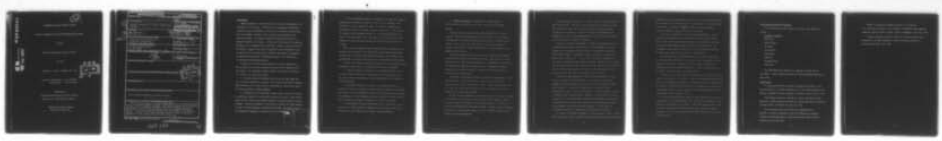
AD-A049 298 CALIFORNIA UNIV BERKELEY ELECTRONICS RESEARCH LAB F/G 17/2
PROSODIC INFORMATION FOR SPEECH UNDERSTANDING SYSTEMS.(U)
SEP 77 M H O'MALLEY DAHC04-75-G-0088

UNCLASSIFIED NL

| OF |
AD
A049298



END
DATE FILMED
3-78
DDC

12

COMPREHENSIVE FINAL TECHNICAL REPORT

PROSODIC INFORMATION FOR SPEECH UNDERSTANDING SYSTEMS

62706E

DAHC04-75-G-0088 and DAAG29-76-G-0250

$121,700

February 1, 1975 to September 30, 1977

Principal Investigator: M. H. O'Malley
Telephone No.: (415) 642-4624

Sponsored by

Defense Advanced Research Projects Agency

ARPA Order No. 2606 Amendment Number 3

Approved for public release;
distribution unlimited

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

Prosodic Information for Speech Understanding Systems

5. TYPE OF REPORT & PERIOD COVERED

Final Report.
Feb. 1, 1975-Sept. 30, 1977

6. PERFORMING ORG. REPORT NUMBER

1 Feb 75-30 Sep 77

7. AUTHOR(s)

M. H. O'Malley

8. CONTRACT OR GRANT NUMBER(s)

DAHC04-75-G-0088
DAAG29-76-G-0250

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Electronics Research Laboratory
University of California
Berkeley, California 94270

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, North Carolina 27709

12. REPORT DATE

30 Sep 77

13. NUMBER OF PAGES

129p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Syntax, Speech Recognition, Prosodies, English

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is a final report of a project whose goal was to use prosodic information to aid speech recognition systems. Naturalistic speech data was collected and used to test and develop hypotheses about the relationship of prosodic information to the syntactic and semantic structure of English sentences. Members of the project interacted closely with other members of the ARPA SUR project. A series of SUR notes describing acoustic features of prosodies, phonological rules, stress rules, etc. were produced.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

127 550

4B

Introduction

Human perception of speech involves the listener's knowledge of his language and of the world. Information from phonological and lexical structure, syntax and semantics, as well as the listener's expectations about the speaker's behavior, can all affect the processing of the acoustical signal. In contrast, machine perception concentrates on physical aspects of the signal. In the case of isolated word recognition machines, such an emphasis seems justified. However, work in Artificial Intelligence suggests that speech recognition system performance could be improved by the incorporation of syntactic and semantic information. To emphasize the Artificial Intelligence approach, such systems are called speech understanding systems.

A number of syntactic parsing systems for natural language have been produced but they have generally involved written language rather than speech. The differences between speech and writing are important to a system in at least three areas.

1. Function words and morphemes (such as is, of, the, -ing, -ed, etc.) are often indistinct in speech. Since parsers for written language make extensive use of function words as delimiters, these parsers cannot be directly applied to spoken language.

2. Prosodies such as intonation, stress, pause, juncture and rhythm are important signals of the syntactic structure of speech. Such signals permit speech to have a more complex syntactic structure than has written language. Prosodic information can be used by a parser for spoken English in order to reduce ambiguity, eliminate false paths, and replace some of the information signaled by function words in writing

3. The relationship between the alphabetic units and their physical representation is much less fixed in speech than in printing. Word boundaries are very difficult to find. The acoustic realization of phonological elements is context sensitive and unstable. Often some of the segmental units (phonemes) which are in the dictionary entry for a word are not pronounced at all. For these reasons, parsers for spoken language are faced with much greater uncertainty than are parsers for writing.

Thus a central problem in speech understanding is to develop a system which takes account of the differences between speech and writing. In particular, the shift of structural information from function words and morphemes to prosodic features must be incorporated into the grammar.

In order to incorporate prosodic information into an automatic speech recognition system, work in the following four areas was proposed:

1. Acoustic analysis of prosodies and phonological rules - we proto collect limited protocols of spontaneous and read speech and to use this data to develop algorithms to locate phonological word, phrase and clause boundaries.

2. Linguistic analysis of prosodies and phonological rules - we proposed to survey and integrate various linguistic studies of intonation and rhythm, cast these hypotheses against our empirical data and generate new hypotheses. We also proposed to collect phonological rules with the goal of developing a system for parsing or inverting such rules.

3. Integration of prosodic information into a SUR system - we proposed to put a simple prosodic component into a SUR system, test it and then extend it as a result of our acoustic research findings.

4. System Development - We proposed to develop a PDP 11 facility which could access various SUR systems over the ARPANET.

RESULTS

The ARPA speech project was originally organized into 5 major projects conducted by organizations with experience in system design and supported by 4 smaller projects with expertese in linguistics and speech. By the end of the 5 year project, one of the systems had actually achieving the original design goals (as re-interpreted by the members of the project).

The project had started with the view that artificial intelligence techniques had advanced to the point where they could offer the technological basis for some practical application. Furthermore, developments in computational linguistics, speech science and signal processing suggested that speech recognition or "speech understanding" might be such a practical applications.

The project started with strength in artificial intelligence, speech and linguistics. The first two years consisted of a great deal of teaching on the part of the 4 smaller groups. There is no question that the result was a much more "linguistic", more principled, less ad hoc design for all of the major systems.

The next 3 years consisted of successive attempts by the major systems to incorporate more and more of what was known about language. However, in a sense none of them came close to incorporating even a fraction of the well known linguistic facts about English. Practical tasks of building systems and of incorporating static rules into a dynamic procedure overwhelmed their good intentions.

The final year, of course, was a frantic effort to cut the crap and make something work. The irony of the project was that the only system that did meet the goals was a simple combination of statistics and low-level speech science -- it had no artificial intelligence and no linguistics, it was a pure engineering system and it was written almost as a side effort by a couple of students.

During the course of the project, I pushed very strongly toward an even more theoretically correct system. In the beginning, I naively thought that the systems could incorporate a larger part of what was known about language. I pushed for studies of dialect, communication modes, natural syntax, etc. -- all of which were quite irrelevant to the types of systems that were finally produced. Had all of us been more goal oriented, we should have produced much more limited, more successful, much less interesting systems. The project which would have resulted would, of course, have had a much smaller long term impact.

Viewed from the prospective of achieving the group goals, my own work was certainly counter productive. It was designed to push the systems in directions which, in retrospect, were the opposite of those which "worked". I still do not know whether a serious, 10 year, theoretically correct project would produce a very good system or merely a slower and less accurate system. However, I do know from the limited success that we did inadvertently achieve that Wizenbaum's criticism of the speech project and its social implications was absolutely correct. The government does not need more word spotters. In any case, the following are some of the things that we did during the project:

We added a toy prosodic component to the toy hearsay I system at CMU. While this showed that we could use the ARPA net better than more linguists

and were not bad at understanding other people's code and then modifying it, it didn't have much effect on the system design. On the other hand, I think that it resulted in heresay I having the only non-empty "prosodic component" among all of the final systems.

We looked in great detail at the various BBN grammars. We did add some pseudo-prosodies to the LUNAR system which at least showed that we knew LISP better than most linguists. We also discovered why the original LUNAR grammar, while a fantastic contribution to text processing, would never have worked as a speech recognition component. The BBN project discovered the same fact independently.

We also looked at the later BBN grammars in terms of what kinds of sentences they would handle. Our results were not especially popular. However, by the end of the project, BBN had caught up with CMU by abandoning the idea of grammar altogether. Neither, however, was able to surpass the higher levels of the SDS system.

We spent a lot of time trying to develop and use a speech processing system on the ARPA net. From this, I learned that systems programming is fun but can overwhelm the attempt to do anything "useful". However, Tovar contributed a great deal to the general net community and to making Unix a reasonable system for speech research. We also learned never to trust anyone (ARPA) with a product to sell.

On a more scientific level, Alan Cole did a great deal of good work using Hearsay II and the CMU speech data. This work on phonological rule analysis is continuing to some extent at IBM where Alan now works. Had the project continued, his work would have had a significant impact -- especially as it was in the spirit of the final statistical process that actually worked.

Participating Scientific Personnel

The following people participated at various times during the project.

Michael H. O'Malley

Alan Cole

Malcah Yaeger

Ron Bader

Greg Shenant

Dean Kloker

John Moch

Richard Gerould

David King

Cole and Kloker will receive Ph.D. degrees for their work on the project. King, Gerould and Bader have received Masters Degrees for their work.

Publications

"A statistical model of low-level phonological processes," Alan Cole and Michael H. O'Malley, presented at the 2nd Annual Meeting of the Berkeley Linguistic Society, Berkeley, California, February 14-16, 1976.

"Phonological Variation," Michael H. O'Malley and Malcah Yaeger, presented to ARPA Phonological Workshop at System Development Corporation in Santa Monica, California from June 3-4, 1974.

"System design issues in prosodic rule implementation," Michael H. O'Malley, presented at the 1975 Conference on Computer Graphics, Pattern Recognition, and Data Structures, Beverly Hills, California, May 14-16, 1975.

"PIXIE: An interactive Graphics System," D. E. King and M. H. O'Malley, presented at the Nrothwest 76 ACM-CIPS Pacific Regional Symposium, Seattle Pacific College, Seattle, Washington, June 24-26, 1976.

"Testing Phonological Rules," Michael H. O'Malley and Alan Cole, presented at the IEEE Conference on Audio and Electroacoustics in Pittsburg from April 15-22, 1974.