

Wellendämpfung in deutlich abgegrenzten Frequenzgebieten, wobei sich die Grenzen der Wellenausbreitung mit steigender Gyrofrequenz zu höheren Frequenzwerten, mit sinkender Gyrofrequenz nach unten verschoben, wie es das normierte Diagramm Abb. 5 anschaulich wiedergibt. Dagegen waren bei Wellen, deren Gruppengeschwindigkeit von der Kathode weg gerichtet war, keine eindeutigen Frequenzbänder zu erkennen, was durch die Translation des Plasmas begreiflich wird. Die Ergebnisse konnten noch nicht in allen Punkten quantitativ gedeutet werden. Trotzdem war es möglich, aus der oberen Grenzfrequenz der an der Anode eingekoppelten Welle die Plasma-Eigenfrequenz zu bestimmen. Es ergab sich gute Übereinstimmung mit den Ergebnissen aus den Langmuir-Messungen.

Die Anregung zu dieser Arbeit verdanke ich Herrn Professor Dr.-Ing. Dr.-Ing. E. h. W. O. SCHUMANN.

Für sein jederzeit gezeigtes Interesse möchte ich ihm herzlich danken. Großen Dank schulde ich auch dem Aeronautical Research Laboratory der United States Air Force, das durch einen Forschungsauftrag die Grundlage für diese Arbeit geschaffen hat. Der Deutschen Forschungsgemeinschaft sowie den Firmen Siemens & Halske AG., Rohde & Schwarz und Osram GmbH. danke ich vielmals für die Bereitstellung von Geräten und Material.

Literatur: [1] SCHUMANN, W. O.: Z. angew. Phys. 8, 482 (1956). — [2] SCHUMANN, W. O.: Z. angew. Phys. 10, 428 (1958). — [3] SCHUMANN, W. O.: Z. angew. Phys. 11, 333 (1959). — [4] GOULD, R. W., and A. W. TRIVELPIECE: Proc. of the Sympos. on Electronic Waveguides, S. 215, New York, April 8, 9, 10 (1958).

Dr.-Ing. JOSEF THELEN,
Elektrophysikalisches Institut
der Technischen Hochschule München

Untersuchungen über die Bedeutung bestimmter Tonfrequenzbänder für die Verständlichkeit synthetischer Sprache und über Änderungen der Sprachverständlichkeit bei Kanalvertauschungen

Von GEROLD SIEDLER

Mit 10 Textabbildungen

(Eingegangen am 27. Dezember 1960)

Die Erforschung der physikalischen Struktur menschlicher Sprache führte von der Analyse sehr bald zu Versuchen, die gewonnenen Vorstellungen von der Entstehung dieser Schallvorgänge durch eine Lautsynthese zu bestätigen. Man entwickelte eine Vielzahl mechanischer Verfahren, um Amplituden und Frequenzen der einzelnen Obertöne sowie die Lage der Formantbereiche festzustellen und im Anschluß daran Sprachlaute künstlich zu erzeugen. Vor allem die Untersuchungen von H. v. HELMHOLTZ [1] in der zweiten Hälfte des vorigen Jahrhunderts erweiterten sehr wesentlich die Kenntnisse vom Aufbau der Vokale. Die mechanischen Verfahren ließen sich jedoch nicht für das Studium der außerordentlich wichtigen Lautübergänge verwenden, so daß erst nach Einführung der Elektronenröhre und entsprechender elektrischer Anordnungen eine objektive Analyse der Sprache und insbesondere der dabei auftretenden nicht-stationären Vorgänge und deren Synthese möglich wurden. Stand zu Beginn bei der Synthese die Absicht im Mittelpunkt, durch die Nachbildung der Sprache die erarbeiteten Vorstellungen von ihrer natürlichen Entstehung zu bestätigen und vor allem eine Entscheidung zwischen der Helmholtzschen „Resonanztheorie“ und der Hermannschen „Anblasetheorie“ herbeizuführen, so schob sich später die Möglichkeit in den Vordergrund, mit Hilfe geeigneter Apparaturen die einzelnen Parameter der Sprache getrennt zu variieren und ihre Bedeutung für Klangcharakter und Informationsgehalt kennenzulernen. Als eines der wichtigsten Geräte zur Analyse und Synthese menschlicher Sprache erwies sich dabei der 1939 von DUDLEY [3] u. Mitarb. entwickelte Vocoder, der auch als Vorbild für die in der vorliegenden Arbeit benutzte Meßapparatur diente.

Prinzip und Arbeitsweise der benutzten Apparatur

Die im Institut für angewandte Physik der Universität Kiel unter der Leitung von Prof. Dr. W.

KROEBEL seit einigen Jahren durchgeführten Arbeiten am Vocoder [4, 5, 6, 7] zeigten, daß es zur Untersuchung des physikalischen Aufbaus der Sprache zweckmäßig ist, ein Gerät zu entwickeln, an das vor allem die Forderung der Vielseitigkeit gestellt werden muß, da nur eine solche Apparatur gestattet, eine genügende Anzahl von Parametern der Sprache getrennt zu variieren. Aus diesem Grunde wurde bei der Entwicklung des im folgenden beschriebenen Vocoders besonderer Wert auf eine große Anzahl von Obertonanalysekanälen gelegt; ferner ist die Möglichkeit vorgesehen, Sprachlaute zu analysieren und zu synthetisieren, die sowohl ein diskretes Linien- als auch ein Rauschspektrum besitzen, und schließlich wurde auf eine möglichst universelle Anwendbarkeit des Tonhöhenfrequenzmessers geachtet.

Das Blockschaltbild der entwickelten Apparatur zeigt Abb. 1. Die von einem Mikrofon oder Tonbandgerät gelieferte Wechselspannung aus natürlicher Sprache gelangt zunächst auf einen Dynamikbegrenzer, der dafür sorgt, daß die Eingangsspannung der darauf folgenden 23 Analysekanäle einen bestimmten Maximalwert nicht überschreitet. Der obere Kanal besorgt die Eliminierung der Grundwelle stimmhafter Laute und die Messung ihrer Frequenz. Die erhaltene frequenzproportionale Spannung wird zur Aussiebung von Frequenzanteilen, die über der Lautfolgefrequenz liegen, durch einen Tiefpaß gefiltert und steht dann der Syntheseapparatur als Steuerspannung zur Verfügung. Die beiden folgenden Kanäle dienen zur Gewinnung getrennter Angaben über Stimmhaftigkeit und Stimmlosigkeit eines Lautes, wobei ausgenutzt wird, daß stimmhafte Laute praktisch nur im unteren Frequenzbereich bis etwa 4000 Hz Obertonanteile besitzen, während man bei stimmlosen Lauten fast nur Anteile im oberen Bereich ab 1000 Hz findet. Es ist daher möglich, aus dem Vorhandensein von Frequenzanteilen unter 1000 Hz auf die Anwesenheit

eines stimmhaften Lautes und aus dem Auftreten von Anteilen über 4000 Hz auf das Vorliegen eines stimmlosen Lautes zu schließen. Der Analysekanal für stimmhafte Laute enthält deshalb einen Tiefpaß mit einer Grenzfrequenz von 1000 Hz und einen Gleichrichter mit anschließendem Tiefpaß mit einer Grenzfrequenz von 20 Hz, der eine mit Lautfolgefrequenz schwankende Gleichspannung liefert, die ein Maß für das Auftreten eines stimmhaften Lautes und in brauchbarer Näherung für die Intensität des diskreten Obertonspektrums ergibt. Entsprechend aufgebaut ist der Analysekanal für stimmlose Laute, wobei der

Umständen gleichzeitig auf die Obertonsynthesekanäle gelangenden Generatorspannungen werden in ihrer Amplitude durch je einen Modulator gesteuert, der seine Signalspannung vom Höhen- bzw. Tiefenanalysekanal erhält. Die Obertonsynthesekanäle enthalten den gleichen Filtersatz wie die entsprechenden Analysekanäle. Jedes Bandfilter siebt aus der Generatorspannung einen Frequenzbereich heraus und bildet so den Träger für den nachfolgenden Amplitudenmodulator, als dessen Signalspannung die Gleichspannung des zugehörigen Analysekanalausganges dient. Die Ausgangsspannungen sämtlicher 20 Modulatoren werden addiert und über einen Verstärker dem Lautsprecher zugeführt.

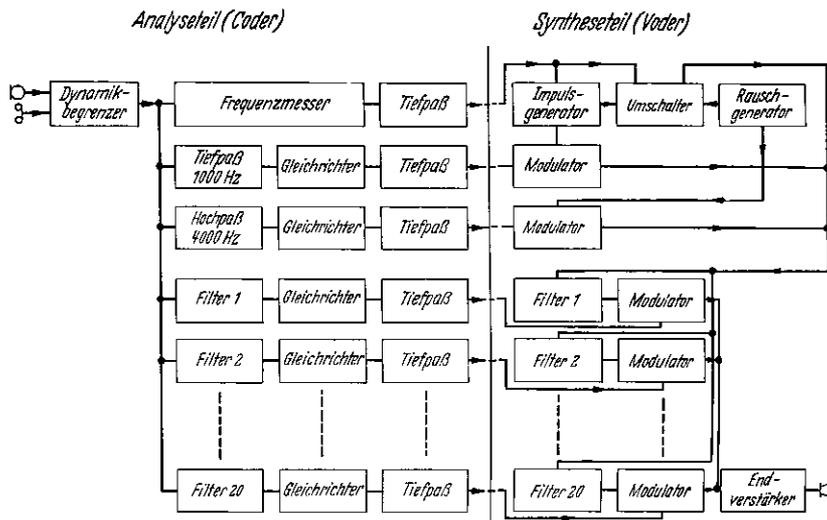


Abb. 1. Blockschaltbild des Vocoders

Tiefpaß von 1000 Hz durch einen Hochpaß der Grenzfrequenz 4000 Hz ersetzt ist. Beim Vorliegen von Ausgangsspannungen in beiden Kanälen gibt deren Verhältnis näherungsweise das Verhältnis von diskretem Oberton- zu Rauschspektrum wieder und macht es dadurch möglich, Laute mit beiden Anteilen nachzubilden.

Die übrigen 20 Analysekanäle dienen zur Ermittlung der Intensitätsverteilung der Obertöne und überdecken mit einer zu höheren Frequenzen hin wachsenden Bandbreite den Bereich von 200 bis 7000 Hz, wobei die Grenzfrequenzen der Bandfilter so gewählt wurden, daß für jeden Formanten mindestens zwei Kanäle zur Verfügung stehen. Der Aufbau dieser Analysekanäle entspricht dem der beiden bereits beschriebenen Anordnungen; sie bestehen aus Bandfilter, Gleichrichter und nachfolgendem Tiefpaß.

Das Blockschaltbild des Syntheseteils zeigt oben Impulsgenerator, Rauschgenerator und Umschalter. Betreibt man die Apparatur nach der Art des Dudley'schen Vocoders, so wird neben den Obertonkanälen nur der Frequenzmesserkanal benutzt, und dessen Tonhöhen-signal steuert einerseits die Grundfrequenz des Impuls-generators, andererseits schaltet es auch die Generatorausgänge um, so daß beim Vorliegen eines Tonhöhen-signals die Ausgangsspannung des Impuls-generators auf die Obertonsynthesekanäle gelangt, während diese sonst die Spannung des Rausch-generators erhalten. Betreibt man die Apparatur mit getrennter Steuerung von stimmhaften und stimmlosen Anteilen, so werden sämtliche Kanäle benutzt. Das Tonhöhen-signal dient jetzt nur zur Einstellung der Impuls-generatorgrundfrequenz, und die jetzt unter

Die Untersuchungen

Mit diesem Gerät wurden Messungen durchgeführt, die zu Aussagen über die Bedeutung der einzelnen Frequenzbereiche für den Informationsgehalt bei der deutschen Sprache führen sollten. Dazu wurde zunächst der Weg gewählt, die Abnahme der Verständlichkeit der synthetischen Sprache beim Fehlen von Obertonkanälen zu beobachten. Da bereits aus einer größeren Anzahl von Untersuchungen die Abhängigkeit der Verständlichkeit von oberer und unterer Grenzfrequenz bekannt ist [9, 10], wurden die Messungen von vornherein nur für das Frequenz-

band von 300 Hz bis 5000 Hz vorgenommen. Aus der nachfolgend aufgeführten Zusammenstellung der Filtergrenzen der Obertonkanäle erkennt man, daß also die Kanäle 2 bis 19 benutzt wurden.

Kanal Nr.	1	2	3	4	5	6
Frequenz [Hz]	200	300	400	500	650	800
Kanal Nr.	7	8	9	10	11	12
Frequenz [Hz]	1100	1300	1500	1700	1900	2100
Kanal Nr.	14	15	16	17	18	19
Frequenz [Hz]	2500	2700	3000	3300	3600	5000

Es zeigte sich nun sehr rasch, daß bei der gewählten Filteranordnung das Fehlen nur eines Kanals eine zu geringe Verständlichkeitsminderung verursachte, um noch mit brauchbarer Genauigkeit erfassbar zu sein. Die Messungen wurden daher in der Weise durchgeführt, daß drei nebeneinanderliegende Obertonkanäle abgeschaltet wurden, wobei diese Dreierkombination jeweils um einen Kanal weiterrückte. Auch dann war die Abnahme der Verständlichkeit noch so schwach, daß ein diesem Problem angepaßtes Testverfahren entwickelt werden mußte, um mit vernünftigen zeitlichem Aufwand zu Ergebnissen zu gelangen.

Der Weg, mit dem von SCHUBERT [11] angegebenen Worttest für Schallpegeluntersuchungen zu arbeiten, erwies sich als nicht gangbar, da die kleinen Verständlichkeitsdifferenzen sich bei den dort auch verwendeten mehrsilbigen Wörtern nicht mehr wesentlich bemerkbar machten. Diese Erscheinung mußte nach den von HAHLBROCK [12] angegebenen Ergebnissen zur Abhängigkeit des Sprachverständnisses vom Testmaterial bei Schwerhörigen als wahrscheinlich zu erwarten sein. Es wurden daher Tests aus einsilbigen

Wörtern zusammengestellt, die bezüglich ihrer prozentualen Lautzusammensetzung der von SCHUBERT benutzten Verteilung unter Vernachlässigung des bei Einsilbern nicht auftretenden auslautenden „e“ entsprechen und daher auf Grund der Lautstatistiken von KAEDING und MEIER [11, 12] als in weitem Maße repräsentativ für die deutsche Sprache gelten konnten. Um den Testaufbau für Relativmessungen möglichst günstig anzulegen und Streuungen, die durch Differenzen in der individuellen Konzentrationsfähigkeit der Testpersonen und durch Einübungseffekte und Ermüdungsercheinungen verursacht werden, so gering wie möglich zu halten, wurde nach einigen Versuchen folgende Reihenfolge gewählt:

1. Zusammenhängender Text zum Einhören (5 bis 10 min),
 2. 30 Einsilber in normaler Vocodersprache zur Einübung.
 3. etwa 70 Einsilber in normaler Vocodersprache,
 4. etwa 200 Einsilber in abgeänderter Vocodersprache,
 5. etwa 70 Einsilber in normaler Vocodersprache.
- Auf diese etwa 40 min dauernde Testgruppe folgte nach einer Pause von 10 min die zweite Gruppe:
6. 20 Einsilber in normaler Vocodersprache zur Einübung.
 7. etwa 70 Einsilber in normaler Vocodersprache,
 8. etwa 200 Einsilber in abgeänderter Vocodersprache,
 9. etwa 70 Einsilber in normaler Vocodersprache.

Um starke Streuungen infolge des unterschiedlichen Bildungsgrades der Testpersonen zu vermeiden, wurde mit einer in dieser Beziehung relativ einheitlichen Gruppe gearbeitet. Die Testpersonen waren ausnahmslos Primaner eines Gymnasiums.

Die Auswertung der Tests geschah folgendermaßen: Als richtig gewertet wurde jedes richtig verstandene Wort. Für die Wortgruppen 3 bis 5 und 7 bis 9 erhält man durch Auszählen eine Angabe über die prozentuale Wortverständlichkeit. Um die Abnahme der Verständlichkeit bei Test 4 bzw. 8 zu ermitteln, wurde zunächst aus 3 und 5 bzw. 7 und 9 eine für die jeweilige Testperson zu diesem Zeitpunkt geltende Normalverständlichkeit durch Mittelbildung bestimmt und die Abnahme in Test 4 bzw. 8 relativ dazu errechnet.

Es zeigte sich, daß die beiden Testgruppen 3 bis 5 und 7 bis 9 wegen des etwas unterschiedlichen Bekanntheitswertes der vorkommenden Wörter nicht völlig gleichwertig waren. Aus dem systematischen Gang der Ergebnisse bei der stufenweisen Verbesserung der Verständlichkeit in den unteren Frequenzbereichen ließ sich jedoch graphisch eine Beziehung zwischen den beiden Testgruppen ermitteln, und die Reduktion auf einen mittleren Verständlichkeitswert ergab die Darstellung der Messungsergebnisse in Abb. 2. Man erkennt den im allgemeinen aus je sechs Messungen gemittelten Wert der relativen Verständlichkeit beim Fehlen von jeweils drei Filterbereichen in Abhängigkeit von deren Lage im Frequenzband. Auffallend ist die beträchtliche Verständlichkeitsminderung beim Fortlassen tiefer Frequenzen, die sich aus der Lage einer Reihe von wesentlichen Formanten in diesem Gebiet erklärt. Es genügen andererseits offenbar bereits ein bis zwei Obertöne im Bereich

von 300 bis 400 bzw. 500 Hz, um die Verständlichkeit der Laute mit tiefem Hauptformanten erheblich zu verbessern.

Es erscheint nach diesen Untersuchungen der Frequenzbereich bis etwa 1300 Hz als besonders empfindlich gegenüber Veränderungen. Das Absinken der Verständlichkeit beim Fehlen des Bereichs von 3000 bis 5000 Hz erklärt sich aus der Bedeutung der stimmlosen Laute für die deutsche Sprache, da wesentliche Anteile vieler dieser Laute in dem genannten Gebiet liegen. Die gefundene Kurve befindet sich in guter qualitativer Übereinstimmung mit den von KROHM [7] bei ähnlichen Untersuchungen festgestellten Werten und zeigt, daß eine Gewichtsfunktion für die Wichtigkeit der Frequenzbereiche ein Maximum zwischen

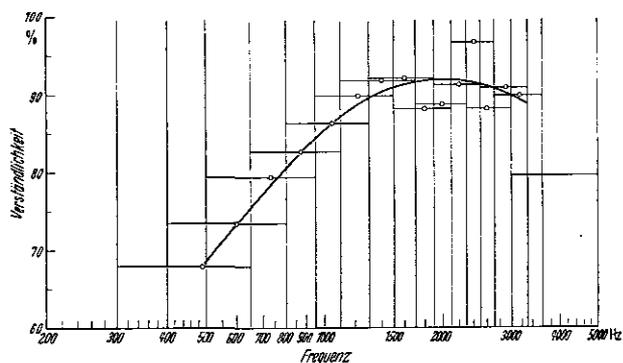


Abb. 2. Verständlichkeitsabnahme beim Fehlen von Kanälen

300 und 1000 Hz besitzen muß. Eine von FLETCHER [10] aus Lautabbauuntersuchungen mittels Hoch- und Tiefpässen gewonnene Gewichtsfunktion zeigt für die englische Sprache ein Maximum bei 600 bis 800 Hz.

Es liegt nun nahe, aus den Ergebnissen der Abb. 2 auf die Abhängigkeit der Verständlichkeit von der Vertauschung von Frequenzbereichen zu schließen. Es ist anzunehmen, daß vor allem die Vertauschung der als besonders wichtig erkannten Bereiche eine starke Verständlichkeitsminderung ergibt. Um das nachzuprüfen, wurde eine Testreihe durchgeführt, bei der — ausgehend vom normalen Frequenzband — nacheinander je zwei Kanäle vertauscht wurden, und zwar einmal in der Mitte und einmal an den Enden beginnend. Den Aufbau des Tests zeigt die folgende Zusammenstellung, wobei die Ziffern die Nummern der Filterkanäle bezeichnen:

1.	2	19
2.	19, 3	18, 2
3.	19 18, 4	17, 3 2
4.	19 .. 17, 5	16, 4... 2
5.	19 16, 6	15, 5 2
6.	19 15, 7	14, 6 2
7.	19 14, 8	13, 7 2
8.	19 13, 9	12, 8 2
9.	19 12,10 11, 9 2
10.	19 2
11.	2 19
12.	2 8,12 9,13 19
13.	2 7,13 8,14 19
14.	2 6,14 7,15 19
15.	2 5,15 6,16 19
16.	2 .. 4,16 5,17 .. 19
17.	2 3,17 4,18 19
18.	2,18 3,19
19.	19 2

Da die Abnahme der Verständlichkeit bei diesen Tests beträchtlich ist, wurde der von SCHUBERT [11] angegebene Satztest verwendet, wobei für jede Messung eine Gruppe von etwa 40 Sätzen benutzt wurde. Den durch die vorangegangenen Tests bereits geübten Testpersonen genügten wenige Sätze, um sich auf den jeweiligen Satztest einzustellen. Die Testpersonen erhielten daher nacheinander 3 Gruppen von 4 Einübungssätzen und etwa 40 Testsätzen vorgesprochen. Die Pausen zwischen diesen Gruppen betragen einige Minuten. Die mittlere Lautstärke lag bei diesen wie bei allen vorangegangenen Tests bei etwa 70 phon. Der Vocoder wurde stets ohne Tiefen- und Höhenanalysekanal betrieben.

Abb. 3 zeigt die ermittelten Werte der Wortverständlichkeit beim Satztest. Es handelt sich um die Mittelwerte aus jeweils drei Messungen; nur der Wert für die unveränderte Vocodersprache ergab sich aus Tests mit sämtlichen 24 Testpersonen. Die in Abb. 4 dargestellte Satzverständlichkeit veranschaulicht besonders deutlich die Empfindlichkeit der Sprache gegen die Vertauschung von hohen und tiefen Frequenzen, wobei allerdings diese beiden Bilder wegen der starken Streuung der Werte nur als qualitativer Überblick gelten dürfen.

Die Schaltung

Abschließend sollen noch Einzelheiten der Schaltung der benutzten Apparatur dargestellt werden. Die Wechselspannung von Mikrofon oder Tonbandgerät gelangt zunächst auf einen Vorverstärker (Abb. 5), der auf fünf mittlere Eingangsspannungsamplituden einstellbar ist. Eingang 1 besitzt einen Anpassungsübertrager für dynamische Mikrofone mit 200 Ω Ausgangsimpedanz, auf den eine gegengekoppelte Verstärkerpentode folgt. Die Eingänge 2, 3, 4 führen auf eine im Kathodenzweig umschaltbare Verstärkertiode, während an Eingang 5 nur ein Spannungsteiler liegt. Über einen Schalter gelangen die Ausgangsspannungen dieser Stufen auf einen Lautstärkereglер, werden anschließend verstärkt und über eine Kathodenstufe niederohmig ausgekoppelt. Der darauf folgende Dynamikbegrenzer sorgt dafür, daß der Wechselspannungspegel einen bestimmten Wert nicht wesentlich überschreitet, so daß es zu keiner Übersteuerung der Analysekanäle kommen kann. Die Schaltung ist in [7] beschrieben und wurde nur geringfügig gegenüber der dort angegebenen verändert. Über einen Kathodenverstärker gelangt die Ausgangsspannung des Begrenzers anschließend auf die verschiedenen Analysekanäle.

Der Frequenzmesser arbeitet nach einem von DOLANSKY [8] angegebenen Prinzip. Das Oszillogramm eines stimmhaften Lautes zeigt infolge des mehr oder

weniger starken Hervortretens der Grundwelle praktisch immer deutlich ausgeprägte Spitzen. Bei einigen Lauten, z. B. bei „o“ und „i“, ist die Amplitude der ersten Oberwelle indessen so groß, daß die durch sie hervorgerufene Spitze nur wenig kleiner als die der Grundwelle ist. Es gelingt nun, durch Gleichrichtung der Spannung mit einer Zeitkonstanten des zugehörigen Siebgliebes, die in der Größenordnung der Grundtonwellenlänge liegt, und anschließende „Differentia-

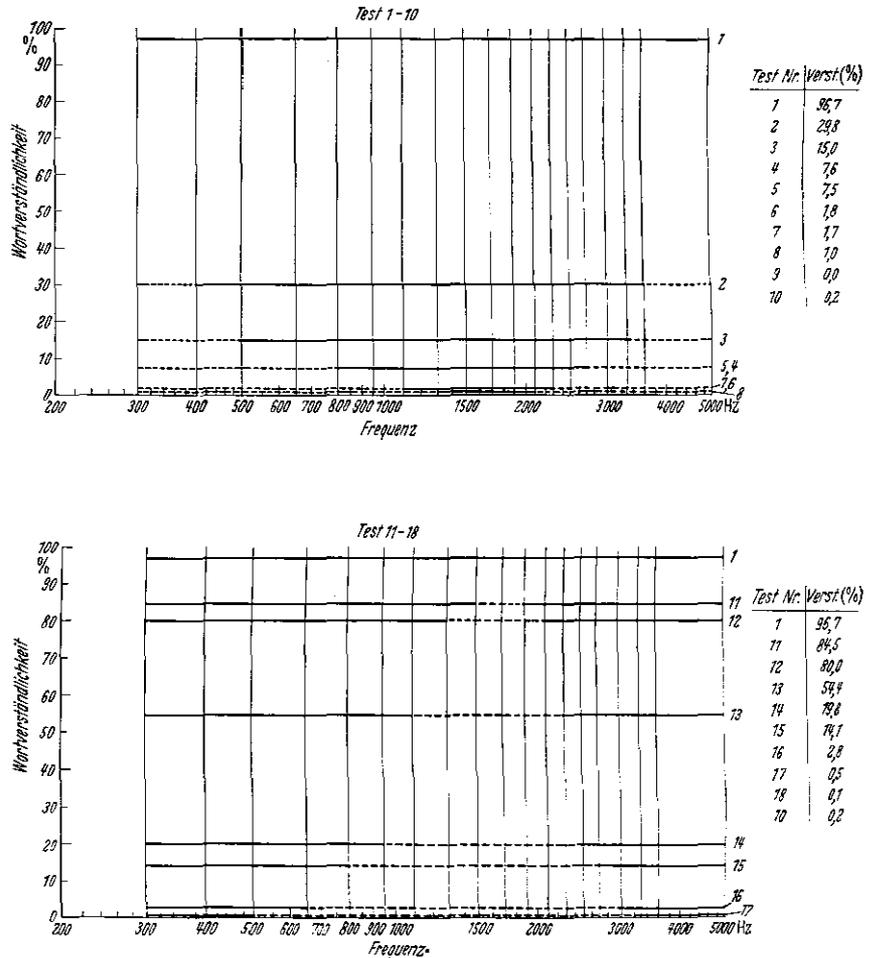


Abb. 3. Abnahme der Wortverständlichkeit bei Kanalvertauschungen

tion“ mit Hilfe eines RC-Gliedes, dessen Zeitkonstante um den Faktor 2 bis 15 kleiner ist, das Verhältnis der Amplituden von Grund- und Oberwellenspitze zugunsten der Grundwelle zu verbessern. Mehrfache Wiederholungen dieses Vorganges führen zu einer vollständigen Unterdrückung der Oberwellenanteile, so daß man schließlich pro Grundwellenlänge eine einzige Spannungsspitze erhält. Benutzt man diese zur Auslösung eines monostabilen Multivibrators, so erhält man durch Integration der entstehenden Rechteckimpulse eine frequenzproportionale Gleichspannung.

Die praktische Ausführung der Schaltung zeigt Abb. 6. Auf einen Eingangsspannungsregler folgt eine frequenzabhängige Verstärkerstufe (Rö 1), die durch die parallel zum Anodenwiderstand liegende Kapazität eine kräftige Abschwächung der Höhen besorgt, so daß die dann noch verbleibenden Rauschanteile von stimmlosen Lauten keine Störungen mehr hervorrufen können. Es folgt ein Kathodenverstärker (Rö 2) als niederohmiger Eingang für die erste Gleichrichter-

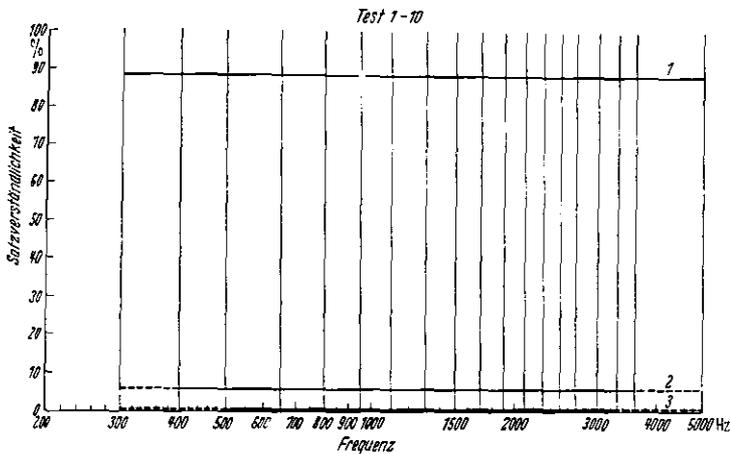
stufe, die negative Spannungsspitzen verarbeitet. Die schwach gesiebte Spannung wird verstärkt (*Rö 3*) und differenziert, wobei die unerwünschten negativen Spannungsspitzen durch eine Diode beseitigt werden. Es folgen nun zwei entsprechend aufgebaute Stufen, die jeweils aus Kathodenverstärker, Gleichrichter mit Siebglied, Anodenwiderstandsverstärker, Differenzierglied und zugehöriger Diode bestehen, wobei die Gleichrichter eine einstellbare Vorspannung erhalten

Anforderungen gestellt: Die frequenzproportionale Ausgangsspannung soll negativ sein und sich für wachsende Frequenz dem Wert Null nähern, wobei die der Frequenz 300 Hz entsprechende Gleichspannung mindestens 2 V betragen soll. Ferner soll die Ausgangsspannung beim Fehlen eines Grundtones den Wert Null annehmen. Das führte zu folgender Schaltung: Der negative Rechteckimpuls an der Anode von *Rö 10* wird differenziert und die positive Spitze des ent-

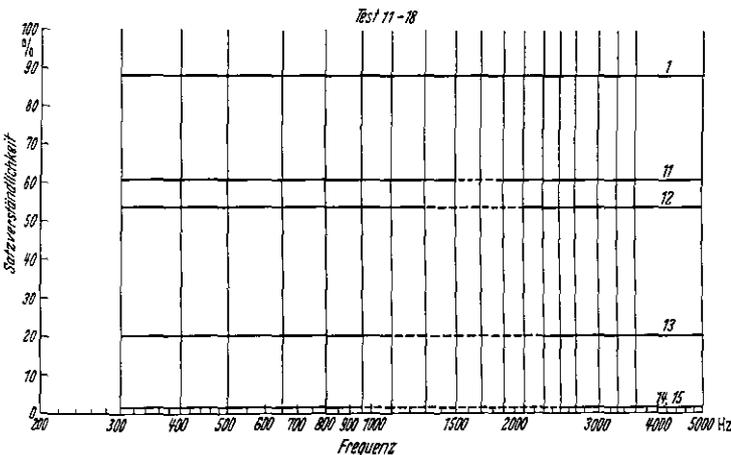
standenen Signals durch eine Diode beseitigt. Der verbleibende negative Nadelimpuls löst einen monostabilen Multivibrator (*Rö 11, Rö 12*) aus, dessen Impulsbreite wegen der Grenzfrequenz von 300 Hz durch den Einstellregler *E 4* auf 30 msec eingestellt wird. Von dem an der Anode von *Rö 11* auftretenden positiven Rechteckimpuls mit negativer Spitze wird der negative Anteil nach Gleichrichtung zur Frequenzanzeige mittels eines Drehspulinstruments benutzt. Am Gitterableitwiderstand von *Rö 12* wird ein in seiner Höhe durch *E 5* einstellbarer positiver Impuls angekoppelt, der eine als Kathodenverstärker geschaltete Leistungspentode steuert, an deren Kathode ein positiver Rechteckimpuls entsteht. Die entstehende Impulsfolge kann nun durch einen Tiefpaß integriert werden.

Um die letzte der oben aufgestellten Forderungen zu erfüllen, wird nun andererseits der an der Anode von *Rö 9* entstehende positive Rechteckimpuls über einen Kathodenverstärker (*Rö 17*) und eine Diode zur Aufladung eines Kondensators in einem RC-Glied benutzt, dessen Zeitkonstante 1 sec beträgt. Man erhält an diesem Kondensator nach seiner Aufladung eine positive Spannung, die während einer Periode der Grund-

schwingung praktisch zeitproportional abnimmt. Sie steuert einen Schmitt-Trigger (*Rö 15, Rö 16*) in der Weise, daß *Rö 15* immer dann geöffnet ist, wenn die Spannung am RC-Glied genügend groß ist. Die Einstellregler *E 9* und *E 10* sind nun so eingestellt, daß der Schmitt-Trigger bei einer Impulsfolgefrequenz von 60 Hz noch hin- und zurückkippt, während bei höheren Frequenzen *Rö 15* stets geöffnet und *Rö 16* gesperrt bleiben. In diesem Falle besitzt die Kathode der nachfolgenden Kathodenverstärkerstufe (*Rö 14*) ein negatives Potential, das beim Fehlen eines Grundtones durch entsprechende Einstellung des Reglers *E 7* so ansteigt, daß am Eingang des Ausgangstiefpasses die Spannung Null auftritt. Man erhält am Ausgang des Frequenzmessers daher bei Anwesenheit stimmhafter Laute eine negative, für wachsende Frequenzen gegen Null gehende Spannung, die am Ende des Lautes infolge des Zurückklippens des Schmitt-Triggers (*Rö 15, Rö 16*) den Wert Null annimmt. Damit sind die gestellten Bedingungen erfüllt.



Test Nr.	Verst. (%)
1	100,0
2	5,8
3	9,6
4	0,0
5	·
6	·
7	·
8	·
9	·
10	0,0



Test Nr.	Verst. (%)
1	100,0
11	100,0
12	53,8
13	20,3
14	1,5
15	1,5
16	0,0
17	·
18	·
19	0,0

Abb. 4. Abnahme der Satzverständlichkeit bei Kanalvertauschungen

und dadurch das Amplitudenverhältnis von Grundton zu Obertonspitze weiter verbessern. Diese drei Stufen reichen aus, um den Einfluß der Oberwellen vollständig zu beseitigen. Es folgt eine Verstärkerröhre (*Rö 8*), die durch das Signal übersteuert wird, so daß sich wegen der vergrößerten Flankensteilheit durch Differentiation ein Dreieckimpuls gewinnen läßt, der zur Auslösung eines Schmitt-Triggers dient. An den Anoden von *Rö 9* und *Rö 10* kann man einen positiven bzw. negativen Rechteckimpuls abnehmen.

Die bisher erläuterte Frequenzmesserschaltung gestattet es, die Grundwellen eines stimmhaften Lautes beliebiger Tonhöhe als Impulsfolge darzustellen. Da bei gesprochenen Lauten die Grundfrequenz etwa 80 bis 200 Hz für Männerstimmen und 200 bis 300 Hz für Frauenstimmen beträgt und gesungene Laute nicht analysiert werden sollten, wurde der nachfolgende Frequenzmesserteil für eine obere Grenzfrequenz von 300 Hz ausgelegt. An diesen Teil der Schaltung wurden mit Rücksicht auf die Syntheseapparatur folgende

analysekanal erhaltenen Gleichspannungen steuern, die beiden resultierenden Wechselspannungen dann summieren und eine genügend große Summenspannung zur Aussteuerung der Obertonsynthesekanäle liefern. In Betriebsart 2 soll er, wie im allgemeinen bei Vocodern üblich, beim Fehlen eines Tonhöhen-signals am Ausgang eine Rauschspannung konstanter mittlerer Amplitude abgeben, während beim Auftreten eines Tonhöhen-signals von der Rausch- auf eine Impulsspannung konstanter Amplitude umgeschaltet wird. Die Ausführung der Schaltung zeigt Abb. 10a.

Zur Gleichspannungsmodulation in Betriebsart 1 werden Anodenmodulatoren mit einer über eine Gitterwiderstandsanzapfung wirkenden Gegenkopplung verwendend. Das Kennlinienfeld dieses Modulators zeigt Abb. 10b. Der Störspannungspegel liegt um mehr als 46 db unter dem maximalen Nutzspannungspegel. Als Anodenspannung dient die Gleichspannung des Tiefen- bzw. Höhenanalysekanals von maximal 10 V. Diese Gleichspannung gelangt über ein Siebglied, das unerwünschte Brummspannungen unterdrückt und ein Überkoppeln der natürlichen Sprache verhindert — diese Störungen sind bei dem hier durchgeführten getrennten Aufbau von Netzversorgung und Analyse- und Syntheseteil auch bei sorgfältiger Führung der Erdleitungen infolge der hohen Empfindlichkeit der Modulatoren schwer vermeidbar —, auf die Anodenmodulatorröhren (*Rö 1* bzw. *Rö 3*), an deren Gitter die Impuls- bzw. Rauschspannung liegt. Die gleichspannungsmodulierten Wechselspannungen werden anschließend verstärkt (*Rö 2* bzw. *Rö 4*), über Entkopplungswiderstände am Gitter von *Rö 5* summiert, durch diese Röhre nochmals verstärkt und über einen Kathodenverstärker (*Rö 6*) ausgekoppelt.

In Betriebsart 2 bekommen die Modulatorröhren eine feste Anodenspannung von 10 V, so daß an ihren Anoden stets eine dem Impuls- bzw. Rauschgeneratorausgang entsprechende Spannung konstanter mittlerer Amplitude zur Verfügung steht. Zur Umschaltung dient ein Transistor-Schmitt-Trigger, der beim Fehlen einer Tonhöhen-gleichspannung die richtige Arbeitspunkteinstellung für *Rö 4* besorgt, beim Auftreten einer Tonhöhen-spannung von mehr als -2 V aber kippt und mit einer Gittervorspannung von -12 V *Rö 4* sperrt, so daß die Rauschspannung nicht mehr auf den Ausgangsverstärker (*Rö 5*, *Rö 6*) gelangen kann. Da mit dem Auftreten der Tonhöhen-gleichspannung auch der Impuls-generator einsetzt und *Rö 2* im Gegensatz zu *Rö 4* stets geöffnet ist, erhält man jetzt am Ausgang statt der

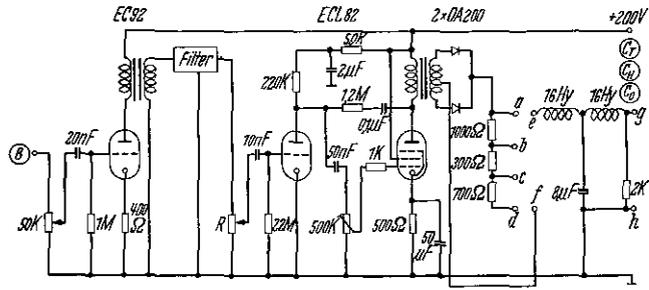


Abb. 7. Analysekanal

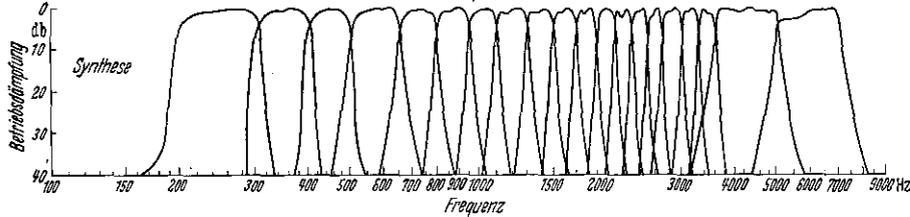
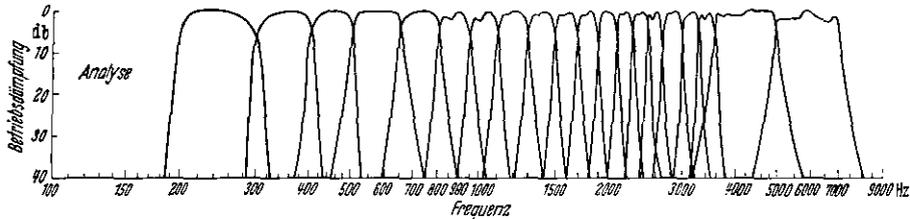
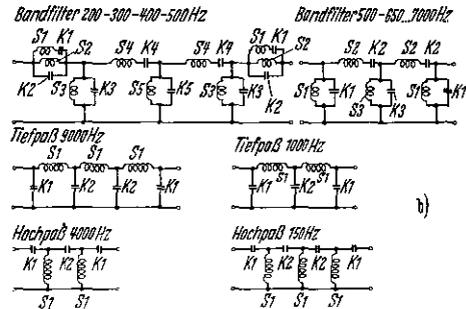
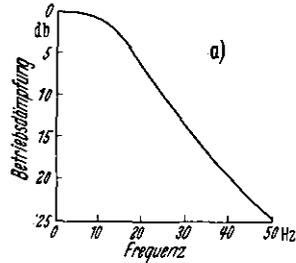


Abb. 8 a—c. Aufbau und Frequenzgang der Filter

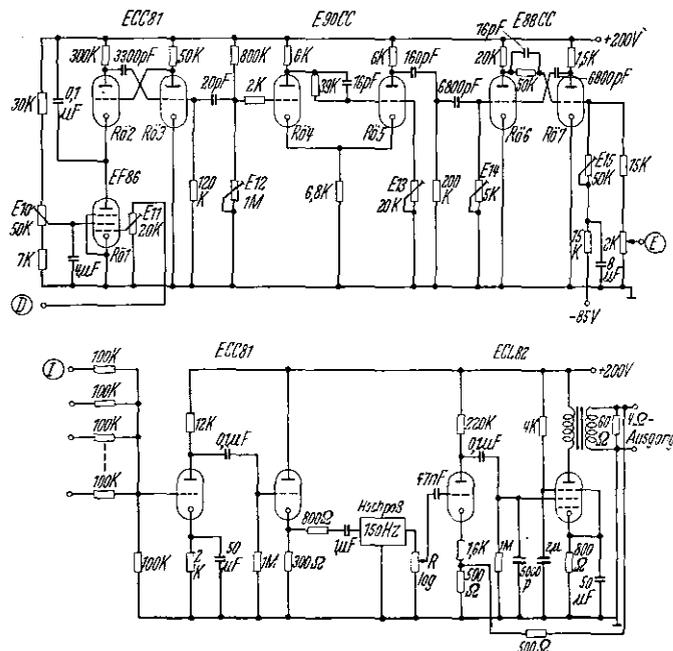


Abb. 9. Impuls-generator und Ausgangsverstärker

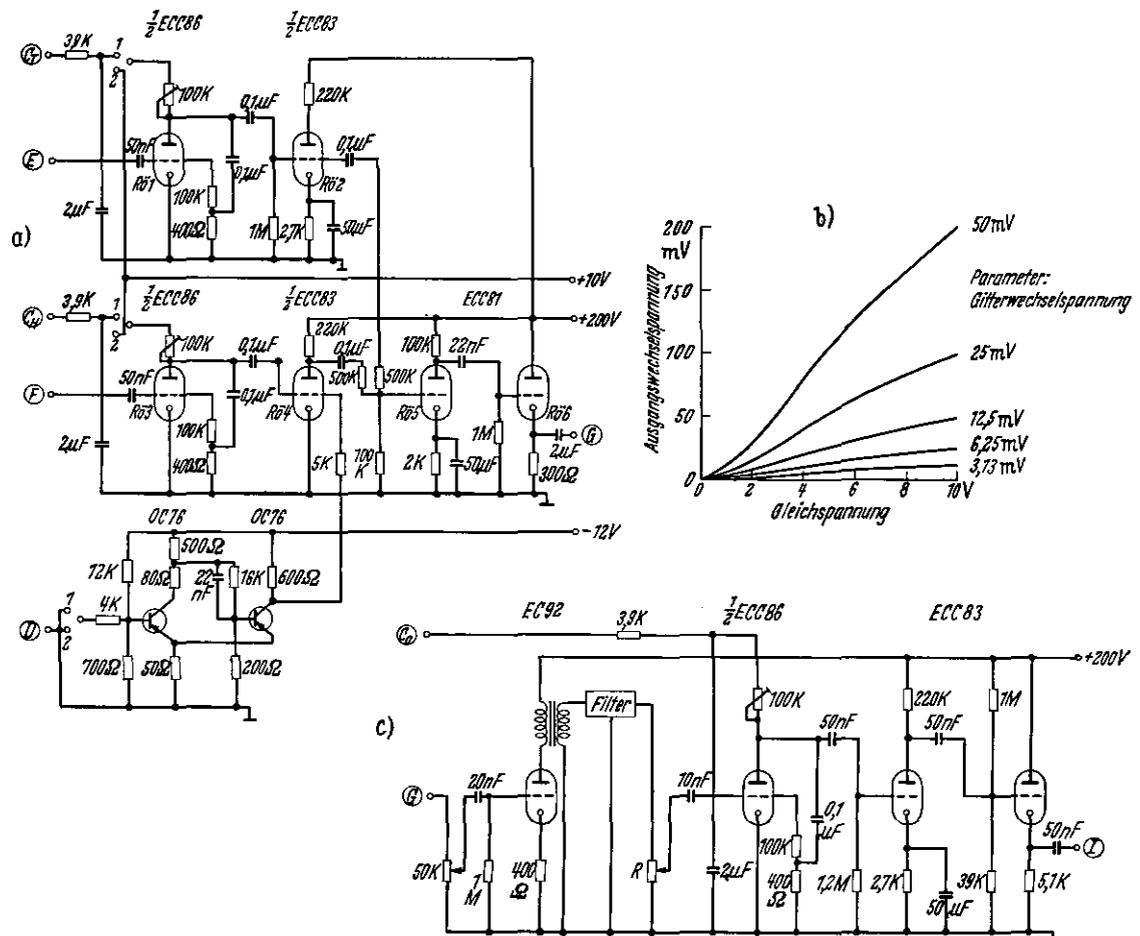


Abb. 10 a—c. Modulator bzw. Umschalter, Modulatorkennlinienfeld und Obertonsynthesekanal

Rauschspannung eine Impulsfolge konstanter Amplitude.

Die Impuls- bzw. Rauschspannung gelangt nun auf die Obertonsynthesekanäle, deren Aufbau Abb. 10 c zeigt. Auf einen Eingangsregler folgt wie bei den Analysekanälen eine Transformatorverstärkerstufe zur Anpassung an das Bandfilter, von dessen Abschlußwiderstand R die Wechselspannung auf das Gitter einer Anodenmodulatorröhre geführt wird, die ihre variierende Anodenspannung aus dem entsprechenden Obertonanalysekanal erhält. Die amplitudenmodulierte Wechselspannung wird verstärkt und über einen Kathodenverstärker ausgekoppelt. Die Ausgangsspannungen sämtlicher 20 Synthesekanäle werden über Entkopplungswiderstände summiert (Abb. 9 b), verstärkt, zur Entfernung tieffrequenter Störspannungen durch einen Hochpaß mit der Grenzfrequenz 150 Hz gefiltert und über den als Lautstärkeregler dienenden Abschlußwiderstand des Passes dem zweistufigen, gegengekoppelten Endverstärker zugeführt.

Zusammenfassung

Es werden Prinzip und Arbeitsweise eines für wissenschaftliche Zwecke entwickelten Vocoders beschrieben und anschließend Ergebnisse von Messungen dargestellt, die Aussagen über die Bedeutung der einzelnen Frequenzbereiche für die Verständlichkeit deutscher Sprache zu machen gestatten. Mit einem speziell entwickelten Worttest wurde die Verständlichkeitsabnahme der Sprache beim Fehlen von Oberton-

bereichen festgestellt. Ferner werden Angaben über den Verständlichkeitsverlust beim Vertauschen von Obertonfrequenzbändern gemacht, die sich aus Untersuchungen mit Satztests ergaben. Abschließend werden Einzelheiten der verwendeten Schaltungen erläutert.

Die Arbeiten zur vorliegenden Dissertation wurden in den Jahren 1958 bis 1960 im Institut für angewandte Physik der Christian-Albrechts-Universität zu Kiel durchgeführt. Meinem hochverehrten Lehrer, Herrn Prof. Dr. W. KROEBEL, danke ich für die Aufgabenstellung und die stetige wissenschaftliche Betreuung der Arbeit. Mein Dank gilt ferner dem Direktor der Universitätsklinik für Hals-Nasen-Ohrenkrankheiten zu Kiel für die Möglichkeit, die Tests in den Räumen dieser Klinik unter Verwendung der dort vorhandenen technischen Hilfsmittel durchführen zu können.

Literatur: [1] HELMHOLTZ, H. v.: Die Lehre von den Tonempfindungen. Braunschweig: Vieweg 1870. — [2] STUMPF, H.: Die Sprachlaute. Berlin: Springer 1926. — [3] DUDLEY, H.: Remaking Speech. J. Acoust. Soc. Amer. 11, 169 (1939). — [4] ROTIGORDT, U.: Entwicklung eines Gerätes zur synthetischen Sprachwiedergabe. Diplomarbeit, Kiel 1954. — [5] HAUSER, O.: Ein Gerät zur synthetischen Erzeugung stationärer Sprachlaute. Diplomarbeit, Kiel 1954. — [6] GÖKE, H.: Die Synthese einer kodifizierten Sprache mittels eines Vocoders und dessen experimentelle Entwicklung. Staatsexamenarbeit, Kiel 1956. — [7] KROHM, G.: Erfahrungen bei der Entwicklung eines Vocoders und Messungen der mit ihm erhaltenen Verständlichkeit. Diss. Kiel 1957. — [8] DOLANSKY, L. O.: An Instantaneous Pitch-Period Indicator. J. Acoust.

Soc. Amer. 27, 67—72 (1955). — [9] TRENDLENBURG, F.: Einführung in die Akustik. Berlin: Springer 1939. — [10] FLETCHER, H.: Speech and Hearing in Communication. New York: D. van Nostrand 1953. — [11] SCHUBERT, K.: Sprachhörprüfmethoden. Stuttgart: Georg Thieme 1958. — [12] HAHNBROCK, K.-H.: Sprachaudiometrie. Stuttgart: Georg Thieme

1957. — [13] RANKE, O.F., u. H. LULLIES: Gehör, Stimme, Sprache. Berlin-Göttingen-Heidelberg: Springer 1953. — [14] MEYER-EPPLER, W.: Elektrische Klangerzeugung. Bonn: Dümmler 1949.

Dr. GEROLD SIEDLER,
Institut für angewandte Physik der Universität Kiel

Zur zerstörungsfreien Messung von Zusammensetzung und Schichtdicke kleiner Bereiche in dünnen Schichten mit Hilfe der Röntgenfluoreszenz

Von REINHARD WEYL

Mit 6 Textabbildungen

(Eingegangen am 18. Januar 1961)

1. Einleitung

Die Bestimmung der Zusammensetzung und Schichtdicke einer dünnen, verschiedenatomigen Schicht beruht auf der Messung der Massenbelegungen der verschiedenen Elemente. Die aus den Massenbelegungen berechnete Schichtdicke setzt allerdings voraus, daß die Schicht homogen ist und dieselbe Dichte hat wie das kompakte Material.

Eine genaue, geeignete Analysenmethode ist die Röntgenfluoreszenzanalyse. Sie gestattet, die Massenbelegung in einfacher Weise zu messen: Polychromatische Röntgenstrahlung regt die Elemente der Schicht zur Emission des charakteristischen Linienspektrums an. Bei hinreichend dünnen Schichten ist die Intensität der Röntgenlinien der Massenbelegung angenähert proportional, da sowohl die anregende als auch die emittierte Röntgenstrahlung innerhalb der dünnen Schicht nur sehr schwach absorbiert werden. Im allgemeinen wird nur die $K\alpha$ -Linie des betreffenden Elementes als stärkste Linie des Fluoreszenzspektrums ausgemessen.

Ein wesentlicher Vorzug der Röntgenfluoreszenzanalyse ist die Möglichkeit, kleine Bereiche der Schicht bis herab zu etwa 1 mm^2 Flächeninhalt mit hinreichender Genauigkeit bestimmen zu können. Es können auf diese Weise lokale Unterschiede in der Schichtzusammensetzung ermittelt werden.

Für die Analyse ist von besonderer Bedeutung, daß die Fluoreszenzintensität pro Einheit der Massenbelegung nahezu unabhängig von der chemischen Zusammensetzung ist. Bei mehratomigen Präparaten besteht nämlich die Möglichkeit, daß die Elemente mit niedriger Ordnungszahl nicht nur primär durch die von der Röntgenröhre kommende Strahlung, sondern auch sekundär durch die Fluoreszenzstrahlung der Elemente mit höherer Ordnungszahl angeregt werden. Wie eine im Anhang II beschriebene theoretische Betrachtung zeigt, ist das Verhältnis ν der sekundären zur primären Fluoreszenzintensität für dünne Schichten vernachlässigbar klein und proportional der Massenbelegung des schwereren sekundärangeregenden Elementes. Als Beispiel ergibt die Rechnung für eine etwa 1000 \AA dicke Nickel-Eisen-Schicht (80/20) das Verhältnis

$$\nu = \text{const} \cdot L_{\text{Ni}} = 0,02 \quad (L_{\text{Ni}} = \text{Nickelbelegung}).$$

Der geringe Einfluß der Sekundärangeregung wird verständlich, wenn man bedenkt, daß bei dünnen

Schichten schon die primäre Fluoreszenzintensität im Verhältnis zur Intensität der primär anregenden Röhrenstrahlung sehr gering ist.

Experimentell kann dies auf folgende Weise demonstriert werden: Gemessen wird die $K\alpha$ -Fluoreszenzstrahlung einer reinen, etwa 300 \AA dicken Eisenschicht. Anschließend wird auf die Eisenschicht eine etwa doppelt so dicke Nickelschicht aufgedampft. Eine wiederholte Messung der Eisenstrahlung ergibt bei einer Meßgenauigkeit von 2% dieselbe Intensität. Das bedeutet: Durch die in der Nickelschicht erzeugten $\text{Ni}K\alpha$ -Quanten wird nur eine vernachlässigbar kleine Menge sekundärer $\text{Fe}K\alpha$ -Quanten gebildet.

Die Abhängigkeit der $K\alpha$ -Linienintensität von der Massenbelegung wurde von RHODIN [1], LIEBHAFSKY und ZEMANY [2] mit Hilfe dünner Schichten bekannter Massenbelegung gemessen. Für den Fall, daß die Herstellung derartiger Eichschichten nicht möglich ist, braucht man eine Methode, die ohne Eichschichten auskommt und in diesem Sinne eine absolute Messung gewährleistet. Im folgenden wird eine derartige Methode für das Beispiel dünner Nickel-Eisen-Schichten bis zu 1000 \AA Dicke beschrieben.

2. Bestimmung der Massenbelegung an dünnen Nickel-Eisen-Schichten

2.1 Theoretische Grundlagen

Für die absolute Bestimmung der Massenbelegung ist es notwendig, den Strahlengang innerhalb einer dünnen Schicht zu diskutieren (Abb. 1) und die Intensität der Fluoreszenzstrahlung zu berechnen.

Unter Massenbelegung wird im allgemeinen die Masse eines Elementes pro Flächeneinheit der dünnen Schicht verstanden.

Das Ziel dieses Abschnittes ist, eine Gleichung zu finden, die die lineare Abhängigkeit der Anzahl gemessener Fluoreszenzquanten $(n_{L:\text{Ni}K\alpha})_{\text{gem}}$ von der Nickelbelegung der Nickel-Eisen-Schicht L_{Ni} beschreibt¹:

$$L_{\text{Ni}} = K_{\text{Ni}K\alpha} \cdot (n_{L:\text{Ni}K\alpha})_{\text{gem}}. \quad (1)$$

Auf die Berechnung der Konstanten $K_{\text{Ni}K\alpha}$ wird im Anhang I näher eingegangen. Hier sei nur der prinzipielle Rechengang wiedergegeben. Für $K_{\text{Ni}K\alpha}$

¹ Da die Sekundärangeregung bei dünnen Schichten bis zu etwa 1000 \AA Dicke vernachlässigt werden kann, gelten die folgenden Gleichungen auch für $\text{Fe}K\alpha$ -Quanten.