The Integration of a Continuous-Speech–Recognition System with the QMR Diagnostic Program

Smadar Shiffman, Christopher D. Lane, Kevin B. Johnson, Lawrence M. Fagan
Section on Medical Informatics, Medical School Office Building X215, Stanford University, Stanford, CA 94305-5479

ABSTRACT

We describe a continuous-speech interface for Quick Medical Reference (QMR), which allows physicians to input spoken descriptions of physical-examination findings, or observations. We analyze the difficulties in designing a continuous-speech interface for systems that use medical terminology. We present a method for matching spoken finding names expressed in natural language to QMR terms. The method is based on a semantic representation of findings that both minimize the effect of misrecognition and derive grammars that are necessary for supporting the recognition process.*

A limiting factor in the acceptance of medical decision-support tools is the effort that people must expend to learn and use them. Previous studies have suggested that the ability to use speech could provide a more satisfactory way to communicate with medical applications, compared to conventional interfaces [1]. We are exploring the use of speech input to increase the usability and acceptability of medical diagnostic systems.

Developments in speech recognition technology have made it feasible to build speech applications in various domains, including medicine [2, 3, 4]. Previously, we designed a speaker-dependent isolated-word interface for entering findings to QMR [3]. Although the resulting interface had adequate functionality, users were required to navigate through a series of menus until they found the desired QMR finding. In this paper, we describe our experience in building a speaker-independent continuous-speech interface to QMR that permits physicians to enter findings using natural language.

BACKGROUND

QMR performs inferences on an updated version of the INTERNIST-1 knowledge base [5, 6]. The knowledge base includes 500 diseases, 3500 patient findings (including physical findings, laboratory-test results, and descriptors of a patient's medical history) and links defining causal, temporal, and logical interrelationships among diseases. QMR accepts a set of findings as input and provides a differential diagnosis of leading disease hypotheses, ranked by their probabilities. The findings in QMR are expressed as compound-noun phrases, some of which are ungrammatical or sound awkward—for example, *skin nevi multiple*.

Linguistic approaches have been used successfully to support applications that use medical terminology. The Linguistic String Project [7] applied semantic techniques for compute-based management of narrative medical data. The SAPHIRE information-retrieval system [8] and the CLARIT project [9] used semantic approaches to perform automatic indexing. In this work, we combined an underlying semantic representation of finding names and a pattern-matching technique to map input utterances to a subset of the QMR terms.

Current speech recognition systems can recognize vocabularies whose sizes range from tens to thousands of words [10]. *Speaker-dependent* systems require that users train the system to recognize their speech, whereas *speaker-independent* systems do not. *Isolated-word* systems require that the speaker pause between words or short phrases, whereas *continuous-speech* systems allow the user to speak long sequences of words without pausing. The vast majority of systems that have been fielded incorporate isolated-word technology.

Continuous-speech recognition systems use a target-language specification in the form of a lexicon and a grammar to decrease the competing interpretations for a given input utterance. Commonly used grammar forms are finite-state networks (out of which the set of allowable sentences can be generated) and trigrams (grammars that indicate the probability that a given word follows its two precedents in one sentence). The form of a grammar represents a compromise between two conflicting trends that affect system usability: As the size and complexity of a grammar increases, the speech recognition system can handle more variable input, but the recognition accuracy decreases.

DESIGN CONSIDERATIONS

Physicians can specify a single medical term in many ways. For example, the QMR finding *skin rash dorsal hand bilateral* might be expressed by a physician as *rash on the back of the hands* or as *a bilateral dorsal hand rash.* QMR, like other medical decision-support programs, uses a controlled vocabulary that is not likely to be known to the system's prospective users. The absence of standard medical terms for expressing concepts requires users to speculate about what terms are known to the diagnostic system. Based on this observation, it is difficult to predict what phrases physicians might speak when they are entering finding names to QMR. We expect users of our speech interface to speak a large variety of expressions, some of which will not designate any QMR term because either they lack details that are relevant to the specification of a

* QMR® is a registered trademark of the University of Pittsburgh.

0195-4210/92/$5.00

finding in QMR, or they include more information than comparable QMR terms. We believe that the language physicians will use will be of manageable variability, because many physicians describe patient findings using phrases that are similar to expressions taught in medical school.

Since the speech recognition system we use, like all other speech recognizers, performs imperfectly, the ASCII string it returns for the input utterance may not include some of the words that were uttered, or may include words that were not spoken. Nevertheless, the interface should still be able to identify the QMR term that the user had in mind. We view the essence of the identification task as assigning the input utterance, based on the underlying meaning of the utterance, to one of the classes designated by the target QMR terms. This classification problem is a simple one because the target classes are predefined.

We learned that, in order to define a target language that would support recognition with reasonable accuracy, we would have to use only a subset of the QMR findings, and to partition that subset into smaller subsets, each of which would have its own target sublanguage and subgrammar. We selected the domain of physical-examination findings because it could be partitioned easily into body parts, which constitute subdomains that are intuitive to the user. Because we excluded all findings for which no words were included in the standard dictionary of the speech recognition system, we used only about one half (518) of the physical-examination findings that are in QMR.

METHODS

We use a semantic representation for finding names that reduces the effect of misrecognition on interpretation accuracy, as misrecognized words that are irrelevant to the meaning of finding names—such as *the, he, patient, notice*—are ignored. We assume that, even if an utterance is partially misrecognized, enough of the semantic content will remain to allow the system to identify a controlled-vocabulary term that is similar in meaning. We assume that the user will enter utterances that encompass only a single finding descriptor (for example, we presuppose that physicians will not use compound sentences to describe findings).

The methods described in this section support the following interaction cycle for adding a finding to a QMR case description. First, the physician specifies a body part and enters a finding name that is translated by the speech recognition system into an ASCII string. Then, the interface program extracts the essence of the input utterance into a semantic canonical form, which it then compares to similar precomposed forms of QMR terms. The program displays the results of the comparison as a rank-ordered list of matches, from which the physician may select findings for adding to the case description by specifying their number and stating whether they were present or absent. When the interface cannot find an appropriate matching QMR term for a finding name because the utterance was misrecognized completely, the user can speak the finding again or edit the string returned by the speech recognizer.

Table 1 shows findings that were included in a case description built using the speech interface to QMR. For each input utterance, the table shows the context, the interpretation returned by the speech recognition system, the QMR term that was added to the case, and the position of the selected finding in the rank-ordered list of matches (in parentheses).

Pattern Matching Using Canonical Forms

The interface program captures the key notions of a finding name in a canonical form by constructing a set of related *keywords* or key concepts. For each finding name, the interface program looks up all the meaningful words or word combinations in a thesaurus, and, for each word or word combination, it includes in the canonical form a representative of the appropriate synonym class.

The program creates reference patterns by converting the QMR findings into canonical forms. Similarly, the program creates test patterns that represent input utterances by converting the input into canonical forms. For each input utterance, the program matches its test pattern to reference patterns that belong to the current system context, and computes a score for each reference pattern based on that pattern's distance from the test pattern. The distance measure is a function of the specificity of concepts that are included in both test and reference patterns. The scoring formula assigns a reward or penalty to the score, depending on whether concepts in the reference pattern are included in the test pattern or are excluded from it. Thus, the formula produces good (high) scores for target terms that include highly specific concepts, all of which are identified in the input utterance. The rationale for using specificity as a weight is the desire to increase the importance of highly specific input terms.

The pattern-matching approach supports recall of relevant findings when there is no exact correspondence

Table 1 Findings from a case built using the speech interface

Context	Spoken phrase	Recognized phrase	QMR finding selected (position in list)
abdomen	there is mild right upper-quadrant mass noted	there is right of a upper under a mass	abdomen mass right upper quadrant (3)
leg	there is edema bilateral severe	there is edema bilateral in the moderate	leg edema bilateral massive (1)
leg	left lower-extremity swelling	left by a extremity swelling	calf swelling unilateral (1)
breast	there are multiple large hard nodules bilaterally	there are a left about her nodule by the breast	breast mass bilateral (2)

between the input and any QMR finding because of an overly narrow or overly general input specification. For example, the input utterance *there is a large mass in the abdomen*, which is represented by the canonical form *{enlargement, mass, abdomen}*, elicits the more general QMR term *abdomen mass present* among the matching findings. The utterance *there is a mass on the right side* (in the context of abdominal findings), which is represented by the canonical form *{mass, unilateral, side}*, elicits the more specific QMR terms *abdomen mass right upper quadrant* and *abdomen mass right lower quadrant*. Figure 1 shows rank-ordered matches that the interface displayed after a user said *rash on the back of the hands*.

Programmatic Grammar Generation

We define the target interaction language as the set of phrases (grammatical or nongrammatical) that might be spoken by physicians to designate any of the target QMR finding names. Identifying all the sentences in the language that could be used to describe physical-examination findings is impossible, because physicians may express findings in numerous ways, which are not governed by any rule. We try to capture the variety of possible input expressions by generating a grammar for the target language automatically from the set of those physical-examination findings that are in QMR. The advantage of this approach is that it is independent of idiosyncratic expressions for finding names; the disadvantage is that the resulting grammar is much too inclusive in that it allows the production of many sentences physicians would never say. For example, a physician would not say *there is a hand of the skin on a rash* to describe a patient who has a rash on his hands, but this sentence is generated by the grammar about the hand. The programmatically generated grammars place a heavy load on the speech recognition system, in that the number of competing sentences that the speech-recognizer has to check is much larger than is required for interacting with the diagnostic system.

We generate grammars for the speech recognition system programmatically by deriving rules from the canonical forms of finding names. Each finding is considered to be a target phrase that the user might express using various words or word orders. We attempt to capture

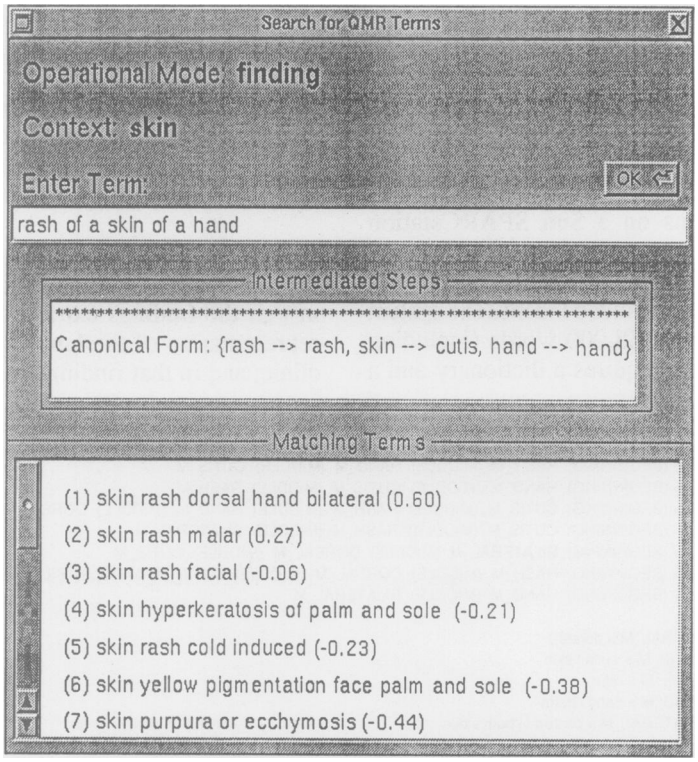

Figure 1 Rank-ordered matches that were displayed for the input utterance *rash on the back of the hands*. The *Enter Term* window at the top displays the string that was returned by the speech recognition system for the input utterance. The *Intermediate Steps* window at the middle shows the canonical form for the interpreted string. The form *A–>B* indicates that word *A* is represented in the canonical form by its synonym *B*. For example, *skin* is represented by *cutis*. The *Matching Terms* window at the bottom of the figure displays the best matches, with their associated relative scores. Negative values result from a high penalty for highly specific concepts in the target term that do not belong to the input term.

the variety of expressions by applying one of the following operations on the set of keywords in a canonical form: power-set generation, permutation, synonym substitution, and insertion (of words that are not keywords). We generate the power set of keywords to account for cases where the user's notion of the particular finding is more general than is the notion in QMR, and therefore we expect the user not to use all keywords in specifying the finding. We permute each element in the power set to account for variations in word ordering. We substitute synonyms for the original keywords in the canonical form to account for differences in choice of words among users. We insert words that can be used in a specification of a finding name before, between, and after keywords. Figure 2 illustrates sample results of applying these operations on a canonical form for deriving grammar rules.

The set of finding grammars is complemented by a command grammar, which includes a few commands for controlling the interaction. For example, the command *select one negative* is used for selecting a finding, specifying that it is negative (or is absent in the patient), and adding it to the case. Other commands are used to enter the next finding, to cancel the last finding or to switch to a new body part. We arranged the commands in a separate grammar to ensure high accuracy for commands, as the commands are crucial for user control of the interaction.

System Configuration

The speech interface runs on a NeXT workstation connected to a Speech Systems, Inc. (SSI) DS200 speech recognition system that runs on a Sun SPARCstation platform (Speech Systems™ DS200 is a trademark of Speech Systems, Inc.) The SSI system recognizes continuous speech and is speaker-independent. The system has a vocabulary of more than 38,000 words (including root forms and inflections). It requires a dictionary and a grammar for each set of phrases that it is expected to recognize. Our configuration includes a dictionary and a grammar for each body part to which the user can refer when specifying a physical-examination finding. The interface communicates with the speech system via a general-purpose speech server that can support several speech-driven applications simultaneously.

A PRELIMINARY TEST OF THE INTERFACE

We performed a preliminary test of the speech interface by speaking 52 finding names to the interface. The findings were extracted from 16 transcribed dictations of history and physical-examination reports that we obtained from a physician affiliated with our laboratory. All the findings were known to be in the scope of the QMR finding set and were known to include only words that were in the vocabulary of the speech recognition system. The response time for a given input utterance (measured from the end of an utterance to the display of matching terms) was about 0.5 second. We compared the first five QMR findings that were displayed after the user spoke the extracted finding names with those displayed after she typed them. For 40 out of the 52 finding names, the sets of the first five QMR findings were equivalent for both spoken and typed input. For seven of the remaining findings, there was partial overlap between the sets. These results demonstrate that, for our data set, the performance of the speech interface was comparable to the performance of the typing interface.

DISCUSSION

Although the initial testing of our interface is biased in that the findings extracted from the test cases are not representative of the language that physicians use in the clinic, and in that findings for the test case were extracted

A

```
S --> (BEGINNING) DORSAL_M (MIDDLE) CUTIS_M (MIDDLE) RASH_M
S --> (BEGINNING) RASH_M (MIDDLE) HAND_M (MIDDLE) CUTIS_M
S --> (BEGINNING) HAND_M (MIDDLE) CUTIS_M (MIDDLE) RASH_M
S --> (BEGINNING) CUTIS_M (MIDDLE) RASH_M (MIDDLE) HAND_M (MIDDLE) DORSAL_M
S --> (BEGINNING) CUTIS_M (MIDDLE) RASH_M (MIDDLE) BILATERAL_M
S --> (BEGINNING) BILATERAL_M (MIDDLE) DORSAL_M (MIDDLE) CUTIS_M
S --> (BEGINNING) RASH_M (MIDDLE) DORSAL_M (MIDDLE) CUTIS_M (MIDDLE) BILATERAL_M
S --> (BEGINNING) HAND_M (MIDDLE) BILATERAL_M

DORSAL_M > dorsal
CUTIS_M > cutis | skin
RASH_M > rash
HAND_M > hand | palm
BILATERAL_M > bilateral | both sides
```

B

```
BEGINNING > there {is | was | were | are} ({ a | some })
BEGINNING > i {found | noticed | notice | saw | see} ({a | some | many | that the })
BEGINNING > {the patient | he | she } {has} { a | some | many }
MIDDLE > is
MIDDLE > PREP {the | her | his | the patient's | a }
PREP == in on about by with at under over of
```

Figure 2 Sample grammar rules that were derived from the canonical form *{cutis, rash, dorsum, hand, bilateral}* that represents the QMR finding *skin rash dorsal hand bilateral.* The rules in group A were generated programmatically; the rules in group B were added manually. All nonterminal symbols are marked with the suffix _M. The symbols *BEGINNING* and *MIDDLE* represent classes of words that may appear at the beginning of a sentence or in the middle of a sentence, respectively.

and spoken by one of the people who developed the interface, the interface still demonstrates the potential of a continuous-speech interface to a medical application that uses a large controlled vocabulary. To test our interface in a clinic, however, we must enhance the range of the terms that the system can identify to include the full set of QMR terms, and we must increase the rate of recognition accuracy by constraining the grammars used by the speech recognition system to include only highly probable utterances. We hope to conduct future tests of the interface to obtain data that will help us to fine tune the grammars.

We realize that automatic grammar generation produces grammars that define large sublanguages (for example, over 2 billion sentences can be generated from the grammar that describes the leg). These grammars generate many phrases that never would be spoken by a physician. We can improve recognition accuracy by eliminating rules that generate unacceptable sentences. Cleaning up the grammars manually is a difficult and tedious job. A possible direction for future research is to find a method that will identify unacceptable sentences programmatically—for example, by comparing word sequences generated by a grammar to entities in a database of medical terms, such as the UMLS metathesaurus.

Our goal—to allow physicians to enter finding names to QMR using continuous speech—was ambitious in light of the capabilities of current continuous-speech recognition systems. We realize that the speech system we used, like other continuous-speech systems that perform with similar recognition accuracy, is still not able to support natural-language communication using a large sublanguage.

Isolated-word speech recognition systems have been used successfully in medical applications with restricted domains—for example, in systems for generating radiology reports. We first approached our task using an isolated-word speech recognition system. Our initial interface required that the user navigate through sequences of menus by saying the names of menu options. Our goal in using a continuous speech interface was to allow a natural description of physical exam findings. We created grammars that allowed terms to be expressed in multiple ways, but these grammars exceeded the capacity of the equipment we were using. We resolved this problem by adding an extra level of selection (e.g., organizing the terms by parts of the body), and, thus, were able to create manageable sized grammars at the expense of naturalness. We are currently exploring ways to improve recognition accuracy, and hence to increase interpretation accuracy—for example, by designing grammars manually and then modifying them automatically.

Acknowledgments

We thank the following people for assisting us in our research: Blackford Middleton, Alex Poon, and William Detmer provided feedback on ideas that we explored in this work; Mark Musen helped us to obtain history and physical reports for preliminary testing. The physician-students in the Medical Information Sciences Program at Stanford University helped us by providing samples of natural-language expressions of finding names. We thank Camdat Corporation for providing us with the synonym dictionaries that are used in QMR. Last but not least, we thank Lyn Dupre for editing earlier versions of this paper.

This work has been supported by the National Library of Medicine under grant LM-04864. Additional support was obtained under contract number 213-89-0012 from the A.H.C.P.R. Computer facilities were provided by the SUMEX-AIM Resource grant LM-05208 and through an equipment loan from Speech Systems™, Inc.

References

(1) Feldman C.A., Stevens, D. Pilot study on the feasibility of a computerized speech recognition charting system. *Community Dent Oral Epidemiol*, 1990, 18, 213–215.

(2) Bergeron, B., Locke, S. Speech recognition as a user interface. *MD Comput.* 7, 329–334. 1990.

(3) Shiffman, S., Wu, A. W., Poon, A.D., Lane, C.D., Middleton, B., Miller, R.A., Masarie, F.E., Cooper, G.F., Shortliffe, E.H., Fagan, L.M. Building a speech interface to a medical diagnostic system. *IEEE Expert*, 6, 41–50. 1990.

(4) Wulfman, C. E., Rua, M., Lane, C. D., Shortliffe, E. H., Fagan, L. M. Continuous-speech recognition in oncology record keeping. Technical Report, KSL-90-67, Knowledge Systems Laboratory, Stanford University, Stanford, CA, 1990.

(5) Miller, R.A., Masarie, F.E. Quick Medical Reference (QMR): An evolving microcomputer-based diagnostic decision-support program for general internal medicine. In Proceedings of the Thirteenth Symposium on Computer Applications in Medical Care, Washington, D.C., 1989, pp. 947–948.

(6) Miller, R. A., Pople, H. E., Myers, J. D. INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med.* 307, 468–476. 1982.

(7) Sager, N., Friedman, C., Lyman, M.S. *Medical Language Processing: Computer Management of Narrative Data*; Reading, Mas: Addison-Wesley, 1987.

(8) Hersh, W.R., Greenes, R.A. SAPHIRE; An information retrieval system featuring concept matching, automatic indexing, probabilistic retrieval, and hierarchical relationships. *Comput Biomed Res.* 23, pp. 410–425. 1990.

(9) Evans, D.A. Concept management in text via natural-language processing: the CLARIT approach. Working Notes of the 1990 AAAI Symposium on Text-Based Intelligent Systems, Stanford University, Stanford, CA, 1990, pp. 93-95.

(10) Lee, K.F. *Automatic Speech Recognition: The Development of the SPHINX System*; Boston: Kluwer Academic Publishers, 1989.