An Evaluation of METAL: the LRC Machine Translation System

Jonathan Slocum
Microelectronics and Computer Technology Corp.

Winfield S. Bennett
Lesley Whiffin
Edda Norcross
Siemens Communication Systems, Inc.

Abstract

The Linguistics Research Center (LRC) at the University of Texas at Austin is currently developing METAL, a fully-automatic high quality machine translation system, for market introduction in 1985. This paper will describe the current status of METAL, emphasizing the results of the most recent post-editors' evaluation, and will briefly indicate some future directions for the system. A 6-page German original text and a raw (unedited, but automatically reformatted) METAL translation of that text into English are included as appendices.

Introduction

The Linguistics Research Center (LRC) at the University of Texas at Austin is currently developing METAL, a fully-automatic high quality machine translation system, for market introduction in 1985. This paper will describe the current status of METAL, including the results of the most recent evaluation, and will briefly indicate some future directions for the system. Exhibits A and B (attached) are, respectively, a German original text and a raw (unedited, but automatically reformatted) METAL translation of that text into English.

History and Status

Machine translation research at the University of Texas began in 1956; the LRC was founded in 1961. For much of the history of this project, funding was provided by the U.S. Air Force's Rome Air Development Center and other U.S. government agencies. In 1979, Siemens AG began funding the development phase of the METAL machine translation system, at which point implementation of the current system was initiated. A prototype has recently been delivered to the sponsor for market testing.

The current system is a unidirectional German-English system, although work to add other target languages, as well as creating an English-German MT system, is now underway. The present staff for the METAL project consists of seven full-time and five half-time personnel.

Application Environment

Software has been developed to handle the formatting problems associated with technical manuals. This software, written in SNOBOL, automatically marks and prepares texts for the METAL translation system [Slocum and Bennett, 1982; Slocum et al., 1984]. The only human intervention prior to translation is checking and correcting the results of the automatic formatting routines. Postediting is expected for the output texts. The system does not expect (or provide for) human intervention during the actual translation phase.

Pre-processing and post-editing are presently done on a DEC-2060; the actual translation, on a Symbolics Lisp machine. The "production system" design envisions a Lisp Machine as the translation unit connected to 4-6 translator workstations, from which the prepared texts will be sent to the translation unit and on which the output texts will be postedited.

METAL uses a transfer approach for translation. The entire process consists of four phases: analysis, integration, transfer, and generation (synthesis). The integration phase works with whole parse tree structures, following analysis and preceding transfer. Until recently, transfer and generation were essentially a single phase, but work is currently underway to separate this single phase into two, with a much more powerful generation phase.

Linguistic Component

The current METAL lexicon consists of over 20,000 German and English monolingual entries, consisting of morphological, syntactic, and semantic features and values, and an appropriately large number of transfer entries. The features and values in monolingual lexical entries supply necessary information for the analysis and/or synthesis of these items during the machine translation process. Most entries are reasonably simple, but entries for verb stems are significantly more complex. Inflected adjectives, nouns, and verbs are parsed by word-level grammar rules, with the stems and endings assigned to appropriate lexical categories.

Each transfer lexical entry is a structure equating the source language canonical form with an appropriate target language canonical form. Certain significant information (i.e., lexical category, subject area, and preference) is coded in the entry to guide the system in selecting the appropriate translation. Furthermore, tests and operations (including transformations) may be included within transfer entries.

The grammar for METAL consists of over 600 augmented phrase structure rules, each of which is used in both analysis and transfer/generation. METAL's grammar rules are used in the parsing of all levels of structure from the word level to the sentence level, including phrases and clauses. A METAL grammar rule consists of five analysis sections, plus an additional section for each target language: a top line describing the phrase structure (with an optional enumeration of each constituent); a series of restrictions, which test the appropriateness of individual constituents on the right-hand side of the rule; TESTs, which enforce agreement among the right-hand constituents; a CONSTR section, which constructs the analysis of the phrase; an INTEGR section, which is executed (once a complete analysis of the sentence is achieved) in order to, e.g., resolve anaphoric references; and one or more target-language-dependent Transfer sections, which control lexical and structural translation into the target language.

Homograph resolution and disambiguation are handled uniformly (i.e., without special passes), in various ways: by orthographic tests, such as the test to ensure that a word that looks like a German noun is not all lower case; by positional constraints, which disallow co-occurrence of ambiguous strings in the same clause location; and, most especially, by the case frame mechanism.

The case (valency) frame mechanism is vital in METAL's analysis of German source language sentences. This mechanism is invoked in clause-level rules and uses features on the verb stem to define the functions of the various central arguments to the predicate. In addition, the case frame mechanism is used to test for such things as subject-verb agreement.

The METAL grammar makes extensive use of transformations to modify structure or perform certain tests. Transformations may be used in the TEST, CONSTR, INTEGR, and Transfer phases of the rules; transformations may also be used in transfer lexical entries. A transformation may be written as part of a rule or called by name.

Computational Component

The lexicon for METAL is maintained via a DBMS written in LISP. Input of lexical entries is facilitated by an INTERCODER, a menu-driven system which asks the user for information in English and encodes the answers into the internal form used by the system. An integral part of the INTERCODER is the "lexical default" program which accepts minimal information about the particular entry (root form and lexical category) and encodes most of the remaining necessary features and values. Entries may also be created using any text editor, without the aid of the INTERCODER or lexical defaulter.

Interfacing with the lexical database is done by means of a number of menu-driven functions which permit the user to access, edit, copy, and/or delete entries individually, in groups (using specific features), or entirely. In order to assure a high degree of lexicon integrity the METAL system includes validation programs which identify errors in format and/or syntax. The validation process is automatically used to check lexical items which have been edited, to ensure that no errors have been introduced during editing.

The grammar is also in a database and may be accessed and/or edited in much the same way as the lexicon. System software and named transformations are stored in individual source files.

METAL's parser is a "some-paths, parallel, bottom-up parser" [Slocum et al., 1984]. It may be considered to be "some-paths" because the grammar rules are grouped into numerically indexed "levels" and the parser always applies rules at a lower level before applying rules at a higher level. Once the parser has successfully built one or more Ss at a given level, it will halt; until it discovers one or more S readings, the parser will continue to apply rules in each successive level. Extensive experimentation with the system has found that the present parser configuration is the most efficient one for METAL [Slocum et al., 1984].

Post-Editors' Evaluation

In June, 1984, the METAL system was used to translate 82.6 pages of text into English; the material varied from a sales prospectus (for a speech recognition system) through various general hardware and software system descriptions to highly technical documentation. The output was then edited by two Siemens revisors (one a member of the METAL project, one not). This section describes the revisors' objective performance and subjective reactions (including comparison with earlier versions of METAL) during this experience.

Post-Editor	1st pass	2nd pass	3rd pass	Min/Pg	Pgs/Hr.	Pgs/Day
#1	9hr 10min	3hr 40min	2hr 10min	10.9	5.5	44.1
#2	13hr 40min	3hr 55min		12.8	4.7	37.6

(N.B. The number of pages of text was computed automatically on the basis of "Siemens standard pages": 26 lines x 55 characters = 1430 characters/pg.)

The table above summarizes the editors' revision times. They employed rather different editing techniques (editor #1 working in three passes, #2 in just two), but their times are relatively close.

Comments by Editor #1:

[The 3rd Pass] tends to be concerned with stylistic improvements, formatting changes and typing errors. The last part of this stage involves running the spelling checker on the file to eliminate remaining typing errors.

The impression of post-editing was that there have been many improvements over previous test runs. This was evidenced by the fact that on this post-editing run less than 5% of sentences were re-translated from scratch. The major task in post-editing is now changing word order, changing verb agreement and re-translating the more idiomatic usages. Considerable improvements in format made post-editing easier, although there is still room for further enhancement.

One of the greatest changes affecting post-editing was the fact that since the initial output [compared to earlier versions] of METAL was deemed to have improved, the different stages of post-editing were more clearly defined. That is to say, it was easier to produce an adequate translation during the first run through the text -- using the reformatted output on the screen and a hardcopy of the source text for reference -- than in previous tests. In the second run through a text -- using a hardcopy of the METAL output upon which preliminary post-editing has been performed -- it was easier to concentrate on polishing the translation. In the third and final post-edit stage, one was able to make a final check for stylistic weaknesses, spelling mistakes and typing errors. This was the same method as used in previous tests but one was better able to distinguish between the stages (initial technical and stylistic post-editing; polishing output; final stylistic check) and the entire process was less tiring than in the past.

Although the overall format of the output has improved...there are still [some] problem areas [with the automatic reformatting program].

As an experiment, the unformatted, interlinear [German-English] version was used for the initial post-editing phase. The text was then reformatted...and the 2nd and 3rd phases of post-editing continued as normal. The previous problems with post-editing a highly formatted text meant that whenever a textual change was made in the the text then the format had to be re-modified. The method of post-editing used in this test proved to be considerably faster and easier to handle. ...[The results] demonstrate that the time saving lies in the initial post-edit phase which is when the most changes are made and which is most time intensive with regard to re-formatting text.

Comments by Editor #2:

As compared to the last run in February 1984, the June 84 output showed considerable improvement. A greater number of sentences was useable and many required a change in word order only. Placement of the determiners has been improved. [Certain] points should be considered to improve future translations.

Future Directions

The METAL German-English configuration was released for market testing in January 1985. Current plans are to continue improvement on the present system and to branch off into other target languages, specifically Spanish and Chinese. We estimate that a German-Spanish system should be ready for testing sometime in 1986, with a German-Chinese system sometime thereafter. We have also begun working on an English-German system. If the planned work is successful, work will begin on English-Spanish and English-Chinese MT systems.

References

Slocum, J., and W. S. Bennett, "The LRC Machine Translation System: An Application of State-of-the-Art Text and Natural Language Processing Techniques to the Translation of Technical Manuals," Working Paper LRC-82-1, Linguistics Research Center, University of Texas, July 1982.

Slocum, J., et al., "METAL: The LRC Machine Translation System," presented at the ISSCO Tutorial on Machine Translation, Lugano, Switzerland, 2-6 April 1984. Also available as Working Paper LRC-84-2, Linguistics Research Center, University of Texas, April 1984.

Exhibit A:
an original German text

CSE Spracheingabe-Geraete
Einfuehrungsschrift

1 Einleitung

Die Entwicklung der Halbleitertechnik, insbesondere der Mikroprozessoren, hat in den vergangenen Jahren neue Perspektiven fuer die EDV eroeffnet. Im Bereich der Datenerfassung wurde mit der Spracheingabe in den Computer ein langgehegter Wunsch erfuellt. Damit steht fuer diesen kostenintensiven Zweig der Datenverarbeitung ein Verfahren zur Verfuegung, das die Benutzerfreundlichkeit entscheidend verbessert, zumal die bisherigen Erfassungsmethoden, die sich der Tastatur bedienen, im Prinzip eine Anpassung des Menschen an die Maschine erforderten.
Um die Datenerfassung schneller und sicherer zu machen, muessen die Daten moeglichst am Ort ihres Entstehens erfasst werden. In einer Reihe von Anwendungen ist dies mit Tastaturen problematisch, wenn nicht gar unmoeglich. Dies ist dann der Fall, wenn der Benutzer mobil sein muss, die Haende fuer andere Taetigkeiten frei bleiben sollen, oder die Umweltbedingungen fuer Tastaturen ungeeignet sind.
Die Spracheingabe bietet hier eine ideale Loesung. Gegenueber den traditionellen Erfassungsmethoden zeichnet sie sich durch folgende Eigenschaften aus:

- Optimale Anpassung an die vom Menschen als natuerlich empfundenen Kommunikationsgewohnheiten
- Geringe Einarbeitungszeit
- Leichte Bedienung auch durch ungeuebtes Personal
- Mehr Bewegungsfreiheit fuer den Benutzer beim Erfassen der Daten
- Vereinfachte direkte Datenerfassung
- Groessere Sicherheit der Eingabe.

Die COMPUTER GESELLSCHAFT KONSTANZ bietet folgende Spracheingabegeraete an:

CSE 1050
CSE 1060*).

Die CSE-Geraete lassen sich an Rechner aller bekannten Hersteller anschliessen.
*) Computer-Sprach-Eingabe

2 Technik der Spracherkennung

2.1 Ueberblick

Unter Spracheingabe verstehen wir die Eingabe von Daten per Sprache in den Computer. Dabei wird das gesprochene Wort durch Spracheingabegeraete, die dem Computer vorgeschaltet werden, in maschinell verarbeitbare Information umgewandelt. Beim Erkennungsvorgang wird das Sprachsignal zunaechst in ein Bitmuster umgesetzt. Das Spracheingabegeraet vergleicht daraufhin dieses Muster mit den gespeicherten Mustern des Wortschatzes. Wird eine ausreichende Uebereinstimmung mit einem der Woerter des Wortschatzes festgestellt, gilt dieses Wort als erkannt. Im andern Fall weist das Geraet die Eingabe zurueck.
Der Wortschatz selbst wird zuvor durch ein Training, bei dem jedes Wort mehrfach einzusprechen ist, eingerichtet.
Der typische Datenfluss bei der Spracheingabe sieht wie folgt aus:
Bild 1 Typischer Datenfluss bei Spracheingabe

Die ueber Mikrofon eingesprochenen Woerter werden nach ihrer Erkennung vom CSE-Spracheingabegeraet in Form eines vereinbarten Codes ueber eine genormte Schnittstelle an den Computer weitergegeben. Dieser fuehrt die anwendungsspezifische Verarbeitung durch. In vielen Faellen ist es dabei nuetzlich, dem Benutzer mitzuteilen, welche Daten im Anwenderprogramm des Computers angekommen sind. Diese Rueckmeldung wird durch eine optische oder andere Anzeige, z.B. des erkannten Wortes, erreicht.

2.2 Der Erkennungsvorgang

Technisch koennen beim Vorgang der Spracherkennung zwei Schritte unterschieden werden:

- Die Vorverarbeitung des akustisch-phonetischen Signals
- Die Klassifizierung

Exhibit B:
a raw METAL translation
(February, 1984)

CSE voice data entry devices
Introduction

1 Introduction

The development of semiconductor technology, in particular the microprocessors has opened the new prospects for EDV in the last years. In the range of data acquisition, a long-cherished wish was filled with voice data entry into the computer. with it, a method of operation which improves user convenience deciding is available for this cost-intensive branch of data processing, because required the previous acquisition methods which handle itself of the keyboard in the principle an adaptation of the human being to the machine.
In order to make data acquisition more faster/rapid and secure/sure, the data must be registered inasmuch as possible at the place of its generation. In a series of applications, this is problematic with keyboards, if not indeed impossible.
this is then case, if the user must be mobil, ought to remain the hands for other activities free, or are the environmental conditions unsuitable for keyboards.
Voice data entry offers an ideal solution here. As compared to traditional acquisition methods, it distinguishes qualities/characteristics following through itself:

- Optimal adaptation to the habits of communication felt by the human being as natural
- Short/low training period
- Easy operation through inexperienced users also
- more freedom of movement for the user during gathering the data
- Simplified direct data acquisition
- Larger security of the input.

The COMPUTER GESELLSCHAFT KONSTANZ offers following voice data entry devices:

CSE 1050
CSE 1060*).

The computer voice data entry devices allow follow the computer of all known manufacturers. *) Computer voice data input

2 Technology of speech recognition

2.1 Summary

By voice data entry, we mean the input of the data per language into the computer. Therewith the spoken word is transformed by the voice data entry devices which are pre-connected to the computer into machine readable information. First the voice signal is converted during recognition procedure into a bit pattern. The voice data entry device compares this pattern then with the stored patterns of the vocabulary. if a sufficient correspondence is determined with one of the words of the vocabulary, this word is valid as recognized. In andern case, the device refuses the input.
the vocabulary itself is previously through a training, with that every word mehrfach to speake ist, set up.
The typical data flow during voice data entry appears as follows:
Figure 1 Typical data flow during voice data entry

the words spoken over microphone are transferred interface after its recognition by the CSE voice data entry device in the form of a determined code via a genormte to the computer. This operates the application-specific processing through. In many cases, it is user-friendly with it to report to the user which data are arrived in the user program of the computer. This feedback is reached by an optical or other display, for example of the recognized word.

2.2 Recognition procedure

Two steps can be distinguished technically during the process of speech recognition:

- Preprocessing of the acoustic-phonetic signal
- Classification

Das Spracherkennungsgeraet wandelt das analoge Sprachsignal in digitale Information um. Dabei wird zunaechst das akustische Signal mit Hilfe einer Filterbank in einzelne Frequenzbereiche zerlegt. Aus diesen Bereichen werden in definierten Zeitsegmenten Merkmale gewonnen. Sie beschreiben akustische und phonetische Eigenschaften des Signals. Diese Merkmale werden dann zeitnormiert als Bitmuster abgelegt. Dadurch koennen unterschiedliche Sprechgeschwindigkeiten ausgeglichen werden.
Die anschliessende Klassifizierung dient der Zuweisung des Signals zu einem bestimmten Wort. Dies stellt den Identifizierungsvorgang im engeren Sinne dar. Das Wort wird durch Vergleich seines Musters mit den Bitwerten des definierten Wortschatzes erkannt.
Bild 2 Technischer Ablauf des Spracherkennungsvorganges

2.3 Betriebsmodi

Die CSE-Geraete der COMPUTER GESELLSCHAFT KONSTANZ arbeiten in zwei Betriebsmodi:

- Trainingsmodus
  Er dient dem Aufbau und der Aktualisierung des Wortschatzes. Dabei werden die einzelnen Woerter 5- bzw. 10mal eingesprochen. Das Spracheingabegeraet bildet aus den gewonnenen Bitmustern einen "Mittelwert", der als Referenzmuster in den Wortschatz aufgenommen wird.

- Erkennungsmodus
  In diesem Modus wird das Geraet zur Dateneingabe benutzt.

3 Eigenschaften der CSE-Geraete

Die CSE-Geraete der COMPUTER GESELLSCHAFT KONSTANZ sind sprecheradaptive Einzelworterkennungssysteme.

3.1 Einzelworterkennung

Einzelworterkennungssysteme erfordern eine erkennbare Pause zwischen zwei gesprochenen Woertern.
Bei den CSE-Geraeten muss diese Pause mindestens 100 msec betragen.

3.2 Eigenschaften des Wortschatzes

Der Wortschatz umfasst je nach Speicherausbau des CSE-Geraetes bis zu 370 Woerter. Er besteht aus den fuer die jeweilige Anwendung ausgewaehlten Begriffen. Jeder Sprecher, der mit dem Geraet arbeiten will, trainiert in einer Trainingsphase das Geraet auf seine Stimme. Dadurch finden die individuellen Merkmale seiner Sprache Beruecksichtigung und tragen zu einer groesseren Erkennungssicherheit bei.
Fuer jeden Begriff des Wortschatzes ist zu unterscheiden nach

- akustisch-phonetischem Signal,
- dessen Bedeutung und
- dem vereinbarten Code bzw. Zeichenstring,
  der an den Anwender-Computer uebergeben wird.

Die Trennung nach akustisch-phonetischem Signal und der Bedeutung dieses Signals ist eine typische Eigenschaft von sprecheradaptiven Geraeten. Die Begriffe mit ihrer Bedeutung werden jeweils fuer eine Anwendung festgelegt. Dies gilt auch fuer die an den Anwender-Computer zu uebertragenden Codes bzw. Zeichenstrings. Das akustisch-phonetische Signal jedoch, also das Wort, wie es ausgesprochen wird, ist sprecheradaptiert. Dadurch spielt es auch keine Rolle, ob der Sprecher das Wort mit Dialektfaerbung oder in einer Fremdsprache ausspricht.
Bei sprecheradaptiven Geraeten wird aus den geschilderten Gruenden pro Sprecher ein Wortschatz eingerichtet. Dieser Wortschatz wird in Form von Bitmustern nach dem Training auf einem Hintergrundspeicher (z.B. Magnetplatte, Magnetbandkassette oder Diskette) abgelegt und bei Inbetriebnahme des CSE-Geraetes in dessen Speicher geladen.

3.3 CSE-Geraete als Teil der Anwenderkonfiguration

Die CSE-Spracheingabegeraete sind selbstaendig arbeitende Prozessoren, die jedoch aus der Sicht des Anwender-Computers die Funktion von Peripheriegeraeten haben.
Die Verbindung zur Anwenderkonfiguration erfolgt ueber die genormte V.24-Schnittstelle, die den Anschluss an Computer aller bekannten Hersteller gewaehrleistet.

3.4 Quicktalk-Einrichtung

Um Daten in rascher Folge einzugeben, koennen die CSE-Geraete mit einer Quicktalk-Zusatzeinrichtung ausgeruestet werden. Die moegliche Eingabegeschwindigkeit kann dabei bis zu 180 Woertern pro Minute sein.

The voice data entry device transforms the similar/analog voice signal into digital information. Therewith the acoustical signal with the help/aid of a filter bank into individual frequency ranges is split first. From these ranges, features are gained in defined time segments. They describe the acoustical and phonetic qualities/characteristics of the signal. Then these features are filed time-normalized as bit pattern. Thereby different rates of speech can be equalized.
The subsequent classification is used for the assignment of the signal to a specific word. This represents strictly speaking recognition procedure. The word is recognized by the comparison of its pattern with the bit values of the defined vocabulary.
Figure 2 Technical run of the speech recognititon procedure

2.3 Modes of operation

The computer voice data entry devices of the COMPUTER GESELLSCHAFT KONSTANZ work in two modes of operation:

- Training mode
  It is used for the format and the updating of the vocabulary. with it are the individual words 5- and/or spoken 10 times. the voice data entry device forms from gained bit patterns an "Mittelwert", is received that as reference pattern into the vocabulary.

- Recognition mode
  In this mode, the device is used for data input.

3 Qualities/characteristics of the computer voice data entry devices

The computer voice data entry devices of the COMPUTER GESELLSCHAFT KONSTANZ are speaker-dependent isolated word recognition systems.

3.1 Isolated word recognition

Isolated word recognition systems require a detectable/recognizable pause between two spoken words.
With the computer voice data entry devices, this pause must be at least 100 msec..

3.2 Qualities/characteristics of the vocabulary

The vocabulary contains up to 370 words according to the storage capacity of the computer voice data entry device. It consists of the terms selected for the respective application. Every speaker who/which wants to work with the device trains the device in a training phase for its voice. By this means, the individual features of its language receive consideration and contribute to a larger reading accuracy.
for every term of the vocabulary is to distinguish after

- acoustic-phonetic signal, dessen meaning/importance
- and the determined code and/or character string which is
- delivered to the host computer.

The separation after acoustic-phonetic signal and the meaning/importance of this signal is a typical quality/characteristic of speaker-dependent devices. The terms with its meaning/importance are determined in each case for an application. Also this applies to the codes transmitting to the host computer to and/or character strings. the acoustic-phonetic signal jedoch, therefore the word, how it is pronounced is speaker-dependently. Thereby it also plays no role whether the speaker pronounces the word with dialect coloration or in a foreign language.
With speaker-dependent devices, a vocabulary is set up from the described reasons per speaker. this vocabulary is filed by the bit patterns after the training on a secondary storage (e.g. magnetic disk, magnetic tape cassette or floppy disk) und during the start up of the computer voice data entry device into whose storage areas loaded.

3.3 Computer voice data entry devices as part of the user configuration

The CSE voice data entry devices are the processors working independently which however have the function from the viewpoint of the host computer of peripheral units.
The connection for the user configuration occurs via the genormte V.24 interface which guarantees the connection at the computers of all known manufacturers.

3.4 Quicktalk feature

In order to input the data in rapid sequence/result, the computer voice data entry devices can be equipped with a quicktalk feature. The possible input rate can be with it up to 180 words per minute.

3.5 Weitere Eigenschaften der CSE-Spracheingabegeraete

Der Benutzer kann bestimmen, dass aktuell nur eine Untermenge des Wortschatzes in den Erkennungsvorgang einbezogen werden soll. Durch diese Strukturierung des Wortschatzes wird die Dialogfuehrung des Benutzers unterstuetzt, da nur die fuer die augenblickliche Phase des Dialoges als aktiv zugelassenen Woerter vom Geraet angenommen werden.
Zusaetzlich wird beim CSE 1060 die Moeglichkeit der Datenpufferung geboten. Dabei puffert das CSE-Spracheingabegeraet mehrere Eingaben. Im Puffer kann der Benutzer Korrekturen vornehmen (z.B. Loeschen einer oder mehrerer Eingaben).

4 CSE-Geraetekonfiguration

4.1 Spracheingabegeraet CSE 1050

Das CSE 1050-Geraet besteht aus

- Mikrofon
- Operateurkonsole
- Vorverstaerker
- Sprachprozessor
- Anzeigegeraet.

Das Mikrofon ist an einem leichten Kopfbuegel befestigt (Headset-Mikrofon). Hintergrundgeraeusche werden weitgehend unterdrueckt, so dass die Spracheingabe auch in geraeuschvoller Arbeitsumwelt moeglich ist.
Die Operateurkonsole dient hauptsaechlich der Identifizierung des Sprechers beim Laden sowie im Trainingsbetrieb zum Aufbau und zur Aktualisierung des Wortschatzes.
Mit Hilfe des Vorverstaerkers kann das CSE-Geraet an die Lautstaerke des Sprechers angepasst werden.
Der Sprachprozessor ist die zentrale Komponente des CSE-Geraetes. Er fuehrt die Spracherkennung durch. Zur Kommunikation mit dem Anwendungscomputer bietet er als E/A-Interface eine V.24-Schnittstelle.
Mit Hilfe des 16- bzw. 32-Zeichen-Anzeigegeraetes werden dem Benutzer Informationen zur Unterstuetzung des Dialogs gegeben (Eingabeaufforderungen, Kontrollmeldungen, Rueckweisungen). Der Umfang der Kommunikation richtet sich nach den Beduerfnissen der jeweiligen Anwendung. Die Steuerung der Anzeige geschieht per Programm im Anwender-Computer.
Der Informationsaustausch zwischen CSE-1050-Geraet und Anwenderprogramm erfolgt ueber eine definierte Programmschnittstelle.
Ueber sie werden sowohl Steuerzeichen als auch Daten uebertragen. Die Uebertragung findet zeichenweise im ASCII-Code oder fuer den Transport des Wortschatzes in Binaerform statt.
Das Anwendungsprogramm kann folgende Funktionen des CSE-1050-Geraetes aufrufen:

- Starten und Beenden des Trainings
- Anfordern von Referenzmustern nach Trainingsabschluss zur Ablage auf einem Hintergrundspeicher
- Normierung des Geraetes
- Laden eines neuen Wortschatzes
- Bestimmung des aktiven Teiles des Wortschatzes
- Ausgabe von Text ueber das 16- bzw. 32-Zeichen-Anzeigegeraet.

Im Erkennungsmodus meldet das CSE-1050-Geraet die zum erkannten Wort vereinbarte Wortnummer bzw. ein Steuerzeichen mit der Bedeutung "Eingegebenes Wort nicht erkannt". Diese Meldungen muessen vom Anwendungs-Computer quittiert werden.

4.2 Spracheingabegeraet CSE 1060

Das CSE 1060-Geraet besteht aus

- Mikrofon
- Vorverstaerker
- Sprachprozessor
- Magnetbandkassettengeraet
- Sichtgeraet.

3.5 Further qualities/characteristics of the CSE voice data entry devices

The user can specify that only a subset of the vocabulary should be included for the present time in recognition procedure. The dialog/prompting of the user is supported by this structuring of the vocabulary, since only the words authorized for the immediate/current phase of the dialog as active are accepted by the device.
The possibility is offered additionally with the CSE 1060 to data buffering. Therewith the CSE voice data entry device buffers several inputs. in buffer could the user of corrections perform (e.g. deleting/clearing of one or several inputs).

4 Computer voice data entry device configuration

4.1 Voice data entry device CSE 1050

CSE 1050-device consists of

- Microphone
- Operator console
- Preamplifier
- Voice processor
- Display.

the microphone is at a lightweight headset attaches (headset microphone).
Background noises are suppressed to a large extent so that also voice data entry is possible in noisy working environment.
The operator console is used mainly for the identification of the speaker during loading as well as in the training for the format and for the updating of the vocabulary.
With the aid of the preamplifier, the computer voice data entry device can be adapted to the volume of the speaker.
The voice processor is the central component of the computer voice data entry device. It operates the speech recognition through. for the communication the host computer offers it as I/O interface a V.24 interface.
with the help/aid of 16 or 32 -character display are the user information for support the dialog given (prompting messages, control messages, rejections). The extent of the communication depends on the requirements of the respective application. The controller of the display occurs per program in the host computer.
The exchange of information between CSE-1050 device and user program occurs via a defined software interface.
Control character as well as data also are transmitted via them. The transmission occurs character by character in ASCII code or for the transportation/transport of the vocabulary in binary form.
Application program can call the following functions of the CSE-1050 device:

- Starting and termination of training
- Requesting reference patterns after training for the file on a secondary storage
- Resetting of the device
- Loading of a new vocabulary
- Regulation/definition of the active part of the vocabulary
- output of text 16 or 32 -character display.

in recognition mode, the CSE-1050 device reports the word number determined for the recognized word and/or a control character with meaning/importance " input word not erkannt". These messages must be acknowledged by the host computer.

4.2 Voice data entry device CSE 1060

CSE 1060-device consists of

- Microphone
- Preamplifier
- Voice processor
- Magnetic tape casette device
- Display device.

Mikrofon und Vorverstaerker entsprechen denen des CSE 1050-Geraetes. Der Sprachprozessor enthaelt ueber die Funktionen des CSE 1050-Geraetes hinaus noch das Trainingsprogramm, so dass das Training zur Erstellung des Wortschatzes unabhaengig vom Anwendungs-Computer durchgefuehrt werden kann.
Das Magnetbandkassettengeraet dient dem Sichern des Wortschatzes nach dem Training und seinem Laden von der Magnetbandkassette in den Sprachprozessor im Erkennungsbetrieb.
Ueber das Sichtgeraet, das mit einer Tastatur ausgeruestet ist, werden die Begriffe des Wortschatzes und die zugehoerigen Ausgabezeichenstrings vereinbart und ausserdem die Benutzerfuehrung waehrend des Trainings realisiert.
Die Programmschnittstelle zum Anwendungs-Computer ist beim CSE 1060-Geraet einfach und damit sehr benutzerfreundlich. Nach Erkennen eines Wortes wird der vereinbarte Zeichenstring vom CSE-Geraet uebergeben. Die Uebernahme durch das Anwendungsprogramm erfolgt wie bei einer Eingabe z.B. durch ein Datensichtgeraet.

5 Betrieb der CSE-Geraete

Die CSE-Geraete bieten eine sehr hohe Erkennungsgenauigkeit und damit eine groessere Sicherheit bei der Datenerfassung gegenueber den herkoemmlichen Erfassungsmethoden.
Bei "kooperativem" Verhalten des Sprechers wird eine Erkennungssicherheit von mehr als 99% erreicht. Die folgenden Benutzungsregeln dienen diesem Ziel:

- richtige Auswahl der Woerter des Wortschatzes
- sorgfaeltiges Training
- deutliches Sprechen im Erkennungsbetrieb.

5.1 Auswahl der Woerter fuer den Wortschatz

Die richtige Auswahl der Woerter fuer den Wortschatz verhindert die Gefahr von Verwechslungen (Substitutionen) im Erkennungsbetrieb. Die Stimme eines Sprechers kann sich taeglich durch verschiedene Einfluesse veraendern. Diese Schwankungen muessen jedoch vom Spracheingabegeraet aufgefangen werden, ohne dass ein erneutes Training notwendig ist. Deshalb ist es sinnvoll, bei der Auswahl der Woerter des Wortschatzes darauf zu achten, dass sie sich in ihrer Aussprache deutlich voneinander unterscheiden. Als Beispiel sei die Ziffer "2" genannt, die man "Zwo" aussprechen sollte, um sie ausreichend gegen die "Drei" abzugrenzen. Dies wird vor allem dann verstaendlich, wenn man beruecksichtigt, dass beim Erkennungsvorgang den Vokalen in den Woertern besonders hoher Informationswert beigemessen wird.

5.2 Gestaltung des Trainings

Sprecheradaptive Geraete erfordern ein Training durch den jeweiligen Sprecher fuer dessen spezifischen Wortschatz. Das Training soll unter denselben Bedingungen erfolgen, wie sie im spaeteren Erkennungsbetrieb anzutreffen sein werden.
Dies gilt sowohl fuer die Arbeitsumwelt als auch fuer die von der Person abhaengigen Spracheigenarten. Der Sprecher sollte es nutzen, dass beim Training jedes Wort mehrfach eingegeben werden kann, um die Aussprache z.B. hinsichtlich der Betonung zu modifizieren, so dass moeglichst viele Varianten fuer den spaeteren praktischen Betrieb erfasst werden.
Bei Betrieb des Geraetes kann es sich ergeben, dass einzelne Woerter in Bezug auf Rueckweisung oder Substitution besonders kritisch sind. In diesen Faellen ist es nicht notwendig, den gesamten Wortschatz neu zu trainieren. Es genuegt in einem Nachtraining das oder die einzelnen kritischen Woerter neu einzugeben.

5.3 Sprachverhalten im Erkennungsbetrieb

Die Kriterien fuer das richtige Sprachverhalten sind die Eingabegeschwindigkeit und die Deutlichkeit der Aussprache. Unterschiedliche Lautstaerke und Geschwindigkeit beim Sprechen des einzelnen Wortes gleicht das CSE-Geraet aus. Die Eingabegeschwindigkeit jedoch, darunter ist das Tempo bei der aufeinanderfolgenden Eingabe mehrerer Woerter zu verstehen, wird vom Sprecher bestimmt. Er wird durch eine optische Anzeige, die ihm die Eingabebereitschaft signalisiert, unterstuetzt. Erfahrungsgemaess ermoeglicht dieses Hilfsmittel dem Sprecher sehr rasch die optimale Eingabegeschwindigkeit zu erreichen.
Bei Verwendung der Quicktalk-Zusatzeinrichtung koennen mehrere Woerter nacheinander eingegeben werden, ohne auf das Eingabebereitschaftssignal achten zu muessen.

Microphone and preamplifier correspond to those CSE of 1050-device. The voice processor includes the training program outward via the functions of CSE 1050-device still so that the training for the generation of the vocabulary can be executed independently of the host computer.
The magnetic tape casette device is used for the storing of the vocabulary after training and its loading of the magnetic tape cassette into the voice processor in recognition mode.
via the terms of the vocabulary and the corresponding output character strings are the display device which is equipped with a keyboard determined und implements in addition user prompting during training.
the software interface to the host computer, 1060-device is single with CSE und with it very user-friendlyally. after recognizing a word is the determined character string deliver of the computer voice data entry device. The transfer through application program occurs as during an input e.g. through a data display terminal.

5 Operation of the computer voice data entry devices

the computer voice data entry devices offer a very high recognition accuracy and with it a larger security during data acquisition as compared to conventional acquisition methods.
with the " cooperative " behaviour of the speaker is a reading accuracy of more than reaches 99%. The following usage rules serve this target/destination:

- Proper/correct selection of the words of the vocabulary
- Careful training
- Clear pronunciation in recognition mode.

5.1 Selection of the words for the vocabulary

The proper/correct selection of the words for the vocabulary prevents the danger of the substitutions (substitutions) in recognition mode. The voice of a speaker can change daily through different influences. However these variations must be compensated by the voice data entry device, without a further training is necessary. Therefore it is meaningful/user-friendly to pay attentione during the selection of the words of the vocabulary on it that they differ clearly in its pronunciation from-one-another/from-each-other. could be called "2" as an example the digit, the one "Zwo" should have pronounce, around it to separate being from those " three ". this is then intelligible in front of each, if one considers that during recognition procedure, especially high value of information is attributed to the vowels in the words.

5.2 Organization of training

speaker-dependent devices require a training through the respective speaker for whose specific vocabulary. training should occur among the same conditions/requirements, as them in later recognition mode to finde will be.
This applies for working environment as well as for peculiarities of speech dependent on the person. the speaker it sollte be useful that during training, every word can be entered repeatedly, modify around the pronunciation for example with regard to accentuation to, so that inasmuch as possible many variants for the later practical operation are registered.
With operation of the device, it can result that the individual words with regard to rejection or substitution are especially critical. In these cases, it is not necessary to train the entire vocabulary again. It suffices in a training update to input that or the individual critical words again.

5.3 Speaking in recognition mode

The criteria for proper/correct speaking are the input rate and the clearness of pronunciation. Different volume and speed during the pronunciation of the individual word equalizes the computer voice data entry device. the input rate jedoch, under that is speed to understande during the consecutive input of several words, is specified of the speaker . It is supported by an optical display which signals availability to it. This aid makes possible those from experience for the speaker very rapidly to reach optimal input rate.
during use of the quicktalk feature, several words can be entered in sequence, without to have to pay attention for the ready signal.

Ein nicht mit ausreichender Deutlichkeit ausgesprochenes Wort wird vom CSE-Geraet zurueckgewiesen. Die Rueckweisung wird optisch und ggfs. akustisch angezeigt, so dass der Benutzer eine erneute Eingabe veranlassen kann.

6 Einsatzmoeglichkeiten fuer die CSE-Geraete

Neben oekonomischen Gesichtspunkten spielen bei der Spracherkennung auch ergonomische Ueberlegungen eine wichtige Rolle. So gewinnt das Argument der Anpassung der technischen Geraete an die menschlichen Kommunikationsgewohnheiten in der Diskussion um die Wahl der geeigneten Datenerfassungsmethoden zunehmend an Bedeutung.

Die Spracheingabe bringt viele Vorteile gegenueber den traditionellen Eingabemoeglichkeiten. Da das gesprochene Wort direkt zur Daten- oder Kommandoeingabe verwendet werden kann, werden Taetigkeiten des Ablesens und der wiederholten Uebertragung von Daten mit ihren negativen Effekten der Ermuedung und der nachlassenden Konzentration beseitigt. Dadurch wird die Fehlerrate auch bei ungeuebtem Personal herabgesetzt. Taetigkeiten, bei denen beide Haende benutzt werden muessen, der Sichtkontakt zur Datenquelle erhalten bleiben muss oder Arbeiten in schmutziger Umgebung stehen einer Dateneingabe nicht mehr hindernd im Wege. Die Erfahrungen beim bisherigen Einsatz der Geraete haben gezeigt, dass es eine Vielzahl von Einsatzmoeglichkeiten in verschiedenen Anwendungsgebieten fuer die Spracheingabe gibt:

- Qualitaetskontrolle
- Lagerverwaltung
- Inventuraufnahme
- Warenein- und -ausgangskontrolle
- Betriebsdatenerfassung
- NC-Programmierung
- Anlagensteuerung
- Steuerung von Industrierobotern
- Hilfe fuer Behinderte
  - Steuerung von Geraeten
  - Behinderten-Programmierplatz
- Grafischer Arbeitsplatz
  - Erfassung kartografischer Daten
  - CAD
- Photogrammetrie
- Labordatenerfassung
- Roentgenbildauswertung
- Fluggepaecksortierung
- Paketsortierung
- Militaerische Fuehrungssysteme
- Nachrichtentechnik

A word pronounced not with sufficient clearness is refused by the computer voice data entry device. Rejection is indicated optically and where appropriate acoustically so that the user can initiate a further input.

6 possible applications for the computer voice data entry devices

Next to economic aspects, also ergonomic considerations play an important role during speech recognition. Then the argument of the adaptation of the technical devices becomes important increasingly at the human habits of communication in the discussion at the choice of the suitabl data acquisition methods.

Voice data entry has advantages. Since the spoken word can be used directly for data or command input, the activities of reading and the repeated transmission of the data with their negative effects of tiredness and the declining concentration are corrected. Thereby the error rate is reduced also with inexperienced users. The activities with which both hands must be used whose sight maintained had to be for the data source, or working in dirty environment are a data input no longer hindering in the way/path. The experiences during the previous application of the devices have indicated that there is a large number of the possible applications in the different fields for voice data entry:

- Quality control
- Warehouse management
- Stock-taking
- Stock control
- Production data acquisition
- NC-programming
- System control
- controller of industrial robots
- help/aid for disabled person
  - controller of devices
  - Programming console for disabled persons
- Graphic work station
  - Acquisition of cartographic data
  - CAD
- Photogrammetrie
- Acquisition of laboratory data
- Interpretation of radiography
- Flight luggage sorting
- Parcel sorting
- Military management information systems
- Communications