EIN SCHÄRFE-ORIENTIERTES VOCODERSYSTEM

H. Knebel

Institut für Elektroakustik, Technische Universität München

1. EINLEITUNG

Die Reduktion der Datenrate von Sprachsignalen mittels Vocodersystemen (Voice-Coder) geht auf eine Entwicklung von H.W.Dudley im Jahre 1939 zurück. Inzwischen gibt es eine Vielzahl von Verfahren, die alle auf dem gleichen Grundprinzip beruhen; der Trennung zwischen Anregungsfunktion und Übertragungsverhalten des Sprechtraktes. Ein Verfahren, das sich zur Gewinnung der Vocoderparameter völlig am Empfänger der Schallsignale, dem menschlichen Gehör, orientiert, soll im folgenden vorgestellt werden.

2. BESTIMMUNG DER VOCODERPARAMETER AUS EMPFINDUNGSGRÖSSEN

Grundlage für das neue Vocoderverfahren ist folgende Überlegung: Die Analyse des Sprachschalles durch das menschliche Gehör führt zu entsprechenden Empfindungsgrößen. Durch Synthese von Empfindungsgrößen, die denjenigen des ursprünglichen Signals entsprechen, läßt sich im Gehör wieder die gleiche Wahrnehmung hervorrufen. Das Verfahren soll also nicht eine möglichst exakte physikalische Kopie des Originalschalles liefern, sondern ein so einfach wie möglich strukturiertes Signal, das aber zur gleichen Hörempfindung führt.

Zur Charakterisierung des Sprachsignals werden das Klangfarbenattribut Schärfe, die Lautheit und die Tonheit verwendet. Die Tonheit wird in diesem Fall grob vereinfachend durch die Grundfrequenz beschrieben. Ein Funktionsmodell zur Gewinnung der Lautheit stark zeitvarianter Schalle wurde von Zwicker [1] angegeben. Funktionsschema bzw. Modell zur Bestimmung der Schärfeparameter von Bismarck [2] und Knebel [3]. Fig. 1 zeigt den prinzipiellen Aufbau des am Gehör orientierten Vocodersystems.

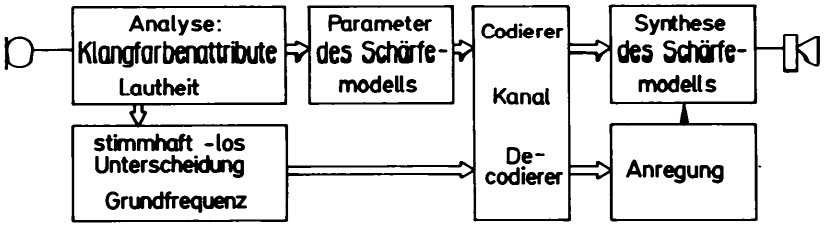

Fig.1. Schärfe-orientiertes Vocodersystem.

In der linken Bildhälfte ist der Analyseteil und in der rechten Hälfte der Syntheseteil dargestellt. Des weiteren wird zwischen den Parametern der spektralen Hüllkurve im oberen Teil und denen der Anregungsfunktion im unteren Teil unterschieden. Im ersten Block, "Analyse", werden zur Beschreibung des Klangfarbenattributes Schärfe drei Schwerpunktstonheiten und die dazu gehörigen Schwerpunktslautheiten bestimmt. Die Empfindungsgröße Lautheit wird ebenfalls gebildet. Einzelheiten sind an anderer Stelle ausführlich beschrieben worden [4], [5].

Im zweiten Block, "Parameter des Schärfemodells", wird die so einfach wie möglich strukturierte spektrale Verteilung berechnet, die bei der Synthese zu denselben Parametern des Klangfarbenattributes Schärfe führt, die auch für die Analyse gelten. Die richtige Wiedergabe der Lautheitsempfindung wird dabei ebenfalls berücksichtigt.

Die einfachste Struktur, mit der die spektralen Eigenschaften der Schärfeempfindung angenähert werden können, ist ein Satz von drei variablen Bandpaßfiltern. Die Berechnung der Parameter dieser Filterstruktur erfolgt, angepaßt an die Empfindungsgrößen im Lautheits-Tonheits-Bereich. Erst nach der Übertragung werden daraus die zu der Steuerung der Synthesefilter nötigen Koeffizienten berechnet.

Die stimmhaft-stimmlos Unterscheidung zum Umschalten der Anregungsfunktion stützt sich bei diesem System ebenfalls auf Empfindungsgrößen. Da stimmlose Sprachanteile schärfer empfunden werden als stimmhafte, genügt eine Schwellwertentscheidung im Schärfemodell, um beide Lautgruppen zu trennen.

Bei diesem experimentellen Vocoder wird die Grundfrequenz auf sehr einfache Weise, nämlich durch Herausfiltern der Grundfrequenz aus dem Sprachsignal, gewonnen. Die bei diesem Verfahren auftretenden Einschränkungen bezüglich Variationsbereich und Genauigkeit werden wegen der einfachen Realisierbarkeit zunächst in Kauf genommen. Genauere Verfahren zur Bestimmung der Tonheit werden von Terhardt angegeben [6].

3. TECHNISCHE REALISIERUNG

Das Vocodersystem soll in Echtzeit arbeiten. Deshalb wurden wesentliche Teile, nämlich die Vorverarbeitungssysteme und das Synthesefilter in analoger Technik erstellt. Digital arbeitet nur der Mikroprozessor zur Nachbildung des Schärfemodells.

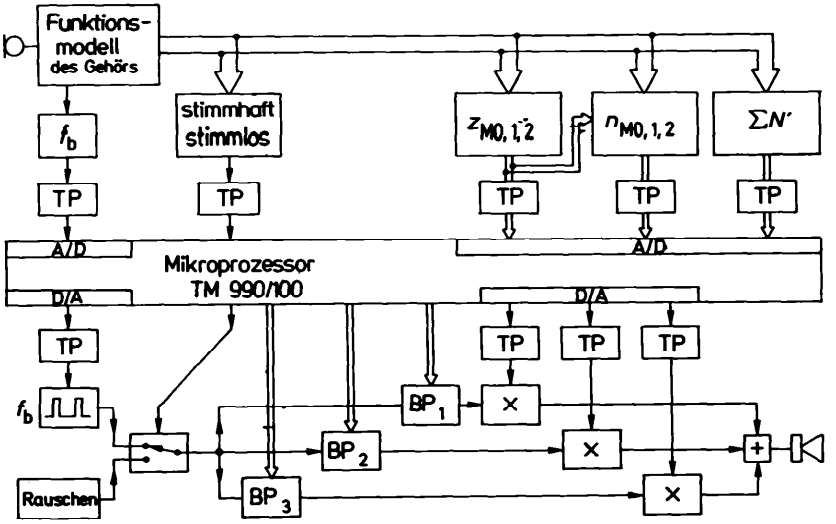

Fig.2. Technisch realisiertes Vocodersystem.

In der oberen Hälfte der Figur ist der Analyseteil des Vocodersystems dargestellt. Ausgangspunkt ist ein Funktionsmodell des Gehörs nach Zwicker [1]. Es transformiert die Schallereignisse in entsprechende Lautheits-Tonheitsmuster. Daraus werden in analog arbeitenden Schaltungen die drei Schwerpunktstonheiten $z_{M0,1,2}$, die dazugehörigen Schwerpunktslautheiten $n_{M0,1,2}$ sowie die stimmhaft-stimmlos Unterscheidung berechnet [3], [5], [7]. Diese zeitvariablen, analogen Größen werden, dem Abtasttheorem entsprechend, tiefpaßgefiltert und alle 7,33 msec in den Rechner eingelesen. Der Mikroprozessor, ein TM 990/100, berechnet, entsprechend dem Schärfemodell, die auf die Lautheits-Tonheits-Verteilung bezogenen Filterparameter. Diese werden dreimal gemittelt, so daß alle 22 msec neue Werte an das Synthesesystem übergeben werden.

Bei räumlicher Trennung von Analyse- und Syntheseteil wird an dieser Stelle die Übertragungsstrecke eingefügt. Im vorliegenden Fall übernimmt jedoch derselbe Mikroprozessor die Aufgabe, die Größen in entsprechende Steuerparameter des technischen Filtersystems umzurechnen. In der unteren Hälfte der Figur 2 ist der Sprachsyntheseteil des Vocoders dargestellt [5]. Das entweder von einem Impulsgenerator mit der Grundfrequenz f_h, oder von einem Rauschgenerator erzeugte ebene Spektrum wird durch drei parallele Bandpässe gefiltert. Mittenfrequenz und Bandbreite eines jeden Filters werden vom Mikroprozessor eingestellt. Die Flankensteilheit der verwendeten Filter beträgt 48 dB/Okt, ein Mittelwert der Steilheit von oberer und unterer Flanke einer entsprechenden Mithörschwelle. Als zusätzliche Randbedingung gilt, daß die Durchlaßbereiche der drei Bandfilter lückenlos aneinander anschließen. Die Verstärkungen der drei Filterkanäle werden ebenfalls vom Rechner gesteuert.

4. HÖRVERSUCHE

Anhand des Freiburger Sprachverständnistests für Einzelworte (DIN 45 621) wurde die Leistungsfähigkeit dieses Verfahrens mit acht Versuchspersonen untersucht. Zum Vergleich mit anderen Systemen beurteilten die Versuchspersonen auch die Synthese-Ergebnisse eines Kanal- und eines LPC-Vocoders für dasselbe Sprachmaterial. Die synthetisierten Worte der einzelnen Vocoder wurden auf gleiche mittlere Lautheit eingestellt (20 sone) und beidohrig über den Kopfhörer DT 48 mit Freifeldentzerrer nach [8] dargeboten. Die Anzahl der vollständig richtig erkannten Einzelworte, angegeben in Prozent, ist in Fig. 3 dargestellt.

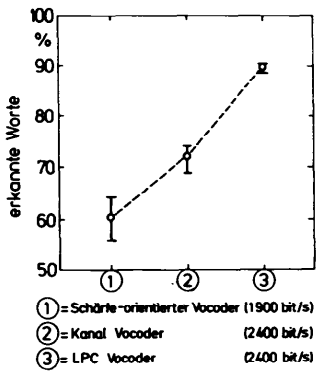

Fig. 3. Ergebnisse der Hörversuche mit dem Freiburger Sprachverständnistest.

Die mit dem vorgestellten, stark vereinfachten Verfahren analysierten und synthetisierten Worte wurden von den Versuchspersonen in 60 % der Fälle richtig erkannt. In Anbetracht der Tatsache, daß die Schärfe nur einen bestimmten Teil der Klangfarbe erfaßt (44 % Varianz) [2], ist dieses Ergebnis als recht brauchbar einzustufen. Durch die Implementierung zusätzlicher Klangfarbenparameter läßt sich vermutlich sowohl die Verständlichkeit als auch die Natürlichkeit der synthetisierten Worte deutlich verbessern.

LITERATUR

[1] Zwicker, E., Procedure for calculating loudness of temporally variable sounds. J.Acoust.Soc.Am. 62, 675 - 682 (1977).

[2] v. Bismarck, G., Extraktion und Messung von Merkmalen der Klangfarbenwahrnehmung stationärer Schalle. Dissertation Technische Universität München (1972).

[3] Knebel, H., Ein elektronisches Modell zur Bildung von Tonheitsmomenten der spezifischen Lautheit. In: Fortschritte der Akustik, DAGA '78, VDE-Verlag, Berlin, 443 - 446 (1978).

[4] Zwicker, E., Terhardt, E., Paulus, E., Automatic speech recognition using psychoacoustic models. J.Acoust.Soc.Am. 65, 487 - 498 (1979).

[5] Knebel, H., Extraktion sprachbeschreibender Parameter aus Lautheits-Tonheitsmustern. In: Fortschritte der Akustik, DAGA '80, VDE-Verlag, Berlin, 671 - 674 (1980).

[6] Terhardt, E., Sprachgrundfrequenz-Extraktion nach Prinzipien der Tonhöhenwahrnehmung. In: Fortschritte der Akustik, DAGA '80, VDE-Verlag, Berlin 667 - 670 (1980).

[7] v.Bismarck, G., Vorschlag für ein einfaches Verfahren zur Klassifikation stationärer Sprachschalle. Acustica 28, 186 - 188 (1973).

[8] Zwicker, E., Feldtkeller, R., Das Ohr als Nachrichtenempfänger. 2. erw. Aufl., Hirzel-Verlag, Stuttgart (1967).

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Sonderforschungsbereiches 50 "Kybernetik" gefördert.