

# EINE SPRECHERUNABHÄNGIGE [a, ɔ] — ERKENNUNG

ROBERT RUPPRATH

Das im folgenden beschriebene Detektionsverfahren der [a, ɔ]-Erkennung basiert nicht mehr auf einer RHO-null-Dichtemessung, sondern bedient sich der lautbezogenen Energiekonzentration im Sprachfrequenzbereich.

Unter den nichtfrikativen Schallsegmenten sind die [a, ɔ]-Segmente eindeutig charakterisiert durch die spektrale Energiekonzentration im Bereich ihrer ersten beiden Formanten zwischen 700 und 1500 Hz. Bei dem Verfahren der RHO-null-Dichtemessung, bei dem die Lage des ersten Formanten von vokalischen Schallsegmenten durch die RHO-null-Dichtefunktion approximiert wurde, zeigen die [a, ɔ]-Segmente zwar ein ausgeprägtes Maximum, die Amplitude dieses Maximums jedoch blieb abhängig von der Frequenz des ersten Formanten, die in Abhängigkeit vom individuellen Sprecher immerhin einen Streubereich bis zu ca.  $\pm 150$  Hz aufweisen können.

Um die [a, ɔ]-Detektion von diesen Schwankungen des ersten Formanten unabhängig zu machen, wurde für das neue Verfahren lediglich die Tatsache berücksichtigt, daß die [a, ɔ]-Segmente im Bereich von 700-1500 Hz maximale Energiekonzentration aufweisen. (Vgl. dazu die in Abb. 1 wiedergegebene Spektralanalyse eines a-Lautes.) Diese Energiekonzentration ist nicht nur durch den ersten, sondern durch den ersten UND zweiten Formanten verursacht. Dabei übertrifft sie die Energiekonzentration im Bereich des Stimmtones. Genau das unterscheidet die [a, ɔ]-Segmente aber von allen übrigen nichtfrikativen stimmhaften Schallsegmenten. Außerdem liefern die [a, ɔ]-Laute die weitaus energiereichsten Segmente des Sprachschalles.

GENERELLE FUNKTIONSWEISE (s. Blockschaltbild, Abb. 2 und 3). Das Detektionsverfahren setzt die Normierung der Intensität des akustischen Sprachsignals voraus. Gemäß Blockschaltbild Abb. 2 wird das eintreffende Sprachsignal mithilfe des Stimmhaftdetektors und anschließendem Niederfrequenzanalogtor auf die stimmhaften Segmente beschränkt und über einen Sprachsignalkompressor in seiner Amplitude normiert. Die Qualität der durch den Sprachsignalkompressor vorgenommenen Intensitätsnormierung ist mitentscheidend für die große Sicherheit der sprecherunabhängigen [a, ɔ]-Lauterkennung, denn nur in ihrer Intensität normierte [a, ɔ]-Segmente können die vorgegebene Bedingung, nämlich Sprecherunabhängigkeit, erfüllen. (Die Darstellung des Stimmhaftdetektors, Niederfrequenzanalogtors und

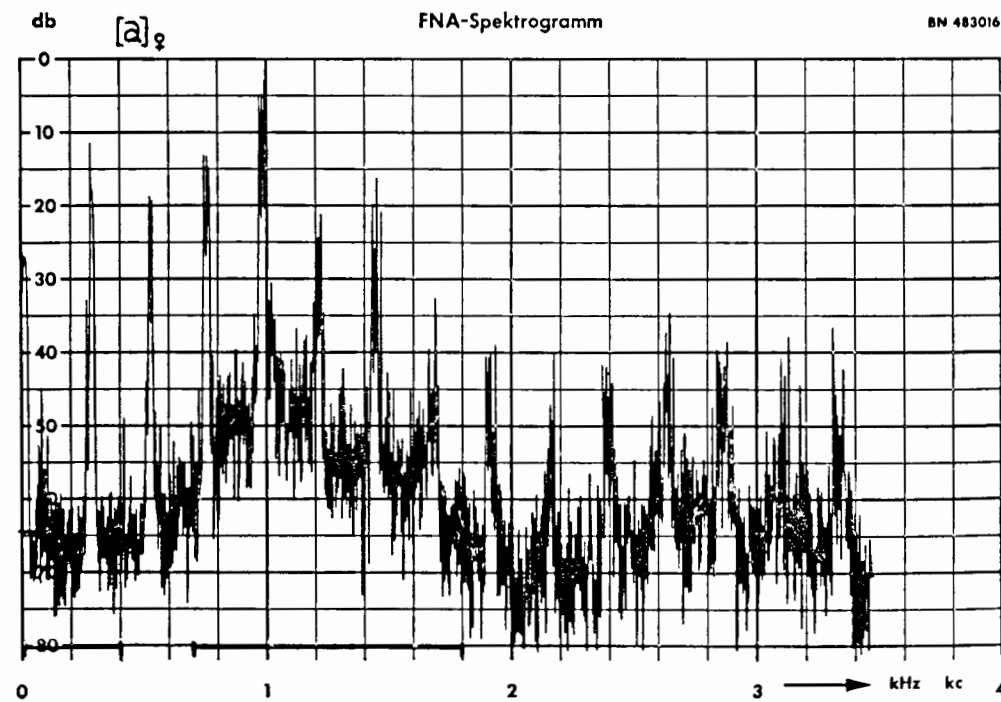
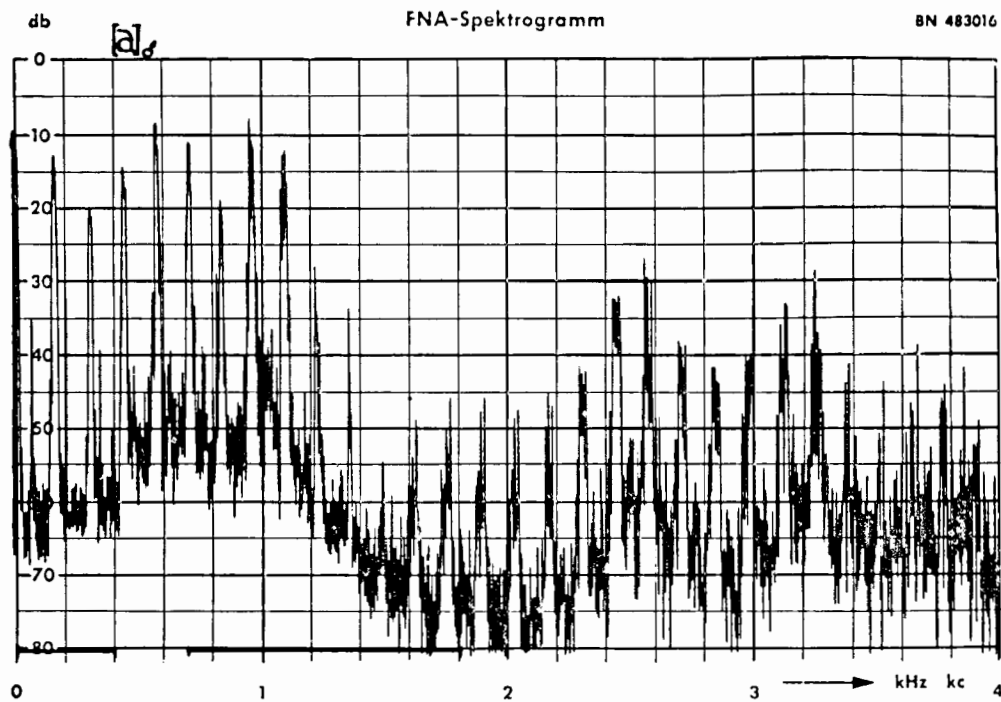


Abb 1

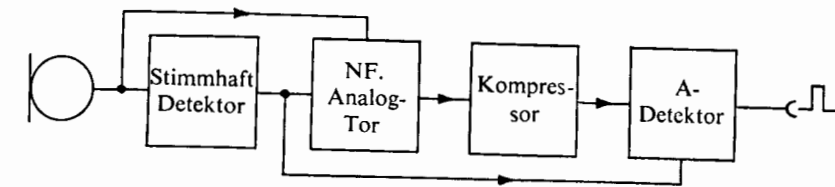


Abb. 2

des Sprachsignalkompressors ist als Forschungsbericht im H. Buske Verlag, Hamburg, erschienen.)

Die vom Kompressor normierten stimmhaften Sprachsignale werden, gemäß Blockschaltbild Abb. 3, über zwei getrennte Zweige weiterverarbeitet. Im oberen Zweig wird mit einem Tiefpaß der tieffrequente Spektralanteil unterhalb 400 Hz herausgefiltert, der bei [a,ɔ]-Schallsegmenten durch die vorgenommene Normierung relativ gering, bei allen anderen stimmhaften Sprachschallsegmenten aber mehr oder weniger mit der genormten Energie des Signals am Kompressorausgang übereinstimmt. Nach der Doppelweggleichrichtung lassen sich also die [a,ɔ]-Segmente von den übrigen stimmhaften Segmenten unterscheiden. Im unteren Zweig wird das vom Kompressor normierte stimmhafte Sprachsignal über einen Bandpaß im Bereich zwischen 700 und 1800 Hz übertragen. Nach der Doppelweggleichrichtung ergibt sich für die [a,ɔ]-Segmente eine maximale Signalenergie, die mit der Signalenergie am Ausgang des Kompressors beinahe übereinstimmt. Für alle anderen stimmhaften Sprachschallsegmente ist kein oder ein nur sehr schwaches Signal am Ausgang des Bandpasses nachzuweisen. Störendes Ausschwingverhalten des Sprachsignalkompressors, verursacht durch Einflüsse des Sprachsignals, wird durch vom Stimmhaftdetektor kontrollierte Und-Tore unterdrückt. Die am Ende der beiden getrennten Verarbeitungswege durch Integration gewonnenen analogen Parametersignale werden durch Subtraktion miteinander verglichen. Die resultierende Differenz

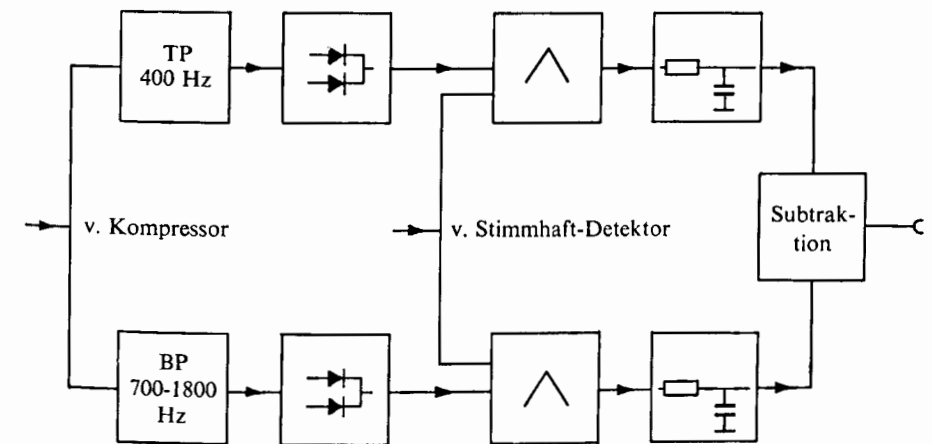


Abb. 3

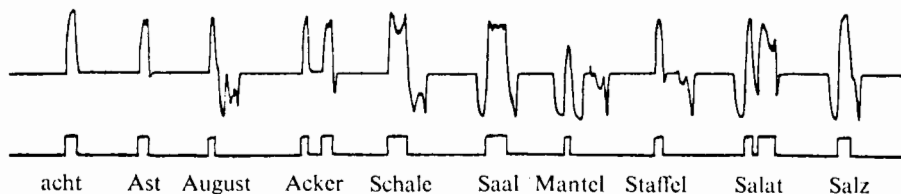


Abb. 4. Testwortreihe.

weist für alle [a,ɔ]-Segmente ausgeprägte Maxima auf, die über einen Schwellwert-schalter eine digitale Merkmalanzeige sicher gewährleisten (vgl. Abb. 4).

Der [a,ɔ]-Detektor kommt im Analysatorteil unseres DAWID-II Projektes zum Einsatz.

*Institut für Kommunikationsforschung und Phonetik  
Universität Bonn*

#### DISCUSSION

HEIKE (Köln)

Haben sie die Sicherheit des Verfahrens getestet und mit welchem Material?

RUPPRATH

Die Sicherheit im Erkennen von [a,ɔ]-Lautfolgen liegt z.Zt. bei 90%. Die Zahl der Testwörter beträgt 640, gesprochen von 13 männlichen und 7 weiblichen Sprechern.