# Themenvorschlag für Magisterarbeit

In den meisten mir bekannten spektralen Modellen wird ein Klang als Summe von Sinustönen plus einem Restsignal modelliert:

$$s(t) = \sum_{r=1}^{R} A_r(t) \cdot \cos[\theta_r(t)] + e(t)$$ (Gleichung aus „DAFX“ Seite 377)

ohne jedoch weitere Bedingungen an die einzelnen sinusoidalen Teiltöne zu stellen. Zur Analyse wird meist eine gefensterte FFT mit fester Block- und Hopsize verwendet. Aus den so gewonnenen Kurzzeitspektren werden dann Frequenz- und Amplitudenverläufe der einzelnen Teiltöne mit Hilfe eines so genannten „Peak Continuation“ Algorithmus extrahiert. Dabei wird nicht zwischen harmonischen und inharmonischen Teiltönen unterschieden. Mein Vorschlag zur Erweiterung dieses Modells enthält eine explizite Unterscheidung zwischen harmonischen und nichtharmonischen Teiltönen. Um bei obiger Notation aus dem DAFX Buch zu bleiben, wird die Summe der Sinuskomponenten also in zwei Teile aufgesplittet:

$$s(t) = \sum_{h=1}^{H} A_h(t) \cdot \cos[h \cdot \omega_0(t) \cdot t + \theta_h(t)] + \sum_{i=1}^{I} A_i(t) \cdot \cos[\theta_i(t)] + e(t)$$

wobei die erste Summe den harmonischen, und die zweite Summe den inharmonischen Anteil darstellt. Hierbei sind:

$A_h(t)$: zeitabhängige Amplitude des Teiltons h

$\omega_0(t)$: zeitabhängige Grundfrequenz des Klangs

$\theta_h(t)$: zeitabhängige Phasenverschiebung des Teiltons h

Es sollen psychoakustische Kriterien zur Definition der Harmonizität dienen:

1. die Grundfrequenz soll sich nicht schneller als mit z.B. 16 Hz ändern („Vibratos sind erlaubt – Frequenzmodulationen nicht“).
2. die Teiltöne sind an sich ganzzahlige Vielfache der Grundfrequenz ( $\omega_h = h \cdot \omega_0$ ). Das Verhältnis zwischen Grundfrequenz und Teiltonfrequenz darf sich aber in gewissen Grenzen (langsam) ändern – siehe 4.).
3. die Teiltonamplituden sollen sich nicht schneller als z.B. mit 16 Hz ändern („Tremolos sind erlaubt – Amplitudenmodulation nicht“).
4. die Phasenverschiebungen der Teiltöne gegenüber dem Grundton sollen sich auch nur langsam ändern (diese Bedingung erzwingt, dass die tatsächlichen Frequenzen der Teiltöne sich nicht allzu weit von den ganzzahligen Vielfachen entfernen können)

Um den harmonischen Anteil vom Rest abzutrennen, schlage ich folgenden Algorithmus vor:

1. Erkennung des Grundtons (Pitch Detection) z.B. mit Hilfe der Autokorrelation
2. bidirektionale Bandpassfilterung (bidirektional, um Nullphasen-Filter zu erhalten) mit Grundtonhöhe als Mittenfrequenz – führt zu Signal, dessen Nulldurchgänge den Nulldurchgängen der Grundfrequenz entsprechen.
3. Segmentierung des Ausgangssignals an den positiven Nulldurchgängen der Grundfrequenz. Das Reziproke der Segmentlänge ist die Grundfrequenz für die Dauer dieses Segments ( $\omega_0$ ).
4. Auf die einzelnen Segmente wird die DFT angewandt, und zwar mit einer DFT-Länge, die genau der Länge des Segments entspricht – also: kein zero-padding bis zur nächstgrößeren Potenz von 2. Damit fallen die DFT-bins genau mit den Frequenzen der zu erwartenden Harmonischen zusammen. Leakage wird somit vermieden, das macht Fensterung unnötig. Die FFT-Funktion von MatLab beispielsweise akzeptiert beliebige Längen und ist trotzdem noch effizient.
5. Aus den (pro Segment) Werten für Grundfrequenz ( $\omega_0$ ), Teiltonamplitude ( $A_h, h = 1...H$ ) und Teiltonphase ( $\theta_h, h = 1...H$ ) können nun (durch Interpolation) Hüllkurven für eine Oszillatorbank gewonnen werden – die Oszillatorbank implementiert genau die erste Summe in obiger Gleichung. Um aber die Kriterien für Harmonizität zu erfüllen werden die Hüllkurven jedoch zunächst bei z.B. 16 Hz tiefpassgefiltert – wieder bidirektional, um ein Filter mit Nullphasencharakteristik zu erhalten – dies erhält die Zeitstruktur der Hüllkurven. Für die Phase

tritt hier das Problem auf, dass diese immer in den Bereich zwischen –pi und pi abgebildet wird, oder – wenn sie ge-unwrapped wird – einen linearen Trend hat. Dieses Problem kann man jedoch umgehen, wenn man nicht Hüllkurven für Amplitude und Phase sondern für Real- und Imaginärteil erstellt.

Im Endeffekt wird die Nicht-Harmonizität einer Spektralkomponente also durch zu schnelle Modulation deren Amplitude und/oder Phase modelliert. Dies entspricht auch der Intuition, dass ein z.B. ein geräuschhafter Anteil von Frame zu Frame sehr unterschiedliche Amplituden in einem bestimmeten FFT-bin haben kann. Nichtharmonische aber dennoch sinusoidale Anteile ändern Ihre Phase von Frame zu Frame relativ schnell (die DFT-bins sind ja genau auf die Harmonischen „gestimmt"). Das Ausgangssignal der Oszillatorbank wird im Zeitbereich vom Originalsignal subtrahiert – aufgrund der Nullphasencharakteristik der verwendeten Filter funktioniert das. Das Residualsignal enthält jetzt noch den inharmonischen, geräuschhaften und transienten Anteil. Als nächstes wird der inharmonische Anteil abgetrennt. Im Prinzip können hier die bekannten Spectral Modeling Algorithmen angewandt werden. Möglicherweise könnte auch eine modifizierte Version der linearen Prädiktion gut funktionieren:

$$\hat{x}[n] = \sum_{k=m}^{K} w_k x[n-k]$$

…das Signal soll also nicht aus allen bekannten vergangenen Abtastwerten vorhergesagt werden, sondern nur aus solchen, die mindestens $m$ samples zurückliegen (normalerweise startet $k$ bei 1). Dadurch werden nur langreichweitige Autokorrelationen (die wahrscheinlich auf periodische Anteile zurückzuführen sind) für die Vorhersage verwendet. $m$ könnte z.B. eine Signalperiode (in samples) sein. Das prediction-error-signal enthält jetzt noch rauschhaften und transienten Anteil. Der vorhersagbare Anteil des Rauschens wird durch ganz normale lineare Prädiktion bestimmt:

$$\hat{x}[n] = \sum_{k=1}^{K} w_k x[n-k]$$

wobei die $w_k$ z.B. durch den LMS-Algorithums bestimmt werden können. Das prediction-error-signal sollte jetzt nur noch white noise und der Transient sein (diese könnte man jetzt auch noch im Zeitbereich voneinander trennen (z.B. einfach schneiden)).

Was ist neu gegenüber bisherigen spektralen Modellen:

- willkürliche Segmentierung des Signals aufgrund von FFT- und Hopsize wird ersetzt durch eine signalangepasste Segmentierung an den positiven Nulldurchgängen der Grundfrequenz.
- dadurch fallen die Analyse-Frequenzen der FFT-bins genau mit den zu erwartenden Harmonischen zusammen. Dies macht Fensterung unnötig.
- Harmonische und inharmonische Teiltöne werden anhand eines psychoakustischen Kriteriums getrennt (Wahrnehmung von langsamen Modulationen in Zeitbereich, Wahrnehmung von schnellen Modulationen im Frequenzbereich).
- Erkennung von inharmonischen Teiltönen (evtl.) über modifizierte lineare Prädiktion
- Sämtliche reynthetisierten Teilsignale werden phasentreu resynthetisiert – Addition im Zeitbereich rekonstruiert das Originalsignal exakt.

Ich habe schon einige „quick and dirty" Simulationen in MatLab gemacht, und das Konzept scheint generell zu funktionieren. Ich würde dies gerne im Rahmen meiner Magisterarbeit ausführlicher ausarbeiten.

Robin Schmidt