

Zeitschrift: Technische Mitteilungen / Schweizerische Post-, Telefon- und Telegrafienbetriebe = 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 Telegrafienbetriebe

Band: 60 (1982)

Heft: 1

Artikel: Mensch-Maschine-Kommunikation mit Sprachsignalen

Autor: Mangold, Helmut / Schenkel, Klaus-Dieter

DOI: <https://doi.org/10.5169/seals-876142>

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: 08.01.2026

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

Mensch-Maschine-Kommunikation mit Sprachsignalen¹

Helmut MANGOLD und Klaus-Dieter SCHENKEL, Ulm

007.51:612.78:534.78:62-5

Zusammenfassung. Die Autoren be-
fassen sich mit dem heutigen Stand der
automatischen Erkennung und der Syn-
these der Sprache. Dabei werden die
Grundsätze der digitalen Sprachverar-
beitung und die Fragen der syntheti-
schen und halbsynthetischen Sprache
erläutert. Im weiteren sind die Probleme
der automatischen Spracherkennung,
deren Lösung zum Dialog mit dem Com-
puter führen wird, ebenfalls behandelt.
Ein Blick in die Zukunft schliesst den Ar-
tikel ab.

Communication orale homme- machine

Résumé. Les auteurs décrivent l'état
actuel de la reconnaissance automati-
que et de la synthèse de la parole. Après
avoir abordé les principes du traitement
numérique de la parole, ils traitent des
questions de la parole semi-synthétique
et synthétique. De plus, les problèmes
de la reconnaissance automatique de la
parole, dont la solution conduira au dia-
logue avec l'ordinateur, sont également
examinés. L'article se termine par des
perspectives d'avenir.

Comunicazione tra l'uomo e la mac- china con segnali linguistici

Riassunto. Gli autori si occupano dello
stato attuale del riconoscimento auto-
matico e delle sintesi della lingua, spie-
gando i principi dell'elaborazione digi-
tale della lingua sintetica e semisinte-
tica. Inoltre trattano anche i problemi del
riconoscimento automatico della lingua
la cui soluzione porterà al dialogo con il
computer. L'articolo termina con delle
prospettive sul futuro.

1 Szenarium 2000 — Utopie oder Wirklichkeit?

Es ist 7 Uhr morgens. Herr Schuberth, Vertriebsach-
bearbeiter in einem grossen Automobilwerk, träumt ge-
rade noch vom letzten Urlaub, da ertönt aus einem klei-
nen Kästchen auf seinem Nachttisch eine zarte Stimme
«Bitte aufstehen», und nach ein paar Minuten etwas lau-
ter «Bitte aufstehen». Herr Schuberth weiss, das ist sein
Wecker, der ihn gerade mit Sprachausgabe geweckt
hat. Er weckt seine Frau — aber auf ganz natürliche
Weise.

Ein kurzes Kommando «Licht an» genügt, um im
Schlafzimmer eine sanfte Morgenbeleuchtung anzu-
schalten. Und nun geht es weiter: Im Badezimmer schal-
tet er mit den Worten «35 Grad» sein Duschwasser auf
die richtige Temperatur. Vorher hatte er noch den Mi-
krowellenherd in der Küche mit Sprachkommandos ein-
geschaltet. Der meldet sich nun bereits mit «Kochvor-
gang beendet». Die Tür zu seiner Garage öffnet Herr
Schuberth mit einem akustischen Schlüssel: Er sagt
«Trebusch» — sein Name von rückwärts. Während der
Fahrt mit dem Auto leitet ihn das Verkehrsweginforma-
tions- und -auskunftssystem VIA. Er bekommt mit
Sprachausgabe Hinweise über die zweckmässige Fahrt-
strecke, nachdem er vorher sein Fahrtziel in ein Mikro-
fon gesprochen hat, das in seinem Auto bereits in das
Steuerrad integriert ist. Kurz bevor er sein Ziel erreicht,
macht ihn sein fahrzeugeigenes Kontrollsystem noch
aufmerksam «Bitte Ölstand überprüfen».

Nachdem Herr Schuberth an der Pforte seines Werkes
den Satz «Meine Name ist Schuberth» gesprochen hat
und das akustische Zugangskontrollsystem ihn als be-
rechtigt passieren liess, erreicht er gegen 8.30 Uhr sein
Büro. Sein Schreibtisch (Fig. 1) sieht recht aufgeräumt
aus, zwei Bildschirme, von denen man den einen fast
nicht sieht, da er in die Tischfläche eingelassen ist, Mi-
krofon und Lautsprecher und ein elektronischer Bleistift,
mit dem er später auf diesem Bildschirm schreiben wird.
Dazu kommen noch zwei kaum sichtbare Öffnungen, in
die man beschriebene Papierstücke eingeben oder ent-
nehmen kann; es sind die Sensoren für die Lese- bezieh-

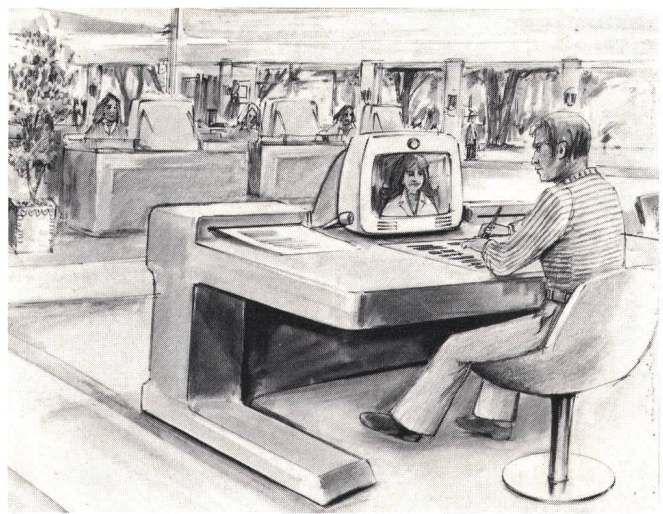


Fig. 1
Schreibtisch mit integrierten Kommunikationssystemen — Utopie
oder Wirklichkeit im Jahre 2000?

ungsweise Schreibmaschine, die direkt mit dem inner-
betrieblichen Kommunikationssystem verbunden ist.

Kaum hat Herr Schuberth in das Mikrofon «Bitte die
heutigen Termine» gesprochen, kommt auch bereits aus
der Ausgabeöffnung ein Stück Papier mit den anstehen-
den Aufgaben. Diese Information erscheint gleichzeitig
auf seinem Bildschirm. Herr Schuberth entschliesst
sich, einen über eine verzögerte Lieferung verärgerten
Autohändler anzurufen. Er sagt dazu in das Mikrofon:
«Telefonverbindung — Herr Obermaier, Autohaus Ober-
maier, München.» Alles andere erledigt sein Telefonsy-
stem, das sich erst wieder meldet, nachdem die Verbin-
dung hergestellt und Herr Obermaier selbst am Apparat
ist. Nachdem er ihn beruhigen konnte, bittet er seine As-
sistentin, Frau Schreiber, zu sich. Mit ihr bespricht er die
inzwischen auf dem Bildschirm und der Papierausgabe
eingelaufene Post. Frau Schreiber, bis vor einigen Jah-
ren noch Sekretärin, heute eine vielgefragte Kommuni-
kationsassistentin, beantwortet im Laufe der nächsten
Stunde die meisten Briefe. Dazu braucht sie keine
Schreibmaschine mehr, sondern sie spricht in ein Mikro-
fon, der Text ist sofort auf dem Bildschirm zu sehen, und

¹ Referat gehalten am 17. internationalen Technischen Presse-Collo-
quium von AEG-Telefunken

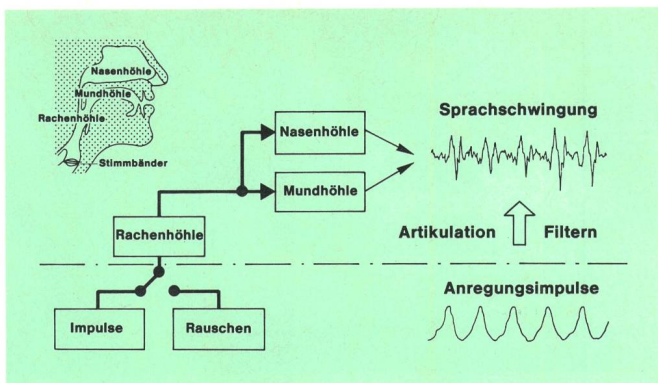


Fig. 2
Modell der Spracherzeugung

er wird mit dem schon seit vielen Jahren eingeführten Teletextsystem zum Adressaten übertragen, nachdem ihn Herr Schubert vorher mit dem Schreibstift direkt auf seinem Bildschirm abgezeichnet hat.

Da ein über einen schadhafte Reifen recht ungehaltener Kunde dem Werk mit juristischen Schritten droht, informiert sich Herr Schubert kurz bei der Rechtsdatenbank über den aktuellen Stand auf dem Gebiet der Herstellerhaftung. Die Kernsätze der jüngsten Entscheidungen erhält er auf dem Bildschirm, während die umfangreichen Kommentare über Lautsprecher dargeboten werden. Die Assistentin, Frau Schreiber, lässt sich in der Zwischenzeit vom Fahrplan-Auskunftssystem «Karlsruhe» — die Bundesbahn hatte gerade das 20jährige Jubiläum der Installation dieses erfolgreichen Systems begangen — eine zweckmässige Bahnverbindung für die morgige Reise von Herrn Schubert nach Rosenheim geben.

Hier wollen wir uns fürs erste aus dem Jahr 2000 wieder ausblenden, denn mit «Karlsruhe», dem derzeit umfangreichsten Auskunftssystem der Welt mit synthetischer Sprache, sind wir direkt in der Gegenwart: Seit fast zwei Jahren tut dieses System, das unter Federführung des Bundesbahnzentralamtes in Frankfurt/M. installiert wurde, zur vollen Zufriedenheit der Benutzer seinen Dienst und gibt täglich rund um die Uhr über das Telefon auf Anfrage gezielte Auskünfte über Zugverbindungen im europäischen Eisenbahnnetz.

Unser Szenarium konnte auch nur einen winzigen Teil dessen andeuten, wo heute und in Zukunft mit den Mitteln der Sprachein- und -ausgabe eine effizientere und doch gleichzeitig auch menschengerechtere, humane Arbeitsweise möglich ist. Viele weitere Anwendungen in der Produktion, in der Verwaltung und im privaten Bereich sind mit einiger Phantasie vorstellbar. Manches, was hier erwähnt wurde, ist schon Realität, vieles aber vorläufig noch Zukunftswunsch.

Wo die Technik der automatischen Erkennung und der Synthese von Sprache heute steht, soll nachfolgend dargestellt werden.

2 Die Prinzipien der digitalen Sprachverarbeitung

Sprache — das ist ein Wort, das eine vielschichtige Bedeutung hat. Die Philosophen und Philologen verstehen darunter die Grundvoraussetzung menschlichen

Denkens, für die Technik ist Sprache zunächst einmal ein Signal, das ganz spezifisch mit dem Menschen verknüpft ist. Betrachtet man das Sprachsignal von seiner Entstehung her (Fig. 2), hat man es mit einem in der Regel periodischen Signal zu tun: Die Stimmbänder im menschlichen Kehlkopf schicken als Anregungssignale kurze Luftstösse — im Abstand von etwa 10 Millisekunden — etwa 100mal in der Sekunde in die Artikulationsorgane Rachen-, Mund- und Nasenhöhle. Dort werden diese noch nicht sprachlichen Signale durch die Artikulation, also die Bewegung der Sprechorgane, so geformt, dass daraus Sprache entsteht. Bei stimmlosen Sprachlauten sind die Stimmbänder an der Schallerzeugung nicht beteiligt, sondern es wird lediglich das Geräusch strömender Luft moduliert.

Statt die Schallschwingung zu betrachten, hilft es gerade bei den Problemen der automatischen Spracherkennung weiter, von Laut zu Laut die wechselnde Vielfalt von Schwingungen im Sprachsignal zu verwenden. Ein solches Spektralmuster (Fig. 3) gibt beispielsweise anschaulich Auskunft über die Eigenresonanzen des Mundes, die sogenannten Formanten, die wiederum charakteristisch sind für den jeweils erzeugten Laut. Während also das Zeitsignal als Schwingungssignal gewissermaßen das natürliche Signal ist, das sich als Luftschwingung fortpflanzt oder als elektrische Schwingung vom Telefon übertragen wird, stellt das Sprachspektrum ein abgeleitetes Signal dar, das genau Auskunft gibt über die charakteristische Zusammensetzung der Oberschwingungen des Sprachsignals. Diese Tatsache macht sich auch das menschliche Gehör zunutze, denn auch im Innenohr wird als erster Schritt des Hörvorgangs zunächst eine solche Spektralanalyse vorgenommen.

Beim synthetischen Erzeugen oder dem automatischen Erkennen von Sprachsignalen wird zweckmässigerweise jeweils die eine oder die andere Betrachtungsweise bevorzugt.

3 Synthetische Sprache

Für die Signalsynthese muss ein obertonreiches Impulssignal erzeugt werden. Dieses Impulssignal, das dem Schallsignal aus dem Kehlkopf ähnlich ist, muss anschliessend in einem steuerbaren Filtersystem, entspre-

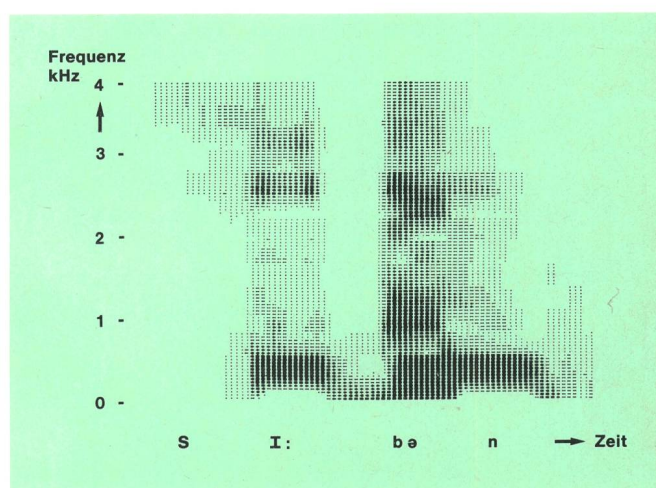


Fig. 3
Digital berechnetes Spektrum der gesprochenen Ziffer «sieben»

chend dem jeweiligen Lautspektrogramm, gefiltert werden. Das Hauptproblem der synthetischen Spracherzeugung liegt also in der richtigen Steuerung dieser Filterfunktion, denn sie hat die Aufgabe, alle komplexen spektralen Vorgänge beim Sprechen möglichst naturgetreu nachzubilden.

Die moderne digitale Filtertechnik bietet vielfältige Möglichkeiten, solche Filtervorgänge zu simulieren. Eine direkte Analogie zu den natürlichen Vorgängen stellt das Wellenfilter dar, bei dem die vor- und rücklaufenden Schallwellen im natürlichen Sprechtrakt direkt nachgebildet werden (Fig. 4). Man denkt sich hierbei die Luftröhre zwischen Kehlkopf und Mund aus einer Anzahl kurzer Röhren von jeweils konstantem Querschnitt zusammengesetzt. Die Filtercharakteristik eines solchen Röhrenmodells wird dann im wesentlichen durch die Reflexion an den Übergangsstellen bestimmt. Diese Wellenvorgänge kann man direkt in einem elektronischen Modell berechnen, das als Wellenfilter bekannt ist. Damit steht ein recht einfaches und elegantes System zum Erzeugen von Sprachsignalen zur Verfügung. Wesentlich und entscheidend für die Art der Sprachsynthese ist die Gewinnung der Steuersignale für diesen «elektronischen Mund». Deshalb wird zwischen halbsynthetischer und vollsynthetischer Sprache unterschieden.

31 Halbsynthetische Sprache

Bei der halbsynthetischen Technik muss der Ansagetext vorher bekannt sein. Es ist dann möglich, aus den entsprechenden von Menschen erzeugten Sprachsignalen mit Hilfe recht komplexer mathematischer Verfahren die optimalen Steuersignale für ein Synthesensystem vorgegebener Struktur zu berechnen. Solche Verfahren werden als «halbsynthetische Verfahren» bezeichnet, da hierbei zwar für das Erzeugen des Sprachsignals ein Synthetisator verwendet wird, Ausgangspunkt für die Gewinnung der Steuerparameter jedoch immer direkt der von Menschen gesprochene Text ist. Durch die Anwendung des Synthetisators lassen sich der Speicher- und der Verarbeitungsaufwand im Steuerrechner um den Faktor 50 reduzieren.

Solche halbsynthetischen Verfahren garantieren höchste Sprachqualität, die bei geschickter Analyse nicht vom Original zu unterscheiden ist. Sprachausgaben, die mit dieser Technik arbeiten, setzen in der Regel aus gewissen Grundelementen, Sätzen, Wörtern oder

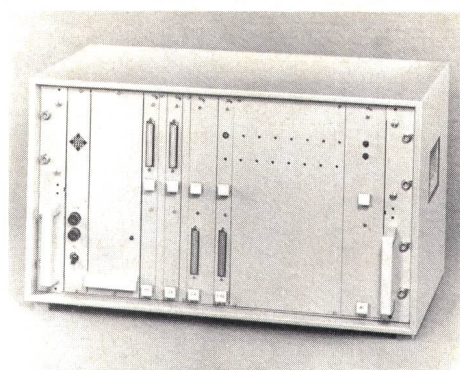


Fig. 5
Halbsynthetisches
Sprachausgabesystem
SPRAUS

Silben ganz unterschiedliche Texte zusammen. Voraussetzung ist jedoch immer, dass die Textvielfalt vor der Generierung der Basiselemente in allen Einzelheiten bekannt ist. Je nach Speicher- und Verarbeitungsaufwand kann auf diese Weise ein Ansagesystem für einen Wecker oder einen Mikrowellenherd oder für ein Fahrplan-Auskunftssystem realisiert werden, bei dem die Namen Hunderter von Zielbahnhöfen und eine Vielzahl verbindender und erläuternder Texte zu erzeugen sind. Das Sprachausgabesystem SPRAUS von AEG-Telefunken stellt ein Beispiel eines solchen sehr fortgeschrittenen halbsynthetischen Systems dar. Es kann entweder in direkter Zusammenarbeit mit einem Rechner arbeiten, der dann gegebenenfalls gleichzeitig die Aufgaben der Informationsbearbeitung (wie beim System «Karlchen») übernimmt, oder es wird als autonomes System mit eigenem Halbleiterspeicher eingesetzt (Fig. 5).

Dieses System kann an unterschiedlichen Schnittstellen betrieben werden. Für Lautsprecherbetrieb ist noch ein Verstärker erforderlich. Das dargestellte Modell kann mit acht Speicherkarten bestückt werden, die jeweils Steuersignale für eine Minute Sprache enthalten. Der Speicherumfang lässt sich durch ein zweites Magazin mit zusätzlichen Speicherkarten fast beliebig erweitern. Neben dem Einkanalmodell lassen sich auch Vielkanalmodelle zusammenstellen, bei denen bis zu 256 Ausgabekanäle simultan auf den gleichen Speicher zugreifen, aber asynchron völlig unterschiedliche Texte produzieren. So werden beispielsweise beim Fahrplan-Auskunftssystem «Karlchen» gleichzeitig zehn Anfragen bearbeitet. Das Sprachausgabesystem SPRAUS wurde so flexibel wie möglich gehalten. Es lässt sich dank seiner Modulbauweise an die unterschiedlichsten Aufgabenstellungen anpassen und bietet dabei gleichzeitig ein Optimum an Sprachqualität, das heisst an Verständlichkeit und Natürlichkeit des erzeugten Signals.

32 Vollsynthese — die Sprachausgabe der Zukunft

Alle gegenwärtig auf dem Markt erhältlichen Sprachausgabesysteme erfordern ein vorheriges Festlegen und Bearbeiten des Vokabulars durch den Menschen. Ebenso muss bei Textänderungen manuell das neue Vokabular erarbeitet werden. Diese Systeme können also bis jetzt Sprache noch nicht vollautomatisch aus geschriebenem Text erzeugen. Das wäre aber der Wunschtraum einer Sprachausgabe, denn dann könnte man fast jede Information, die in einem Computer vorhanden ist, mit Sprache ausgeben. Eine solche Ma-

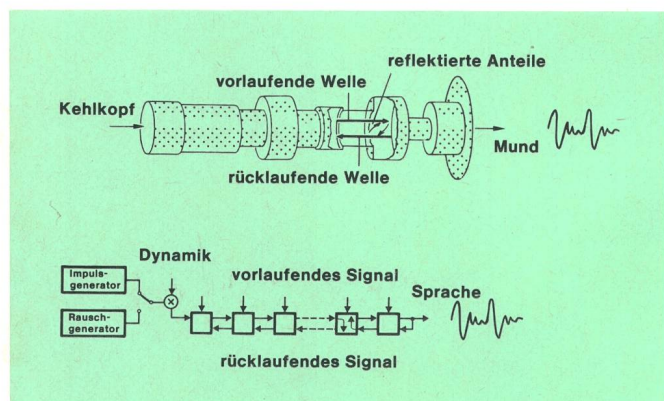


Fig. 4
Röhrenmodell und digitales Wellenfilter

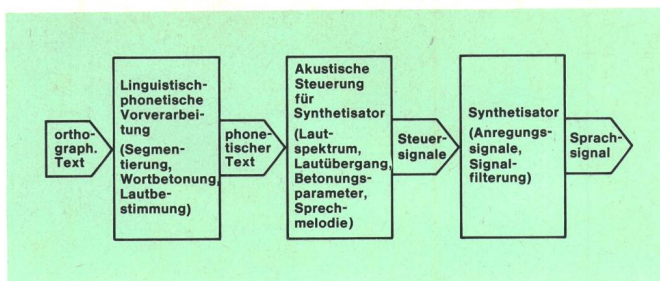


Fig. 6
Prinzip der Vollsynthese von Sprachsignalen

schine entspräche einem Vorleser, der vom Inhalt des Vorgelesenen keine Ahnung hat. Das Sprachsignal muss gut verständlich sein und soll darüber hinaus noch möglichst natürlich wirken. Um dieses Ziel zu erreichen, müssen linguistische und phonetische Regeln erarbeitet werden, die Hinweise geben auf die Aussprache in einer bestimmten Sprache. *Figur 6* zeigt die drei wesentlichen Stufen einer solchen Vollsynthese, bei der in einer linguistisch-phonetischen Vorverarbeitung eine Transkription aus dem orthographischen in den phonetischen Bereich vorgenommen wird und in der anschließenden zweiten Stufe aus den phonetischen Parametern die eigentlichen Steuerparameter für den nachfolgenden Sprachsynthetisator ermittelt werden. Für alle diese Verarbeitungsstufen lassen sich Regeln erarbeiten, die recht gut das beschreiben, was der Mensch beim Lernen einer Sprache im Gedächtnis behalten hat oder einfach intuitiv aus dem Sprachgefühl heraus richtig macht.

Welche Probleme dabei auftreten, zeigen einige Beispiele aus dem Bereich der Betonungsregeln. So kann man aus einem «Nacht-Eilzug» durch falsche Betonung leicht einen «Nachteil-Zug», aus einem «Hühner-Ei» eine «Hühnerei» oder aus «Erst-Ehen» etwas «erstehen» lassen. Dabei ist das, was Betonung genannt wird, ein vielschichtiges Ergebnis der Steuerung von Lautstärke, Lautlänge und Lauthöhe. Welche Kombination dieser drei Parameter im vorliegenden Fall zu einer auch vom Sinn her semantisch richtigen Betonung führt, ist gegenwärtig noch Gegenstand der Forschungen. Wo diese Problematik vermutlich in naher Zukunft auf Grenzen stossen wird, ist leicht zu erkennen, dann nämlich, wenn man sich die zahlreichen, meist semantisch begründeten Möglichkeiten der Wortbetonung in einem einfachen Satz wie «Ich spreche heute zu Ihnen über Sprachsynthese» vorstellt. Derartige Feinheiten des Ausdrucks werden wohl auch in Zukunft dem Menschen und seiner willkürlichen Intention vorbehalten bleiben.

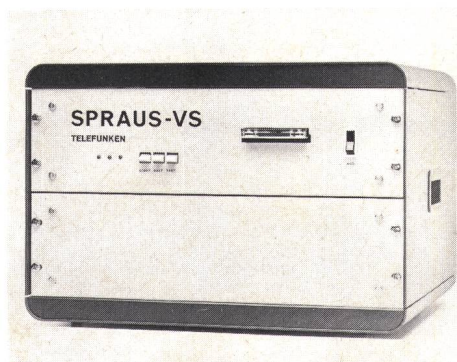


Fig. 7
Vollsynthetisches Sprachausgabesystem SPRAUS-VS

Die Vollsynthese muss sich einstweilen noch mit den viel primitiveren Fragen der zweckmässigen Gestaltung der Lautübergänge in der zweiten Stufe von *Figur 5* plagen. Im *Ulmer Forschungsinstitut* existiert schon ein recht perfekt arbeitender Sprachsynthetisator (*Fig. 7*), aber über die eigentlichen Bewegungsgesetze des menschlichen Sprachtraktes besteht nur ein lückenhaftes Wissen. Damit Sprache aber genügend natürlich klingt, müssen auch die Übergänge zwischen den einzelnen Lauten hinreichend natürlich sein. Auf diesem Gebiet wurden zwar im Laufe der letzten Jahre beachtliche Fortschritte erzielt, doch haben eben diese Ergebnisse leider auch gezeigt, dass noch viele Einzelprobleme gelöst werden müssen, bis eine perfekte, vollsynthetische Sprachausgabe machbar ist.

4 Automatische Spracherkennung

Nach diesen ausführlichen Betrachtungen zu den Techniken und Möglichkeiten der automatischen Sprachausgabe drängt sich die Frage nach Lösungsmöglichkeiten für die umgekehrte Richtung, die automatische Spracherkennung, auf. Grundlage und Ausgang einer solchen Erkennung ist das Signalspektrum, wie es etwa in *Figur 3* dargestellt ist. Ein solches Muster, das nach dem heutigen Wissensstand gewisse Ähnlichkeit mit den beim menschlichen Hörprozess verarbeiteten Mustern hat, gibt die Möglichkeit, die zahlreichen Normierungen einfacher als am Zeitsignal (*Fig. 2*) vorzunehm-

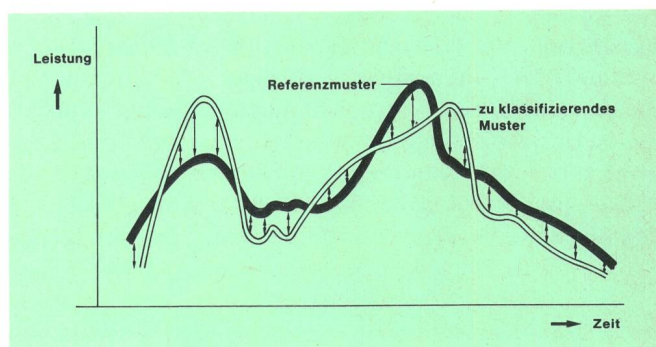


Fig. 8
Klassifikation durch Abstandsmessung bei einem eindimensionalen Muster

men. Solche Spektralmuster lassen sich als fortlaufende Muster für einzelne Laute oder Wörter bilden, je nachdem, ob Laute oder Wörter erkannt werden sollen. Natürlich liesse sich aus der Folge von Lauten jedes Wort erkennen, aber bereits ein Blick auf *Figur 3* zeigt deutlich, dass die Trennung der einzelnen Laute voneinander grosse Schwierigkeiten bereitet. Aus diesem Grunde sind die technisch bedeutsamen Erkennungssysteme bisher immer als Ganzworterkenner realisiert.

Die Aufgabe der Erkennung besteht nun in einem Vergleich zwischen bereits bekannten, gelernten Referenzmustern und den neuen, noch zu klassifizierenden Mustern. Dieser Vergleich geschieht mit Hilfe mehr oder weniger raffinierter mathematischer Methoden, die den Abstand zwischen den bekannten Referenzmustern, den Klassifikatoren, und den zu erkennenden Mustern ermitteln. Als erkanntes Muster wird jenes ausgegeben, zu dem dieser Abstand minimal ist und gleichzeitig ein vorgegebenes Höchstmass nicht überschreitet (*Fig. 8*).

Die Klassifikation stellt einen zwar wichtigen, aber doch nur kleinen Teil der gesamten Signalerkennung dar. Schon die Entscheidung, ob etwa Einzellaute, isoliert gesprochene Befehlswörter oder gar zusammenhängende Sprache erkannt werden soll, kann eine Fülle von weiteren Problemen aufwerfen. Dazu zählt auch die zweckmässige Normierung der Sprechlautstärke und der Sprechgeschwindigkeit. Es ist keineswegs so, dass eine erhöhte Sprechlautstärke einfach einem proportionalen Anheben aller Lautstärkepegel entspricht oder dass eine schnellere Sprechgeschwindigkeit durch proportionales Dehnen rückgängig gemacht werden könnte. Alle diese Parameter ändern sich stark nicht-linear. Wegen der grossen Variationen in der Sprechweise müssen Spracherkennungssysteme stets eine gute Adaptionsfähigkeit aufweisen. Das gilt natürlich ganz besonders dann, wenn verschiedene oder gar beliebige Sprecher erkannt werden sollen. Die am meisten fortgeschrittenen Systeme adaptieren sich während des Betriebs an die stets wechselnde Sprechweise des Benutzers.

Bisher praktisch einsetzbare Spracherkennungssysteme können mit wenigen Ausnahmen lediglich isoliert gesprochene Wörter zuverlässig erkennen. Bisher nur wenige, aber schon in naher Zukunft immer mehr Systeme werden auch verbunden gesprochene Folgen von Einzelwörtern gut erkennen. Damit passt sich die Erkennung besser der natürlichen Sprechweise an. Die im praktischen Einsatz häufig vorkommenden Folgen von Einzelziffern können so sehr flott gesprochen werden. Auch einfache Sätze, die freilich eine wohldefinierte Syntax besitzen müssen, lassen sich so erkennen. Das Problem der Erkennung beliebiger, kontinuierlicher Sprachsignale ist damit freilich noch lange nicht gelöst. Hierzu ist ein hierarchisch gestaffelter Ablauf des Erkennungsprozesses nötig (Fig. 9). Dieser Ablauf beginnt zunächst ebenso wie einfachere Erkennungstechniken mit einem Ermitteln typischer Messdaten, beispielsweise

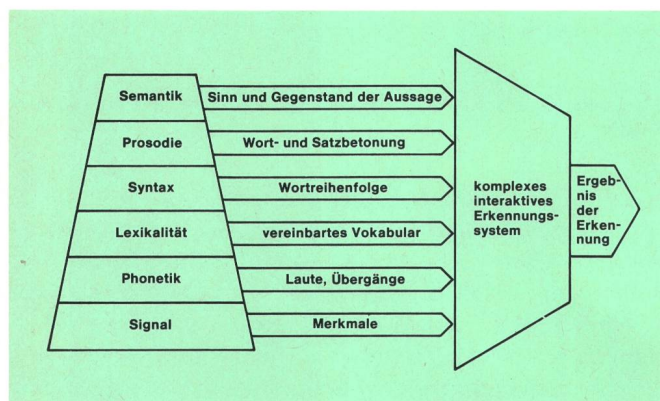


Fig. 9
Hierarchische Analyse der Sprache

Spektrum, Tonhöhe und Lautstärke. Daraus lassen sich Lautkategorien definieren. Mit Hilfe lexikalischen Wissens können in der nächsten Stufe Wörter gebildet werden. Schliesslich gibt die Syntax, möglicherweise unter Zuhilfenahme der prosodischen¹ Parameter wie etwa der Betonung, Hilfestellung, ganze Sätze zu klassifizie-

¹ Prosodie = Lehre von der metrisch-rhythmischen Behandlung der Sprache



Fig. 10
Einsatz des akustischen Datenerfassungssystems ADES bei der Aufnahme komplexer Datenlisten mit akustischer Rückmeldung und Anzeige auf dem Display

ren. Alle diese Stufen können interaktiv zur Verbesserung, in ungünstigen Fällen aber auch zur Verschlechterung des Erkennungsergebnisses beitragen. Man denke nur daran, wie auch der Mensch aus ein paar Lauten ein ganzes Wort subjektiv rekonstruieren kann, wie aber auch bei nicht genauem Hinhören Fehlinterpretationen auftreten können.

Schon aus diesen wenigen Beispielen lässt sich bereits deutlich erkennen, dass das Problem der automatischen Spracherkennung sehr vielschichtig anzugehen ist. Zu dieser Vielschichtigkeit gehört auch das Einbetten der Spracherkennung in eine ganz bestimmte Problemlösung. Die technische Lösung des Problems kann also nie unabhängig von der speziellen Aufgabenstellung gesehen werden.

5 Der Dialog mit dem Computer

Welche technische Ausstattung dazu nötig werden kann, soll ein kurzer Überblick über das akustische Datenerfassungssystem ADES von AEG-Telefunken deutlich machen (Fig. 10). Es kann im wesentlichen folgende Aufgaben übernehmen:

- Erkennen und Erfassen isoliert gesprochener Eingabewörter
- Ausgabe von Sprachsignalen zur Rückmeldung oder als Führung des Benutzers
- Vorauswertung und Speichern der eingegebenen Information zum weiteren Auswerten in einem Hintergrundrechner
- direkte Steuerung peripherer Systeme, etwa als Maschinensteuerung

Damit ist ADES ein System, das recht universell bei den verschiedensten Anwendungen der Datenerfassung eingesetzt werden kann. Lediglich durch Ändern der Software lassen sich unterschiedliche Anwendungsbereiche erschliessen. Durch zwei eingebaute Plattenlaufwerke ist auch das Wechseln der Software im Betrieb leicht möglich. Gleiches gilt für den Wechsel des Vokabulars oder das Anpassen an unterschiedliche Sprecher

oder Ausgabesysteme. Das System kann bis zu 600 Wörter zur Erkennung speichern. Freilich ist es nicht sinnvoll, aus 600 Wörtern in einem Erkennungsschritt zu klassifizieren. Praktische Anwendungsfälle werden in der Regel auch mit einem viel kleineren Wortvorrat auskommen. Bei grossen Vokabularen empfiehlt es sich, aus Gründen der Erkennungssicherheit Teilvokabularen zu definieren und diese dann syntaktisch zu ordnen, so dass jeweils nur ein kleiner Teil des Wortvorrats benutzt wird.

Wesentlich für den praktischen Einsatz ist darüber hinaus auch noch, dass eine zuverlässige Erkennung auch bei hohem Geräuschpegel, etwa in Fabrikhallen, möglich ist. Hierfür ist eine Sprechgarnitur für Nahbesprechung oder ein geräuschkompensierendes Mikrofon vorgesehen. In Zukunft werden hier adaptive Methoden zur Geräuschunterdrückung wichtig werden, da in zahlreichen Anwendungen das Mikrofon verhältnismässig weit vom Benutzer entfernt ist. Um schliesslich dem Benutzer eine völlig uneingeschränkte Bewegungsmöglichkeit zu geben, ist die Verwendung eines drahtlosen Mikrofons, falls erforderlich mit einer optischen Grossbildanzeige als Rückmeldung, vorgesehen.

6 Wie geht es weiter?

Bereits an dieser Skizzierung des Anwendungsbereiches von ADES wird deutlich, dass allein die technische Lösung des Problems, Sprache automatisch zu erzeugen oder zu erkennen, nicht genügt. Diese neuartigen Möglichkeiten greifen so tief in den Ablauf wohlvertrauter Prozesse ein, dass vielfältige ergonomische Fragen mit zu berücksichtigen sind. So haben sich beispielsweise mit dem ungeheuren Anwachsen der Zahl der Bildschirmterminals in den letzten Jahren ihre Gestalt und Technik stark gewandelt. Während aber die Schnittstelle zwischen Terminal und Rechner dank wesentlich verbesserter Übertragungsprotokolle fast perfekt ist, hat sich die Schnittstelle zwischen Mensch und

Terminal hier nur andeutungsweise geändert. Freilich wurden Anzeigeeinstrumente umkonstruiert — vielleicht zeigen sie heute Ziffern an, Tastaturen wurden bedienerfreundlicher gestaltet, und integrierte Darstellungen für Prozessabläufe gehören heute zur üblichen Technik. Aber gleichzeitig zeigen die zunehmenden Klagen über die physische und die psychische Belastung durch die Arbeit an neuen Arbeitsplätzen, dass die Ergonomie hier noch gewaltige Aufgaben zu lösen hat. Hier kann die Sprache auf weiten Gebieten den Informationsaustausch zwischen Mensch und Maschine vereinfachen und menschenwürdiger gestalten.

Die technischen Möglichkeiten sind heute bereits soweit fortgeschritten, dass viele Dinge am Arbeitsplatz und im Privatleben mit Sprachsignalen gesteuert werden könnten. ADES kann heute beispielsweise für die Qualitätsdatenerfassung in der Fertigung eingesetzt werden. «Karlchen» als automatisiertes Fahrplan-Auskunftssystem hat sich bewährt. Hier wird in Zukunft auch eine Spracheingabe möglich sein. Schon in Kürze werden die ersten Autos mit Sprachausgabe in Europa gekauft werden können. Ein grosser Teil der Anwendungen im Konsumbereich scheitert bisher noch daran, dass die technologische Integration der komplexen Verarbeitungsalgorithmen noch nicht genügend fortgeschritten ist. Hier wird in den nächsten Jahren eine rege Entwicklungstätigkeit erfolgen, die zu einer Vielzahl neuer Produkte führen wird. Aber auch die Erkennung beliebiger verbaler kontinuierlicher Texte wird intensiv erforscht werden. Bereits für Mitte der achtziger Jahre sind erste Prototypen sprachgesteuerter Schreibsysteme angekündigt. Die Ausnutzung der Möglichkeiten der Sprachein- und -ausgabe wird helfen, die Scheu zu überwinden, die bisher vor der Anwendung der Datenverarbeitung bestand.

Adressen der Autoren: Dipl.-Ing. H. Mangold, Leiter der Abteilung Sprachsignalverarbeitung und Videotechnik, Forschungsinstitut Ulm, D-7900 Ulm/Donau. Dr.-Ing. K.-D. Schenkel, Leiter der Entwicklung im Fachbereich Trägerfrequenztechnik, c/o AEG-Telefunken, D-7150 Backnang.

Die nächste Nummer bringt unter anderem

Vous pourrez lire dans le prochain numéro

2/82

B. Szentkuti	Compatibilité électromagnétique — CEM
Chr. Bärffuss	Les PTT et la protection contre les perturbations électromagnétiques
R. Bersier	Les points essentiels de la nouvelle ordonnance sur la protection contre les perturbations électromagnétiques: Aspects techniques
M. Schaeren	Das Pilotnetz Telepac Le réseau pilote Télépac
P.-A. Probst, P. Vörös	Synchronisierung digitaler Netze: Synchroner Betrieb (1. Teil) Synchronisation des réseaux numériques: Exploitation synchrone (1 ^{re} partie)
