SPEECH UNDERSTANDING AND Ai
AI AND SPEECH UNDERSTANDING

Donald L. Walker (Session Chatrman) tee D. Erman. Allen Newell..
Nils J. Nilsson, William H. Paxton, Terry Winograd, William A. Woods

AIJ5TRACT

This panel will review research in speech understanding (SU) and in artificial intelligence (AI) from two perspectives:

the contributions that AI has made to SU -- the? resources in AI that have been used in the development of SU systems.

the contributions that SU has made to AI •• the results of the SU program that have affected or arc likely to affect futuie AI research.

Four topics are identified for major consideration:

Multiple sources of Knowledge which are required; how should they be oiganized,

System control how to manage the complex inter actions involved.

Ianguage undetstanding comparisons of text and speech input.

Organization of research -creating complex, multisource, knowledge-based systems.

INTRODUCTION

Donald L Walker, SKI International

The Advanced Research Projects Agency of the Department of Defense sponsored a five year research program with the goal of developing a speech understanding system capable of engaging a human operator in a natural conversation concerning a specific task domain. Among the justifications fot undertaking the enterprise in 1971 were the developments in artificial intelligence taking place at that period. The culmination of this program in 1976 with the demonstration of a system that met the target specifications set out at the beginning[1] provides an opportunity for an overall review of the contributions of artificial intelligence to speech understanding research and of the implications for artificial intelligence of the research on speech understanding.

REFLECTIONS ON OBIAINING SCILNCE;

Allen Newell, Carnegie Mellon University

Each science builds its own methodology, primarily by practice rather than preaching. However, sometimes reflections help a little. These are purely my personal opinion.

One element in AFs methodology is that progress is made by building systems that perform: synthesis before analysis. Why we have this methodological tenet is not far to seek, or why other sciences do not, giving primacy to analysis in various forms.

The recently concluded ARRA SUS program is an important data point in assessing the synthesis methodology. My purpose here is to state the data point (but not argue it) and describe some of its conditions.

Q1. The question: Can scientific progress in AI be made by building performance systems? The increment of evidence from the SUS program is definitely yes,

Pursuing this (overbrief) answer leads to other questions.

02. What is science in AI? It is knowledge -• theories, data, evaluations -- that describes the means to reach a class of desired ends given certain structures and situations. Science leaches beyond the situation of its generation and become", a source of knowledge for future scientists and technologists - investment rather than consumption. Knowledge of moaiv.-onds relations is what characterizes the artificial sciences, of which AIis a part.

Q3. What then is The science that came from the SU5 effort? this note can only summarize, since its purpose is to focus on the conditions. Other panel members must provide the details. Speaking, broadly, we now have (1) the knowledge from winch to build a technology of limited SUSs; (?.) a wide scattering of knowledge along the way to higher capability SUSs; and (3) several important additions to the general stock of knowledge about perceptual systems.

Q4. What about the effort actually permitted the results to be obtained[7] (1) The performance specifications. (?) That several groups were attempting the same goal, inhibiting unilateral judgments of unattainability. (3) The scientific caliber and goals of the major scientists involved. (4) The intensive community.

Q5. Would these results have been obtained anyway if the funding had been spent in a standard style[7] Not at all. Other results would have been, no doubt. But the base level of the field would not have been lifted.

Q6. How did the striving for performance yield the science rather than detract from it? In forcing attention to aspects that really made a difference to performance, and to performance feedback to indicate what was really true. Compared to other AI efforts, the SUS effort was data and analysis rich.

Q7. Didn't the performance focus distort by de-emphasizing speech results and inhibiting the front-ends from

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

being much bettor? Limited resource*, imply choices.: Bettor front-ends would have meant worse total systems — and a loss of the scientific knowledge of what total systems are sufficient to yield what performance and in what way. In the artificial sciences, scientific knowledge does not reside in the components alone.

Q8. Didn't the performance forcus distort by making everyone throw systems together to meet, the specs? The deadline flavor of the last six months of the effort was unmistakable. Yet, the scientific yield from that same period was very high. Givon the continuation of the program, the salvage value would have been even higher.

Q9. Didn't the performance focus inhibit scientific cooperation and produce too much competition? Yes, that seemed to be a cost, though it was kept under moderately good control.

Q10. Can this technique be used in other situations? Several unique conditions obtained for -this effort that see in to be critical. (!) There was a single performance system (specs) which several groups could accept as their operatmp, goal. These specs have extieme face validity and have survived well the inevitable sniping. *(?)* 1 he groups that accepted the task were not being redirected, with a strong desire to continue to do their old things under new labels. (3) Substantial amounts of local autonomy and stability were granted to the effort (at the end this faltered, but too late to keep the program from succfeeding.) "I he SUS organizational paradigm was paradially attempted several times, but these conditions were not satisfied and the organization did not work. These failures reflect limitations on the scope of the organizational technique, not on whether attempting performance goals is a successful methodological tenet for AI.

Q11. Would I favor using the technique again to generate more AI science? Absolutely, if the conditions were right and one could tinker with the arrangements a bit.

WHAT HAVE. WE: ILANNLD?

lee D. Oman; Carnegie Mellon University

I would like to concentrate on those experiences of the recent speech understanding research that might be relevant to other areas of AI.

Speech undeistanding is a domain with a highly variable input signal, and it requires multiple sources of knowledge which are not well enough understood to be specified accurately. Thus the problem solution of necessity must include errorful and approximate processing. The requirement of building systems with specific, performance levels insured that these problems could not be ignored. The performance requirements also provided relatively objective measures of success; these were especially useful in helping select among alternative designs.

These projects extended over five (or more) years. It was realized from the start that several iterations of system design and implementation would be needed. An important aspect was keeping each iteration balanced with respect to difficulty of the task (that is, constraints provided by limithing vocabulary, syntax, number and types of speakers, quality of the signal, etc.) and its effect on performance (accuracy and speed). When systems became unbalanced, usually because of opening up some constraint precipitously, performance dropped to such low levels, that it was difficult to analyze the system and suggest directions for incremental improvement.

A strong emphasis has been on systems design, resulting in system organizations less- ad hoc than is usual. At least two of these (Dragon/Haipy *and* Hearsay) are being.explored in other domains (including image understanding, signal interpretation, protein c rystallographic analysis, complex learning, and modelling of human reading) and thus appear to be somewhat genet al.

The Harpy system raises, an interesting issue for AI. Its Organization, in which all knowledge is flattened into a single, simple, common graph representation, and its search method (the "locus" method), in which the graph is matched against the input data using a heunstic form of dynamic programming, are radically different from those used in most other AI systems. The success of Haipy forces the field to understand this approach. A few other rerenl successful AI systems (i.e., the Northwestern Chess 45-progi am and the Univeisity of Pittsburgh Internist diagnostic piogi-am) seem, to have similar characteristics, and thus accentuate the possible importance of doing a largo amount of sean h over a large but simply represented knowledge stiucture.

The *need* for efficient systems while under development (for experimentation) and the need for iterating (both on the system design and the implementation of the knowledge sources) led to sensitive problems in system construction and maintenance. The experiences with Hearsay-11, in particular, have shown that modern techniques of structured programming can be *nsed* to implement (and reimplement) complex AI systems in a research environment, striking reasonable balances among efficiency, flexibility, and user (i.e., reseaicher) convenience. In particular, good system implementation design led to the ability to reimplement selectively those aspects of the system found to be too inefficient without adversely effecting other parts of the system. This was compatible with the conceptual stiategy of starting with a very general design for the system and selectively specializing it as the need was discovered.

One development common to most of the systems is the separation of the problem-solving strategy from the domain-specific (i.e., speech) knowedge. (For example, see the three papers by Hayes-Roth and Lesser, Paxton, and Woods in these Proceedings.) These represent some of the first steps of being able to specify complex control strategies in a clean way.

Another common development, which contrasts with many other current AI efforts, is the compilation of particular kinds of knowledge into forms that *ave* specialized for application but which may be very different from their "natural" external forms (i.e., different from the way in which we specify the knowledge to the system in the first place). In some cases the same knowledge is transformed into several different forms, each appropriate for a different kind of application.

The use of "semantic" or "pragmatic" grammars, which contain semantically oriented categories instead of the (fewer number of) syntactically motivated categories of conventional

grammai s was highly effective for the most succ essful of these systems. This development parallels similar developments in several other current natural language efforts (e.g., see The paper by Hendrix in these Proceedings).

SPINOITS FROM SPIECH UNDERSTKSIML RESEARCH

William A. Woods, Bolt Beranek and Newman

The focus of my piescalation will be on what one can abstract horn the experience of the Speech Understanding ("Yogi am that can have general applicability or at least suggestive directions of approach for other Artificial Intelligence problems --- specifically other high-level perceptual tasks such as visual scene interpretation and the analysis of dialog and discourse structures in natural language, logelher with the early robot vision projects, the SUR project ranks as one of the few instances where a total system has been constituted and faced with real wot ld data that has not been alv.lt acted and simplified to eliminate noise and make the problem easiei. Although a great many things were learned dui ing the. project about the low level signal characteristics of speech sounds, I do not expect these to have great carryover into other areas. However, in the techniques for interfacing high level hypothesis formation and evaluation to such low-level sources, I believe the speech understanding work has made some generalizable contributions.

The major two such contributions, in my opinion, are the explicit exploration of control struc tines and shategies, and the. demonstration of the power of what 1 will call *factored* knowledge stiuc lures In the former area, 1 think that the* discoveiy of the density scoring sti ategies has interesting consequences in the aiea of search and optimization, since it represents a new technique not subsumed by the A* algorithm/ In the latter area, the pervasive use of various factored knowledge 11 presentation structures, in which the common parts of many different patterns or schemata are merged, has significant import for the problem?; of generalized perception and knowledge representation.

By factored struc tines, I refer to such structures as the phonetic segment latliee and the tree structured dictionary representations that are used in the BUN UWIM system, the tree structuied grammar used in CMIfs Harpy system, and the AIN grammar formalism used in many systems by now, In general, a factoi od repi esentation is any knowledge structure in which common parts of different knowledge elements are merged in such a way that retrieval processes can access them incrementally to create more and more specific hypotheses as additional data or measurements on those data are obtained. The simplest examples of such structures are decision trees or discrimination nets.

Such representations are used many places in the BBN speech system to organize its internal information about alternative theories, and is of course the major organizing principle of the Harpy system at CMU. If one shifts to the context of frame-based language understanding systems and considers the problem of determining the frame that one should be in at any given point (in a system that contains thousands of such frames, any number of which might match some initial poilion of a dialog), then it seems clear that similar such factored stiuclures can he useful. That is, one would like to have internal states corresponding to the results of sequences of measurements on the input stimuli that constrain the possible' interpretations; of the input •without enumerating all of the possibilities explicitly. These states can then indicate further measurements to be made on the input, and transitions to new internal states corresponding to more specific hypotheses that can be made as a result of such measurements. I expect this kind of factored structure to have more and more application in artificial intelligence - -- especiaJly in vision and natural language undeistanding.

Other aspects of the speech understanding systems that I think will have wider applicability include the use of Bayesian probability estimates to combine information from different knowledge sources, the use of analysis by-synthesis verification as a source of information in perception systems, and the development of middle out parsing algorithms for ATN and other phrase-structure grammars.

EXPERIMLNI AT ION IN AIM INCIAL INTILL1GLNC.E

William M. Paxton, SHI International[3]

My presentation will focus on the contributions of the SUR program to AI methodology, in particular on the value of experiments to aid in understanding the effects and interactions of system design features. The various speech systems are large and complex. It is often difficult for the designers themselves to understand the operation of the systems, and traditional techniques such as traces of sample runs arc of little use because of the complexity, of the control strategies and the vai iatinn among utterances. A particular system feature may improve performance in one case but make it worse in another, so judgments about the value of a feature must not be based on intuition or casual tests of a. few sentences. Moreover, the complexity *and* size of the systems make analytic methods of little use.

These considerations lead us to adopt an experimental approach, hut what kind of experiments should we perform? How should we cany out an experimental study so that it will help us to understand how the system works -- help us to sec which design featuics are important and why they have the effects they do? The system designers will of course have ideas about which are the important features, and, if they have clone their job well, they will also know what the main alternatives are. For example, in most of the speech systems the designers felt that it was important for the systems to be able to *island-dtirie* that is, to construct interpretations starting with words found anywhere in the input. The alternative to island driving is a more constrained control strategy such as strict left-to-right processing of the input. In this instance, a simple method to determine the effect of island-driving is to look at the difference in performance with island-driving versus with left-to-right processing.

Such comparative tests are often possible for the major

2 See my paper, "Shortfall and Density Scoring Strategies for Speech Understanding Control" in this conference proceedings.

3 Now at Xerox Palo Alto Research Center.

design features, particularly if the system is constructed with testing in mind. As an experimental approach, comparative tests have several attractive attributes, including independence of absolute performance levels and compatibility with powerful statistical methods. Consequently, we suggest that such trsts should be a standard technique to aid in understanding complex systems.

Since the aim of the tests is to understand how the system works, and not jusl to optimise performance, auxiliary measurements of system operation must he made in addition to measurements of primary criteiia such as speed and accuracy. Again, we rely on the designer's knowledge of the system to decide what' to measure so that we will have the necessary information available to explain the observed results. In a speech understanding system, auxiliary measures would include a variety of things such as number of correct and incorrect words accepted and storage usage. These auxiliary measures provide the intermediate steps in explanations of the effects of the system features. Thus, for example, the effect of island-driving on system accuracy might be explained by reference to its effect on storage usage in conjunction with information about the relation between storage usage and accuracy.'

Typically, there will be several design features that are believed to be important, and it will be desirable to test the features simultaneously to see how they interact: Moreover, the features will usually have good effects in some cases and bad effects in otheis. The statistic al method for dealing with such a situation is called analysis of variance. This-technique makes it possible to compute the probability that observed effects and interactions are really caused by the experimental variable*- rather than by chance. We will briefly illustrate the use of this technique by sketching some *oi* our experiments on control strategy design choices.[4]

In conclusion, if it is worth building a system as part of an AI reseatch project, it is certainly wotth making an effort to understand how the system actually works, and experimentation is an important technique for doing this. Simply demonstrating a working system should no longer be enough; let us begin to demand that the AI system designer specify the supposedly important system features and their alternatives, do the experiments to show the features' effects, and provide explanations of why the features have those effects.

LANGUAGE

Terry Wmograd, Stanford University

Although I have not been directly involved in research on speech understanding, 1 have followed the work closely, since I believe that it is a harbinger of things to come in AI research in general. The speech projects are the first major AI efforts which have placed primary emphasis on the system organization required for making use of diverse sources of knowledge in a task whose structure defies simply structured programs. Over the next few years, othei AI researchers will begin to attack problems which demand this kind of robustness -- the ability to tome up with an answer when the input data are messy, the combinatorial possibilities are explosive, and high level knowledge can be of great influence in determining the answer. In programs for vision, scientific analysis (a la Dondral), and language, I foresee a shift in this direction.

In programs for comprehending language — even those that deal with text .....we need to move toward handling more natural inputs, with all of the inaccuracy, incompleteness, and ill-formedness we have long ignored. This will demand program and knowledge organizations that are based on ideas that the speech work has begun to expfore: multi-process communication; hypothesis formation *and* verification; the intermixture of goal-driven an*d* event-driven processing; careful attention lo the *interface Languages* that make it possible to give meaningful structure to the communication between components; and the importance of performance evaluation tools that help us make sense of what is happening in a complex multi- process environment.

In the panel, I would like to see a discussion along two lines: What are the major insights to be gained from the speech project experience that can be of use in organizing other AI programs? an*d* what are the obvious gaps to be filled in the next round of experimentation? Since my major current concern is with representation languages, I would like to take part in a discussion concerning the problems that were encountered with the representational systems available for speech work, and the features that will be important for systems of the same degree of complexity in the future.

UNDERSTANDING RESEARCH

Nils J. Nilsson, SRI International

Artificial Intelligence has recently completed an extensive and coordinated exploration into the terra incognita of large scale, knowledge-based systems. To the brave and resourceful explorers, we stay at homes must say, "Congratulations and well-done! We enjoyed your slide shows and marveled at your specimens". But when the celebrations are over it will be important for us all to digest the new knowledge uncovered by these explorations.

We need to ask more than "What have we learned?" It is too tempting to answer that question using our current vocabulary. We might for example fall into adopting the rough and ready frontier parlance and metaphors of the explorers themselves and start speaking about "multiple cooperating sources of knowledge", "blackboards", and "island growing". Or we might attempt to describe the new' vistas with older and perhaps inadequate phrases such as "rule-based systems", "left-to-right parsing", and "heuristic search".

The question before us, I think, is harder than "What have we learned?" It is "How are we going to express what we have learned?" A major expedition just completed is too precious an occasion to let pass heralded only by accounts from the explorers. It is an opportunity for attempts at

4 Further details are given in Walker and Paxton, et al., "Procedures for Integrating Knowledge in a Speech Understanding System" in the proceedings of this conference; a full description of the experiments appears in Paxton, A Framework for Speech Understanding, Ph.D. Dissertation, Stanford University, 1977.

synthesis and for inventing new concepts and new paradigms. We should not be discouraged meiely because there is no guarantee that these attempts will ho successful or because the odds against useful new paradigms are always high. We have just spent about 100 man-years on exploring. We can affoid to follow this up with a few man-years of thinking about how to say whal we have learned.

As good as they are at speech understanding, it is unfortunately true that HARPY and HEARSAY and friends cannot speak for themselves. The major product of the 100 man-years is not the total body of code that was produced nor is it what that code accomplished in the demonstrations. The memorable output, what can be taught to future generations of students, will be a description of that code. It is not even necessary that the descriptions be completely accurate. Simplifications and even fabrications are justified if they have pedagogic value and do not overly mislead posterity. *My* ma|or point is that it is important that these descriptions be elegant and that they have a certain,, hard-to-define, esthetic appcal so that they will be memorable, easy to use for teaching purposes, and provocative for the design of new systems.

In creating the kinds of descriptions that I think will be important, inventive latent will be more important than reportorial skill. Suppose, for example, that one could invent some imaginary system that was something like one of the actual speech understanding systems but different in many details. Since our imaginary system doesn't really have to run on a computer we can strip it of the various ad hoc features of real systems so necessary for efficiency. Now maybe .we can reorganize it a bit to give it a more coherent internal organization and to relate it more closely to existing well-understood AI mechanisms. There may he some tension in trying to do this. Maybe the existing AI mechanisms aren't so well-understood or as general as we thought. Perhaps the effort of trying to build our imaginary system out of these mechanisms stretches then) a hit. Maybe we'll be fortunate enough to think of a major generalization of some of these mechanisms to make them more useful for our fictional system. Now, maybe we'll reorganize the fictional system some more and go through the loop again. Once in a decade or so, and if our interests are broad, we might notice that the new AI concepts just invented could also be profitably used to describe the results of other explorations. At the very least our new synthesis will geatly simplify the process of designing new systems of a similar kind.

These steps are important if a field is to grow into a mature scientific or engineering discipline. Artificial Intelligence has to take several such steps before it can be as productive as we all would like it to be. AI has not yet really developed what could be called a set of universally adopted methodologies that can be followed in the design of new systems. If six different AI laboratories were given the task of building a *rule-based system* for some well-understood application, 1 would not be surprised to see several quite different designs. Much of the terminology used by AI people is still pre-technical at best and meaningless jargon at worst. Let's try to use the plentiful and excellent experiences of the speech understanding projects to climb a rung or two in the conceptual understanding of our field.