DARSTELLUNG DES SPRECHVERHALTENS ALS STATISTISCHE TONHÖHEN- UND FORMANT-VERTEILUNG MITTELS LANGZEITANALYSE

F. WINCKEL

Wenn man davon ausgeht, daß die Lautproduktion der Sprache einen stochastischen Prozeß darstellt, so muß man erwarten, daß man in einer statistischen Auswertung der Lautparameter Merkmale der Sprecher von Sprachgemeinschaften ermitteln kann. Die Auftragung der Amplituden-Häufigkeitsverteilung über dem Sprachfrequenzband, ermittelt aus dem Sprechvorgang von mehreren Minuten, müßte Merkmale einer Sprachgruppe erkennen lassen.

Jedoch nur bei gedrucktem Text werden die statistischen Stichproben an verschiedenen Abschnitten des Textes den gleichen Mittelwertsverlauf ergeben (ergodisches Verhalten), während in der Lautsprache die prosodemischen Merkmale die ergodische Lautstruktur verdecken, wobei diese wiederum von emotionalen Faktoren abhängen.

Wenn es gelingt, identische normale Verteilungskurven von Sprechern einer Gruppe zu ermitteln, dann müssen die Abweichungen von diesen Kurven die prosodemischen Merkmale als quantitative Größen aufweisen. Die Möglichkeit läßt sich anhand der hier gezeigten Histogramme nachweisen.

Hierzu wurde ein Gerät entwickelt, das in einer Terzfilteranalyse die Häufigkeit der Amplituden, die einen mittleren Lautpegel — hier z. B. 30 dB — unter Vollaussteuerung überschreiten, mittels eines Zählers anzeigt.

Die relative Terzpegelhäufigkeit ergibt sich, indem die Impulszahl Z auf die Mittenfrequenz des jeweiligen Terzfilters f_m und die Gesamtsprechdauer t_{ges} bezogen wird.

Terzpegelhäufigkeit

$$H = \frac{Z \,.\, 100}{f_m \,.\, t_{ges}} \qquad (\%) \tag{1}$$

Es ergibt sich im allgemeinen eine mit der Frequenz abnehmende Häufigkeitsverteilung (Abb. 1a). In einer anderen Darstellung wird der Mittelwert der Impulszahl über der Gesamtmeßzeit als Funktion der Terzfilter-Mittenfrequenz gezeigt (Abb. 1b).

Mittlere Zählrate

$$Z_m = \frac{Z}{t_{ges}} \tag{2}$$

In diesem Falle kommen die höheren Frequenzgebiete besser zur Geltung, wodurch das Übergangsgebiet vom Vokal- zum Konsonantbereich genauer untersucht werden

kann. Im Fall (1) tritt der tieffrequente Bereich besser hervor, weshalb er über die Tonhöhenverteilung besser Auskunft gibt. Die Häufigkeitskurven lassen oft einen Einbruch im Bereich von 200 Hz (männliche Stimme) erkennen, was eine

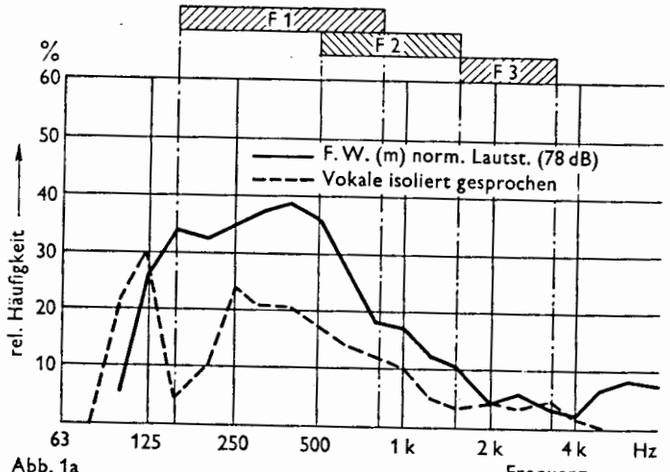

Abb. 1a

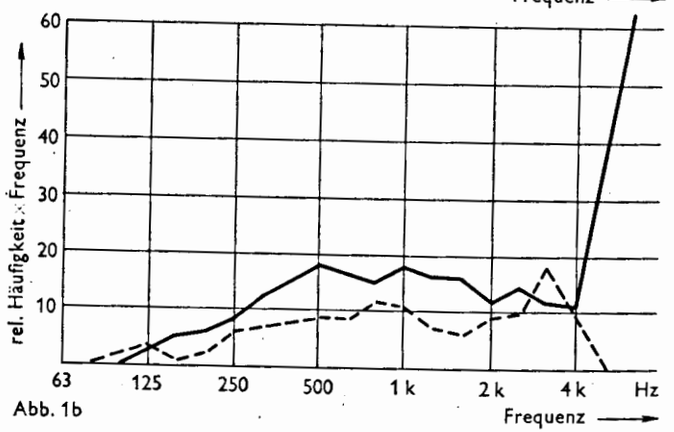

Abb. 1b

Abtrennung des Grundtonbereichs vom Formantbereich bedeutet. Zur Bestimmung der Tonhöhen-Häufigkeitsverteilung wurde in einer getrennten Messung eine feinere Filterteilung im Bereich unterhalb 200 Hz vorgenommen (Abb. 2b). Die Amplitudenverteilung unabhängig von der Frequenz wurde schon früher als eine Gaußverteilung festgestellt.

Abb. 3 zeigt das Sprechverhalten von zwei männlichen und einem weiblichen Sprecher beim Lesen von belanglosem Text über 3 Min. Die Abweichung von der ergodischen Struktur ist vor allem in der Tonhöhenverteilung zu erkennen: Die breite Verteilung bei F. W. geht in den Formantbereich über (Zeichen für Vitalität), während bei den anderen beiden Sprechern Kumulationen bei 125 bzw. 250 Hz zu

beobachten sind, was auf eine mehr monotone Sprechweise schließen läßt. Die Proben von anderen Abschnitten des gelesenen Textes ergeben Deckungsgleichheit der Kurven.

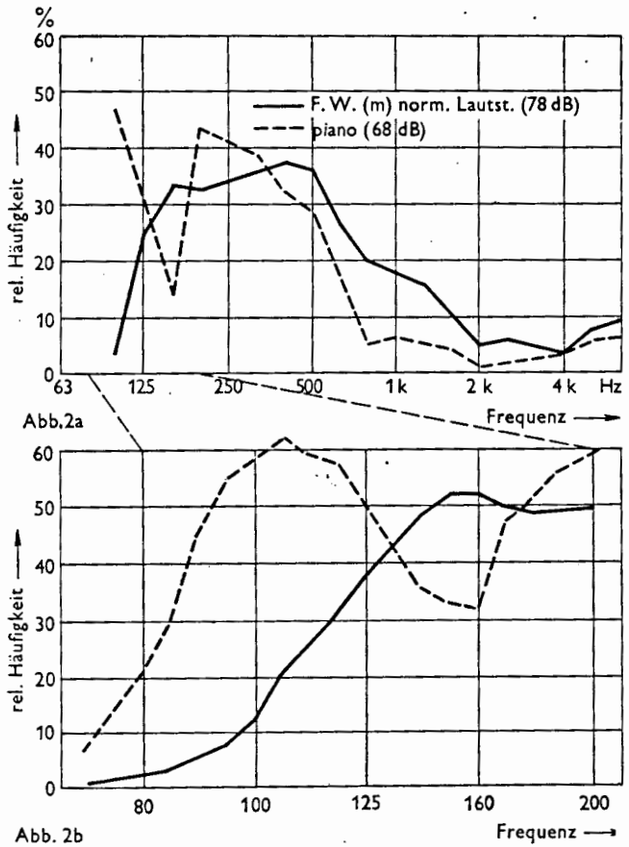

Abb.2a

Abb. 2b

Neben diesem prosodischen Merkmal wird ein Hinweis auf die Lautstärke bzw. Dynamik durch den Versuch gegeben, normal und dann denselben Text lauter zu lesen (Abb. 2a). Der Flächenunterschied zwischen den beiden Kurven ergibt ein Maß für Lautstärke bei gleichzeitig veränderter Intonation.

Es sind ferner noch folgende Sprechweisen untersucht worden: zu schnell, zu langsam, zu hoch, ferner Störungseinflüsse auf den Sprecher wie Geräusch und Lee-Effekt, sprechen mit und ohne Zähne, Gesangsstimmen, Stottern und sonstige Stimm- und Sprachfehler.

Eine gewisse Grundtendenz ist darin zu erblicken, daß jede zusätzliche Beanspruchung, sei es durch Erschwerung des Sprechaktes selbst oder durch Ablenkung der Aufmerksamkeit (Echo oder Lärm), die Artikulationstiefe bzw. den Artikulationsfaktor reduziert. Die Frequenzhäufigkeitsverteilung tendiert allerdings nicht

in dem erwarteten Maß zur Kumulation in den Formantbereichen des neutralen Vokals. Die Terzfilteranalyse ist dafür außerdem zu grob. Ferner führt das Nachlassen der Spannung in der Artikulationsmuskulatur zur Amplitudenverringerung des Formantbereichs um 2500—3500 Hz (Vokal i). Eine erhöhte Spannung in diesem Bereich ist bei ausgebildeten Sprechern stets vorhanden.

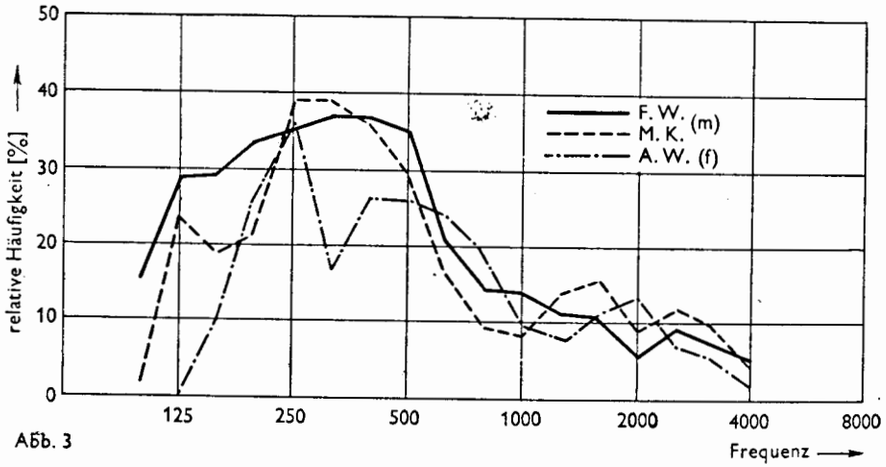

Abb. 3

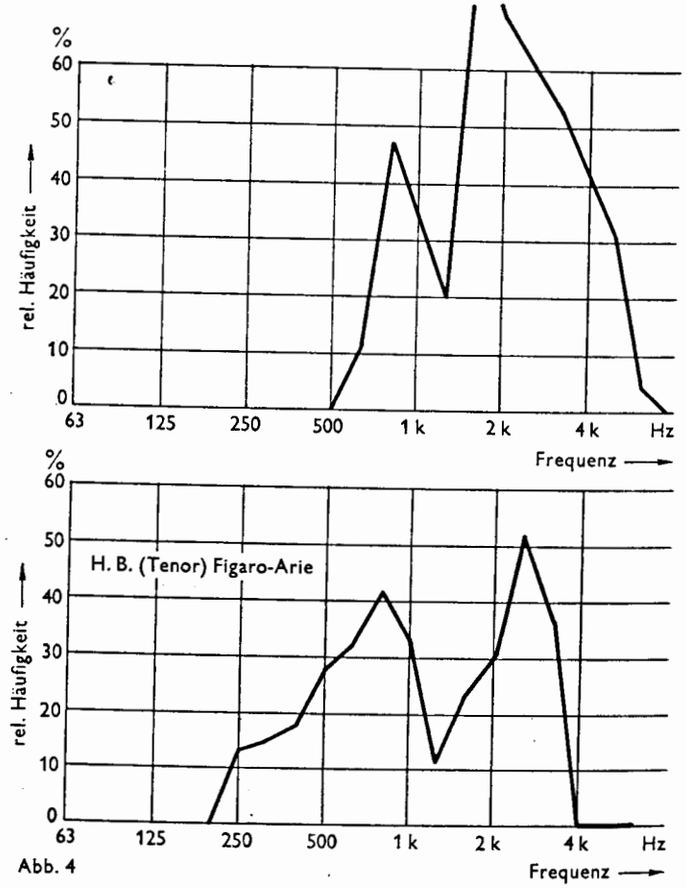

Abb. 4

Weiterhin bewirken die zusätzlichen Beanspruchungen beim Sprechen durch Störungseinflüsse eine Verringerung des Intonationsumfangs im Sinne zunehmender Monotonisierung. Die Individualunterscheidung beim normalen Sprechen drückt sich in den Tonhöhen-Diagrammen durch den Grad der Monotonie aus, außerdem in der Amplitude des 3000-Hz-Formantbereichs. Bei lebhaften Sprechern wird der Einschnitt zwischen Tonhöhen- und Formantbereich überbrückt wegen der breiteren Skala der Melodiehöhe.

Aus dem Diagramm „Tenorstimme — Hahnruf" (Abb. 4) geht eine weitgehende Übereinstimmung der Häufigkeitsmaxima hervor. Daraus läßt sich die Erkenntnis ableiten, daß in der spektralen Häufigkeitsverteilung wesentlich die Charakteristik des Stimmerzeugers vorhanden ist. Man kann dann in dem gezeigten Beispiel der ausgeprägten Formantspitzen, die im Verhältnis 2 : 1 und 3 : 1 vorkommen, auf die Form der Generatorschwingung schließen, im letzteren Fall also auf eine Rechteckschwingung, was im Prinzip auf eine andersartige Lösung gegenüber dem Verfahren des „Inverse Filtering" hinausläuft.

LITERATUR

F. Winckel u. M. Krause, Ermittlungen von Spracheigenschaften aus statistischen Eigenschaften von Amplitude und Tonhöhe, *Int. Akustik-Kongreß* Lüttich (A 41) 1965.