Development of an Expert System for Pediatric Auditory Brainstem Response Interpretation

Anne Marie Tharpe*
Gautam Biswas†
James W. Hall III ‡

Abstract

Expert systems are computer programs which incorporate artificial intelligence technology and are created to emulate the decision-making abilities of human experts. The advantage of such systems lies in their ability to capture and model expert problem solving knowledge in a domain and make it available to an unlimited number of consumers in an economic and efficient way. The purpose of this project was to develop an expert system to interpret infant auditory brainstem response data as entered by the user. The resulting system provides diagnostic conclusions regarding hearing status, type of hearing loss, and brainstem function at an accuracy level equal to that of a human expert.

Key Words: Artificial intelligence, auditory brainstem response (ABR), computer technology, expert system, pediatric audiology

Although computer systems have been used in the medical setting for decades, their use has primarily been limited to medical record keeping and interpretation of laboratory data. In the mid-1970s, however, a flurry of research introduced expert system technology to medicine (Shortliffe et al., 1975; Pauker et al., 1976; Kunz et al., 1978; Weiss et al., 1978; Kulikowski and Ostroff, 1980; Fieschi et al., 1982; Jaganathan et al., 1982; Miller et al., 1982; Puppe et al., 1983). Expert systems are computer programs created to emulate the decision-making abilities of human experts. The goal in developing such a system is for the computer to mimic an expert's decision-making process in complex problem solving. The advantage of such a system lies in its ability to capture and model expert problem solving knowledge in a domain and make it available to an unlimited number of consumers in an economic and efficient way.

The domain of audiology appears to be well-suited to expert system applications. The scope of practice in audiology is an ever-expanding one that includes pediatrics, geriatrics, neurodiagnostics, and rehabilitation among others. It is quite likely, therefore, that audiologists are often presented with cases that require detailed expertise from more than one of the above domains. Since the complexity in reasoning increases significantly when one has to draw upon and combine knowledge from multiple domains, an expert consultation system could provide a computer-based second opinion that alerts the audiologist to appropriate test procedures and specialized methods for collecting relevant information, and then helps in the interpretation of results.

Unfortunately, there is a dearth of information on expert system development in audiology. The primary work in this area has been conducted by Bareiss (1989) who developed PROTOS, a knowledge-based learning apprentice that aids users in classifying adult hearing disorders such as Meniere's disease or acoustic neuroma. The user provides information on the patient's pure-tone classification, speech discrimination ability, history, and impedance data. PROTOS classifies a new case by using an exemplar, i.e., a case which has been previously...
processed and retained and has some similarity to the new case, as a model for interpreting its features. Features are attributes used to describe a case, or test results abstracted into qualitative descriptors. PROTOS then uses a case-based reasoning approach to match these data against a repository of typical cases to find one that best matches the given data. Domain knowledge acquisition, therefore, involves the building of a repository of typical cases and forming general concepts from similar cases. When the system cannot find a good match between data provided and case prototypes, and cannot explain the differences, it initiates an interactive session with a human expert. In this case, and when the human expert determines that it has an incorrect solution to a problem, a focused dialogue is conducted with the expert to acquire new domain knowledge or to remove inconsistencies in existing knowledge. PROTOS, therefore, learns from cases and explanations. In addition, it learns as a by-product of attempting to perform classification under a teacher’s, or expert’s, guidance. Its accuracy in classifying hearing disorders was 82 percent on 200 training cases and 100 percent on 26 test cases after training (Bareiss, 1989).

A second diagnostic prototype system in audiology was developed as a rule-based expert system at the National Institutes of Health using a commercial expert system shell called GURU (DeLeo et al., 1988). This is a small system that interprets auditory brainstem responses (ABRs) using fifteen rules. The system is provided with adult ABR data and concludes whether the response is normal, questionable, or abnormal.

Whereas both of the above systems deal with adult disorders, pediatric audiology poses unique diagnostic challenges to audiologists. The difficulty in making diagnoses of hearing loss in this population is apparent when one considers that severe to profound hearing losses are typically not identified in the United States until children are approximately 2 years of age (Gustafson, 1989). Part of the difficulty lies in the interpretation of audiometric data obtained on this young population. One must consider developmental age, gestational age, and chronological age when interpreting test data used in diagnosis of hearing loss. In other words, adjustments must be made if a child is premature or developmentally delayed. This is crucial for correct interpretation of subjective and objective test data. These considerations must be applied to behavioral as well as electrophysiological data, thus requiring a certain level of expertise in the domains of electrophysiology and behavioral pediatric testing.

The two systems described above, PROTOS and the one developed from GURU, each took a single area within the domain of audiology, i.e., the traditional adult test battery and ABR respectively, from which to build the knowledge base. As indicated, the diagnostic process in pediatric audiology requires expertise in several areas of audiology including behavioral assessment and electrophysiology. A unique challenge in the development of an expert system for pediatric audiology is the need to combine the expertise of two human experts, that is, one in behavioral assessment and one in electrophysiological assessment of infants. The first step in this endeavor was to develop a diagnostic expert system that could interpret infant ABR data at a level equal to that of a human expert. The pediatric ABR model consists of several diagnostic steps, each requiring its own development. Specifically, these include determination of hearing sensitivity utilizing threshold and latency values of waveforms and patient history, type of hearing loss utilizing threshold and latency values of waveforms and tympanometric data, and brainstem status utilizing absolute and interwave latencies and patient history data (Fig. 1). This system is designed and implemented as a rule-based expert system. The rest of this paper describes the design, development, and testing of this system.

**METHOD**

The development of this expert system, named SIMON,1 required a multi-step incremental development process that included designing the pertinent domain models, building the knowledge base, coding the system using EXSYS (1988), an expert system shell, running a number of test cases for knowledge verification and validation, and modifying the system to achieve correct diagnostic results. We describe each of these steps in detail below.

**Constructing the Domain Model**

**Knowledge Base Development**

The process of building an expert system is termed knowledge engineering. This refers to
the tasks of acquiring problem-solving knowledge from human experts and other sources, and organizing and coding it into an expert system. The task of organizing and coding knowledge involves two steps. One is determining appropriate knowledge representation schemes (i.e., rules, a network representation, and/or frames). The other is either choosing an appropriate tool in which to encode the system or developing appropriate algorithms and code for encoding the knowledge and reasoning mechanisms that form the two key components of the system. In designing SIMON, we decided to use an existing expert system shell called EXSYS (1988) to avoid significant amounts of programming effort. This enabled us to focus primarily on knowledge base design in the initial phases of the project.

The development of the knowledge base for SIMON was three-tiered. First, expert knowledge was elicited through a regular series of dialogues with a human expert and from published data (Jerger, 1970; Gorga et al, 1987, 1989; Eggermont and Salamy, 1988; Joint Committee on Infant Hearing, 1991; Hall, 1992). Second, the knowledge was explicitly coded in the knowledge base. Unlike the traditional approach to the building of expert systems where the human knowledge engineer serves as an intermediary between the expert and the system, the coding was conducted by a professional in the domain of audiology (the first author). As discussed earlier, this was feasible because the well-developed expert system shell, EXSYS, that was used provides reasonable interfaces for creating the rule base. Finally, the expert evaluated the knowledge base by running a number of representative test cases and providing critiques on the diagnostic conclusions. This process was iterated until the system's performance was judged to be satisfactory by the expert. It is important to note that these stages were implemented in an iterative and incremental manner, and were not necessarily sequential in nature. In other words, the knowledge base construction was incremental. A set of rules were created and tested. Additions, deletions, and modifications were made to these rules until their performance was satisfactory, and then the next set of rules was added to account for new situations.

**System Coding**

The primary characteristics of the EXSYS shell are that it: (1) is designed for rule-based programming; (2) incorporates forward and backward reasoning mechanisms; (3) allows representation of uncertainty in rules and allows the system designer to build evidence combination mechanisms; and (4) runs on an IBM compatible microcomputer using MS DOS or UNIX operating systems.

**Rules.** Rule-based programming arose from the inference schema known as modus ponens. Modus ponens, a logical rule, states that when A is known to be true, and a rule states if A then B, then it is valid to conclude that B is true (Harmon and King, 1985; Giarratano and Riley, 1989). In other words, if we find the premises of a rule to be true, it is reasonable to assume the conclusions are true. Rules represent modular chunks of knowledge. In a particular problem domain, rules represent heuristic and experiential associations that human experts make between available data or facts in the form of observations and test measurements and diagnostic conclusions of interest. Often, in complex problem solving situations, a system may go through a chain of intermediate conclusions to derive a final conclusion. Medical diagnostic systems are particularly amenable to rule-based programming because reasoning based on rules is intuitive and easy to follow. In order to gain acceptability, the user of a diagnostic expert system must be able to access an explanatory capability that clearly describes the reasoning process employed to reach a conclusion.

As discussed earlier, rules were formulated as members of three general classifications including hearing sensitivity rules, type of hear-
ing loss rules, and brainstem status rules. For example, type of hearing loss rules included tympanometric test results, wave I latency values, and wave V threshold values. Brainstem status rules included information on presence or absence of wave forms, absolute and interwave latency values, and neurologic history of the patient.

Since rules usually express the experiential or judgmental conclusions of human experts, the link between facts and conclusions may have a degree of belief, or certainty associated with them. Degrees of belief, or certainty factors (CF), are not probabilities but are subjective measures of confidence that we have in our data or conclusions.

Forward and Backward Chaining. A sequence of inferences that leads from a problem statement to its solution via a set of intermediate concepts is called a chain. A chaining process that traverses a problem from given data to its solution is called forward chaining. In other words, forward chaining reasons from facts to conclusions, which result from the facts. A backward chaining approach traverses from a hypothesis back to the facts, which support the hypothesis (Giarratano and Riley, 1989). In a diagnostic system, both chaining techniques are useful. For example, initially a program starts with a clean slate having no information with which to make a conclusion. The user can be prompted for patient data and then the system forward chains toward partial conclusions. Just as in human decision making, the program formulates an intermediate or partial hypothesis, and then works backward to find facts that support the hypothesis, or backward chains.

Uncertainty Management. An important issue arises in making inferences from rules that have confidence factors associated with them. As discussed earlier, the certainty factor (CF) is a subjective measure of the expert's confidence in the conclusions implied by the rule, given the evidence that forms the premise of the rule. The goal of the expert system is to derive relevant conclusions with as high a CF as can be achieved given the data provided by the user. The process of maximizing confidence, and minimizing or controlling these uncertainties, is known as uncertainty management. The primary method for achieving this is to look for multiple evidence to support a particular hypothesis or conclusion. An evidence combination scheme, similar to the one in the MYCIN² expert system (Shortliffe et al, 1975), is used to update belief values when multiple rules support a conclusion or conclusions.

Certainty factors are also used to represent the confidence that we have in a piece of evidence (Shortliffe et al, 1975). In SIMON, the user is prompted to enter a level of confidence in a piece of data. For example, an audiologist may be absolutely certain that the wave form identified is wave V or may question that, perhaps, it is a delayed wave III. The user selects one of the following descriptors of certainty:

1. definite;
2. very probably correct;
3. probably correct; or
4. somewhat uncertain.

The program assigns each of these descriptors with a certainty factor (CF), such as, 90 percent, 70 percent, 50 percent, and 30 percent, respectively.

In addition, as discussed above, rules have certainty factors associated with them. The program handles indefinite or uncertain information by propagating certainty factors. It also combines certainty factors obtained from different rules. In other words, a rule that states "if A and B and C then E" has an overall CF = C. SIMON then uses the following general formula to update the certainty factor in conclusion E:

\[
(CF_{E_{new}}) = (CF_{E_{old}}) + (100 - (CF_{E_{old}})) \cdot C_r \cdot \min(CF_A, CF_B, CF_C) / 100,
\]

where \((CF_{E_{old}})\) is the previous value of the certainty factor for E, \((CF_{E_{new}})\) is the updated value of the certainty factor for E given the new evidence (A and B and C), and \(\min(CF_A, CF_B, CF_C)\) takes the minimum certainty factor associated with the three pieces of evidence, A, B, and C, to weight the overall confidence. Note that the min function is a conservative estimate; it implies the overall belief in evidence with multiple components is the minimum of the confidence values of the individual evidences.

²MYCIN is a landmark medical diagnostic system for identifying and prescribing therapy for meningitis-related infections. It has served as a model for numerous subsequent rule-based expert systems in medicine and other domains.
For example, in order to conclude "Type of loss is conductive" and determine a CF for that conclusion, the following conditions must be met:

If:
- Wave I is present
- Wave I is delayed
- Hearing loss is present

Then:
- Type of loss is conductive

Let us assume that a previous rule based on tympanometric results has also concluded "Type of loss is conductive" with an overall CF of 65 percent. The user has provided a CF for confidence in the presence of wave I, for example 60 percent. "Wave I is delayed" has a CF of 90 percent and "Hearing loss is present" has a CF of 70 percent. The overall CF for the rule is 80 percent. Therefore,

\[
(CF_{E_{\text{new}}}) = 65 + (100 - 65) \times 0.8 \times 60/100
\]

i.e., the current belief that hearing loss is conductive is now 82 percent. If the user had provided a CF for confidence in the presence of wave I as 90 percent, the outcome for confidence in E would be:

\[
(CF_{E_{\text{new}}}) = 65 + (100 - 65) \times 0.8 \times 70/100
\]

a slight increase because belief in the evidence has risen from 60 percent to 70 percent.

These confidence values should be interpreted by the user as relative rankings. For example, if the diagnosis of conductive hearing loss has an associated confidence value of 60 percent, the user may look at the individual CFs of the rules and conclude that if more data is made available, such as tympanometry, the confidence in the diagnosis could be improved. Note that the assignment of certainty factors to individual rules is a difficult and time-consuming task that is hard to set correctly in a single attempt. The assignment of certainty factors to rules was adjusted throughout the program testing phase until the results satisfied the expert.

**Representative Case Dialogue**

This project resulted in a diagnostic expert system that interprets infant ABR data. At the start of a diagnostic session, the user is prompted for the following data:

1. Was wave I present or absent?
2. Was wave V present or absent?
3. Was tympanometry conducted?

Depending on his or her responses, the user is led through a series of queries that chart the following data:

1. Patient's age in weeks (conceptional age from 33 to 40 weeks if premature or chronological age if full-term);
2. Wave I latency in msec (at 60 or 80 dB nHL);
3. Wave V latency in msec (at 60 or 80 dB nHL);
4. Wave V threshold in dB HL;
5. User's confidence in the presence of waves I and V;
6. Case history data; and
7. Tympanometric data.

At the beginning of a run, SIMON provides an instruction screen to the user that specifies the test parameters that must be followed in order to receive accurate diagnoses. For example, specific stimulus parameters (i.e., rate, filters, intensity levels) must be adhered to in order for the normative values in the knowledge base to be accurate. The user is then prompted to enter the required data. As discussed above, queries were formulated carefully, so users have a fixed set of options from which to choose their response. In the actual program, a menu of choices appears with a query, and the user enters a response number, and associated CF if relevant (Fig. 2). Within a few seconds of providing all the relevant data requested by the system, the user is provided with the following information:

1. Hearing sensitivity classification based on click or tone burst thresholds (i.e., normal, mild, etc.);
2. Type of hearing loss, if present (i.e., conductive, sensorineural, or mixed);
3. auditory brainstem status (i.e., normal, abnormal); and
4. certainty factors associated with each of these conclusions.

The program also allows the user to query SIMON about what rule it is trying to affirm and how it reached certain conclusions. For example, SIMON may ask the user if the patient's age is term or pre-term. If the user wants to know why that question is being asked, she or he may enter the query, "Why?" SIMON responds by displaying the rule it is trying to fire with that data. This makes it clear to the user what conclusion the system is attempting to establish with this data. In addition, the user may enter, "How?" in response to an intermediate or final conclusion. SIMON will then reveal the confirmed rules that led to the conclusion at question. In addition, a hypertext file, which accompanies the program highlights certain key words on the screen and allows the user to inquire about their definition or meaning. This is a particularly useful feature for providing definitions and explanations of terms, such as conceptional age and gestational age.

Knowledge Verification and Validation

In a rule-based system, determining if new rules are in the correct form is termed verification of rules, while determining if a chain of inferences leads to a correct answer is called validation (Giarrantano and Riley, 1989). In order to verify and validate the SIMON knowledge base, it was placed in the Audiology Clinic at Vanderbilt University Hospital for a 2-month period. The users of SIMON were five certified audiologists, four master's level and one doctoral level, all of whom had extensive experience in ABR measurement with young children. They were instructed to use the program after they completed an ABR assessment and made a diagnosis. In this first prototype, the system was tested with only those cases for which behavioral data were not available. The audiologists were five certified audiologists, four master's level and one doctoral level, all of whom had extensive experience in ABR measurement with young children. They were instructed to use the program after they completed an ABR assessment and made a diagnosis. In this first prototype, the system was tested with only those cases for which behavioral data were not available. The audiologists were asked to record their original diagnosis and the computer's diagnosis for later evaluation. If there was a discrepancy between the two, the audiologists reported whether they still believed their original diagnosis or agreed with SIMON. A complete trace of the interaction between system and user for all cases was saved for later evaluation.

The purpose of this testing was two fold. The first was to record problems and make suggestions for improvement of the program. The second was to record diagnostic data that was later analyzed to determine the accuracy of the program. The test analysis focused on the following major problems:

1. incorrect answers;
2. incomplete answers; and
3. inconsistent answers.

Ideally, one might be interested in determining the accuracy of this tool to correctly diagnose hearing loss and brainstem function. Our purpose, however, was to determine the accuracy of this program in making the same diagnosis as an expert faced with the same set of data. For that reason, long-term follow-up of the patients from whom data were collected was not necessary for the current phase of the project.

System Modification

The final stage of this project was system modification. A first step of this stage was to summarize the results of the verification validation testing and determine needed changes to the system. Expert dissatisfaction with a diagnosis may be the result of several factors. First, there may be a missing rule that needs to be added to the program in order to make an accurate diagnosis. Another possibility is that the weights given to the intermediate confidence factors may result in a final confidence factor that is too high or too low when compared to the confidence that the expert would have in the same conclusion. In that case, the weights may need to be adjusted or one or more of the rules may need to be made more specific or more general in order to control when it fires.

Specifically, in this stage our efforts were directed to pinpointing which rule caused the problem that resulted in a bad conclusion, or diagnosis. When the rule was identified, adjustments were made until a satisfactory result was obtained. Such adjustments underwent repeated evaluation to ensure that test cases run previously which resulted in satisfactory conclusions were not adversely affected by the change. We continued this process of testing and modifying until we reached an acceptable level of system functioning.

RESULTS

The knowledge verification and validation stage was conducted with data from 55 cases that included children between the ages of
34 weeks gestational age and 3 years chronological age. The cases included infants with a full range of degrees and types of hearing impairment. As noted in Figure 3, discrepancies occurred between the audiologists’ diagnoses and those of the computer in 49 percent of the cases. A user made an error in diagnosis in only one case (2%), while incomplete diagnoses were made in 32 percent of the cases. A diagnosis was considered incomplete if all possible conclusions were not included in the final diagnosis. For example, if the audiologist concluded that a hearing loss was conductive, based on a delayed wave I latency, but failed to state that the loss could also be mixed, then the diagnosis was considered incomplete.

In 15 percent of the cases, SIMON made an incorrect diagnosis. The range of these problems is outlined in Table 1. The problems were the result of incomplete, missing, or illogical rules. Incomplete rules would be those that failed to include all the factors that were needed to establish that conclusion. For example, if a rule concludes that brainstem function is abnormal because wave V is delayed but fails to consider tympanometry results, then it is an incomplete rule. A missing rule usually resulted in the program being unable to come to a conclusion when the audiologist did, whereas an illogical rule resulted in an erroneous conclusion by SIMON.

After analyzing these cases, and re-examining the pertinent rules, we modified the knowledge base to correct such problems. A total of 73 test cases were analyzed including the original 55 cases and an additional 18 cases which were run to ensure that changes made produced correct results in different situations. In a second evaluation of SIMON, the expert agreed with the computer diagnosis for 100 percent of the cases. These results are illustrated in Figure 4.

**DISCUSSION**

Although it has been available for two decades, expert system technology has rarely been utilized in the field of audiology. The reasons for the lack of further development in this area are varied. First, the diagnostic problem is to infer a specific disorder from a set of manifestations (i.e., history, observation, test results). The set of manifestations can usually be associated with several alternatives. The clinician usually conducts a preliminary assessment of the most likely disorders and rules out as many of the alternatives as possible. Unfortunately, this diagnostic process is not a stable one. It is dynamic, varying from instant to instant for the same clinician. Something a parent says or a simple observation of a child's activity may change the direction in which the diagnosis was going. Also, two clinicians do not perform in exactly the same way or necessarily arrive at the same diagnosis under the same set of circumstances. One clinician's experiences may lead him or her to suspect a disorder that another clinician would not even think about. Finally, the clinical process is not invariant. It is different today from what it was a decade, or even a year, ago. Indeed, the very effort to elucidate the process may well change it. We found that during the coding of the knowledge base for SIMON, it was sometimes quite difficult to express in terms of a rule how certain knowledge was used to assist in a diagnosis.

![Figure 4](image-url)
Experienced clinicians often use expert intuition in decision making. We may not have thought about what makes us have confidence in identifying a waveform, for example. At what point do we stop having high confidence in the presence of wave V and start having moderate confidence? Is it based primarily on how replicable the waveform is or on how close the peak is to other normative values? These are the types of questions that must be answered in developing a rule-based system.

Another difficulty in the development of expert systems is the vast number of concepts and facts that are necessary for the formulation of a diagnosis. The representation of knowledge ranges from global concepts, such as those taught to beginning students of audiology, to specific diagnostic details gleaned only through years of clinical experience. The development of the knowledge base, therefore, posed one of the greatest challenges in the development of this expert system. The conversion of the knowledge base to rules, however, was probably less difficult than is usual since the coding of the rules was accomplished by a professional in the domain of interest, which is most often not the case in expert system development. In addition, the assignment of certainty factors to the rules is a tedious and on-going process. The addition of every rule to SIMON resulted in a change to the overall certainty of a diagnosis and, thus, would require the re-evaluation of test cases to assess its impact.

Some medical expert systems of the past have not been readily accepted by physicians (Teach and Shortliffe, 1981). One reason is that the programs frequently offered advice dogmatically as opposed to offering some explanation for conclusions. In order to gain acceptability, the user of a diagnostic expert system must be able to access an explanatory capability that clearly describes the reasoning process employed to reach a conclusion. For this reason, SIMON allows the user to question why the program is asking a particular question, how the program arrived at a conclusion, and what key words and phrases used by the program mean. In addition, we believe that these features will help to make SIMON a valuable educational tool for audiology students in the process of developing clinical expertise.

**CONCLUSIONS AND FUTURE DIRECTIONS**

We believe the SIMON expert system holds promise for assisting in the interpretation of ABR data in young children for several reasons:

1. Infant ABR interpretation requires the use of specific normative values for premature to full-term infants through 18 months of age. SIMON provides a rapid second opinion, equivalent to that of an expert, for the interpretation of this data by automatically utilizing infant normative values.  
2. Unlike human diagnosticians, SIMON never forgets to ask certain questions or to consider their answers. It never has an “off day” that may result in inconsistencies in interpretation.  
3. Unlike expert systems of the past, this program can be run on an IBM compatible micro-computer such as those commonly found in most audiology clinics.  
4. Because of the explanatory capabilities of this system, we believe that students of audiology can benefit from their interactions with SIMON.

We are currently expanding this program to include behavioral data, to incorporate case-based reasoning as used by Bareiss (1989), and to provide interactive intelligent tutoring. A potential weakness in SIMON’s ability to make an accurate diagnosis is linked to the reliability and validity of the data entered by the user. To this end, we are exploring ways to assess the user’s competence at correctly identifying ABR wave forms prior to the start of a run. With that information, we will be better able to assign appropriate certainty factors to the intermediate conclusions and final diagnosis.

The ability to incorporate information from multiple experts in the expansion of this system is a unique, and potentially powerful, contribution to expert system technology. Finally, we plan to explore the potential benefits of SIMON in audiology training programs.

**Acknowledgment.** This paper was presented, in part, at the American Academy of Audiology Annual Meeting in Nashville, Tennessee in April, 1992.

The authors wish to express their gratitude to Drs. Fred H. Bess, Wesley Grantham, and Marleen Ochs for their valuable input throughout this project. In addition, the first author expresses her thanks to Dr. Thomas Klee for originally directing her attention to this area of research.
REFERENCES


