

## Diphone Studio

### Teil 3 — Analyse resonanter Modelle und Chant-Synthese

Hans Tutschku  
tutschku@ircam.fr

Der Artikel bezieht eine Sammlung von Beispielklängen ein, die auf folgender web-adresse angehört werden können (Quicktime 4.0 erforderlich):

<http://www.multimania.com/hanstutschku/diphone/diphone.htm>

Der dritte und letzte Teil der Artikelserie über Diphone wird nun in komprimierter Form den sehr umfangreichen Bereich der Chantsynthese und die Analyse resonanter Modelle beschreiben. (Auch ohne die Kenntnis des Programms Diphone, ist es möglich, sehr interessante Klänge mit den resonanten Modellen zu erzeugen.)

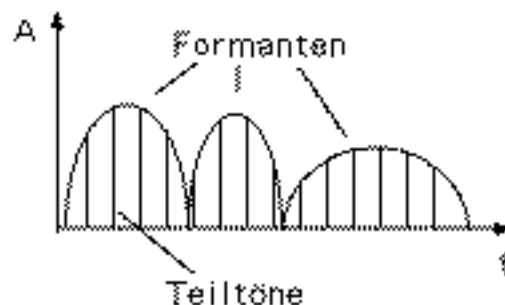
Die Chantsynthese ist in zwei Funktionsbereiche zu teilen, wobei jeder von ihnen jeweils mit der additiven bzw. subtraktiven Synthese verglichen werden kann.

Im ersten Fall werden Spektren durch Addition von Formanten generiert, im zweiten Fall werden die Formantdefinitionen genutzt, um vorhandenes Klangmaterial zu filtern.

Die Parameter der Formanten können in beiden Fällen entweder frei bestimmt oder durch Analyse vorhandener Klänge ermittelt werden.

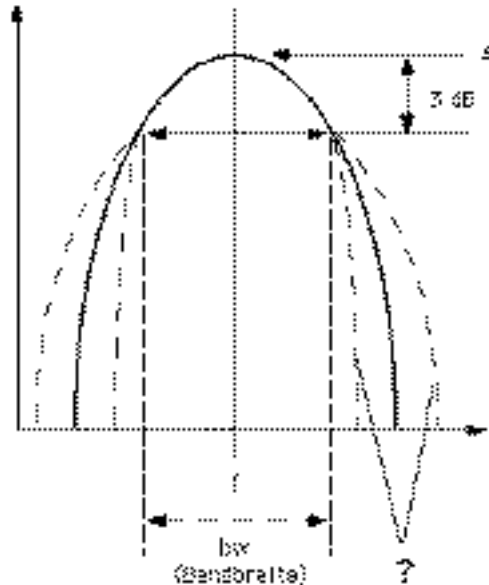
#### Formanten

Die landläufige Definition eines Formanten beschreibt einen Bereich im Spektrum, der mehr Energie aufweist, als seine benachbarten Frequenzbereiche. Es ist also eine Gruppe von spektralen Teiltönen, die im Spektrum hervortritt.



Dabei wird ein Formant mit drei Parametern beschrieben:

- zentrale Frequenz ( $f$ )
- Amplitude an der zentralen Frequenz ( $A$ )
- Bandbreite (Öffnung des Formanten an der Stelle, wo die Amplitude 3 dB leiser ist, als die maximale Amplitude - die Bandbreite wird demnach in Hz angegeben)



Wir sehen hieran, daß die Definition eines Formanten eigentlich sehr ungenau ist, da mit diesen drei Parametern nicht zu definieren ist, wie sich die spektrale Öffnung in den leiseren Bereichen verhält.

Analyse resonanter Modelle:

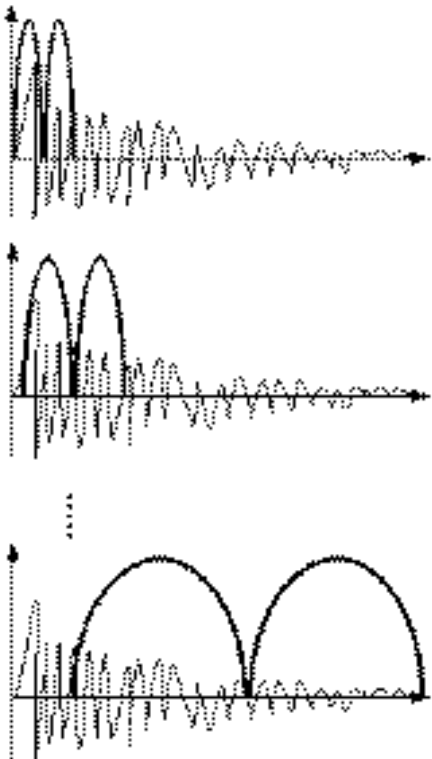
Unter resonanten Modellen versteht man Vibrationsmodelle, die durch eine kurze Energiezufuhr angeregt werden und ausklingen. Dabei ändern sich die Frequenzen während des Ausklingens nicht. Zu diesen Modellen gehören zum Beispiel geschlagene bzw. gezupfte Klangkörper.

Die Analyse solcher Modelle versucht nun, ausgehend von einer Klangaufnahme, die Resonanzqualitäten zu bestimmen und eine Serie von Formanten mit ihren Frequenzen, Amplituden und Bandbreiten zu ermitteln. Dies ist eine feststehende Serie von Formanten, die sich in der Zeit nicht ändert.

Die Analyse, die in mehreren Durchgängen errechnet wird, muß dabei sowohl den Moment des Attacks, als auch die Resonanz beschreiben. Deshalb wird die Analyse mit zwei benachbarten

Fenstern durchgeführt, wobei das linke direkt den Attack analysiert und das rechte die anschließende Resonanz.

Beide Klangausschnitte (Fenster) werden mit der FFT analysiert (das Prinzip dieser Analyse wurde im ersten Teil über Diphone beschrieben) und die ermittelten Frequenzen (spektrale "peaks") beider Fenster miteinander verglichen.



Wenn eine Resonanzfrequenz in beiden Analysefenstern auftritt, geht man davon aus, daß sie Spektrumsbestandteil des analysierten Modells ist. Durch den Vergleich der Amplituden kann die Ausschwingzeit der jeweiligen Frequenz ermittelt werden.

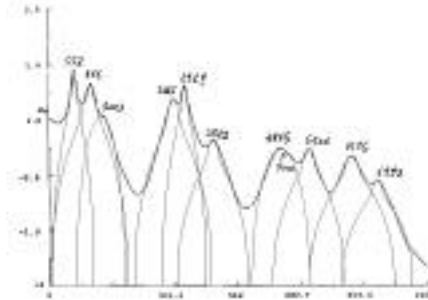
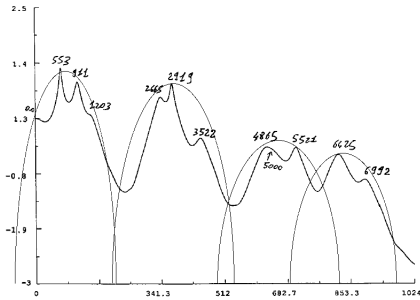
Dieser erste Durchgang erzeugt ein sogenanntes "elementares Modell".

Im den darauffolgenden Durchgängen vergrößert man die Fenster und verschiebt sie etwas vom Attack in Richtung Resonanz.

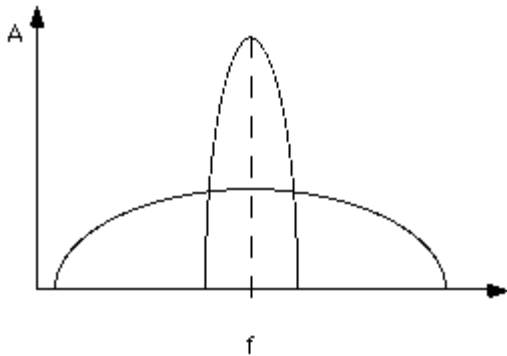
Jeweils am Ende eines Durchgangs wird das Modell des aktuellen Durchgangs mit dem vorherigen Modell verglichen. Sind Frequenzen übereinstimmend (oder sehr nah aneinander), werden sie in das daraus resultierende "angereicherte Modell" übernommen. Ihre Parameter für Amplitude und Bandbreite werden nach Vergleich der alten und neuen Werte angepaßt. Treten bestimmte Resonanzen nicht mehr auf, oder kommen neue hinzu, verbleiben sie im Modell bzw. werden hinzugefügt.

Da die Fenstergröße (und deren FFT) der ersten Analysedurchgänge sehr klein ist, bekommt man keine sehr genaue Frequenzabbildung für den Moment des Attacks. Daraus resultieren sehr breite Formanten, die sehr viel Energie durchlassen. Dies entspricht dem natürlichen Verhalten der Modelle, die im Moment des Attacks ein sehr reiches Spektrum haben und mit allen "Bestandteilen" des Vibrationskörpers schwingen. Die Energie des Attacks verklingt sehr schnell und im Folgenden bleiben lediglich die Resonanzen des Modells übrig.

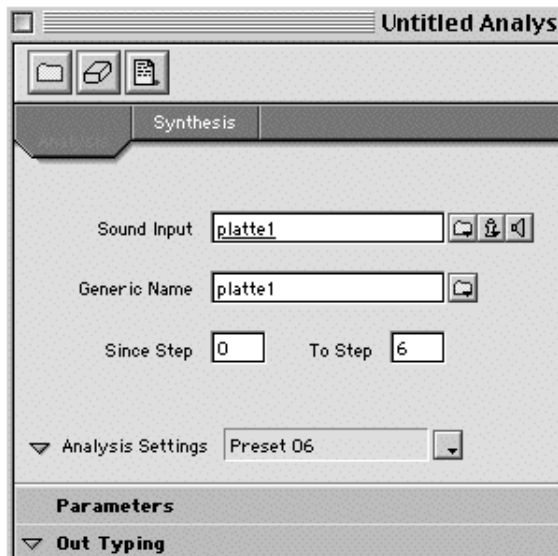
Durch die Vergrößerung der Fenster in jedem Analysedurchgang können nach und nach auch genauere Frequenzen unterschieden und damit feinere Abstufungen der Formanten erreicht werden.



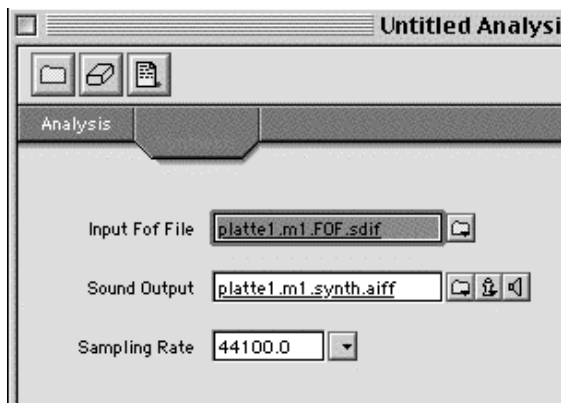
Die Analyse eines gegebenen Klanges mit kleinen Fenstern ergibt wenige, breite Formanten. Die Analyse mit größeren Fenstern erlaubt die Ermittlung genauerer Formantdaten. Breite Formanten lassen die Energie sehr schnell durch und resonieren kaum, enge Formanten "benötigen mehr Zeit", um eine gegebene Energiemenge passieren zu lassen und resonieren demnach länger. Dies ist also eine "simple" Erklärung für die Beziehung von Bandbreite und Resonanzzeit: je enger der Formant, desto länger klingt er. Bei den angereicherten Modellen kann es durchaus vorkommen, daß eine gleiche Frequenz mit mehreren Formanten abgebildet wird, die unterschiedliche Amplituden und Bandbreiten haben.



Dabei modelliert der breitere Formant den Moment des Attacks und der schmalere die darauf folgende Resonanz.



Alle Parameter für den Vergleich der elementaren und angereicherten Modelle können im einzelnen geregelt werden. (pro Analysedurchgang ca. 25 Parameter). Einen relativ einfachen Einstieg erhält man, wenn man die vorbereiteten Presets verwendet. Ich bevorzuge Preset 6, das mit 6 Durchläufen sehr gute Ergebnisse für eine breite Palette von Ausgangsklängen erzeugt.



Nach Beendigung der 6 Analysedurchgänge findet man im Ordner "ImpExport:Chant" die 6 SDIF-Files, die den angereicherten Modellen der Einzeldurchgänge entsprechen. Auf der Funktionsseite "Synthesis" von ResAn können diese Einzeltappen resynthetisiert werden. (Beispielklänge: Original: crotale Resynthese: crotale.m1, crotale.m2, crotale.m6). Bei diesem Beispiel hört man, daß mit jedem

Analysedurchgang das Formantspektrum angereichert wird.

### Kreation "ungewöhnlicher" resonanter Modelle

Das beschriebene Analyseverfahren läßt sich nicht nur für echte Resonanzklänge verwenden. Auch längere Klangsequenzen können analysiert werden, die dem eigentlichen Modell nicht entsprechen. Dies ergibt dann sehr überraschende Resonanzklänge, die der Summe der hervortretensten Formanten des analysierten Klanges entsprechen. Auch hierzu habe ich einige meiner Lieblings-Beispielklänge auf der Webseite.

Original: huhn, Resynthese: huhn.m1, huhn.m2, huhn.m5

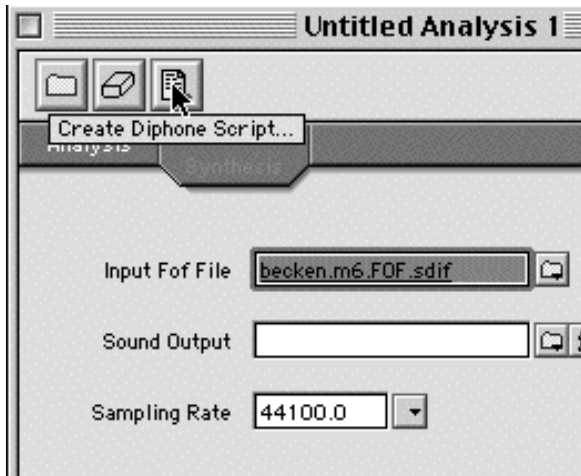
Original: bulg-2 Resynthese: bulg-2.m5

Original: bulg-frau Resynthese: bulg-frau.m5

Original: sundanesie Resynthese: sundanesie.m5

## "Dictionary"

Wie schon im Teil über die additive Synthese beschrieben, müssen die Analysedaten in sogenannten Bibliotheken (dictionaries) gespeichert werden, die jeweils Instrumente und Segmente enthalten, aus denen neue Sequenzen erstellt werden können.



Ist man mit einer der Resynthesen in ResAn zufrieden, kann man ein Skript generieren, das zur Erzeugung des "dictionary" benötigt wird. (Es sind die selben Schritte, wie für AddAn).



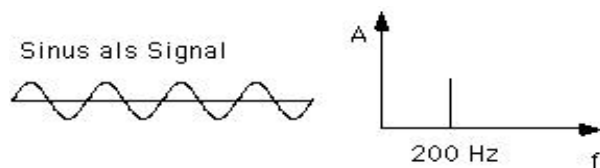
In Diphone generiert man dann das "dictionary", das in diesem Fall dem Modell Chant entspricht.



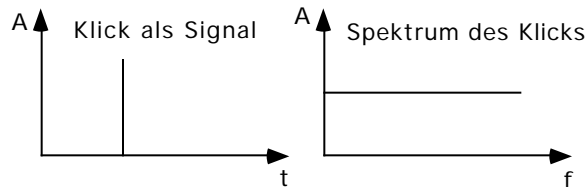
Da Formantdefinitionen sich im zeitlichen Verlauf der Segmente nicht ändern, ist keine Segmentierung des Ausgangsklanges in Einzelsegmente nötig. Es wird ein einziges Segment pro Klang erzeugt, das alle Formantdaten enthält.

## FOF-Synthese

Bei der FOF-Synthese werden die analysierten Formantdaten zur Generierung von Formanten auf rein synthetische Art verwendet. Ein FOF (*forme d'onde formantique*) ist eine Sinusschwingung mit einer Amplitudenhüllkurve. Auch hier werden wir wieder sehen, daß es eine direkte Beziehung zwischen Zeitdauer eines Klangereignisses und seiner spektralen Definition (Frequenz) gibt. Eine ausführlichere Beschreibung hierzu finden Sie im ersten Teil dieses Artikels über Diphone - ich wiederhole hier nur noch einmal die Grundprinzipien.

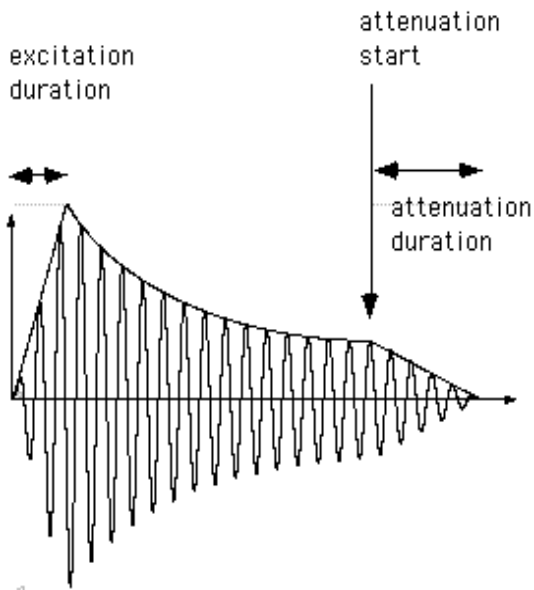


Theoretisch hat eine Sinusschwingung die spektrale Definition einer einzelnen Frequenz.

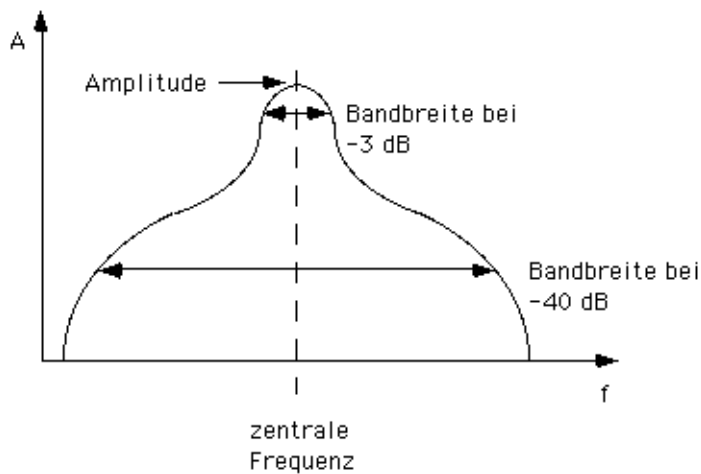


Der kürzeste Klang, ein Klick, hat ein völlig flaches Spektrum. Eine Frequenzdefinition ist nicht möglich, da der Klang keine Periode hat und nur aus einem einzelnen Sample besteht.

Die oben genannte spektrale Definition für einen Sinus gilt aber nur, wenn er schon immer existierte und auch immer weiter dauern würde. In dem Moment, wo man ihn beginnen läßt und beendet, multipliziert man das Signal mit einer Amplitudenhüllkurve. Diese Hüllkurve kann ebenfalls als Signal verstanden werden, das selbst ein eigenes Spektrum hat. Durch die Multiplikation beider Signale (Sinus \* Hüllkurve) kommt es zu einer Convolution beider Spektren und die simple Definition einer einzelnen Frequenz des Sinus gilt nicht mehr. Dies macht man sich jetzt zunutze und kann über die Dauer des Sinus seine spektrale Definition kontrollieren.

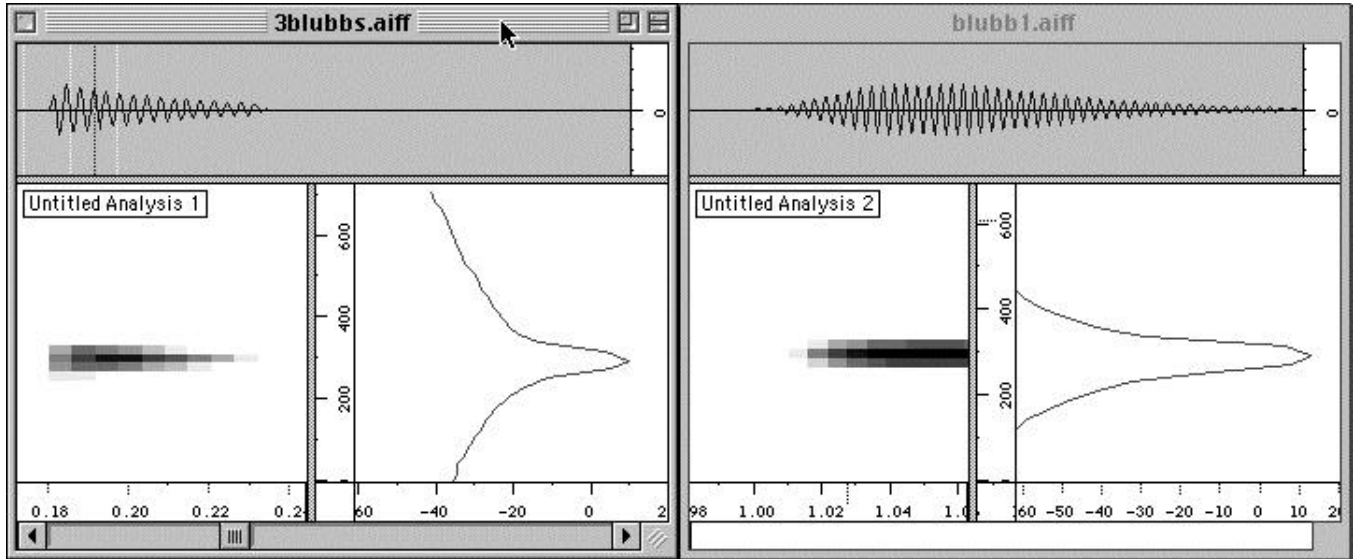


Die Sinusschwingung eines FOF wird dabei mit folgender Hüllkurve gesteuert: lineares Einschwingen mit regelbarer Zeit "excitation duration"; sofort beginnendes exponentielles Ausschwingen; ab "attenuation start" lineares Ausschwingen in der Zeit "attenuation duration".



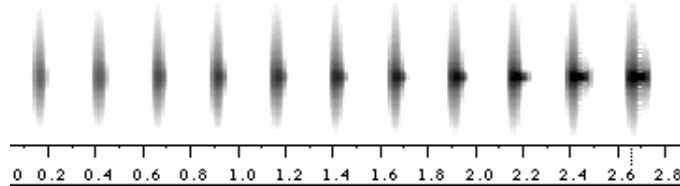
Dieses Signal ergibt in der spektralen Darstellung einen Formanten. Seine spektrale Definition (Bandbreite) hängt von der Länge des Signals ab. Außerdem ergibt sich noch ein zusätzlicher Faktor. Es ist die Bandbreite bei -40 dB, die von der Einschwingzeit abhängt und wiederum reziprok zu ihr ist: je länger die "excitation duration", desto schmaler die Bandbreite bei -40 dB.

Spielen wir einen solchen FOF, erhalten wir einen kurzen "Blubb", mit dem beschriebenen Formantspektrum.

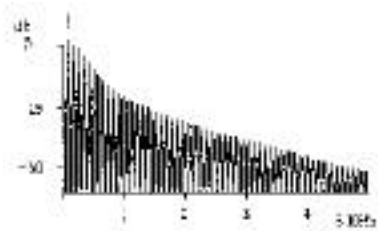


In dieser Abbildung sehen wir das Signal mit seiner Sonagrammdarstellung und den resultierenden Formanten. Die Zentralfrequenz des Formanten entspricht hierbei der Frequenz der gespielten Sinusschwingung. Für die rechte Darstellung habe ich die "excitation duration" erhöht und man sieht deutlich, daß die Bandbreite bei -40 dB wesentlich schmaler ist. (-40 dB relativ zur maximalen Amplitude des Formanten)

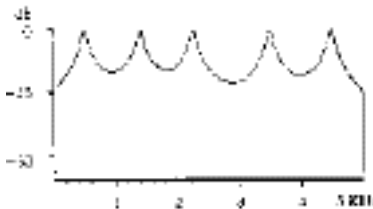
Verringert man die Bandbreite des Formanten, wird seine Resonanzzeit länger. (Beispielklang "bandwidth")



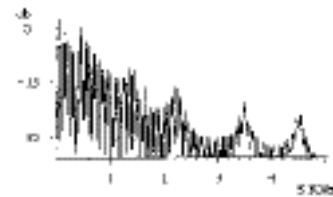
Chant wurde Ende der 70er Jahre am IRCAM entwickelt und war in erster Linie zur Synthese der menschlichen Gesangsstimme gedacht. Die Stimmbänder erzeugen bei einer gesungenen Note ein energiereiches harmonisches Spektrum, das dann durch den Vokaltrakt gefiltert wird.



energiereiches harmonisches Spektrum

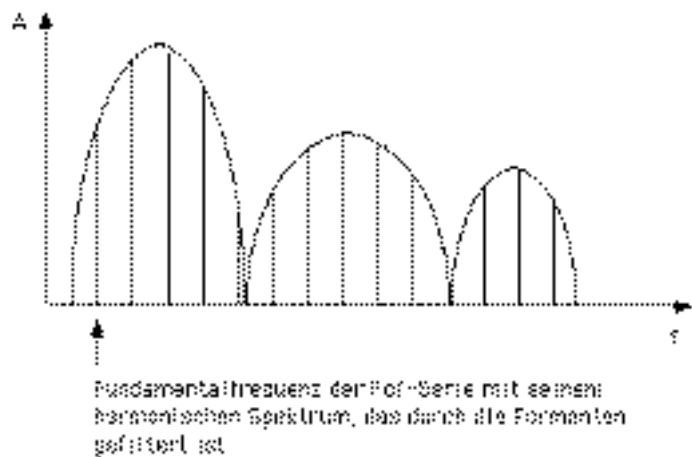
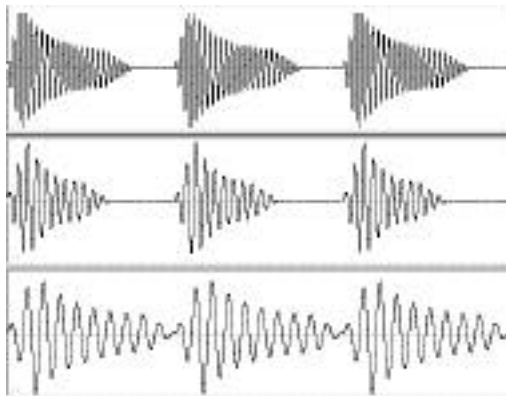


Filterung des Vokaltraktes



resultierendes Spektrum

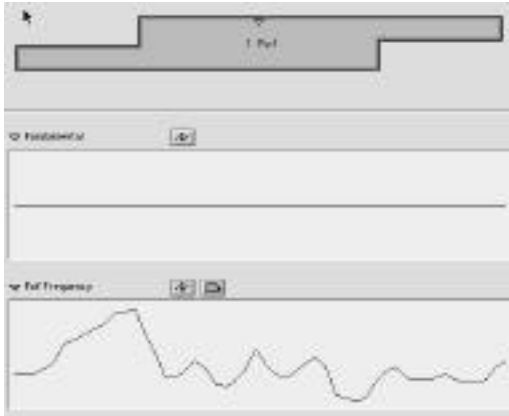
Möchte man mehrere Formanten mit den FOFs erzeugen, benötigt man parallel mehrere Sinussignale, die zeitgleich gestartet werden - also die gleiche Fundamentalfrequenz haben. Jeder Sinus-Strom erzeugt einen Formanten.



Die Fundamentalfrequenz (für alle FOFs gleich) ist die Basis eines harmonischen Spektrums, das durch die Formanten gefiltert wird. Jeder FOFstrom kann dabei seine eigene Hüllkurve haben, wodurch unterschiedliche Bandbreiten der Formanten entstehen. Die Amplituden der Formanten ergeben sich aus den Amplituden der Sinussignale.



Im Diphone-Ordner "dico&seq" befindet sich das Chant-dictionary "Chant-Base.dico", das ein Instrument mit 16 "leeren" Segmenten enthält. Aus diesen "default"-Segmenten kann man Sequenzen mit 1-16 Formanten von Grund auf konstruieren.

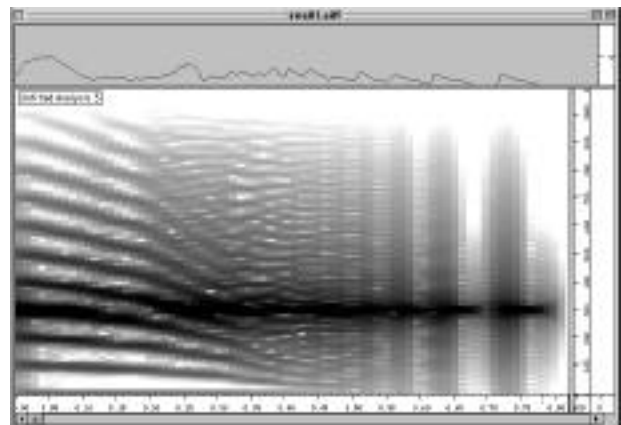
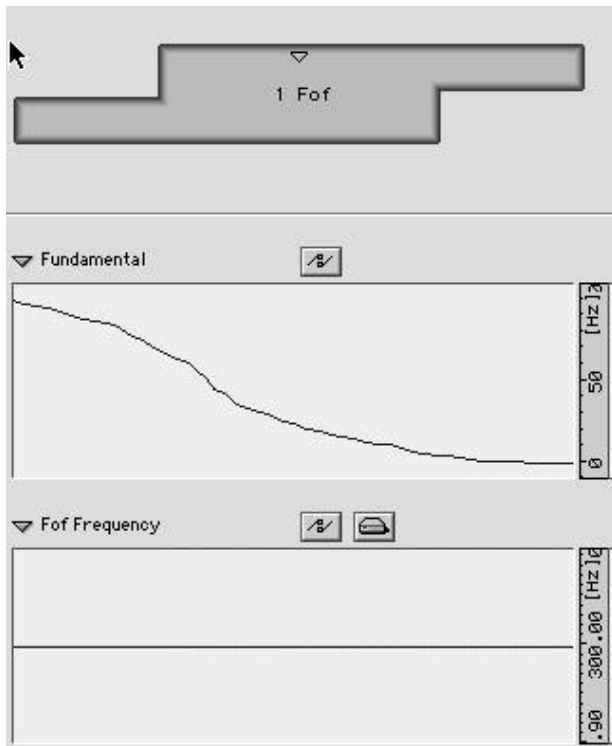


Alle Parameter sind, wie auch im Bereich der additiven Synthese von Diphone als aufklappbare Editoren unter der Sequenz zu erreichen. Auch alle Bearbeitungen der Segmente innerhalb der Sequenz mit Artikulationen, Interpolationsbereichen, Composit-Segmenten treffen genauso auf den Bereich der Chant-Synthese zu.

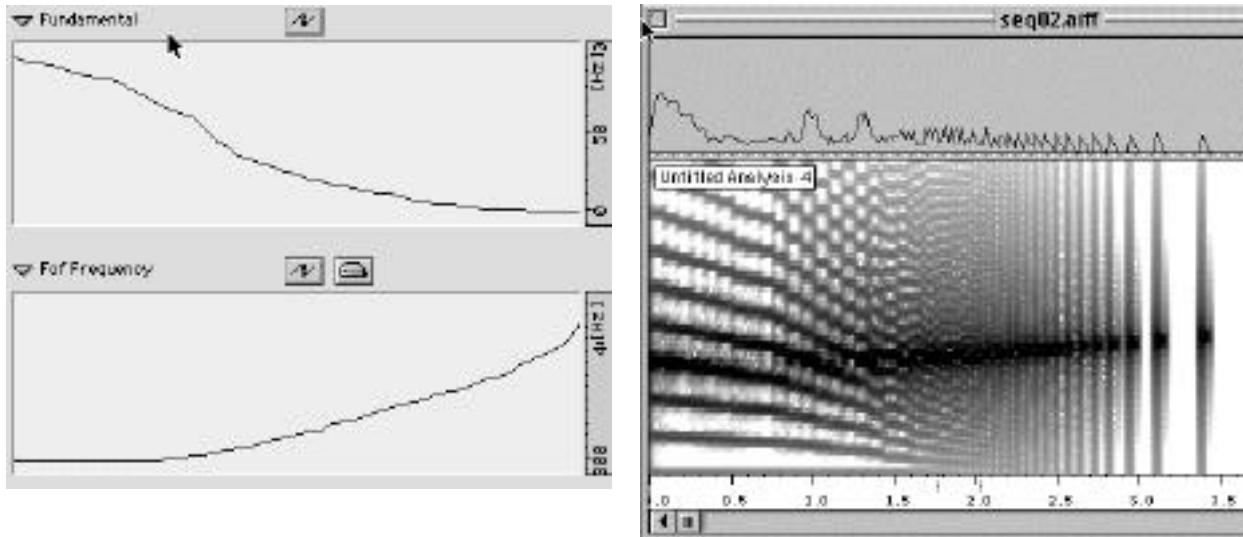
Ich gebe im Folgenden eine kleine Serie von Beispielklängen, die die wichtigsten Parameter beschreibt:

Beispiel: "chant-seq01" ist ein Segment, dessen zentrale Frequenz bei 300 Hz liegt. Lediglich die Fundamentalfrequenz, das heißt die Frequenz mit der die einzelnen FOFs gestartet werden, fällt von 100 Hz bis 0 Hz. Unterhalb 16 Hz hören wir den Rhythmus der einzelnen FOF-Einsätze.

Die spektrale Darstellung zeigt deutlich das harmonische Spektrum, das auf 100 Hz basiert und ein Glissando abwärts beschreibt. Die zentrale Frequenz des Formanten um 300 Hz ist ebenfalls zu sehen.



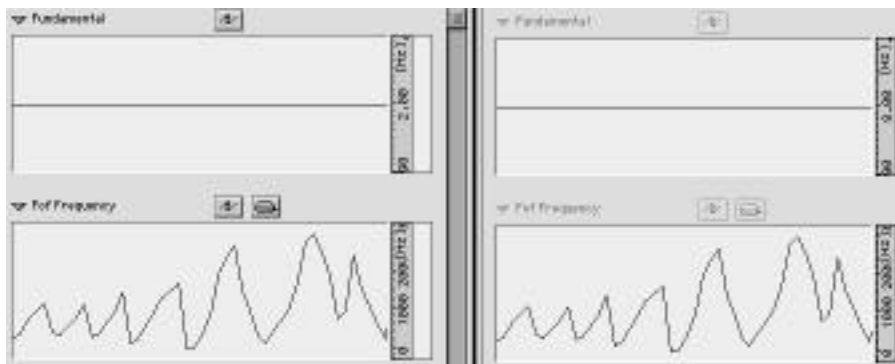
In Beispiel: "chant-seq02" fällt die Fundamentalfrequenz wie im ersten Beispiel. Die Frequenz des Formanten steigt von 300 nach 400 Hz.

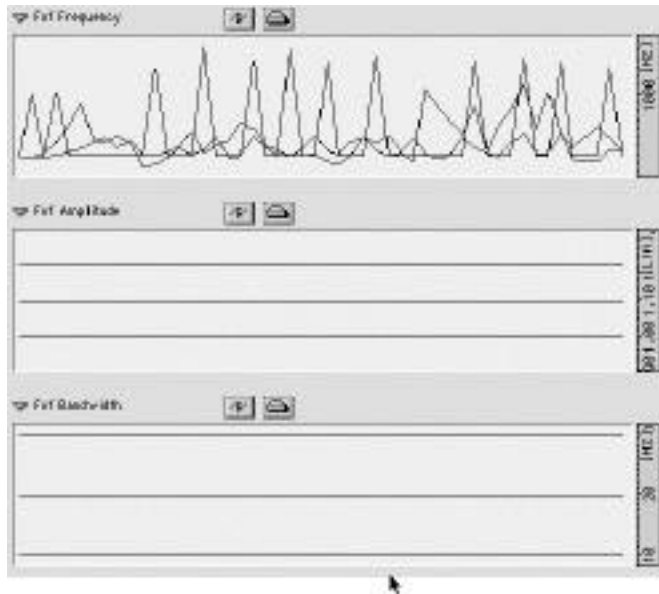


Parameteränderungen für FOFs sind nicht kontinuierlich. Dies bedeutet, daß beim Start eines FOFs die aktuellen Parameter angewendet werden, die zu diesem Zeitpunkt gültig sind. Während seiner Dauer bleiben die Werte konstant.

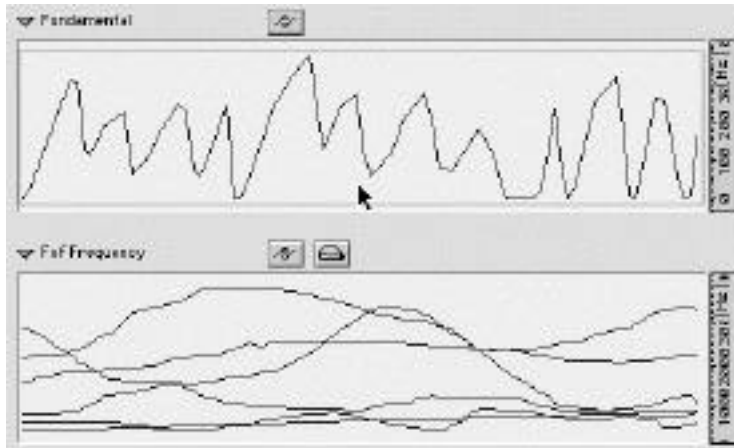
"chant-seq03" hat den gleichen Frequenzverlauf wie "chant-seq04". Da die Fundamentalfrequenz in "chant-seq03" aber nur bei 2 Hz liegt, werden aus der Frequenzkurve nur alle 500 ms die aktuellen Werte ausgelesen.

"chant-seq04" hat eine Fundamentalfrequenz von 8 Hz.





"chant-seq05" ist nun eine Sequenz mit drei Formanten, deren Frequenzen sich ständig ändern. Die drei Amplituden und Bandbreiten bleiben konstant.



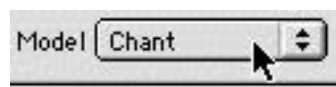
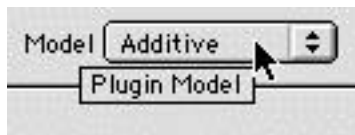
In "chant-seq06" variiert die Fundamentalfrequenz zwischen 0 und 500 Hz. Da es sich um ein Segment mit 6 FOFs handelt, werden bei einer Fundamentalfrequenz von 500 Hz 3000 FOFs gestartet. Dies ergibt sehr reiche Klänge, die Rechenzeit erhöht sich aber dementsprechend.

### Verwendung von additiven Analysen zur Chant-Resynthese

Reduziert man die Bandbreite eines Formanten gegen Unendlich (0.001 Hz), bekommt man praktisch die Abbildung eines einzelnen Teiltones.

Die folgende Serie von Beispielklängen vergleicht die additive Synthese und die Chant-FOF-Synthese. "java.additive" ist die additive Resynthese einer Sequenz. Die selben Analysedaten kann man auch als Formantdefinitionen verwenden. Dabei werden die Frequenzdaten der Teiltöne zu den zentralen Frequenzen der Formanten und die Amplitudendaten der Teiltöne zu den Amplituden der Formanten. Lediglich die Bandbreite der Formanten muß "erfunden" werden.

Wenn man mit einer Sequenz arbeitet, die additive Segmente enthält, sieht man im oberen Bereich des Fensters das Modell "Additive".



Schaltet man es auf "Chant" um, werden automatisch die Frequenz- und Amplituden-Parameter für die Formanten übernommen. "java-as-chantfof" ist die Chant-FOF-Synthese der selben Sequenz mit einer sehr Bandbreite von 1 Hz. Für den Klang "java-as-chantfof-bw1000" wurde die Bandbreite auf 1000 Hz erhöht.

Läßt man die Fundamentalfrequenz nicht mehr der originalen Frequenz folgen, erhält man nur noch die Veränderung der Formanten, was zu sehr interessanten Klängen führt. "java-as-f0-500-bw10" ist die gleiche Sequenz mit einer Bandbreite von 10Hz, lediglich die Fundamentalfrequenz wurde auf 500 Hz gesetzt und bleibt konstant.

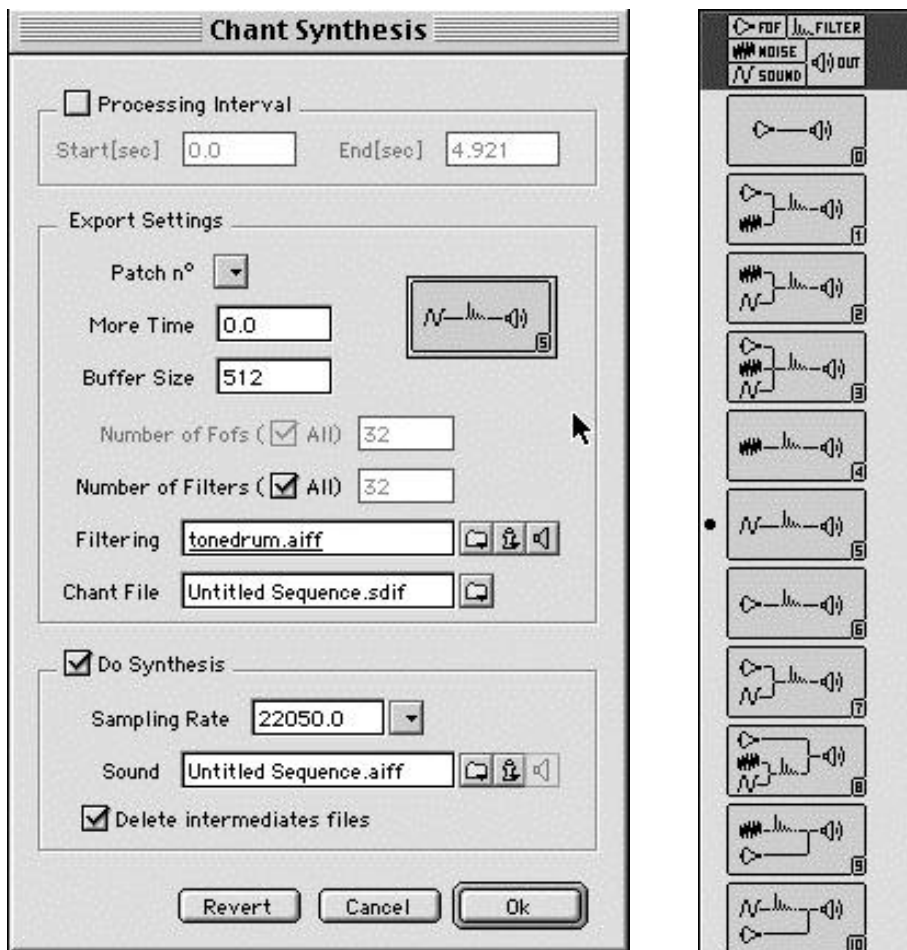
## Filter

Die Formantdefinitionen können nicht nur zum Steuern der FOFs verwendet, sondern auch als Filter benutzt werden. Filtern kann man entweder weißes Rauschen (es ist ein interner Rauschgenerator vorhanden) oder ein externes Soundfile.

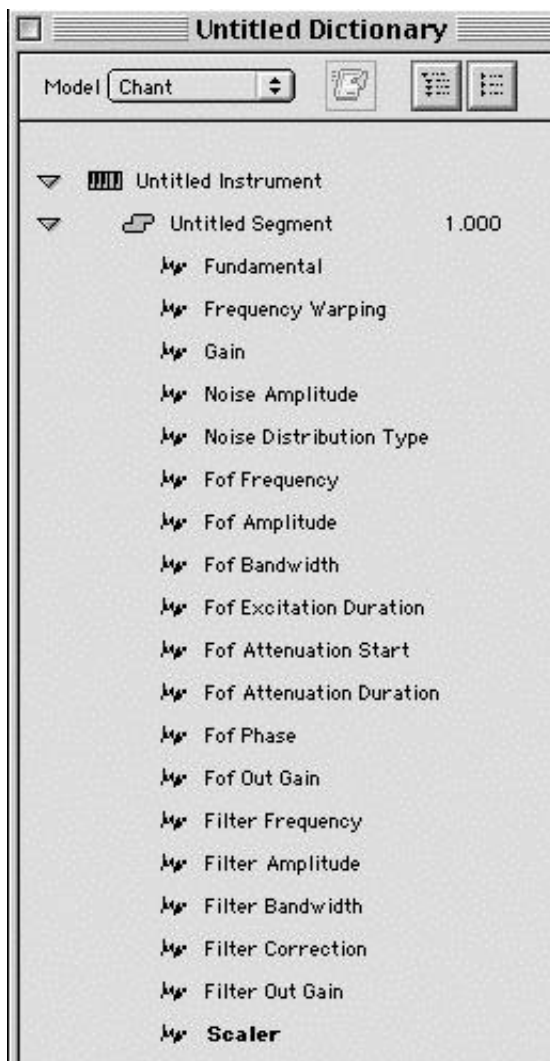
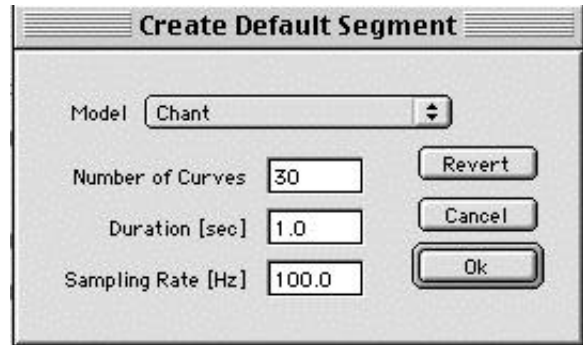
"java-as-filter-noise" ist das Ergebnis der Filterung von weißem Rauschen mit den Formantdaten der vorangegangenen Sequenz.

Das Soundfile "tonedrum" wurde mit den selben Formanten gefiltert, wie das Beispiel "chant-seq06". Im Unterschied zu den FOFs sind die Parameteränderungen bei den Filtern kontinuierlich, d.h. es ergeben sich Glissandi der Formanten. Das Ergebnis ist "tonedrum.filtered".

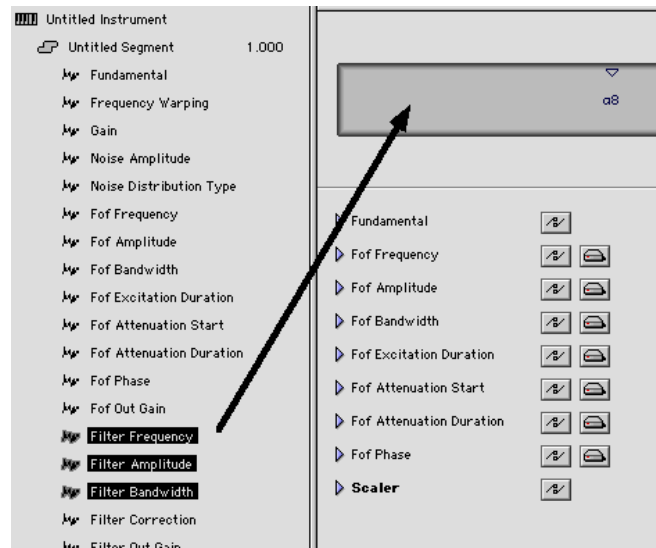
Während der Resynthese muß festgelegt werden, mit welchem "Patch" Chant die Resynthese durchführt. Bisher haben wir lediglich den ersten Patch verwendet, der die FOF-Synthese direkt mit dem "Lautsprecher" verbindet. Es können alle möglichen Kombinationen aus FOFs, Rausch-Filterung und Soundfile-Filterung ausgewählt werden.



Um eine Sequenz als Formant-Filter verwendet zu können, müssen die Segmente die Parameter für die Filter enthalten. Ist dies nicht der Fall, kann man ein sogenanntes "Default Segment" erzeugen, das alle Chant-Parameter enthält. Wichtig ist dabei, daß man das Modell Chant für das zu erzeugende Default Segment auswählt.



Im daraus resultierenden neuen Dictionary findet man ein Segment mit allen Chant-Parametern. Die benötigten können ausgewählt und in das gewünschte Segment gezogen werden.



Auf diese Weise kann man Segmente mit allen erdenklichen Parameterkombinationen erstellen. Die benötigten Parameter hängen selbstverständlich vom ausgewählten Chant-Patch ab.

### **Schlußbemerkung**

Ich hoffe, daß ich mit diesen drei Artikeln einen Einblick in die Arbeit mit Diphone geben konnte - ein Programm, das ich für sehr flexibel halte und das zum Teil sehr ungewöhnliche Klangbearbeitungsmöglichkeiten bietet. Es ist immer noch in Entwicklung und die (internen) Versionen wechseln manchmal innerhalb von Wochen. Ich denke aber, daß mit den jetzt zur Verfügung stehenden Möglichkeiten der Resonanzmodell- und der additiven Analyse und den Bearbeitungsmöglichkeiten in Diphone schon eine breite Palette zu erschließen ist. Die aktuelle Programmierung konzentriert sich auf die Integration neuer Synthesealgorithmen und eine bessere Editierung der multi-BPFs.