



Universität der Künste Berlin

Sommersemester 2008  
Zentralinstitut für Weiterbildung  
Sound Studies  
Seminar: Geschichte der Klangkunst  
Dozent: Dr. M. Supper

# Sprechende Instrumente

Die Entwicklung der **Sprachsynthese** und  
ihre Umsetzung in der zeitgenössischen **Musik**

---

Seminararbeit  
verfasst von  
Kathrin Grenzdörffer  
Matrikelnr.: 358425  
Bauhofstr. 5  
10117 Berlin  
k.grenzdorffer@gmx.de

Datum der Abgabe: 08.10.2008

## **Sprechende Instrumente - Inhalt**

1 Einleitung.....	2
1.1 Begriffliches.....	2
1.2 Historische Überlegungen.....	3
2 Grundlagen akustischer Phonologie.....	5
2.1 Unterschied zur musikalischen Akustik.....	5
2.2 Akustische Eigenschaften der Laute.....	6
3 Sprachsynthese.....	7
3.1 Voraussetzungen.....	7
3.2 Von Stimmgabeln zum Vocoder.....	8
3.3 Linear Predictive Coding.....	11
4 Umsetzung in zeitgenössischen Kompositionen.....	12
4.1 Paul Lansky: Idle Chatter Junior (1999).....	12
4.2 Klarenz Barlow: Im Januar am Nil (1984).....	14
5 Schlussfolgerungen und Ausblicke.....	15
6 Literaturverzeichnis.....	17
6.1 Literatur.....	17
6.2 Online-Quellen.....	17
6.3 Klangbeispiele.....	18
7 Abbildungsverzeichnis.....	18

*It seemed simple:  
any sound we could describe,  
we could create.<sup>1</sup>  
Paul Lansky*

## **1 Einleitung**

### **1.1 Begriffliches**

In der Musikwissenschaft existiert das Genre *Sprachkomposition*.<sup>2</sup> Hierzu zählen Stücke, die nicht darauf angelegt sind, Sprache zu vertonen, sondern vielmehr *Klangfarbenkompositionen* sind, die das phonetische Material der Lautäußerung zum Gegenstand haben. Ein semantischer Gehalt geht ihnen weitestgehend ab.

Um das Thema meiner Arbeit besser eingrenzen zu können, schlage ich vor, eine Unterscheidung zwischen Sprachkompositionen zu treffen, die sich an dem Medium orientiert, das zur Lautproduktion verwendet wurde. Die

---

1 Lansky: *The Importance of Being Digital*.

[www.music.princeton.edu/~paul/lansky\\_beingdigital.htm](http://www.music.princeton.edu/~paul/lansky_beingdigital.htm) am 04.10.08

2 vgl. Beiheft CD *Sprachkomposition*. Andere Klassifizierungen hingegen zählen auch Lautgedichte mit darunter, z.B. Lentz: *Lautpoesie*.

Klassifizierung ist nicht als starr zu betrachten, denn Mischformen sind denkbar. Doch im Wesentlichen führt der Einsatz einer Gesangstechnik zu anderen Ergebnissen als der Einsatz eines Computers.

### Sprachkomposition

mit Einsatz des Vokaltraktes

a) *physiologisch* – als Gesangstechnik in der Neuen Musik

ohne Einsatz des Vokaltraktes

a) *mechanisch* – durch Einsatz obertonarmer, nicht-elektrischer Klangerzeuger

b) *elektrisch* – durch Apparatur von Generatoren und Filtern

c) *computer-generiert* – durch digitale Resynthese

Meine Arbeit soll sich mit Sprachkompositionen beschäftigen, bei denen die Laute *ohne den Vokaltrakt* synthetisiert werden. Dabei ist es unerheblich, ob die Quelle aufgezeichnete oder künstliche Laute sind. Ausschlaggebend für derartige Kompositionen ist, dass die Sprachelemente vom Medium signifikant modifiziert oder generiert werden.

Um Sprachlaute nachzubilden, ist es von großer Bedeutung, ihre akustischen Eigenschaften zu kennen und ein passendes *Kodierungs-verfahren* entwickelt zu haben, damit die Synthese erfolgreich ist. Deshalb halte ich es für wichtig, in dieser Arbeit einen kurzen Einblick in die Sprachsynthese zu geben.

## 1.2 Historische Überlegungen

Auch historisch betrachtet scheint diese Unterscheidung sinnvoll zu sein. So haben die neueren Vokalkompositionen, angefangen bei Kurt Schwitters' *Ursonate*, eher literaturgeschichtliche Wurzeln im Dadaismus und Futurismus<sup>3</sup>. Dagegen stellen sich Kompositionen der zweiten Art eher in die Tradition einer technischen Akustikgeschichte, die ausgehend von Helmholtz' *Lehre von den*

<sup>3</sup> vgl. Lentz: *Lautpoesie*.

*Tonempfindungen*, in die Entwicklung der elektronischen Musik mündet. Zwar gab es frühere Versuche, Sprache synthetisch zu erzeugen. Doch bahnbrechend in seiner Popularität und damit einflussreich für die Klangsynthese war wohl Werner Meyer-Epplers *Elektrische Klangerzeugung. Elektronische Musik und synthetische Sprache* von 1949 – ein Standardwerk, das minutiös die Grundlagen der Akustik und ihre praktische Anwendung anhand der frühen Klangerzeuger beschreibt.

Ab dem Ende der 1940er Jahre ermöglichte der Einsatz von Tonbändern und die damit verbundene Schnitttechnik die Demontage aufgenommenen Materials. Sprachaufnahmen konnten so beinahe beliebig geschnitten und wieder zusammengefügt werden. Wurden zusätzlich Generatoren und Filter eingesetzt, ergab sich die Möglichkeit, das Material zusätzlich zu erweitern<sup>4</sup>. Als Computer immer leistungsfähiger wurden, wurden auch sie in den Kompositionsprozess integriert.

Zentral für die synthetische Sprache sind darüber hinaus phonetische Grundlagen, denn aus den umgekehrten Analyseverfahren der Phonetik schöpfen die elektronischen Kompositionen letztlich ihr Material. Historisch betrachtet bestand bereits in den 1920er Jahren eine Kooperation zwischen dem Berliner Institut für Schwingungsforschung und den Bell Telephone Laboratories in New Jersey. Ersteres befasste sich hauptsächlich damit, die akustische Phonetik zu erforschen, während letztere eher die Anwendung auf die Nachrichtentechnik anstrebten. Im Dunstkreis beider Felder begannen Komponisten, diese Ergebnisse für sich nutzbar zu machen. Für die frühen elektronischen Musiker war es aufgrund des hohen technischen Aufwands nahezu unabdingbar, sich mit den ingenieurstechnischen Grundlagen vertraut zu machen. In der mir bekannten Literatur jedoch wird kaum auf die technischen Voraussetzungen computergenerierter Sprachkomposition eingegangen, ebenso wenig wird der Einfluss der akustischen Phonetik tiefer beleuchtet.

Mit dieser Arbeit möchte ich versuchen, diese Aspekte zu berücksichtigen und anhand ausgewählter Beispiele von zeitgenössischen Komponisten illustrieren. Vorrangig interessiert mich dabei, wie ein musik-fremdes Gebiet in Musik

---

<sup>4</sup> Als prominentestes Beispiel einer solch kombinierten Arbeit sei hier Stockhausens *Gesang der Jünglinge* erwähnt.

überführt werden kann.

Für das Feld der Sound Studies könnte die Arbeit einen interessanten Beitrag leisten, indem sie sich mit dem Prinzip besonderer Klangfarbenkomposition beschäftigt. Ich möchte untersuchen, welche Voraussetzungen den Boden dafür bereiten, mit elektronisch modifizierter Sprache kreativ umzugehen.

## 2 Grundlagen akustischer Phonologie

### 2.1 Unterschied zur musikalischen Akustik

Sprache ist ein komplexes Schallereignis. Würde man gebeten werden, den Schall einer Rede auf einer Gitarre nachzubilden, gelangte man schnell an die Grenzen. Nicht nur, dass die Klangfarbe der Gitarre nicht einer Stimme entspricht, nein, auch die Sprachmelodie lässt sich nicht so einfach zusammensetzen. Woran liegt das? Zum Einen daran, dass wir uns im Alltag nicht entsprechend der vorhandenen zwölfstufigen Tonleiter artikulieren, zum Anderen sind geäußerte Laute mehr oder weniger aperiodische Schallereignisse – im Gegensatz zu der angenäherten Periodizität unserer Instrumentenklänge. So muss dieser Versuch scheitern.<sup>5</sup> Doch damit ist die Geschichte der Übersetzung von Sprachklängen in Musik noch lange nicht zuende erzählt.

Fasst man ein Schallereignis auf, als wäre es aus unterschiedlichen periodischen Schwingungen zusammengesetzt, die miteinander interferieren, so ergeben sich mittels einer Fouriertransformation Möglichkeiten, Sprache in diese Schwingungen aufzuschlüsseln. Für aperiodische Signale wird dabei eine unendlich kleine Grundfrequenz des Signals angenommen, deren Harmonische sich über alle Frequenzen verteilen können. Die Erforschung der Quasi-Teilharmonischen liefert die Grundlage für die Synthese von Sprache.

Für die frühe Psychoakustik war es mit immensem Aufwand verbunden, Sprachschall zu analysieren. Carl Stumpf beschreibt in *Die Sprachlaute* eindrucksvoll, wie zu seiner Zeit verschiedene Resonanzröhren an das Ohr oder

---

<sup>5</sup> Steve Reich unternimmt mit *Different Trains* einen bedeutenden Schritt, aus den Sprachmelodien (Prosodie) ein Streicherarrangement abzuleiten. Festzuhalten ist jedoch, dass Sprache *nicht natürlicherweise* musikalisch artikuliert wird. Zudem verfolgt Reich nicht das Ziel, Formanten nachzubilden.

Stimmgabeln vor den Mund des Sprechers gehalten wurden oder man diesem gar den Kopf abklopfte um die Teiltöne zu messen<sup>6</sup>.

## **2.2 Akustische Eigenschaften der Laute**

Phoneme sind die kleinsten bedeutungsunterscheidenden Lautelemente einer Sprache. Die Phonetik gruppiert sie in Plosive, Frikative, Affrikative, Nasale, Liquide, Gleitlaute und Vokale.<sup>7</sup> Alle Phoneme decken unterschiedliche Frequenzbereiche ab. Das liegt an der Art und Weise, wie sie vom *supralaryngealen Filter*, d.h. durch das Zusammenspiel von Stimmlippen, Luftröhre sowie dem Mund- und Rachen- und Nasenraum, geformt werden. Darauf komme ich im nächsten Abschnitt zurück.

Das Frequenzspektrum stimmloser Konsonanten, wie bei /f/ in /elektrisch/, ist diffus; auf allen Frequenzen erscheint im Spektrum stufenlos die gleiche Ausprägung. Je stimmhafter ein Phonem ist, desto stärker treten bestimmte Frequenzen hervor, welche als *Formanten* bezeichnet werden. Diese konstituieren die Klangfarbe. Für Vokale gilt: Sie bestehen aus weniger Formanten als Konsonanten. *Zwei Hauptformanten* sind für ihre Beschreibung grundlegend. Im Spektrum lässt sich die Verteilung der Vokalformanten folgendermaßen veranschaulichen:

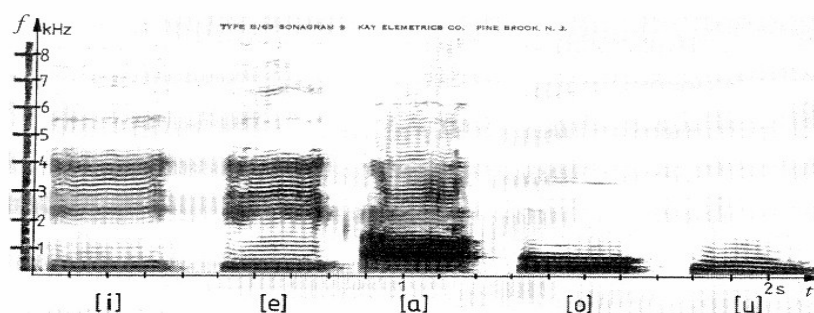


Abb. 1: Spektren der Vokale

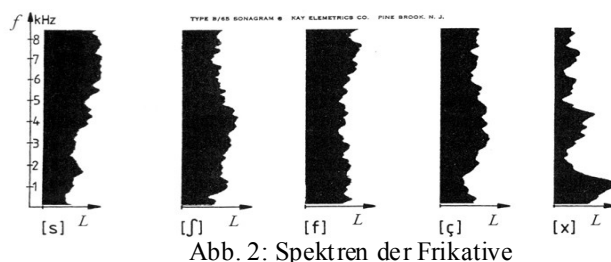
---

6 Stumpf, S. 9.

7 Aus Platzgründen verzichte ich hier auf weitere Spezifizierungen. Als weiterführende Literatur sei jede Einführung zur Phonetik empfohlen.

Reine stimmlose Konsonanten wie die Zischlaute besitzen dagegen *unendlich viele Formanten*.

Abb. 2



Bei allen Sprachlauten sind diese konstituierenden Frequenzen unabhängig von der Tonhöhe des Sprechers immer vorhanden<sup>8</sup>. Verfügt man über Informationen, wie das Spektrum eines Lautes beschaffen ist, gilt es, die beteiligten Frequenzen nachzubilden, um zur Sprachsynthese zu gelangen. Sprachsynthese ist damit eine besondere Form der *Klangfarbenmodulation*. Meyer-Eppler bezeichnet dies vielleicht exakter als *spektrale Modulation*<sup>9</sup>.

## 3 Sprachsynthese

### 3.1 Voraussetzungen

Schwingen die menschlichen Stimmlippen periodisch, so wird ein Ton produziert. Bleiben sie geöffnet, so kann nur ein Rauschen produziert werden. Außerdem gibt es Phoneme, die sowohl einen Rauschanteil als auch einen tonalen Anteil aufweisen. Individuell verschieden ist die Eigenschwingung der Luftröhre. Die Luft, die die Glottis verlässt, wird durch das *superglottale Filter*, bestehend aus Nasen-, Mund- und Rachenraum, moduliert. So werden der Grundfrequenz bzw. dem Rauschen weitere Frequenzanteile beigemischt<sup>10</sup>. Sämtliche Kodierungsverfahren für Sprache funktionieren analog zur Sprachproduktion nach dem *Quelle-Filter-Model (source-filter-model)*.

Der zweite Eckpfeiler der Sprachsynthese ist durch die *Darstellung im*

8 Um zu diesen allgemeinen Frequenzen zu gelangen, ist es erforderlich, den Einfluss der Harmonischen der Grundschwingung aus dem Spektrum herauszurechnen, denn die Grundschwingung ist sprecherabhängig. Zu bemerken ist hier der Unterschied zu den Formanten von Instrumentalklängen, die sich im Gegensatz zur sprachlichen Äußerung mit der Tonhöhe verschieben.

9 Meyer-Eppler: *Elektrische Klangerzeugung*. S. 78 ff.

10 Der Einfluss des resonierenden Brustkorbs wird im Allgemeinen vernachlässigt.

Spektrum gegeben. Im Spektrum lässt sich ein komplexes Schallereignis als ein aus mehreren Frequenzen zusammengesetztes darstellen. Die Frequenzanteile mitsamt ihrer Amplitude werden durch eine Fouriertransformation gewonnen und erfahren je nachdem, was das Ziel der Analyse ist, eine spezifische Auflösung im Abbild. Eine jede Sprachsynthese geht davon aus, dass die Spektralansicht für die Lautäußerung repräsentativ ist.

Daher erscheint eine Quelle-Filter-Modell-artige Nachbildung des Spektrums vielversprechend, um Sprache verständlich zu synthetisieren. In der Fachliteratur werden solche Verfahren unter dem Kurzwort SMS (*spectral modeling synthesis*) zusammengefasst.

### **3.2 Von Stimmgabeln zum Vocoder**

Zunächst beschränkte sich die Sprachsynthese auf Vokale, da man aufgrund technischer Voraussetzungen noch nicht in der Lage war, komplexe Klänge hinreichend zu analysieren. So hätte es, die Gabelmethode vorausgesetzt, einer unendlichen Anzahl von Stimmgabeln bedurft um ein einziges /f/ zu generieren. Jedoch war die Gabelmethode geeignet um Vokale herzustellen, da sie mit obertonarmen Klangerzeugern arbeitete. Dem Akustiker Herrmann von Helmholtz gelang dies offenbar bereits im 19. Jahrhundert. Zu Beginn des 20. Jahrhunderts etablierte Carl Stumpf die Vokalsynthese durch ein komplexes System miteinander verbundener Röhren. So konnte man Frequenzen gezielt zuschalten bzw. eliminieren<sup>11</sup>. Er kam bereits 1926 zu dem Ergebnis,

*[...] daß sich zwischen Vokal- und Instrumentalklängen eine völlig scharfe Grenze nicht ziehen läßt, daß aber gerade dieser Umstand die Unterordnung der stimmhaften Vokale unter den Allgemeinbegriff der Klangfarben nur noch entschiedener rechtfertigt.<sup>12</sup>*

Die massenhafte Nutzung der Elektrizität verhalf der Klangsynthese zu neuem

---

<sup>11</sup> Stumpf: Die Sprachlaute. S. 171 ff.

<sup>12</sup> Ebd., VI.



Aufschwung: Da sich durch elektrische Spannung periodische Schwingungen erzeugen lassen, deren Frequenz modulierbar ist, kann man durch die Verbindung mehrerer Generatoren bis zu einem gewissen Grad komplexe Klänge erzeugen (additive Synthese). Integriert man in einen Schwingkreis einen Widerstand, so lässt sich eine Rauschspannung herstellen.

Lassen sich nun beliebige Schwingungen elektrisch erzeugen, so lässt sich auch Sprache synthetisieren. Das eindrucksvolle erste Gerät für die Sprachsynthese ist die Urform des Vocoders, der *Voder*. Die spieltischartige Apparatur wurde in den 1920er Jahren von Dudley an den Bell Telephone Laboratories entwickelt. Mit dem Voder gelang es erstmals, auch Konsonanten zu generieren. Durch manuelles Bedienen können verschiedene Schwingkreise angesteuert werden. Zunächst muss man dabei entscheiden, ob es sich um einen stimmhaften oder stimmlosen Laut handelt und dementsprechend Impulsgenerator oder Rauschgenerator zuschalten. (Diese Unterscheidung taucht auch im weiter unten beschriebenen LPC-Verfahren auf.) Indem verschiedene Formantfilter zugeschaltet werden, lässt sich der Frequenzbereich der Laute bestimmen. Die Tonhöhe stimmhafter Laute wird über ein Pedal gesteuert. Ein zusätzlicher Schalter, die Leise-Taste, ermöglicht es, die Amplitude eines Lautes um 20 dB abzusenken und somit zwischen betonten und unbetonten Silben zu unterscheiden<sup>13</sup>

Dass die Analyse von Sprache nicht manuell geschehen muss, zeigte die Entwicklung des *Vocoders*, der für die Nachrichtentechnik von erheblicher Bedeutung war. Zunächst wird über ein Mikrofon das Schallsignal aufgenommen. Die Frequenzanteile werden durch Filter analysiert, während die Grundschiwingung durch einen einzelnen Frequenzmesser ermittelt wird. Auf der Wiedergabeseite wird das Signal durch den Einsatz von gewöhnlich 12 Bändern sowie einer periodischen Schwingung bzw. Rauschen resynthetisiert.

---

13 Vgl. Meyer-Eppler: *Elektrische Klangerzeugung*, S. 119 ff.

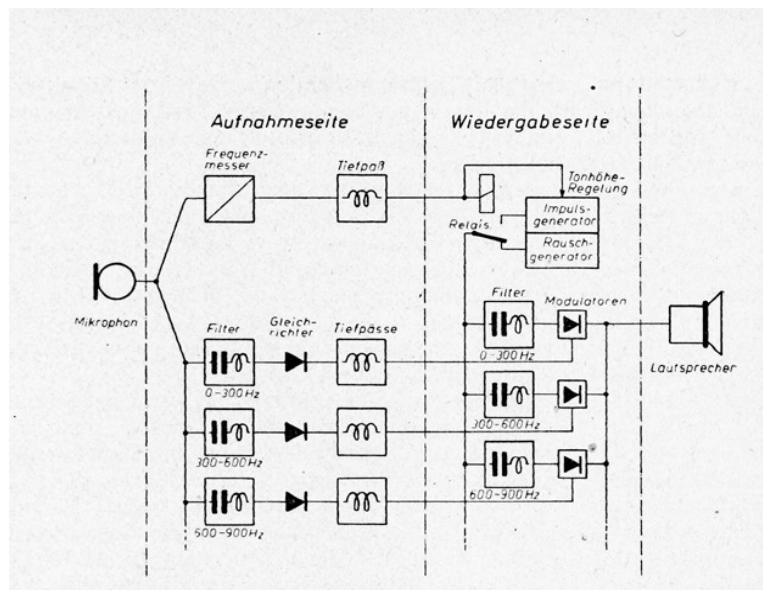


Abb. 3: Schaltplan eines Vocoder

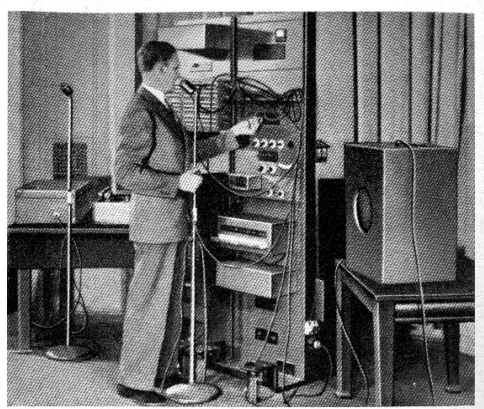


Abb. 4: Früher Vocoder im Einsatz

Ein Nachteil des Verfahrens war jedoch, dass stets die Informationen für alle Filterströme übertragen werden müssen und damit das Codierungsverfahren als nicht besonders effizient gilt. Gleichzeitig kommt es zu gewissen Unschärfen bei der Wiedergabe, da die 12 Bänder eben auch nur 12 Frequenzanteile resynthetisieren können und beispielsweise bei der Wiedergabe von Nasalen, die sowohl Rauschanteile als auch periodische Frequenzen besitzen, Schwachpunkte aufweisen.

### 3.3 Linear Predictive Coding

Das Verfahren des Linear Predictive Coding wurde in den 1960er Jahren von Schroeder<sup>14</sup> und Atal<sup>15</sup> an den Bell Telephone Laboratories entwickelt. Als nachrichtentechnisches Verfahren wurde es dazu verwendet, akustische Daten zu verschlüsseln und weniger rechenintensiv übertragen zu können. Vor allen Dingen die mobile Telekommunikation beruht auf diesem Prinzip. Bei LPC handelt es sich im Gegensatz zum Vocoder nur um *einen* Filter, der effizient arbeitet und durch eine mehrpolige Hüllkurve ein Spektrum simuliert.

Ein Eingangssignal wird mit einem Abstand von 10 ms<sup>16</sup> nach seinen Frequenzen analysiert. Das geschieht durch sog. Pole, die wie Filter funktionieren. Von Zeitfenster zu Zeitfenster werden Korrelationen zwischen den Polen errechnet, woraus sich auf statistischem Wege

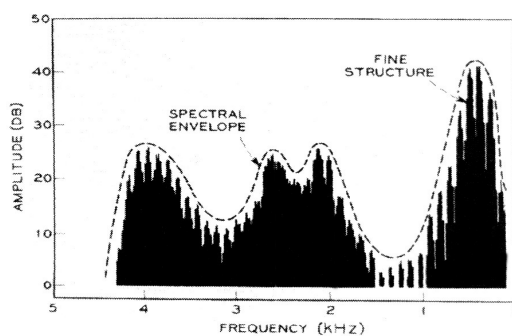


Abb. 5: Ausgabe eines mehrpoligen Filters

eine gewisse Vorhersagekraft für den weiteren Signalverlauf ableiten lässt. Durch die Ausprägung der Pole lassen sich Aussagen zu den Formanten des Sprachsignals treffen. Die Formanten werden mit der analysierten Tonhöhe des Sprechers (*pitch detection*) kombiniert und statt des ursprünglichen Signals übertragen. Auf der Empfängerseite wird das Sprachsignal mithilfe dieser Informationen resynthetisiert. Für Sprachlaute, deren Tonhöhe nicht anzugeben ist, die also *stimmlos* sind, wird ein Rauschgenerator eingesetzt; für *stimmhafte* Phoneme kommt ein Oszillator zum Einsatz, bei *stillen* Abschnitten nichts von beiden.<sup>17</sup> Ob man bei der Resynthese als Anregungsfrequenz nun eine periodische Schwingung oder ein Rauschen hinzunimmt, lässt sich ermitteln, wenn das linear kodierte Signal vom originalen abgezogen wird. Diese Restgröße (*residual*)

---

14 Schroeder: *Vocoder: Analysis and Synthesis of Speech*.

15 z. B. Atal & Hanauer: *Speech Analysis and Synthesis by Linear Prediction of the Speech Wave*.

16 In einigen Fällen beträgt die Fenstergröße bis zu 20 ms.

17 Dieses Element kann man als Analogon zu dem Zustand der Stimmlippen betrachten, die entweder eine periodische Schwingung oder ein Rauschen hervorbringen.

steuert, welcher Generator verwendet wird. Zusätzlich dient das Residuum dazu herauszufinden, welche Koeffizienten nun am besten das Eingangssignal beschreiben, d.h. wenn die Restgröße klein ausfällt, kann die Analyse – im statistischen Sinne – viel von ihm erklären.

Ein LPC-System funktioniert – ähnlich wie ein Vocoder – immer zweiseitig, d.h. Es gibt eine Kodierseite (Analyse) und eine Dekodierseite (Resynthese). Wohl existieren inzwischen avanciertere Formen von LPC<sup>18</sup>, die sich durch bessere Übertragungsqualität auszeichnen, doch scheint mir dies eher Gebiet der Nachrichtentechnik zu sein und für die weitere Argumentation weniger relevant. Ebenso wenig soll hier diskutiert werden, inwiefern Sprachsynthese-Systeme mangelhaft sind, wenn menschliche Sprache möglichst exakt nachgebildet werden soll. Wichtig für meine Arbeit sind eher die künstlerischen Möglichkeiten, die sich aus dem Verfahren ergeben. Der Komponist Paul Lansky setzt mit seinen Sprachkompositionen bei der Resynthese an.

## **4 Umsetzung in zeitgenössischen Kompositionen**

### **4.1 Paul Lansky: Idle Chatter Junior (1999)**

Für die insgesamt vier *Idle-Chatter*-Stücke (*Idle Chatter Junior* ist das jüngste unter ihnen) kombinierte Paul Lansky das LPC-Verfahren mit der Granularsynthese. Die Granularsynthese, in die ein Zufallsalgorithmus integriert war, diente dazu, die aus dem LPC gewonnenen Sprachsegmente zu einer neuen Einheit zusammenzufügen. In diesem Abschnitt möchte ich mich damit befassen, wie Lansky das LPC nutzt, um Sprachklänge zu simulieren, die sehr musikalisch anmuten, einer erkennbaren Semantik jedoch entbehren.

Treffend erklärt Moorers in seinem Artikel *The Use of Linear Prediction of Speech in Computer Music Applications* das ästhetische Prinzip einer solchen auf LPC beruhenden Sprachkomposition<sup>19</sup>:

---

18 Z. B. *Code-excited linear prediction*, womit Schwankungen der Tonhöhe besser ausgeglichen werden können.

19 Diesen Umgang mit LPC könnte man als *modifizierte Resynthese* bezeichnen.

*In general, there is no point in directly resynthesizing a piece of speech or singing. One could just use directly the original segment. The only point is to be able to modify the speech in ways that would be difficult or impossible for the speaker to do.*"<sup>20</sup>

Wer Nachrichtentechnik für die künstlerische Arbeit verwendet, dem geht es also nicht um die höchste Wiedergabetreue des Signals. Welche Möglichkeiten ergeben sich nun aus dem LPC? Die kreative Manipulation geschieht bei Lanskys *Idle-Chatter-Stücken* auf mehreren Wegen:

- a) *Tonhöhe*. Bei der Resynthese werden die Koeffizienten nicht durch die ursprüngliche Anregungsfrequenz ergänzt, sondern durch einen beliebigen anderen Impuls.
- b) *Zeit*. Es ist möglich, die Fensterrate so zu beeinflussen, dass die Koeffizienten sich schneller oder langsamer verändern.
- c) *Klangfarbe*<sup>21</sup>. Eine *Formantenmanipulation* lässt sich erreichen, indem der Filterausgang modifiziert wird und sich damit die Koeffizienten ändern. Dieser Vorgang lässt sich bis zum *physical modeling* weiterentwickeln, so dass rein virtuelle Klangobjekte geschaffen werden können. Außerdem ist es möglich, einen Klang in einen anderen zu überführen (*Interpolation*), wenn das Spektrum zweier Signale gemittelt wird und aus diesen Koeffizienten eine Vielzahl von Übergangsfenstern berechnet wird. Hybride Klangfarben können durch sog. *cross-synthesis* erzeugt werden, d. h. wenn man die Spektren mehrerer Fenster übereinander legt. Um den Übergang zwischen zwei Fenstern fließend zu gestalten, verwendet Lansky das Prinzip der Überlappung: So enthält ein jedes Fenster zu 50 % Informationen des vorhergehenden und des nachfolgenden Abschnitts.

In der Realität sind diese Vorgänge komplexer als oben dargestellt. Beispielsweise stellte es für die frühen LPC-Anwendungen ein Problem dar, die

---

<sup>20</sup> Moorer, The Use of Linear Prediction of Speech in Computer Music Applications. S. 134.

<sup>21</sup> Um tiefer in das Gebiet einzusteigen, sei hier *DAFX. Digital Audio Effects* von Udo Zölzer empfohlen.

sog. *buzzyness* zu eliminieren. Damit ist gemeint, dass eine resynthetisierte Stimme leicht als solche zu erkennen war, da der Oszillator beständig bei den stimmhaften Lauten eine periodische Schwingung und damit eine Art Summen produzierte. Bei der Sprachproduktion jedoch wird diese Schwingung ab und zu unterbrochen oder ein geringer Rauschanteil beigemischt. Lansky gleicht eine artifizielle Stimme einer natürlichen an, indem er beim Resynthetisieren der Frequenz einen Zufallsalgorithmus einfügt, der Rauschen generiert oder die Anregungsfrequenz nicht vollkommen periodisch macht.

Paul Lansky bediente sich für diese drei Kompositionen eines selbst entwickelten Computerprogramms namens Cmix, welches die Synthese ausführt. Es ist heute ebenso wie ein Programm zum LPC frei erhältlich<sup>22</sup>.

## 4.2 Klarenz Barlow: Im Januar am Nil (1984)

Barlow verfolgt in seiner Komposition einen tonaleren Ansatz der Sprachkomposition. Für das Stück verwendete er ein Verfahren, das er *Synthrummentation* nennt. Dabei handelt es sich um die *additive Resynthese* von Sprache durch Instrumente. Der Komponist geht von Textfragmenten aus, die eingesprochen und dann anhand ihres Spektrums in Frequenzen transformiert werden<sup>23</sup>. Die Grundschwingung des Sprechers - in diesem Falle ist er wahrscheinlich männlich - wird dabei nicht herausgerechnet, sondern von einem Instrument gespielt. Ebenso finden sich die Harmonischen dieses Grundtons in der Partitur. Durch Runden passt Barlow alle ermittelten Frequenzen dem diatonischen Notensystem an und macht die sprachlichen Äußerungen damit für Instrumente spielbar. Da die Anzahl der Instrumente begrenzt ist und diese eben nur die Frequenzen des diatonischen Systems hervorbringen können, beschränkt sich die Synthese auf Vokale und Konsonanten mit starkem stimmhaften Anteil, z.B. Nasale. Das wird an der von Barlow verwendeten Textvorlage deutlich, von der z. B. die Zeile „An Müllmänner in Armenien nun ein Jahr lang erinnern“<sup>24</sup> arrangiert wird.

---

22 Cmix ist erhältlich unter [www.music.princeton.edu/winham/cmixon.html](http://www.music.princeton.edu/winham/cmixon.html),  
LPC unter [soundlab.cs.princeton.edu/software/rt\\_lpc](http://soundlab.cs.princeton.edu/software/rt_lpc) oder [www.hawksoft.com/hawkvoice](http://www.hawksoft.com/hawkvoice)  
(alle drei Websites aufgerufen am 23.09.08)

23 Aus dem Visuellen das Musikalische abzuleiten, ist typisch für Barlows Arbeitsweise.

24 Zit. nach CD-Beiheft: *Sprachkomposition*. S. 18.

Im *Januar am Nil* existiert sowohl in einer Version für Synthesizer, realisiert durch einen Akai-Sampler, als auch als Ausgabe für Streicherensemble. Letztere Umsetzung kommentiert Barlow mit: „Das Ergebnis ist nicht so deutlich wie auf dem Sampler.“<sup>25</sup> Doch wie lässt es sich erklären, dass Streichinstrumente die Sprachlaute nicht verständlich simulieren können? Ein Grund wäre, dass Sprachformanten unterschiedlich laut auftreten und dies hinreichend repräsentiert werden müsste. Ob die dynamischen Anweisungen von *pianissimo* bis *fortissimo* dazu ausreichen, ist fragwürdig. Auch durch die Anzahl der Instrumente könnte man die Amplitude der Formanten kontrollieren. Bei Barlowes Besetzung durch vier Violinen, zwei Celli, einen Kontrabass und eine Bassklarinette ist das nicht gewährleistet. Dennoch: „Unter Idealbedingungen sollte durch eine genaue Einhaltung der Partitur Sprachlaute zu erkennen sein“<sup>26</sup>, so Barlow. Bei der Umsetzung durch den Akai-Sampler, der eine genauere Kontrolle der Amplitude ermöglicht, scheinen die phonetischen Qualitäten jedoch noch immer nur bruchstückhaft durchzudringen. Das mag an den Rundungen liegen, denen die Sprachfrequenzen unterzogen wurden, doch soll das Thema hier nicht weiter erörtert werden.

Um die Noten im Sampler zusammenzufügen, greift Barlow auf das von ihm entwickelte *Autobusk*<sup>27</sup> zurück. Dieses dient der Verarbeitung von Midi-Daten innerhalb eines *Atari*-Systems und ist heute frei erhältlich.

## 5 Schlussfolgerungen und Ausblicke

Es wurde anhand einiger Beispiele gezeigt, wie aus den Prinzipien der Sprachsynthese kompositorische Ansätze entwickelt werden können. Die Qualität einer jeden Sprachsynthese hängt eng mit dem *Stand der Analyse* zusammen. Für Paul Lanskys *Idle Chatter*-Kompositionen war die Nähe zu den Bell Telephone Laboratories, der Entwicklungsstätte für LPC, mutmaßlich ausschlaggebend<sup>28</sup>.

Die Sprachkompositionen der elektronischen Musik setzen zumeist ein

---

25 Barlow: Methoden algorithmischer Klangerzeugung. S.76.

26 Barlow: Werkkommentar. Zit. nach CD-Beiheft *Sprachkomposition*, S. 18.

27 Bei [www.musikwissenschaft.uni-mainz.de/Autobusk](http://www.musikwissenschaft.uni-mainz.de/Autobusk) (03.09.08).

28 Lansky: The Importance of Being Digital.

[www.music.princeton.edu/~paul/lansky\\_beingdigital.htm](http://www.music.princeton.edu/~paul/lansky_beingdigital.htm) am 04.10.08

Computerprogramm voraus. Die vorgestellten Komponisten besitzen fundierte Programmierkenntnisse. Heute jedoch lassen sich Kompositionen nach Lansky oder Barlow relativ problemlos nachbauen: Im Internet existieren LPC-Programme genauso wie das von Lansky verwendete *Cmix* und Barlows *Autobusk* zum kostenlosen Herunterladen<sup>29</sup>. Programmierumgebungen wie *Max/MSP* oder *Pure Data* ermöglichen weiterführende Optionen für Sprachkompositionen, werden aber meines Wissens nach kaum dazu eingesetzt.

Es ist eher die spektrale Modulation im Allgemeinen, die in Form von Effekten breiten Einzug in die Audiotbearbeitung gehalten hat, um Klangfarben jeglicher Natur zu modifizieren. Für die Sprachmanipulation in der Popmusik wird das *Vocoding* häufig verwendet, obwohl es wie oben beschrieben gewisse Kodierungsmängel aufweist. Jedes Programm zur Audiotbearbeitung verfügt mittlerweile über einen digitalen Vocoder, der gerade den für die Nachrichtentechnik als störend empfundenen Buzzyness-Effekt als Charakteristikum nutzt.

Es scheint für die künstlerische Arbeit auf dem Gebiet der elektronischen Sprachkomposition lohnenswert, sich eingehender mit akustischer Phonetik und Formantmustern zu beschäftigen, um die Ergebnisse besser steuern oder erklären zu können. Um die Möglichkeiten der Manipulation zu entdecken, bieten sich nicht nur die Kodierungsverfahren der Nachrichtentechnik an. Auch die medizinische Akustik mit ihren neuen Methoden, Sprachsignale vom Rauschen zu trennen, könnte fruchtbar sein.

---

<sup>29</sup> *Cmix* ist erhältlich unter [www.music.princeton.edu/winham/cmix.html](http://www.music.princeton.edu/winham/cmix.html),  
LPC unter [soundlab.cs.princeton.edu/software/rt\\_lpc](http://soundlab.cs.princeton.edu/software/rt_lpc) oder [www.hawksoft.com/hawkvoice](http://www.hawksoft.com/hawkvoice)  
(alle drei Websites aufgerufen am 23.09.08),  
*Autobusk* gibt es bei [www.musikwissenschaft.uni-mainz.de/Autobusk](http://www.musikwissenschaft.uni-mainz.de/Autobusk) (03.09.08).



## 6 Literaturverzeichnis

### 6.1 Literatur

- ATAL, B. UND HANAUER, S.: Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. In: The Journal of the Acoustical Society of America. Vol. 50, No. 2, Part 2. 1971. S. 637 – 655.
- BARLOW, K.: Methoden algorithmischer Klangerzeugung anhand eigener Arbeiten: Von Klangglomeration und Schallwellenreiten zu Synthrummentation und Spektastik. In: Stelkens, J. und Tillmann, H. (Hrsg.): Klangforschung '98 / Symposium zur Elektronischen Musik vom 26.10. bis 30.10.1998 in München. Saarbrücken: 1999.
- LANSKY, P.: Compositional Applications of Linear Predictive Coding. In: Mathews, M. und Pierce, J. (Hrsg.): Current Directions in Computer Music Research. London: 1989. S. 5 – 8.
- LENTZ, M.: Lautpoesie. Lautmusik nach 1945. Eine kritisch-dokumentarische Bestandsaufnahme. Wien: 1998.
- MEYER-EPPLER, W.: Elektrische Klangerzeugung. Elektronische Musik und synthetische Sprache. Bonn: 1949.
- MOORE, B.: An Introduction to the Psychology of Hearing. London: 1994. 4. Aufl.
- MOORER, J.: The Use of Linear Prediction of Speech in Computer Music Applications. In: Journal of the Audio Engineering Society. Vol. 27, No. 3. 1979. S. 135. – 140.
- NEPPERT, J.: Elemente einer akustischen Phonetik. Hamburg: 1999. 4. Aufl.
- REETZ, H.: Artikulatorische und akustische Phonetik. Trier: 2003. 2. Aufl.
- SCHROEDER, M.: Vocoders: Analysis and Synthesis of Speech. In: Proceedings of the IEEE. Vol. 54, No. 5. S. 720 – 734.
- STUMPF, C.: Die Sprachlaute. Experimentell-phonetische Untersuchungen. Nebst einem Anhang über Instrumentalklänge. Berlin: 1926.
- SUPPER, M.: Elektroakustische Musik und Computermusik. Geschichte – Methoden – Systeme. Darmstadt: 1997.
- ZÖLZER, U. (HRSG): DAFX. Digital Audio Effects. Chichester: 2002.

### 6.2 Online-Quellen

- Programm Cmix: [www.music.princeton.edu/winham/cmixon.html](http://www.music.princeton.edu/winham/cmixon.html) (04.10.08)
- LPC-Programm: [soundlab.cs.princeton.edu/software/rt\\_lpc](http://soundlab.cs.princeton.edu/software/rt_lpc) (04.10.08)
- LPC-Programm: [www.hawksoft.com/hawkvoice](http://www.hawksoft.com/hawkvoice) (04.10.08),
- Autobusk-Programm: [www.musikwissenschaft.uni-mainz.de/Autobusk](http://www.musikwissenschaft.uni-mainz.de/Autobusk) (03.09.08).

Paul Lanskys Website: [www.paullansky.org](http://www.paullansky.org) (07.10.08)

Frequenz-Midi-Ton-Wandler:

[http://tomscarff.tripod.com/midi\\_analyser/midi\\_note\\_frequency.htm](http://tomscarff.tripod.com/midi_analyser/midi_note_frequency.htm) (08.10.08)

### **6.3 Klangbeispiele**

Barlow, K.: Im Januar am Nil. In: Deutscher Musikrat (Hrsg.): Musik in Deutschland 1950 – 2000. Sprachkomposition. Compact Disc. 2001.

Lansky, P. Not Just More Idle Chatter. In: [www.ubu.com/sound/tellus\\_22.html](http://www.ubu.com/sound/tellus_22.html) (20.09.08)

Lansky, P.: Idle Chatter Junior. In: [www.paullansky.org](http://www.paullansky.org) (01.10.08)

## **7 Abbildungsverzeichnis**

Abb. 1 in: NEPPERT, J.: Elemente einer akustischen Phonetik. Hamburg: 1999. 4. Aufl., S. 107.

Abb. 2 in: Ebd.S. 102.

Abb. 3 in: MEYER-EPPLER, W.: Elektrische Klangerzeugung. Elektronische Musik und synthetische Sprache. Bonn: 1949. S.120.

Abb. 4 in: Ebd., Tafel XIV.

Abb. 5 in: SCHROEDER, M.: Vocoders: Analysis and Synthesis of Speech. In: Proceedings of the IEEE. Vol. 54, No. 5. S. 722.