REALFORMANTEN IN DER AUTOMATISCHEN SPRACHERKENNUNG
ERSTE ERGEBNISSE DER EXTREMWERTANALYSE IM ZEITBEREICH

Patrick Schweisthal, Thomas Schweisthal, Walter Kopetzky

Arbeitskreis Spracherkennung Sprachgenerierung und Phonetische Datenbanken
der Gesellschaft für linguistische Datenverarbeitung (GLDV)
Institut für Phonetik und sprachliche Kommunikation
Universität München, Schellingstr. 3, D-8000 München 40

Im LDV-Forum 5(1987)1 vom Juni 1987 wurde die Extremwertanalyse im Zeitbereich als ein Verfahren der akustischen Sprachschallanalyse vorgestellt. Darauf aufbauend werden nun die ersten Schritte in Richtung auf eine Transkriptionsmaschine als phonetischer Beitrag zur Spracherkennung dokumentiert. Den Autoren, Mitgliedern des GLDV-Arbeitskreises "Spracherkennung Sprachgenerierung und Phonetische Datenbanken" ist es gelungen, analysierte Realformanten im Bild des Intervallogramms so darzustellen, daß sie mit entsprechenden herkömmlichen Spektralformanten des Sonagramms verglichen werden können. Hierbei wird deutlich sichtbar: Die Strukturen der Realformanten sind im Nasalbereich genau so klar erkennbar wie im Vokalbereich. Dies gilt auch für die Analyse von Frauenstimmen.

In unserem letzten Beitrag postulierten wir analog zu den Formanten des Frequenzspektrums die sogenannten Realformanten (RF). Als Realformant bezeichnen wir die Konzentration zeitlicher Extremwertabstände vergleichbarer Dauern. Im Gegensatz zur herkömmlichen linearen Filterung beläßt ein symmetrisches Glättungsverfahren die Extremwerte in ihrer ursprünglichen Lage. Die zeitlichen Dauerverhältnisse bleiben dadurch erhalten. Dies ist Voraussetzung für die Bestimmung der RF. Wird das digitalisierte Signal weniger stark geglättet, so ergeben sich Konzentrationen kurzer Zeitdauern der Extremwerte: die sogenannten RF 2.

Wird das digitalisierte Signal stark geglättet, so ergeben sich Konzentrationen langer Zeitdauern der Extremwerte: die sogenannten RF 1.

Die folgende Darstellung bezeichnet man als Intervallogramm. Vertikal ist der Kehrwert der zeitlichen Dauerverhältnisse angetragen (in Hz bei logarithmischem Maßstab), horizontal die zeitliche Dauer der Äußerung (linearer Maßstab, Gesamtdauer einer Darstellung 1,2 s).

Die Graphik enthält den ersten Realformantenormanten (dargestellt von 200 bis 1400 Hz Extremwertfrequenz), sowie den zweiten Realformanten (von 1400 bis 20.000 Hz). Oberhalb des Intervallogramms befindet sich jeweils das Oszillogramm der Äußerung.

Die drei Intervallogramme stellen die Äußerung: "Unser Programm" dreier unterschiedlicher Versuchspersonen (A, B und C) dar, darunter eine weibliche (A) und zwei männliche (B und C).

Zum Vergleich ist neben jedem Intervallogramm das entspechende Sonagramm derselben Äußerung abgebildet (allerdings bei linearem Frequenzmaßstab, von 0 bis 6000 Hz auf der Vertikalachse).

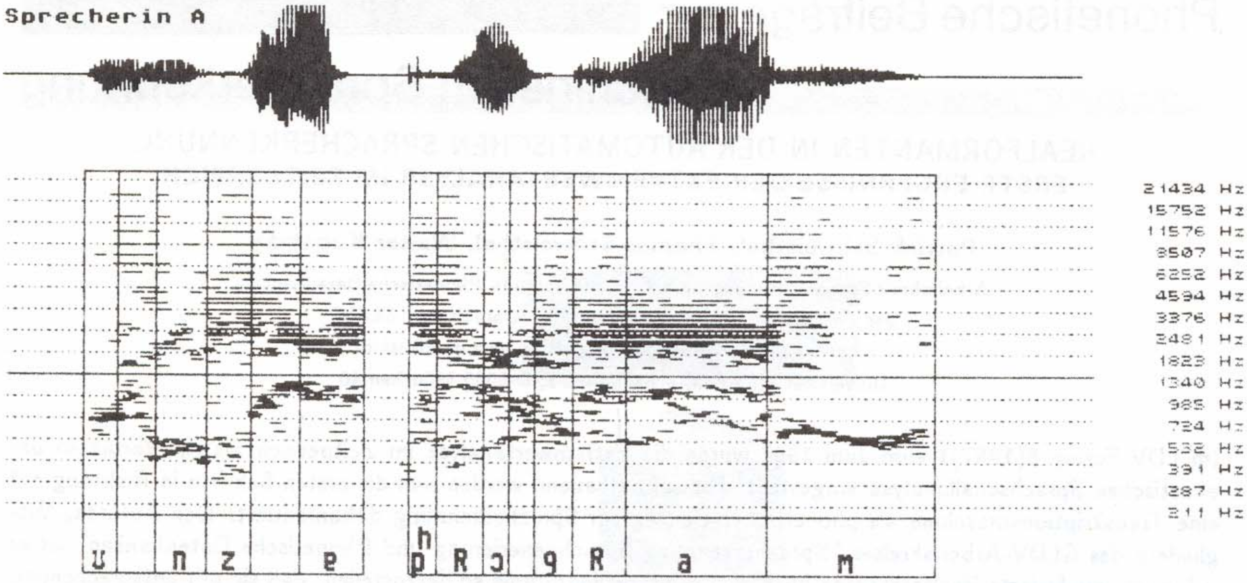
Beim Vergleich ist zu beachten, daß die Frequnenzangaben in Sonagramm und Intervallogramm unterschiedliche Bedeutung haben: Das Sonagramm markiert die Schwingungsanteile (der einzelnen Frequenzen) des Sprachsignals, während das Intervallogramm die zeitliche Extremwertdichte ("Extremwertfrequenz") der Signalkurve bezeichnet.

Die Intervallogramme lassen Realformantverläufe erkennen, die sich mit Spektralformantverläufen im Sonagramm vergleichen lassen. Ausgehend von den Realformantverläufen haben wir die Intervallogramme segmentiert und den einzelnen Segmenten phonetische Transkriptionszeichen zugeordnet.

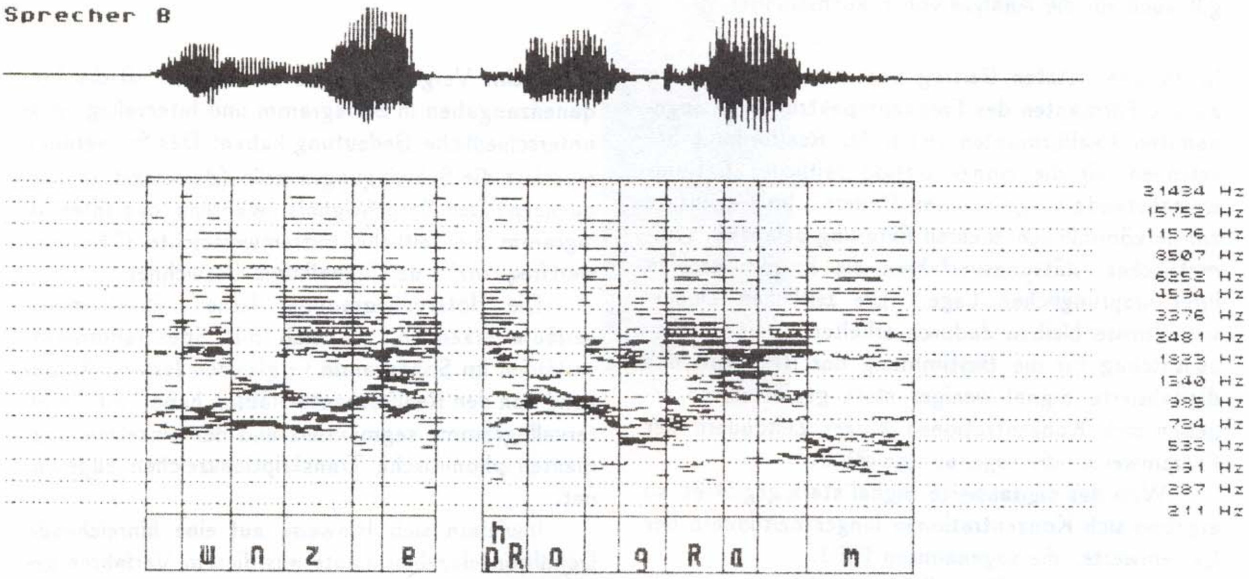
Inwiefern sich Hinweise auf eine hinreichende Definition einzelner Laute aus diesem Verfahren gewinnen lassen, muß sich bei weiteren Untersuchungen erweisen.

Zumindest eine objektive Transkription zeitlicher Segmente sollte bei geeigneten statistischen Verfahren möglich sein. Dies gilt auch für den stimmlosen Bereich (z.B. Frikative, Flüstersprache).

Sprecherin A



Sprecher B



Sprecher C