

<b>Zeitschrift:</b>	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 telegraфи svizzeri
<b>Herausgeber:</b>	Schweizerische Post-, Telefon- und Telegrafenbetriebe
<b>Band:</b>	72 (1994)
<b>Heft:</b>	5
<b>Artikel:</b>	Automatische Rufnummernauskunft
<b>Autor:</b>	Pfister, Beat / Schaub, Arthur
<b>DOI:</b>	<a href="https://doi.org/10.5169/seals-874710">https://doi.org/10.5169/seals-874710</a>

### 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:** 11.01.2026

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

# Automatische Rufnummernauskunft\*

Beat PFISTER, Zürich, und Arthur SCHAUB, Hombrechtikon

## 1 Einleitung

Sprachgesteuerte automatische Auskunftssysteme, wie sie in diesem Artikel behandelt werden, kommunizieren mit dem Benutzer über das Medium der gesprochenen Sprache. Ziel dieses Dialogs ist es, den Informationswunsch des Benutzers zu eruieren und anschliessend die gewünschte Information abzugeben.

Ein solches Auskunftssystem braucht neben den Fähigkeiten, die gesprochene Sprache zu verstehen (Spracherkennung) und Sprache zu erzeugen (Sprachsynthese) auch Zugriff auf die verlangten Informationen wie Wetterprognosen, Telefonnummern, Aktienkurse, Fahrplaninformationen usw.

Die Telecom PTT ist verständlicherweise besonders an einem System interessiert, das automatisch über Telefonnummern von Abonnenten Auskunft geben kann. Sie hat deshalb 1991 in Zusammenarbeit mit der *Eidgenössischen Technischen Hochschule (ETH) Zürich* und der *Ascom Tech AG* ein Gemeinschaftsprojekt mit dem Ziel begonnen, einen ersten Schritt in Richtung einer Automatischen Rufnummernauskunft (ARA) zu vollziehen.

## 2 Die ideale Automatische Rufnummernauskunft

Ein ARA-System müsste idealerweise ein vollwertiger Ersatz für eine bediente Auskunft sein. Das heisst vor allem, dass es jedem Benutzer ohne Vorwissen möglich sein sollte, dem System mit Redewendungen mitzuteilen, welche Information gewünscht wird. Das System, das Zugriff auf eine geeignete Datenbank hat, würde die benötigte Information suchen und diese schliesslich als natürlichsprachliche Äusserung dem Benutzer mitteilen.

Zusätzlich wird von einem automatischen System erwartet, dass es schnell arbeitet, nie überlastet ist, keine Fehler macht, stets über die neuesten Informationen verfügt und zudem für den Benutzer günstig und für den Systembetreiber einträglich ist.

Ein automatisches Auskunftssystem, das all diese Erwartungen erfüllt, bleibt jedoch vorläufig noch eine Wunschvorstellung. Eine ganze Reihe verschiedenster Probleme stehen heute noch der Verwirklichung im Wege. Viele davon haben damit zu tun, dass die natürliche Sprache, im Gegensatz zu formalen Sprachen (z. B.

Programmiersprachen), äusserst schwierig zu handhaben ist:

- Die Forschung auf dem Gebiet der Spracherkennung hat bei weitem noch nicht den erforderlichen Stand erreicht. Ein System, das den Spracherkennungsfähigkeiten des Menschen nahekommt, das also gesprochene Sprache in die entsprechende textliche Form umsetzt, und zwar für einen uneingeschränkten Wortschatz und für beliebige Satzkonstruktionen oder sogar für Dialekt, liegt noch in ferner Zukunft.
- Für ein Auskunftssystem genügt es aber nicht, dass die gesprochene Anfrage eines Benutzers in den entsprechenden Text umgewandelt wird. Das System muss verstehen, was der Benutzer will, und muss wissen, wie das Informationsbedürfnis unter Einsatz der verfügbaren Datenbanken im konkreten Fall zu befriedigen ist. Die Analyse der Bedeutung natürlichsprachlicher Ausdrücke ist heute nur für einfache Anwendungen, nur mit stark eingeschränktem Vokabular und für simple Formulierungen machbar.
- Bei jeder Kommunikation mit natürlicher, gesprochener Sprache entstehen Verständigungsprobleme, die darin begründet sein können, dass entweder etwas nicht gehört oder nicht verstanden worden ist, oder dass etwas im vorliegenden Zusammenhang keinen Sinn ergibt. Für eine erfolgreiche Kommunikation ist es erforderlich, dass die Maschine diese Situationen erkennen und darauf richtig reagieren kann. Zudem muss es für den Benutzer jederzeit möglich sein, der Maschine mitzuteilen, wenn er etwas nicht verstanden hat.
- Nebst diesen in der natürlichen Sprache begründeten Schwierigkeiten sind auch Probleme zu lösen, die mit dem Wissen des täglichen Lebens zusammenhängen. So ist es beispielsweise für gewisse Auskünfte nötig, dass die Maschine mit zeitlichen und geografischen Aspekten umgehen kann, also weiss, was früher, vorher, später, nahe usw. bedeutet.

Die Lösung dieser und weiterer Probleme ist Voraussetzung für ein automatisches Auskunftssystem, das eine bediente Auskunft vollwertig ersetzen könnte. Es ist hier jedoch weder nötig noch sinnvoll, nach dem «Alles-oder-nichts»-Prinzip vorzugehen. Einfachere automatische Auskunftssysteme sind durchaus machbar und stehen im Ausland auch bereits im Einsatz. Sie verlangen allerdings, dass sich der Benutzer drastisch den Möglichkeiten der Maschine anpasst. Zur Hauptsache be-

\* Gemeinschaftsprojekt der Eidgenössischen Technischen Hochschule Zürich, der Ascom Tech AG und der Telecom PTT

trifft dies die Art, wie zur Maschine gesprochen werden muss, oder anders ausgedrückt, welche Möglichkeiten die eingesetzte Spracherkennung bietet.

Die Leistung heute verfügbarer Spracherkennungssysteme ist im Vergleich zur oben formulierten Wunschvorstellung ziemlich bescheiden. Sprecherunabhängige Systeme zur Erkennung einzeln gesprochener Wörter aus einem bis zu mehrere hundert Termini umfassenden Vokabular können heute als anwendungsreif gelten. Bei der Erkennung kontinuierlich gesprochener Sprache, wie sie ja für ein Wunschsystem verfügbar sein sollte, beschränkt sich der Erfolg hingegen auf Laborversuche mit stark eingeschränktem Vokabular.

Im folgenden wird dargelegt, wie mit derartigen Einschränkungen im Rahmen des ARA-Projekts ein erster Schritt in Richtung eines anspruchsvollen automatischen Auskunftssystems vollzogen wird.

### 3 Das Ziel des laufenden ARA-Projekts

Von einem ARA-System kann man die Rufnummer eines Abonnenten erhalten, indem sein Name und seine Adresse eingegeben werden. Umgekehrt können auch unter Eingabe der Rufnummer Name und Adresse eines Abonnenten abgefragt werden. Mit den Elektronischen Teilnehmerverzeichnissen (ETV) der Telecom PTT ist dies über eine Bildschirmkonsole verhältnismässig einfach möglich. Statt über Tastatur und Bildschirm wäre es für viele Leute jedoch nützlicher, mit gesprochener Sprache über das normale Telefon mit dem Verzeichnis zu kommunizieren.

Dass dieses Vorhaben nicht so einfach zu verwirklichen ist, liegt unter anderem daran, dass der Wortschatz für diese Kommunikation nicht beschränkt ist, bedingt durch die Eigennamen (Namen und Vornamen der Abonnenten, Strassen- und Ortsnamen). Der Einsatz eines konventionellen Wörterkenners ist deshalb nicht möglich, um dem System die Daten eines Abonnenten einzugeben, von dem die Rufnummer gewünscht wird.

Für die erste Version eines ARA-Systems (als ARA-V1 bezeichnet), das mit einem sprecherunabhängigen Wörterkennner arbeitet, ist deshalb von der Annahme ausgegangen worden, dass der Benutzer die Eigennamen nicht zusammenhängend ausspricht, sondern mit der offiziellen Buchstabiertabelle, die in jedem Telefonbuch zu finden ist, buchstabiert. Der für das ARA-System nötige Wortschatz konnte dadurch drastisch auf die Wörter der Buchstabiertabelle, die Ziffern und einige Kommando- bzw. Dialogwörter verkleinert werden.

Damit ist selbstverständlich das System weit davon entfernt, sich einem Benutzer als vollwertiger Ersatz für eine bediente Auskunft anzubieten. Trotzdem ist es überaus nützlich, diesen ersten Schritt zu vollziehen. Viele Fragen der Mensch-Maschine-Kommunikation und das Zusammenwirken mit den Elektronischen Teilnehmerverzeichnissen können mit dieser stark vereinfachten Anordnung mindestens ebensogut studiert werden wie mit einem sehr viel komplexeren System. Zudem wird es auch in einem ARA-System, das mit einem viel weiter fortgeschrittenen Spracherkennner ausgestattet ist, immer wieder vorkommen, dass gewisse Eigennamen ganz oder teilweise buchstabiert werden müssen,

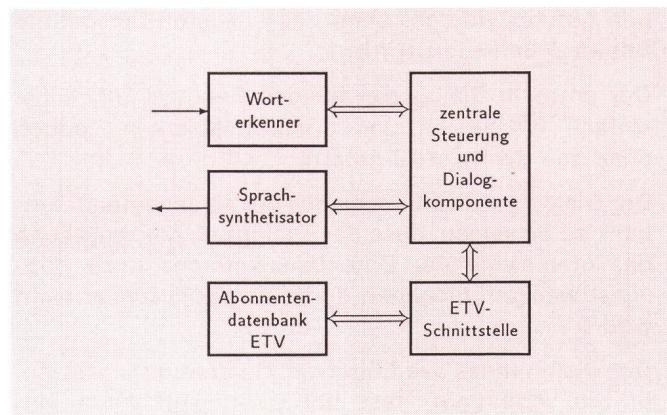


Fig. 1 Blockscheme des ARA-Systems  
ETV, Elektronische Teilnehmerverzeichnisse

sei es, weil verschiedene geschriebene Namen gleich ausgesprochen werden (z. B. Maier, Meier, Meyer), oder sei es, weil die Verständlichkeit beeinträchtigt ist.

Ein grobes Blockschaubild von ARA-V1 ist in *Figur 1* zu sehen. Die vier Hauptteile des Systems sind:

#### Wörterkennung

Einheit zur sprecherunabhängigen Erkennung einzeln gesprochener Wörter (siehe Kap. 7).

#### Sprachsynthesizer

Hier wird das an der ETH Zürich entwickelte Sprachsynthesesystem SVOX eingesetzt, mit dem das ARA-System dem Benutzer Anweisungen geben, Rückfragen stellen oder Antworten geben kann (siehe Kap. 5).

#### Datenbankanfragekomponente

Intelligente Schnittstelle zum Elektronischen Teilnehmerverzeichnis der Telecom PTT (siehe Kap. 6).

#### Zentrale Steuerung

Flexibler Dialog- und Entscheidungsprozess, der unter Einbezug allen verfügbaren Wissens mit dem Benutzer mit Hilfe von Spracherkennung und Sprachsynthese einen zielgerichteten Dialog führt.

Mit dieser Systemkonfiguration lässt sich der Dialog mit dem Benutzer frei gestalten, selbstverständlich unter Berücksichtigung der Randbedingung, dass der Spracherkennung erstens nur einzeln gesprochene Wörter erkennen kann und zweitens je nach Sprecher mehr oder weniger oft Erkennungsfehler macht. Im Konzept des Dialogs von ARA-V1 werden deshalb hauptsächlich folgende Gesichtspunkte berücksichtigt:

- Das ARA-System übernimmt von Anfang an die Dialogführung, um zu verhindern, dass der Benutzer frei zu sprechen beginnt, statt die erlaubten Dialogwörter zu verwenden. Zusätzlich sind Fragen und Anweisungen so zu formulieren, dass der Benutzer nicht dazu verleitet wird, mit einem Satz statt mit einem Wort zu antworten.
- Das System muss ohne jede Vorinformation benutzbar sein. So muss einerseits der Neuling mit Erklärungen

gen geführt werden, ohne dass dadurch der geübte Benutzer behindert wird.

- Der normale Dialog soll für den Benutzer möglichst einfach erscheinen und speditiv ablaufen, jedoch ohne den Benutzer zu hetzen.
- Der Dialog soll das Wissen aus der Abonentendatenbank so einsetzen, dass das Dialogziel, die Ermittlung des interessierenden Datenbankeintrages, nach möglichst wenigen Eingaben durch den Benutzer erreicht wird.
- Der Dialog muss das Mittel der Quittierung einsetzen, um zu verhindern, dass bei Erkennungsfehlern der Dialog scheitert. Quittierung und Korrektur müssen jedoch so gestaltet sein, dass der Benutzer beim Buchstabieren oder bei der Eingabe einer Folge von Ziffern nicht die Übersicht verliert.

Etliche dieser Gesichtspunkte haben sich bei der Entwicklung als widersprüchlich herausgestellt (z. B. verzögert jedes Quittieren den Dialog). Sie können deshalb nur bedingt berücksichtigt werden, und es ist stets eine Frage des Ermessens, welcher Gesichtspunkt an welcher Stelle des Dialogs wie stark zum Tragen kommen soll.

Klarheit über Fragen der Dialoggestaltung und über bisher noch gar nicht entdeckte Probleme, die durch unvorhergesehenes Verhalten des Benutzers hervorgerufen werden können, ist letztlich nur im nächstens durchzuführenden Versuchsbetrieb zu erhalten.

#### 4 Die Weiterentwicklung des ARA-Systems

Der Grund dafür, dass der Benutzer der ersten Version des Systems zum Buchstabieren mit der Buchstabiertabelle gezwungen werden muss, ist hauptsächlich beim Worterkenner zu suchen. Nur so ist es möglich, dass der Benutzer eine Folge von Buchstaben eingeben kann, die mit einiger Sicherheit vom ARA-System richtig verstanden wird.

Erheblich komfortabler wäre selbstverständlich das normale Buchstabieren (ohne Buchstabiertabelle). Da sich aber gesprochene Buchstaben am Telefon nur wenig zuverlässig erkennen lassen, muss zur Verbesserung der Erkennung zusätzliche Information angewandt werden. So kann beispielsweise ausgenutzt werden, dass die eingegebene Buchstabenfolge einem Ortsnamen entsprechen muss. Dies kann mit einer Optimierung, die die Folge der Eingaben unter Berücksichtigung der Verwechslungsmatrix des Erkenners dem passendsten Namen zuordnet, relativ einfach gelöst werden. Dazu sind allerdings eine Ortsliste oder Listen mit Abonenten- bzw. Strassennamen je Ort usw. erforderlich. Mit der gegenwärtig verfügbaren ETV-Schnittstelle lässt sich dies jedoch nicht machen.

Ein wesentlicher Schritt zur Steigerung des Benutzerkomforts eines ARA-Systems kann mit einem vokabular-unabhängigen Worterkenner erreicht werden, weil dieser vom Prinzip her normal ausgesprochene Wörter aus einem beliebig grossen Vokabular erkennen kann. Dafür müssen jedoch nicht nur die bereits erwähnten Listen verfügbar sein, sondern auch für jeden Eigennamen die gebräuchlichen Aussprachen.

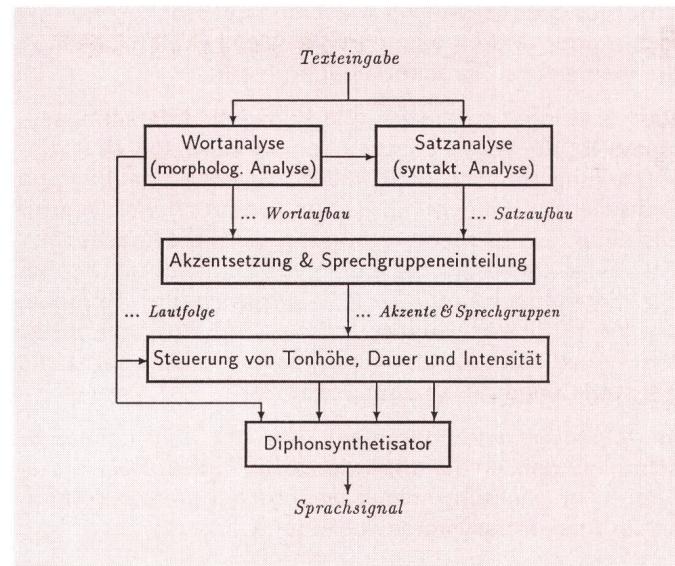


Fig. 2 Blockscheme des Sprachsynthesesystems SVOX

Mit diesem Schritt beim Worterkenner und mit einem optimal angepassten Dialog dürfte der Punkt erreicht werden können, wo ein ARA-System breite Akzeptanz findet und somit auch häufig benutzt wird. Die Frage, wieso als Ziel des Projekts nicht von Anfang an das direkte Erreichen dieses Punktes angestrebt worden ist, lässt sich wie folgt beantworten: Zum einen ist der Schritt von null zu diesem Punkt zu gross, d. h. mit zu vielen Unbekannten behaftet, um seriös beurteilt werden zu können, und zum andern wird das Buchstabieren mit und ohne Tabelle auch im einem künftigen System zur Anwendung kommen, vergleichbar mit der bedienten Auskunft, wo zur Verständigung sehr oft und spontan buchstabiert wird.

#### 5 Sprachsynthesesystem SVOX

Das System SVOX (Fig. 2) erzeugt für jeden deutschen Text satzweise ein entsprechendes Sprachsignal, gehört also zu den Text-zu-Sprache-Systemen (Text-to-Speech Systems). Es besteht aus zwei Hauptteilen, dem *Transkriptionsteil* und der *phono-akustischen Stufe*.

Die Transkription setzt die orthographische Eingabe in eine phonologische Repräsentation (abstrakte Darstellung des zu erzeugenden Sprachsignals) um, die eine dem Text entsprechende Lautfolge, die Wort- und Satzakzente und die Sprechgruppengrenzen umfasst. Dafür werden zuerst die Wörter eines Satzes morphologisch analysiert (Bestimmung von Wortart und Wortform) und die phonetische Umschrift ermittelt. Das Ergebnis der morphologischen Analyse wird verwendet, um unter Anwendung einer Satzgrammatik den syntaktischen Aufbau des Satzes zu ermitteln. Der Satzaufbau dient dazu, die Einteilung in Sprechgruppen (Phrasen) vorzunehmen und die relativen Akzente der Wörter zu bestimmen.

Die phono-akustische Stufe bestimmt aufgrund dieser abstrakten Darstellung die Dauer der Laute, den Tonhöhen- und den Intensitätsverlauf des zu synthetisierenden Sprachsignals, was als Prosodiesteuerung bezeichnet wird. Schliesslich verändert der Diphonsynthetisator die natürlichsprachlichen Lautübergangselemente (Diphone) so, dass sie die verlangte Dauer, Grundfrequenz

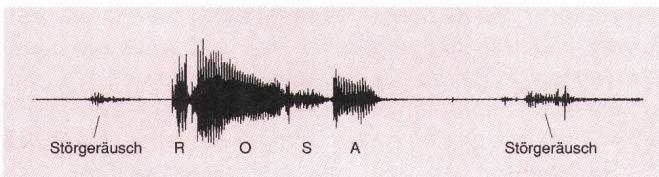


Fig. 3 Sprachsignal des Wortes «Rosa»

und Intensität aufweisen, und verknüpft sie zum gewünschten Ausgangssignal (Konkatenationssynthese).

## 6 ETV-Schnittstelle

Die Schnittstelle zu den Elektronischen Teilnehmerverzeichnissen hat hauptsächlich die Aufgabe, die Anfragen der Dialogsteuerung in ETV-kompatible Eingabefolgen (Strings) umzuwandeln und die Ausgabefolgen in eine für die Dialogsteuerung brauchbare Form zu bringen. Dies ist mit dem heutigen ETV nur sehr beschränkt möglich, weil die Ausgaben nicht so strukturiert sind, dass sie ohne eine semantische Interpretation verarbeitet werden können. Ein derartiger Interpreter ist grundsätzlich machbar, aber nur mit sehr grossem Aufwand.

Für die erste Version des ARA-Systems ist nur eine einfache Ausführung des Interpreters entwickelt worden, in der Annahme, dass die Elektronischen Teilnehmerverzeichnisse der Telecom PTT in der näheren Zukunft ohnehin besser strukturiert werden.

## 7 Wörterkennner

Die Aufgabe und Schwierigkeiten der Spracherkennung werden anhand der *Figuren 3* und *4* erläutert. In *Figur 3* ist ein Sprachsignal von zwei Sekunden Dauer dargestellt. Der Sprecher spricht den Namen «Rosa» aus. Vor und nach dem Namen sind für Telefonverbindungen typische Störgeräusche sichtbar. Aus *Figur 4* sind die Spektren von je 30 ms langen Zeitsegmenten der vier Laute des Namens «Rosa» ersichtlich.

Die Auswertung von Merkmalen des Sprachsignals – im wesentlichen der spektralen Verteilung der Signalleistung sowie zeitlicher Ableitungen dieser Grössen – ermöglichen eine automatische Spracherkennung. Zur Verarbeitung solcher Sprachmerkmale sind Verfahren mit «versteckten Markov-Modellen» (Hidden Markov Models) [1] und «neuronalen Netzen» [2] bekannt. Obwohl es sich dabei grundsätzlich um klassische Verfahren handelt, sind zum Gelingen einer praktischen An-

wendung umfangreiche Optimierungen hinsichtlich Struktur und Dimensionierung erforderlich.

Im Gegensatz zur Trainings- und ersten Testphase, bei denen in der Regel manuell markierte bzw. korrigierte Wortanfänge und -enden verwendet werden, muss ein Wörterkennner im praktischen Einsatz die einzelnen Wörter aus dem fortlaufenden Signal heraus erkennen. Dies ist besonders bei Telefonverbindungen schwierig, weil zu ungünstigem Sprecherverhalten, wie Einatmen oder Schnalzgeräuschen vor Wörtern, noch Leitungsrauschen und zusätzliche Störgeräusche hinzukommen.

Im Rahmen des Projekts ARA sind bei Ascom Tech AG verschiedene Arbeiten zur Spracherkennung durchgeführt worden, nämlich

- im Labor Hombrechtikon: die Weiterentwicklung einer Wörterkennungssoftware aus einem Vorprojekt und der Aufbau eines PC-gestützten Wörterkenners mit ISDN-Schnittstelle
- im Labor Solothurn: die Erkennung einzelnen gesprochener Buchstaben mit neuronalen Netzen.

## 71 Wörterkennungssoftware

Ausgangspunkt für die Arbeiten an der Erkennungssoftware war ein Programm, das bei vorgegebenen Wortanfängen und -enden die Ziffern null bis neun mit 95 % Zuverlässigkeit erkannte. Mit ersten Arbeiten erhöhte man die Zuverlässigkeit der Erkennung auf 99 %. Die anschliessenden Experimente mit den 26 Wörtern der Buchstabiertabelle bestätigen die 99 %. In einem zweiten Schritt wurde angestrebt, aufeinanderfolgende Wörter aus dem kontinuierlichen Signal heraus zu erkennen, also ohne Vorgabe manuell markierter Wortanfänge und -enden. Dabei wurde schliesslich eine Zuverlässigkeit der Erkennung von 96 % erzielt.

## 72 PC-gestützter Wörterkennner mit ISDN-Schnittstelle

Für den praktischen Einsatz der Wörterkennung im ARA-System ist aus Gründen der zurzeit erhältlichen Systemkomponenten eine PC-gestützte Lösung gewählt worden. Eine ISDN-Karte als Schnittstelle zum SwissNet 2 erlaubt es, kontinuierliche Sprachsignale über den PC-Bus an eine Karte für digitale Signalverarbeitung (Digital Signal Processing DSP) mit einer Rechenleistung von 30 MFLOPS (Millionen Gleitkomma-Rechenoperationen je

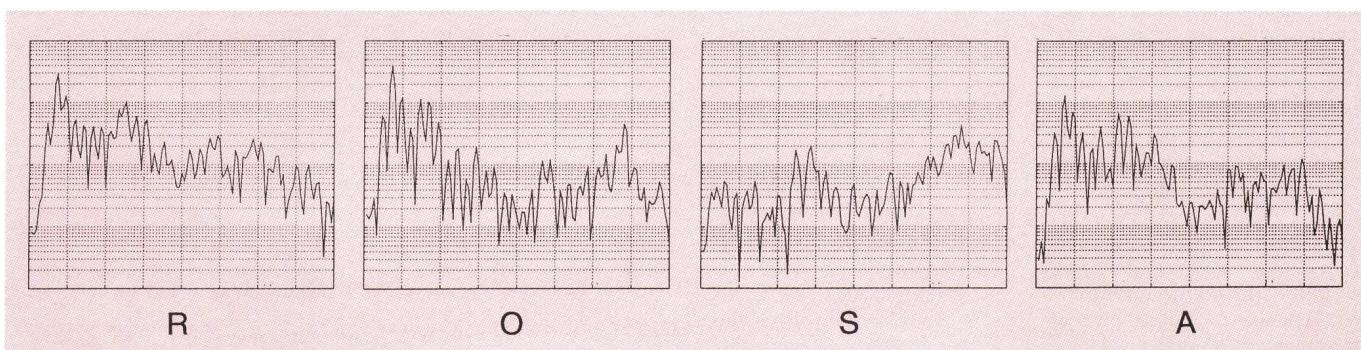


Fig. 4 Frequenzspektren der Laute R, O, S und A

Sekunde, Mega Floating Point Operations per Second) zur Durchführung der Wörterkennung weiterzuleiten. Auf dem PC sind die Steuerung der ISDN- und der DSP-Karte sowie der Dialog mit der zentralen Anwendungssteuerung als parallele Prozesse verwirklicht.

### 73 Erkennung einzelner gesprochener Buchstaben mit neuronalen Netzen

Bei diesen Arbeiten mussten zuerst Aufnahmen von insgesamt 200 Sprecherinnen und Sprechern durchgeführt und Wortanfänge und -enden markiert werden. Anschliessend wurden Experimente mit neuronalen Verzögerungsnetzen (Time-Delay-Neural-Networks, TDNN) durchgeführt, die auf einzelne Gruppen von Buchstaben mit hoher Verwechslungswahrscheinlichkeit spezialisiert werden, und sich anschliessend zu einem Gesamtnetz zusammenfügen lassen. Am Beispiel der Buchstabengruppe (f, l, m, n, r, s) wurde eine Erkennungsrate von 86 % erzielt.

## 8 Der Nutzen des ARA-Projekts für die Telecom PTT

Wie erwähnt, sind im Ausland bereits etliche sprachgesteuerte Systeme im Einsatz, wenn auch für viel einfachere Anwendungen als dies eine Rufnummernauskunft ist. Es soll hier nicht über die möglichen Gründe argumentiert werden, wieso in der Schweiz bisher noch keine solchen Systeme zur Anwendung gebracht worden sind.

Im Zusammenhang mit dem ARA-Projekt ist es jedoch wichtig, darauf hinzuweisen, dass es selbstverständlich ein leichtes ist, in das aufgebaute ARA-System einfache Dienste zu integrieren. Dies wurde bereits in ARA-V1 mit dem Wetterbericht, der sprechenden Uhr und dem auto-

matischen Weckdienst gemacht. Nennenswerte Schwierigkeiten sind eigentlich nur bei Anwendungen zu erwarten, in denen der Sprachsynthesator Texte mit unbekannten Eigennamen oder fremdsprachigen Wörtern in Sprachsignale umsetzen muss. Dies ist sicher bei den Nachrichten und den Sportmeldungen der Fall.

Die erste Phase des ARA-Projekts steht kurz vor dem Abschluss, und die erste Version ARA-V1 kann bereits zu Testzwecken eingesetzt werden. Um erste Erfahrungen mit dem System zu sammeln und diese für den Schlussbericht auswerten zu können, *wird vom 30. Mai bis zum 15. Juni 1994 unter der Telefonnummer 01 260 19 75 ein Testbetrieb eingerichtet*. Die Projektbearbeiter möchten die Leserschaft ermuntern, das System anzurufen und zu versuchen, eine Telefonnummer oder eine Adresse abzufragen. Am Schluss kann ein Kommentar abgegeben werden. Der gesamte Dialog wird für die Auswertung aufgezeichnet.

### Adressen der Autoren

Beat Pfister  
Institut für Technische Informatik  
und Kommunikationsnetze, ETH Zentrum  
CH-8092 Zürich

Arthur Schaub  
Ascom Tech AG, H 32 412  
CH-8634 Hombrechtikon

### Bibliographie

- [1] Rabiner L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. of the IEEE, Febr. 1989, S. 257–286.
- [2] Weibel A. and Hampshire J. Building Blocks for Speech. Byte, Aug. 1989. S. 235–242.

## Zusammenfassung Résumé

### Automatische Rufnummernauskunft

Sprachgesteuerte automatische Auskunftssysteme kommunizieren mit dem Benutzer über das Medium der gesprochenen Sprache. Das Ziel dieser Kommunikation ist, den Informationswunsch des Benutzers zu eruieren und anschliessend die gewünschten Auskünfte zu geben. Ein solches System braucht neben den Fähigkeiten, gesprochene Sprache zu verstehen (Spracherkennung) und Sprache zu produzieren (Sprachsynthese) auch Zugriff auf die verlangten Informationen. Dies können beispielsweise Wetterprognosen, Telefonnummern, Aktienkurse, Fahrplaninformationen usw. sein.

### Renseignements par numéros d'appel à réponse automatique

Les systèmes de réponse automatique à commande vocale communiquent avec l'usager en langage parlé. Le but de cette communication est de déceler l'information que l'utilisateur désire et de lui fournir ensuite les renseignements voulus. Un tel système doit non seulement comprendre le langage parlé (reconnaissance de la parole) et produire des signaux vocaux (synthétiseur de parole), mais encore pouvoir accéder aux informations désirées. Il peut s'agir par exemple de météo, de numéros téléphoniques, de cours de la Bourse, d'informations sur des horaires de transport, etc.

## Riassunto

### Servizio telefonico automatico d'informazione

I sistemi d'informazione automatici con guida fonica comunicano con l'utente mediante la voce. Lo scopo della comunicazione è di scoprire che cosa desidera il cliente e di dargli quindi le informazioni richieste. Oltre ad essere in grado di capire la voce (riconoscimento della voce) e di generarla (sintesi della voce), un simile sistema deve poter anche accedere alle informazioni richieste. Quest'ultime possono essere di tipo diverso: previsioni del tempo, numeri telefonici, quotazioni di borsa, orari di partenza e di arrivo, ecc.

## Summary

### Automatic Telephone Directory

Speech controlled automatic information systems communicate with the user by means of the spoken language. The aim of this communication is to find out the information request of the user and then to give the required information. Such a system needs, apart from the capabilities of understanding the spoken language (speech recognition) and producing speech (speech synthesis), access to the required information. This could be for example weather forecasts, telephone numbers, stock quotations, timetable information etc.