinanderfolgenden Darbietung** zweier Töne sind Schwebungen ausgeschlossen. Meinen viele Psychoakustiker. Entscheidend ist hierbei aber nicht, welche Schalle generiert werden, sondern was bei der Versuchsperson ankommt. Schalldarbietungen in einem Raum sind immer von **Reflexionen** begleitet, die, so sie in Massen auftreten, **Hall** genannt werden. Wenn die Pause zwischen den sukzessiv dargebotenen Tönen zu kurz ist, entsteht im Übergang vielleicht doch eine kurze Schwebung, die u.U. gehört wird. Deshalb sollten derartige Versuche nur mit Kopfhörerdarbietung erfolgen. Der Raum als Übertragungssystem birgt noch weitere Gefahren: Wegen der Überlagerung zeitlich verschachtelter Reflexionen wird die Impulsantwort in die Breite gezogen; ihre Fouriertransformierte, die Übertragungsfunktion, bekommt selektive Maxima und Minima, und dazwischen einen steilflankigen Verlauf. Eine Frequenzänderung von 1 Hz, die als solche vielleicht unhörbar wäre, erfährt hierdurch eine Pegeländerung von mehreren dB, und das Gehör bemerkt die Änderung. Ursache ist dann zwar die Frequenzänderung, für die Detektion maßgeblich ist aber die Amplitudenunterschiedsschwelle des Gehörs.

Die **Frequenzunterschiedsschwelle** von (sukzessiv über Kopfhörer dargebotenen) Sinustönen, deren Dauer 0,2 s nicht unterschreitet, beträgt im Frequenzbereich unter 500 Hz ca. 1 Hz; über 500 Hz ca. 2‰. Verkürzung der Dauer  $< 0,2$  s verschlechtert die Unterschiedsschwelle. Diese Daten sind Mittelwerte aus einer größeren Zahl psychoakustischer Experimente.

Bei einem Sinuston lässt sich leicht feststellen, ob er zum 1-Hz-Kriterium gehört, oder zum 2‰-Kriterium: Die Grenze liegt bei 500 Hz, hier gehen beide Grenzwerte\* ineinander über. Bei einem aus mehreren Teiltönen zusammengesetzten Ton (Klang) ist diese Entscheidung nicht mehr so einfach. Bei der E<sub>2</sub>-Saite liegen die ersten 6 Teiltonfrequenzen unterhalb von 500 Hz, alle weiteren oberhalb. In solchen Fällen gilt: Frequenzänderungen werden hörbar, wenn bei mindestens einem (hörbaren) Teilton die Frequenzunterschiedsschwelle überschritten wird. Im Falle der E<sub>2</sub>-Saite ist somit nicht von 1 Hz / 82,4 Hz  $\hat{=}$  12‰ auszugehen, sondern vom 2‰-Obertonkriterium. Dieser Wert ist in guter Übereinstimmung mit dem bei Stimmgeräten gefundenen Toleranzbereich. Umgerechnet auf die bei Musikern gängige Einheit cent erhält man für den Tonhöhentoleranzbereich **3 – 5 cent**. Mit cent wird der hundertste Teil eines Halbtonintervalls bezeichnet, 1 cent  $\hat{=}$  0,58‰. Die manchmal geforderte 1-cent-Genauigkeit ist übertrieben: Bei der Gitarre müsste hierfür die Saitentemperatur auf ca. 0,1°C konstant bleiben, und das wird schwierig bei heißen Rhythmen. Lässt sich die Gitarre mit einer Genauigkeit von  $\pm 2\%$  stimmen, ist man auf der sicheren Seite (bzw. Saite). Woraus aber nicht geschlossen werden darf, dass alle größeren Abweichungen bereits als falsch gestimmt wahrgenommen werden. Das Gehör ist im Einzelfall nämlich auch zu großen Kompromissen bereit.

### 8.2.3 Die Tonhöhenwahrnehmung

Es wurde schon darauf hingewiesen, dass Tonhöhe und Frequenz zwei verschiedene Größen sind (Kap. 8). Das Gehör ermittelt die Tonhöhe nach komplizierten Algorithmen, deren umfassende Darstellung den Rahmen dieser Arbeit sprengen würde (hierüber existiert Spezialliteratur). Erster wichtiger Verarbeitungsschritt ist die Frequenz/Orts-Transformation im Innenohr (Cochlea): Im Inneren der schneckenförmigen Cochlea läuft eine Wanderwelle, deren Maximum von Amplitude und Frequenz der Schallwelle abhängt. Winzige Sinneshärchen reagieren auf die Bewegung dieser Wanderwelle und senden Nervenimpulse über verschiedene Nervenfasern ans Gehirn, wo eine weitergehende Verarbeitung stattfindet. Ein normal angeschlagener Gitarrenton besteht aus einer Vielzahl fast harmonischer Teiltöne. So ungefähr die ersten 6 – 8 dieser Teiltöne führen zu unterscheidbaren lokalen Wanderwellenmaxima, die höheren Teiltöne werden zu Gruppen zusammengefasst verarbeitet.

Normalerweise hören wir beim Anzupfen einer Saite nicht die einzelnen Teiltöne, sondern einen komplexen Ton mit *einer* Tonhöhe. Mit etwas Aufwand lassen sich aber einzelne Teiltöne heraushören: Hierzu unterdrückt man zunächst mit einem Sperrfilter (Notchfilter) den betreffenden Teilton, und schaltet dann die Filterwirkung ab, sodass wieder das Originalsignal wiedergegeben wird. Ab diesem Moment wird der betreffende Teilton für einige Sekunden hörbar, um dann wieder mit seinen Kollegen zu einem integralen Klangereignis zu verschmelzen. Voraussetzung hierfür ist ausreichende Teiltonamplitude; der Teilton darf nicht von seinen spektralen Nachbarn so stark verdeckt sein, dass er zum Höreindruck gar nichts beiträgt. Wie Einzelelemente zu Gruppen zusammengefasst werden, ist seit langem ein Forschungsthema der Gestaltpsychologen, das zunächst für das visuelle System die **Gestaltgesetze** ergab (siehe Kap. 8.2.4). Insbesondere das "Gesetz des gemeinsamen Schicksals" spielt auch im auditiven System eine dominierende Rolle, wenn es darum geht, die einzelnen Teiltönen eines komplexen Klangereignisses zu gruppieren, sie Schallquellen zuzuordnen, und denen wiederum Attribute wie z.B. eine Tonhöhe zuzuordnen.

---

\* Sowohl 1 Hz als auch 2‰ sind als Richtwerte zu interpretieren, die individuellen Streuungen unterliegen.

Bei komplexen Klängen, deren Teiltöne *exakt harmonisch* liegen, funktioniert die Tonhöhen-erkennung in der Regel ganz gut – insbesondere, wenn viele Teiltöne vorhanden sind. Wie beim visuellen System mit seinen optischen Täuschungen sind aber auch im auditiven Bereich spezielle Klänge bekannt, die zu scheinbar paradoxen Wahrnehmungen führen. Liegen die Teiltonfrequenzen *unharmonisch*, was z.B. bei Glocken die Regel ist, so erarbeitet der Tonhöhenalgorithmus Schätzwerte mit Wahrscheinlichkeiten. Hierbei kann durchaus der Fall eintreten, dass sich eine Versuchsperson nicht zwischen zwei Tonhöhen entscheiden kann, oder dass zwei Versuchspersonen demselben Schall völlig unterschiedliche Tonhöhen zuweisen. Saitentöne sind aber nur schwach inharmonisch, im schlimmsten Fall sind da Oktavverwechslungen vorstellbar. In der Regel wird als **Tonhöhe eines Saitentones** ein Wert ermittelt, der nahe beim Grundton liegt, mit diesem aber nicht identisch ist. Das Gehör weist in einem ersten Schritt allen nichtverdeckten Teiltönen ihre **Spektraltonhöhen** zu, und ermittelt daraus über eine spektrale **Bewertungskurve**, die bei **ca. 700 Hz** ein flaches Maximum aufweist\*, die **virtuelle Tonhöhe**. Tief- und hochfrequente Teiltöne tragen somit zur Tonhöhe weniger bei als mittelfrequente Teiltöne. Dass insbesondere nicht der Saitengrundton die wahrgenommene Tonhöhe definiert, zeigen Versuche von Plomp<sup>†</sup>: Hierbei wurde bei einem Klavierton die Frequenz des Grundtons um 3% erniedrigt, während gleichzeitig die Frequenzen aller anderen Teiltöne um 3% erhöht wurden; die wahrgenommene Tonhöhe erhöht sich dadurch um 3%. Zwar kann der Grundton (je nach Pegel) großen Einfluss auf die **Klangfarbe** haben, für die Tonhöhe ist er – sofern genügend Obertöne vorhanden sind – eher unwichtig.

Bei der **Gitarre** sind nun aber die Obertöne zu hohen Frequenzen hin progressiv verschoben (bei 1 kHz durchaus +15 cent). Rechnet man hiervon auf die Tonhöhe zurück, ergibt sich ein Wert, der höher ist als der eines **Stimmgerätes**, das nur den Grundton analysiert. Bevor nun aber der Eindruck entsteht, dann müsse beim Stimmen das Stimmgerät eben um 15 cent mehr anzeigen (was nicht der Fall ist), sei zumindest in Stichworten darauf verwiesen, dass die gehörte Grundtonhöhe (bzw. -frequenz) nicht einfach der *n*-te Teil der Frequenz des *n*-ten Teiltones ist: Fastl/Zwicker [12] berichten von Hörversuchen mit harmonisch-komplexen Tönen, deren wahrgenommene Tonhöhe *niedriger* ist als die objektive Grundfrequenz – der Fehler des o.a. Stimmgerätes tendiert somit in dieselbe Richtung wie das Gehör. Ergänzend ist auch noch zu berücksichtigen, dass die Tonhöhe (trotz konstanter Teiltonfrequenzen) vom **Schallpegel** abhängt: Mit zunehmendem Pegel nimmt die Tonhöhe ab, und zwar durchaus bis zu 5 cent pro 10 dB. Noch größere Effekte können **Zusatzschalle** hervorrufen, die sich dem Gitarrenton überlagern: Die Literatur [12] nennt hierfür Verschiebungen, die im Extremfall bis zum Halbtonabstand gehen können! Auch wenn derart große Verstimmungen nicht zum Gitarrenalltag gehören: Alles in allem erkennt man ein weites Feld, das viel Raum für Grundlagenversuche lässt. Das aber auch klar zum Ausdruck bringt: **Ein centgenaues Stimmen ist nicht möglich**. Auch wenn einzelne Teiltonfrequenzen mit hoher Präzision messbar und einstellbar sind – ob's "stimmt", entscheidet das Gehör, und zwar nach komplizierten, sehr individuellen und situationsbedingten Kriterien. Dass in Laborversuchen Tonhöhenunterschiede von **3 – 5 cent** erkannt werden, bedeutet noch nicht, dass diese Genauigkeit generell einzuhalten ist. Ein verbindlicher Grenzwert für hörbar verstimmte Töne kann nicht angegeben werden, weil zu viele Parameter den Einzelfall bestimmen, aber für die Praxis hat sich die Faustregel bewährt: **5 cent maximale Verstimmung sind wünschenswert**, 10 cent häufig akzeptabel. Hörer mit goldenen Ohren dürfen diese Zahlen aber auch gerne halbieren.

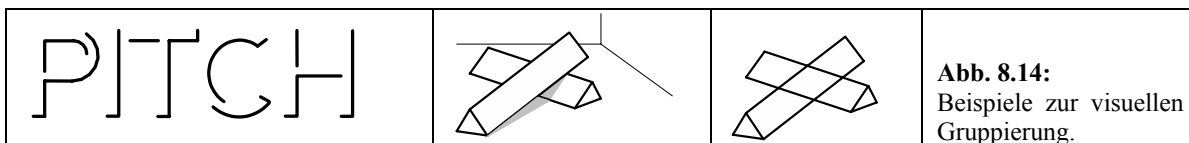
\* Terhardt E.: Pitch, Consonance, and Harmony. JASA 55 (1974), 1061–1069.

† Plomp R.: Pitch of complex tones. JASA 41 (1967), 1526–1533.



### 8.2.4 Teilton-Gruppierung

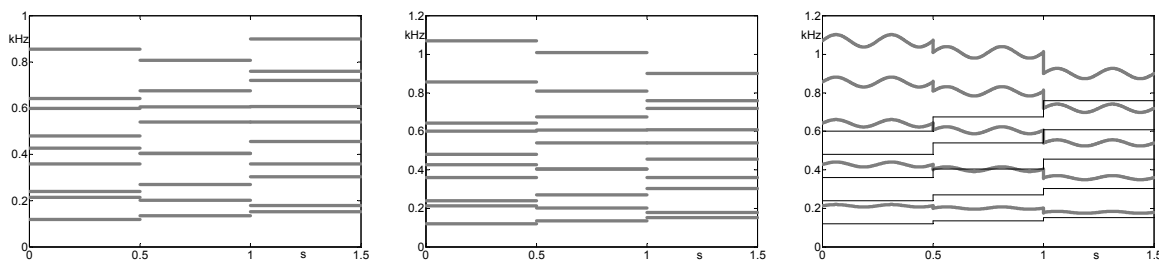
Es ist üblich, Saitenschwingungen als Summe unterschiedlich abklingender Teiltöne zu beschreiben. Diese "Entwicklung nach harmonischen Reihengliedern" ist nicht zwingend, aber das Standardwerkzeug der Spektralanalyse, das seine Berechtigung u.a. aus der Hydromechanik des Innenohrs\* ableitet. Auch wenn es letztlich ein Modell ist: Der Gitarrenton "besteht aus" Teiltönen. Nun hören wir aber beim Anzupfen einer Saite nicht eine Vielzahl einzelner Töne, sondern *einen* Ton – es gibt bei der auditiven Wahrnehmung folglich **Gruppierungs-Mechanismen**, die aus den auf niedriger Verarbeitungsebene gewonnenen Spektraltonhöhen (der nichtverdeckten Teiltöne) Gruppen zusammengehöriger Teiltöne bilden. Das Gehirn, die menschliche CPU, erhält über die Sinnesrezeptoren Informationen und wertet sie aus, d.h. reduziert durch Kategorisierungs- und Entscheidungsprozesse diese immense Datenflut. Als Beispiel: Nur eine Sekunde einer Musik-CD enthält 1,4 Millionen bit Information! Ob davon tatsächlich 50 bit (pro Sekunde) ins Bewusstsein gelangen, wie Psychologen vermuten, oder ein paar mehr oder ein paar weniger: Der überwiegende Teil der ankommenden Information muss "weggeworfen" werden. Aber welcher?



Gestaltpsychologen wie z.B. Max Wertheimer haben aus Versuchen zur visuellen Wahrnehmung **Gestaltgesetze** formuliert, die auch auf Hörwahrnehmungen übertragbar sind. Offenbar wird im Erkennungsprozess die Vielfalt der von den Rezeptoren gelieferten Daten anhand bereits im Gedächtnis gespeicherter Gruppierungsprozesse und -muster reduziert, wird Bekanntem und Plausiblen eine höhere Priorität zugewiesen als Unbekanntem bzw. Unlogischem. Die in **Abb. 8.14** in der Mitte dargestellte Anordnung von zwei Holzstreifen ist auf einen Blick räumlich interpretierbar, obwohl die Zeichenebene ja nur eine ebene Fläche darstellt. Nach kleinen Änderungen (rechtes Bild) will sich der Raumeindruck aber schon nicht mehr einstellen. Es würde zu weit führen, die Gesetze der Nähe, der Ähnlichkeit, des glatten Verlaufs, der Geschlossenheit und des gemeinsamen Schicksals im Detail zu erläutern, hier muss auf die Literatur der Wahrnehmungspsychologie verwiesen werden. Als ein Beispiel, das den Bogen zur Akustik schließt, ist in **Abb. 8.14** das Wort "Pitch" in einer unvollständigen Outline-Schrift gedruckt. Trotz erheblicher Bilddefizite gelingt es unserem visuellen Sinnessystem ohne Probleme, die Linien sinnvoll, d.h. nach gespeicherten Gesetzen, zu vervollständigen und ihnen eine Bedeutung zuzuordnen. "Pitch" wird als Wort erfasst, und nicht als Linienhaufen. Obwohl auch das möglich ist, unser Sehsinn ist da flexibler als unser Gehör. Was beim Sehen willkürlich trennbar ist (Linien oder gruppiertes Objekt), wird beim Hören schwierig bis unmöglich: So leicht wie bei "Pitch" fällt es beim Gitarrenton nicht, zwischen Einzelobjekt (Teilton) und Gruppe (Gitarrenton) hin- und herzuwechseln. Beim Anschlagen der Saite hören wir *einen* (musikalischen) Ton, können aber nur schwer einzelne Teiltöne heraushören. Das ist zwar nicht ganz unmöglich, fällt aber wesentlich schwerer als die Trennung eines abgelesenen Wortes in seine Buchstaben und deren Linien und Bögen. Insofern bestehen also Unterschiede zwischen der visuellen und der auditiven Verarbeitung, aber eben auch Gemeinsamkeiten, wie z.B. bei der Fähigkeit zur Gruppierung, oder die hierarchische Struktur. Nach dem Terhardtschen Tonhöhenmodell werden zuerst (im Innenohr) Spektraltonhöhen ermittelt, und daraus (auf höherer Ebene) virtuelle Tonhöhen.

\* Frequenz-Orts-Transformation, [12] Kapitel 3.

Der unterste (periphere) Verarbeitungsschritt dieser Hierarchie ähnelt einer Kurzzeit-Fourieranalyse (mit gleichwohl sehr speziellen Parametern). Bereits auf dieser Verarbeitungsebene werden Teiltöne ausgesondert, deren Energie so gering ist, "dass man ihr Fehlen nicht bemerken würde". Denn nicht jeder Teilton trägt zum Höreindruck bei: Ist sein Schallpegel im Vergleich zu dem seiner spektralen Nachbarn zu gering, wird er unterdrückt, was die Psychoakustik **Verdeckung** oder Maskierung nennt. Den nicht- bzw. teilverdeckten Teiltönen werden Spektraltonhöhen zugeordnet, die in höheren Hierarchieebenen gewichtet und zu einer virtuellen Tonhöhe zusammengefasst werden. Hierbei stört es nicht, wenn der Grundton eines harmonisch komplexen Klanges im Schallsignal völlig fehlt: So können beim Telefon mit seiner Bandbegrenzung auf 300 – 3400 Hz die ersten beiden Teiltöne einer Männerstimme ( $f_G = 120$  Hz) gar nicht übertragen werden – trotzdem ist die Grundtonhöhe bei Anhören rekonstruierbar, und es entsteht keinesfalls der Höreindruck eines sprechenden Kindes.

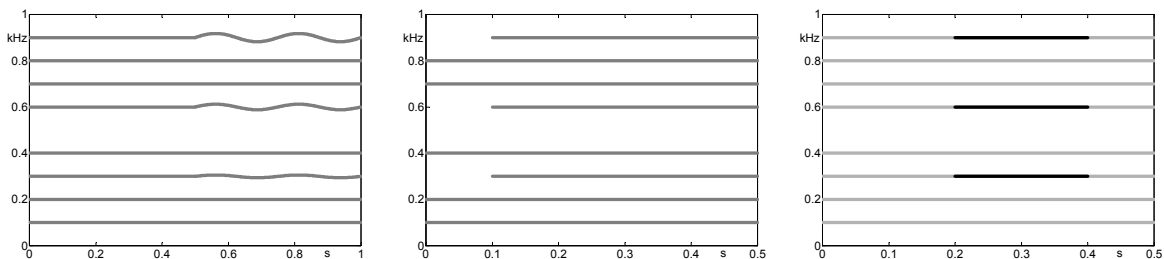


**Abb. 8.15:** Spektrogramme zweier Tonfolgen; im rechten Bild ist die absteigende Folge frequenzmoduliert.

Eine (von mehreren) Gruppierungsregeln besagt, dass *gleichzeitig* beginnende sinusförmige Töne, deren Frequenzrelation *ganzzahlig* ist, wahrscheinlich von derselben Schallquelle stammen und deshalb zu einem Objekt zu gruppieren sind. Natürliche Schallquellen, und nur mit diesen konnte das Gehör während seiner Evolution trainieren, erzeugen so gut wie nie reine Töne. Und sollte es doch einmal vorkommen, es wäre äußerst unwahrscheinlich, wenn im selben Augenblick noch weitere derartige Schallquellen zu tönen begännen, und überdies noch in ganzzahliger Frequenzrelation. Wenn folglich ein derartiger *harmonisch-komplexer Klang* ertönt, kann er nur von *einer* Quelle stammen, und deshalb ist es im Sinne der Informationsreduktion zweckmäßig, die zugehörigen Spektrallinien zusammenzufassen, wie eben auch die beiden Linien der Buchstaben L, V oder T als zusammengehörig erachtet werden. Und wie die visuelle Signalverarbeitung die Überlagerung zweier Buchstaben trennen kann, folgt auch das Gehör *einem* Sprecher – selbst wenn gleichzeitig noch ein zweiter spricht. Das funktioniert nicht perfekt, aber erstaunlich gut: Chucks "Long distance information" ist klar zu verstehen, trotz konkurrierender Begleitinstrumente, und ähnlich ist's bei "O sole mio" und "im fernen Land". Mehr oder weniger, je nach Orchester und Sänger. Der sich bei "der Gral" schon mächtig ins Zeug legen sollte, damit das Publikum nicht erstaunt zur Kenntnis nehmen muss, es sei "der Karl", der dort in der Monsalvat bewacht würde. Denn die Gruppierung der Harmonischen und somit deren Dekodierung gelingt nicht immer fehlerfrei. **Abb. 8.15** vermittelt einen Eindruck, welche Schwierigkeiten entstehen können: Im linken Bild ist das Spektrogramm einer kleinen zweistimmigen Melodie dargestellt. Schwer zu sagen, welche Linien zusammengehören. Im mittleren Bild mit seinem größeren Frequenzumfang beginnt das Bildungsgesetz sichtbar zu werden, aber erst im rechten Bild herrscht Klarheit: Mit unterschiedlicher Linienstärke und frequenzmodulierter Oberstimme fällt die Trennung leicht. Das Gehör, insbesondere das musikalisch ausgebildete, trennt schon ohne **Vibrato** die beiden Stimmen in eine auf- und eine absteigende, mit Vibrato geht's noch leichter. Das ist ein Grund, warum Sänger und Instrumentalsolisten ihre Töne häufig mit Vibrato erzeugen: Sie können damit leichter aus der Vielfalt der Begleittöne herausgehört werden. Da die Modulation aller (Solisten-) Teiltöne gleichartig verläuft, erhält das Gehör eine Gruppierungshilfe.

Der Wahrnehmungspsychologe spricht hier vom **Gesetz des gemeinsamen Schicksals** (law of common fate): Alles, was gleichzeitig beginnt und in gleicher Weise die Tonhöhe ändert, gehört "vermutlich" zusammen. Und damit das Erkennen bzw. Gruppieren besonders leicht fällt, wählt der Solist seine Modulationsfrequenz bei ungefähr 4 – 7 Hz, denn für derartige Modulationen ist das Gehör besonders empfindlich (Schwankungsstärke [12]). Auch die begleitenden Musiker (Orchester bzw. Chor) bedienen sich häufig des Vibratos – teils, weil sie schon gar nicht mehr anders können, vor allem aber, um unschönen Schwebungen zu vermeiden, die bei mehrstimmigem Spiel sonst automatisch entstünden. Für die "Subs" ist bezüglich Vibrato allerdings Zurückhaltung erforderlich, sonst entsteht ein heilloses Durcheinander.

Wie ein Vibrato die Teilton-Gruppierung beeinflusst, zeigt auch das linke Bild in **Abb. 8.16**: Zunächst erklingt ein 100-Hz-Ton, gebildet aus seiner 1., 2., 3., 4., 6., 7., 8. und 9. Harmonischen. Ab der Hälfte ertönt aber **ein zusätzlicher Ton** in Quintlage (streng genommen ist's eine Duodezime), weil die 3., 6. und 9. Harmonische leicht moduliert werden – sie bilden jetzt, neu gruppiert, die 1., 2. und 3. Harmonische des zusätzlichen 300-Hz-Tones.



**Abb. 8.16:** Teiltöne mit gemeinsamem Schicksal (common fate) werden zu Objekten gruppiert.

Im mittleren Bild beginnen einige Teiltöne verspätet: Es erklingt zuerst ein 100-Hz-Ton, gefolgt von einem 300-Hz-Ton. Aber nur, wenn die Verzögerung lang genug ist (z.B. 100 ms). Bei ca. 30 – 50 ms entsteht eine Art Anschlag-Akzent, die verzögerten Teiltöne sind kurz hörbar, als "Belebung" des 100-Hz-Tones. Für noch kürzere Verzögerungen (z.B. 5 ms) verliert dieser Akzent an Bedeutung, es erklingt nur mehr *ein* Ton. Trotz objektiver Verzögerung entsteht dann eine subjektive Gleichzeitigkeit, der *eine* gemeinsame Ursache zugeordnet wird.

Im rechten Bild wird der Pegel der 3., 6. und 9. Harmonischen abrupt geändert, hier durch Schwärzung angedeutet. Man hört einen 100-Hz-Ton, und zwischen 0,2 – 0,4 s einen zusätzlichen 300-Hz-Ton. Ändert sich der Pegel der 3., 6. und 9. Harmonischen aber kontinuierlich, so hört man nur *einen* Ton mit sich ändernder Klangfarbe. Eine abrupte Änderung, so lehrt die Erfahrung, kann nur von einem neu hinzukommenden Objekt verursacht worden sein, während langsame Veränderungen auch Einzelobjekten zugestanden werden.

Das Entdecken und Verstehen der auditiven Gruppierungsalgorithmen, die hier nur anhand weniger Beispielen erläutert wurden, interessiert nicht nur Musiker und Psychoakustiker, sondern zunehmend auch Neurowissenschaftler. Wer sich vertieft mit der kortikalen Hard- und Software beschäftigen möchte, findet in Spitzers "Musik im Kopf" [ISBN 3-7945-2427-6] eine fundierte Ergänzung.

### 8.2.5 Teilton-Inharmonizität

Bedingt durch die dispersive Transversalwellenausbreitung sind die Teiltöne der Gitarrensaite nicht harmonisch\*, sondern spektral gespreizt: Die Frequenz des  $i$ -ten Teiltons ist nicht  $i \cdot f_G$ , sondern etwas größer. Im Detail wurde der analytische Zusammenhang zwischen Biegesteifigkeit und Teiltonspreizung schon in Kap. 1.3 dargestellt, hier sollen nun die damit zusammenhängenden Auswirkungen auf den wahrgenommenen Klang diskutiert werden.

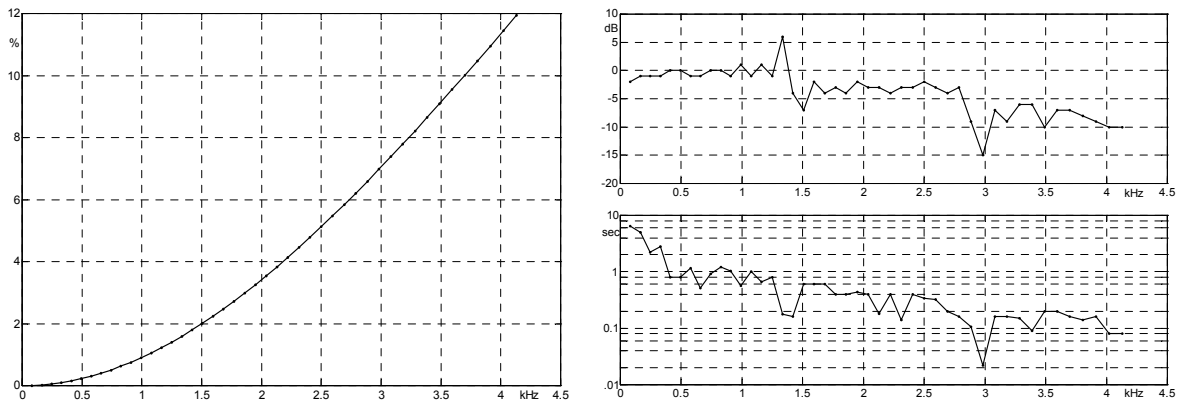
In den folgenden Analysen wird ein reales Gitarrensinal mehreren synthetischen Signalen gegenübergestellt. **Das reale Signal** wurde (ohne Klangfilterung) vom Piezo-Tonabnehmer einer Ovation-Gitarre (Viper EA-68) abgenommen und im Rechner gespeichert. Für die Aufnahme wurde die leere  $E_2$ -Saite (D'Addario EJ-26, 0.052") direkt neben dem Steg mit einem Plektrum griffbrettnormal angezupft; die erste Sekunde des Ausschwingens wurde für Hörversuche verwendet. Für **das synthetische Signal** wurden exponentiell abklingende sinusförmige Schwingungen überlagert und als WAV-File gespeichert.

Die DFT-Analyse des realen Signals ergab den Spreizungsparameter in sehr guter Präzision zu  $b = 1/8000$ ; die Teiltonfrequenzen  $f_i$  berechnen sich hiermit zu:

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

$f_i$  = Teiltonfrequenz;  $f_G$  = Grundtonfrequenz.

**Abb. 8.17** zeigt für die gespreizten Teiltöne die prozentuale Frequenz-Spreizung; Abszisse ist hierbei  $f_i$ , und nicht  $i \cdot f_G$ . In der rechten Bildhälfte sind oben die Pegel der Teiltöne dargestellt, darunter ihre Abklingzeitkonstanten. Bei vielen Teiltönen findet man in guter Näherung ein exponentielles Abklingen, einige zeigen aber starke Hüllkurven-Schwabungen. Für die ersten Versuche wurden diese Schwabungen allerdings ignoriert und durch exponentielles Abklingen angenähert.



**Abb. 8.17:** Prozentuale Teiltonspreizung (links), Teiltonpegel und -Abklingzeitkonstante (rechts).

Die Pegel- und Abklingdaten der Teiltöne des realen Signals waren die Basis zur Erzeugung unterschiedlicher synthetischer Signale:

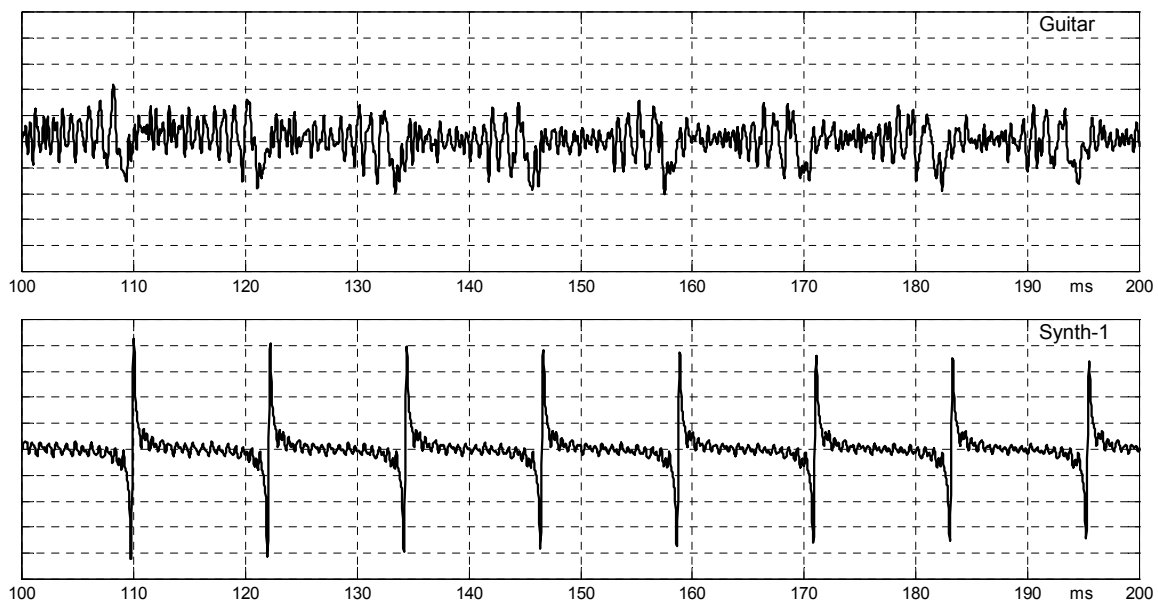
$$u_{\text{synth}} = \sum_i A(i) \cdot e^{-t/\tau(i)} \cdot \sin[2\pi \cdot f_i(i) \cdot t + \varphi(i)]; \quad \text{synthetisches Signal}$$

\* Harmonisches Spektrum: Die Frequenzen der Teiltöne stehen zueinander in ganzzahligen Verhältnissen.

In der Formel steht  $A$  für die Amplitude,  $\tau$  für die Abklingzeitkonstante,  $f_i$  für die gespreizte Frequenz, und  $\varphi$  für die Phase; alle diese Parameter sind Funktionen der Teiltonordnung  $i$ . Die Teiltonphasen des realen Signals waren nicht gemessen worden – im Gegensatz zu Pegelspektren erfordern Phasenspektren eine nicht unerhebliche Nachbearbeitung, um eine gut interpretierbare Kurve zu erhalten.

**Für einen ersten Hörversuch** wurde ein synthetisches Signal erzeugt, dessen Teiltonamplituden und -Abklingzeitkonstanten dem realen Signal entsprachen; alle Teiltonphasen waren aber zu null gesetzt, alle Teiltonfrequenzen ganzzahlig vielfach zur Grundtonfrequenz, d.h. nicht gespreizt. Das derart synthetisierte Signal klingt anders als das reale Signal. Angesichts der in Abb. 8.17 dargestellten Frequenzverschiebungen könnte man spontan an einen Tonhöhenunterschied denken – dieser wurde beim ersten Hörversuch auch tatsächlich bemerkt. Die "exakte" Grundfrequenz des realen Signals ist bei nur 1 s Signaldauer allerdings gar nicht genau genug ermittelbar, und sie verändert sich auch während des Ausschwingens (Saitenmechanik). Deshalb wurde das synthetische Signal nach Gehör auf  $f_G = 81,9$  Hz gestimmt, und damit war die Tonhöhe ausreichend gut angeglichen. Danach konnte der **wesentliche Klangunterschied** im Hörversuch ermittelt werden: Der synthetische Klang wird mit "klarer, schnarrender, räumlich kleiner" beschrieben, der reale Klang mit "raschelder, metallener, räumlich größer". Bei Schalldarbietung über Lautsprecher (Breitbandchassis, normal reflektierendes Zimmer) ergab sich zusätzlich ein interessanter Abstandseffekt: Mit zunehmendem Lautsprecherabstand werden reales und synthetisches Signal subjektiv immer ähnlicher.

Das Gehör hat keinen Rezeptor, der den im Gehörgang ankommenden Schalldruck zeitlich analysiert. Vielmehr wird das Schallsignal zuerst in einem hydromechanischen Filter in spektrale Bänder (sog. Frequenzgruppen) zerlegt [12], und erst daraufhin in elektrische Nervenimpulse (Aktionspotentiale) umcodiert. Gleichwohl macht es Sinn, einen Blick auf die Zeitfunktionen der Schallsignale zu werfen – zumindest, solange man die im Gehör stattfindende Bandpassfilterung nicht ganz aus den Augen verliert. In **Abb. 8.18** sind die Zeitfunktionen des realen und des synthetischen Signals dargestellt – sie unterscheiden sich erheblich.



**Abb. 8.18:** Zeitfunktion des realen und des synthetischen Signals; E<sub>2</sub>-Saite.

Das in Abb. 8.18 dargestellte synthetische Signal ist periodisch, das reale Signal nicht. Der Hauptunterschied der beiden Signale ist aber nicht in der Periodizität zu finden, sondern im Crestfaktor (Spitzenwert : Effektivwert). Die große Impulshaltigkeit des synthetischen Signals zeigt sich auch in einer gehörbezogenen Spektralanalyse (Abb. 8.19), wie sie z.B. von der CORTEX-Software "VIPER" erzeugt wird: Hier ist als Abszisse die Zeit dargestellt, als Ordinate die in der Einheit Bark skalierte Tonheit (eine nichtlinear abgebildete Frequenz [12]), und als Farbe codiert eine spezifische Erregungsgröße, die aus der im Innenohr durchgeführten Signalfilterung abgeleitet wird. Während beim synthetischen Signal das Gehör über die ganze Signalbandbreite gleichzeitig erregt wird, tritt diese Synchronität beim realen Signal nur im tieffrequenten Bereich auf. Beim Anblick dieser Bilder wird verständlich, warum das synthetische Signal als "schnarrend", und das reale Signal als "raschelnd" bezeichnet wurde, und man kann auch vermuten, warum der Abstand zwischen Lautsprecher und Zuhörer so einen großen Einfluss auf den Klang hat: Bei größerem Abstand werden beim synthetischen Signal die Lücken zwischen den Impulsen mit Echos aufgefüllt, es nähert sich damit dem realen Signal an. Offensichtlich ist nicht die Inharmonizität per se das Besondere am realen Signal, sondern das Fehlen einer strengen zeitperiodischen Struktur mit großer Impulshaltigkeit.

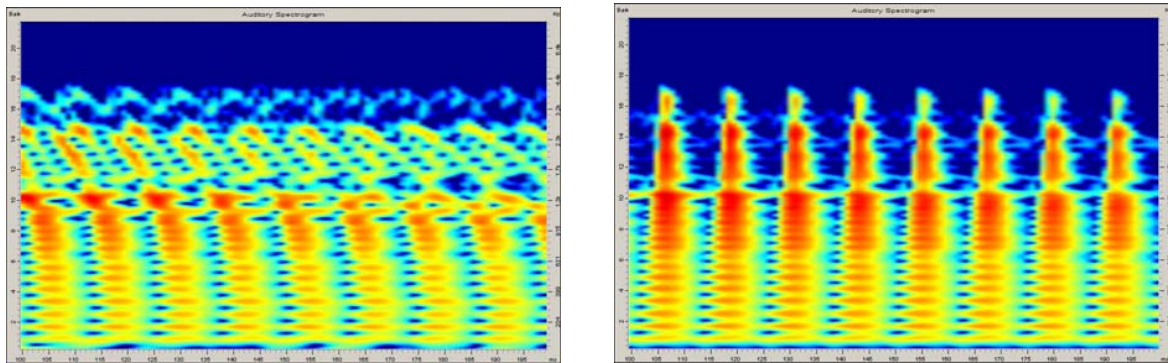


Abb. 8.19: Auditory Spectrogram (CORTEX-VIPER), reales Signal (links), Synth-1 (rechts).

Es gibt einen einfachen Weg, diese Hypothese zur Impulshaltigkeit bzw. Harmonizität zu überprüfen: Wählt man nicht alle **Teiltonphasen** zu null, sondern statistisch gleichverteilt, so erhält man ein sog. Pseudo-Noise-Signal: Aufgrund seiner streng harmonischen Teiltonstruktur ist es periodisch, der Verlauf innerhalb einer Periode (die in diesem Fall ca. 12 ms beträgt) ist aber zufallsbedingt. Abb. 8.20 zeigt das Auditory Spectrogram, Abb. 8.21 die Zeitfunktion. Obwohl auch dieses Signal (wie schon Synth-1) nicht die Frequenzspreizungen des realen Signals besitzt, klingt es doch fast genau so wie dieses. Einige Probanden mit trainiertem Gehör werden zwar immer noch kleine Unterschiede entdecken – insbesondere der Anschlag

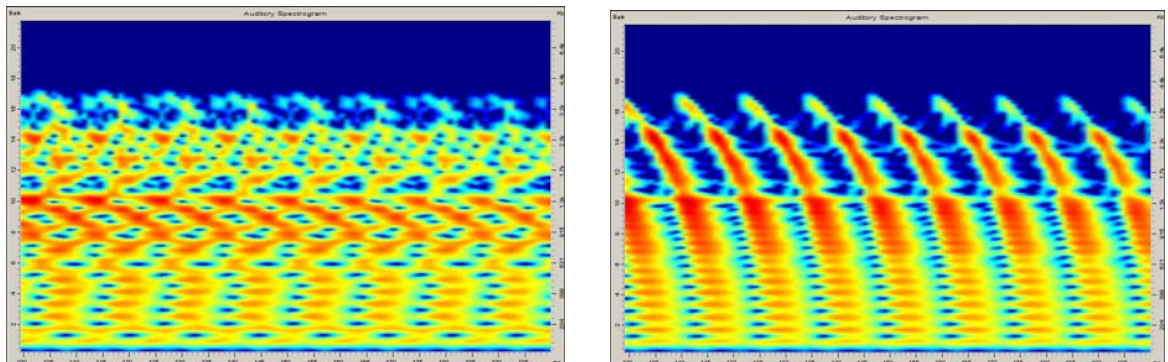


Abb. 8.20: Auditory Spectrogram (CORTEX-VIPER), Synth-2 (links), Synth-3 (rechts).

klings nicht ganz so präzise – der Unterschied zwischen dem realen Signal und Synth-1 ist aber wesentlich größer als der Unterschied zwischen dem realen Signal und Synth-2. Das dem realen Signal eigene *Rascheln* ist auch bei Synth-2 vorhanden, hingegen fehlt das für Synth-1 charakteristische *Schnarren*. Sehr kritische Probanden könnten bei Synth-2 vielleicht "eine Spur zu viel Rascheln" heraushören, die meisten werden im Vergleich zum realen Signal aber gar keinen Unterschied bemerken. Eine Alternative zur gleichverteilten Phase ist ein von M. R. Schröder\* vorgeschlagener Phasenfrequenzgang, der ebenfalls einen kleinen Crestfaktor garantiert. Das als Synth-3 bezeichnete Signal hat ein harmonisches (d.h. nicht gespreiztes) Spektrum, dessen Teilton-Phasen nach folgender Formel definiert sind:

$$\varphi(i) = i^2 \cdot \pi \cdot 0,04;$$

Schröder-Phase

Beim ersten Anhören unterscheiden sich das reale Signal, Synth-2 und Synth-3 nur wenig; Synth-1 klingt hingegen deutlich anders. Bei Kopfhörerwiedergabe wird das trainierte Gehör zwar zwischen allen vier Signalen Unterschiede bemerken, bei Lautsprecherwiedergabe (Nahbereich) klingt hingegen nur noch Synth-1 anders, und bei großem Lautsprecherabstand klingen alle vier Signale praktisch gleich.

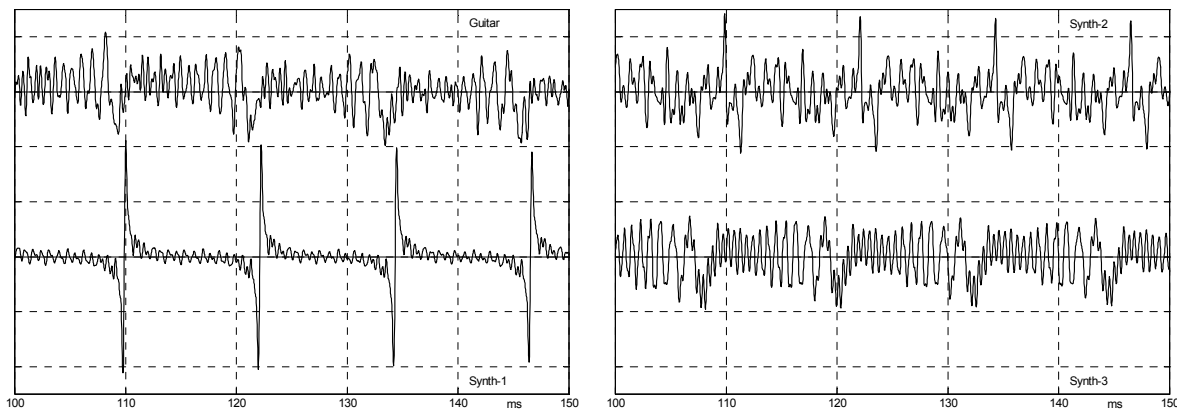


Abb. 8.21: Zeitfunktionen des realen und der drei synthetischen Signale.

Da alle drei synthetischen Signale identische Amplitudenspektren besitzen, und teils ähnlich, teils unterschiedlich klingen, kann der Frequenzauflösung des Gehörs diesbezüglich keine Bedeutung zukommen. Grundlage des Klangunterschiedes ist einzig und allein der Phasenunterschied; nur in diesem Parameter unterscheiden sich die zur Synthese verwendeten Formeln. Wird eines der Signale über Lautsprecher abgestrahlt, ändern sich die Teiltonfrequenzen nicht, die Teiltonphasen schon. Diese plakative Aussage ist aus signaltheoretischer Sicht zwar nicht ganz korrekt, weil ein abklingender Teilton ja nicht durch eine einzige Frequenz, sondern durch ein kontinuierliches Spektrum beschrieben wird, das sehr wohl von Lautsprecher und Raum verändert werden kann – gleichwohl als Näherung brauchbar. Das direkte Auswerten von Phasenfrequenzgängen hilft allerdings noch nicht weiter: Das Gehör hat keinen Rezeptor, der a priori die Phase ermittelt. Vielmehr tasten im Cortischen Organ kleine Sinneshäärchen die frequenzselektiv schwingende Basilarmembran ab, und deren Schwingungs-Hüllkurve liefert die Grundlage für Schwankungs- und Rauigkeitsempfindungen [12]. Das dem Signal Synth-1 bescheinigte *Schnarren* ist ein typisches Attribut eines rauen Schalles. Die klassische Psychoakustik definiert **Rauhigkeit** als die zu einer schnellen Signalmodulation

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

gehörende Empfindung; "schnell" sind Modulationen, deren Modulationsfrequenz im Bereich zwischen 20 und 200 Hz liegt. Der Linienabstand aller drei synthetischer Signale liegt mit 82,3 Hz ganz nahe bei 70 Hz, der Standardfrequenz für Rauigkeitsskalierungen. Neben der Modulationsfrequenz sind aber auch die zeitlichen Erregungsfunktionen aus benachbarten Basilmembran-Bereichen zu bewerten: Ihre Kreuzkorrelationsfunktionen (KKF) sind eine Art Gewichtungsfunktion für die aus einzelnen Teilrauigkeiten gebildete **Gesamtrauhigkeit\***. Bei Synth-1 sind alle Frequenzbänder gleichzeitig aktiv, was in Abb. 8.19 daran zusehen ist, dass die roten Bereiche übereinander liegen (bei gleichen  $t$ -Werten). Gleichzeitigkeit ist eine notwendige Bedingung für Rauigkeit. Bei Synth-2 (Abb. 8.20) sind hingegen die roten Bereiche zerstreut, sie treten in den einzelnen Frequenzbändern zu unterschiedlichen Zeiten auf, und deshalb ergibt sich kein schnarrender, rauer Klang, sondern ein raschelder♦.

Neben Angaben zur Rauigkeit der Signale beurteilten Probanden auch die wahrgenommene Größe der Schallquelle. Ein typisches Phänomen der Wahrnehmungspsychologie: Während ja die objektive Schallereignisgröße, nämlich die Lautsprecherabmessungen, unverändert bleibt, ändert sich mit den Phasenänderungen die **Hörereignisgröße**: Synth-1 scheint punktquellenartig aus der Mitte der Lautsprechermembran zu kommen, Synth-2 kommt hingegen aus einem Raumbereich. Nicht sehr groß, vielleicht 10cm x 10cm, aber eben nicht punktförmig. Und noch etwas fällt auf: Alle Schalle außer Synth-1 scheinen hinter dem Lautsprecher zu entstehen, sie haben mehr räumliche Tiefe. Dieser Eindruck entsteht vor allem, wenn man zuerst Synth-1, und danach eines der anderen synthetischen Signale anhört. Eine Erklärung hierfür könnte sein, dass das Gehör bei Synth-1 keine Echos entdecken kann, und im Vergleich dazu die anderen beiden synthetischen Schalle als ähnlich, aber mit sehr frühen Echos behaftet interpretiert. Und Echos verleihen Räumlichkeit und Größe, auch wenn sie aus derselben Richtung kommen wie der Primärschall.

**Zusammengefasst** ergibt sich: Die Teiltonfrequenzen des realen Signals sind gespreizt, diese Spreizung hat auf die Tonhöhe aber nur einen untergeordneten Einfluss. Vergleicht man das reale Signal mit einem synthetischen, dessen Teiltonpegelverläufe denen des realen entsprechen, dessen Teiltonfrequenzen aber harmonisch, d.h. nicht gespreizt sind, so ergibt sich ein sehr ähnlicher Klangeindruck, wenn die Teiltonphasen so gewählt werden, dass der Crestfaktor nicht zu groß wird. Setzt man hingegen alle Teiltonphasen zu null, entsteht ein andersartiger, schnarrenderer Klang, der (bei Lautsprecherwiedergabe) aus einem Raumpunkt zu kommen scheint, während alle anderen Schalle aus einem Raumbereich kommend wahrgenommen werden.

Nun wird die Synthese dahingehend modifiziert, dass die Teiltonfrequenzen über die weiter oben angegebene Spreizungsformel definiert werden ( $b = 1/8000$ ). **Synth-4** ist ein synthetisches Signal, dessen Teiltonfrequenzen und Teiltonpegelverläufe denen des realen Signals entsprechen. Unterschiede bestehen in den Teiltonpegelphasen (bei Synth-4 sind alle Phasen null), und im detaillierten Teiltonpegelverlauf – wie schon erwähnt, werden schwebend abklingende Teiltöne bei allen synthetischen Signalen durch exponentiell abklingende Teiltöne ersetzt. Die inharmonische Synthese überzeugt auf Anhieb, Synth-4 ist selbst bei Kopfhörerdarbietung vom realen Signal kaum unterscheidbar. Dabei offenbaren sowohl die Zeitfunktionen als auch die Spektrogramme (**Abb. 8.21**) Unterschiede, was aber zu erwarten ist: Die Synthese ist auf lediglich 45 exakt exponentiell abklingende Teiltöne ( $f < 4,1$  kHz) begrenzt.

\* W. Aures: Ein Berechnungsverfahren der Rauigkeit, *Acustica* 58 (1985), 268-281.

♦ Der Begriff "rauschähnlicher Klangeindruck" ist aus signaltheoretischer Sicht naheliegend, mit Blick auf eine spezielle Untermenge der angesprochenen Zielgruppe aber u.U. irreführend – deshalb "Rascheln".



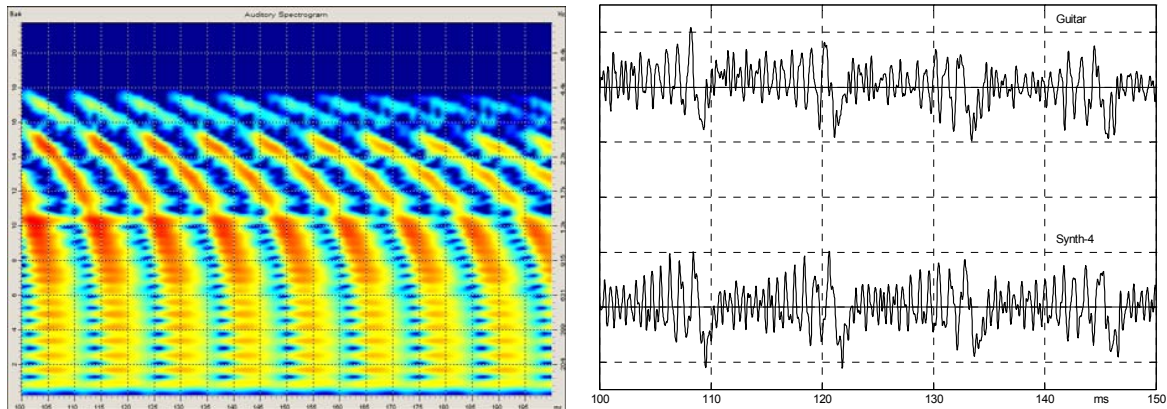
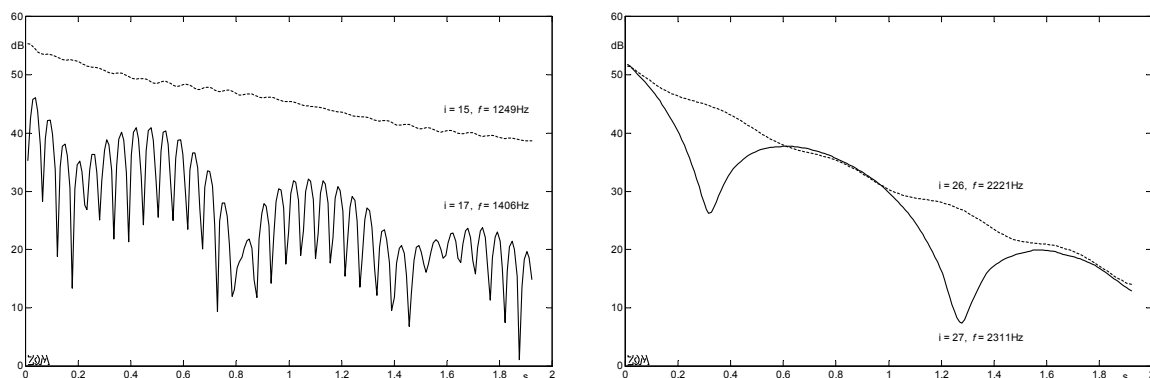


Abb. 8.22: Synthetisches Signal mit gespreiztem Spektrum (Synth-4).

Die Spreizung der Teiltöne führt im Zeitbereich zu einem fortschreitenden Synchronisationsverlust. Im Moment des Anzupfens (Attack) müssen alle Teiltöne phasengleich zusammenarbeiten, um die abrupte Signaländerung zu bewerkstelligen. Bei Synth-1 wird der Attack in identischer Form periodisch wiederholt, die spektralen Maxima treten zeitgleich auf (pancochleäre **Synchronität**), der Ton schnarrt. Bei Synth-2 und Synth-3 ist diese pancochleäre Synchronität durch Laufzeitverzerrungen weitgehend zerstört, die Erregungsperiode bleibt aber in allen Frequenzgruppen konstant. Bei Synth-4 nimmt die Erregungsperiode mit zunehmender Frequenz ab, die für die Rauigkeitsbildung zugrunde liegende KKF wird zeitvariant. Dass eine echte pancochleäre Synchronität wegen der (maximal) ca. 6ms dauernden intracochleären Laufzeit gar nicht zustande kommt, stört nicht: Das Gehör hat sich daran gewöhnt, alle Impulse erleiden dieses Schicksal – und bleiben doch ein Objekt.

Es ist nicht selbstverständlich, dass Änderungen im **Phasenspektrum** hörbar werden. Würde man denselben Versuch mit 500 Hz Grundfrequenz wiederholen, die o.a. Phasendrehungen würden zwar die Zeitfunktion verändern, wären aber nicht wahrnehmbar. Es hat sich als zweckmäßig erwiesen, die auditive Zeitauflösung zu ca. 2 ms anzunehmen; bei 82 Hz Grundfrequenz kann das Gehör somit noch "in die Zwischenräume hören", bei 500 Hz nicht mehr. Eine besondere Sensibilität für die Zeitverläufe der frequenzgruppenspezifischen Lautheit ist aber offensichtlich nicht vorhanden: Synth-1 wird eindeutig als andersartig erkannt, Synth-2 und Synth-3 klingen aber sehr ähnlich, trotz unterschiedlicher KKF. Wobei diese Ähnlichkeit durchaus intraindividuellen Streuungen unterliegt: So kann es vorkommen, dass ein spezieller Schall als zu tief gestimmt empfunden wird. Änderung der Grundfrequenz (von z.B. 81,9 Hz auf 82,3 Hz) behebt diese Diskrepanz, jetzt stimmt's. Perfekt. Ein paar Minuten später ist derselbe Ton aber plötzlich zu hoch – und muss auf z.B. 81,9 Hz runtergestimmt werden. Im besten Fall kann das Gehör 0,2% Frequenzunterschied bemerken [12]. Kann – muss aber nicht. Die Hörversuche vermittelten den Eindruck, als ob die Aufmerksamkeit des Probanden selektiv arbeitete: Einmal wird der Tonhöhe mehr Beachtung geschenkt, ein anderes Mal der Rauigkeit. Oder Attributen, die den Rahmen allgemeinverständlicher Adjektive sprengen. "Topfig" (auch bei WORD unbekannt), "drahtig", "metallisch", "rollend" ... irgendwie möchte man die noch verbleibenden Klangfarbenunterschiede beschreiben, indes, die Semantik lässt einen im Stich. Und hoffentlich muss das Ganze nie in eine Fremdsprache übersetzt werden. Wer hätte gedacht, dass "**kinzokuseino**" metallisch heißt? Oder "**hakuryokunoaru**" kräftig? Nach was klingt "**namerakadadenai**"? Hört man da "Rauhigkeit" heraus? Beziehungsweise "**chruuschness**", so man aus multinationalen Forschungsprojekten franklikanisch gewöhnt ist.

Die meisten Teiltöne des realen Gitarrensingals klingen in guter Näherung exponentiell ab, bei einigen sind aber **Schwebungen** zu beobachten. Deren Ursache an dieser Stelle jedoch nicht näher untersucht wird – geht es doch vor allem um Hörwahrnehmungen. Bereits der zweite Teilton lässt kurz nach dem Ende der 2 s dauernden Aufnahme ein Schwebungsminimum vermuten; also ca. 5 s Schwebungsperiode. Für die auf 1 s begrenzten Hörversuche ist dieser Abfall noch gut mit einem exponentiellen Abklingen zu approximieren, aber schon beim 17. Teilton sind zwei Schwebungen kombiniert: Eine langsame mit 1,6 Hz, und eine schnellere mit 18,4 Hz Schwebungsfrequenz (**Abb. 8.23**). Der (insbesondere im Vergleich zum 15. Teilton) relativ geringe Teiltonpegel sorgt aber dafür, dass diese Schwankung praktisch nicht wahrgenommen werden kann – sie ist verdeckt [12]. Wieder andere Verhältnisse findet man beim 27. Teilton: Hier liegt eine klassische Schwebung mit 950 ms Schwebungsperiode vor. Auf den ersten Blick ist keine eklatante Verdeckung zu erkennen, die benachbarten Teiltöne haben ähnliche Pegel – aber sie klingen alle relativ glatt ab, sodass der gesamte Frequenzgruppenpegel, der ja aus 4 Teiltonpegeln zu bilden ist, so gut wie keine Schwankung aufweist. Die aus einer schmalbandigen **DFT-Analyse** gewonnenen Teiltonpegel liefern objektive Signalparameter, ermöglichen für sich betrachtet aber noch keine Aussagen über die Hörbarkeit spezieller Klangattribute. Aber auch psychoakustische Berechnungsverfahren, wie z.B. die Rauigkeits- oder die **Schwankungsanalyse**, sind mit Vorsicht zu gebrauchen: Zu unvollständig ist noch das Wissen über die Interaktionen inharmonischer Klänge. Die besten Ergebnisse über die Hörbarkeit von Teiltenschwebungen liefert der Hörversuch – nicht verwunderlich, natürlich. Im Falle des o.a. Gitarrentons führt er zu einer klaren Aussage: Trotz inharmonischer Teiltöne sind praktisch keine Schwebungen zu hören.



**Abb. 8.23:** Abklingkurven einzelner Teiltonpegel; Ovation-Gitarre, Piezo-Tonabnehmer.

Aus der Tatsache, dass bei dem hier vorliegenden Gitarrenton keine Schwebungen hörbar sind, darf aber nicht geschlossen werden, dass Schwebungen generell unhörbar sind. Sie sind vorhanden, und sie werden hörbar, wenn sich ihre Teiltonpegel ausreichend aus ihrer spektralen Nachbarschaft abheben. Ursachen können Tonabnehmer-Magnetfelder sein (Kap. 4.11), oder Modenkopplungen in den Saitenlagern (Kap. 1.6.2). Teilton-Inharmonizitäten können hingegen (für sich betrachtet) nur geringe Schwankungs-Wahrnehmungen erzeugen. Oktav-Schwebungen [Plomp, JASA 1967] oder zeitvariante Kreuzkorrelationen [Aures, Acustica 1985] erklären nur sehr dezente Schwankungen – deutlich schwebende Teiltöne erfordern im Spektrum jedoch zwei nahe benachbarte Linien gleichen oder ähnlichen Pegels. Und die können alleine durch Inharmonizität nicht erzeugt werden. Man möchte noch "**im LZI-System**" hinzufügen, um wirklich an alles gedacht zu haben. Und merkt schlagartig, dass diese Einschränkung gerade bei Gitarrenverstärkern in vielen Fällen nicht erfüllt ist. Spektrale Inharmonizität kann sehr wohl Nachbartöne erzeugen, wenn man **Nichtlinearitäten** zulässt!

**Nichtlineare Verzerrungen** treten im Gitarrenverstärker in unterschiedlicher Stärke auf. Die mit dem Piezo-Tonabnehmer abgenommene Akustikgitarre wird man meistens nicht hörbar verzerren wollen, die über Magnettonabnehmer abgenommene Elektrogitarre – je nach Stilrichtung – vielleicht schon. Eine Nichtlinearität, d.h. vereinfacht ausgedrückt eine krumme Übertragungskennlinie, bereichert das Spektrum um zusätzliche Töne. Ein aus zwei Primärtönen bestehendes Tongemisch (Eingangssignal der Nichtlinearität)

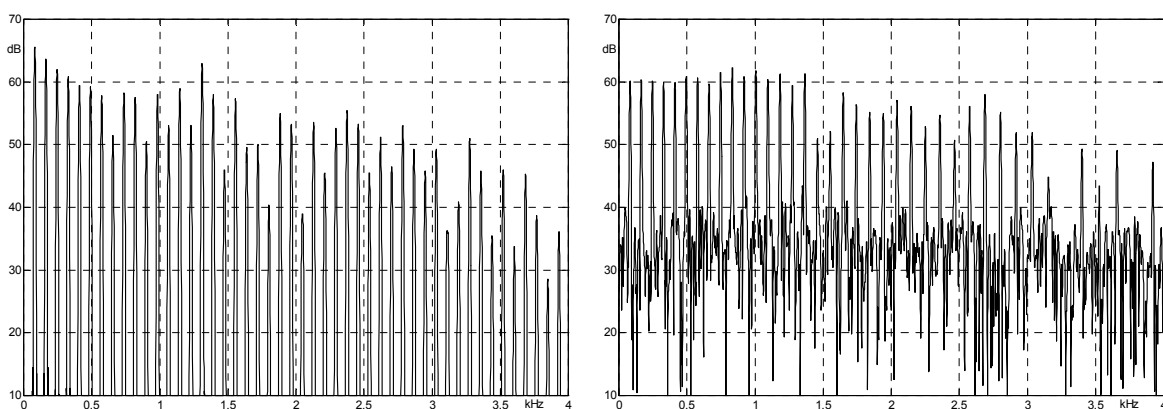
$$x(t) = \hat{x} \cdot [\cos(2\pi \cdot f_1 \cdot t) + \cos(2\pi \cdot f_2 \cdot t)]$$

wird an der nichtlinearen, in eine Reihe entwickelten Übertragungskennlinie

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

auf das Ausgangssignal  $y(t)$  abgebildet. Für rein quadratische oder rein kubische Verzerrungen lässt sich das zu  $y(t)$  gehörende Spektrum leicht berechnen [z.B. 3], für Verzerrungen beliebiger Ordnung entsteht für das o.a. Eingangssignal ein Verzerrungsspektrum, das harmonisch zur neuen Grundfrequenz  $\text{ggT}(f_1, f_2)$  ist. Die Operation  $\text{ggT}(f_1, f_2)$  ermittelt den *größten gemeinsamen Teiler* der beiden Frequenzen  $f_1$  und  $f_2$ . Für  $f_1 = 500$  Hz und  $f_2 = 600$  Hz ergibt sich beispielsweise ein Verzerrungsspektrum mit Spektrallinien bei ganzzahlig Vielfachen von 100 Hz, für  $f_1 = 510$  Hz und  $f_2 = 610$  Hz ein Verzerrungsspektrum mit Spektrallinien bei ganzzahlig Vielfachen von 10 Hz.

Verallgemeinert man nun das Zweitonsignal  $x(t)$  zum  $n$ -Tonsignal, so ist dessen Verzerrungsspektrum im allgemeinen Fall harmonisch zu einer Grundfrequenz, die dem größten gemeinsamen Teiler aller  $n$  beteiligten Primärtonfrequenzen entspricht. Ist  $x(t)$  ein zeitperiodisches Signal mit der Periodendauer  $T$ , so ist sein Spektrum harmonisch, d.h. alle Teiltonfrequenzen sind ganzzahlig vielfach zu  $f_G = 1/T$ . Der größte gemeinsame Teiler aller Teiltonfrequenzen ist ebenfalls  $f_G$ , und somit ändert eine Nichtlinearität nichts an der Harmonizität (bzw. an der zeitlichen Periodizität). Bei einem gespreizten Spektrum erhält man hingegen eine unüberschaubare Vielfalt neuer Frequenzen (die Wurzelfunktion ist irrational), die ein rauschendes oder prasselndes Zusatzgeräusch erzeugen. **Abb. 8.24** zeigt für ein zeitperiodisches Signal (Synth-1) und ein synthetisches Signal (ähnlich Synth-4, aber  $b = 1/3000$ ) das an einer punktsymmetrischen Verzerrungskennlinie entstehende Spektrum. In diesem Konglomerat aus überlagerten Primär- und Verzerrungstönen ist alles möglich – auch hörbare Schwebungen.



**Abb. 8.24:** Spektren der nichtlinear verzerrten Signale. Harmonisches Primärsignal (links), gespreizt (rechts). Vergleiche auch: Kap. 10.8.5.

**Fazit:** Aufgrund ihrer Biegesteifigkeit haben Saiten kein harmonisches, sondern ein gespreiztes, inharmonisches Spektrum; die zugehörige Zeitfunktion ist deshalb nicht periodisch. Vergleicht man den inharmonischen Klang mit einem harmonischen Klang (gleicher Grundfrequenz), dessen Teiltonpegel denen des inharmonischen Klanges zumindest näherungsweise entsprechen, so kommt der *Teiltonphase* große Bedeutung zu. Ein harmonischer Klang, dessen Teiltonphasen alle null (oder  $\pi$ ) sind, klingt schnarrend, und deutlich anders als ein realer Gitarrenton. Mit einer geeigneten Phasenfunktion, die einen kleinen Crestfaktor erzeugt, kann man aber harmonische Klänge synthetisieren, die sich klanglich kaum von den inharmonischen unterscheiden. Auch wenn das trainierte Gehör bei Kopfhörerwiedergabe noch kleine Unterschiede bemerken kann – bei der für Gitarre typischen Lautsprecherwiedergabe sind praktisch keine Unterschiede mehr zu hören. Erst wenn der Spreizungsparameter  $b$  wesentlich größer als etwa  $1/5000$  angenommen wird (bei Gitarrensaiten untypisch), bemerkt man die Inharmonizität deutlich. Bei  $b = 1/500$  entsteht z.B. das für den Stundenschlag eine Wanduhr charakteristische "Bong", mit  $b = 1/100$  nähert man sich synthesesizerähnlichen Klängen. Wird allerdings ein stark nichtlineares System (Verzerrer) in den Signalweg geschaltet, können auch schwach inharmonische Signale ihr Spektrum und damit ihren Klang total verändern; harmonische Klänge ändern dabei nur ihre Amplituden und Phasen, bleiben aber harmonisch.

Diese Aussagen sollten als Ergebnisse einer kleinen Versuchsreihe interpretiert, und nicht auf alle beliebigen Instrumentenklänge verallgemeinert werden. Ziel dieser Untersuchungen war nicht, die absolute Wahrnehmbarkeitsschwelle für Inharmonizität zu finden, sondern die eher geringe Bedeutung gitarrentypischer Inharmonizitäten aufzuzeigen. Bei andersartigem Abklingen der Teiltöne hoher Ordnung können sehr wohl Inharmonizitäten bemerkt werden, denen ein viel kleinerer Inharmonizitätsparameter zugrunde liegt (Järveläinen, JASA 2001).

Formel-Zusammenstellung:

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

Die Winkelfunktion wurde als Sinus angesetzt, um den Crestfaktor nicht noch größer zu machen.

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

Die Phasenwinkel  $\varphi(i)$  sind im Intervall  $[0 \dots 200^\circ]$  gleichverteilt.

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

Die Phasenwinkel  $\varphi(i)$  berechnen sich nach Schröder zu:  $\varphi(i) = 0,04 \cdot \pi \cdot i^2$ . Dies entspricht einer über der Frequenz linear zunehmenden Gruppenlaufzeit.

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

Die Teiltonfrequenzen sind inharmonisch gespreizt.

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