In Pursuit of Effective Handsfree Decision Support: Coupling Bayesian Inference, Speech Understanding, and User Models

Eric Horvitz

Decision Theory and Adaptive Systems
Microsoft Research
Redmond, WA 98052
horvitz@microsoft.com

Michael Shwe

Knowledge Industries
Oak Grove Avenue
Menlo Park, CA 94025
shwe@kic.com

Abstract

We describe our efforts to develop an effective handsfree human–computer interface that allows users to access the results of diagnostic inference without interfering with the typical patterns of interaction for complex, procedure-oriented domains. For the prototype, we integrated a decision-theoretic diagnostic system, that makes inferences about testing strategies and pathophysiology in emergency medicine, with a speech-understanding system, and a head-mounted personal display. The personal display projects inference results in a decision maker's field of view. We explored the nature and timing of feedback about speech recognition as well as how knowledge about the structure and content of displayed information might be employed to enhance speech recognition. After presenting the prototype system, we present research on extensions motivated by the prototype, centering on elucidating principles for developing user models that infer the likelihood of a user's next utterances, given information about the content of displays, recent user actions, and context.

Keywords: Human–computer interface, decision-theoretic diagnosis, speech understanding, user modeling.

1 Introduction

Despite significant advances in representations and inference strategies for computer-based diagnosis, there has been relatively little use of decision support systems in the real world. We believe that a key factor in the slow diffusion of automated inference into real-world settings is the fundamental disparity between the human-computer interface provided by traditional computing platforms and the ergonomics of the daily tasks and procedures of the real-world applications.

Addressing the problem of efficiency, availability, and naturalness of the human–computer interface to diagnostic reasoning can enhance the delivery of intelligent systems to decision makers. For example, healthcare personnel typically spend a majority of their work days interviewing and examining patients, performing procedures, and communicating information to patients, colleagues, and supporting staff. To date, accessing information from computers generally requires the use of a keyboard, mouse, and computer display. The patterns of activity and typical pressures of medical practice diminish the incentive or opportunities for physicians to have continuous interaction with computer-based resources—especially in such contexts as time critical contexts. The delivery of computer-based information into real-time decision making requires that workers be able to access computer-based information wherever they may be working, and in a manner that does not significantly interfere with their ability to perform procedures and communicate with colleagues wherever domain problems may arise.

We address in this paper the problem of making computer-based information available for interactive use during real-time procedures by creating handsfree diagnostic systems. The goal of this work is to leverage automated inference as well as innovations in speech understanding and display technology to move beyond the traditional computer display, keyboard, and mouse-controlled cursor associated with the traditional human–computer interface. We believe that the basic approach holds promise for bridging the gap between traditional human–computer interfaces and the needs of real-time diagnostic work for a variety of applications, but especially for use in *procedure-oriented* specialties such as emergency medicine and surgery.

We focused on the domain of medical decision problems. In procedure-oriented medical specialties, automated reasoning and decision support should not interfere with a physician's hands, and cannot obscure the physician's view of the patient. This work has extended earlier efforts on combining decision-theoretic systems with personal displays for medical diagnosis and machine repair (Horvitz 1992; Horvitz and Shwe, 1995).

We initiated our project with the intuition that the problem of working on a handsfree interface for emergency medicine would be a straightforward problem in human-interface engineering. We discovered that the application motivated work in several areas reaching beyond the simple meshing of speech understanding and an automated reasoning system. In particular, we found that it can be critical to design effective feedback and recovery for failures of speech understanding. We also were stimulated to begin new investigation on principles for developing and integrating models of how the information being displayed to users can influence their utterances. Such modeling promises to be useful for better predicting utterances, and, thus, adapting the speech understanding system to dynamically changing diagnostic contexts. In the first half of this paper, we shall describe the initial system and its components and review some basic design issues. In the second half of the paper, we will present new research problems and directions for solving problems catalyzed by the application.

2 Handsfree User Interface

At the core of the handsfree diagnostic system are Bayesian-network–based inference procedures. We constructed and assessed Bayesian-network knowledge bases for problems in time-critical medicine, including a knowledge base for diagnosing potntially time-critical abdominal pain in the emergency department and criticality assessment for trauma care at the site of an injury (Horvitz and Seiver, 1997).

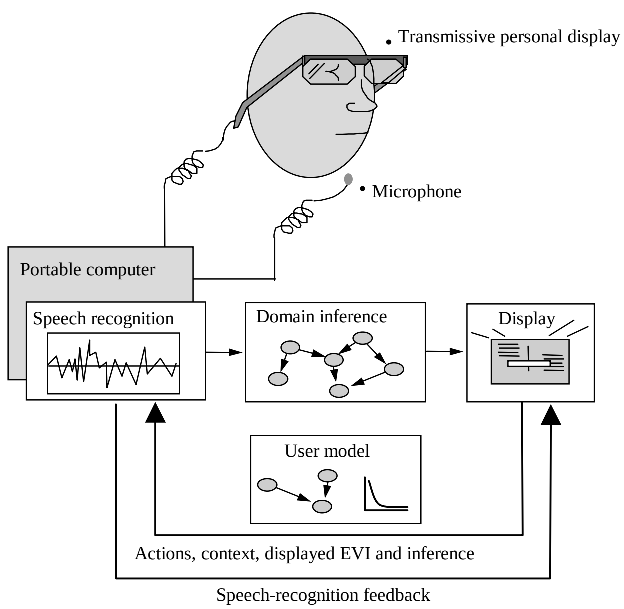

Figure 1. Key components of handsfree decision-support system. A transmissive personal display relays speech recognition feedback as well as core information on the differential diagnosis and the best information to gather. As indicated in the figure, user models promise to extend the accuracy of systems in noisy environments by predicting the probability of utterances given the displayed output of diagnostic inference.

The inference system was combined with a speaker-independent speech-understanding system, and a personal display system. We experimented with several speech understanding systems including packages from Speech Systems, IBM, and Dragon. We also experimented with several display systems including a transmissive personal display, manufactured by Virtual I/O. A schematic overview of the handsfree decision-support system is displayed in Figure 1.

The speech-understanding system allows a user, through voice commands, to navigate through menus and to enter information about signs, symptoms, and test results considered by the diagnostic system. The transmissive display overlays the expert system's output on the clinician's view of the world. The display continues to provide the user with a variety of input and inference options. In our tests, we provided users with recommendations on best additional information to acquire and the computed differential diagnosis with associated probabilities of diseases represented with an adjacent bar graph.

We explored design issues for hands-free decision support, including the design of the user-interface controls for managing information in a personal, heads-up display, and controlling the display of information with speech and context cues. Beyond run-time functionality, we investigated tools for constructing handsfree systems. As part of this effort, we built a development environment for reducing the cost of constructing and training the speech-understanding component of the system. The tool builds grammers for language models automatically from a specification of the functionality of the interface and predicates defined in the knowledge bases.

3 Decision-Theoretic Diagnostic Systems

We focused our handsfree decision-support efforts on developing an effective user interface for interacting with *decision-theoretic* diagnostic systems. There have been significant developments over the last decade in representing and performing inference with probabilities and utilities for diagnosis (Breese et al., 1994; Heckerman, et al. 1992; Horvitz, et al. 1988). In particular, there have been strides in the use of the Bayesian network representation to capture uncertain relationships between findings and explanatory hypotheses. Bayesian networks are directed acyclic graphs describe the joint probability distribution of a domain (Pearl, 1989). Such Bayesian diagnostic systems perform probabilistic inference given any set of findings observed in a patient or other complex system, assigning a probability to feasible hypotheses. For medical applications, Bayesian inference procedures yield a ranked list of diseases, or a *differential diagnosis*, in response to the signs and symptoms input by a decision maker. Such systems can be extended with additional procedures that employ an information-theoretic or decision-theoretic analyses for computing the best information to gather next, given the current state of knowledge (Ben-Bassat, 1978). Such procedures center on the computation or approximation of *the expected value of information* (EVI) (Howard, 1967). Typically, such information-gathering is based on a myopic analysis, assuming that a patient will be treated immediately after that information is collected. The systems operate with an iterative refinement pattern called *hypothetico-deductive reasoning* or the *method of sequential diagnosis* (Gorry 1968). With this approach, depicted in Figure 2, salient findings are input to the system and a differential diagnosis is constructed. Then, the best next findings are computed through an expected value of information procedure. When no test or observation is available that has informational value greater than its cost, a final diagnosis is rendered or a therapeutic action is recommended. Psychologists have found that human diagnostic behavior tends to follow a hypothetico-deductive pattern (Elstein and Sprafka, 1978). The behavior of these systems provides a set of crisp, well-defined speech understanding and display tasks.

4 Speech Recognition and the User Interface

After construction knowledge bases for emergency medicine and trauma care, we began work to weave the inference system into a handsfree system. To support this effort, we explored relevant work on speech-recognition interfaces studied by in the human factors and ergonomics communities. We found previous results useful in designing interfaces for handsfree decision support. We shall briefly review this relevant literature.

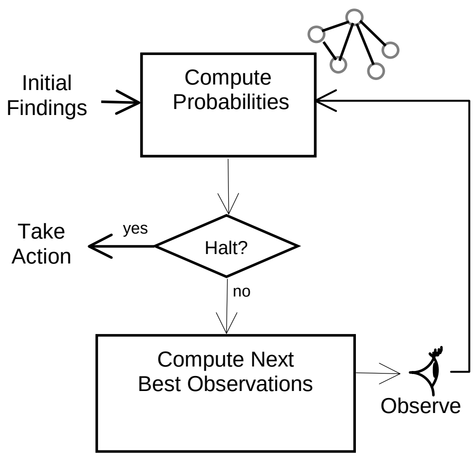

Figure 2. The hypothetico-deductive cycle. After initial findings are input to the system, Bayesian inference procedures operating on a Bayesian network knowledge base compute probabilities of competing hypotheses. Then the next best findings are computed. The cycle continues until action is taken.

4.1 Relevant Work on Speech and the User Interface

Jones, et al. (Jones, et al. 1989) present guidelines for speech-based user interfaces. The study recommends that speech commands should be used consistently, and that speech should not be used to manipulate graphical objects or to note screen position. The study also indicated that speech input should have a special vocabulary, and that commands should be designed so as to maximize acoustical distinctness. Perhaps most importantly, the investigators found that explicit feedback should be provided to users about the activity of the speech recognizer. In particular, the timing of the feedback is significant: the user should be appraised when the recognizer has recognized or rejected an utterance. Auditory feedback was found to interfere with the user's memory of complex control tasks such as sequences of commands.

Martin (Martin 1989) describes work on efficiency gains and losses with speech-based interfaces. She reviews evidence indicating that humans can operate computers more quickly using speech interfaces than with traditional inputting. Moreover, the gains in efficiencies are greater than explained by merely eliminating the hands as a data-entry bottleneck; Martin contends that the use of the human auditory channel in addition to the normal visual channels allows significant parallel processing of information. This result suggests that speech-understanding systems could enhance the efficiency with which healthcare personnel work with computers while responding to time-critical decision problems.

Zajicek and Hewitt (Zajicek and Hewitt 1990) investigated the value of different types of dialogue for error-recovery in speech-driven graphical user interfaces. They found that approaches to recovering from failures employing such tools as a set of questions requiring a yes or no answer were confusing. The natural tendency for people is simply to repeat a word that is not recognized. Users tend to be confused by methods for error recovery that ask questions to clarify misunderstood words. We took into consideration such findings in designing the handsfree system.

A related study in speech-driven applications for clinical data input (Shiffman, et al 1995; Johnson, et al 1992) have utilized elaborate grammars for interaction. These grammars and synonym lists become necessary when the list of findings becomes so large that the speaker could not be expected to know the exact names of the findings. Indeed, in the application areas of internal medicine and general history taking, the number of findings can be well into the thousands. The QMR knowledge base of internal medicine, which (Shiffman, et al 1995) use in their application, contains over 4,000 findings.

4.2 Speech Recognition Systems

Speech recognition systems can be *continuous* or *discrete*. Continuous speech recognition programs attempt to recognize words in flowing speech. They are based on consideration of an acoustic model of words based on phonemes. By contrast, discrete recognition systems are designed to recognize specific words and phrases. In many continuous speech understanding systems, acoustic models are based on the spelling of a word, and, thus, no training of the words in the vocabulary is required. Most approaches to discrete speech understanding make use of stored sound models of the words or word phrases.

Given the well-defined task of understanding a predefined set of commands and discrete knowledge-base distinctions displayed to a user, we initially explored a discrete speech recognition system. We selected the Dragon VoiceTools software for discrete speech modeling. We later experimented with a continuous speech systems from IBM and Speech Systems. We will focus mainly on the discrete speech implementation.

5 Design of a Handsfree User Interface for Diagnosis

The user interface for the handsfree decision-support prototype was based on the preexisting graphical user interface developed for a commercially available decision-theoretic diagnostic system for Microsoft Windows named WIN-DX.[2] We incorporated results from the earlier studies of speech recognition into a modification of the graphical user interface for the handsfree system. A view of the user interface is displayed in Figure 3.

As an example of integrating results from prior studies, we took into consideration the suggestion by Jones, et al.'s that voice is a poor means of controlling window placement. Since overlapping windows will lead to inefficiencies in data input, we based all of our data input on dialog boxes that expand in place, within larger, scrollable windows. For example, in Figure 3, the question box prompting for information on a surgeon's evaluation of the number of air-fluid levels on a three-way abdominal film expands when the user is setting the value of this finding. The question box contains the span of possible values and a set of explanatory bar graphs representing the log-likelihood ratios associated with the different answers. The bar graphs shows how different finding values will affect the diagnostic differential.

5.1 Introducing Feedback and Error Recovery

Given the findings of the Zajicek and Hewitt study, we were motivated to develop and integrate simple speech-recognition feedback and error-recovery mechanisms. Also, to ensure that all aspects of the system were handsfree, rather than require a user to manually press a button to invoke listening, we developed a simple command, "*mic*," (for *microphone*) to toggle the speech recognition on and off. When the speech recognition is in the off state, the status bar indicates that the system is sleeping, and will reject all utterances besides the "*mic*" command.

When listening for commands, WIN-DX places a pop-up box prominently in the middle of the screen displaying the current state of recognition. For phrases that are delivered in two segments, a status bar at the bottom of the screen displays the currently recognized phrase and expected class of words that can be used to complete the input. The status bar and pop-up window are displayed in Figure 3. When the system does not recognize an utterance, the pop-up box relays "Unknown word."

Also, the two-part status bar at the bottom of the screen provides real-time feedback to the speaker. The right-hand side of the status bar reads "Idle" when the speech recognizer does not detect speech and reads "Listening" when it does. The left- hand side shows the current command and the options for completing the command. When word rejection is indicated, a user can immediately repeat a word or phrase.

[2] WIN-DX is a trademark of Knowledge Industries. MS Windows is a trademark of Microsoft Corporation.

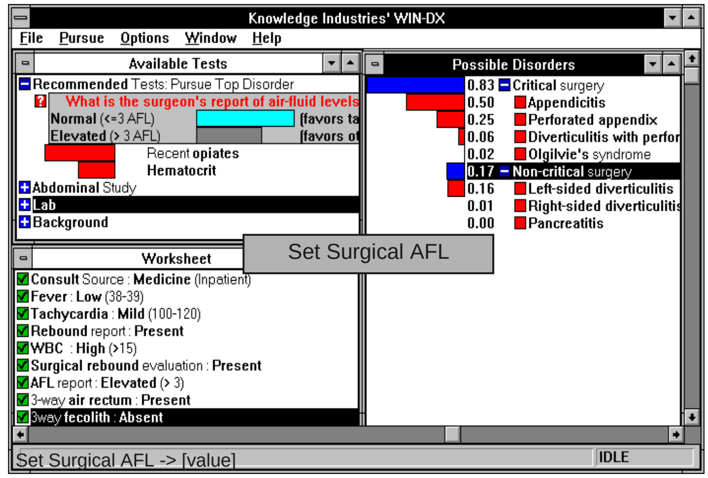

Figure 3. Design of speech-controlled version of WIN-DX. The pop-up box in the middle of the screen displays the current state of speech recognition. The status bar, at the bottom, left corner of the screen, displays the currently recognized phrase and expected completion.

5.2 A General Command Syntax for Diagnostic Reasoning

We developed a command syntax for WIN-DX to support efficient data entry access to diagnostic feedback. The syntax allows users to observe findings; change the values of observed findings; select different test-recommendation strategies, and scroll windows (page up and page down operations). Commands can also be issues to collapse and expand disease categories. Managing abstractions of disorders can be important in diagnosis. The WIN-DX system has the ability to display and perform inference about disease categories rather than all of the diseases (Horvitz, et al. 1989). Speech commands can also be used to change the level of abstraction at which the system performs inference and displays diagnoses pattern (Elstein and Sprafka, 1978).

6 Automated Generation of Language Models

To maximize the efficiency of handsfree systems, we built a functionality to automatically build a language model with input distinctions derived from the functionality of the diagnostic system, as well as the observations and distinctions considered in specific Bayesian network models. We designed a software module that automatically generates a speech vocabulary file from the words embedded in the user-interface menus, the way users interact with the functionality of the system, and distinctions in the knowledge base. The vocabulary file is augmented with the command syntax that we developed for efficient navigation in the system. A software tool, named Speech DX Wizard, was constructed to populate a language

model. The system also guides the training of the system by iterating through the vocabulary during a training phase, directing the recording of utterances of multiple users for the vocabulary generated from a knowledge base. We then use the commercial VoiceTools package to aggregate the utterances of the multiple speakers, to create a speaker-independent model. Speak-DX employs several cross-validation functions, providing statistics on the accuracy of recognition

We explored methods that dynamically prune the vocabulary based on the displayed information generated by the hypothetico-deductive inference cycle. We also explored methods that would allow WIN-DX to place a higher recognition priority on terms that it expects the user to speak. We also ensured that WIN-DX would flexibly allow speakers to interrupt the current command, allowing them to initiate a new command. For example, a user would ordinarily would say "*Set, Surgical AFL*" to herald the forthcoming input about the finding of a surgeon's evaluation of air-fluid level on a three-way abdominal study. A user would be expected to report either "*Normal* or "*Elevated*" to set the value of the finding. WIN-DX places a high recognition priority on the values presented as options for the finding. We can interrupt the command by explicitly uttering "*Cancel*," or by initiating another command. In this manner, we retain command flexibility while providing context-sensitive recognition.

7 Performance of the Implemented System

We did not perform a formal validation in an actual clinical setting. However, we validated the speech component of the system with a Bayesian-network knowledge base for diagnosing severe abdominal pain in patients in several contexts, including emergency department and inpatient settings. We found that the user-interface—focused language model performed well in quiet settings. On a two-hundred word vocabulary, trained with 6 speakers, we found that the overall recognition was greater than 96% before addition of the menu-driven focusing provided by the diagnostic system. For use with the diagnostic system, we found that we could adapt the discrete speech paradigm to follow natural patterns of pauses between phrases that users employ to enter data to the expert system by setting timer thresholds so that users can issue a command such as "*Set surgical AFL, elevated*" with only slight pauses between "set", "surgical AFL," and "elevated."

We were initially content with the performance of the speech recognition software in conjunction with the feedback conventions we employed in the system. However, we noticed that the accuracy of recognition could be diminished significantly by operation in noisy environments. The paradigm established by the handsfree application—where users could be assumed to be influenced in an ongoing manner by the context established by the displayed results of hypothetico-deductive inferential process---stimulated additional research on principles for dynamically updating the language models employed in speech recognition. In particular, our pursuit of a handsfree diagnostic system for time-critical medicine catalyzed our pursuit of means for exploiting synergies between expert system inference and speech understanding to make the voice recognition more accurate and noise tolerant.

8 Promise of User Modeling for Dynamic Language Models

Our efforts to increase the accuracy of speech recognition in noisy environments focused us on the problem of introducing context dependency into the language model being used by the system. A straightforward approach to introducing context-sensitivity and reducing the space of recognition possibilities is to dynamically prune the vocabulary based on the displayed information generated by the hypothetico-deductive inference cycle. A more principled approach is to develop strategies for assigning likelihoods to alternate utterances that the system might hear from the user. The promise of this methodology focused us on task of developing principles for estimating the probability of different utterances based on the results of inference, the configuration of displayed information, and the typical behaviors of users in responding to displayed information. Estimating the probabilities of utterances would enable a speech recognition system to employ a dynamically focused, context dependent language model. We shall review the foundations of generating a context-dependent language model for diagnostic systems.

In probabilistic approaches to speech understanding employed by a number of speech recognition systems, the probability of words or a stream of words associated with an utterance is computed with the following Bayesian strategy,

$$p(words \mid speech) = \frac{p(words)\, p(speech \mid words)}{p(speech)} \tag{1}$$

where *speech* refers to the acoustic signals detected by a speech system and *speech* is the string of words a user generates when attempting to interact with the speech recognition system. If we could infer the probability of the user's *intended* communicated concept, given evidence about goals and context provided by the state of a case or of a user's recent actions, we could update the likelihood that specific utterances will be used. Instead of assuming that a user's utterance is simply one of a set of equally likely possible utterances in a context, we attempt to update the probability of utterances given evidence about the context, the nature and configuration of objects that are being displayed, and about the user, including such variables as a user's expertise and recent actions. Let us use *E* for a vector of evidence about the context and user. Rather than assuming a static database of word strings allowed by a context, we update the probability of word sequences as follows:

$$p(words \mid E) = \sum_i p(words \mid concept\ i, E)\ p(concept\ i \mid E) \tag{2}$$

where *concept i* is the concept the user intends to communicate to the system. We can now revise Equation 1 to yield the Bayesian update on the probability of phrases given evidence about context, actions, and displayed information *E*,

$$p(words \mid speech, E) = \frac{\sum_i p(words \mid concept\ i, E) \times p(concept\ i \mid E)\, p(speech \mid words)}{p(speech, E)} \tag{3}$$

More generally, in real time, we can continue to fold in the last *n* words recognized in a phrase, and continue to compute the probability of the next word, or set of words, *i.e.*, computing the probability, *p*(*concept i* | *E, words already recognized*).

Hypothetico-deductive diagnostic systems provide a relatively structured dialog with users about determining the nature of disorders in complex systems. Let us consider an example of how we can harness knowledge that a user is viewing a diagnostic display to generate a dynamic language model.

After the analysis of new findings in a diagnostic session, a hypothetico-deductive system recomputes the list of recommended observations and displays an updated list of hypotheses. A probabilistic *user model* can take advantage of the overall structure of this cycle of diagnosis to predict the likelihood that a user will make alternate utterances about the status of new observations. We wish to compute the probability, *p*(*speech* | *E, display, context*) of different utterances, given the previously observed set of evidence *E,* the information displayed to the user, and the current state of the diagnostic session. To compute this probability, we need to know the likelihood that a new observation, E', from the set possible tests will be selected by the user, and the likelihood that a value, e, of a range of potential values of the finding will be noted upon evaluation of the observation. More precisely, we need to model the probability that each finding will be evaluated, *p*(*E' selected* | *E, display, context*), the probability that a specific value will be reported for that finding, *p(E'=e|E, display, context),* and the probability that the user will produce one of a set of different utterances for the concept represented by the finding and test result, *p*(*speech* | *concept(E'*,*e*)), where *concept(E'*,*e*), refers to the user's intention to communicate to the diagnostic reasoner that making a specific new observation *E'* has revealed state *e*. Assuming that the probabilities of utterances for findings and values are independent of the display and the context, we have

$$p(concept(E',\mathrm{e})|\, E, display, context) = \mathrm{p}(E'\ selected \mid E, display, context) \times p(E'=e|\, E'\ selected, E, display, context) \qquad (4)$$

and,

$$p(speech \mid E, display, context) = \sum_i p(speech \mid concept_i(E',e)) \times p(concept_i(E',\mathrm{e})|\, E, display, context) \qquad (5)$$

Thus, by encoding for each test the likelihoods of different feasible utterances for describing the test, and considering the probability that each test will be selected, we can compute the likelihood of utterances for all available tests.

Let us consider how we can generate key probabilities required for the dynamic language modeling on the fly. A simple example of the use of a user model for estimating the probability, *p*(*E' selected* | *E*, *display, context*), is a parametric function that provides an estimate of the probability that user will input a finding from a displayed ordered list of recommended findings, taking as arguments the current display of inferred information as well as information derived from the dynamically computed expected value of information used to prioritize the recommended findings. Such a function is depicted in Figure 4. We model the likelihood that a test will be selected as a function of the order of the test in a list of tests sorted by expected value.

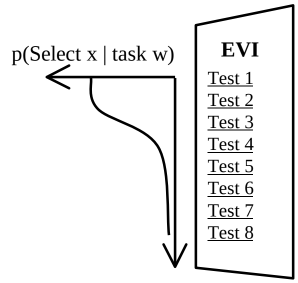

Figure 4. Use of a function that estimates the probability of an intended input as a function of the nature and configuration of displayed information. In this case, we highlight a probability distribution over the utterance of a recommended test as a function of the order in a list sorted by the expected values of the tests.

The probability that a *specific value* for that finding *e* will be reported, *given* that a finding is evaluated, *p(E'=e|E, display, context),* is directly available from the inference engine; this probability is available from the domain-level inference. It is computed as part of the expected value of information procedure. That is, for each new finding *E'* potentially input by the user, we have available from the Bayesian inference model the probability of seeing each new result, *e*, in the context of a set of previously observed evidence, *p(E'=e|E)*.

We have assessed functions that capture the probability that a user will utter a phrase as a function of the expected value of evaluating the finding and the position of that finding on the recommendation list, as well as the probability of different tests results being reported current context available from the inference system.

We can extend the principles described for generating language models for the input of findings to other tasks in the context of a structured interface for hypothetico-deductive diagnosis. As portrayed in Figure 5, we consider the probability of each high-level task (inputing a finding versus accessing information about a disease) as a function of the state of the case, and then consider the likelihood of each subtask (e.g., test or disease) and specification (e.g., test result, disease information class) as a function of the displayed information, inferred beliefs, and expected value of information.

We similarly employ probabilistic user models in conjunction with likelihoods computed by domain-level Bayesian inference for computing the probability that different available tasks and substasks will be invoked by a user. Coupling the probabilities that users will intend to communicate a target concept with assessed or inferred likelihoods that different utterances would

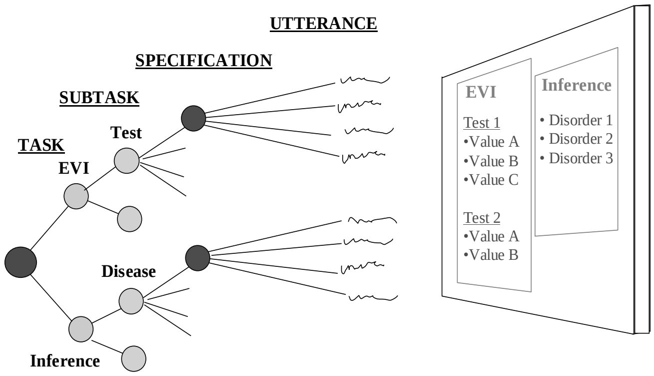

Figure 5. We can model the likelihood of different utterances as a function of the context, defined in the case of a medical decision support system, as the phase of the case and displayed distinctions. We highlight here the modeling of probability of utterance as a function of the task and subtask for accessing information about lists of diseases and recommended tests.

be used to express these concepts yields a probability distribution over utterances.

Moving beyond simple functions of expected value and ordering of a list of items, we can consider building richer probabilistic user models that take into consideration such variables as user expertise, and the influence of prior utterances and changes in displayed information on the next utterances. Such richer user models can be represented as Bayesian networks. A Bayesian user model that takes into consideration the influence of changes in displayed information and prior interactions is displayed in Figure 6. The Bayesian user model expresses the influence on utterances at the present moment, t_o , of the current display, the prior interaction, the previous display, and user expertise.

7 Research Directions and Applications

Our initial goal centered on building and demonstrating an integrated handfree diagnostic system by coupling decision-theoretic inference with a speech understanding system and personal display. Our efforts to leverage the results and display of decision-theoretic advice to enhance speech understanding led first to straightforward mechanisms for manipulating the language models based on the functionality of the diagnostic system, and on disorders and observations represented in the diagnostic knowledge bases. Although our handfree prototype based on such focusing performed satisfactorily in controlled environments, our interest in further bolstering the ability of the system to perform in noisy environments stimulated us to explore handsfree decision support systems. We hope to see these principles employed for developing even more effective handsfree systems. We believe that richer probabilistic user modeling hold opportunities for leveraging domain level inference and knowledge about user behavior to make yield more effective handsfree systems. In particular, we believe that dynamic probabilistic language models hold opportunity for enhancing significantly general speech recognition for complex

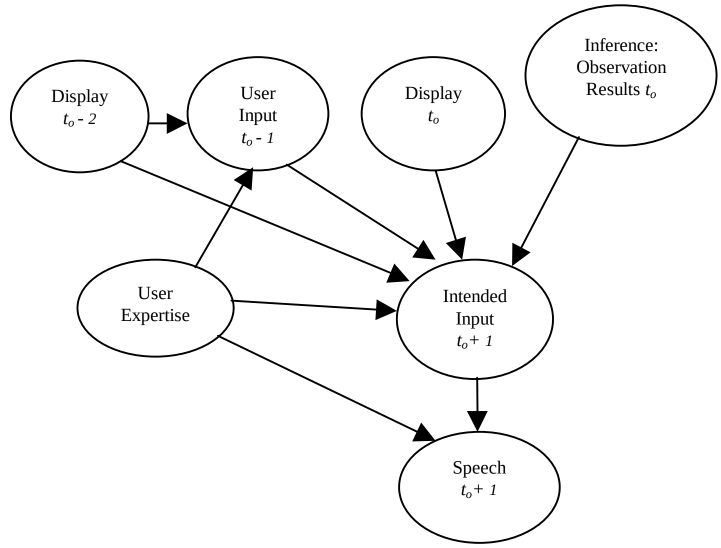

Figure 6. Bayesian network representation of a user model for utterances. The richer dependency model can be used to infer the probability distributions over utterances, given the last and current displayed information as well as the last user action. User expertise and uncertain inference about the state of currently unobserved results are also considered.

and noisy domains. Additional attention to user modeling and exploitation of knowledge of the task and display will likely be valuable for enhancing the efficiency of interaction for more complex problems. We plan to pursue the development and integration of such richer user models.

Other areas of investigation include innovations in display and input techniques. For example, there is opportunity for designing new methods for controlling the nature and configuration of multiple windows of information in a handsfree system. Windows can be controlled through additional interface modalities, such as systems that track a user's gaze. Making use of such additional information as a user's gaze can also reveal valuable evidence about a user's attention which can be considered in user models.

We suspect that refined version of handsfree decision-support systems will be valuable in a variety of applications. The systems may be particularly useful for assisting decision makers with time-critical decision making where time delays can be associated with significantly increased rick, morbidity, and mortality (Horvitz and Rutledge, 1991; Horvitz, et al. 1992).

We are excited about the many possibilities for using handsfree user interfaces to integrate computer-based systems into interactive clinical activities. We foresee handsfree methods as providing one of several innovations at the human–computer interface that will enhance the diffusion of computer-based reasoning and information into the real world.

Acknowledgments

We are indebted to Dan Sommerfield for assistance with programming on the handsfree system and development environment. We thank Paul Dagum for discussions on applications of personal displays for use in surgery, Humam Kakish and Adam Seiver for providing expertise in assessing the Bayesian networks, Rick Satava for ideas on the value of personal displays for telemedicine applications, and John Silva for enthusiastic support for user-interface innovations for clinicians.

References

(Ben-bassat, 1978) M. Ben-Bassat, Myopic policies in sequential classification, IEEE Transactions on Computers, 27:170-178, 1978.

(Breese et al., 1994) D. Heckerman, J. Breese, K. Rommelse, Decision-Theoretic Troubleshooting, Communications of the ACM, 38(3), 1995.

(Elstein and Sprafka, 1978). A.S. Elstein, L.S. Shulman, S.A. Sprafka, Medical Problem Solving: An Analysis of Clinical Reasoning, Harvard University Press:Cambridge, MA, 1978.

(Gorry 1968) G.A. Gorry and G.O. Barnett, Experience with a model of sequential diagnosis, *Computers and Biomedical Research*, 1:490-507, 1968.

(Heckerman, et al., 1992) D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani. Toward normative expert systems: Part I. The Pathfinder project. *Methods of Information in Medicine* 31:90-105, 1992.

(Horvitz 1992) E.J. Horvitz. Innovations at the Human–Computer Interface: A Medical Informatics Perspective. *Medicine Meets Virtual Reality: Discovering Applications for Three-Dimensional Multimedia in the Health Sciences.* Department of Surgery, University of California at San Diego, San Diego, June 1992.

(Horvitz, et al. 1988) E.J. Horvitz, J. Breese, and M. Henrion. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, Special Issue on Uncertainty in Artificial Intelligence, 2:247-302.

(Horvitz, et al. 1989) E.J. Horvitz, D.E. Heckerman, K. Ng, and B.N. Nathwani. Heuristic abstraction in the decision-theoretic Pathfinder system. In *Proceedings of the Thirteenth Symposium on Computer Applications in Medical Care*, Washington, DC, pages 178-182. IEEE Computer Society Press, Silver Spring, MD, 1989.

(Horvitz, et al. 1992) E. Horvitz, C. Ruokangas, S. Srinivas, M. Barry, Decision-Theoretic Approach to the Display of Information for Time-Critical Decisions: The Vista Project, *Proceedings of SOAR-92, Conference on Space Operations Automation and Research*, National Aeronautics and Space Administration, Houston, TX, August 1992.

(Horvitz and Rutledge 1991) E.J. Horvitz and G. Rutledge. Time-dependent utility and action under uncertainty. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, Los Angeles, CA, pages 151-158. Morgan Kaufmann, San Mateo, CA, July 1991.

(Horvitz and Seiver, 1997) E. Horvitz and A. Seiver. Time-Critical Action: Representations and Application. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, Providence, RI, pages 250-257. Morgan Kaufmann, San Francisco, CA, July 1997..

(Horvitz and Shwe, 1995) E. Horvitz, and M. Shwe. Handsfree Decision Support: Toward a Noninvasive Human—Computer Interface. *Proceedings of the Symposium on Computer Applications in Medical Care*, IEEE Computer Society Press, November, 1995.

(Howard, 1967) R.A. Howard, Value of Information Lotteries, IEEE Transactions of Systems Science and Cybernetics, SSC-3(1):54-60, 1967.

(Jones, et al. 1989) D. Jones, K. Hapeshi, and C. Frankish, Design guidelines for speech recognition interfaces, Applied Ergonomics, 1989, 20:47-52.

(Martin 1989) G. Martin, The utility of speech in user-computer interfaces, International Journal of Man-Machine Studies, 1989, 30:355-75.

(Pearl, 1988) J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers:San Francisco, 1991.

(Zajicek and Hewitt 1990) M. Zajicek and J. Hewitt, An investigation into the use of error recovery dialogues in a user interface management system for speech recognition, Human-Computer Action — INTERACT '90, Elsevier Science Publishers, 1990, 755-60.

(Shiffman et al 95) S. Shiffman, W.M. Detmer, C.D. Lane, and L.M. Fagan, A continuous-speech interface to a decision-support system: I. Techniques to accomodate for misrecognized input, , J Am Med Informatics Assoc. 1995; 2:36-45.

(Johnson et al 92) K. Johnson,., A. Poon, S. Shiffman, R. Lin, L.M. Fagan, A history-taking system that uses continuous speech recognition. Proc Ann Symp Comput Appl Med Care. 1992:757-61.