

Die Tonhöhenempfindung

Citation for published version (APA):

Ritsma, R. J., & Cardozo, B. L. (1962). Die Tonhöhenempfindung. *Philips' Technische Rundschau*, 24(11/12), 357-364.

Document status and date:

Gepubliceerd: 01/01/1962

Document Version:

Uitgevers PDF, ook bekend als Version of Record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

an den arbeitenden Menschen, die wir zu Beginn dieses Aufsatzes als für die Industrie wichtig genannt haben, die aber ebensoviel Bedeutung hat für die Arbeit in einem Laboratorium, und sogar für die Art und Weise, wie jedermann in der privaten Sphäre sein Leben organisiert.

Literatur

Colin Cherry, On human communication, Chapman & Hall, London 1957.
D. E. Broadbent, Perception and communication, Pergamon Press, London 1958.

W. A. Rosenblith, Sensory communication, Wiley, New York 1961.

S. S. Stevens, Handbook of experimental psychology, Chapman & Hall, London 1951.

Heinz Zemanek, Elementare Informationstheorie, Oldenbourg, Wien 1959.

Zusammenfassung. Einleitung über Wesen, Zielsetzung und Arbeitsgebiete des Instituts für Perzeptionsforschung, das als gemeinschaftliche Stiftung vom Philips-Forschungslaboratorium Eindhoven und von der Technischen Hochschule Eindhoven errichtet worden ist und im September 1962 auf fünfjähriges Bestehen zurückblicken konnte. Der Aufsatz ist gleichzeitig eine Einführung zu fünf Artikeln über verschiedene in dem Institut im Gang befindliche Untersuchungen und enthält eine allgemeine Betrachtung über die Probleme psychophysischer Messungen.

DIE TONHÖHENEMPFINDUNG

von R. J. RITSMA *) und B. LOPES CARDOZO *).

534.321

Übersicht; Sinusschwingungen und zusammengesetzte Klänge

Das Tonhöhenempfinden, das bei geübteren Personen sehr ausgeprägt sein und mit erstaunlicher Genauigkeit arbeiten kann, beruht auf einem Mechanismus, der noch nicht vollständig bekannt ist. Wir wollen in diesem Artikel kurz darlegen, was wir heute mit Sicherheit über die Tonhöhenempfindung wissen, und eine Anzahl Experimente beschreiben, auf denen diese Erkenntnisse beruhen.

Um Mißverständnisse zu vermeiden, weisen wir einleitend darauf hin, daß die wahrgenommene Tonhöhe eine subjektive Größe darstellt und deshalb mit psychophysischen Methoden gemessen werden muß¹⁾. Auch darf die Tonhöhe nicht ohne weiteres mit einer Tonfrequenz identifiziert werden. Im Gegenteil: Das Ziel der Untersuchung über die Tonhöhenempfindung ist es gerade zu finden, in welcher Weise die wahrgenommene Tonhöhe von den verschiedenen Parametern abhängt, mit denen ein bestimmter Ton beschrieben werden kann. Eine diesbezügliche Untersuchung ist möglich dank der Tatsache, daß die Tonhöhe eine „eindimensionale“ Größe darstellt: von zwei Tönen kann immer festgestellt werden, ob sie gleich hoch empfunden werden, und falls dies nicht der Fall ist, welcher der höhere ist. Mit Hilfe dieser Eigenschaft kann jemand die Tonhöhe, die er einem gewissen Klang zuerkennt, immer mit derjenigen eines Bezugstones vergleichen, und auf diese Weise kann man der Schwie-

rigkeit aus dem Wege gehen, daß die Tonhöhe, wie jede Empfindung, einer direkten messenden Beobachtung nicht zugänglich ist.

Aus den folgenden Ausführungen wird sich ergeben, daß man einen Unterschied machen muß zwischen Klängen, die aus einer einzigen Sinusschwingung bestehen, und Klängen, die ein zusammengesetztes Spektrum aufweisen. Bei den erstgenannten ist es zulässig, in der gebräuchlichen Weise die Frequenz als Maß für die Tonhöhe zu betrachten. Dies ermöglicht es beispielsweise, eine musikalische Tonfolge als eine Zahlenfolge niederzuschreiben. Ein zusammengesetzter Klang jedoch kann eine derartige Tonhöhe haben, daß die Frequenz eines gleich hoch klingenden Sinustones nicht einer der in dem Spektrum vorkommenden Frequenzen entspricht, und auch nicht daraus einfach abgeleitet werden kann. Beispiele hierfür sind der Schlagton von Kirchenglocken und auch die menschliche Stimme²⁾.

Im nachstehenden wollen wir zeigen, daß die letztgenannte Erscheinung sich nicht erklären läßt, wenn man von der bis vor kurzem allgemein akzeptierten sog. Ortstheorie ausgeht, und sodann angeben, in welcher Richtung die Lösung wahrscheinlich gesucht werden muß. Zunächst wollen wir anhand von *Abb. 1* diese Theorie kurz rekapitulieren³⁾.

²⁾ Siehe z.B. A. Cohen, Phonetische Forschung, Philips' techn. Rdsch. 24, 364-370, 1962/63 (Nr. 11/12).

³⁾ Man findet die Ortstheorie ausführlich behandelt u.a. in H. Fletcher, Speech and hearing in communication, Van Nostrand, New York 1953.

*) Institut für Perzeptionsforschung, Eindhoven.

¹⁾ Siehe den ersten Artikel in diesem Heft: J. F. Schouten, Philips' techn. Rdsch. 24, 353-357, 1962/63 (Nr. 11/12).

Abb. 1a ist eine schematische Darstellung des menschlichen Gehörorgans. Links ist der äußere Gehörgang angedeutet, durch das Trommelfell vom Mittelohr getrennt, und rechts das Innenohr. Letzteres besteht aus einem länglichen Raum (35 mm lang), der mit einer Flüssigkeit gefüllt ist und in der Längsrichtung in zwei Hälften geteilt wird durch die

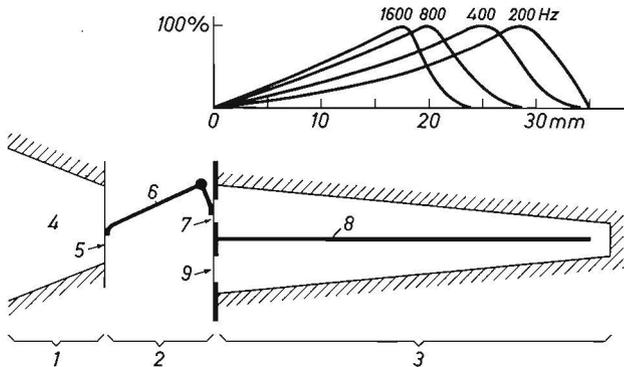


Abb. 1. Schematische Darstellung des menschlichen Ohres; von links nach rechts: äußerer Gehörgang (1), Mittelohr (2) und Innenohr (3). Akustische Schwingungen der Luft erreichen über den Gehörgang 4 das Trommelfell 5. Dessen Bewegung wird über die Ohrknöchelchen 6 (schematisch als Winkelhebel mit einem langen und einem kurzen Hebelarm gezeichnet) auf das ovale Fenster 7 übertragen, das das Innenohr abschließt. Durch die Flüssigkeit, mit der das Innenohr gefüllt ist, wandert demzufolge eine Druckwelle, die die Schneckentrennwand 8 in Bewegung versetzt und auf diese Weise die in ihr liegenden Enden des Gehörnervs reizt. Die zu Sinustönen verschiedener Frequenzen gehörende Amplitudenverteilung dieser Bewegung findet man oben skizziert. Das Maximum einer solchen Kurve liegt um so näher beim Ende der Trennwand, je niedriger die Frequenz ist. Die zur Bewegung von 7 erforderliche Flüssigkeitsbewegung wird ermöglicht durch das Vorhandensein des (elastischen) runden Fensters 9. In Wirklichkeit ist der Raum 3 zu einer Spirale aufgewickelt (die sog. Schnecke).

Schneckentrennwand, die durch Schallwellen in Schwingung gerät. Diese Trennwand besteht u.a. aus der Basilarmembran, in der die sog. Cortischen Haarzellen liegen. Diese bilden die Ausläufer des Hörnervs und stellen die Auslenkung des betreffenden Punktes der Basilarmembran fest. In Wirklichkeit ist der gezeichnete Teil des Innenohrs aufgewickelt, weshalb er als Schnecke bezeichnet wird⁴⁾. Gemäß der Ortstheorie nun entspricht kurz zusammengefaßt erstens jeder Frequenz ein bestimmter Ort auf der Zwischenwand der Schnecke und zweitens jedem Ort eine bestimmte Tonhöhe.

Für Sinustöne kann man den Vorgang schematisch tatsächlich in dieser Weise zusammenfassen: dank den Untersuchungen von von Békésy⁵⁾ wissen wir jetzt, daß die fortschreitenden Wellen, die über das ovale Fenster in der Flüssigkeit des Innenohrs

hervorgerufen werden, wenn eine Schallschwingung das Trommelfell trifft, um so schneller „gedämpft“ werden, je höher die Frequenz ist; nur Wellen mit ziemlich niedriger Frequenz „laufen weiter“ bis ans Ende des Innenohrs. Man findet dies in Abb. 1b veranschaulicht, wo für einige Frequenzen der Verlauf der Amplitude in Abhängigkeit von dem Ort auf der Trennwand (schematisch) aufgetragen ist. Der soeben erwähnte, der Frequenz entsprechende Ort auf der Schneckentrennwand kann mit der Abszisse des Maximums der betreffenden Kurve identifiziert werden, d.h. mit dem Punkt, in dem die Trennwand am stärksten erregt wird. Man darf das Innenohr also tatsächlich als einen Frequenzindikator auffassen.

Es zeigt sich, daß man es innerhalb gewisser Grenzen sogar als einen Frequenzanalysator auffassen darf: In beschränktem Maße ist das Ohr imstande, in einem beliebigen Klang die Sinustöne, aus denen er zusammengesetzt ist, gesondert zu identifizieren. Man kann dies beispielsweise bei folgendem Experiment feststellen. Hört man den zusammengesetzten Klang, der durch einen periodischen Impuls hervorgerufen wird, so nimmt man diesen Klang

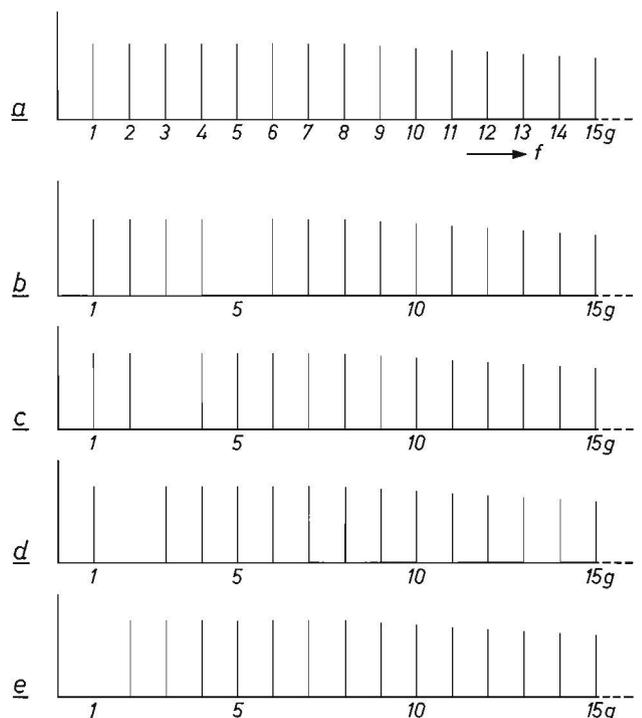


Abb. 2. a) Spektrum eines Impulstones, d.h. der Ton, der durch einen periodischen Impuls von sehr kurzer Dauer hervorgebracht wird. Aufgetragen ist die Intensität der Spektrallinien gegen ihre Frequenz (f), gemessen als Vielfaches der Grundfrequenz g . Das Spektrum besteht aus einer sehr großen Anzahl ungefähr gleich starker Harmonischer.

b), c), d) und e) Dasselbe nach Entfernung der 5., 3., 2. bzw. 1. Harmonischen. In allen Fällen entspricht die Tonhöhe derjenigen des Grundtons; die Töne unterscheiden sich lediglich im Timbre.

⁴⁾ Eine anatomische Beschreibung gibt z.B. E. G. Wever und M. Lawrence, *Physiological acoustics*, Princeton University Press, Princeton 1954.

⁵⁾ Sieh G. von Békésy, *Experiments in hearing*, McGraw-Hill, New York 1960.

beim ersten Hören als einen einzigen Ton mit einem scharfen Timbre wahr. Wird eine der niedrigen Harmonischen aus dem Spektrum eines solchen Impulstones weggenommen (Abb. 2), so wird das Timbre noch etwas schärfer. Wird sodann diese Harmonische wieder hinzugefügt, so hört man sie, da nun die Aufmerksamkeit auf ihr Vorhandensein gelenkt worden ist, danach als gesonderten Ton.

Aus Experimenten des folgenden Typs geht hervor, daß für zusammengesetzte Klänge die zweite Behauptung der Ortstheorie — „jedem Punkt auf der Schneckentrennwand entspricht eine gewisse Tonhöhe“ — nicht stimmt. Man bildet beispielsweise eine Melodie (siehe Abb. 3a) aus Tönen, die in derselben Weise erzeugt werden wie beim vorigen

(Abb. 3b), so nimmt der Hörer die Änderung des Frequenzbereiches nur als eine Änderung des Timbres wahr, nicht als eine Änderung der Tonhöhe. Die Tonhöhen entsprechen immer denjenigen der aufgezzeichneten Noten.

III

In allen Fällen wird also eine Tonhöhe wahrgenommen, die gleich derjenigen eines Sinustones mit einer Frequenz ist, die weit unterhalb der unteren Grenze des durchgelassenen Bereiches liegt, und die also überhaupt nicht dem erregten Ort der Schneckenzwischenwand entspricht.

Einen zusammengesetzten Klang, der dem Ohr als ein einziger Ton erscheint, dessen Höhe nicht einer der vorhandenen Frequenzen entspricht, bezeichnen wir als Residuum⁶⁾. Diese Bezeichnung ist dadurch entstanden, daß die beschriebene merkwürdige Erscheinung hinsichtlich der Tonhöhe zum ersten Mal an einem Impulston (siehe oben) wahrgenommen wurde, dem alle einzeln wahrnehmbaren (niedrigen) Harmonischen entzogen waren; die restierende Gruppe von Harmonischen durfte dann mit Recht als „Residuum“ bezeichnet werden. Später hat man gefunden, daß eine Gruppe von nur drei aufeinanderfolgenden Harmonischen, z.B. die achte, neunte und zehnte, den Residuum-Effekt bereits aufweist. Wir weisen noch darauf hin, daß es andererseits nicht notwendig ist, daß alle einzeln hörbaren Harmonischen eliminiert worden sind.

Nähere Betrachtung des Residuum-Effektes

Das Bestehen des Residuum-Effektes als selbstständiges Phänomen wurde lange Zeit in Zweifel gezogen. Man dachte, daß ein Ton der wahrgenommenen Tonhöhe doch reell im Ohr vorhanden wäre; dieser Ton würde nichts anderes sein als ein durch nichtlineare Verzerrung im Ohr entstandener Differenzton. Wäre diese Erklärung richtig, dann könnte die Ortstheorie in vollem Umfang ihre Gültigkeit behalten.

Bei nichtlinearer Verzerrung läßt sich der Zusammenhang zwischen dem Eingangssignal $V_i(t)$ und dem Ausgangssignal $V_u(t)$ allgemein formulieren als:

$$V_u(t) = C [V_i(t) + \delta_1 V_i^2(t) + \dots], \dots \dots (1)$$

worin C, δ_1, \dots Konstanten sind. Hat man ein Eingangssignal, das aus zwei gleich starken Sinusschwingungen besteht,

$$V_i(t) = A(\sin \omega_1 t + \sin \omega_2 t), \dots \dots (2)$$

so findet man nach (1) für das Ausgangssignal (wir setzen $C=1$):

$$V_u(t) = V_i(t) + \delta_1 A^2 [1 - \frac{1}{2} \cos 2\omega_1 t - \frac{1}{2} \cos 2\omega_2 t + \cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t] + \dots$$

Das ursprüngliche Signal wird also durch die nichtlineare Verzerrung u.a. „verunreinigt“ mit einer Schwingung mit der Frequenz $(\omega_1 + \omega_2)$ und mit einer Schwingung mit der Frequenz $(\omega_1 - \omega_2)$.

⁶⁾ Siehe J. F. Schouten, Die Tonhöhenempfindung, Philips' techn. Rdsch. 5, 294-302, 1940.

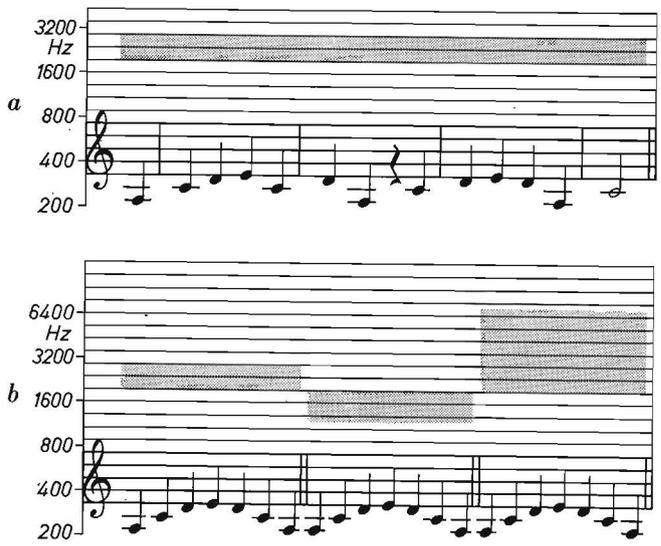


Abb. 3. a) Bildet man eine Melodie aus Impulstönen (Wiederholungsfrequenz 200-300 Hz), von denen nur die zwischen 2000 und 3000 Hz liegenden Harmonischen durchgelassen werden, so entspricht dennoch die Tonhöhe derjenigen der notierten Noten. Das Gebiet der durchgelassenen Harmonischen ist der schattierte Streifen in dem zu dem Notensystem gehörenden Hilfsliniensystem. (Ein Notensystem, einschließlich der Hilfslinien, kann aufgefaßt werden als ein rechteckiges Koordinatensystem mit dem Logarithmus der der Tonhöhe entsprechenden Frequenz als Ordinate.)
 b) Bei Verschiebung des durchgelassenen Frequenzgebietes (siehe die verschiedenen schattierten Streifen) ändert sich, wie sich im Experiment ergibt, nur das Timbre der Töne, nicht die Höhe.

Experiment. Die Tonhöhe wird variiert, indem man die Wiederholungsfrequenz der Impulse ändert. Mittels eines Filters jedoch werden von dem Klangspektrum dieser Töne nur die Komponenten durchgelassen, deren Frequenzen zwischen etwa 2000 und 3000 Hz liegen. Trotzdem entsprechen die wahrgenommenen Tonhöhen denjenigen der in Abb. 3a aufgezeichneten Noten, d.h. der Höhe von Sinustönen von etwa 200 bis 300 Hz.

Verschiebt man bei einem derartigen Experiment die Lage des durchgelassenen Frequenzbereiches

In der Akustik spricht man hier von dem Summenton bzw. dem Differenzton. Das Bestehen dieser Töne wurde von Musikern bereits in der Mitte des 18. Jahrhunderts beschrieben (G. A. Sorge 1744; G. Tartini 1754).

Man kann gegen die Differenztonhypothese eine Anzahl voneinander sehr abweichender Argumente anführen:

- 1) Der Residuum-Effekt tritt auch bei Klängen auf, die so schwach sind, daß nichtlineare Verzerrung als ausgeschlossen angesehen werden muß.
- 2) Der Residuumton gibt mit einem Sinuston von ungefähr gleicher Tonhöhe keine Schwebungserscheinungen.
- 3) Man kann den Residuumton, im Gegensatz zu einem gleich hohen Sinuston, nicht durch Rauschen überdecken (sog. Maskieren), dessen Frequenzspektrum sich um die Frequenz f_p erstreckt, die der wahrgenommenen Tonhöhe entspricht; er kann dagegen durch Rauschen maskiert werden, dessen Spektrum die Komponenten des Residuums umfaßt ⁷⁾; siehe Abb. 4.

IV

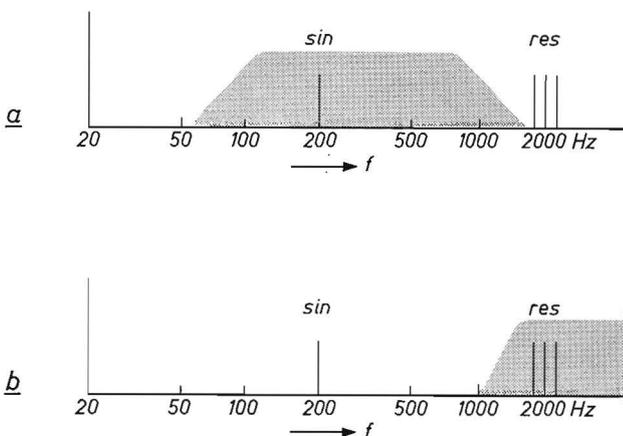


Abb. 4. Maskieren mittels Rauschen zeigt, daß der Residuumton nicht als Differenzton im Ohr entsteht. Wählt man das Rauschspektrum ungefähr so, wie es schattiert in (a) dargestellt ist, so verschwindet ein Sinuston von 200 Hz (*sin*), aber nicht der gleich hoch klingende Residuumton, der von dem Komplex *res* (Frequenzen 1800, 2000 und 2200 Hz) herrührt. Wählt man das Rauschspektrum wie in (b), so verschwindet gerade der Residuumton und bleibt der Sinuston hörbar.

- 4) Ein unumstößliches Argument gegen die Differenztonhypothese bildet die Erscheinung, daß die Residuumtonhöhe nicht immer der Differenzfrequenz entspricht. Dies kann man u.a. bei einem Experiment folgender Art feststellen.

Man geht z.B. aus von einem Residuum aus drei Komponenten mit den Frequenzen 1800, 2000 und 2200 Hz, d.h. der 9., 10. und 11. Harmonischen von 200 Hz. Die Tonhöhe entspricht

200 Hz ($f_p = 200$ Hz). Sodann erhöht man alle Frequenzen um einen gleichen Betrag, z.B. um 10 Hz, und wiederholt dies einige Male, z.B. bis zu ungefähr 1850, 2050 und 2250 Hz. Obwohl der Frequenzunterschied der Komponenten bei dieser Operation konstant bleibt (200 Hz), hört man die Residuumtonhöhe ansteigen. Haben die Frequenzen die zuletzt genannten Werte erreicht, so ist f_p auf beinahe 205 Hz angestiegen (Abb. 5). Läßt man in derselben Weise die Frequenzen abnehmen, so sinkt die Residuumtonhöhe.

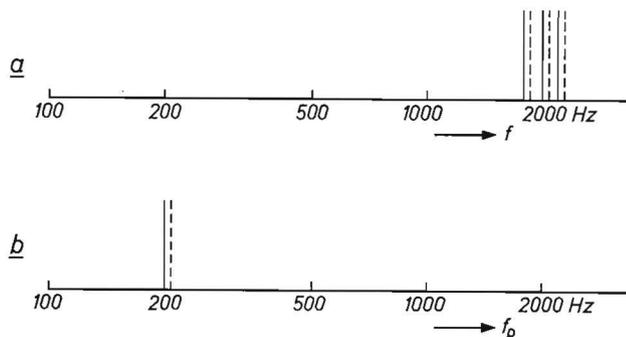


Abb. 5. Bei äquidistanter Verschiebung eines Komplexes (a) bleibt die Residuumhöhe (dies ist die in (b) aufgetragene Frequenz f_p eines gleich hoch klingenden Sinustones) nicht konstant, sondern ändert sich proportional zur mittleren Frequenz f der Komponente des Komplexes.

Die Tonhöhe eines Komplexes aus drei Komponenten

Der einfachste Komplex, an dem der Residuum-Effekt wahrgenommen werden kann, besteht aus drei Sinusschwingungen mit gleichem gegenseitigen Frequenzunterschied. Abgesehen von der Lautheit und dem gegenseitigen Lautstärkeverhältnis, die wir beide einen Augenblick außer Betracht lassen, läßt sich ein solcher Komplex mit nur zwei Parametern beschreiben, der Frequenz f der mittleren Komponente und dem Frequenzunterschied g . Versuche von der soeben beschriebenen Art haben gelehrt, daß eine Änderung von f bei konstant bleibendem g (äquidistante Verschiebung) eine f proportionale Änderung der Tonhöhe f_p zur Folge hat ⁸⁾. Dies bedeutet jedoch nicht, daß bei zunehmendem f die Tonhöhe fortwährend ansteigt. Geht man von einem Komplex aus, der z.B. ebenso wie in dem oben beschriebenen Experiment aus einer 9., 10. und 11. Harmonischen besteht ($f = 10g$), so springt bei $f \approx 10,5g$ die Tonhöhe herab ⁹⁾. Bei weiterem Anstieg von f steigt f_p wieder, um bei $f = 11g$ von neuem den Wert g zu erreichen. Anfangs ist $f_p =$

⁷⁾ Dies wurde zum ersten Mal nachgewiesen von J. C. R. Licklider, J. Acoust. Soc. Amer. 26, 945, 1954.

⁸⁾ J. F. Schouten, R. J. Ritsma und B. Lopes Cardozo, Pitch of the residue, J. Acoust. Soc. Amer. 34, 1418-1424, 1962 (Nr. 9 II).

$f/10$, nach dem Sprung $f/11$ usw. Der Verlauf von f_p mit f hat also den Charakter einer Sägezahnkurve; die Tonhöhe entspricht nur dann dem Frequenzunterschied ($f_p = g$), wenn f ein ganzes Vielfaches von g ist (Abb. 6).

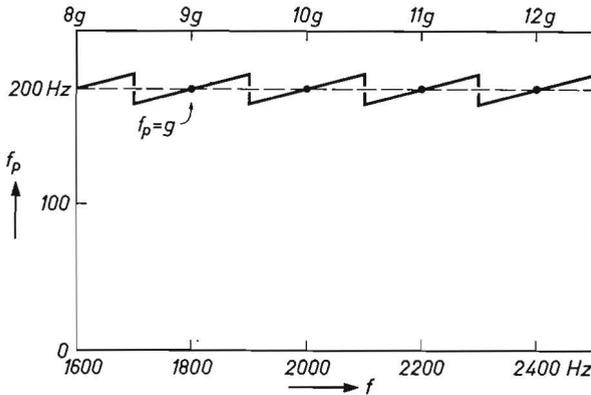


Abb. 6. Bei äquidistanter Verschiebung eines Komplexes aus drei Komponenten (Frequenz der mittleren Komponente f , Frequenzunterschied g) nimmt f_p nicht unbeschränkt weiter mit f zu. Hat f die Strecke zu einem folgenden Vielfachen von g ungefähr bis zur Hälfte zurückgelegt, so macht f_p einen Sprung, so daß beim Erreichen dieses Vielfachen f_p wieder gleich g ist.

Die Proportionalität zwischen f_p und f läßt sich in folgender Weise aus der Feinstruktur des akustischen Signals erklären, die bei unseren Versuchen die in Abb. 7 skizzierte Form aufwies. Es besteht aus einer Schwingung mit der Frequenz f (die ausgezogene Kurve), deren Amplitude mit einem Modulationsgrad von 100% mit der Frequenz g (strichpunktierte

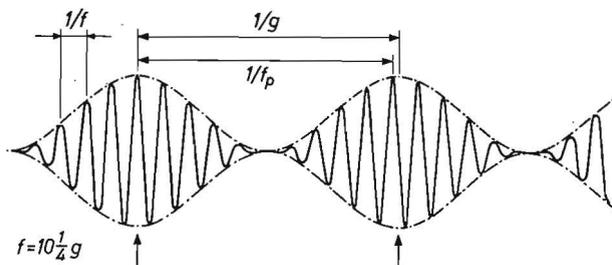


Abb. 7. Zur Erklärung der Erscheinung, daß bei äquidistanter Verschiebung eines Komplexes von drei Komponenten die Residuumentonhöhe f_p sich proportional zur zentralen Frequenz f ändert. Das Klangsignal wird durch die ausgezogene Kurve dargestellt (Sinusschwingung mit der Frequenz f , zu 100% moduliert mit der Frequenz g). Das Ohr leitet die Tonhöhe zwar aus der Periodizität eines Signals ab, jedoch in dem Sinne, daß es als solche den Abstand der Maxima betrachtet, die innerhalb jeder Periode $1/g$ den Maxima (siehe Pfeile) der strichpunktierten Kurve am nächsten liegen. Dieser Abstand ist natürlich ein Vielfaches des Abstandes $1/f$. Ist f nicht zugleich ein Vielfaches von g (anharmonischer Komplex), so ist f_p also von g verschieden.

Kurve) moduliert wird; auf diese Weise wurde das Signal übrigens auch erhalten. Offenbar entspricht nun die wahrgenommene Tonhöhe der Periodizität der Amplitudenmodulation, in dem Sinne, daß das Gehör diese aus dem Abstand der Maxima ableitet, die den Maxima der strichpunktierten Linien am nächsten liegen⁹⁾. Dieser Abstand ist natürlich dem Abstand $1/f$ proportional; als Formel geschrieben: $1/f_p = n/f$. Nimmt f so weit zu, daß der Abstand $1/g$ der Maxima der Frequenz g dadurch besser angenähert wird, daß man auf ein benachbartes Maximum von f übergeht, d.h. für n eine Zahl wählt, die um eins größer ist, so macht die Tonhöhe einen Sprung. Dies erklärt auch, daß $f_p = g$ ist, wenn f ein ganzes Vielfaches von g ist.

Das Obenstehende führt zu dem Schluß, daß das Organ, das die Tonhöhe komplexer Klänge feststellt, sich nicht in dem mechanischen Teil des Ohres befindet, sondern neuraler Art ist; es muß sich im Gehörnerv oder im Gehirn befinden. Offenbar ist es nicht so sehr ein Spektrograph als vielmehr eine Art Zeitmesser, der die Feinstruktur des Signals analysiert¹⁰⁾.

Am Schluß dieses Abschnittes sei darauf hingewiesen, daß der Residuumentonhöhe nicht bei jeder beliebigen Kombination von Frequenzen auftritt¹¹⁾. In Abb. 8 sind die Grenzen des Existenzgebietes des Residuumentonhöheffektes für einen Komplex von drei Komponenten graphisch dargestellt. Längs der Abszisse ist hier die Frequenz f aufgetragen, längs der Ordinate der Quotient n aus f und g . Die geeigneten Geraden sind Linien für konstantes g . Wir betrachten vorläufig nur die ausgezogene Kurve $M = 100\%$. Bei den Kombinationen von f - und n -Werten, deren graphische Darstellung innerhalb des von dieser Kurve umschlossenen Gebietes fällt, ist eine Residuumentonhöhe wahrnehmbar; außerhalb desselben nicht. Obwohl die Form der Kurve in ihren Einzelheiten nicht für alle Personen dieselbe ist, so gilt doch wohl allgemein, daß der höchste f -Wert bei einem n von annähernd zehn auftritt, und der höchste n -Wert bei $f = 2000$ bis 3000 Hz. Die niedrigste Frequenz g — d.i. also zugleich ungefähr das niedrigste f_p — ist ungefähr 35 Hz, die höchste ungefähr 800 Hz.

Experimentiert man mit einem hinsichtlich der Frequenzen identischen Komplex, bei dem die Komponenten mit den Frequenzen $(f-g)$ und $(f+g)$ weniger stark sind, so ist das Existenzgebiet kleiner. Die Schwingung eines solchen Komplexes sieht im Prinzip genau so aus wie die von Abb. 7,

⁹⁾ Vgl. auch E. de Boer, On the „residue“ hearing, Diss. Amsterdam 1956.

¹⁰⁾ R. J. Ritsma, A model of human pitch-extraction based on additive correlation, Proc. 4th int. Congr. Acoust. I, Kopenhagen 1962, im Druck.

¹¹⁾ R. J. Ritsma, Existence region of the tonal residue I, J. Acoust. Soc. Amer. **34**, 1224-1229, 1962 (Nr. 9 I).

mit der Abweichung, daß der Modulationsgrad M kleiner ist als 100%. Die gestrichelten Kurven in Abb. 8 geben die Grenze des Existenzgebietes für die danebengeschriebenen Werte von M an. Das Kleinerwerden des Existenzgebietes mit abnehmendem M erklärt, warum man den Residuumton in einem gewissen Augenblick verschwinden hört, wenn man von einem Komplex aus drei Komponenten die beiden äußeren an Stärke abnehmen läßt.

Die Änderung von f_p mit f bei äquidistanter Verschiebung eines Komplexes aus drei Sinusschwingungen braucht nicht immer in der in Abb. 6 skizzierten Weise zu verlaufen. Eine Versuchsperson, die sich auf die Änderung der Residuumtonhöhe konzentriert, kann den Sprung in f_p verzögern⁸⁾. In Abb. 9 sind einige Messungen hierüber graphisch dargestellt. Man sieht, daß die Frequenzbereiche, für die die verschiedenen f_p - f -Kurven gelten, einander so weit überlappen, daß z.B. bei $f = 1800$ Hz nicht weniger als vier Werte f_p zuerkannt werden können. Die Tonhöhe, die das menschliche Gehör einem bestimmten Klang zuerkennt, wird also nicht in allen Fällen eindeutig durch die physikalischen Parameter des Klanges bestimmt. Man bezeichnet dies als die Ambivalenz der Tonhöhenempfindung.

Hört man dagegen arglos einen derartigen Komplex ab, der äquidistant z.B. von $f = 10g$ nach $f = 11g$ verschoben wird, und vergleicht man die wahrgenommene Tonhöhe nicht regelmäßig mit derjenigen eines Bezugstones, so macht man den Sprung unserer Erfahrung nach unbewußt. Man meint einen Ton zu hören, der fortwährend ansteigt, stellt jedoch nach

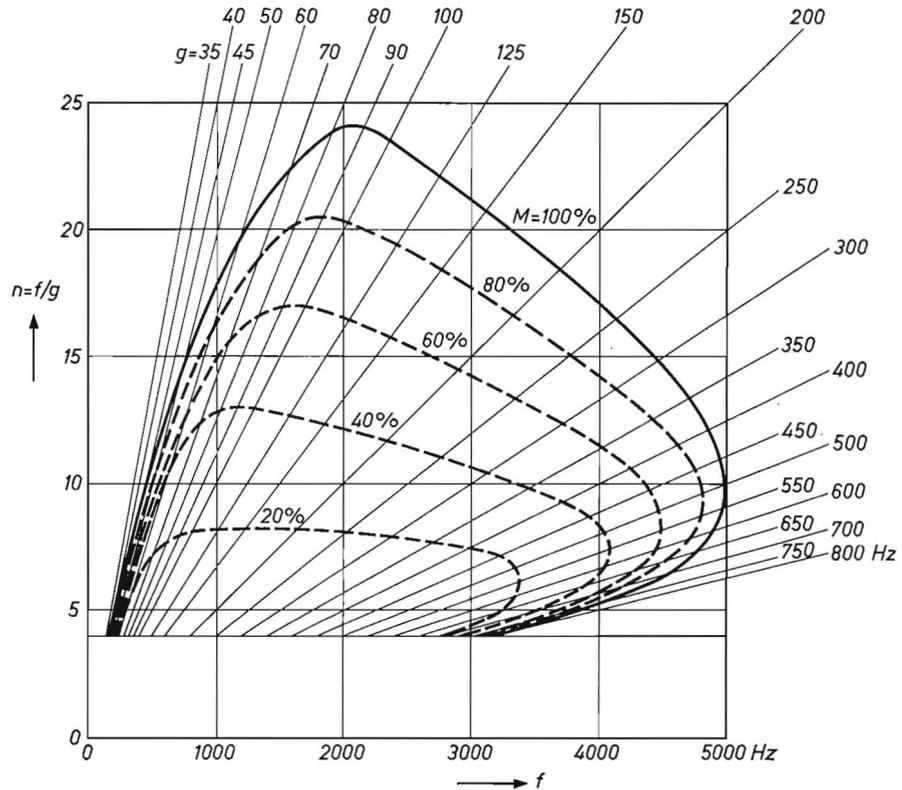


Abb. 8. Existenzgebiet des Residuumtones eines Komplexes dreier benachbarter Harmonischen (Frequenz der mittleren f , Grundfrequenz g): der Residuumton wird im Gebiet innerhalb der ausgezogenen Kurve wahrgenommen. Macht man die beiden äußeren Komponenten schwächer (experimentell: indem man die Schwingung f mit einem unter 100% liegenden Modulationsgrad moduliert), dann wird das Existenzgebiet kleiner.

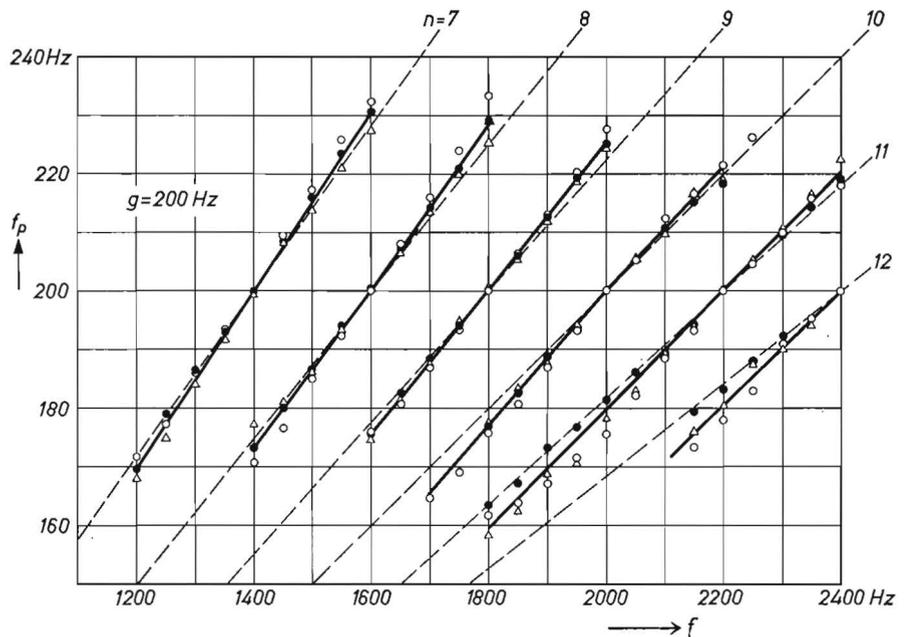


Abb. 9. Die Änderung der Tonhöhe f_p eines Komplexes aus drei Komponenten bei sich ändernder zentraler Frequenz f kann, wenn man sich darauf konzentriert, noch eine gewisse Zeit über die Frequenz hinaus wahrgenommen werden, bei der für einen nicht eingeweihten Hörer f_p einen Sprung macht (vgl. Abb. 6). Dies bedeutet, daß die Tonhöhe eines solchen Komplexes nicht eindeutig festliegt. In dem Fall, auf den sich das Diagramm bezieht, kann man bei $f = 1800$ Hz dem Komplex nicht weniger als vier Tonhöhen zuerkennen. (Nach der Hypothese über den Verlauf von f_p in Abhängigkeit von f (siehe Abb. 7) müßte die Änderung in f_p genau gleich dem n -ten Teil derjenigen in f sein und würden die Kurven mit den gestrichelten Geraden zusammenfallen. Wie man sieht, tritt eine kleine systematische Abweichung auf; eine Erklärung hierfür wurde noch nicht gefunden.)

Ablauf des Experimentes zur eigenen Überraschung fest, daß f_p gleich g ist wie zu Anfang.

Wie man aus Abb. 9 ablesen kann, ist die Änderung in f_p nicht genau gleich dem n -ten Teil der Änderung in f . Es zeigt sich, daß die Größe der Abweichung von der Lautheit abhängt. Eine Untersuchung über die Ursache dieses Effektes ist noch im Gange.

Die Tonhöhe kurzzeitiger Klänge

Bisher haben wir nur über Klänge gesprochen, die dem Ohre relativ lange dargeboten werden. Jetzt wollen wir uns etwas mit kurzzeitigen Klängen beschäftigen.

Nach den obenstehenden Darlegungen über die Residuumtonhöhe eines Komplexes aus drei Komponenten und die f proportionale Änderung derselben bei äquidistanter Verschiebung liegt die Frage nahe, wieviel Perioden $1/g$ das Gehör benötigt, um die Residuumtonhöhe festzustellen. Eine Untersuchung in unserem Institut hat gelehrt, daß dies für Residuumtonhöhen mit f_p zwischen 200 und 475 Hz immer vier sind. Bei einem g -Wert von z.B. 200 Hz ist also offenbar das Gehör imstande, in 20 ms die Tonhöhe ziemlich genau festzustellen.

Während also bei einem Residuumton die Zahl der Perioden, $1/g$, den bestimmenden Faktor für das Zuerkennen einer Tonhöhe darstellt, ist dies bei Sinustönen die zeitliche Dauer des Stoßes. Einen relativ langen Stoß beurteilt man deutlich als Ton. Wird die Dauer verkürzt, so ändert sich anfangs kaum etwas, aber bei einem bestimmten kritischen Wert beginnt der Klang seinen Charakter zu ändern: er geht allmählich von einem Ton in ein Tickgeräusch über.

Messungen über die Tonhöhenempfindung kurzzeitiger Stöße lassen sich am besten in der Weise durchführen, daß man der Versuchsperson zwei gleich lange Stöße von verschiedener Frequenz kurz nacheinander darbietet. Man beginnt mit einem sehr kleinen Frequenzunterschied und läßt diesen allmählich größer werden. Dabei untersucht man, bei welchem Frequenzunterschied die Versuchsperson die beiden „Töne“ gerade nicht mehr als gleich hoch betrachtet.

Die Ergebnisse einer derartigen in unserem Institut durchgeführten Untersuchung findet man in Abb. 10 zusammengefaßt. Es handelt sich hier um Stöße einer Sinusschwingung von 1000 Hz, wobei sowohl der Anfang als auch das Ende mit einem Nulldurchgang der Schwingung zusammenfiel¹²⁾.

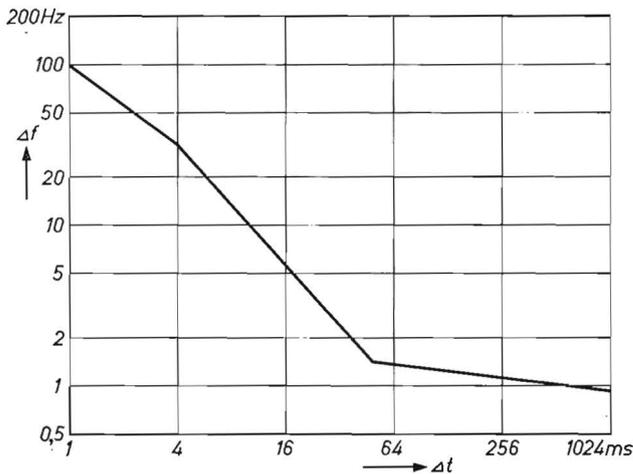


Abb. 10. Die Deutlichkeit, mit der man einen kurzen Sinusstoß als Ton vernimmt, hängt von der Dauer Δt des Stoßes ab. Läßt man Δt abnehmen, so wird der Hörer stets unsicherer hinsichtlich der dem Klang zuzuerkennenden Tonhöhe. Dargestellt ist hier das Ergebnis von Messungen, bei denen der Versuchsperson immer zwei Tonstöße (1000 Hz) von der Länge Δt dargeboten wurden, die einen kleinen Frequenzunterschied aufwiesen, der langsam gesteigert wurde. Auf der Ordinate ist der Frequenzunterschied Δf aufgetragen, bei dem die Versuchsperson die Töne gerade nicht mehr als gleich hoch beurteilt.

Die Größe Δf ist der soeben eingeführte kritische Frequenzunterschied und Δt ist die Dauer des Tonstoßes.

Wie man sieht, ist Δf nahezu konstant und sehr klein (≈ 1 Hz), solange Δt länger ist als 50 ms. Verkürzt man Δt noch weiter, dann nimmt Δf zu, aber nicht besonders stark. Bei $\Delta t = 2$ ms ist Δf noch nicht größer als 50 Hz, d.i. nur $2\frac{1}{2}\%$ von f . Das menschliche Gehör ist also bei Sinusschwingungen — Schwingungen, bei denen die Periodizität identisch ist mit der reziproken Frequenz — offenbar imstande, einen Frequenzunterschied von einigen Prozent in einigen Millisekunden festzustellen. Das Diagramm Abb. 10 für $\Delta t < 50$ ms läßt sich mit guter Näherung mit der Formel $\Delta f \Delta t = \text{konst.}$ beschreiben. Die Konstante, die von Person zu Person etwas verschieden ist, liegt in der Größenordnung von 0,1.

Zum Schluß noch einige Bemerkungen über noch unregelmäßigere Klangkomplexe als die besprochenen Gruppen von drei Schwingungen sowie über Klänge, deren Charakter sich nach dem Einsetzen schnell verändert.

Über die stationären harmonischen Komplexe mit mehr als drei Komponenten läßt sich sagen, daß ihre Eigenschaften aus den erwähnten Eigenschaften der besprochenen Komplexe von drei Komponenten abgeleitet werden können. Komplexe, deren Komponenten nicht äquidistant sind, wurden noch nicht untersucht.

¹²⁾ B. Lopes Cardozo, Frequency discrimination of the human ear, Proc. 4th int. Congr. Acoust., Vortrag H 16, Kopenhagen 1962.

Auch über Klänge, die sich nach dem Einsetzen schnell verändern, ist unser Wissen noch lückenhaft. Trotz der großen Bedeutung des Einsatzeffektes — ein Klavierton, dessen charakteristischer Einsatz abgeschnitten ist, ist als solcher kaum erkennbar¹³⁾ — ist dieser Effekt für Musikinstrumente noch nicht ausführlich untersucht worden. Wohl werden eingehende Untersuchungen an der menschlichen Stimme durchgeführt. Diese fallen jedoch in der Hauptsache in das Gebiet der Phonetik.

Beispiele für die Experimente, die in diesem Aufsatz mit römischen Ziffern (I bis VII) am Rande bezeichnet sind, wurden auf einer vom Institut für Perzeptionsforschung hergestellten Schallplatte fest-

¹³⁾ Vgl. z.B. die Klangbeispiele bei dem Artikel: H. Badings und J. W. de Bruyn, Elektronische Musik, Philips' techn. Rdsch. 19, 208-219, 1957/58.

gelegt¹⁴⁾. Mit Hilfe dieser Platte kann der Leser die besprochenen Eigenschaften der Tonhöhenempfindung selbst wahrnehmen.

¹⁴⁾ Diese Schallplatte (die außer den Klangbeispielen Erläuterungen in englischer Sprache enthält) wird kostenlos zugesandt nach Einsendung des Bestellscheins, der sich auf dem lose beigelegten Blatt mit Kurzzusammenfassungen in diesem Heft befindet.

Zusammenfassung. Der bekannte Zusammenhang zwischen Frequenz und Tonhöhe von Sinustönen gilt in vielen Fällen nicht für komplexe Klänge: eine Gruppe von drei (oder mehreren) benachbarten Harmonischen (Frequenzen $f-g$, f und $f+g$) hat in einem großen Bereich von f , g -Kombinationen dieselbe Tonhöhe wie der Grundton (Residuum-Effekt). Aus Maskierungsversuchen mit Rauschen und aus der Tatsache, daß sich die Tonhöhe ändert, wenn man die drei Komponenten bei konstantem g in der Frequenz verschiebt, geht hervor, daß der Residuerton nicht im mechanischen Teil des Ohres entsteht, sondern neuralen Ursprungs ist. Aus den letzteren Versuchen ergibt sich ferner, daß die Tonhöhe nicht genau der Periodizität der Umhüllenden des Signals (Frequenz g) entspricht, sondern aus seiner Feinstruktur abgeleitet wird. Bei einem kurzen Tonstoß wird die Residuertonhöhe bereits gehört, wenn der Stoß vier Perioden $1/g$ lang ist. Bei einem Sinusstoß ist die Dauer die bestimmende Größe.

PHONETISCHE FORSCHUNG

von A. COHEN *).

534.4

Die Phonetik beschäftigt sich mit der Untersuchung von Sprachlauten; seit langem gehört die Hervorbringung der Sprachlaute dazu, und seit kurzem auch ihre Perzeption. Man kann sagen: die Phonetik untersucht die Kommunikation von Mensch zu Mensch mittels des Sprechens, wobei dann die Sprache, welche gesprochen wird, nicht als Objekt der Untersuchung, sondern als etwas Gegebenes betrachtet wird¹⁾. Man kann die Phonetik nicht ausschließlich als ein Teilgebiet der Sprachwissenschaft, der Biologie oder der Physik auffassen; alle diese Wissenschaften liefern ihren besonderen Beitrag. Verfolgen wir nämlich einmal den ganzen Übertragungsweg, den eine gesprochene Nachricht durchläuft, so begegnen wir der Reihe nach 1. dem menschlichen Stimmorgan (Untersuchungsobjekt der Anatomie und der Physiologie), 2. der Art und Weise, wie dieses Organ die Sprach-

laute hervorbringt (Artikulationslehre), 3. den hervorgebrachten Lauten, als Schallschwingung der Luft betrachtet (Akustik), 4. dem Ohr und den zu dem Hörorgan gehörenden neuralen Elementen (Anatomie und Physiologie). Daß dieser letzte Teil des Übertragungsweges — die Perzeption — jetzt in die phonetische Forschung mit einbezogen wird, ist gerade der bewußten Anwendung der Konzeptionen der noch jungen Informationstheorie zu verdanken.

Im Mittelpunkt der phonetischen Forschung, die seit 1959 im Institut für Perzeptionsforschung durchgeführt wird, steht die Frage: *Welche physikalischen Merkmale der Sprachlaute sind wesentlich zum Erkennen des in ihnen enthaltenen Sprachinhaltes?* Ebenso wie die Zeichen vieler anderer Kodierungssysteme, wie z.B. unsere Buchstaben, enthalten die Sprachlaute nämlich eine ziemlich große Menge an Information, die zum Erkennen nicht absolut notwendig ist. Dies geht bereits aus der auffälligen Tatsache hervor, daß man mit dem Gehör nicht nur den Inhalt des Gesprochenen aufnehmen kann, sondern auch Schlüsse hinsichtlich der Identität des Sprechers und eventuell seines Gemütszustandes zu ziehen vermag. Es ist wohl klar, daß man das genannte Problem, welche Merkmale wesentlich zum Erkennen des Sprachinhaltes sind, nicht auf dem Wege der akustischen Analyse allein lösen

*) Institut für Perzeptionsforschung, Eindhoven.

¹⁾ Man findet die Phonetik in verschiedenen Hand- und Lehrbüchern behandelt, wie z.B. L. Kaiser, Manual of phonetics, North-Holland Publ. Co. Amsterdam 1957. Der Wissenschaftszweig, der sich mit der Frage beschäftigt, wie man sich eine bestimmte Sprache (oder Dialekt) aus einer bestimmten Anzahl wortunterscheidender Klangformen (den sog. Phonemen) aufgebaut denken kann, bildet einen Teilbereich der Sprachwissenschaft und trägt den Namen Phonologie. Da nur die Unterscheidbarkeit der Worte als Maßstab gilt, macht man im phonologischen System der holländischen Sprache z.B. keinen Unterschied zwischen dem *ee* von *mees* und demjenigen von *meer*. Für die Phonetik sind dies verschiedene Laute.