

Vocoder für Sprachverfremdung und Klangeffekte

Vocoder VSM 201

Published in FUNKSCHAU 1978, Heft 7
by Dipl.-Ing. Dietrich Buder



Dipl.-Ing. Dietrich Buder, (45), Studium und Dipl.-Prüfung an der TH-Hannover. Seit 1959 bei Sennheiser electronic als Laborgruppenleiter, massgeblich an der Entwicklung des Vocoders VSM 201 beteiligt

Die Erschaffung einer künstlichen Stimme ist ein uralter Menschheitstraum, und bereits im 17. Jahrhundert entstanden rein mechanische Sprechmaschinen. Aber erst in den dreissiger Jahren dieses Jahrhunderts gelang es Homer Dudley, auf elektronischem Wege eine synthetische Sprache zu erzeugen. Er nannte sein Gerät Vocoder, eine Abkürzung des englischen Ausdrucks voice coder (Stimmcodierer). Dudley hoffte damals sogar, nur durch geschicktes Betätigen einer geeigneten Vocoder-Tastatur regelrecht Sprache erzeugen zu können. Diese Hoffnung erfüllte sich zwar nicht; die Vocoder erlangten jedoch für verschiedene Anwendungsfälle grosse Bedeutung. Der folgende Beitrag beschreibt einen neuartigen elektronischen Vocoder von Sennheiser Electronic.

Vocoder für Sprachverfremdung und Klangeffekte 1. Teil

■ Menschliche Stimme

Zum Verständnis des Vocoderprinzips ist es erforderlich, zunächst die menschliche Spracherzeugung zu betrachten. Unter den verschiedenartigen menschlichen Lauten gibt es zwei Hauptgruppen oder Stimmarten: stimmhafte Laute wie a, e, i, o, u und stimmlose oder Zischlaute wie f, sch, z. Bei den stimmhaften Lauten unterbrechen die Stimmbänder im Kehlkopf periodisch den Luftstrom der Luftröhre und erzeugen dabei dreieckförmige Druckimpulse. Die Zahl der Unterbrechungen pro Sekunde bzw. die Folgefrequenz bestimmt die musikalische Tonhöhe des stimmhaften Lautes: es ist die sogenannte Melodiegrundwelle oder Melodiefrequenz. Sie liegt bei tiefen Bässstimmen bei etwa 80 Hz und reicht bei Sopransängerinnen bis über 1000 Hz. Die normalen Stimmlagen für Männer und Frauen umfassen den Bereich von etwa 90 Hz bis 400 Hz.

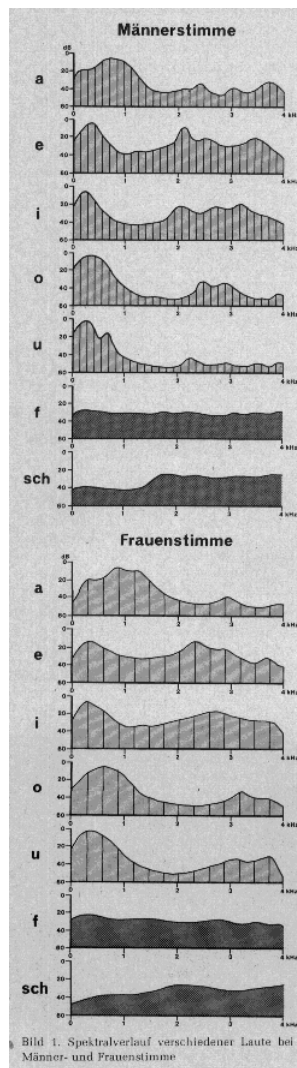
Die Druckimpulse des Kehlkopfes enthalten neben der Grundschwingung noch viele harmonische Oberschwingungen bis über 4 kHz, in dem üblichen Frequenz/Amplituden-Diagramm ein Linienspektrum. Die Mund-, Nasen- und Rachenhöhlräume stellen Filter dar, die aus dem Kehlkopfspektrum bestimmte Bereiche selektieren und den typischen Spektralverlauf der einzelnen stimmhaften Laute formen (Bild 1). Dieser Verlauf lässt sich sehr übersichtlich durch eine Umhüllende darstellen.

Die deutlich ausgeprägten Frequenzbereiche mit hoher Energie heissen Formanten. Sie sind bei Männer- und Frauenstimmen sehr ähnlich. Aufgrund der höheren Stimmlage der Frauenstimmen unterscheiden sich die Spektrogramme deutlich durch die Frequenzabstände der Oberschwingungen.

Bei stimmlosen Lauten schwingen die Stimmbänder nicht, sondern lassen den Luftstrom ungehindert passieren. Die Zischlaute sind Windgeräusche, die durch das Vorbeiströmen der Luft an Gaumen, Zunge, Zähnen und Lippen entstehen. Ihnen entsprechen gefilterte Rauschspektren mit ebenfalls typischen Hüllkurven für die einzelnen stimmlosen Laute. Hier bestehen zwischen Männer- und Frauenstimmen keine charakteristischen Unterscheidungsmerkmale.

Auf die Beschreibung weiterer Laute, wie z. B. Explosivlaute, kann hier verzichtet werden.

Bild 1. Spektralverlauf verschiedener Laute bei Männer- und Frauenstimme



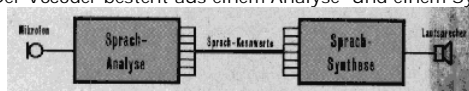
Synthetisches

Bis zum heutigen Tage lebt die elektroakustische Übertragungstechnik davon, akustische Eindrücke in entsprechende elektrische Signale umzusetzen, diese dann zu verstärken, einzelne Frequenzbereiche anzuheben, abzusenken, herauszufiltern, um dann wieder die elektrischen Signale in akustische zu wandeln. Auf diesem Weg wird einiges mit dem Signal gemacht, was dieses dann auch verändert; z.B. wenn man damit Hochfrequenz moduliert, wie bei Sendern. Hinzu kommen jedoch meistens noch störende Veränderungen, wie z. B. Klirrfaktor und Störgeräusche. Wenn man es aber genau nimmt, fließt das Signal doch mehr oder weniger ungehindert - wohl über ein paar Fallstricke - vom Anfang bis zum Ende der Übertragungskette. In unserem heutigen Hauptbeitrag stellen wir dagegen mit einem Vocoder ein Gerät vor, bei dem der Eingang nicht mehr mit dem Ausgang verbunden ist. Hier wird am Ausgang alles neu gemacht. Aus geeigneten intern - oder auch extern - erzeugten Signalen wird das ursprüngliche "Schallereignis" neu generiert - vollständig neu erzeugt. Ich habe schon einmal über einen solchen Vocoder telefoniert. Sieht man einmal von der unüberhörbaren Synthetik des Wiedergehenden ab, so ist es doch erstaunlich, wie weit schon die Ähnlichkeit mit dem Original getrieben ist. Einen einigermaßen bekannten Gesprächspartner kann man - am Telefon! - wiedererkennen.

Hier konnte vielleicht ein neuer Weg für die Zukunft der Übertragungstechnik aufgezeigt sein, in welcher Form auch immer dieses Verfahren dabei angewendet wird.

Bernhard Krieg

Bild 2. Der Vocoder besteht aus einem Analyse- und einem Syntheseteil



Grundprinzip der Vocoder

Jeder Vocoder besteht aus einem Analyse- und einem Syntheseteil (Bild 2).

Die Analyse ermittelt aus dem eingespeisten Sprachsignal verschiedene Steuersignale, die in der Vocoder-Terminologie Sprachparameter oder Sprachkennwerte heißen. Diese werden in einer je nach Anwendungsfall geeigneten Form übertragen. Sie steuern im Syntheseteil eine elektronische Nachbildung des menschlichen Sprechorgans, die eine gutverständliche künstliche Sprache liefert.

Die Bedeutung der Vocoder für die Sprachübertragung erklärt sich aus zwei typischen Eigenschaften: Alle Sprachkennwerte zusammengefasst benötigen eine wesentlich geringere Bandbreite als das ursprüngliche Sprachsignal. Ausserdem lassen sich die Kennwerte sehr leicht beeinflussen und somit bestimmte Klangeffekte in der synthetischen Sprache erzielen. Diese zweite Möglichkeit wird beim Sennheiser-Soundeffect-Vocoder VSM, .201 (Bild 3) und bei emigen anderen Anwendungsfällen

ausgenutzt

Eine für fast alle Vocoder gültige Unterteilung der Analyse und Synthese in Funktionsblöcke zeigt Bild 4. Die Analyseseite besteht normalerweise aus drei Hauptblöcken, der Spektralanalyse, der Stimmartanalyse und der Melodieanalyse.

Die Spektralanalyse ermittelt auf eine geeignete Weise den jeweiligen spektralen Kurvenverlauf der Sprechlaute nach Bild 1, die sogenannte Umhüllende des Sprachspektrums. Sie liefert eine grössere Anzahl von Spektral-kennwerten, die eine Codierung für die Stellung der Mund-, Nasen- und Rachenfilter darstellen.

Die Stimmartanalyse stellt fest, ob die Stimmbänder schwingen oder den Luftstrom ungehindert passieren lassen bzw. ob der Laut stimmhaft oder stimmlos ist. Sie heisst deshalb auch Stimmhaft/Stimmlos-Detektor. Der Ausgang der Stimmartanalyse liefert ein Ja/Nein-Signal für die Funktionsart des Kehlkopfes, den Stimmartkennwert.

Die Melodieanalyse schliesslich sucht aus dem Sprachsignal die Grundwelle heraus und liefert den Melodiekennwert, eine Codierung der Grundfrequenz des Stimmbändchensignals.

Die Synthese besteht aus zwei Funktionsblöcken, der Spektralsynthese und dem Stimmgenerator. Der Stimmartkennwert bestimmt die Art des Stimmsignals: bei stimmlosen Lauten ein Geräusch oder Rauschen, bei stimmhaften Lauten eine Impulsfolge mit einem geeigneten Obertonspektrum. Der Melodiekennwert schliesslich steuert Folgefrequenz des Stimmimpulsgenerators.

Aus den Spektralkennwerten und dem Stimmsignal erzeugt die Spektralsynthese das synthetische Sprachsignal. Die Spektralsynthese stellt somit eine Nachbildung der menschlich Stimmorganfilter dar.

Bild 3. Der neue Soundeffekt-Vocoder VSM 201 von Sennheiser Electronic

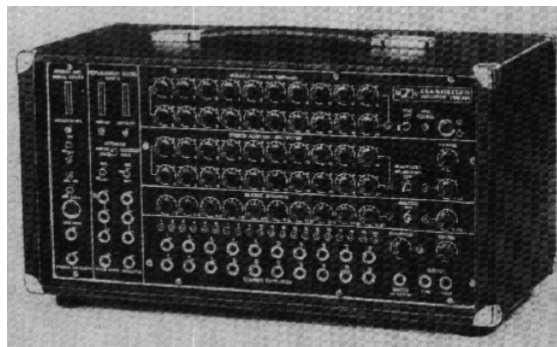


Bild 4/5. Die beiden Funktionsblöcke des Vocoders; die Analyse-Seite besteht aus drei Hauptblöcken, die Synthese-Seite aus zwei

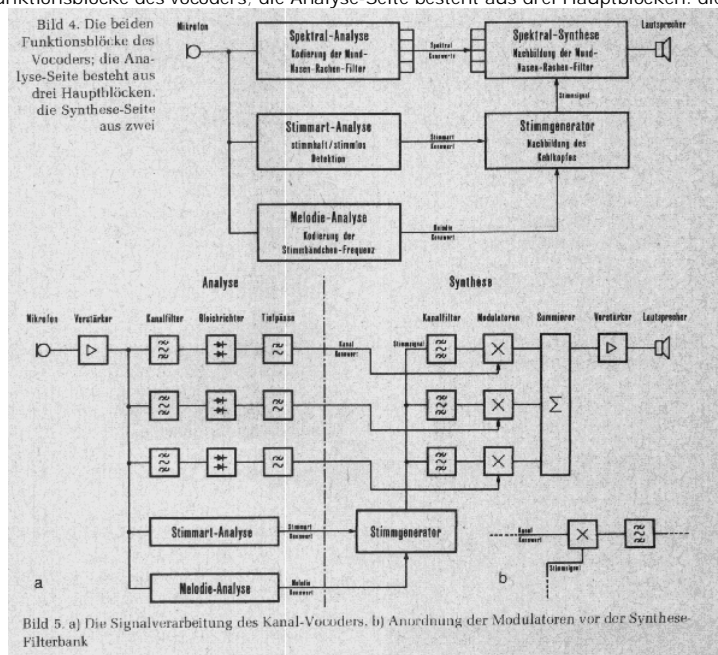
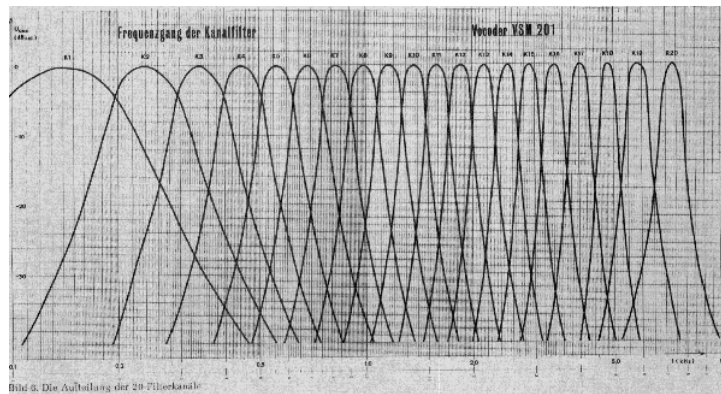


Bild 6. Die Aufteilung der 20 Filterkanäle



■ Grundprinzip des Kanal-Vocoders

Heute gibt es eine Vielzahl von Lösungen für die einzelnen Funktionsblöcke eines Vocoders, und die Entwicklung ist noch lange nicht abgeschlossen.

Die verschiedenen Vocodertypen werden klassifiziert nach der Art der Spektralanalyse. Das von Dudley vorgestellte Gerät gehört danach zu der Gruppe der Kanal- oder Filterbank-Vocoders. Zu diesen zählt auch der Vocoder der VSM 201 von Sennheiser.

Bild 5 zeigt die Signalverarbeitung eines Kanal-Vocoders. Das vom Mikrofon kommende Sprachsignal gelangt über einen Verstärker an einen Satz von Bandfiltern, die sogenannte Filterbank. Jedes der Kanalfilter analysiert einen kleinen Bereich des Sprachbandes. Die Zahl der Filter von praktisch ausgeführten Vocodern reicht von etwa 10 wie beim Dudley-Vocoder, bis hin zu 24.

Mit höherer Filterzahl steigt die Genauigkeit der Analyse und damit die Wiedergabequalität. Andererseits erhöhen sich entsprechend der Schaltungsaufwand und die Zahl der Spektralkennwerte was eine höhere Übertragungsbandbreite zur Folge hat.

Der Sennheiser-Vocoder enthält 20 Kanäle; in Bild 6 sind ihre Aufteilung und die Frequenzgänge dargestellt. Jedem Kanalfilter folgen ein Doppelweg-Gleichrichter und ein Tiefpass. Die Grenzfrequenz dieser Kanaltiefpässe liegt im Bereich von 20 Hz bis 50 Hz. Diese niedrige Grenzfrequenz ist deshalb zulässig, weil sich die menschlichen Sprachlaute aufgrund der relativ langsamen Mund- und Rachenbewegungen gar nicht schneller ändern können. Die Ausgangsspannungen der Tiefpässe sind die Spektral- oder Kanalwerte des codierten Sprachsignals.

Eine kurze Zwischenbetrachtung soll die Bandbreitenverminderung eines Vocoders gegenüber der Originalsprache verdeutlichen. Bei einem 16-Kanal-Vocoder beispielsweise treten an die Stelle des Sprachbandes von 300 Hz bis 4000 Hz ein Satz von 16 Spektralkennwerten mit 20 Hz Grenzfrequenz, der Melodiekennwert mit ebenfalls 20 Hz Grenzfrequenz und noch einige Zusatzsignale, deren Bandbreitenbedarf mit zwei Spektralkennwerten veranschlagt sei. Das ergibt eine rechnerische Bandbreite von $(16+1+2) \cdot 20 \text{ Hz} = 380 \text{ Hz}$ und somit eine Bandbreiten-Reduktion von über 90%. Bei Vocodern mit digitaler Übertragung der Kennwerte ist die erreichbare Reduktion sogar noch etwas höher.

Um die Arbeitsweise eines Kanal-Vocoders besser verstehen zu können, ist es nützlich, die Signalverarbeitung in den einzelnen Baugruppen von Bild 5 anhand einiger Original-Oszillogramme zu betrachten. Bild 7a zeigt die Eingangssprache vor der Filterbank, hier den Anfang des Wortes „Bast“. Das „a“ enthält einen kräftigen ersten Formanten im Bereich von 500 Hz bis 1000 Hz (siehe Bild 1). Das Analysefilter des Kanals 6 mit der Mittenfrequenz von etwa 700 Hz erfasst diesen Bereich und gibt während des „a“ eine hohe Ausgangsspannung für dieses Sprachteilband ab (Bild 7b). Gut erkennbar ist die Amplitudenmodulation des Teilbandes (Schwankungen der Spannung) mit der Sprach-Grundfrequenz, hier etwa 2,5 Rasterteile auf dem Bildschirm. Der nachgeschaltete Gleichrichter beseitigt die negativen Halbwellen (Bild 7c). Darauf folgt der Kanaltiefpass, der den Kanalkennwert liefert (Bild 7d). Der Tiefpass unterdrückt das Sprach-Teilband von 700 Hz weitgehend und die Modulation mit der Sprachgrundfrequenz etwas. Die Umhüllende des Sprach-Teilbandes aber bleibt im wesentlichen erhalten, speziell der deutliche Spannungsanstieg am Anfang des „a“. Alle anderen Spektralkanäle verhalten sich entsprechend und werden deshalb nicht dargestellt.

Die Synthese ist die Umkehrung des Analyse-Ablaufs. Das Stimmsignal Bild 7e besteht hier aus Nadelimpulsen mit einer Folgefrequenz entsprechend der Sprachgrundfrequenz, wieder etwa 2,5 Rasterteile auf dem Bildschirm. Die Impulse enthalten ein kräftiges Oberschwingungsspektrum, aus dem jedes Kanalfilter der Syntheseseite ein synthetisches Sprach-Teilband gewinnt (Bild 7f). Dieses enthält bereits die charakteristische Amplitudenmodulation mit der Sprachgrundfrequenz, läuft aber unabhängig von der Eingangssprache kontinuierlich weiter. Der nachfolgende Modulator moduliert das kontinuierliche Sprach-Teilband mit dem Kanalkennwert und liefert das modulierte Sprach-Teilband (Bild 7g), das dem der Analyse von Bild 7b weitgehend gleicht.

In einer anschließenden Additionsschaltung entsteht aus den Teilbändern aller Kanäle das vollständige, synthetische Sprachsignal (Bild 7h). Es unterscheidet sich im Oszillogramm aufgrund von Phasenverschiebungen und gewissen Ungenauigkeiten des Vocoder-Prinzips erheblich von dem ursprünglichen Sprachsignal von Bild 7a. Der Mensch jedoch versteht die synthetische Sprache gut, da die wichtigen Formanten mit genügender Genauigkeit erhalten bleiben.

Es ist ebenso möglich, die Modulatoren vor die Synthese-Filterbank zu schalten (Bild 5b). Das Stimmsignal von Bild 7i entspricht dem Bild 7e. Vor dem Wort „Bast“ steht der Stimmgenerator hier auf „Rauschen“. Der Modulator moduliert mit dem Kennwert jetzt die Amplitude des Stimmsignals (Bild 7k), und das Synthesefilter gewinnt daraus wieder das synthetische Sprach-Teilband (Bild 7l, entsprechend Bild 7g).

Das synthetische Sprachsignal von Bild 7m unterscheidet sich etwas von Bild 7h aufgrund der zufälligen Phasenlage des Stimmimpulses. Die Schaltungsanordnung nach Bild 5b ändert nichts am Grundprinzip des Kanal-Vocoders, sie ist nur etwas schwieriger zu verstehen. Dafür bietet sie aber schaltungstechnische Vorteile und wurde deshalb auch für den VSM 201 gewählt.

Bild 7. Sie Signalverarbeitung der einzelnen Baugruppen des Vocoders

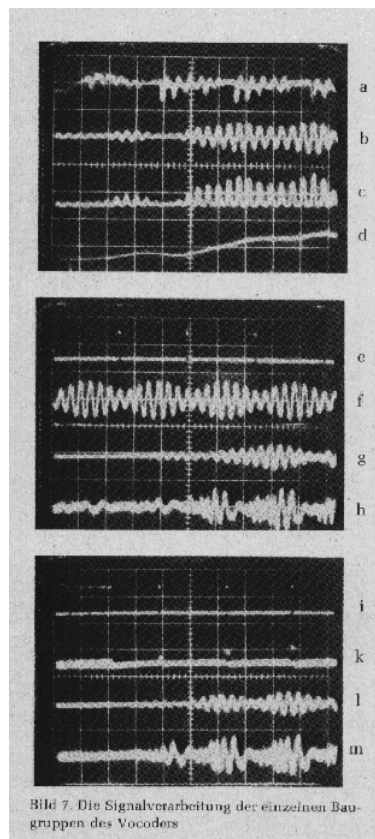


Bild 7. Die Signalverarbeitung der einzelnen Baugruppen des Vocoders

▪ Stimmartanalyse

Für die bereits genannte Stimmartanalyse gibt es ebenfalls mehrere, sehr unterschiedliche Verfahren. Einige von diesen beruhen auf einer Amplituden-Auswertung des Sprach-Spektrums, andere werten in Verbindung mit einem geeigneten Melodie-Analyseverfahren die Periodizität des Spektrums aus. Letztere sind recht aufwendig und kommen hier kaum in Betracht.

Die Stimmartanalysen mit Spektrum-Auswertung nutzen aus, dass sich die spektrale Verteilung bei den beiden Stimmarten deutlich unterscheidet. Stimmhafte Laute haben im Frequenzbereich bis 1,5 kHz immer deutlich höhere Pegel als im Bereich über 4 kHz, bei stimmlosen Lauten ist es genau umgekehrt.

Von den verschiedenen Varianten der Stimmartanalyse ist eine für den Soundeffect-Vocoder besonders gut geeignet: der Vergleich der Spektralwerte eines hohen Kanals über 5 kHz mit denen eines speziellen, tiefen Kanals bis 1 kHz. Diese Schaltung funktioniert sehr zuverlässig, wenn ein breitbandiges Sprachsignal anliegt, was in dem vorliegenden Anwendungsfall leicht erfüllbar ist.

▪ Melodieanalyse

Die Melodieanalyse bestimmt aus dem sehr komplexen Sprachsignal die jeweilige Grundfrequenz. Hierfür gibt es verschiedene Grundprinzipien mit einer größeren Zahl von Varianten.

Das bekannteste Verfahren beruht auf der direkten Aussiebung der Grundwelle mit einem Tiefpass. Aufgrund der verschiedenartigen Übergänge zwischen den einzelnen Lauten kommt es bei allen Verfahren gelegentlich zu fehlerhaften Messungen der Grundfrequenz, was die Sprachqualität der synthetischen Sprache erheblich mindert. Daher sind immer zusätzliche Logikschaltungen erforderlich, die alle falschen Messergebnisse erkennen und korrigieren sollen.

Für einen Klangeffekt-Vocoder ist eine Melodieanalyse nur in Sonderfällen erforderlich. Aus diesem Grunde soll die Schaltungstechnik hier nicht weiter behandelt werden.

▪ Vocoder-Anwendungen

Die nachfolgende Aufzählung einiger typischer Einsatzfälle soll die Möglichkeiten des Vocoder-Prinzips weiter verdeutlichen und die bisherigen Ausführungen ergänzen. Es folgen deshalb auch nur sehr kurze Erläuterungen zu den einzelnen Beispielen.

Vocoder für wissenschaftliche Forschungszwecke

Mit Vocodern konnten viele wissenschaftliche Erkenntnisse auf dem Gebiet der menschlichen Artikulation und Stimmerzeugung gewonnen werden. Selbst auf dem Gebiet der Hörphysiologie wurden Vocoder eingesetzt.

▪ Sprechende Computer

Computer speichern alle Informationen nur in digitaler Form. Die digitalisierten Sprachkennwerte eines Vocoders erfordern nur 5% bis 10% der Speicherplätze, die ein direkt digitalisiertes Sprachsignal (PCM) benötigt, etwa 1200 bis 2400 Speicherplätze für eine Sekunde Sprachwiedergabe.

▪ Abhörsichere Sprachübertragung

Abhörsichere Verschlüsselungs-Geräte können nur digitale Daten verschlüsseln. Die Bitraten (Datenmenge pro Zeiteinheit) der direkt digitalisierten Sprache sind für die verschlüsselte Datenübertragung auf normalen Telefonleitungen und auf Kurzwellen-Funkstrecken viel zu hoch, nicht aber die Bitraten von 1200 bis 2400 bit/s der digitalisierten Sprachkennwerte eines Vocoders.

- Umwandlung der Helium-Sprache
 Die Stimmen von Tauchern in Tauchglocken bei Helium-Atmosphäre klingen unnatürlich wie eine "Donald-Duck-Stimme" und sind daher schlecht zu verstehen. Infolge des geringeren spezifischen Gewichtes des Heliums gegenüber der Luft steigt die Grundtonhöhe der Stimme an, die spektrale Verteilung der Laute ändert sich aber nur wenig. Durch Zwischenschalten eines Vocoders mit geeigneter Umcodierung des Melodiekenntwertes lässt sich der Effect beseitigen, und die Taucherstimme klingt wesentlich natürlicher.
 - Sprachverfremdung und Klangeffekte
 Bei Verwendung eines externen Stimmsignals aus verschiedenartigen Signalquellen sowie durch zusätzliche Umcodierung der Sprachkenntwerte lassen sich viele interessante Effekte erzeugen, die dramaturgisch und musikalisch verwendbar sind. Hierfür ist der Soundeffect-Vocoder VSM 201 speziell konzipiert.
- Für die hier aufgezählten Anwendungsfälle kommen grundsätzlich alle der bekannten Vocoder-Verfahren in Betracht. Zur Erzeugung von Klangeffekten ist jedoch der Kanal-Vocoder besonders gut geeignet und deshalb sollen die anderen, zum Teil sehr interessanten Vocoder-Verfahren hier unerwähnt bleiben.
- Die bisherigen ausführlichen Erklärungen stellen allgemeine Grundlagen dar, die jeder Benutzer eines Klangeffekt-Vocoders kennen sollte, um die Arbeitsweise und die Möglichkeiten seines Gerätes richtig beurteilen zu können. (Schluss folgt)

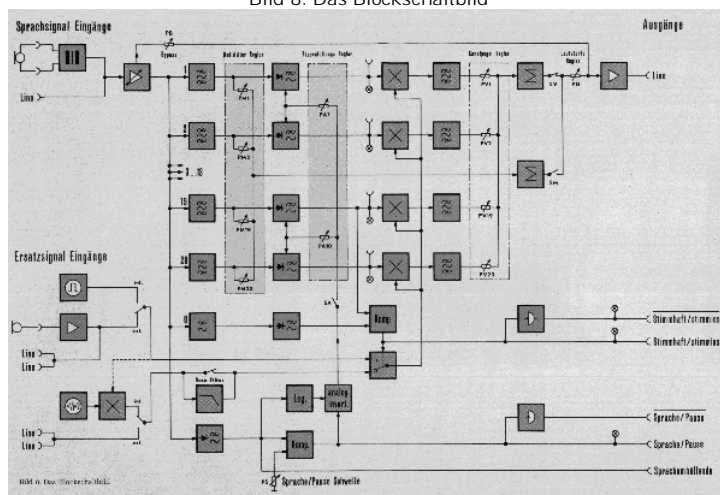
Im vorliegenden zweiten Teil des Beitrags beschreibt der Autor die Bedienung des Vocoders anhand des Blockschaltbildes.

Published in FUNKSCHAU 1978, Heft 8
by Dipl.-Ing. Dietrich Buder

Vocoder für Sprachverfremdung und Klangeffekte, Teil 2.

- Klangeffekte
 Mit dem Sennheiser-Soundeffect-Vocoder wurde bereits eine Reihe von sehr wirkungsvollen Effekten erprobt, und es ist zu erwarten, dass kreative Künstler und Techniker noch weitere Varianten entdecken. Die wichtigsten bereits erprobten Effekte sind:
- a) Sprache oder Gesang von extrem tiefen bis zu sehr hohen Stimmlagen aus gesprochenem Text, zusätzlich mit ungewöhnlichem Vibrato oder verfremdeter Sprachmelodie.
- b) Heisere Flüstersprache aus Gesang oder normaler Sprache.
- c) Sprechende monophone oder polyphone Musikinstrumente.
- d) Ersatz einer menschlichen Stimme durch eine andere.
- e) Mehrstimmiger polyphoner Gesang aus gesprochenem Text von einer einzigen Stimme.
- f) Formantverschiebungen (Donald-Duck-Effekte) in Echtzeit.
- g) Stimmen mit unharmonischen Oberwellen-Spektren.
- h) Artikulation natürlicher Geräusche wie sprechender Wind, schimpfender Motor, mahrende Glocken, Artikulation einer tosenden Fußball-Zuschauermenge.
- i) Queue von sprachabhängigen Steuersignalen für externe Geräte wie Synthesizer, Lichtorgel o.ä.
- k) Verwendung des Vocoders als Multifilter zur Einstellung eines weitgehend frei wählbaren Frequenzganges.
- 1) Bisher nicht produzierte Instrumental-Effekte wie getrommelte Orgel oder ähnliches (Multimodulation).

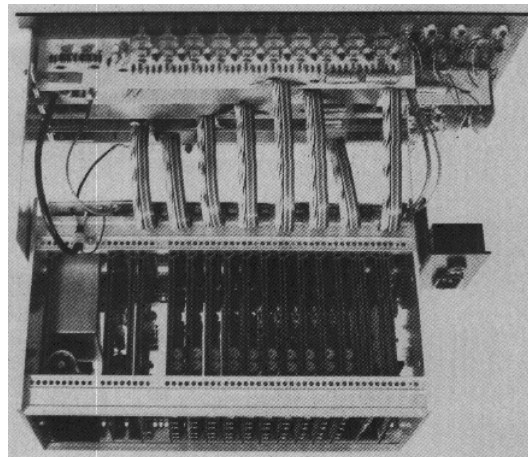
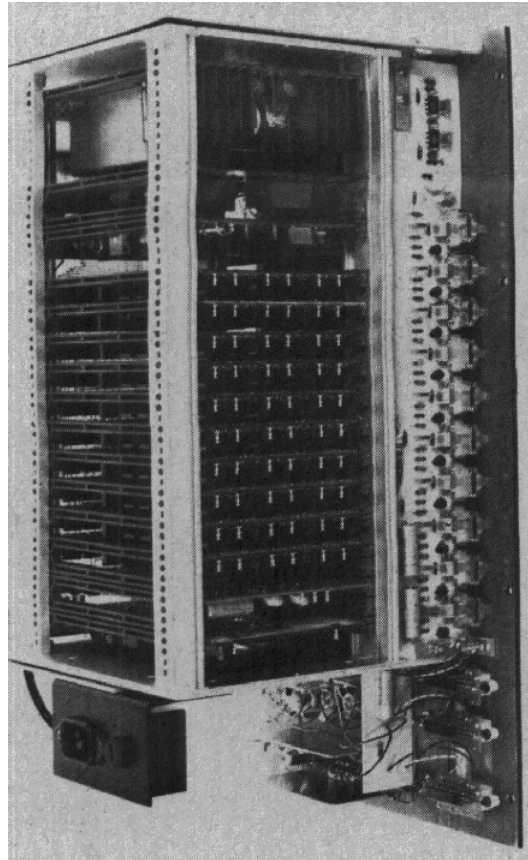
Bild 8. Das Blockschaltbild



- Die Entstehung der angeführten Effekte ist bei Kenntnis des Vocoder-Prinzips nun leicht zu durchschauen. In den Beispielen 8 a bis h handelt es sich um die typische Sprachverfremdung. Es erfolgt eine Einspeisung des eigentlichen Sprachsignals in den Mikrofon- oder Sprachsignal-Eingang des Vocoders. Das Stimmsignal jedoch liefert die sogenannte Ersatzquelle, z. B. ein Synthesizer bei den Beispielen 8 a, b und g, ein Synthesizer in Verbindung mit einem Frequenz-Shifter bei e und f, ein Bandgerät bei h oder ein zweites Mikrofon bei c und d.
- Nur im Beispiel 1 ohne Sprachanteile kommt das Signal des einen Instruments an den Spracheingang, das des anderen, z. B. die Orgel, an den Ersatzsignal-Eingang.
- Bei dem Beispiel k arbeitet der eigentliche Vocoder nicht mit, das Eingangssignal gelangt von der Filterbank der Analyse direkt an den Ausgang des Vocoders; bei i wird nur der Analyseteil des Vocoders verwendet.

Bei der Erzeugung aller Sprachverfremdungs-Effekte ist eine Besonderheit des Vocoders zu beachten: Die Ersatzsignale für stimmhafte und stimmlose Laute müssen breitbandig sein, also Frequenzanteile über den ganzen Sprachbereich enthalten. Wenn diese Forderung nicht erfüllt ist, fehlen bestimmte Frequenzbereiche in der synthetischen Sprache und das Ergebnis entspricht nicht den Wünschen des Benutzers. Um es drastisch zu formulieren: Ein einzelner Sinus-Ton ist als Stimmsignal völlig ungeeignet. Das Ersatzsignal für stimmlose Laute, normalerweise ein Rauschspektrum, sollte hinsichtlich Lautstärke und Frequenzgang dem Ersatzsignal für stimmhafte Laute näherungsweise entsprechen. Für den Fall, dass kein geeignetes externes Rauschsignal zur Verfügung steht, kann der interne Rauschgenerator eine dem stimmhaften Ersatzsignal angepasste Rauschspannung liefern.

Bild 9/9a. Der Haugruppenträger in Normalstellung und aufgeklappt



- Bedienung des Vocoders VSM 201

Die Bedienung des Vocoders VSM 201 lässt sich am leichtesten an Hand des vereinfachten Blockschaltbildes erläutern (Bild 8). Um die Übersichtlichkeit zu gewährleisten sind hier verschiedene Schalter, Regler und LED (Leuchtdioden) nicht eingezeichnet.

- Regelverstärker

Die zu verfremdende Sprache gelangt vom Mikrofon oder über einen + 6-dB-Line-Eingang zunächst an einen Regelverstärker, der einen normierten Sprachpegel liefert. Infolgedessen braucht sich der Sprecher nicht darauf zu konzentrieren, immer mit gleicher Lautstärke zu sprechen oder zu singen.

Zur optischen Überwachung des Sprachpegels dienen ausserdem eine LED-Zeile und eine weitere LED, die das Herabregeln des Regelverstärkers anzeigt. Das Einregeln eines normierten Sprachpegels ist erforderlich, um die maximale Wiedergabedynamik des Vocoders von mehr als 60 dB zu erzielen. Für eventuelle Sonderfälle lässt sich der Regelteil des Regelverstärkers abschalten; es leuchtet dann eine rote LED zur Warnung

- **Spektralkanäle**
 Die 20 Spektral-Kanäle sind in herkömmlicher Weise aufgebaut und angeordnet: Bandfilter, Gleichrichter, Tiefpaß, Modulator, Bandfilter und Summierer. Hinter den Ausgängen der syntheseseitigen Bandfilter liegen die 20 Kanalpegelsteller PV 1 bis PV 20 für die Frequenzgang-Korrektur der Vocoder-Sprache. Ein individuelle Einstellung ist immer dann notwendig, wenn das Einsatzsignal stark überbetonte Anteile im Frequenzgang hat und dadurch bestimmte Bereiche des synthetischen Sprachsignals im Pegel nicht stimmen. Ebenso dienen die Einsteller aber auch zur Korrektur des Frequenzgangs der Eingangssprache. Beim Einstellen ist es oft wünschenswert, zwischen der Korrektur-Einstellung und der Grund- bzw. Normal-Stellung der Regler vergleichen zu können. Dieses ermöglicht ein spezieller Umschalter; zwei LED zeigen die jeweilige Schalterstellung an. Dieser Umschalter hat zusätzlich eine Mittelstellung zum völligen Abschalten der Vocoder-Synthese. Diese Schaltmöglichkeit ist im Blockschaltbild dargestellt als Schalter SV zwischen dem Summierer und dem Endverstärker.
- **Sprach -Addition oder Multifilter**
 An den Ausgängen der Analyse-Filter liegen die 20 Multifiltersteller PM 1 bis PM 20 für eine frequenzabhängige Zumischung der direkten Sprache zu der synthetischen Sprache. Sofern der Schalter SV des Syntheseteils geöffnet und der Schalter SM der Sprachzumischung geschlossen ist, arbeitet das Gerät als Multifilter. Auch hier ist ein Umschalter vorhanden, der einen Vergleich zwischen dem korrigierten Frequenzgang und dem normalen Frequenzgang ermöglicht. Zwei LED Signalisieren die jeweils eingeschaltete Sprachzumischung. Die Mittelstellung dieses Umschalters ist die, "Aus"-Stellung des dargestellten Schalters SM.
- **Kanal -Kennwerte**
 Allen 20 Kanal-Kennwerten ist je eine LED zugeordnet. Diese beginnt schwach zu leuchten, wenn der betreffende Kanal einen Pegel überschreitet, der etwa 20 dB unter der Vollaussteuerung des betreffenden Kanals liegt. Die Spannungen der Kanal-Kennwerte sind ausserdem an Ausgangs-Buchsen geführt, um sie für beliebige Zwecke zur Verfügung zu haben, z. B. zur Steuerung eines Synthesizers oder einer Lichtorgel.
- **Stimmart-Analyse**
 Die bereits beschriebene Stimmart-Analyse-Schaltung enthält den zusätzlichen Spektralkanal 0 (Null) für den Bereich bis etwa 1 kHz. Ein Komparator vergleicht die Kennwerte der Kanäle 0 und 19, und sein Ausgang steuert den Stimmsignal-Umschalter. Das jeweilige Ergebnis der Stimmart-Analyse zeigen zwei LED an, ebenso sind zwei Ausgangsbuchsen für beide Schaltzustände vorhanden.

 Durch einen speziellen Umschalter lässt sich der Stimmsignal-Umschalter in beiden Stellungen blockieren.
- **Stimmsignale**
 Üblicherweise arbeitet ein Klang Effekt-Vocoder immer mit einem externen Ersatzsignal für die stimmhaften Laute. Der interne Impulsgenerator mit fester Folgefrequenz ist nur für Testzwecke gedacht. Der interne Rauschgenerator dagegen eignet sich gut zur Erzeugung des stimmlosen Ersatzsignals. Seine Ausgangsspannung gelangt zu einem Modulator (VCA), der den Pegel des Rauschens automatisch der Spannung des stimmhaften Ersatzsignals anpasst. Das Spektrum des Rauschens lässt sich ausserdem durch ein zuschaltbares "Rosa-Filter" verändern und damit auch im Frequenzgang dem stimmhaften Ersatzsignal angleichen. Selbstverständlich kann auch ein externes Signal für die Erzeugung der stimmlosen Laute eingespeist werden. Zur Pegelkontrolle der Ersatzsignale sind zwei LED-Zeilen vorgesehen.
- **Sprache-/Pause-Detektor und Sprach-Umhüllende**
 Für die verschiedenen Steuer- und Schaltzwecke dienen der Lautstärke-Kennwert, auch Sprach-Umhüllende genannt, und das Sprache-/Pause-Signal des Sprache-/Pause-Detektors. Im Blockschaltbild sind die dafür erforderlichen Funktionseinheiten eingezeichnet. Das Sprachsignal läuft über einen Gleichrichter mit Tiefpass. Die Tiefpass-Ausgangsspannung ist der Lautstärke-Kennwert. Der Lautstärke-Kennwert gelangt ausserdem an einen Komparator, der das Sprache-/Pause-Signal in beiden Schaltzuständen liefert. Die gewünschte Schaltschwelle ist durch den Einsteller PS wählbar. Eine am Ausgang angeschlossene LED zeigt die jeweilige Stellung des Sprache-/Pause-Detektors an.
- **Pausen-Auffüllung**
 Die menschliche Sprache besteht häufig aus relativ kurzen Silben und ziemlich langen Pausen zwischen den Silben und Worten. Für eine Artikulation von natürlichen Geräuschen (Effekt h) genügt es nicht, nur das betreffende Geräuschsignal an den Ersatzsignal-Eingang anzuschalten. Dabei entsteht lediglich eine eigenartig klingende Sprache, nicht jedoch der beabsichtigte dramaturgische Effekt, weil das Ersatzsignal aufgrund der relativ häufigen Silbepausen kaum erkennbar ist.

 Im Vocoder VSM 201 ist deshalb eine Pausen-Auffüllung (silence bridging) vorhanden, die das Ersatzsignal in den Pausen zwischen den Silben und Worten hörbar macht. Dabei läuft das Ersatzsignal während der Pausen auch über den Syntheseteil des Vocoders. Somit sind die synthetische Sprache und das Pausenfüllsignal hinsichtlich Frequenzgang und Phasengang optimal angenähert.

 Eine direkte Wiedergabe des Ersatzsignals in den Silbepausen ergibt meist nicht den gewünschten Effekt, sondern es klingt so, als ob er eine eigenartige Stimme vor einem bestimmten Hintergrundgeräusch spricht.

 Das Blockschaltbild zeigt die gewählte Schaltung mit einem Logarithmierer und Inverter. Das Pausen-Auffüllen geschieht durch Addition eines Auffüll-Kennwerts zu den Kanal-Kennwerten. Zur individuellen Frequenzgang-Korrektur dienen die Einsteller PA 1 bis PA 10, wobei ein Potentiometer jeweils zwei benachbarte Kanäle beeinflusst.

 Wenn der Lautstärke-Kennwert einen bestimmten, hohen Pegel überschreitet, erreicht der Auffüll-Kennwert die Spannung Null und bleibt dort auch bei Erhöhen der Eingangsspannung stehen. Andererseits stellt sich der Auffüll-Kennwert auf den eingestellten maximalen Pegel ein, wenn der Sprache-/Pause-Detektor auf "Pause" schaltet. Diese Verknüpfung ist im Blockschaltbild eingezeichnet. Die richtige Einstellung aller Steller und eine geschickte Wahl der externen Signalspannungen erfordern eine gewisse Sorgfalt und bereiten etwas mehr Mühe als die Bedienung der anderen Vocoder-Funktionen. Eine genauere Kenntnis über das Zusammenspiel der einzelnen Funktionsblöcke ist hier nützlich.
- **Ausgang und Bypass**
 Der Vocoder hat einen Ausgang Lautstärkesteller PG und zwei Ausgangsbuchsen, eine mit einem Pegel von +6dB und eine mit -20dB. Schliesslich ist mit dem Einsteller PB auch noch ein abschaltbarer Bypass vorhanden, der für Kontrollzwecke eine direkte Verbindung zwischen Spracheingang und Endverstärker ermöglicht.

Bild 10. Die Kanalkarte (zwei Kanäle)

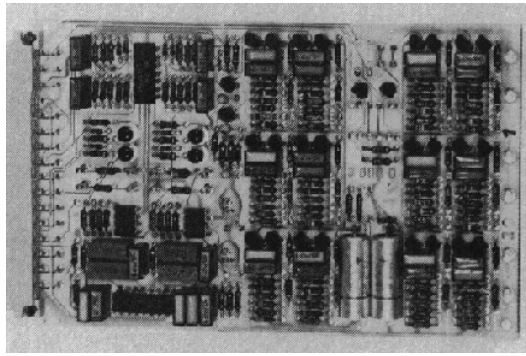
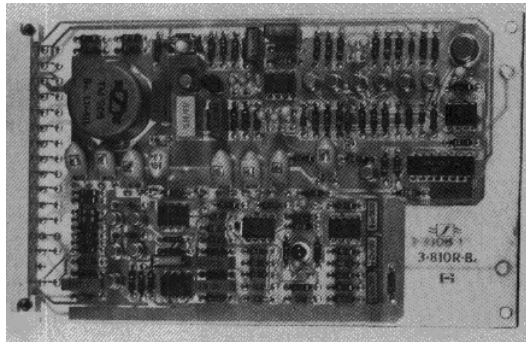


Bild 11. Der Eingangsverstärker mit Stimmart-Detektor



- Kontruktion

Sennheiser stellt mit dem Sound-Effect-Vocoder VSM 201 einen vielseitig einsetzbaren Zusatzbaustein vor, der für Tonstudios, Film- und Fernseh-Produktionen, avantgardistische Musikgruppen und experimentierfreudige Alleinunterhalter vorgesehen ist. Als Quelle für die Ersatzsignale eignen sich elektronische Instrumente und Synthesizer aller Fabrikate.

Das Äussere des Vocoders ist den bekannten Moog-Studio-Synthesizer angepasst. So entsprechen die Abmessungen von Gehäuse und Frontplatte den Massen eines Moog-Bausteins, und auch das Design ist sehr ähnlich. Die Frontplatte passt mit ihren Befestigungslöchern in die Nussbaum-Gehäuse der Moog Studiobausteine. Das mit kunstleder bezogene Gehäuse ist für einen Bühneneinsatz geeignet, es hat einen Tragegriff und vorgezogene zum Schutz der Bedienungs-rente. Ein spezieller Holzkoffer mit Schaumstoffpolsterung für den Transport gehört zum Lieferumfang.

Die Konstruktion ist sehr servicefreundlich. Der eigentliche Vocoder befindet sich in einem Baugruppenträger auf Steckkarten im Europaformat 160mm x 100mm. Die Einsteller mit dem zugehörigen Bauelementen sind auf speziellen Karten hinter der Frontservicefreundlich angebracht.

Die Bedienungselemente sind auf der Frontplatte nach logischen Gesichtspunkten angeordnet. Die beiden Signaleingänge mit den zugehörigen Schaltern und Anzeileuchten befinden sich auf den zwei senkrechten Feldern links auf der Frontplatte. Die Bedienungselemente für die Verarbeitung und die Ausgänge sind auf den grossen, waagerechten Feldern an der rechten Seite untergebracht. Die ausführliche Beschriftung gestattet den Betrieb des Vocoders auch ohne Benutzung eines Handbuches. Lediglich die richtige Einstellung der Pausenauffüllung setzt die Kenntnis der Beschriftung voraus. Bilder 9 bis 11 zeigen die gewählte mechanische Konstruktion.