

**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:** 72 (1994)

**Heft:** 12

**Artikel:** Servizio telefonico automatico d'informazione

**Autor:** Pfister, Beat / Schaub, Arthur

**DOI:** <https://doi.org/10.5169/seals-874741>

### **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:** 24.11.2025

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

# Servizio telefonico automatico d'informazione\*

Beat PFISTER, Zurigo, e Arthur SCHAUB, Hombrechtikon

## Zusammenfassung

*Automatische Rufnummernauskunft*

Sprachgesteuerte automatische Auskunftssysteme kommunizieren mit dem Benutzer über das Medium der gesprochenen Sprache. Das Ziel dieser Kommunikation ist, den Informationswunsch des Benützers zu eruieren und anschliessend die gewünschten Auskünfte zu geben. Ein solches System braucht nebst 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.

## Résumé

*Service téléphonique automatique d'informations*

Les systèmes de réponse automatique à commande vocale communiquent avec l'utilisateur en langage parlé. Le but de cette communication est de détecter 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 (synthèse de la parole), mais encore pouvoir accéder aux informations désirées. Il peut s'agir par exemple de bulletins 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 (language recognition) and producing speech (speech syntheses), access to the required information. This could be for example weather forecasts, telephone numbers, stock quotations, timetable information, etc.

## 1 Introduzione

I sistemi d'informazione automatici con guida fonica dei quali si parla in questo articolo comunicano con l'utente mediante la voce. Lo scopo è di stabilire quali informazioni desidera il cliente e quindi di fornirglielle.

Simili sistemi d'informazione devono essere in grado non solo di comprendere la voce (riconoscimento della voce) e di generarla (sintesi della voce), ma anche di accedere alle informazioni richieste, per esempio alle previsioni meteorologiche, ai numeri telefonici, alle quotazioni borsistiche, agli orari di partenza e di arrivo ecc.

Evidentemente Telecom PTT è interessata soprattutto ad un sistema capace di fornire automaticamente informazioni sui numeri di telefono degli abbonati. Per questo motivo nel 1991 ha dato l'avvio, in collaborazione con il *Politecnico federale di Zurigo (ETH)* e la

ditta *Ascom Tech AG*, ad un progetto comune allo scopo di compiere un primo passo verso i servizi telefonici automatici d'informazione (ARA = Automatische RufnummernAuskunft).

## 2 Il servizio telefonico automatico d'informazione ideale

Il sistema ARA ideale dovrebbe essere in grado di sostituire in modo completo il servizio d'informazione prestato da personale umano. Ciò significa in primo luogo che un utente qualsiasi, senza conoscenze preliminari, dovrebbe poter comunicare al sistema mediante la voce quale informazione desidera. Il sistema, che può accedere ad una banca dati appropriata, dovrebbe cercare l'informazione richiesta e fornirla all'utente con un linguaggio simile alla voce naturale.

Da un sistema automatico si esige anche che lavori rapidamente, che non sia mai sovraccarico, che non faccia errori, che disponga sempre delle informazioni più aggiornate e che sia economico per l'utente e redizizio per il suo gestore.

\* Progetto sviluppato in comune dal Politecnico federale di Zurigo, dalla Ascom Tech AG e da Telecom PTT

Un sistema d'informazione automatico che soddisfi tutte queste esigenze rimane tuttavia per il momento solo un'utopia. Una lunga serie di problemi impedisce ancor oggi che questa utopia diventi realtà. Molti problemi sono dovuti al fatto che il linguaggio naturale, al contrario di quelli di tipo formale (p. es. i linguaggi di programmazione), è estremamente difficile da «trattare»:

- Le ricerche svolte nel settore del riconoscimento della voce sono lungi dall'aver raggiunto la meta. Un sistema che si avvicini alla facoltà tipica dell'uomo di riconoscere la voce, che in pratica sia in grado di convertire il parlato nella corrispondente forma testuale per un lessico illimitato e per qualsiasi costruzione sintattica o addirittura il dialetto, è decisamente avveniristico.
- Un sistema d'informazione non ha però solo il compito di trasformare la richiesta orale di un utente nel corrispondente testo. Il sistema deve anche riuscire a capire ciò che l'utente desidera e sapere come soddisfare concretamente la richiesta d'informazione, impiegando le banche dati a disposizione. L'analisi del significato del linguaggio naturale è attualmente realizzabile solo per le applicazioni semplici con un lessico fortemente limitato e per formulazioni elementari.
- In qualsiasi comunicazione in cui si utilizza il linguaggio parlato naturale possono sorgere problemi di comprensione causati sia dal fatto che qualcosa non è stato sentito o non è stato capito, sia dal fatto che una parte risulta sconnessa rispetto al resto. Affinchè una comunicazione venga recepita è necessario che il sistema riconosca queste situazioni e sia in grado di reagire in maniera adeguata. L'utente inoltre deve in ogni momento avere la possibilità di informare il sistema del fatto che non ha capito qualcosa.
- Oltre a queste difficoltà che risiedono nel linguaggio naturale, si devono anche risolvere problemi legati alla vita quotidiana. Così è necessario che per certe informazioni l'apparecchio sia in grado di riconoscere dati cronologici e geografici, in altri termini di capire nozioni quali «prima», «dopo», «vicino» ecc.

Per poter realizzare un sistema automatico d'informazione capace di sostituire pienamente il servizio d'informazione prestato da personale umano si dovrebbero risolvere questi ed altri problemi. Tuttavia non è necessario nè sensato procedere secondo il principio del «tutto o niente»; infatti si possono senz'altro realizzare sistemi automatici d'informazione più semplici, come già avviene all'estero. A tal fine l'utente deve però adeguarsi drasticamente alle possibilità offerte dalla macchina, soprattutto al modo in cui deve parlare alla macchina, o in altri termini, alle possibilità offerte dal sistema di riconoscimento della voce.

Le prestazioni dei sistemi di riconoscimento della voce attualmente a disposizione sono relativamente modeste se confrontate con le summenzionate esigenze. Stanno già comparando sul mercato sistemi che indipendentemente da chi parla sono in grado di

riconoscere singole parole parlate costituenti un lessico di poche centinaia di termini. Per quanto riguarda invece la capacità di riconoscere il linguaggio parlato in modo continuo, come si richiederebbe al sistema ideale, si è ancora fermi a prove di laboratorio con un lessico fortemente limitato.

Qui di seguito viene mostrato in che modo le summenzionate prove di laboratorio, effettuate con un lessico limitato nell'ambito del progetto ARA, consentano di fare un primo passo in direzione di un sistema automatico d'informazione altamente valido.

### 3 Obiettivi del progetto ARA

Da un sistema ARA si può ottenere il numero di chiamata di un abbonato del quale si sono immessi il cognome e l'indirizzo, oppure si possono ottenere il cognome e l'indirizzo dell'abbonato di cui si è immesso il numero di chiamata. Con l'ETV (elenco elettronico degli abbonati) di Telecom PTT ciò è possibile in modo relativamente semplice attraverso la consolle di uno schermo. Sarebbe però assai più comodo se le persone, per comunicare con l'elenco elettronico, potessero parlare al telefono invece di utilizzare tastiera e schermo.

Questo progetto non è tuttavia facile da realizzare, in quanto il lessico concernente il tipo di comunicazione in questione è praticamente illimitato poichè consiste principalmente di nomi propri (nomi e cognomi di abbonati, nomi di strade e di località). Non è perciò possibile impiegare un dispositivo di riconoscimento delle parole convenzionale per immettere nel sistema i dati di un abbonato del quale si desidera sapere il numero di chiamata.

Per realizzare la prima versione di un sistema ARA (designato quale ARA-V1) funzionante con un dispositivo in grado di riconoscere le parole indipendentemente da chi parla, si è partiti dall'ipotesi che l'utilizzatore non pronuncia i nomi propri per intero, bensì li compita secondo la tabella ufficiale di compitazione che si trova in qualsiasi elenco telefonico. Si è cioè potuto ridurre drasticamente il lessico necessario al sistema ARA alle parole figuranti nella tabella di compitazione, a cifre, ad alcune parole di comando e di dialogo.

Un sistema simile non può certamente essere considerato un'alternativa valida al sistema d'informazione prestato da personale umano. Tuttavia è assai utile compiere questo primo passo. Molti problemi legati alla comunicazione fra uomo e macchina e all'interazione con gli elenchi elettronici degli abbonati possono essere studiati altrettanto bene sia con questa struttura fortemente semplificata che con un sistema più complesso. Può inoltre succedere che anche in caso di un sistema ARA dotato di dispositivo di riconoscimento della voce si debbano sillabare per intero o parzialmente determinati nomi propri, in particolare quelli aventi una grafia diversa ma una pronuncia identica (p. es. Maier, Meier, Meyer) oppure quelli di difficile comprensione.

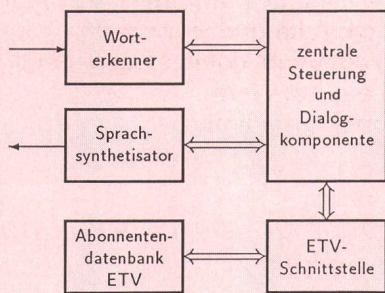


Fig. 1 Schema a blocchi del sistema ARA

ETV, elektronisches Teilnehmerverzeichnis — ETV, elenco elettronico degli abbonati  
 Worterkenner — Dispositivo di riconoscimento delle parole  
 Sprachsynthesator — Sintetizzatore della voce  
 Abonentendatenbank ETV — Banca dati degli abbonati ETV  
 Zentrale Steuerung und Dialogkomponente — Comando centrale e componenti del dialogo  
 ETV-Schnittstelle — Interfaccia ETV

La figura 1 mostra in modo sommario lo schema a blocchi del sistema ARA-V1. Le quattro parti principali sono:

#### Il dispositivo di riconoscimento delle parole

Unità in grado di riconoscere singole parole parlate indipendentemente da chi parla (cfr. punto 7).

#### Il sintetizzatore della voce

Nel nostro caso si tratta del sistema di sintesi della voce SVOX sviluppato dal Politecnico federale di Zurigo, che consente al sistema ARA di dare istruzioni all'utente, di porre a sua volta domande o di dare risposte (cfr. punto 5).

#### Il dispositivo per l'interrogazione della banca dati

Un'interfaccia intelligente verso l'elenco elettronico degli abbonati ETV di Telecom PTT (cfr. punto 6).

#### Il comando centrale

Processo dialogico e decisionale flessibile che, sfruttando tutti i dati a disposizione, è in grado di svolgere con l'utente un dialogo finalizzato ad un obiettivo, con l'aiuto del riconoscimento e della sintesi della voce.

Questa configurazione del sistema consente all'utente di svolgere il dialogo liberamente, ovviamente nei limiti noti: ciò significa che il dispositivo di riconoscimento della voce è in grado di riconoscere solo singole parole parlate e può commettere errori di riconoscimento più o meno frequenti a seconda di chi parla. Per quanto riguarda il dialogo ARA-V1 si devono considerare in particolar modo i seguenti aspetti:

- Il sistema ARA assume sin dall'inizio la guida del dialogo per impedire che l'utente parli liberamente invece di utilizzare solo le parole autorizzate. Inoltre formula le domande e le istruzioni in modo tale da non indurre l'utente a rispondere con una frase invece che con una parola.

- Il sistema deve poter essere utilizzato senza conoscenze preliminari: chi lo utilizza per la prima volta dev'essere guidato per mezzo di spiegazioni che però non devono intralciare l'utente già abile.
- Il dialogo deve scorrere in modo semplice e normale, senza per questo sollecitare l'utente.
- Durante il dialogo il sistema deve riuscire a impiegare i dati della banca dati degli abbonati in modo tale che l'obiettivo, cioè la determinazione dell'iscrizione cercata nella banca dati interessata, possa essere raggiunto dall'utente con il minor numero possibile di interrogazioni.
- Nel corso del dialogo il sistema deve impiegare il «metodo della conferma» per impedire che gli errori di riconoscimento lo facciano fallire. Le conferme e le correzioni devono essere strutturate in modo che l'utente non perda il filo quando compita o quando immette una sequenza di cifre.

Alcuni di questi aspetti si sono rivelati in contraddizione l'uno con l'altro (p. es. ogni conferma rende più lento il dialogo) nel corso dello sviluppo del progetto. Per questo motivo essi possono essere presi in considerazione solo limitatamente ed è pertanto ogni volta una questione opinabile stabilire di quale di essi si debba tenere conto in quale punto del dialogo e in che misura.

## 4 Ulteriore sviluppo del sistema ARA

Il motivo per cui chi utilizza la prima versione del sistema è costretto a compitare in base alla tabella di compitazione è riconducibile in primo luogo al dispositivo di riconoscimento delle parole. Solo così è possibile che l'utente possa immettere una serie di lettere che vengano comprese correttamente, con una certa sicurezza, dal sistema ARA.

Molto più comodo sarebbe evidentemente poter compitare in modo normale (senza ricorrere alla tabella di compitazione). Poiché le lettere parlate al telefono risultano difficilmente comprensibili, occorrono informazioni supplementari per migliorare il loro riconoscimento. Si può sfruttare il fatto, per esempio, che la sequenza di lettere immessa debba corrispondere ad un nome di località. Ciò può essere risolto in modo relativamente semplice con un metodo che consiste nell'attribuire alla sequenza di dati immessi il nome più somigliante, in base alla matrice di cui è dotato il dispositivo di riconoscimento. A tal fine sono tuttavia necessarie liste di località, rispettivamente liste con i nomi degli abbonati e delle strade di ogni località. Questo non è però possibile con l'interfaccia ETV attualmente disponibile.

Un passo considerevole in direzione di una maggiore comodità per chi utilizza un sistema ARA può essere fatto se si impiega un dispositivo in grado di riconoscere le parole pronunciate normalmente di un vocabolario illimitato. A tal fine non devono però essere disponibili solo le summenzionate liste, ma anche la pronuncia corrente di tutti i nomi propri.

Con questo passo in avanti, per quanto riguarda il dispositivo di riconoscimento delle parole e un dialogo

adeguato in modo ottimale, è stato raggiunto il punto in cui un sistema ARA riscuote vasti consensi e di conseguenza viene utilizzato sovente. Il motivo per cui tale aspetto non è stato preso in considerazione fin dall'inizio quale obiettivo del progetto è il seguente: da un canto, il passo fatto da zero fino a questo punto è troppo lungo, in altre parole ancora caratterizzato da troppe incognite per poter essere preso seriamente in considerazione; d'altro canto la compilazione, con o senza tabella, continuerà ad essere utilizzata anche nei sistemi futuri, proprio come avviene per il servizio d'informazione prestato da personale umano, dove essa viene usata di frequente e spontaneamente.

## 5 Il sistema di sintesi della voce SVOX

Il sistema SVOX (fig. 2) genera per ogni testo tedesco frase per frase un segnale vocale corrispondente: fa dunque parte dei sistemi «text to speech». Esso consiste di due parti principali, la *parte di trascrizione* e il *livello fonoacustico*.

La trascrizione trasforma l'immissione ortografica in una rappresentazione fonologica (una rappresentazione astratta del segnale vocale da generare) che comprende una sequenza di suoni corrispondente al testo, gli accenti di parola e di frase e i confini dei gruppi fonetici. A tal fine le parole di una frase vengono dapprima analizzate dal punto di vista morfologico (determinazione di genere e forma della parola) e poi viene determinata la loro trascrizione fonetica. Il risultato dell'analisi morfologica viene utilizzato per determinare la struttura sintattica in base ad una grammatica. La struttura sintattica serve ad effettuare la ripartizione in gruppi fonetici (frasi) e a determinare i relativi accenti di parola.

Il livello fonoacustico determina, in base a questa rappresentazione astratta, la durata dei suoni, l'andamento dell'altezza del suono e dell'intensità del segnale vocale per la sintesi: ciò viene chiamato «comando prosodico». In conclusione il sintetizzatore difono modifica gli elementi di transizione del suono relativi alla lingua naturale (difoni), in modo tale che essi abbiano la durata, la frequenza fondamentale e l'intensità richieste e li concatena al segnale d'uscita desiderato (sintesi di concatenazione).

## 6 L'interfaccia ETV

L'interfaccia verso l'elenco elettronico degli abbonati ETV ha il duplice compito di trasformare le richieste della guida dialogica in una sequenza di immissioni compatibili ETV (Strings) e di ritrasformare la sequenza in uscita in una forma utilizzabile dalla guida. L'attuale ETV permette ciò solo in misura molto limitata, poichè le emissioni non sono strutturate in modo da poter essere elaborate senza un'interpretazione semantica. In linea di massima un simile dispositivo di interpretazione è realizzabile, ma solo con un considerevole dispendio di costi.

Per la prima versione del sistema ARA è stato sviluppato un dispositivo di interpretazione molto semplice, nell'ipotesi che in un prossimo futuro l'elenco elettronico degli abbonati di Telecom PTT venga strutturato meglio.

## 7 Il dispositivo di riconoscimento delle parole

I compiti e le difficoltà del dispositivo di riconoscimento della voce sono illustrati nelle figure 3 e 4. La figura 3 rappresenta un segnale vocale della durata di 2 s. Chi parla pronuncia il nome «Rosa». Prima e dopo il nome sono visibili dei rumori disturbatori, tipici delle comunicazioni telefoniche. Nella figura 4 sono invece visibili gli spettri di segmenti temporali, lunghi ciascuno 30 ms, dei quattro suoni che compongono la parola «Rosa».

L'analisi delle caratteristiche del segnale vocale — essenzialmente della distribuzione spettrale della potenza del segnale e delle derivazioni temporali di queste grandezze — permette il riconoscimento automatico della voce. Per l'elaborazione di simili caratteristiche sono noti i procedimenti con «modelli Markov nasco-

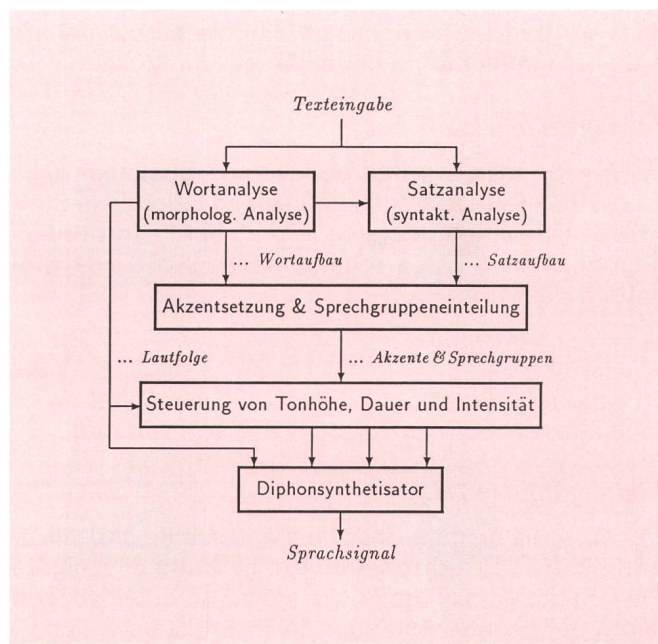


Fig. 2 Schema a blocchi del sistema di sintesi della voce SVOX

Texteingabe — Immissione del testo  
 Wortanalyse (morphologische Analyse) — Analisi della parola (analisi morfologica)  
 Satzanalyse (syntaktische Analyse) — Analisi della frase (analisi sintattica)  
 Wortaufbau — Struttura delle parole  
 Satzaufbau — Struttura delle frasi  
 Akzentsetzung und Sprechgruppeneinteilung — Accentazione e suddivisione dei gruppi fonetici  
 Lauffolge — Sequenza di suoni  
 Akzente und Sprechgruppen — Accenti e gruppi fonetici  
 Steuerung von Tonhöhe, Dauer und Intensität — Comando dell'altezza, della durata e dell'intensità del suono  
 Diphonsynthetisator — Sintetizzatore difono  
 Sprachsignal — Segnale vocale

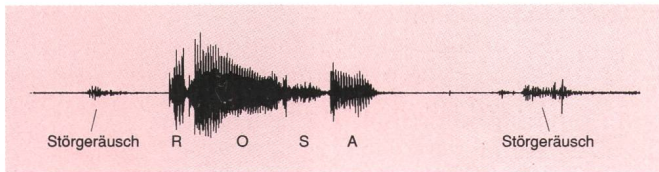


Fig. 3 Segnale vocale della parola «Rosa»  
Störgeräusch — Rumori disturbanti

sti» (Hidden Markov Models) [1] e con «reti neurali» [2]. Nonostante si tratti di procedimenti fondamentalmente classici, per la riuscita dell'applicazione pratica è necessario migliorare ancora di molto la struttura e le dimensioni.

Diversamente dalla fase di training e dalla prima fase di test in cui di regola vengono utilizzati i segmenti estremi (inizio e fine) delle parole marcate manualmente rispettivamente corrette, un dispositivo di riconoscimento delle parole deve essere in grado di distinguere le singole parole dal segnale continuo. Ciò è particolarmente difficile per le comunicazioni telefoniche poichè ai rumori provocati da chi parla, come i respiri o lo schioccare prima delle parole, si aggiungono anche rumori dovuti alle linee e altri rumori disturbatori.

Nel quadro del progetto ARA, sono stati svolti presso la ditta Ascom Tech AG diversi lavori per il riconoscimento della voce:

- nel laboratorio di Hombrechtikon si sono studiati il perfezionamento di un software per il riconoscimento delle parole e la realizzazione di un dispositivo di riconoscimento della voce basato su PC con interfaccia ISDN
- nel laboratorio di Soletta si è studiato il riconoscimento con reti neurali di singole lettere parlate.

### 71 Il software per il riconoscimento delle parole

Il punto di partenza per i lavori riguardanti il software per il riconoscimento delle parole è stato un programma in grado di riconoscere, con un grado di affidabilità

del 95 %, le cifre da 0 a 9 in caso di segmenti estremi (inizio e fine) delle parole precedentemente fissati. Durante i primi lavori il grado di affidabilità è salito fino al 99 %. I successivi esperimenti con le 26 parole della tabella di computazione hanno confermato un indice del 99 %. L'obiettivo della seconda fase era di distinguere da un segnale continuo una sequenza di parole senza che l'inizio e la fine delle parole fossero marcate manualmente. Ne è risultato un grado di affidabilità pari al 96 %.

### 72 Il dispositivo di riconoscimento delle parole basato su PC con interfaccia ISDN

Per l'impiego pratico del riconoscimento delle parole nel sistema ARA è stata scelta una soluzione basata su PC conformemente alle tendenze attuali del mercato. Una scheda ISDN quale interfaccia verso la rete SwissNet 2 permette di trasmettere, per il riconoscimento delle parole, segnali vocali continui attraverso il bus del PC a una scheda per l'elaborazione digitale dei segnali (Digital Signal Processing, DSP), con una capacità di calcolo di 30 MFLOPS (milioni di operazioni di calcolo a virgola mobile per ogni secondo, Mega Floating Point Operations per Second). Il comando delle schede ISDN e DSP e il dialogo con il comando centrale di applicazione sono realizzati sul PC quali processi paralleli.

### 73 Il riconoscimento con reti neurali di singole lettere parlate

Durante i lavori sono state effettuate le registrazioni di circa 200 parlatrici e parlatori e sono stati marcati i segmenti estremi (inizio e fine) delle parole. Sono stati svolti anche esperimenti con reti neurali a differimento di tempo (Time-Delay-Neuronal-Networks, TDNN), le quali vengono specializzate in singoli gruppi di lettere con un'alta probabilità di essere capite male e che possono essere ricondotte ad una rete globale. Per il gruppo di lettere «f, l, m, n, r, s» è stato raggiunto un tasso di riconoscimento dell'86 %.

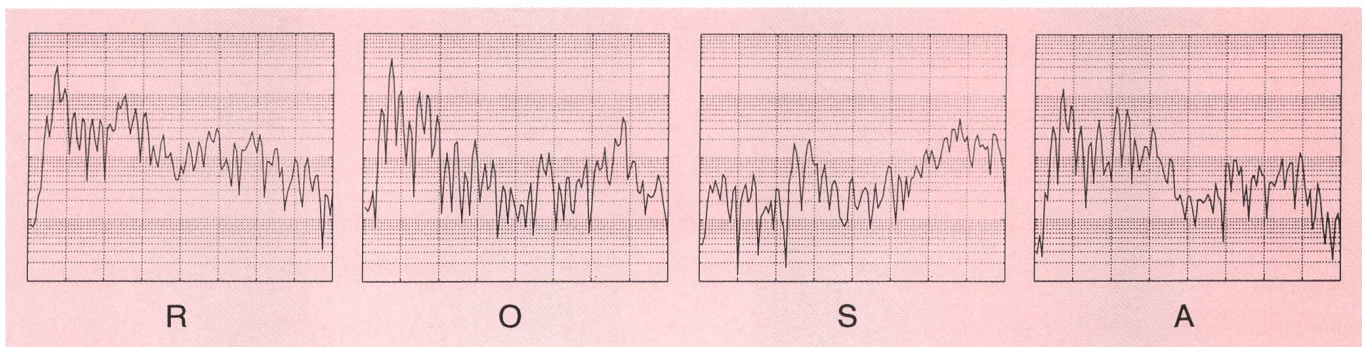


Fig. 4 Spettri di frequenza dei suoni R, O, S e A

## 8 Vantaggi del progetto ARA per Telecom PTT

All'estero vengono già impiegati alcuni sistemi di riconoscimento della voce, anche se per servizi molto più semplici di quelli costituiti dall'informazione sui numeri di chiamata. Non è questa la sede adatta per discutere dei possibili motivi per cui in Svizzera simili sistemi non sono ancora stati impiegati.

È tuttavia importante sottolineare quanto sia sostanzialmente facile integrare servizi semplici nel sistema ARA. Questo è già stato compiuto nel sistema ARA-V1 con le previsioni del tempo, l'ora esatta e il servizio automatico di sveglia. Praticamente solo le applicazioni nelle quali il sintetizzatore della voce deve trasformare in segnali vocali testi con nomi propri sconosciuti o con parole straniere possono fare sorgere difficoltà di rilievo. È il caso del notiziario e dei risultati sportivi.

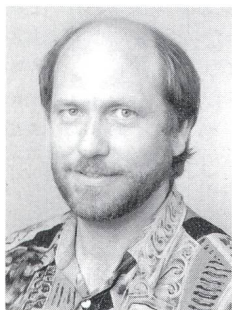
### Indirizzi degli autori:

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

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

### Bibliografia

- [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.



**Beat Pfister** è nato il 22 dicembre 1950 ed è cresciuto a Dagmersellen nel canton Lucerna. Dopo avere conseguito la maturità scientifica presso il liceo di Lucerna, ha frequentato il politecnico federale di Zurigo laureandosi in elettrotecnica. All'inizio del 1976 è entrato a far parte del nuovo gruppo impegnato nell'elaborazione della voce presso l'istituto per la fisica tecnica, in seno al quale si è occupato principalmente della realizzazione di calcolatori di grande potenza per l'elaborazione dei segnali. Dal 1981 dirige il gruppo per l'elaborazione della voce, annesso all'istituto per l'informatica tecnica e le reti di comunicazione. Negli ultimi anni, il gruppo composto attualmente di otto persone, si è occupato soprattutto della sintesi e del riconoscimento della lingua tedesca come pure della sua applicazione nel campo dei sistemi di informazione automatici.



**Arthur Schaub** è nato nel 1950 e ha studiato fisica sperimentale all'università di Berna. Oltre a dedicarsi alla simulazione digitale di circuiti analogici in diversi progetti industriali, si è perfezionato nell'elaborazione — classica, adattativa e digitale con varie velocità di trasmissione — dei segnali. Dalla primavera del 1988 lavora presso l'Ascom dove dirige un laboratorio per l'elaborazione digitale della voce e si occupa principalmente delle seguenti attività: la codificazione della voce per memorizzare e trasmettere in modo efficiente la voce digitale, il riconoscimento della voce indipendentemente dall'interlocutore e nel caso in cui la voce dell'interlocutore è conosciuta come pure la preparazione di segnali vocali per gli apparecchi acustici. Egli si interessa inoltre della statistica matematica e della teoria dei test come pure della codificazione che riconosce gli errori e li corregge.