VERFAHREN UND ERGEBNISSE DER ANALYSE DEUTSCHER VOKALE UND KONSONANTEN

W. TSCHESCHNER

I. EINFÜHRUNG

Die Sprache, der Mechanismus des Sprechens und Hörens ist schon seit langem Gegenstand der Untersuchung der verschiedensten wissenschaftlichen Disziplinen. Aber erst die Entwicklung der modernen Elektroakustik konnte in den letzten Jahrzehnten die Voraussetzungen für eine hochwertige Aufnahme- und Messtechnik schaffen, die eine exakte Aufzeichnung sowie eine genaue physikalische Analyse der Sprachtlaute gestatten. Nachdem es mit Hilfe der Informationstheorie gelang, den Wirkungsgrad von Übertragungssystemen zu ermitteln, gewann die Untersuchung der physikalischen Lautkriterien auch für die Nachrichtentechnik ein besonderes Interesse. Zeigte es sich doch dabei, dass die bisher üblichen Übertragungssysteme bei der Übertragung von Sprache nur mit einem ausserordentlich geringen Wirkungsgrad ausgenutzt werden[1].

II. OBJECTIVE ANALYSE VON SPRACHLAUTEN

Von den verschiedenen Verfahren, die zur Untersuchung der Sprachstrukturen herangezogen werden, sind besonders die für die Nachrichtentechnik von Interesse, die eine unmittelbare Auswertung der Signaleigenschaften der Sprachlaute gestatten. Ein dafür zugeschnittenes Verfahren, welches die Abbildung und Auswertung der Sprachstrukturen erlaubt – hier objektive Analyse genannt – geht von dem Übertragungssystem der interhumanen Kommunikationskette aus (Bild 1). Hierbei sind die durch elektroakustische Wandler in elektrische Schwingungen übersetzten Sprachsignale der Ausgangspunkt zur objektiven Analyse. Das abgeleitete Signal besteht aus einem komplizierten Schwingungsgemisch, das sowohl kontinuierliche als auch diskontinuierliche Frequenzanteile besitzt, die des weiteren in ihrer Amplitude als auch in ihrer Frequenz moduliert sein können. Untersucht man z.B. die Silbe "te", so findet man folgende mit einem normalen Schleifenoszillografen messbare Zeitfunktion (Bild 2). Diese Funktion enthält neben der speziellen Lautinformation auch Informationen, aus denen das menschliche Empfangsorgan den Sprecher und auch seine Gemütsverfassung erkennen kann. Aufgabe der objektiven Analyse ist es nun,

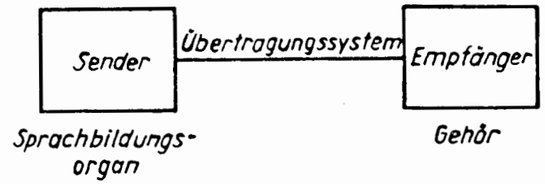

Bild 1. Interhumane Kommunikationskette.

aus der vorliegenden Zeitfunktion die relevanten Merkmale herauszufinden, die es gestatten, die einzelnen Sprachlaute physikalisch exakt zu beschreiben.

Die Zeitfunktionen von Lauten lassen, wie auch aus Bild 2 ersichtlich, bereits einige linguistische Besonderheiten erkennen, die eine Unterscheidung zumindest von Lautgruppen ermöglichen. So lässt sich auf Grund des Hüllkurvenverlaufes am Lautanfang sagen, ob Explosiv-oder Dauerlaute vorliegen. Auch die Unterscheidung der Lautgruppen "stimmhaft" – "stimmlos" lässt die Betrachtung der vorliegenden Zeitfunktion zu.

Zur Unterscheidung und eindeutigen Beschreibung einzelner Sprachlaute ist jedoch eine Zerlegung der Zeitfunktion unbedingt erforderlich. Diese Zerlegung muss so erfolgen, dass die Abhängigkeiten der Amplitude und der Frequenz von der Zeit unmittelbar beobachtet werden können. Gemäss den Geräteeigenschaften der üblichen Analysatoren kann jedoch nur eine zweidimensionale Darstellung erzielt werden. Lediglich die Geräte, die zur Ableitung der Visible Speech-Diagramme verwendet werden, weisen eine dreidimensionale Darstellung auf. Diese lässt dann jedoch nur noch qualitative Angaben über die jeweilige Amplitude zu.

Die für alle technischen Frequenzanalysatoren entscheidende Gesetzmässigkeit ist die Ungenauigkeitsrelation $\triangle f. \triangle t \approx 1$ [2]. Aus dieser Beziehung wird die ausserordentlich grosse Schwierigkeit der technischen Analyse von Sprache offenbar, denn das menschliche Ohr als der entscheidende Indikator ist ein Empfangsorgan, welches nicht dieser Relation gehorcht. In dem menschlichen Ohr arbeitet ein System, welches sehr grosse Frequenzauflösung mit sehr kleinen Einschwingzeiten verbindet [3].

Um also die Analysiereigenschaften des Ohres, die ja letzten Endes dem anzustrebenden Auflösungsvermögen der Analysiereinrichtungen zugrunde zu legen sind, in irgendeiner Form berücksichtigen zu können, müssen verschiedene Kompromisse eingegangen werden.

Die technischen Analysatoren verwenden im wesentlichen das Parallelfilter- oder das Suchtonprinzip. Das Parallelfilterverfahren ist hinsichtlich einer kleineren Analysierzeit wesentlich dem Suchtonverfahren überlegen. Um eine dem menschlichen Hörvermögen ähnliche Frequenzauflösung zu erzielen, bedarf es aber eines sehr grossen Aufwandes an Filtern und Anzeigevorrichtungen. Verwendet man das Abtast- und Speicherprinzip, so reduziert sich zwar der Aufwand an Anzeigevorrichtungen, jedoch vergrössert sich gleichzeitig die Analysierzeit. Ausserdem müssen bei diesem

Verfahren die Fehler, die stets in den Filterüberlappungsbereichen entstehen, in Kauf genommen werden. Das Suchtonverfahren weist diesen Fehler nicht auf, dagegen braucht es bei allerdings wesentlich kleinerem Aufwand Analysierzeiten bis zu ca. 5 min bei Realisierung einer hohen Frequenzselektion und einem Frequenzbereich von O Hz bis 20 kHz.

Für das vorliegende Vorhaben wurde nun folgendes Verfahren angewendet. Zur Erzielung einer hohen Frequenzselektion wurde das Suchtonprinzip angewendet. Um damit auch Analysen von kurzen Lauten bzw. Explosivlauten durchführen zu können, wurde eine zusätzliche Einrichtung gebaut, die eine periodische Abtastung von Zeitfunktionen gestattet [4]. Dabei wurde eine Abtastfrequenz von 50 Hz gewählt, womit Zeitbereiche von jeweils 20 ms periodisch abgetastet werden können. Mit dieser Einrichtung erhält man bei der Integrationszeit von 20 ms mittlere Frequenzspektren, die durch die Abtastfrequenz und deren Harmonischen moduliert sind. Die Zahl und die Grösse dieser Harmonischen wird durch die von der Abtastvorrichtung abhängigen Spaltfunktion bestimmt.

Das Zeitzerlegungsgerät, welches die periodische Abtastung der auf Magnetband gespeicherten Zeitfunktion besorgt, besteht aus einer Serie von Hörköpfen, die auf einer rotierenden Scheibe untergebracht sind – Bild 3. Die Hörkopfspannungen werden über einen Kollektor abgenommen und dem Suchtongerät zugeführt. Ein Zeitausschnitt wird jeweils solange abgetastet, bis der Suchtonanalysator seine Analyse, die 2½ Minuten dauert, beendet hat. Im Anschluss daran wird in dem benachbarten Zeitbereich die Analyse fortgesetzt.

Bild 4 zeigt die komplette Versuchsanordnung, bestehend aus dem Zeitzerlegungsgerät, dem Suchtongerät mit Zusatzeinrichtungen, sowie dem Anzeigeoszillografen mit der Registriereinrichtung. Ein weiterer Oszillograf ermöglicht die Zeitfunktion bzw. den jeweils untersuchten Lautausschnitt als stehendes Bild darzustellen, während die Analysiereinrichtung unmittelbar das dazugehörige Spektrum liefert. Mit Hilfe der Zusatzeinrichtungen kann auch eine dreidimensionale Darstellung erreicht werden.

Hinsichtlich der Dimensionierung konnten bei dem Suchtonanalysator folgende technische Werte realisiert werden:

Bandbreite: $\triangle f_{o,r_{N_p}} = 30$ Hz oder 80 Hz
Amplitudenbereich bei direkter Anzeige: 60 db
Frequenzbereich: 100 Hz – 20 kHz
Amplitudenteilung: linear oder log., wobei bei der logarithmischen Teilung zwischen dem Log. der Eingangsspannung und der Anzeige ein linearer Zusammenhang besteht.
Frequenzteilung: Nach Frequenzgruppen.

Welches Auflösungsvermögen sich praktisch erzielen lässt, geht auch aus Bild 5 hervor. Diese Aufnahmen zeigen Spektralzerlegungen einer gering verzerrten Sinusschwingung und einer rechteckförmigen Zeitfunktione. Es ist zu ersehen, dass die

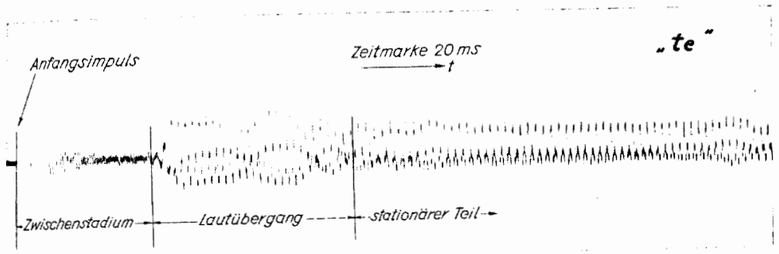

Bild 2. Zeitfunktion der Silbe "te".

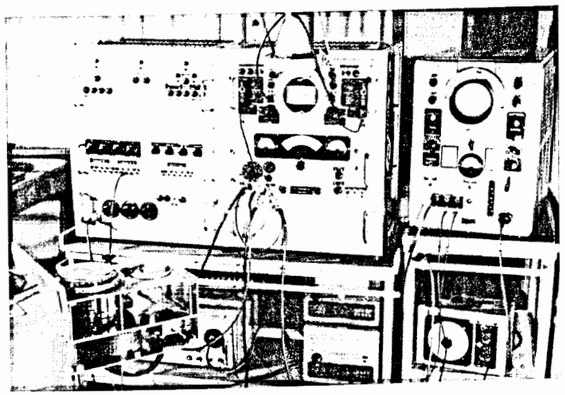

Bild. 3. Aufbau des Zeit-Zerlegungsgerätes.

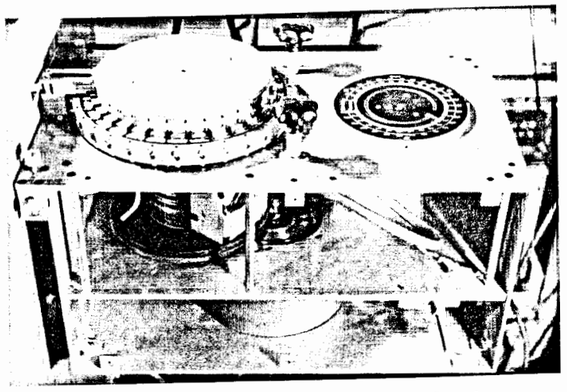

Bild 4. Versuchsanordnung zur Analyse von Sprache.

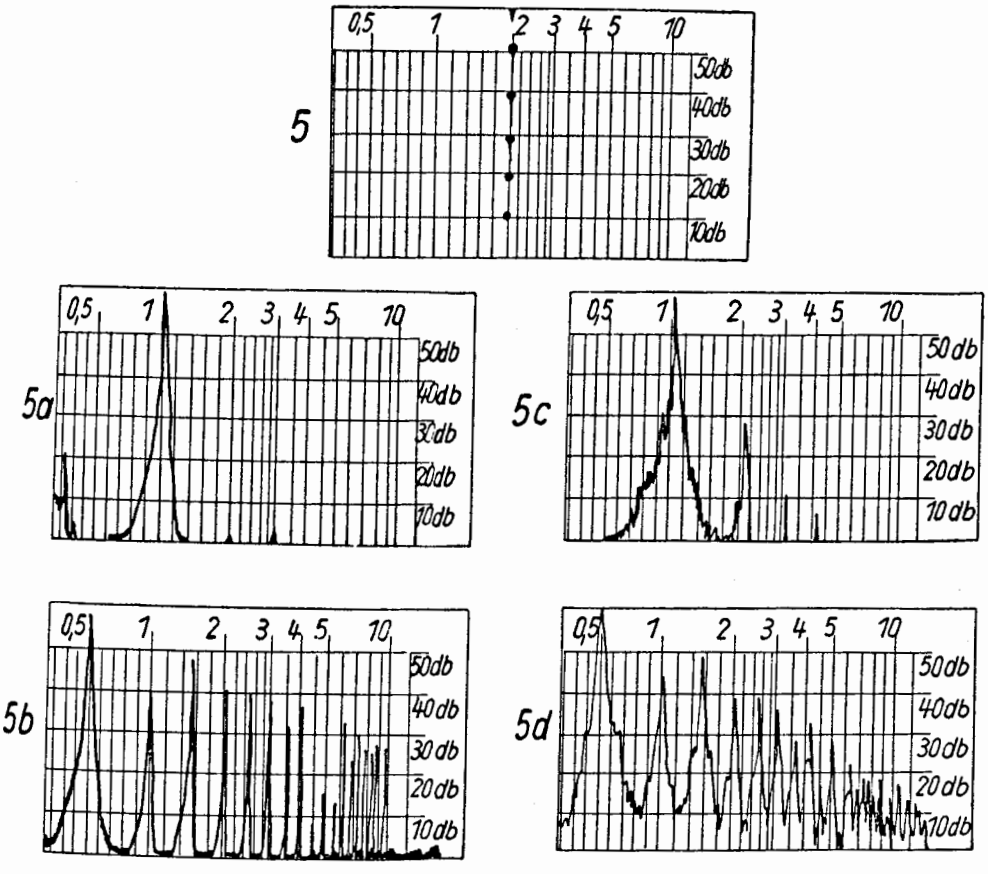

Bild 5. Spektralanalysen stationärer Vorgänge. 5a: Analyse einer Sinuschwingung f = 1 kHz, u = 1,5 mV, k = 2‰; 5b: Analyse einer rechteckförmigen Zeitfunktion f = 0.5 kHz; 5c: Aufnahme wie unter 5a auf Magnetband gespeichert und mit dem Zeit-Zerlegungsgerät abgetasted; 5d: Aufnahme wie unter 5b auf Magnetband gespeichert und mit dem Zeit-Zerlegungsgerät abgetastet.

systemeigenen Verzerrungen ausserordentlich klein sind und dass auch die bei der Zeitzerlegung entstehenden Störkomponenten die Analyse nur unwesentlich beeinflussen. Die bei der Aufnahme 5 c in beträchtlicher Grösse erscheinenden Harmonischen sind auf die dabei vorliegende starke Übersteuerung des Magnetbandes zurückzuführen. Wie aus der Aufnahme 5 d hervorgeht, verschlechtert sich die Linearität des Frequenzganges bei Anwendung des Zeitzerlegungsgerätes. Der Abfall, der bei Frequenzen um 15 kHz etwa 8 db beträgt, muss deshalb bei Lautanalysen korrigiert werden. Werden nun einzelne Lautausschnitte analysiert, so lassen sich die jeweils gefundenen Spektralverteilungen zu einem Gebirge zusammensetzen, wie es z.B. für die Silbe "to" aus Bild 6 ersichtlich ist. In dieser Abbildung ist die Frequenz und die Amplitude in Abhängigkeit von der Zeit dargestellt. Die hier besonders hervortretenden Frequenzgebiete, die als typisch für die Laute anzusehen sind, werden als Formanten der Laute bezeichnet. Für die Lautgruppe der Sonanten geht das bei

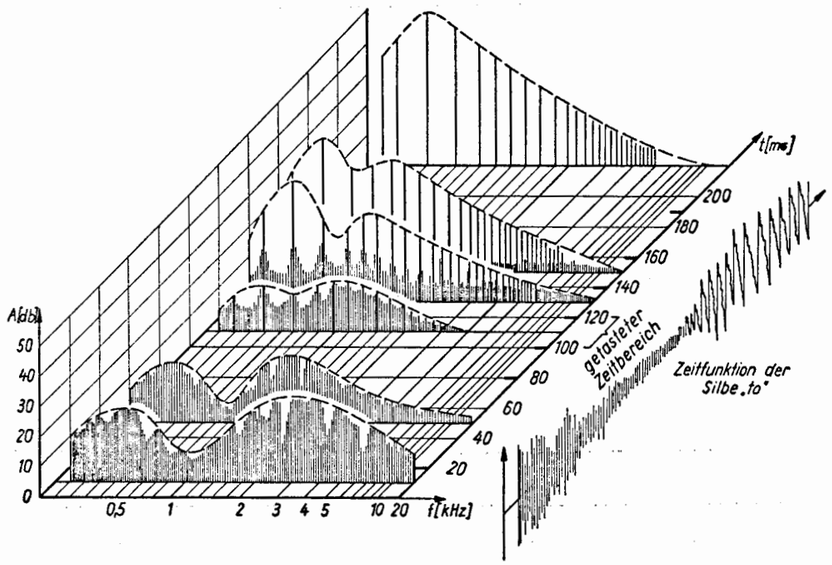

Bild 6. Frequenz – Amplituden – Zeitdiagramm der Silbe "to".

allen Geräusch-und Explosivlauten vorhandene kontinuierliche Spektrum in ein diskontinuierliches über, das umso weniger kontinuierliche Anteile besitzt, je grösser die Zahl der kongruenten Perioden der Zeitfunktionen ist, d.h. je reiner und je länger ein Sonant gesprochen bzw. gesungen wird.

Wie Bild 6 zeigt, besitzt das beschriebene Verfahren ein solches Auflösungsvermögen, dass es geeignet ist, auch die akustische Struktur der Explosivlaute zu untersuchen. Diesbezügliche Untersuchungen erbrachten für am Lautanfang (Lautausschnitt 20 ms) analysierte Explosivlaute unterschiedlicher Strukturen, die bereits eine objektive Unterscheidung ermöglichen. Die gefundenen Frequenzverteilungen – Bild 7 und 8 – stellen dabei Mittelwerte von 20 Lautproben der Lautkombination Vokal – Konsonant zweier Sprecher dar. Da einzelne Aufnahmen stets eine Vielzahl von individuellen Besonderheiten zeigen, müssen die prinzipiellen Strukturen stets aus einer grösseren Anzahl von Lautproben gemittelt werden. Die Frequenzstrukturen der stimmhaften Explosivlaute weisen dabei noch zusätzliche harmonische Anteile unterhalb 1 kHz auf.

Werden Lautproben der Lautkombination Konsonant – Vokal analysiert, so ergeben sich für die Laute "b", "d", "p" und "t" prinzipiell ähnliche Verhältnisse, während bei den Lauten "g" und "k" die Lage der Oberformanten sehr stark von dem nachfolgenden Vokal beeinflusst wird.

Da die Empfindlichkeit der Einrichtung so gross ist, dass auch die Analysen des

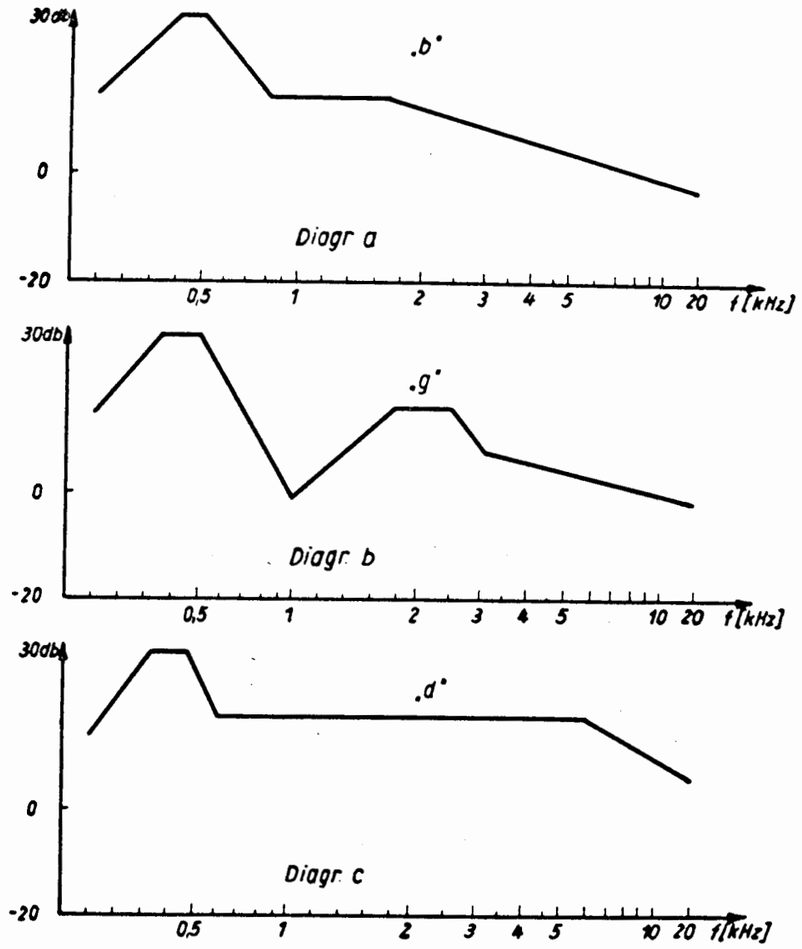

Bild 7. Mittlere Spektralverteilung stimmhafter Explosivlaute.

stimmlosen Frikativlautes "h" gute Ergebnisse liefern, lassen sich somit alle Laute der deutschen Sprache auf ihre akustischen Strukturen hinsichtlich ihrer Frequenz – Amplituden – und Zeitabhängigkeiten untersuchen.

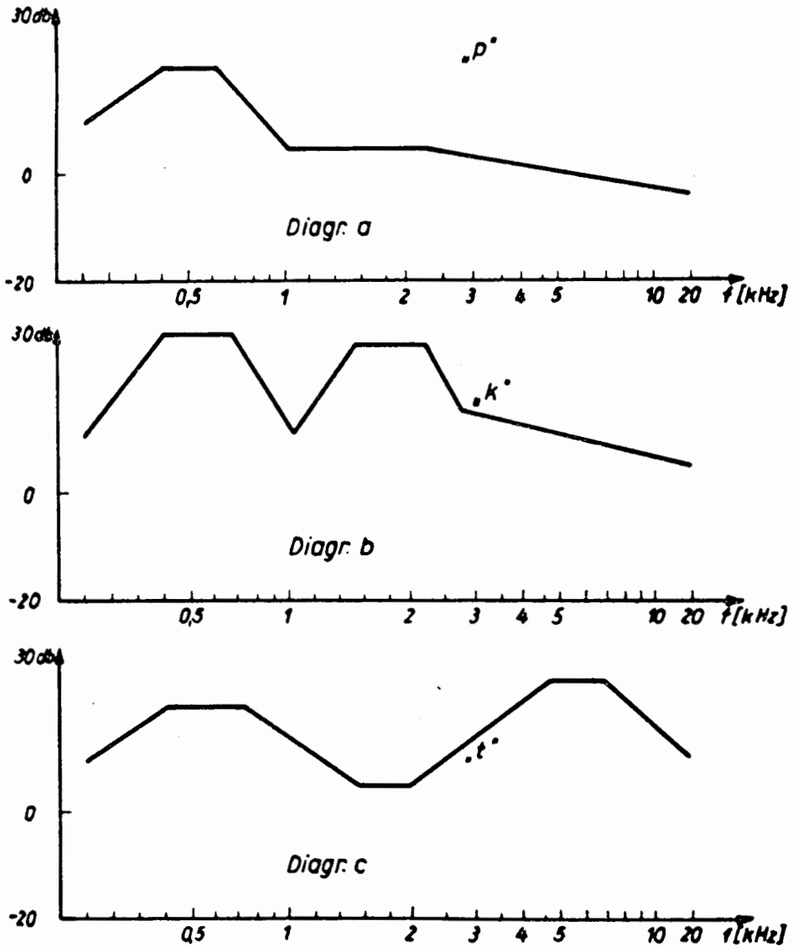

Bild 8. Mittlere Spektralverteilung stimmloser Explosivlaute.

III. SUBJECTIVE ANALYSE VON SPRACHLAUTEN

Die Merkmale, die sich mittels der objektiven Analyse aus den Strukturen der Sprachlaute ablieten lassen, können zwar schon zum Bau von Geräten zur automatischen Spracherkennung herangezogen werden. Es lässt sich jedoch nicht erkennen, inwieweit solche Merkmale von dem menschlichen Hörsystem tatsächlich zur Unterscheidung und Erkennung von Lauten verwendet werden. Deshalb ist es erforderlich,

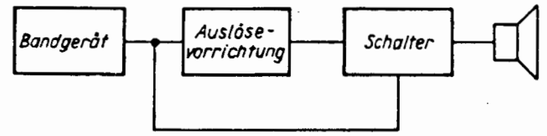

Bild 9. Prinzip des Zeitfilters.

die objektiven Kriterien auf ihre subjektive Wertigkeit zu untersuchen, d.h. die objektive Analyse muss durch die subjektive Analyse ergänzt werden. Dabei wird ebenfalls von dem Übertragungssystem der interhumanen Kommunikationskette ausgegangen. Durch physikalisch genau fixierbare Eingriffe in die Sprachstrukturen werden die Einflussgrössen hinsichtlich ihrer Auswirkung auf die subjektive Verständlichkeit untersucht.

Bei diesen Untersuchungen ist es sinnvoll, innerhalb des physikalischen Raumes Frequenz – Amplitude – Zeit solche Eingriffe vorzunehmen, dass nur jeweils eine Dimension möglichst rückwirklungsfrei beeinflusst wird, d.h. das physikalische Sprachsignal (s.z.B. Bild 6) muss durch Schnitte quer zur Frequenz –, Amplituden- bzw. Zeitachse zerlegt und die so reduzierten Signale hinsichtlich ihrer subjektiven Verständlichkeit untersucht werden. An technischen Einrichtungen müssen somit

1. Frequenzfilter
2. Amplitudenfilter*
3. Zeitfilter*

eingesetzt werden.

Diese Einrichtungen ermöglichen es, im wesentlichen die Bedeutung bestimmter Signalanteile für die subjektive Erkennung und Unterscheidung zu untersuchen. Daneben können jedoch auch weitere technische Verfahren angewendet werden, die physikalisch definierte Eingriffe in die Struktur der Sprachsignale erlauben. Von den verschiedenen Möglichkeiten wurde hier besonders von einem Verfahren zur Deformation der Zeit-bzw. der Frequenzachse Gebrauch gemacht. Mit diesem lassen sich wichtige ergänzende Versuche machen, so z.B. über die Bedeutung der Lage und der Breite der Formanten, oder über die Bedeutung der Anfangsimpulse von Explosivlauten u.a. mehr. Von der apparativen Seite her stehen dafür Zeit-und Frequenztransformationseinrichtungen zur Verfügung.

An technischen Einrichtungen fanden in dem vorliegenden Fall als Frequenzfilter Tiefpässe und Hochpässe mit umschaltbarer Grenzfrequenz Verwendung. Als Amplitudenfilter wurden normale regelbare dreistufige Diodenbegrenzer benutzt, während für das Zeitfilter eine Einrichtung verwendet wurde, die eine Austastung bestimmter vorgegebener Zeitausschnitte längs der Zeitachse erlaubt. Diese Einrichtung – Bild 9 – besteht aus einer Auslösevorrichtung, die den Zeitpunkt t = O am Anfang des zu

* Zwecks einheitlicher Darstellung ist es hier sinnvoll, die in der Impulstechnik üblichen Ausdrücke [5] zu übernehmen.

untersuchenden Lautes fixiert, sowie 2 monostabilen Kippschaltungen, die wahlweise so geschaltet werden können, dass ein Zeittiefpass (ZTP), ZHP, ZBP und ZBS verwirklicht werden kann. Je nach Wahl der Schaltung können bestimmte Lautschnitte getastet werden – Bild 10 –, deren Breite im Bereich von 10 ms bis 1 s beliebig einstellbar ist.

Ein Gerät, mit dem sich eine Zeitdehnung bzw. eine Zeitkompression realisieren lässt, ist der von Springer [6] angegebene akustische Temporegler. Dabei besteht die Möglichkeit, die Bandaufnahmen bei der Wiedergabe nicht mit der Aufnahmegeschwindigkeit abzuspielen, so dass eine längere bzw. kürzere Wiedergabedauer erzielt werden kann. Um der damit verbundenen Frequenztransformation entgegenzuwirken, muss die Relativgeschwindigkeit zwischen Band und Hörkopf konstant gehalten werden. Das lässt sich durch rotierende Wiedergabeköpfe erreichen, wobei allerdings eine Segmentierung der Zeitfunktion in Kauf genommen werden muss. Für diese Untersuchungen lässt sich das bereits für die objektive Analyse herangezogene Zeit-Zerlegungsgerät verwenden, sobald die Absolutgeschwindigkeit der Abtastköpfe regelbar gemacht wird. Das wurde dadurch erreicht, dass der Antrieb der Hörköpfe durch einen Synchronmotor erfolgt, der über einen regelbaren RC-Generator mit nachfolgendem Kraftverstärker gespeist wird.

Bei dem praktischen Einsatz dieser Verfahren muss beachtet werden, dass bei der subjektiven Beurteilung von Sprachqualitäten bzw. bei Verständlichkeitsmessungen die Streuungen sehr gross sind. Nur durch geeignete Auswahl der Hörer nach ihrem Hörvermögen sowie durch statistische Auswertung einer Vielzahl von Urteilen pro Messpunkt, sind brauchbare Ergebnisse zu erzielen.

IV. ERGEBNISSE BEI DER ANALYSE DEUTSCHER SPRACHLAUTE

Führt man Verständlichkeitsuntersuchungen bei einer Amplitudenbegrenzung durch, so reduziert sich die Lautverständlichkeit selbst bei hohen Begrenzungen nur in geringem Masse. D. H. die Gesamtdynamik bzw. der Hüllkurvenverlauf spielt für die Erkennbarkeit eine untergeordnete Rolle. Eine stärkere Beeinflussung der Verständlichkeit wird bei Anwendung von Tiefpässen und Hochpässen variabler Grenzfrequenz erreicht. Hierbei ist Z. B. bei Explosivlauten die Tendenz der Änderung der Verständlichkeitswerte in Abhängigkeit von der Grenzfrequenz so, wie auf Grund der Ausbildung der Oberformanten – Bild 7 und 8 – zu erwarten ist. So ist der Abfall der Verständlichkeit bei einer Hochpassbegrenzung umso geringer, je höher der Oberformant des Explosivlautes ist. Diese Abhängigkeit ist jedoch nicht so eindeutig, dass die Ausbildung der Formanten der Explosivlaute allein für die subjektive Unterscheidung verantwortlich gemacht werden kann. Mit Hilfe der Zeitfilter kann geklärt werden, dass man sich bei der subjektiven Erkennung eines weiteren wesentlichen Kriteriums bedient.

Untersucht man die Lautkombination, Explosivlaut – Vokal, so wird mit immer

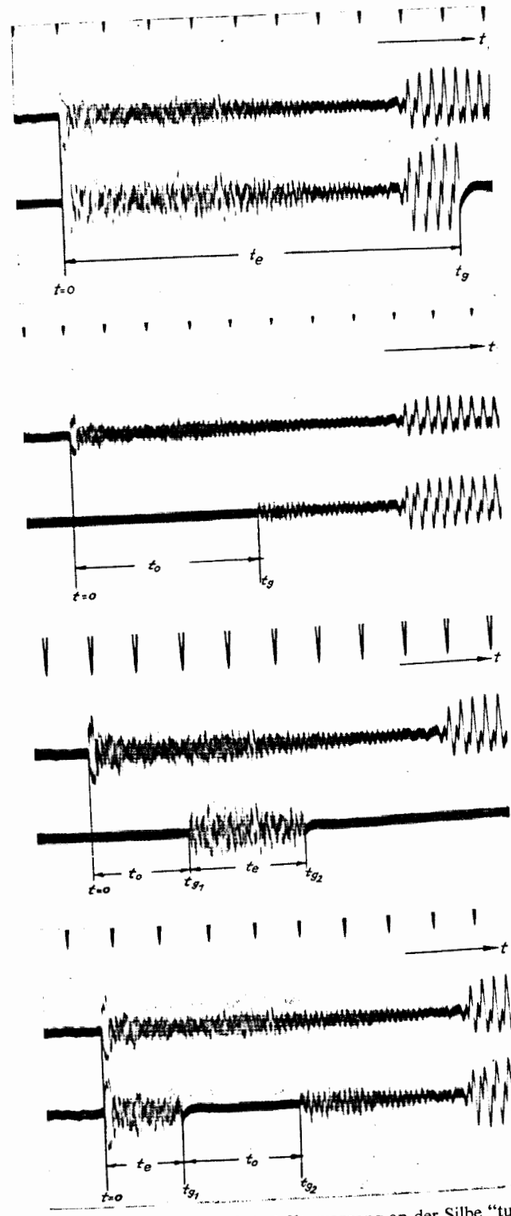

Bild 10. Zeitfiltervorgang an der Silbe "tu".

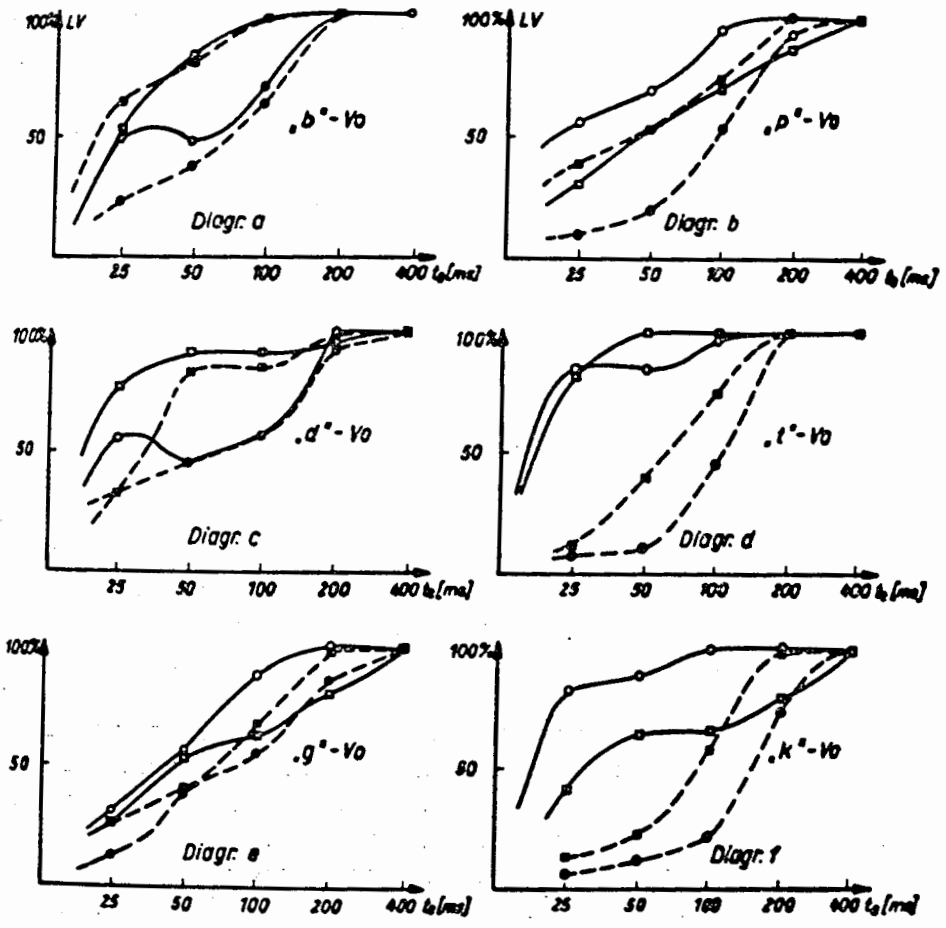

Bild 11. Lautverständlichkeit von Explosivlauten und Vokalen bei einer Zeitfilterbegrenzung.

—○— Konsonantenverständlichkeit weiblicher Sprecher
—□— „ „ männlicher Sprecher
—●— Vokalverständlichkeit weiblicher Sprecher
—■— „ „ männlicher Sprecher

kürzer gewähltem Silbenausschnitt bei einer ZTP-Begrenzung (s. auch Bild 10) erst die Zeitfunktion des Vokales und danach die des Konsonanten beeinträchtigt. Trägt man die dafür ermittelten Verständlichkeitswerte über der Länge der Silbenausschnitte auf, so ergeben sich die im Bild 11 gezeigten Abhängigkeiten. Während die Lautlänge der einzelnen Explosivlaute vor unterschiedlichen Vokalen nur um ca. 20% toleriert, weichen die Lautlängen bei verschiedenen Sprechern um über 100% voneinander ab. Aus diesem Grund kann bei Zeitfilteruntersuchungen nicht über die Messpunkte von Sprachproben verschiedener Sprecher gemittelt werden.

Auf der Ordinate der gezeigten Diagramme sind die Lautverständlichkeitswerte der Explosivlaute und gleichzeitig die der nachfolgenden Vokale aufgetragen. Die Abszisse

enthält dabei die Länge des getasteten Silbenausschnittes. Pro Messpunkt wurden jeweils 40 Hörerurteile berücksichtigt. Für die Beurteilung der Diagramme ist noch die objektive Lautlänge der Explosivlaute wichtig. Diese beträgt bei den vorliegenden Sprachproben für die stimmhaften Laute 4–30 ms und für die stimmlosen Laute 60–120 ms. Steigt die Verständlichkeit der Konsonanten mit zunehmendem Lautausschnitt erst dann an, wenn bereits die Vokale zu erkennen sind, d.h. fallen die Kurven für die Explosivlaute mit denen der Vokale näherungsweise zusammen, so kann der Explosivlaut nicht selbständig erkannt werden. Zu seiner Erkennung bzw. Unterscheidung ist ein spezielles Kriterium erforderlich, das hier in der Übergangsfunktion zu dem nachfolgenden Vokal gesucht werden muss. Der Verlauf der Übergangsfunktion kann mit Hilfe der Einrichtungen der objektiven Analyse bezüglich des Amplituden-Zeit-bzw. des Frequenz-Zeitverhaltens untersucht werden. Zur Klärung der Frage, welche speziellen Besonderheiten der Übergangsfunktion in dem menschlichen Hörsystem den Erkennungsprozess beeinflussen, bedraf es jedoch weiterer Untersuchungen. – Während im vorliegenden Fall die stimmhaften Explosivlaute nur mittels der Übergansfunktion zu erkennen sind, sind die stimmlosen Explosivlaute nicht in diesem Umfang darauf angewiesen. Das wesentliche Merkmal der stimmlosen Explosivlaute ist hier der Impuls am Anfang der Zeitfunktion und die unterschiedliche Frequenzstrukturierung der Geräuschanteile des Zwischenstadiums, wie sich mit Hilfe einer inversen Zeittransformation nachweisen lässt. Bei den stimmlosen Explosivlauten tritt hier der Beitrag des Lautüberganges zur Lautverständlichkeit zurück, da bei den vorliegenden Sprachproben Wert auf eine besonders deutliche Aussprache gelegt wurde. In der Umgangssprache trägt der Lautübergang wesentlich stärker zur Erkennbarkeit auch der stimmlosen Explosivlaute bei.

Welche grosse Bedeutung gerade die in dem Lautübergang liegenden Signalanteile für die Lautverständlichkeit besitzen, lässt sich auch bei den Vokalen nachweisen. Bild 12 zeigt die Verständlichkeit der Grundvokale bei einer Begrenzung mit einem umschaltbaren Tiefpass. Als Sprachproben wurden isolierte Vokale und die Lautkombinationen Konsonant – Vokal, Vo – Ko. und Ko. – Vo – Ko eines weiblichen und eines männlichen Sprechers zugrunde gelegt. Zum Vergleich sind in Bild 13 die Spektralverteilungen, die über 20 Laute gemittelt wurden, der quasistationären Lautteile von 200 ms Länge der isolierten Vokale und der Vokale aus den Lautkombinationen Ko. – Vo. und Vo – Ko. angegeben. Aus Bild 13 ist zu erkennen, dass bei den gewählten Gruppen der Lautkombinationen nur relativ geringfügige Abweichungen der mittleren Spektralverteilung auftreten. Dagegen zeigen die Darstellungen in Bild 12 ganz erhebliche Abweichungen in der Lautverständlichkeit der Vokale bei den verschiedenen Lautkombinationen. Unterschiedliche An- bezw. Ablautkonsonanten innerhalb einer Gruppe führen nur zu geringfügigen Abweichungen der Vokalverständlichkeit. Demsgegenüber wirkt sich die Stellung der Konsonanten ganz entscheidend auf die Verständlichkeit aus.

Bei der Tiefpassbegrenzung isoliert gesprochener Vokale sinkt die Lautverständlichkeit in dem Masse, in dem der typische Formant unterdrückt wird. Wird also

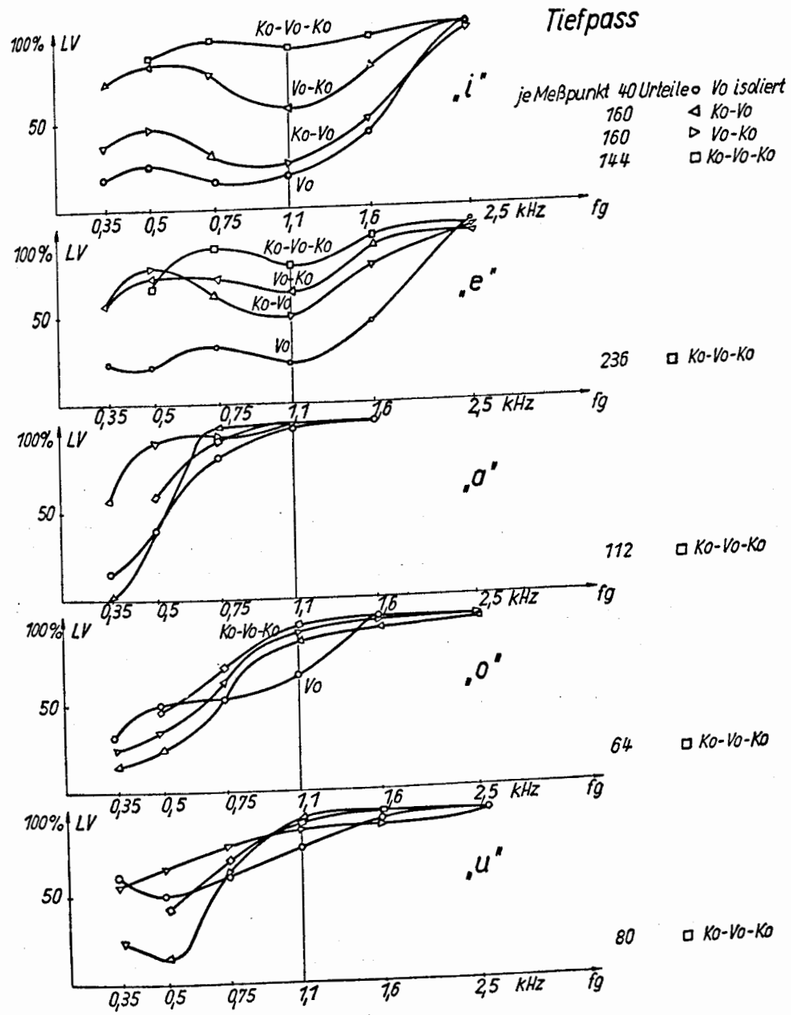

Bild 12. Lautverständlichkeit von Vokalen bei einer Frequenzbandbegrenzung.

z.B. der Oberformant des Vokals "i" bei 2.5 kHz unterdrückt, so sinkt die Verständlichkeit auf wenige Prozent ab und entsprechend dem verbleibenden Unterformanten geht "i" mit hoher Natürlichkeit in den Vokal "u" über. Dieser Zusammenhang zwischen Spektrum und Lautverständlichkeit liegt nur bei isolierten Vokalen, besonders bei gesungenen Vokalen vor. Bei Vokalen, die mit Konsonanten verknüpft sind, ist der Zusammenhang zwischen Spektrum und Lautverständlichkeit nicht mehr in diesem Masse gegeben. Einmal ist die Verständlichkeit der Vokale im wesentlichen höher und Lautverwechslungen treten in weit geringerem Umfang auf. Für diese Erscheinung sind 2 Ursachen verantwortlich zu machen.

1. Bei den Vokalen der Lautkombination Vo. – Ko. liegt die Verständlichkeit bei einer Tiefpassbegrenzung < 750 Hz bei den Vokalen "i" und "e" etwa 40–60% über der der isolierten Vokale und bei den Vokalen "o" und "u" etwa 20–30% darunter. Dieser Sachverhalt lässt sich auch auf Grund weiterer Untersuchungen, die hier nicht weiter angeführt werden sollen, auf nichtlineare Verzerrungen im menschlichen Ohr und den damit durch Oberwellen bzw. Kombinationstöne vorgetäuschten Oberformanten zurückführen. Dadurch ergeben sich bei den Vokalen "i" und "e" eine höhere Verständlichkeit und bei den Vokalen "u" und "o" treten leicht Lautverwechslungen nach "i" und "e" hin auf, die Verständlichkeit sinkt somit ab.

2. Bei den Vokalen der Lautkombination Ko. – Vo. und Ko. – Vo. – Ko. liegt die Lautverständlichkeit bei einer Begrenzung von f_{gTp}=750 Hz bis zu 80% über der der isolierten Vokale. Dieser Verständlichkeitsgewinn ist ausschliesslich durch eine zusätzliche Information bedingt, die in dem Lautübergang Ko. – Vo. zu suchen und die unabhängig von der Spektralverteilung des quasistationären Lautteiles ist.

Wie bereits aus den in Bild 13 enthaltenen Diagrammen hervorgeht, ist bei einer Grenzfrequenz f_{gTp} = 750 Hz das resultierende Spektrum so weit entartet, dass die zumindest für die Laute "e" und "i" wichtigen Oberformanten völlig unterdrückt sind.

Dieses Ergebnis soll durch einen weiteren Versuch deutlich gemacht werden. Bei einer Frequenzbandbegrenzung von Logatomen sinkt die Lautverständlichkeit der Grundvokale von 100% (ohneTiefpass) in der in Bild 14 dargestellten Weise auf LV ca. 83–96% (f_{gTp} = 750 Hz) ab. Wird dem durch den Tiefpass vorverzerrten Spektrum zur Beseitigung jeglicher informationstragenden Frequenzanteile, die durch Kombinationstöne oder Oberwellen entstehen können, ein zusätzliches weisses Bandpassrauschen im Bereich von 750 Hz bis 5 kHz hinzugefügt, so sinkt die Lautverständlichkeit auf LV ca. 82–93% ab. Daraus folgt, dass die hier erzielte hohe Lautverständlichkeit nur durch die im ungestörten Frequenzband von 0–750 Hz verbleibenden Signalanteile bewirkt wird. Dass diese Information speziell im Lautübergang zu suchen ist, folgt aus einem weiteren Versuch. Werden die aus den gleichen Logatomen mit einem Zeitbandpass ausgeschnittenen quasistationären Lautanteile der Vokale von 200 ms einer Frequenztiefpassbegrenzung (f_{gTp} = 750 Hz) unterworfen, so reduziert sich die Verständlichkeit auf LV ca. 5–64%. An dem Verständlichkeitsverlust von △LV etwa 20–80% wird deutlich, welcher massgebliche Beitrag der Lautübergang

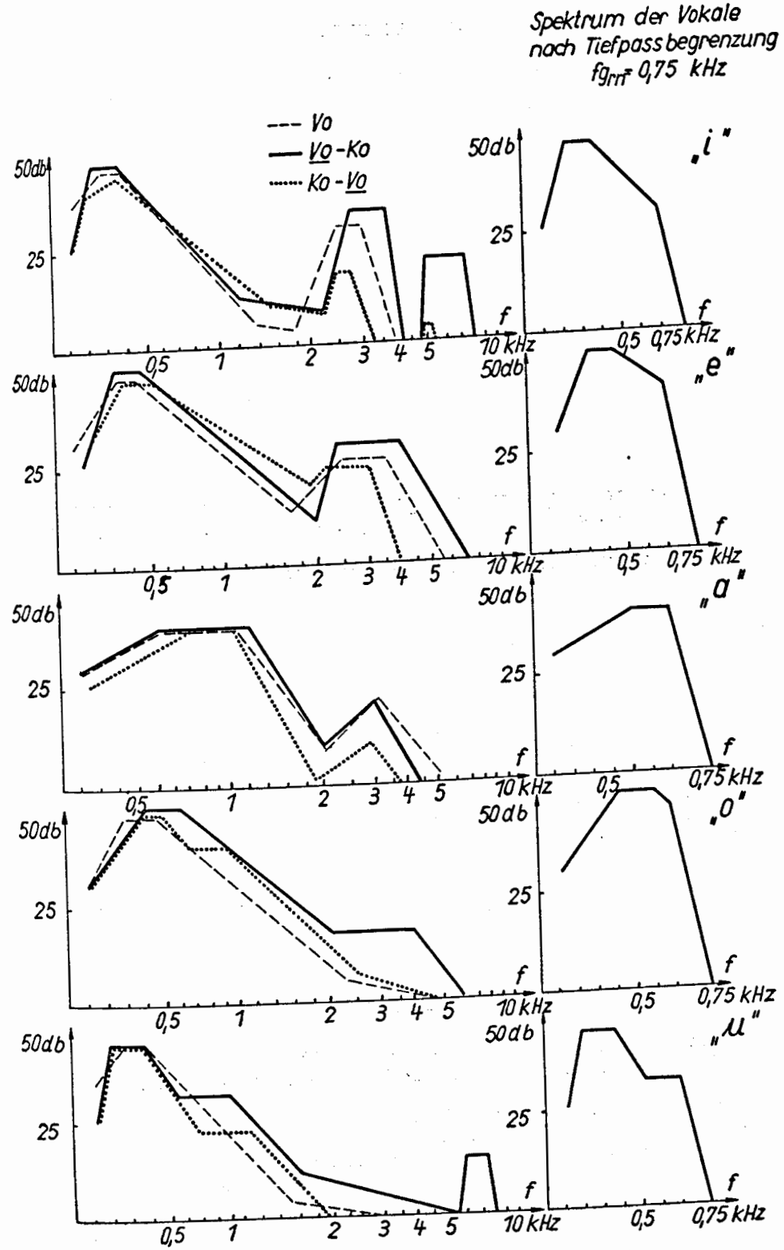

Bild 13. Mittlere Spektralverteilung von natürlichen Vokalen.

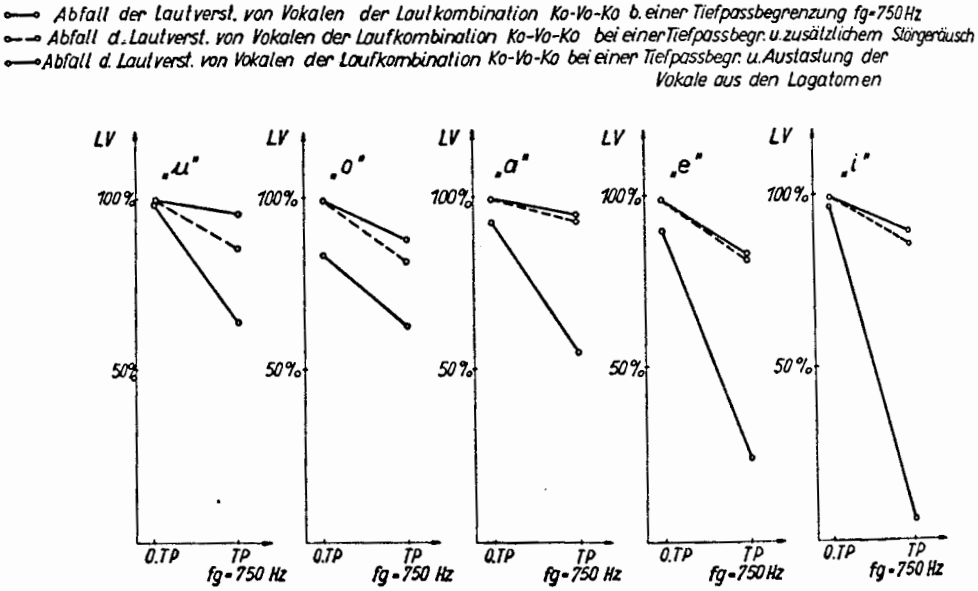

Bild 14. Lautverständlichkeit von Vokalen bei einer Frequenz- und Zeitfilterbegrenzung.

zum subjektiven Lauterkennungs-und Unterscheidungsprozess liefert. Gleichzeitig zeigen die genannten Versuche, dass bei dem subjektiven Erkennungsprozess der Merkmalcharakter der Formantstrukturierung durch die dem Lautübergang innewohnenden Gesetzmässigkeiten nicht nur ergänzt, sondern auch ersetzt werden kann.

Zusammenfassend kann gesagt werden, dass die objektive wie auch die subjektive Analyse in dem beschriebenen Rahmen in der Lage sind, wesentliche Beiträge zur Klärung der akustischen Lautstrukturen sowie deren subjektiven Unterscheidung und Erkennung zu liefern.

Institut für Fernmeldetechnik
Technische Hochschule
Dresden

LITERATUR

1 Küpfmüller, K.: "Die Entropie der deutschen Sprache", *FTZ*, 7 (1954), H. 6.
2 Fischer, F. A. "Die grundlegenden Begriffe der Frequenzanalyse". *Der Fernmelde-Ingenieur*, 6 (1952), H. 10.
3 Scudrzyk, E., *Die Grundlagen der Akustik* (Springer, Wien, 1954), Kap. XXV
4 Günther, W. *Frequenzanalyse akustischer Einschwingvorgänge* (Juris, Jürich, 1951).
5 Hölzler, E. Holzwarth, H., *Theorie und Technik der Pulsmodulation* (Springer, Berlin, 1957).
6 Springer, M. A. "Dehnung und Raffung von Schallaufnahmen", *Akustika*, 5 (1955), S. 279.