

Zeitschrift:	Technische Mitteilungen / Schweizerische Post-, Telefon- und Telegrafenbetriebe = Bulletin technique / Entreprise des postes, téléphones et télégraphes suisses = Bollettino tecnico / Azienda delle poste, dei telefoni e dei telegrafi svizzeri
Herausgeber:	Schweizerische Post-, Telefon- und Telegrafenbetriebe
Band:	51 (1973)
Heft:	6
Artikel:	Subjektive Versuche zur Ermittlung der Sprachübertragungsqualität
Autor:	Kündig, Albert
DOI:	https://doi.org/10.5169/seals-875294

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 16.01.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Subjektive Versuche zur Ermittlung der Sprachübertragungsqualität

Albert KÜNDIG, Bern

621.391.883:621.395

Zusammenfassung. Nach einer Übersicht der benutzerorientierten Methoden zur subjektiven Ermittlung der Sprachübertragungsqualität wird auf zwei spezielle Verfahren näher eingegangen, nämlich jenes der paarweisen Vergleiche sowie die Reimtests. Es wird eine Methode zur Erfassung der Zuverlässigkeit von Testpersonen vorgeschlagen und über praktische Resultate berichtet. Für die Reimtests werden Wortkataloge in deutscher Sprache angegeben.

Essais subjectifs destinés à déterminer la qualité de transmission de la parole

Résumé. Après avoir brièvement rappelé les méthodes faisant appel aux usagers pour déterminer subjectivement la qualité de la transmission de la parole, l'auteur explique en détail deux procédés spéciaux, soit celui des comparaisons par couples et les tests sous forme de rimes. Il propose une méthode d'enregistrement de la fiabilité des personnes choisies pour les tests et fait état des résultats pratiques. Pour les tests sous forme de rimes, il indique une liste de mots en langue allemande.

Esperimenti per la valutazione soggettiva della qualità di trasmissione dei segnali vocali

Riassunto. A un breve sunto dei metodi tradizionali per determinare la qualità soggettiva di trasmissione dei segnali vocali, che si basano su criteri orientativi dell'utente, segue una relazione più dettagliata su due metodi speciali, cioè sull'analisi comparativa a coppie di fonemi e sull'analisi basata su parole rimate. Quindi si spiega un metodo per la determinazione dell'affidabilità delle persone che eseguiscono i «test» e si commentano i risultati pratici ottenuti. Per le parole rimate troviamo inoltre una tabella di termini in lingua tedesca.

1. Einleitung

Mit der Entwicklung neuer Sprachübertragungsverfahren hat sich in den letzten Jahren ein vermehrtes Bedürfnis nach geeigneten Methoden zur Beurteilung der Übertragungsqualität ergeben. Da letztlich immer der Mensch Empfänger der Sprachsignale ist, können nur subjektive Methoden über die Auswirkung von Störungen und Verzerrungen zuverlässig Auskunft geben. Leider ist über die Physiologie des Hörorganges immer noch viel zu wenig bekannt, als dass, mit Hilfe eines entsprechenden Modells für die Wahrnehmung der Laute, Charakteristiken, wie Verständlichkeit, Wiedergabegüte oder gar Sprechererkennbarkeit, theoretisch hergeleitet werden könnten. Es bedeutet dies allerdings nicht, dass für eine ganz bestimmte Klasse von Störungen – nach Durchführung entsprechender subjektiver Versuche – keine spezifische Theorie entwickelt werden könnte. Beispielsweise lässt sich für den Fall von verschiedenen Übertragungsdämpfungen und weissem Rauschen als Störung mit den in [1] und [2] dargestellten Methoden ein Qualitätswert bestimmen, der mit subjektiven Verständlichkeitsmessungen recht gut übereinstimmt. Es wurde aber schon bald erkannt, dass für andere Arten von Störungen oder Verzerrungen die rechnerischen Methoden unzuverlässig sind beziehungsweise nur mühsam mit Hilfe von Korrekturfaktoren angepasst werden können.

Digitale Modulationsmethoden sowie Verfahren zur extremen Bandkompression (Vocoder [3] und Voraussagecodierung [4]) gewinnen beispielsweise an Bedeutung. In den meisten Fällen lassen sich die dabei auftretenden Störungen oder Verzerrungen kaum mit dem Gewohnten vergleichen, so dass – mindestens am Anfang einer Begutachtung – eine subjektive Beurteilung kaum zu umgehen ist.

Neben den bereits erwähnten theoretischen Verfahren zur Ermittlung der Übertragungsgüte, die immerhin gewisse Erkenntnisse über den Hörvorgang mitbenutzen, wird auch immer wieder versucht, rein aufgrund von Stör- und Verzerrungssignalleistungen auf die Qualität zu schliessen.

Diese Betrachtungsweise ist aber vielfach im voraus zum Scheitern verurteilt, da sie gewisse Tatsachen sowohl bei der Sprachsignalerzeugung wie bei der Wahrnehmung von Lauten missachtet:

- das Störempfinden des Gehörs ist nicht notwendigerweise der Störleistung proportional; eine Addition von Störleistungen verschiedenartigen Ursprungs ist meistens unzulässig. In extremen Fällen können Verzerrungen die Qualität sogar scheinbar erhöhen.
- für die Berechnung der Störleistungen müssen Annahmen über das Signal getroffen werden. Meistens handelt es sich dabei um eine Langzeitstatistik, beispielsweise das Leistungs-Spektrum, die als Mittelwert völlig vom Charakter des Sprachsignals während eines einzelnen Lautes abweichen kann.

Die Notwendigkeit, bei neuen Übertragungsverfahren eine subjektive Beurteilung durchzuführen, sollte mit dem Vorstehenden genügend begründet sein. Die subjektiven Methoden werden heute, einem Vorschlag von Hecker und Guttman entsprechend [5], meistens in zwei Klassen geteilt:

- *Benutzerorientierte Methoden* sollen eine «Über-alles»-Beurteilung eines Systems gestatten, ohne auf die psychologischen und physiologischen Aspekte einzugehen.
- *Analytische Methoden*. Von entsprechenden Versuchen erwartet man Erkenntnisse über die psychologischen Komponenten der «Sprachqualität» und hofft, diese in Beziehung zu akustischen oder elektrischen Kenngrössen setzen zu können [6, 7, 8].

Wir wollen uns im folgenden der ersten Gruppe zuwenden, die für Telephonieübertragung von unmittelbarerem praktischem Interesse ist. Nach einer Übersicht über verschiedene Verfahren soll kurz auf zwei Methoden eingegangen werden, die im Zusammenhang mit PCM-Übertragungsversuchen verwendet wurden.

2. Benutzerorientierte Methoden: eine Übersicht

Die Qualität der Sprachübertragung kann offensichtlich nicht auf einfache Art umschrieben werden, und es ist unmöglich, sie etwa durch einen einzigen Zahlenwert zu charakterisieren. Vielmehr weist die Sprachübertragungsqualität mehrere verschiedene Komponenten auf, die zudem gegenseitig abhängig sein können. Beispiele von Charakteristiken sind:

- (1) Verständlichkeit,
- (2) Lästigkeit von Störungen,
- (3) Notwendige Anstrengung, um ein Gespräch zu führen,
- (4) Natürlichkeit,
- (5) «timbre»,
- (6) Sprechererkennbarkeit.

Welche dieser Komponenten wichtig ist, wird in erster Linie durch den Benutzer – oder besser vielleicht «Endverbraucher» – eines Übertragungssystems bestimmt und in zweiter Linie durch die Art der Verzerrungen und Störungen. Zum Beispiel könnte man sich in einem militärischen Übertragungssystem von bescheidener Qualität vor allem auf die Punkte (1) und (3) konzentrieren, während bei einer Konferenzanlage (3), (4) und (6) im Vordergrund stehen mögen. Da die neuen heute für öffentliche Telephonnetze verwendeten Übertragungsmethoden abschnittsweise eine recht gute Qualität aufweisen, interessieren für diese mehr und mehr Verfahren, mit denen die Natürlichkeit der Wiedergabe erfasst werden kann. Auch ergibt die Messung der Verständlichkeit mit herkömmlichen Methoden (sinnlose Silben, sogenannte Logatome) so hohe Werte, dass zu verfeinerten und empfindlicheren Verständlichkeitstests gegriffen werden muss.

So gross die Vielfalt an Übertragungsverfahren und deren Charakteristiken ist, fast so zahlreich sind auch die subjektiven Testmethoden, die mit mehr oder weniger Erfolg bisher eingesetzt worden sind. *Figur 1* zeigt dazu eine systematische Übersicht, wobei mit den Zahlen auf die entsprechenden Literaturstellen hingewiesen wird. Selbstverständlich konnten nicht sämtliche bekannten Methoden berücksichtigt werden; insbesondere fehlen auch Versuche spezieller Art, beispielsweise solche zur Erfassung der Verständlichkeitsschwelle von Nebensprechen. Im folgenden soll ein kurzer Abriss der verschiedenen Methoden von *Figur 1* gegeben werden.

2.1 Paarweise Vergleiche (Präferenztests)

Bei diesen Methoden hören die Versuchspersonen Paare A/B von «Sprachproben», und zwar meistens zeitlich nacheinander in der Form A-B-A-B. Sie müssen dann angeben, welche der beiden Versionen – A oder B – sie bevorzugen würden. Der prozentmässige Anteil der Bevorzugung wird als Mass für die Qualität ausgewertet. Je nach Art der Ver-

suchsanlage und Auswertung haben sich verschiedene Varianten der Methode entwickelt, wovon zwei vom *Institute of Electrical and Electronics Engineers* (IEEE) empfohlen worden sind [22] und nachstehend beschrieben werden, während auf weitere Varianten im Abschnitt 3 eingegangen wird.

2.1.1 Isopräferenzmethode

Bei der vom IEEE empfohlenen Methode handelt es sich um eine vereinfachte Form [10], wobei im Gegensatz zu [9] alle Proben mit der gleichen Lautstärke – die zuvor mit Versuchen (ebenfalls paarweise Vergleiche) als am angenehmsten bestimmt wurde – abgespielt werden. Das zu beurteilende System wird mit einem einzigen Referenzsignal verglichen, aus dem man durch Hinzufügen eines Rauschsignals verschiedener Pegel eine Reihe von «Bezugsproben» herstellt. Die Qualität des geprüften Systems wird dann als das Signal/Rausch-Verhältnis angegeben, für das eine bestimmte Bezugsprobe gleich häufig wie das Testsignal bevorzugt wird.

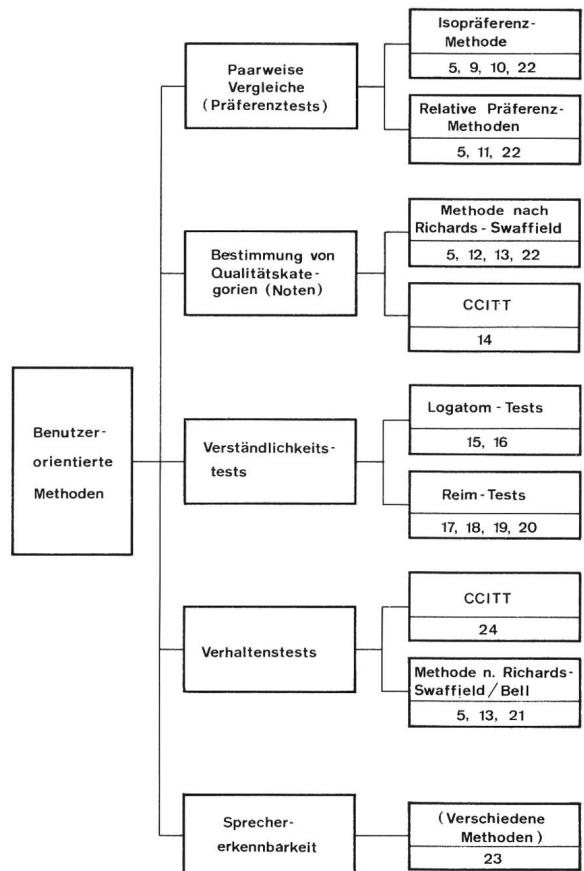


Fig. 1

Benutzerorientierte Methoden zur subjektiven Beurteilung der Sprachübertragungsqualität

Bei vielen modernen Sprachübertragungs- und Verarbeitungsverfahren treten Störungen nur zusammen mit dem Nutzsignal auf (spezielle Arten von Verzerrungen, wie Quantisierungsrauschen). In diesen Fällen ist es angezeigt, bei den Bezugssignalen nicht additives, sondern multiplikatives Rauschen zu verwenden [22].

2.1.2 Relative Präferenzmethode

Es ist möglich, dass bei der Isopräferenzmethode die Versuchspersonen von ihrer eigentlichen Aufgabe abgelenkt werden und sich vor allem auf die Unterschiede im Signal/Rausch-Verhältnis des Bezugssignals konzentrieren. Daher wurde in [11] vorgeschlagen, eine ganze Reihe von Bezugssignalen mit verschiedenartigen Störungen zu verwenden, zum Beispiel:

- A Bandbreite 50...10 000 Hz, S/N = 45 dB
- B Bandbreite 800...3000 Hz, Flanken 24 dB/Oktave
- C Bandbreite ...3000 Hz, Flanken 24 dB/Oktave, additives Rauschen (...500 Hz; 9 dB/Oktave), Spitzenwert S/N = 10 dB
- D Mehrfaches Echo (1. Echo nach 150 ms, Dämpfung 6 dB)
- E Beidseitige Begrenzung (30 dB) mit nachfolgender Filterung 300...2000 Hz, 24 dB/Oktave

Die Versuchsreihe wird so aufgebaut, dass die Testpersonen im ungewissen sind, ob sie ein Test- oder ein Bezugssignal hören. Die Bezugssignale werden idealerweise in einer Art gewählt, dass sie unter sich eine eindeutige Qualitätsreihenfolge besitzen. Die Lage des Testsignals in diesem Bereich ist dann ein Mass für seine Qualität.

2.2 Einordnen in Qualitätskategorien

Bei diesen Methoden wird die Qualität direkt mit einer Note beurteilt (im allgemeinen 1...5). Es bedingt dies nicht nur eine Umschreibung der Noten, wie in der nachfolgenden Tabelle, sondern auch ein Einüben (sogenanntes Verankern) der Testpersonen an Beispielen, für die zum voraus Noten gegeben werden.

Note	Mögliche Umschreibungen
1	unbrauchbar
2	schlecht
3	befriedigend
4	gut
5	vorzüglich

Gegenüber den paarweisen Vergleichen und den Verständlichkeitstests haben die Noten-Methoden den Vorteil, auch bei Zweiweg-Verbindungen angewendet werden zu können. Damit wird der Einfluss der Rückfragemöglichkeit

mit einbezogen, was vor allem bei qualitativ schlechten Systemen von Bedeutung sein könnte. Dem stehen als Nachteile gegenüber, dass im allgemeinen nur gut geübte Testpersonen zuverlässige Resultate abgeben sowie die mässige Eignung bei qualitativ guten Übertragungsmethoden. Auch wurde bemerkt, dass bei länger andauernden Versuchen die Testpersonen ihr Verhalten ändern können [3].

2.3 Verständlichkeitstests

2.3.1 Logatom-Tests

Bei den Logatom-Tests versucht man die Verständlichkeit mit sinnleeren Silben, sogenannten Logatomen, zu bestimmen. Für jeden Messpunkt (bestimmte Dämpfung, Signal/Rausch-Abstand usw.) wird eine zufällige Auswahl aus einer Logatomliste [16] vorgelesen und dafür der Prozentsatz der richtig verstandenen Silben bestimmt. Die Logatome sind alle vom Typ Konsonant-Vokal-Konsonant (CVC). Da aus Tabellen der Anfangs- und Endkonsonanten sowie der Vokale beliebige Kombinationen gebildet werden können, sind die entstehenden Silben im allgemeinen ohne Sinn in irgendeiner Sprache. Die Erfahrung hat gezeigt, dass zuverlässige Resultate daher nur mit Testpersonen erhalten werden, die auf diese Methode eingeübt sind.

2.3.2 Reim-Tests

Bei den modernen, in öffentlichen Telephonesystemen eingesetzten Übertragungssystemen ist die Logatomverständlichkeit so hoch, dass sie (und unter Umständen auch die Noten-Methoden) kaum mehr für die Beurteilung verschiedener Systeme herbeigezogen werden kann. Es sind empfindlichere Methoden notwendig. Eine solche bietet sich mit den sogenannten Reim-Tests an, die für die englische Sprache in [17, 18, 19] und für das Italienische in [20] beschrieben worden sind. Im Gegensatz zu den Logatom-Tests verwendet man ein-, allenfalls zweisilbige Wörter, die in den jeweiligen Sprachen einen bekannten Sinn besitzen.

Da es sich um eine für unsere Anwendungen besonders interessante Methode handelt, soll in Abschnitt 4 noch ausführlicher darauf eingegangen werden.

2.4 Verhaltenstests

Die Verhaltenstests gehen darauf aus, Versuchspersonen eine bestimmte Aufgabe mit Hilfe einer Telefonverbindung lösen zu lassen und durch Beobachtung ihres Verhaltens Rückschlüsse auf die Übertragungsqualität zu ziehen. Naturgemäß eignet sich eine solche Methode besonders für Zweiwegverbindungen oder gar Konferenzgespräche. Mehr noch als bei andern Methoden muss hier die Gestaltung und Auswertung der Versuche mit grösster Sorgfalt vorgenommen werden, wenn man vermeiden will, dass die Resultate durch Nebeneffekte (Nichteignung der Versuchspersonen,

Einfluss der gestellten Aufgabe, Zufälligkeit des sich Herausbildens eines Gesprächsleiters usw.) zu stark beeinflusst werden.

2.4.1 CCITT-Empfehlung P. 74

Es handelt sich dabei um eine der einfachsten möglichen Formen von Verhaltenstests. Während einer längeren Gesprächsdauer oder einer grösseren Zahl von Gesprächen wird die Zahl der Rückfragen beobachtet, die wegen einer Beeinträchtigung der Übertragungsqualität notwendig werden.

2.4.2 Methode nach Richards und Swaffield

Bei dieser Methode lösen die beiden Gesprächspartner eine Aufgabe in der Form eines Zusammensetzungsspiels. Nach abgeschlossener Lösung erfolgt eine Bewertung der Übertragungsqualität entweder gemäss 2.2 (Noten), also in rein subjektiver Art, oder indirekt durch eine Zeitmessung.

2.4.3 Methode der Bell-Laboratorien

Zur Beurteilung des Einflusses von Laufzeiten (hervorgerufen durch Satellitenverbindungen) auf Konferenzgespräche wurde ein Test für Dreiergruppen entwickelt. Jeder Teilnehmer erhält je eine Serie von Karten mit aufgedruckten Symbolen. Die Aufgabe besteht darin, herauszufinden, welche Karten an sämtliche Teilnehmer verteilt und welche nur individuell abgegeben wurden. Als Mass für die Übertragungsqualität werden die beiden objektiv erfassbaren Grössen Zeitbedarf bis zur Lösung und Anzahl Fehler je Lösung verwendet.

2.5 Sprechererkennbarkeit

Es handelt sich hier um ein sehr vielschichtiges Problem, das bei weitem noch nicht gelöst worden ist. Die bisherigen Versuche beschränkten sich mehrheitlich auf kleine Gruppen von Sprechern und wurden weniger zur Beurteilung der

Übertragungsqualität als aus Interesse am Problem an sich durchgeführt. Eine sehr gute Übersicht des heutigen Standes findet sich in [23].

3. Präferenztests

Im letzten Abschnitt wurde bereits kurz auf zwei Methoden des paarweisen Vergleichs eingegangen. Beide gehen davon aus, dass als Vergleichsbasis reproduzierbare und in irgendeiner Form normierte Bezugssignale verwendet werden können, meistens in der Form «Signal+Störung» oder «Signal×Störung». Man sucht dann jenes Bezugssignal, bei dem das Testsignal als gleichwertig beurteilt wird, und benutzt diese Aussage als indirektes Mass für die Qualität. Nun sind aber auch noch andere, einfachere Formen des paarweisen Vergleichs möglich, nämlich:

- Der direkte vollständige Vergleich: Sämtliche m Testsignale werden paarweise jedes mit jedem verglichen. Diese Methode ist natürlich nur dann anwendbar, wenn m klein ist, da immerhin mindestens $\frac{1}{2}m(m-1)$ Vergleiche notwendig sind. Sie ist dann angezeigt, wenn entweder ein Referenzsignal schwer zu definieren oder eine Reihe ganz ähnlicher Testsignale zu rangieren ist.
- Direkter Vergleich sämtlicher Testsignale mit einem einzigen Bezugssignal. Diese Methode drängt sich auf, wenn als Vergleichsbasis ein wohldefiniertes Signal, zum Beispiel das «theoretisch ideale System», herangezogen werden kann und es zum Beispiel lediglich um die Fixierung einer Störschwelle geht.

3.1 Der direkte vollständige Vergleich

Führt man jeden Vergleich sowohl in der Form A-B wie auch B-A durch, um Bevorzugungen aufgrund der Reihenfolge zu vermeiden, so können die $m(m-1)$ Resultate in der Form einer Präferenzmatrix aufgezeichnet werden, die für jedes Versuchspaar die Zahl der Bevorzugungen – beispielsweise B gegen A – angibt. *Figur 2* zeigt ein Beispiel einer solchen Matrix für $m = 6$ und $m = 10$ Testpersonen. Daraus kann für jedes Testsignal die Summe der Bevorzugungen berechnet werden, beispielsweise für das Signal a mit

$$\frac{100}{2 \cdot 10 (m-1)} (4+7+8+2+5+6+8+7+3+6) = 56\%.$$

Es ergibt sich dann eine Rangfolge wie in *Figur 3*.

Beispiele von Anwendungen finden sich in [3] für Vocoder-Signale sowie [27] für Abklärungen im Blick auf eine optimale Bandbegrenzung für Telefonieübertragung.

3.2 Vergleich mit einem einzigen Bezugssignal

Besonders einfach ist eine Versuchsreihe, wenn sich ein Vergleich mit einem einzigen Bezugssignal aufdrängt. Dies ist häufig der Fall in einer ersten Versuchsstufe, bei der es zunächst nur darum geht, die Testsignale in grobe Gruppen

a	b	c	d	e	f	
x	4	7	8	2	5	a
4	x	9	8	4	7	b
2	3	x	6	0	3	c
3	1	4	x	1	2	d
7	5	10	10	x	8	e
4	3	6	5	1	x	f

Fig. 2

Präferenzmatrix (Beispiel)

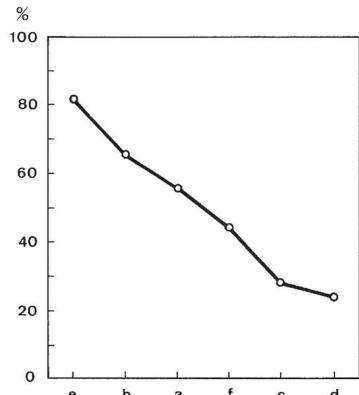


Fig. 3

Prozentuale Bevorzugung der Systeme a...f aufgrund der Präferenzmatrix von Figur 2

zu klassieren. Als Bezugssystem könnte zum Beispiel eine «idealisierte Telefonübertragung» mit 300...3400 Hz Bandbreite und definiertem Geräusch dienen; in einem ersten Schritt würden dann mit paarweisen Vergleichen alle Testsignale ausgeschieden, die schlechter beurteilt werden.

Als Beispiel soll nun ein Ausschnitt aus einer Versuchsreihe dargestellt werden, mit der der Einfluss von unvollkommenen Sende-Tiefpassfiltern in PCM- oder PAM-Systemen untersucht wurde. Dabei kamen spezielle Filter zur Anwendung, wie sie in *Figur 4* gezeigt sind. Bei einer Abtastfrequenz von 8 kHz wird das ideale Tiefpassfilter, das gerade noch keine Faltungsverzerrungen erzeugt [26], praktisch durch die Charakteristik '00' dargestellt. Dieses drängt sich daher als Bezugssystem auf.

Da Vorversuche zeigten, dass unter Umständen auch scheinbare Verbesserungen gegenüber theoretisch idealen Tiefpassfiltern auftreten könnten, wurden beim paarweisen Vergleich drei Antwortmöglichkeiten zugelassen:

- (1) A besser als B
- (2) kein Unterschied feststellbar
- (3) A schlechter als B.

Bei der Auswertung werden zunächst für jedes Filter die prozentualen Häufigkeiten der drei Antworten H1, H2 und H3 berechnet. Es sind verschiedene Methoden denkbar, um aus diesen wiederum einen resultierenden Gesamteindruck zu berechnen. Es ist zum Beispiel zweckmäßig, kontroverse Antworten (H1 und H3 etwa gleich gross) dem Entscheid «kein Unterschied» gleichzustellen. Diese Annahme wird mit dem Ausdruck

$$q = (1-H2)(H3-H1) = (H3+H1)(H3-H1)$$

berücksichtigt, der tatsächlich für die Fälle $H2 \rightarrow 1$ und $H1 \rightarrow H3$ unabhängig 0 wird, während für $H2 \neq 1$ und $H3 \neq H1$ q mit Vorzeichen und Grösse die Bevorzugung angibt. Es fragt sich, welcher Bereich um $q=0$ noch als

«gleichwertig» angesehen werden kann. Zur Abklärung dieser Frage wurde in eine normale Versuchsreihe eine Serie von 110 Paaren eingestreut, bei denen die beiden Signale A und B absolut gleich waren. Dabei zeigte sich folgendes Ergebnis:

$$(H1, H2, H3) \approx (1/4, 1/2, 1/4)$$

Streuung $\sigma(q) = 0,023$

Dementsprechend kann um den Wert $q=0$ ein Band von zum Beispiel $\pm 2\sigma(q)$ gelegt werden, das die Bereiche «besser als Referenz» und «schlechter als Referenz» trennt. Das Auswertungsergebnis für die Filterreihe in dieser Art ist in *Figur 5* dargestellt. Aufgrund dieser Darstellung können die Filter für weitere Versuche in zwei Gruppen eingeteilt werden, nämlich die Reihe 30...80, die eindeutig schlechter als das Referenzfilter sind, und die Filter 10 und 20.

3.3 Zur Beurteilung von Versuchspersonen

Subjektive Versuche schliessen naturgemäß das Risiko ein, dass durch ungeeignete Versuchspersonen die Ergebnisse verfälscht werden. Nun ist es aber sehr schwierig, wenn nicht unmöglich, eine «Normaltestperson» zu defi-

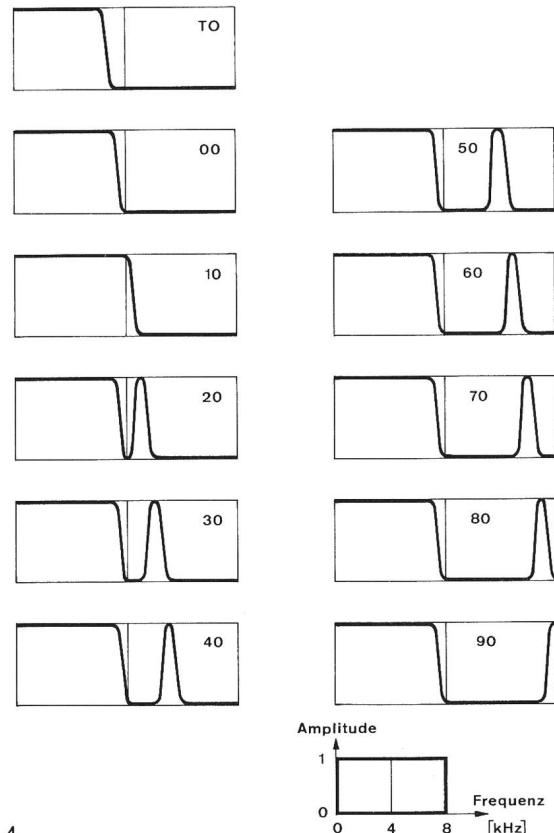


Fig. 4
Versuchs-Tiefpassfilter

nieren. Wohl können in einem ersten Schritt Personen ausgeschieden werden, deren Hörvermögen mangelhaft ist; die Eignung wird aber auch durch Komponenten bestimmt, die unter Umständen von Versuch zu Versuch ändern können. Dazu gehören Müdigkeit, Einstellung zu den Versuchen, vorübergehende Hörschwächen usw. Es lag daher nahe, nach einer Methode zu suchen, die eine gewisse Einschätzung der Aussagekraft von Resultaten gestattet und möglichst viele Unsicherheitsfaktoren erfasst.

3.3.1 Grundlagen der Beurteilung

Zu diesem Zweck werden alle Vergleiche zweimal angeboten, wobei die Testpersonen natürlich nicht wissen, in welchem Zeitpunkt ein Versuch wiederholt wird. Entsprechend den bereits erläuterten drei Möglichkeiten bei einer Bewertung ergeben sich für zwei Vergleiche insgesamt neun mögliche Ergebnisse, denen man je nach Grad der gegebenenfalls unterschiedlichen Bewertung «Strafpunkte» zuordnen kann. *Figur 6* zeigt die entsprechende Beurteilungsmatrix, die für die Versuche verwendet wurde. Offensichtlich totalisiert damit eine Versuchsperson bei n Versuchspaaren eine Strafpunktzahl s , die im Prinzip irgendwo zwischen 0 und n liegen kann.

3.3.2 Ein statistisches Modell

Die aufgrund von Versuchen ermittelte Strafpunktzahl s sagt natürlich für sich allein, ohne Vergleich mit Richtwerten, nicht allzuviel aus. Ein besseres Bild ergibt sich, wenn man mit einer fiktiven Testperson vergleicht, die ihre Resultate rein zufällig, also ohne Bezug auf das Gehörte, abgibt. Für dieses Modell kann mit Hilfe der Wahrscheinlichkeitsrechnung die Verteilungsfunktion für die Strafpunktzahl s berechnet werden. Da jede der Positionen der Beurteilungsmatrix in diesem Falle mit der Wahrscheinlichkeit $1/9$ belegt wird, muss man für jedes Versuchspaar den drei Fällen $s=0, 0,5$ und 1 die Wahrscheinlichkeiten $1/3, 4/9$ und $2/9$ zuordnen. Bei n Versuchspaaren trete x -mal das Ergebnis «1», y -mal «0,5» und $(n-x-y)$ -mal «0» auf. Dafür gibt es total

$$m(x, y) = \frac{n!}{x! y! (n-x-y)!}$$

Möglichkeiten.

Anderseits sind bei der Strafpunktbewertung mit n Versuchen insgesamt 9^n verschiedene Ergebnisfolgen möglich, so dass die Wahrscheinlichkeit, mit n Versuchen die Ergebniskombination $[x, y, (n-x-y)]$ zu erhalten, gleich

$$p(x, y) = \frac{n!}{x! y! (n-x-y)!} \left(\frac{2}{9}\right)^x \left(\frac{4}{9}\right)^y \left(\frac{1}{3}\right)^{(n-x-y)}$$

ist (Trinomialverteilung). Eine bestimmte Strafpunktzahl s kann im allgemeinen mit verschiedenen Kombinationen $[x, y, (n-x-y)]$ verwirklicht werden. Man erhält dementsprechend für die Wahrscheinlichkeit, in n Paaren von Zufallsbeurteilungen s Strafpunkte zu erzielen, den Ausdruck

$$P(s) = \sum_{x+y=s} p(x, y)$$

$$x + \frac{y}{2} = s$$

Für den Vergleich mit den tatsächlichen Versuchsergebnissen ist die kumulative Verteilung

$$Q(s) = \sum_{z=0}^s P(z)$$

aufschlussreich. Man ist nun in der Lage, folgende Aussage zu machen: «Bei s Strafpunkten ist die Wahrscheinlichkeit $\leq Q(s)$, dass die Versuchsperson nur Zufallsentscheidungen gefällt hat». Damit ist die gesuchte Referenz gefunden. *Figur 7* zeigt die Verteilungsfunktionen $Q(s)$ für verschiedene Werte von n . Der Figur kann entnommen werden, dass sich $P(s)$ mit zunehmendem n rasch einer Normalverteilung nähert. Dies lässt sich tatsächlich auch theoretisch ableiten. Man erhält als Näherung nämlich:

		Erster Vergleich:		
		1	2	3
Zweiter Vergleich:	1	0	0,5	1
	2	0,5	0	0,5
3	1	0,5	0	

Fig. 6
Beurteilungsmatrix

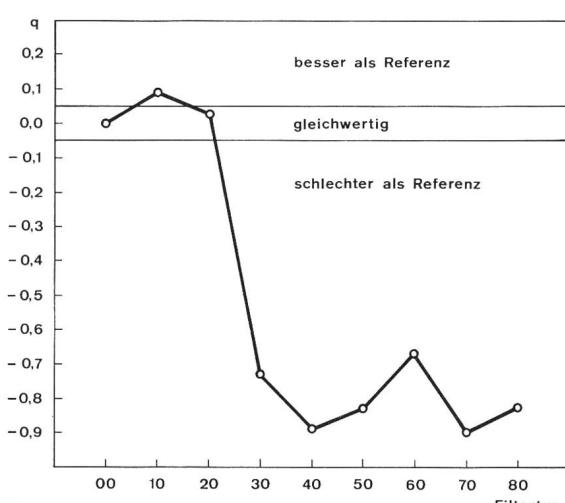


Fig. 5
Vergleich der VersuchsfILTER mit dem idealen Tiefpass '00'

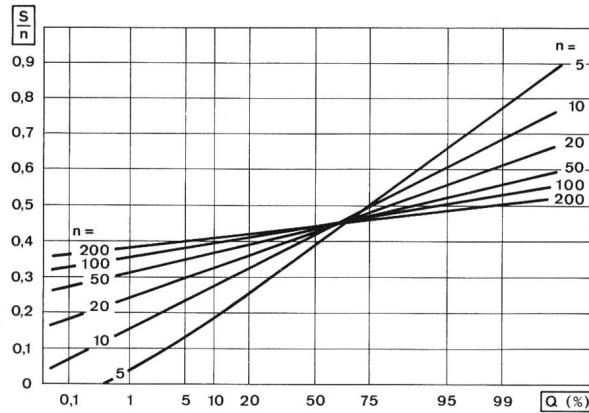


Fig. 7
Verteilungsfunktionen

$$P(s) = \frac{9}{\sqrt{22\pi n}} \cdot e^{-\frac{(9s-4n)^2}{22n}}$$

Mittelwert und Streuung dieser Funktion sind

$$E(s) = \frac{4n}{9} \quad \sigma(s) = \frac{11n}{9}$$

Für die Auswertung von Versuchen dient *Figur 8*, aus der die Grenzwerte s_{\max} für verschiedene Q entnommen werden können.

In der statistischen Praxis hat sich die Faustregel bewährt, dass zwei Verteilungen signifikant verschieden sind, wenn ihre Mittelwerte mehr als 3σ auseinanderliegen. In unserem Falle handelt es sich um die Verteilung für die Zufallsversuchsperson einerseits und um die praktisch in Versuchen erzielte Verteilung andererseits. Dementsprechend kann man einen praktischen Wert von $Q = 3\%$ annehmen.

3.3.3 Praktische Ergebnisse

Bei einer recht ausgedehnten Versuchsreihe mit 28 Testpersonen und $n = 75$ Versuchspaaren wurde die Strafpunktdistribution ermittelt. Sie ist in *Figur 9* wiedergegeben und zeigt, dass lediglich 3 Personen den Wert $s_{\max} = 27,4$ (aus *Fig. 8*) überschreiten. Die Mehrheit der Testpersonen beurteilte die Aufgabe bei diesem Versuch als recht schwierig, da viele Vergleiche annähernde Übereinstimmung ergaben. Umso mehr zeigt das Ergebnis, dass von der grossen Mehrheit der Testpersonen in Präferenztests signifikante Resultate erwartet werden können.

4. Reimtests

4.1 Versuchsprinzip

In den Reimtests werden im allgemeinen einsilbige Wörter verwendet, die den Aufbau Konsonant – Vokal – Konsonant (CVC) aufweisen. Im Gegensatz zu den Logatomen handelt es sich aber um Ausdrücke, die in den je-

weiligen Sprachen einen bekannten Sinn besitzen. Dies ist mit ein Grund, warum der Reimtest erfolgreich auch mit verhältnismässig ungeübten Testpersonen durchgeführt werden kann, was in Anbetracht des bei subjektiven Versuchen immer beträchtlichen Aufwandes sehr ins Gewicht fallen kann.

In der ursprünglichen, von *Fairbanks* eingeführten Form [17] wird eine Reihe der Worte abgespielt, wobei die Testpersonen auf ihrem Antwortblatt den Wortstamm -VC mit dem vermeintlich gehörten Anfangskonsonanten C' ergänzen müssen (Gruppe der Antwortmöglichkeiten also «offen»). Die Zahl der richtig bestimmten Wortanfänge ist direkt ein Mass für die Verständlichkeit, während die Art der Verwechslungen (zum Beispiel in der Form einer Verwechslungsmatrix) unter Umständen Hinweise auf Verzerrungsmechanismen geben kann. Damit kann der Reimtest in zweierlei Hinsicht aufschlussreiche Resultate liefern. In einer durch *House* u. a. sowie *Griffiths* [18, 19] modifizierten Form des Reimtests wird den Versuchspersonen auf dem Resultatblatt eine beschränkte Zahl von vollständigen Wörtern zur Auswahl gegeben (Gruppe der Antwortmöglichkeiten «geschlossen», sogenannte erzwungene Wahl). Zu-

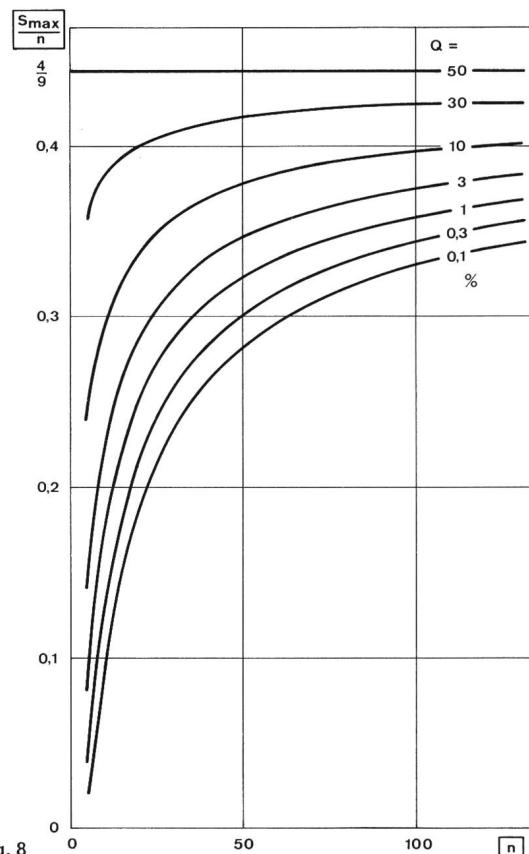


Fig. 8
Grenzwerte

sätzlich wird in einer zweiten Liste statt des Anfangselementes der Schlusskonsonant variiert. Als Beispiel wird ein Ausschnitt aus den englischen Listen der Publikation [18] in *Tabelle I* gezeigt.

Die Testpersonen erhalten als Resultatblatt eine Liste, die im wesentlichen das gleiche Aussehen wie Tabelle I hat, bei der aber in jeder Zeile die Wörter zufällig vertauscht sind. Vorsichtshalber wird mit mehreren Varianten der Zufallslisten gearbeitet, um eine mögliche Bevorzugung bestimmter räumlicher Anordnungen auszumitteln. Als Antwort wird jenes Wort angestrichen, das die Testperson zu vernehmen glaubte.

Die Zusammenstellung der Listen kann nach verschiedenen Aspekten geschehen. Während zum Beispiel in [17] auf eine «phonetische Ausgeglichenheit» innerhalb jeder Reihe, gleiches Schriftbild für gleiche Laute sowie der Umgangssprache angepasste Worthäufigkeiten geachtet wurde, liegt das Schwergewicht bei den Versuchen von *Griffiths* [19] bei der qualitativen Analyse der Verwechslungen, was die Auswahl der Worte beeinflusst hat. In Sprachen, die vom Englischen verschieden sind, kann es außerdem notwendig werden, vom Konzept der einsilbigen Worte abzugehen, wie das zum Beispiel die Arbeiten von *Ibba* und *Bianchi* [20] für das Italienische gezeigt haben. Diese verwenden zweisilbige Wörter vom Typ C'VCV oder CVC'V mit C' als dem in einer Zeile variablen Konsonanten. Dabei werden teilweise auch Doppelkonsonanten zugelassen, so dass zum Beispiel folgende Zeilen möglich werden:

C'VCV: trema – prema – brema – frema – crema
 CVC'V: cerno – cerbo – cervo – certo – cerco

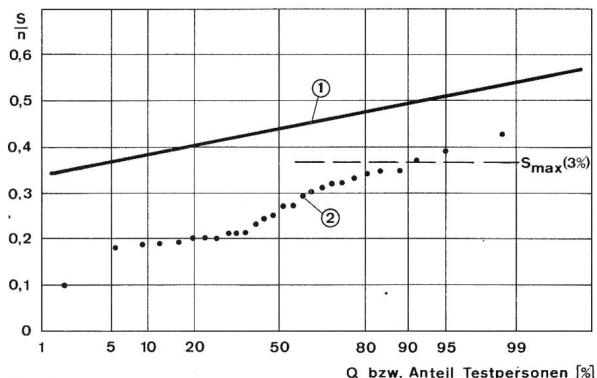


Fig. 9

Beurteilung der Versuchspersonen in einem praktischen Versuch

n = 75

① Zufallsverteilung

② Versuchsreihe mit 28 Testpersonen

Wichtig für die praktische Anwendung sind die Feststellungen, die aufgrund von Versuchen vor allem in [18] gemacht wurden:

- Reimtests können erfolgreich auch mit verhältnismässig ungeübten Testpersonen durchgeführt werden.
- Bereits mit Gruppen von 10...20 Personen werden zuverlässige Resultate erzielt.
- Die Resultate sind sehr gut reproduzierbar; es kann auch kein Lerneffekt nachgewiesen werden.

Tabelle I. Reimtest-Listen / Auszug aus [18]

	A	B	C	D	E	F
CVC'-Liste	1 bat	bad	back	bass	ban	bath
	2 beam	beach	beat	beam	bead	beak
	3 bun	bus	but	buff	buck	bug
	4 came	cape	cane	cake	cave	case
	25 teach	tear	tease	teal	team	teak
C'VC-Liste	1 led	shed	red	bed	fed	wed
	2 sold	told	hold	fold	gold	cold
	3 dig	wig	big	rig	pig	fig
	4 kick	lick	sick	pick	wick	tick
	25 way	may	say	gay	day	pay

Tabelle II. C'VC-Liste

Sieb	lieb	Dieb	gib	Hieb	schieb
Bach	Fach	wach	Dach	Sach	mach
Schicht	Sicht	Gicht	Wicht	dicht	nicht
weich	gleich	reich	Laich	Teich	Deich
fliegt	biegt	siegt	schmiegt	wiegt	liegt
Speck	streck	Dreck	Reck	Geck	Schreck
dick	Trick	Blick	nick	flick	Klick
Troll	Moll	Zoll	Groll	scholl	soll
Schaum	kaum	Baum	Saum	Traum	Raum
Kahn	Schwan	Hahn	Bahn	Kran	Wahn
dein	nein	Hain	Wein	Schein	fein
Land	Hand	Stand	fand	Sand	Tand
find	Rind	lind	Kind	blind	sind
Sund	wund	Fund	Schund	Mund	Schlund
Rang	Hang	lang	Gang	bang	Drang
frank	Dank	Wank	Schrank	Schwank	trank
Star	Zar	Schar	Paar	wahr	gar
Heer	Teer	sehr-	Wehr	Meer	kehr
gern	lern	Kern	Stern	fern	Bern
Schluss	Schuss	Nuss	Fluss	Kuss	Guss
preist	Geist	meist	kreist	schreist	weisst
mäht	säht	näht	bläht	lädt	spät
Beet	fleht	geht	seht	dreht	steht
Not	Boot	tot	Schot	Kot	Lot
Wut	Brut	Blut	Mut	Nut	Hut

Tabelle III. CVC'-Liste

Bann	Band	bald	Ball	Balg	Balm
bieg	Bien'	bis	Bier	Biest	Biet
denk	der	den	Deck	des	dem
Dorsch	Dock	Dorn	Dorf	doch	dort
dumm	Dunst	Durst	Dung	durch	dumpf
Fett	Fels	fesch	Feld	feg	Fell
Genf	Geld	Ger	Gen	geht	gelb
Heft	Helm	hell	Hemd	Held	Hecht
hohl	Hof	hob	hoch	Hohn	Holz
kennt	kess	keck	Kehl	Kelch	kehr
lebt	lehr	Lehm	Lehn	lehnt	Lenz
Mann	Markt	matt	Mast	Macht	Magd
Mist	Mink	Milz	mit	Milch	mir
leicht	Lein	leis	Leim	Leib	Leid
weiss	weit	Wein	Weib	weil	weich
Rad	Rahm	Rand	Rat	rar	Rang
raunt	Rausch	Rauch	raus	rauh	Raub
seit	seicht	Seil	seid	Seim	sein
selb	Senf	sehr	sechs	Senn	Sekt
sich	Sims	Sitz	sind	sing	Sinn
fahr	Fach	fand	Falt	fast	Fass
Tor	Tod	top	Ton	toll	Topf
Wald	Wahl	wach	Waag	Wand	was
Wind	Wirt	Wink	Wirz	wir	Wild
Wucht	Wurm	Wunsch	Wulst	Wurf	wund

4.2 Reimtests für die deutsche Sprache

4.2.1 Wortlisten

Die bereits erwähnten Vorteile haben uns bewogen, auch für die deutsche Sprache Reimtests zu entwerfen. (Mit ein Grund war die Art der geplanten Anwendung, da die betreffenden Verzerrungen – wie sich in Vorversuchen zeigte – unter Umständen Konsonantenverwechslungen begünstigen.) Neben den gebräuchlichen deutschen Wörterbüchern war dabei das Häufigkeitswörterbuch von *Kaeding* [25] eine besondere Hilfe, da damit in beschränktem Umfang auch die Wort- und Silbenhäufigkeiten berücksichtigt werden konnten. Es zeigte sich bald, dass es wohl möglich ist, Listen mit einsilbigen sinnvollen Wörtern vom Typ CVC zu konstruieren, dass dies aber im erwünschten Umfang von mindestens 5×25 bereits nur noch mit einigen zusätzlichen Freiheiten möglich ist. Tabelle II zeigt die entsprechende C'VC-Liste und Tabelle III die CVC'-Liste. Insbesondere wurden anstelle von einfachen Konsonanten auch mehrfache zugelassen, die in der deutschen Sprache recht häufig auftreten. Zum Teil geht dadurch die Eigenschaft, dass in einer Zeile lediglich Wörter, die sich in einem Phonem¹ unterscheiden,

¹ Phonem: Kleinstes Element lautlicher Art, das die Unterscheidung von zwei Wörtern erlaubt.

verloren. Es ist aber möglich, Gruppen zu konstruieren, bei denen nur ein einziges Element variiert (Tabellen IV und V). Dabei muss in Kauf genommen werden, dass die Vokal- und Silbenhäufigkeiten etwas weniger den Mittelwerten entsprechen.

Da gewisse Verzerrungen auch zu Vokalverfälschungen führen können, haben wir zusätzlich versucht, eine Wortliste von der Art CV'C zu entwerfen. Es zeigte sich aber bald, dass dies für die deutsche Sprache mit grossen Schwierigkeiten verbunden ist; das Resultat ist entsprechend bescheidenen Umfangs und in Tabelle VI dargestellt.

4.2.2 Ein Anwendungsbeispiel

Die C'VC-Liste von Tabelle 2 wurde – neben anderen Versuchen – eingesetzt, um den Einfluss der bereits in Abschnitt 3 erwähnten Tiefpassfilter auf die Verständlichkeit zu untersuchen. Ein Ausschnitt aus einer Versuchsreihe soll zum Anlass genommen werden, um für den Reimtest ein konkretes Anwendungs- und Auswertungsbeispiel zu beschreiben. Während bei Filter T0 noch keinerlei Verzerrungen auftreten, führen die Filter 20 und 40 zu Faltungsverzerrungen bei 3,5 beziehungsweise 2,5 kHz (Spiegelung an der halben Abtastfrequenz von 4 kHz).

Die Versuche wurden mit 6 Gruppen zu je 4 Testpersonen durchgeführt, wobei als Antwortblatt 8 verschiedene Zufallsversionen der Tabelle II verwendet worden sind. Jede Kolonne wurde für jedes Filter 4 mal verwendet, je mit verschiedenen Antwortblättern.

Die folgende Tabelle enthält das Resultat einer Auswertung der Verständlichkeit:

Filter	Mittelwert der falsch verstandenen Wörter	Streuung der sechs verschiedenen Kolonnen
	%	%
T0	4,8	3,1
20	2,5	1,6
40	4,5	1,3

Diese Tabelle zeigt zunächst:

- Die Verständlichkeit wird trotz unangenehm empfundenen Faltungsverzerrungen (siehe Abschnitt 3.2, Filter 40) nur unwesentlich beeinflusst und ist so gross, wie die in [17...20] bei hohen Signal/Rausch-Verhältnissen gemessenen Werte.
- Eine «dosierte» Faltungsverzerrung führt im Falle des Filters 20 zu optimalen Verständlichkeitswerten. Die Unterschiede sind aber gemäss Tabelle nicht sehr bedeutend, wie die Streuungen zeigen.

Tabelle IV. C'VC-Liste mit minimalen Phonem-Differenzen

Sieb	Dieb	gib	Hieb	schieb
Bach	Fach	wach	Dach	Sach
Schicht	Sicht	Gicht	Wicht	dicht
weich	reich	Laich	Teich	Deich
Schaum	kaum	Baum	Saum	Raum
dein	Hain	Wein	Schein	fein
Land	Hand	fand	Sand	Tand
find	Rind	lind	Kind	sind
Sund	Fund	wund	Schund	Mund
Rang	Hang	lang	Gang	bang
Zar	Schar	Paar	wahr	gar
Heer	Teer	sehr	Wehr	kehr
gern	lern	Kern	fern	Bern
mäht	säht	näht	lädt	rät
Not	Boot	Lot	Kot	tot
Wut	Hut	gut	tut	Nut
Kuss	Guss	Bus	muss	Nuss
Kahn	Hahn	Bahn	Wahn	Zahn
voll	Moll	Zoll	scholl	soll
fahl	Saal	Schal	Zahl	Wahl
Gier	Tier	dir	wir	hier
feil	Seil	Teil	Beil	geil
bog	log	Sog	zog	wog
laust	baust	Faust	schaust	saust
Hatz	Katz	Latz	Satz	Schatz

Tabelle V. CVC-Liste mit minimalen Phonem-Differenzen

weil	Wein	weiss	weich
Tor	Ton	Tod	Topf
Laich	leis	Leib	Leid
Reiz	Reich	Reif	Reim
bald	Ball	Balg	Balm
fein	feil	feiss	feig
bieg	bis	Bier	Biet
Pack	Pass	Patt	Pan
Lob	log	Lohn	Lot
weiss	weich	Weib	weit
schad	Schal	Schaf	Schar
der	den	des	dem
Dorsch	Dorn	Dorf	dort
raus	Rausch	Rauch	Raub
Hof	hoch	hohl	Hohn
dann	Damm	dass	Dach
Russ	Ruhm	Ruhr	Ruf
kess	keck	Kehl	kehr
Gischt	girrt	Gicht	Gift
Rat	rar	Rad	Rahm

Aufschlussreicher ist die sogenannte *Verwechslungsmatrix*, die ein Bild der aufgetretenen Verwechslungen wiedergibt. In *Tabelle VII* ist ein Ausschnitt (nur betroffene Konsonanten) aus dieser Tabelle dargestellt. Man kann daraus folgendes entnehmen:

- Die Abnahme der Verwechslungen bei Filter 20 gegenüber Filter T0 ist nicht zufällig, sondern erfolgt für die Mehrheit der Laute gleichmässig.
- Das Ergebnis für die Filter 40 wird ganz einseitig durch die Verfälschung des Konsonanten S bestimmt. Dies deckt sich mit den qualitativen Aussagen der Testpersonen in einem andern Versuch, indem diese die Verzerrungen mit Bezeichnungen wie «Lispeln, Zischen, verfälschte Zischlaute» charakterisieren.

Tabelle VI. CV'C-Liste

Senn	Sinn	Sonn	sein
lieg	leg	lüg	log
wäg	wog	Weg	wag
reisst	Rast	Rost	Rest
leb	Leib	lieb	Lob
hob	heb	Heib	Hub
biet	Boot	Beet	böt
wahr	wir	Wuhr	wer
Deck	duck	Dock	dick
feil	Fell	Fall	voll
Most	maust	meist	Mist
Brut	Braut	brat	breit

Tabelle VII. Verwechslungsmatrizen für 3 Tiefpassfilter

		Ausgesprochener Laut																
		D	F/V	FR	G	K	KR	PR	R	S	ST	SCH	SCHL	SCHR	T	TR	W	Z
Filter 10	B																	
	D																	
	DR																	
	G																	
	H																	
	K																	
	KR																	
	L																	
	N																	
	P																	
Filter 20	R																	
	S																	
	SCH																	
	SCHW																	
	T																	
	TR																	
	W																	
	D																	
	DR																	
	FL																	
Filter 40	GR																	
	H																	
	K																	
	KP																	
	S																	
	SCH																	
	T																	
	TR																	
	W																	

Der Reimtest ist also eine recht empfindliche Versuchsmethode. Sie gestattet, nicht nur quantitative Aussagen über die Verständlichkeit zu machen; eine Analyse der Verwechslungen ermöglicht es unter Umständen, die Auswirkungen von Verzerrungen konkreter zu umschreiben (Störmechanismus). In diesem Sinne stellt sie auch eine wertvolle Ergänzung anderer Methoden dar.

5. Zur praktischen Durchführung von Versuchen

5.1 Versuchsanordnung

Die Versuche wurden in Gruppen von 4 Testpersonen durchgeführt. Wie Figur 10 zeigt, dient als Signalquelle ein Zweikanal-Tonbandgerät, wobei die beiden Kanäle A und B gleichzeitig den 4 Schutznischen zugeführt werden können. Die Nischen bewirken einerseits eine bescheidene zusätzliche Abschirmung gegen externe Geräusche, vor allem aber dienen sie dazu, dass die Versuchspersonen nicht abgelenkt werden können. Die Testsignale werden mit Kopfhörern abgehört, wobei für gewisse Versuche (siehe 5.3) die Möglichkeit besteht, mit einem Taster zwischen den beiden Kanälen A und B zu wählen.

5.2 Sprachmaterial

Offensichtlich stehen oder fallen Untersuchungen an Sprachsignalen mit der Qualität der verwendeten Originale, da diese den Ausgangspunkt aller Versuche darstellen. Es empfiehlt sich daher, die entsprechenden Aufnahmen nicht nur in technisch einwandfreier Qualität durchzuführen, sondern auch Texte und Sprecher sorgfältig auszuwählen. Vor allem hat sich bewährt, Sprachmaterial in einem größeren Umfange bereitzustellen, als das zunächst notwendig erscheint. Bei späteren Versuchen kann dann beispielsweise umso rascher verifiziert werden, dass die gemessenen Resultate nicht vom gewählten Text, dem Sprecher oder der verwendeten Sprache abhängen.

Für die Präferenztests kommen neben Ausschnitten aus Prosatexten besonders ausgewählte Sätze in Frage. In verschiedenen Sprachen wurden Listen von sogenannt «phonetisch ausgeglichenen» Sätzen konstruiert, deren lautliche Zusammensetzung einer Statistik der Umgangssprache möglichst nahe kommt. Im Englischen wird zum Beispiel häufig die *Harvard-Liste* [22] verwendet.

Tonbandgerät Revox A77

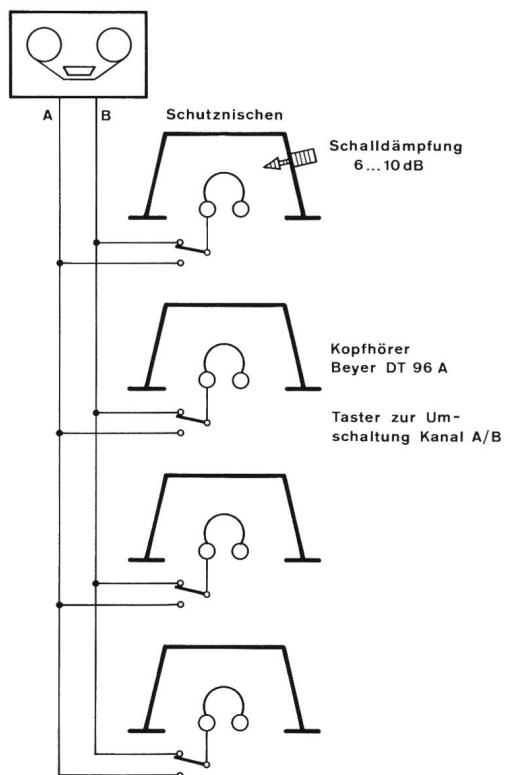


Fig. 10
Versuchsanordnung

5.3 Varianten des paarweisen Vergleichs

Üblicherweise werden Präferenztests durchgeführt, indem man die beiden Signale A und B zeitlich nacheinander abspielt (A-B-A-B). Damit noch ein richtiger Vergleich möglich ist, dürfen die entsprechenden Texte nicht zu kurz, aber auch nicht allzu lang sein, typischerweise in der Größenordnung von 5 Sekunden. Der Trennzeiten wegen dauert daher ein solcher Vergleich mindestens 30 Sekunden, obwohl nur ein Bruchteil davon als Sprachausschnitt benutzt wird. Es liegt daher nahe, statt eines Vergleichs «nacheinander» einen solchen «nebeneinander» zu versuchen, indem die Testpersonen mit einer Taste auswählen können, welche der beiden synchron ablaufenden Varianten A und B sie gerade hören wollen (Fig. 10). Bei einem mehrmaligen Wechsel während einer Textzeit von 20...40 Sekunden sollte dann ebenfalls ein eingehender Vergleich möglich werden, der namentlich in Fällen, wo Verzerrungen oder Störungen verhältnismässig selten auftreten, unter Umständen sogar gesicherter wird. Wiederum mit den bereits in Abschnitt 3 vorgestellten Filtern wurden drei Versuchsreihen durchgeführt (analog 3.2), und zwar

- (1) Methode «nebeneinander», deutscher Text von etwa 30 s Dauer, Sprecherin
- (2) Methode «nacheinander», englischer Satz von etwa 4 s Dauer, Sprecher
- (3) Methode «nacheinander», englischer Satz von etwa 4 s Dauer, Sprecherin

Nach der Auswertung gemäss Abschnitt 3 ergaben sich zwischen den drei Reihen von je 10 Vergleichspaaren folgende Korrelationskoeffizienten:

$$\begin{aligned} K_{12} &= 0,95 \\ K_{13} &= 0,80 \\ K_{23} &= 0,90 \end{aligned}$$

Die beiden Methoden können also bei dieser Anwendung als ebenbürtig bezeichnet werden. Interessanterweise wertete die Mehrheit der Versuchspersonen die sonst übliche Methode «nacheinander» als angenehmer, aber anstrengender. Als weiteres Ergebnis zeigt der Vergleich, dass Präferenztests unabhängig von Sprache und Sprecher – mindestens bei den untersuchten Verzerrungen – sehr gut übereinstimmende Resultate liefern können.

Bibliographie

- [1] Fletcher H. Speech and Hearing in Communication. D. Van Nostrand-Verlag 1957. French N.R., Steinberg J.C. Factors Governing the Intelligibility of Speech Sounds. J. Acoust. Soc. Am. 19 (1947) 1.
- [2] CCITT Red Book Vol V. Part II, Annex 30 (Method Used by the Swiss Administration for the Determination of Transmission Quality Based on Objective Measurements).
- [3] Pachel W.P., Urbanek G.E., Rothauser E.H. Preference Evaluation of a large set of Vocoded Speech Signals. IEEE Trans. AU-19 (1973) 3.
- [4] Atal B.S., Hanauer S.L. Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. J. Acoust. Soc. Am. 50 (1971) 2.
- [5] Hecker M., Guttman N. Survey of Methods for Measuring Speech Quality. J. Audio Eng. Soc. 15 (1967) 4.
- [6] McGee V.E. Determining Perceptual Spaces for the Quality of Filtered Speech. J. of Speech and Hearing Res. 8 (1965) p. 3.
- [7] Mc Dermott B.J. Some Factors that Influence Judgement of Circuit Quality. J. Acoust. Soc. Am. 39 (1966) p. 1257.
- [8] Voiers W.D. Perceptual Spaces of Speaker Identity. J. Acoust. Soc. Am. 36 (1964) p. 1065.
- [9] Munson W.A., Karlin J.E. Isopreference Methods for Evaluating Speech-Transmission Circuits. J. Acoust. Soc. Am. 34 (1962) p. 762.
- [10] Rothauser E.H., Urbanek G.E., Pachl W.P. Isopreference Methods for Speech Evaluation. J. Acoust. Soc. Am. 44 (1968) p. 408.
- [11] Hecker M.H.L., Williams C.E. Choice of Reference Conditions for Speech Preference Tests. J. Acoust. Soc. Am. 39 (1966) p. 946.
- [12] Swaffield J., Richards D.L. Ratings of Speech Links and Performance of Telephone Networks. Proc. IEE 106 (1959) Part B, p. 65.
- [13] Richards D.L., Swaffield J. Assessment of Speech Communication Links. Proc. IEE 106 (1959) Part B, p. 77.
- [14] CCITT Red Book, Vol. V. Part II, Annex 2, p. 595.
- [15] CCITT Red Book, Vol. V, Part I, Recommendation P. 45.
- [16] Sotschek J. Über verschiedene Lautevorräte zur Bildung von Logatomen für Sprachverständlichkeitsmessungen. Techn. Ber. des Forschungsinstitutes im FTZ (Deutsche Bundespost) A 13 TBr, 6. Juli 1970.
- [17] Fairbanks G. Test of Phonemic Differentiation: The Rhyme Test. J. Acoust. Soc. Am. 37 (1963) 1.
- [18] House A.S. u.a. Articulation-Testing Methods: Consonantal Differentiation with a Closed Response Set. J. Acoust. Soc. Am. 37 (1965) 1.
- [19] Griffiths J.D. Rhyming Minimal Contrasts: A Simplified Diagnostic Articulation Test. J. Acoust. Soc. Am. 42 (1967) 1.
- [20] Ibbi G., Bianchi F. Studio preliminare per l'addattamento alla lingua italiana del metodo delle parole rimate. Note Recensioni e Notizie 20 (1971) 3.
- [21] Vartabedian A.G. The Effects of Transmission Delays in Four-Wire Teleconferencing. Bell Syst. Tech. J. 45 (1966) 10.
- [22] IEEE Recommended Practice for Speech Quality Measurements. IEEE Trans AU-17 (1969) 3.
- [23] Bolt R.H. u.a. Speaker Identification by Speech Spectrograms; a Scientist's View of its Reliability for Legal Purposes. J. Acoust. Soc. Am. 47 (1970) 2, Part 2.
- [24] CCITT White Book, Vol. V, Part I, Recommendation P. 74 (Repetition Observation Test).
- [25] Kaeding F.W. ed. Häufigkeitswörterbuch der deutschen Sprache. Berlin-Steglitz 1897 (Selbstverlag des Herausgebers).
- [26] Kündig A. Digitale Telephonie. Verlag Hallwag Bern 1972 (Blaue Reihe Heft 103).
- [27] Gleiss N. The Effect of Bandwidth Restriction on Speech Transmission Quality in Telephony. 4th Internat. Symp. on Human Factors in Telephony, Bad Wiessee 1968. VDE-Verlag, Berlin.