N93-72615

[redacted]

RESEARCH ON SPEECH UNDERSTANDING AND RELATED AREAS AT SRI

DONALD E. WALKER

SRI INTERNATIONAL
MENLO PARK, CALIFORNIA

S3-32
176339

I INTRODUCTION

SRI International has a long history of research on natural language and on speech. The groups working in these areas were brought together specifically to work on the development of a speech understanding system, but their activities range much more broadly. As a result, SRI is qualified to engage in a variety of projects relating to voice technology for interactive systems applications:

- The design and development of speech understanding systems varying widely in complexity and context of application.

- Research on syntax, semantics, and discourse as they relate to speech recognition and speech understanding systems.

- The integration of practical natural language interface capabilities into systems for speech recognition and voice control.

- Acoustic-phonetic research.

- The development of procedures for speech analysis and speech synthesis.

- The evaluation of the intelligibility and quality of speech, both human and computer-generated, and of communication systems that carry it.

- The identification of properties of speech that contribute to specific qualities, such as "naturalness" and talker identification, and the development of computer procedures for using these properties.

- Studies of the effects on speech of abnormal physiological and psychological states and the development of voice analysis algorithms to detect those effects.

- The conduct of experiments to study the relationships among parameters like the quality of computer speech and computer understanding, the effectiveness of task performance, and the psychological and physiological states of the users.

The technical review of previous work in Section II will concentrate exclusively on our research on speech understanding. It will be followed in Section III by a brief description of our current capabilities for research on speech understanding, speech recognition, and voice control. Section IV will present relevant current research activities; although some of these activities involve text input rather than speech, it would be possible to adapt them to voice control.

II. TECHNICAL REVIEW OF PREVIOUS WORK

A. Introduction

From 1971 to 1976, SRI International participated in a major program of research on the analysis of continuous speech by computer sponsored by the Advanced Research Projects Agency of the Department of Defense.* The goal was the development of a speech understanding system capable of engaging a human operator in a natural conversation concerning a specific task domain (see Newell et al., 1973). A rather complex set of specifications defined the parameters more precisely. The program culminated in the demonstration of a system that did meet the target specifications (see Reddy et al., 1976; Medress et al., 1977). However, more important for the future of this technology are developments in the various constituents or sources of knowledge used in the systems--particularly phonetics, phonology, syntax, semantics, and discourse--and in the system architecture necessary for coordinating them efficiently and effectively.

At SRI, we have made signigicant advances in the development both of the components that provide knowledge for use in a speech understanding system and of a framework for coordinating and controlling them. Our work in the ARPA Program was conducted in two phases. During the first phase, we were responsible for the entire system. During the second phase, we worked cooperatively with the System Development Corporation (SDC). For this joint system development effort, SRI provided capabilities for system organization and control, syntax, semantics, and discourse analysis; SDC provided capabilities for signal processing, acoustics, phonetics, and phonology. In this paper, only the SRI work is considered. In the following description, we discuss first our more recent work, since it represents the latest results of our research on

----

*This research was funded under the following ARPA contracts, all administered through the Army Research Office: DAHCO4-72-C-0009, DAHCO4-75-C-0006, and DAAG29-76-C-0011.

speech understanding, and since we have conducted experiments that have enabled us to provide a partial evaluation of its effectiveness. The subsequent presentation of our earlier efforts considers only the acoustic processing components; it is included to illustrate the research we have done in the speech sciences in the context of speech understanding.

B. Recent Research on Speech Understanding

1. Introduction

Our research on speech understanding has been designed specifically to handle naturally occurring speech, conversations that would take place as a person uses the system on a regular basis as an adjunct to his regular technical activities. A distinctive characteristic of our approach is its emphasis on the relevance of contributions from computational linguistics and artificial intelligence. In processing ordinary conversational dialog, the various sources of acoustic uncertainty combine with the large number of linguistic choices to create an extremely large number of alternative hypotheses that must be considered during the interpretation of an utterance. To control the combinatorial explosion and limit the number of choices that has to be considered, we have introduced sophisticated components for combining information about the structure of English sentences (syntactic knowledge), about the task being considered (semantic knowledge), and about previous utterances in the dialog (dicourse knowledge). To cope with the added complexity of these extra sources of knowledge, we have provided special procedures for coordinating their interactions.

The syntactic component of our speech understanding system is a performance grammar; it describes the syntax of the English occurring in spontaneous dialog rather than the English of edited text. Semantic knowledge about the task domain is encoded in a partitioned semantic network. Partitioning the network allows us, among other things, to represent multiple alternative parses without using excessive storage and to associate syntactic units directly with their semantic counterparts. The discourse component uses the context of the preceding dialog to identify the entities referred to by pronouns and definite noun phrases and to expand incomplete (elliptical) utterances.

Our approach to the coordination of these knowledge sources, and those containing acoustic, phonetic, and phonological information, stresses integration--the process of forming a unified system out of a collection of components--and control--the dynamic direction of the overall activity of the system during the processing of an input utterance. Our approach to integration

- Allows specifying the interactions of information from various sources of knowledge in a procedural representation.

- Provides a means for adjusting the language accepted as input for different tasks without loss of generality.

- Avoids commitment to a particular system control strategy.

Our approach to system control

- Allows processing an input left-to-right, right-to-left, or from the middle out.

- Enables combining top-down, predictive procedures, with bottom-up, data-directed procedures.

- Allows evaluating partial results (phrases) within the larger linguistic contexts (sentences) in which they could be embedded.

A review of the total project is beyond the scope of this paper. After discussing the task domain and presenting an overview of the operation of the system to provide context, we will consider each of the system knowledge sources together with discussions of a facility for language definition that provides the basis for coordinating them and of the executive routines that control them. A brief statement on the results of our experiments with alternative system control strategies also is included.

A more complete statement of this work is contained in our final project report (Walker, 1976). A somewhat expanded description of the language definition system and executive and of the experiments conducted to test them is presented in Paxton (1977; see also Walker and Paxton et al., 1977). The discourse component is treated more fully in Grosz (1977a; see also 1977b for a discussion of the concept of focus). Fikes and Hendrix (1977) summarize the scheme for semantic representation and the procedures for deductive retrieval used in the system. References to other papers are included in the final project report.

2. The Task Domain

The domain of discourse for the speech understanding system is defined by a data base of information about the ships of the U.S., Soviet, and British fleets. The system data base contains such characteristics as owner, builder, size, and speed for several hundred ships. Utterances can be formulated that relate to attributes of a particular ship or of ships meeting a certain description; to part-subpart relations between a ship and, for example, its crew; to set membership and kind relationships between various individuals and classes (such as "all ships" and "Are all ships diesels?"). It is possible to specify an object on the basis of its properties ("What country owns the Skate?";

"What American destroyer has a speed of 33 knots?") or of the number of individuals meeting a given description ("How many diesel submarines are owned by the U.S.?"). Queries may be quantified to seek information over classes of individuals ("What is the speed of each American sub?"). Dialog sequences can be processed, with previous utterances serving as context so that pronouns can be used, the referents of determined noun phrases can be identified, and it is not necessary to use complete utterances if the reference is clear ("What is the speed of the Lafayette?"; "The Ethan Allen?"; "Do both ships belong to the U.S.?"; "Are they both submarines?").

3. The Operation of the System

When a speaker records an utterance, it is analyzed acoustically and phonetically, and the results are stored in a file. When these data are available, the executive begins to predict words and phrases, guided by the rules for phrase formation in the language definition, and to build up phrases from words that have been identified acoustically in the utterance. When a word is predicted at a specified place in the utterance, alternative phonological forms of that word are mapped onto the acoustic data for that place, and a score indicating the degree of correspondence is returned. As each phrase is constructed, relevant semantic and discourse information is checked, and if appropriate, a semantic network representation of the phrase is developed. When an interpretation for the entire utterance is complete, relevant structures from the semantic model of the domain and from an associated relational data base are processed to identify in semantic network form the content of an appropriate response. This response is then generated either in text form or through the use of a speech synthesizer.

4. The Language Definition

The input language is a subset of natural, colloquial English that is suitable for carrying on a dialog between a user and the system regarding information in the data base. The definition of this language consists of a lexicon containing the vocabulary and a set of composition rules for combining words and phrases into larger phrases. This language definition is translated by a definition compiler into an efficient internal representation, which is used by the executive to process an utterance. The lexicon is separated into categories, such as noun and verb, and the words in each category are assigned values for various attributes, such as particular grammatical features and semantic representations. The composition rules are phrase-structure rules augmented by a procedure that is executed whenever the rule constructs a phrase. Information provided by the procedure includes both attributes of the phrase based on the attributes of its constituents, and factors for use in judging the acceptability of the phrase.

An attribute statement may compute values that specify acoustic properties related to the input signal, syntactic properties such as mood (declarative or interrogative) and number (singular or plural), semantic properties such as the semantic network representation of the meaning of the phrase, and discourse properties such as the entity a pronoun refers to. The values of constituent attributes are used in computing the attributes of larger phrases, and the attributes of complete interpretations are used in generating responses.

The factor statements compute acceptability ratings for an instance of the phrase. The factors are non-Boolean; that is, they may assume a wide range of values. As a result, a proposed instance of a phrase is not necessarily simply accepted or rejected; it may be rated as more or less acceptable, depending on a combination of factor values. Like attributes, factors may be acoustic, syntactic, semantic, or discourse related. Acoustic factors reflect how well the words match the actual input; syntactic factors deal with tests like number agreement between various constituents; semantic factors assure that the phrase has a meaning in the task domain; and discourse factors indicate whether a pronoun or definite noun phrase makes sense in the given dialog context. The values of factors are included in a composite score for the phrase. The scores for constituents are combined with the factor scores to produce the scores of larger phrases, and the scores of complete interpretations are used in setting executive priorities.

The attribute and factor statements in the procedural parts of the rules contain specifications for most of the potential interactions among system components. The form of the rules is designed to avoid commitments to particular system control strategies. For example, the rule procedures can be executed with any subset of constituents, so incomplete phrases can be constructed to provide intermediate results, and it is not necessary to acquire constituents in a strictly left-to-right order.

5. Syntax

The syntactic knowledge in the system is represented both in the phrase structure part of the language definition rules and in the attribute and factor statements in the procedure part of the rules. Syntax provides computationally inexpensive information about which words or phrases may combine and how well they go together. In testing word or phrase combinations, syntactic information alone often can reject an incorrect phrase without requiring costly semantic and discourse analysis. Factors are used for traditional syntactic tests, such as agreement for person or number, but factors also are used to reduce the scores of unlikely phrases. For example, questions that are negative (e.g., "What submarine doesn't the U.S. own?") are not likely to occur. A factor statement lowers the value for this interpretation but does not eliminate it completely, so that if no better hypothesis can be formed to account for the input utterance, this interpretation will be accepted. Since the language definition system provides the capability for evaluating phrases in context by means of non-Boolean factors, the grammar can be tuned to particular discourse situations and language users simply by adjusting factors that enhance or diminish the acceptability of particular interpretations. It is not necessary to rewrite the language definition for each new domain.

6. Semantics

The system's knowledge about the task domain is embodied in a partitioned semantic network. A semantic network consists of a collection of nodes and arcs where each node represents an object (a physical object, situation, event, set, or the like) and each arc represents a binary relation. The network model of the task serves as a foundation on which the structures corresponding to new utterances are built. It is used to assess the feasibility of combining utterance constituents to form larger phrases. And it is a source of information for answering queries, supplemented by a relational data base, which can be accessed directly from the network.

The structure of our semantic networks differs from that of conventional networks in that nodes and arcs are partitioned into spaces. These spaces, playing in networks a role roughly analogous to that played by parentheses in logical notation, group information into bundles that help to condense and organize the network's knowledge. Network partitioning serves a variety of purposes in the speech understanding system:

- Encoding logical connectives and higher-order predicates, especially quantifiers.

- Associating syntactic units with their network images.

- Interrelating new inputs with previous network knowledge while maintaining a definite boundary between the new and the old.

- Simultaneously encoding in one network structure multiple hypotheses concerning alternative incorporations of a given constituent into larger phrases.

- Sharing network representations among competing hypotheses.

- Maintaining intermediate results during the question-answering process.

- Defining hierarchies of local contexts for discourse analysis.

7. Discourse

The discourse knowledge in the speech understanding system is used to relate a given utterance (or a portion of it) to the overall dialog context and to entities and structures in the domain. The procedures we have developed are based on systematic studies of dialogs between two people performing some activity together. Contextual influences were found to operate on two different levels in a discourse. The global context--the total discourse and situational setting--provides one set of constraints on the interpretation of an utterance. These constraints are used in identifying the referents of pronouns and definite noun phrases. The second set of constraints is provided by the immediate context of closely preceding utterances. These constraints are used to expand utterance fragments into complete utterances. Since the task domain of the system is data base retrieval, the discourse context is limited to a linear history of preceding interactions. For complex task-oriented dialogs, the linear discourse history can be replaced by a more structured history related to the organization of the task being performed.

8. Deduction

Along with the ability to represent entities and their interrelationships in a task domain, it is necessary to reason about them. Thus, the system also contains an inference mechanism for retrieving information from the semantic network. This mechanism serves a dual purpose: (1) during the interpretation of an utterance, it supplies information needed to produce the appropriate semantic structure corresponding to each phrase and to relate it to the dialog context; (2) after an interpretation has been found for a question, it is used to find an answer. This inference capability can retrieve information explicitly stored in the networks, can derive information using general statements, or theorems, in the network, and can invoke user-supplied functions to obtain information from knowledge sources other than the network, such as data files.

9. Generation

We also have developed the capability of generating, as a response from the system, an English phrase or sentence that corresponds to a semantic network substructure. This substructure usually is the answer to a question asked by the user. Words and phrases are chosen to express the semantic content; a syntactic frame for their organization is selected; and the response is expressed in text form, although we have sometimes used a commercial speech synthesizer to produce a spoken output.

10. Executive

The executive has three main responsibilities:

- It coordinates the work of the other parts of the system calling acoustic processes and applying language definition rules.
- It assigns priorities to the various tasks in the system.
- It organizes hypotheses and results so that information common to alternative hypotheses is shared, avoiding duplication of effort.

When a successful interpretation has been found, the executive invokes the response functions, which produce a reply.

The principal data structure used by the executive is called the parse net. It is a network with two types of nodes: phrases and predictions. Phrases are built from words or from smaller phrases by applying composition rules from the language definition. Phrases can be complete, containing all their constituents, or incomplete, with some or all of their constituents missing. A prediction is for a particular category of phrase associated with a particular location in the utterance. As the interpretation of an utterance progresses, new phrases that have been constructed from existing phrases or from words found in the utterance are added to the parse net. At the same time, new predictions are made as more information is obtained. Thus, as the interpretation process advances, the parse net, which holds intermediate hypotheses and results, grows. A complete root category phrase (usually a sentence) with its attributes and factors constitutes an interpretation of the utterance.

There are two tasks entailed in maintaining and evolving this parse net: the word task and the predict task. The role of the word task is to look for a particular word in a particular location in the utterance. If the acoustic mapper has not been called previously for that word in that location, the word task calls it. If a word is found successfully in the specified location, the word is used to build new phrases. The role of the predict task is to make a prediction for a word or phrase that can help complete an incomplete phrase. Whenever a new constituent is inserted into an incomplete phrase, any adjacent constituents that had been missing can be predicted. New predictions can include predictions for particular words, leading to new instances of calls on the word task.

Establishing the priority of a task begins with determining the score of the phrase involved. The score is computed from the results of the acoustic mapping of any of the words contained in the phrase, from the factor statements for the phrase, and from the scores of the constituents. The score is thus a local, context-free piece of information about how good the phrase is. After the score is determined, the phrase is given a rating that is an estimate of the best score for a phrase of the root (sentence) category that uses the given phrase. The rating for a phrase does depend on the other phrases in which it may be embedded to form a sentence. This rating is then modified depending on the control strategy being used, and the result is the priority of the task to be performed for that phrase.

Both the word and the predict task can work either left-to-right through an input or bidirectionally from words selected at arbitrary positions within an utterance. This ability to add constituents to phrases in any order has made it possible to experiment with a variety of control strategies. Also important for experimental studies is the fact that each task does a limited amount of processing and then stops after scheduling further operations for later. The scheduling does not specify a particular time, but instead gives each operation a certain priority. The operation is performed when its priority is highest. Since the executive sets the task priorities, changing the way these priorities are set alters the overall system strategy.

11. Experimental Results

Loss of the computer facility at the System Development Corporation shortly after the system was implemented prevented extensive exercising of the complete system with the acoustic processing components. However, using a simulation of those components, we were able to perform a variety of experiments to analyze the effect of variations in control strategy on system performance. We used an analysis of variance procedure to study four variables:

- To check context or not: use the effects of sentential context based on attribute and factor information in setting priorities versus using only constituent structure information. Context checking should provide more information for setting priorities and should lead to better predictions, but it could prove costly and result in poorer performance.

- To island drive or not: go in both directions from arbitrary starting points in the input versus proceeding strictly left to right from the beginning. Island driving allows interpretations to be built up around words that match well anywhere in the input, but the process is more complex.

- To map all or one: test all the words at once at a given location versus trying them one at a time and delaying further testing when a good match is found. Mapping all at once identifies the best acoustic candidates and reduces the chances of following false paths, but it takes substantially more time.

- To focus or not: assign priorities for tasks focusing on selected alternatives by inhibiting competion versus proceeding each time with the task with the highest score. Focusing prevents frequent switching among alternatives, but it may result in continuing along false paths.

All combinations of the four control-strategy variables were tested on 60 sentences that varied in length, vocabulary, and sentence type.

The results of most interest are those relating to the effects of context checking, that is, using the attribute and factor information. Significant increases in accuracy were found; there was a higher percentage of utterances for which the correct sequence of words was found. Fewer phrases were constructed, so there was less work for the system to do. Rule factors blocked 27% of the attempts on the average; and whereas the average number of phrases constructed over all system configurations was 267, the most accurate system with context checking averaged 158. The percentage of incorrectly identified words was reduced; there was a lower priority for looking at words adjacent to such false alarms than there was for looking at words adjacent to correct words. Finally, the total processing time was reduced, in spite of the extra executive processing required.

These experiments did not provide unequivocal data on how the system would perform with actual rather than simulated acoustic processing components. However, for a lexicon of over 300 words, the most accurate system configuration identified 73% of the utterances. If minor errors that would have no effect on the response of the system are ignored, the figure is increased to 82%. Modifications in the executive alone could increase this latter figure to 90%. Improvements in the acoustic processing components, which the loss of the SDC computer never allowed time to refine, could be expected to increase this figure further. For example, a 7% downward shift in the distribution of scores for words incorrectly accepted by the acoustics would result in a 13% increase in accuracy.

Much more work would be necessary to provide a comprehensive evaluation of our research on speech understanding. However, it is clear that we have produced system control concepts and a set of system components that are well-suited for further research on unconstrained naturally occuring speech.

C. SRI Research on Acoustic Processing for Speech Understanding

Our earlier work on acoustic processing for speech understanding was conducted in the context of a system design concept that was similar to but simpler than the one described above. The control strategy was exclusively top down; that is, syntactic and semantic information relevant for the current discourse context was used to predict the set of words that could possibly occur at a given place in the utterance. Using data derived from a speech analysis subsystem, a word verification subsystem determined for each proposed word: (1) the confidence that the proposed word did in fact exist at the specified place in the utterance, and, if it could be present, (2) where the word began and ended. The parser, in this version of the system, proceeded through the utterance from left-to-right according to a search strategy that kept track of all possible paths, at any particular moment following the one with the highest priority.

The speech analysis subsystem classified each 10-ms portion of the digitized signal into one of ten classes based on a classification algorithm using digital filter information. The classes were chosen because they would give reliable information in a context-free manner. In addition, a linear predictive coding (LPC) analysis of the voiced intervals provided frequency and bandwidth information for the first five formants. All of the acoustic data in this preprocessing step were stored for each utterance.

The word verification subsystem consisted of a set of algorithms representing the words in the vocabulary. Each such word function was prepared after a detailed examination of acoustic data for that word in selected contexts from a variety of utterances. A word function consisted of a series of Fortran subroutines that used data from a variety of sources: the acoustic preprocessing of the utterance; algorithms for level (volume) detection, formant smoothing, detecting formant discontinuities, fitting formant trajectories, and identifying formant bandwidths; and specially designed digital filters or LPC analyses.

The system that incorporated these acoustic components was not tested extensively, so no conclusions can be made regarding its performance. Of 71 utterances processed by the system, 62% were understood correctly, 10% misunderstood, and 28% not understood at all. We were encouraged by the results of these early efforts, and the experiences influenced our subsequent work.

III. CURRENT CAPABILITIES FOR SPEECH UNDERSTANDING RESEARCH

A. Facilities Available

The major computer facility used for our research on speech understanding was a Digital Equipment Corporation PDP/KA-10. It provided time-shared computing capabilities supporting a large variety of programming languages, LISP and Fortran being the ones used most frequently. Currently, SRI has a DEC PDP/KL-10 (System 1090T), which is a larger, faster computer with similar characteristics; it is being used in most of the projects described under Current Research Activities in the following section.

For our early acoustic research, we developed a very powerful interactive speech analysis system. This system, based upon a Vector-General Display controlled by a DEC PDP-15 connected to the PDP/KA-10 computer, allowed scientists to digitize speech, present speech both aurally and visually, edit and mark time series, calculate and display Fourier transforms and LPC analyses of selected portions of speech, calculate and display the results of classification algorithms, plot Formant trajectories, etc. The system was the major tool in the development of the acoustic-phonetic analysis algorithms used in the speech understanding system.

The speech analysis system currently is being upgraded to employ a Hughes Conographics Display controlled by a PDP-11/40 connected over the ARPANET to the PDP/KL-10 computer. The PDP-11 also is connected to an SPS-41 fast array processor that provides real-time calculations of complex speech algorithms such as LPC spectral analysis.

Complementing this system is a PDP-11-controlled psychophysiological laboratory with facilities for digitizing and recording 64 channels of voice and electrophysiological data, including beat-by-beat heart rate, skin conductance response, peripheral pulse volume, respiration rate, and electroencephalographic and electromyographic data from as many as eight subjects simultaneously. We also have a PDP-11-controlled psychophysics laboratory that is used for automated presentation of auditory and/or visual stimuli to subjects and automated recording of responses. Both of these PDP-11 computers are connected to the PDP/KL-10 computer, so that data can be analyzed on the time-shared system.

A Threshold Technology VIP-100 system, which is interfaced to a PDP-11, provides capabilities for isolated word and phrase recognition. We also have a Federal Screw Works VOTRAX ML-I Multi-Lingual Voice System for synthesizing speech.

B. Personnel

Computational Linguistics and Artificial Intelligence:

Barbara J. Grosz--natural language understanding, discourse analysis, knowledge representation

Gary G. Hendrix--natural language semantics, knowledge representation, semantic network architecture, practical natural language interfaces

Jerry R. Hobbs--text processing, natural language semantics

Gordon S. Novak--question-answering systems, data-base semantics

Ann E. Robinson--language understanding systems, semantic representation and problem solving

Jane J. Robinson--syntax, semantics, phonology, discourse, case and performance grammars, prosodics

Earl D. Sacerdoti--natural language systems for data access, decision aids for command and control

Jonathan Slocum--language generation, semantic network architecture, syntax, semantics, and case systems

Donald E. Walker--language understanding systems, natural language systems for data access, text processing

Speech Sciences:

Richard W. Becker--acoustic-phonetics, speech and speaker recongition by computer, design of large-scale interactive computer systems for speech analysis

Earl J. Craighill--integrated data and voice communication networks, interactive graphic display programming, application of packet radio technology to command and control

Michael H. Hecker--acoustic-phonetics, speech and speaker recognition by computer, forensic applications of speaker identification, effects of pathologies on speech

Fausto Poza--acoustic-phonetics, speech and speaker recognition by computer forensic applications of speaker identification, effects of physiological states on speech

James R. Young--speech and speaker recognition by computer, speech signal analysis and signal processing

IV. RELEVANT CURRENT RESEARCH ACTIVITIES

A. Natural Language Understanding Using Text Input

Under ARPA support (Contract DAAG29-76-C-0012), we are providing natural language capabilities in a Navy command and control context. The objective is to develop the technology needed to support a series of increasingly sophisticated systems that provide natural language access to multiple data base management systems over the ARPANET in real time. Each system in the series accepts natural language questions about the data--currently in text form, plans a sequence of appropriate queries to the data base management system to answer each question, determines on which computer to execute the queries, establishes links to those machines over the ARPANET, monitors prosecution of the queries, recovers from certain errors in execution, and prepares a relevant answer to the original question.

Under National Science Foundation support (Grant MCS76-22004), we are developing natural language capabilities for use in intelligent systems that can function as experts, advising and supporting human efforts over a range of problem areas. The objective of the research is to define formally the knowledge necessary for effective communication in natural language between a person and a computer, when they are cooperating on a shared task. Our major emphasis in the project are: (1) the investigation of the structure of dialogs about a task, and (2) the use of the contexts provided by the dialog and the task as aids in understanding utterances. Our activities center on the development of representations for the various kinds of knowledge necessary for understanding utterances and on the development of effective computational procedures for using that knowledge to interpret a sequence of such utterances in a dialog.

A distinctive feature of the project is its concern with understanding the language that occurs in dialogs which take place in a dynamically changing environment. Most other current research either analyzes independent questions or statements within a static environment, as in information retrieval from a computer data base, or considers narratives rather than dialogs, as in story understanding. In contrast, we are interpreting a coherent dialog in relation to an ongoing or previously executed task in which the context can be continually changing. Capabilities are being developed for representing structural features of dialogs and tasks and for dealing explicitly with utterances that relate to past and future, as well as present, and to hypothetical, as well as actual, conditions. Attainment of the goals of this research is essential for the development of intelligent systems that can function as experts, advising and supporting human efforts over a critical range of problems.

B. Practical Natural Language Interfaces with Text Input

Under SRI Internal Research and Development support, we have been developing and testing LIFER (Hendrix, 1977), a practical system for creating English language interfaces to other computer software (such as data base management systems and expert consultant programs). Its purpose is to make the competence of other computing systems more readily accessible by overcoming the language barriers separating these systems from potential users. Emphasizing human engineering, LIFER has bundled natural language specification and parsing technology into a single package, which includes an automatic facility for handling inputs that do not form complete sentences, a spelling corrector, a grammar editor, and a mechanism that allows even novices, through the use of paraphrase, to extend the language recognized by the system. Offering a range of capabilities that supports both simple and complex interfaces, LIFER allows beginning interface builders to rapidly create workable systems and gives ambitious builders the tools needed to produce powerful and efficient language definitions. Experience with LIFER has shown that for some applications, very comfortable interfaces can be created in a matter of days. The resulting systems are directly usable by such people as business executives, office workers, and military officials whose areas of expertise are outside of computer science. The initial system developed for the ARPA project, referenced above, used LIFER. Other applications provide access to a medical data base and to an interactive photointerpretation system.

C. Speech-Related Research at SRI

The acoustic facilities at SRI are being used in a variety of research projects. Using the first version of our interactive speech analysis system, we developed a Semi-Automatic Voice Verification System for the Law Enforcement and Administration Agency (Grant NI 71-078-G) that is currently being put into operation. The new speech analysis system, while not yet complete, has already been used in an ARPA project (Contract N00039-76-C-0363) to simulate a Packet Switched Speech Network in a study of the effects of varying system parameters, such as delay and loss of packets, on the efficiency of two-person communication. Under U.S. Government support (Contract 10123-6281770047-7WR) the facilities of the psychophysiological laboratory are being employed to obtain voice and physiological data on 150 subjects to form a data base that can be used to relate speech characteristics to the physiological state of a person in various situations. Within the psychophysics laboratory, we have investigated the effects of phase in human hearing and are currently evaluating the intelligibility and qualtiy of various kinds of machine-processed speech under Defense Communications Agency support (Contract DCA 160-77-C-004).

D. Practical Uses of Voice Control in Industrial Automation

Under NSF support (currently, Grant APR75-13074), we have been conducting exploratory research into advanced automation. The object of the project is to develop a programmable and adaptable computer-controlled system of manipulators, end-effectors, and contact or non-contact sensors that can be easily trained to perform material handling, inspection, and assembly tasks of the kind that are encountered in industrial settings. The VIP-100 provides voice control to guide a Unimate manipulator in this process. For example, to establish a particular fastening operation the operator, using only spoken words or phrases for control, can train the hand to go through a sequence of positions at several spots in a desired pattern. After training the system in this manner, a single spoken command will cause the system to retrace its sequence of stored actions.

We have just begun a research effort to adapt the parsing techniques of the LIFER system for use with the VIP-100. The resulting prototype system will provide a much more sophisticated capability for responding to complex spoken commands.

V. REFERENCES

Fikes, Richard E., and Hendrix, Gary G. A Network-Based Knowledge Representation and its Natural Deduction System. Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, 22-25 August 1977.

Grosz, Barbara J. The Representation and Use of Focus in Dialog Understanding, Ph.D. Dissertation, University of California, Berkeley, California, 1977. (a)

Grosz, Barbara J. The Representation and Use of Focus in a System for Understanding Dialogs. Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, 22-25 August 1977. (b)

Hendrix, Gary G. Human Engineering for Applied Natural Language Processing. Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, 22-25 August 1977.

Medress, Mark F., et al. Speech Understanding Systems: Report of a Steering Committee. SIGART Newsletter, April 1977, 62, 4-8.

Newell, Allen Et Al. Sɒeech Understanding Systems. North-Holland Publishing Company, Amsterdam, 1973.

Paxton, William H. A Framework for Speech Understanding. Ph.D. Dissertation, Stanford University, Stanford, California, 1977.

Reddy, D. Raj. Speech Understanding Systems: Summary Results of the Five-Year Research Effort. Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania, September 1976.

Walker, Donald E., (Ed.) Speech Understanding Research. Final Report, Project 4762, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, October 1976.

Walker, Donald E., and Paxton, William H., with Grosz, Barbara J., Hendrix, Gary G., Robinson, Ann E., Robinson Jane J., and Slocum, Jonathan. Procedures for Integrating Knowledge in a Speech Understanding System. Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, 22-25 August 1977. Pp. 36-42.

BIOGRAPHICAL SKETCH

Donald E. Walker

Senior Research Linguist
Artificial Intelligence Center
Information Science and Engineering Division

SPECIALIZED PROFESSIONAL COMPETENCE

Computational linguistics; language understanding systems--speech and text: interactive systems for data access through natural language; artificial intelligence strategies for information integration; text processing

REPRESENTATIVE RESEARCH ASSIGNMENTS AT SRI (Since 1971)

- Project leader, research on natural language communication with computers for task performance
- Project leader, research on integrating tactical information from multiple sources
- Project leader, research on speech understanding (ARPA program)

OTHER PROFESSIONAL EXPERIENCE

- Head, language and text processing, the MITRE Corporation: Computer-based syntactic analysis procedures for transformational grammars; question-answering systems; personal text file systems
- Research affiliate, Linguistics Group, Research Laboratory of Electronics, Massachusetts Institute of Technology
- Assistant professor of psychology, Rice University

Visiting assistant professor and research associate, University of Chicago

Research psychologist, Houston Veterans Administration Hospital, and Research associate, Baylor University College of Medicine: Language variability in psychiatric patients

ACADEMIC BACKGROUND

Ph.D. (1955), University of Chicago

Social Science Research Council Fellow in linguistics, Yale University

PROFESSIONAL ASSOCIATIONS AND HONORS

American Federation of Information Processing Societies (secretary; member, board of directors); Association for Computational Linguistics (president, vice-president, secretary-treasurer); Association for Computing Machinery (national lecturer); American Society for Information Science (chairman, Special Interest Group on Automated Language Processing); International Federation for Documentation (chairman, Committee on Linguistics in Documentation); International Joint Conferences on Artificial Intelligence (general chairman, program chairman, trustee, secretary-treasurer); Linguistic Society of America

Editorial Boards: Artificial Intelligence, American Journal of Computational Linguistics (managing editor)

Phi Beta Kappa: Sigma Xi

PUBLICATIONS

More than 25 papers on computational linguistics, artificial intelligence, computer & information science, linguistics, psychology, and anthropology

Editor: _Natural Language in Information Science: Interactive Bibliographic Search_; _Proceedings of the International Joint Conference on Artificial Intelligence_; _Information System Science and Technology_; _Information System Sciences_