EINTHOVEN and Tolerance for Human Error: Design Issues in a Decision Support System for Cardiac Arrhythmia Interpretation

Lawrence E. Widman, M.D., Ph.D. and David A. Tong, Ph.D.
Cardiovascular Section, University of Oklahoma Health Sciences Center
and Medical Service, Oklahoma City VA Medical Center
Oklahoma City, OK

Decision support systems are becoming increasingly accepted in medical practice in the United States. Clinicians recognize the need for aid in interpretation of complex cardiac rhythms. The EINTHOVEN *system is being developed to meet that need. In this paper, we address the need to deal with errors in the input due to inaccuracies in hand annotations by the inexperienced user and to interact with the user to correct them. Four specific types of input errors are described: missing waves, mispositioned waves, mislabeled waves, and extra waves. General and specific mechanisms by which these errors can be recognized and remedied are described. These results may be interesting as an example of the practical problems that arise in the design of real-world expert systems.*

Keywords: knowledge-based systems, model-based diagnosis, cardiac arrhythmias, human error tolerant algorithms

INTRODUCTION

The classical motivation for development of computer-based electrocardiogram (ECG) interpretation systems was the need to improve automated real-time and near real-time rhythm interpretation, as in intensive care units and Holter monitoring equipment. While these are still important applications, the recent increase in access to Internet communications from rural areas suggests another application area: the rural health care provider who has a patient with a complex arrhythmia and no readily-available expert in electrocardiography. We are extending the EINTHOVEN system to support the user who has a sophisticated medical background but is relatively inexperienced in the area of ECG interpretation [1, 2, 3]. This paper explores design issues concerning errors that such users may make, and implementation of tolerant strategies in EINTHOVEN that recognize and compensate for these classes of errors. Compensation for errors is particularly important because inexperienced users may rely uncritically on decision support systems [4].

*This publication was supported in part by grant numbers R01 LM05530 and R29 LM06004 from the National Library of Medicine, by project number HR4-092 from the Oklahoma Center for the Advancement of Science of Technology, and by a grant from the Whitaker Foundation. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of these granting agencies.

The types of errors expected from inexperienced users differ from those expected from automatic rhythm interpretation systems. In both cases, errors in identification of waves can occur. In the case of automated systems, the errors will be due to artifact in the signal, while the noise-free signal can be analyzed accurately by very sophisticated algorithms. In the case of the wave annotation by hand by humans, errors will be due to failure to recognize unusual, yet clear, appearances of waves, while artifact will usually be recognized and ignored. In this report, we focus on the types of errors that are commonly made by clinicians with limited expertise in electrocardiography.

RELATED WORK

The general topic of human error in medicine has been an active area of research in cognitive science, human factors engineering, and systems analysis, and has been reviewed recently in medical publications [5, 6]. To our knowledge, the specific problem of human errors in interpreting electrocardiograms has not been analyzed previously.

METHODS

Overview of the EINTHOVEN System

This section presents the cardiac conduction model, the input to the system, the reasoning mechanisms, and output produced by the EINTHOVEN system. EINTHOVEN has been validated using simple rhythms from introductory ECG textbooks [7] and more complex rhythm strips from clinical records and intermediate-level ECG textbooks [8].

Cardiac Conduction Model In EINTHOVEN the cardiac conduction system is modeled as a graph [9], a mathematical construct defined as a collection of objects with connections between them. The objects are called "nodes" and in our model represent the anatomic or functional elements of the cardiac conduction system. Nodes also have properties such as refractory periods, transit times from entrance to exit (that are related to the conduction velocities within the element), and observability on the electrocardiogram or on intracardiac electrograms. The connections between the objects represent the connections between the elements of the cardiac conduction system.

The most complex and comprehensive model, or "base model", implemented in EINTHOVEN can

0195—4210/96/$5.00 © 1996 AMIA, Inc.

represent anatomic abnormalities consisting of up to four accessory atrioventricular (AV) pathways, three slow AV nodal pathways, and one atriofascicular ("Mahaim") pathway, as well as up to seven ectopic foci in each of the atria, atrioventricular node, and ventricle. These values were chosen as practical limits based on the number of abnormal pathways observed in patients. There are no theoretical limits on the number of these pathways in the model.

System Input The input to the system is a text sequence describing the observed events (P waves, QRS complexes, and T waves) that comprise a given cardiac rhythm. Each element in the sequence describes the morphology, onset time, and width of an event. These elements are obtained from annotations by hand of the ECG by the user. Thus, errors in annotation create errors in the input.

The Reasoning Mechanism Based on these data, models of cardiac electrophysiology are generated by selectively activating and inactivating individual nodes in the base model. The models that are created are specified through heuristic pattern-matching of the observed events using deterministic production rules [10]. Reasoning with the models and observed events is implemented as a variation of the hypothesize-and-test paradigm [11, 12]. Hypothesized model behaviors are generated using discrete-event simulation. These behaviors are the events, either a P wave or QRS complex, that may be observed on the ECG should that model be valid for the given data.

Expected model behaviors for P waves and QRS complexes are generated by traversing all "paths" of electrical activation present in the model. Paths begin with the node that produced the observed event and progress through the model until a node is reached whose activation may be observed on the ECG. The output of the discrete-event simulation algorithm is the set of all possible behaviors (observed events) for a given model. These behaviors are described by: (1) the chamber in which the predicted event will occur (either Atria or Ventricles), (2) characteristics of the shape of the predicted wave on the surface ECG (a shape class and whether the class is normal or abnormal), and (3) a time window specifying the earliest and the latest times at which the event may occur. These events are generated from the model by traversing it starting with chamber that produced the current observed event, passing in turn through all possible paths, and ending with the postulated next event for each path.

The expected model behaviors are compared to the next observed event. Models with at least one expected behavior matching the observed event are updated to better reflect the observed event using a set of "Transition" rules. Examples of model parameters that are updated in this manner include the spontaneous cycle length and transit times. Models that fail to have at least one behavior that matches the observed data, are marked as invalid and new models are created using a set of "Abnormal Finding" rules. The updated valid models or the newly created models are simulated again, continuing the cycle. Simulating and testing in this event-by-event manner prevents combinatoric explosion of possible expected behaviors.

System Output The output of EINTHOVEN is a set of one or more causal diagrams, called "ladder diagrams," that together constitute the complete differential diagnosis of the rhythm. This format is a standard format used by clinicians [13] and completely describes the cause-and-effect relationships between individual events within a rhythm, as well as transitions between distinct rhythms.

When more than one rhythm diagnosis is possible for a given data set, the complete differential diagnosis is produced with a ladder diagram for each possible rhythm. The rhythms are ranked roughly by likelihood. Ambiguous rhythms have on the order of five possible mechanisms.

Analysis of Errors Made by Inexperienced Users

We obtained data describing these errors from two sources. The first source was a preliminary study [14] of the ability of physicians and nurses to interpret rhythms when the presentation was in the form of a copy of an electrocardiogram or a reconstructed tracing that incorporated only the timing and shape of the P waves and QRS complexes. The overall error rate for all readers was surprisingly high. Review of the data showed (1) failure to recognize unambiguous patterns and (2) inattention to detail, as evidenced by success in recognizing a pattern on one specimen and failure to recognize the same pattern on another. The second source was the experience of one of us (L.E.W.) in teaching medical students, residents in Internal Medicine, fellows in Cardiovascular Diseases and Clinical Cardiac Electrophysiology, and consulting colleagues. Our analysis is consistent with the experience of other teachers of clinical ECG interpretation, but has not been published previously to our knowledge.

The types of errors in feature extraction during interpretation of the body surface electrocardiogram to which inexperienced health care providers are prone include the following misclassifications:

1. *Overlapping events.* Failure to recognize P waves that fall on T waves during tachycardias and during second-degree AV block, flutter waves that overlap the QRS complex, and atrial premature complexes as causes of pauses;
2. *Timing errors.* Incorrectly positioning the bounding box for a given wave, thus making the onset of the wave too early or too late, and the width too narrow or too wide.
3. *Misidentification.* Overinterpretation or underinterpretation of peaked T waves as P waves;

4. *Misclassification of P wave shape.* Mistaking the shape of P waves that fall on T waves because the time of onset of the P wave is not known or because the P shape is distorted by the shape of the T wave; and
5. *Misclassification of QRS complex shape.* Misclassifying narrow QRS complexes as normal when they are wide, and visa versa; and wide QRS complexes as ventricular when they are more likely supraventricular with aberrant conduction or with ventricular preexcitation, or visa versa.

Other aspects of the rhythm, such as intervals between waves and regional patterns such as group beating, are derived automatically by EINTHOVEN and so can be misinterpreted indirectly due to the above errors.

RESULTS

Errors in input may be divided into four general categories: missing waves, extra waves, mispositioned waves, and mislabeled waves. Strategies for compensation by EINTHOVEN are tailored to each category.

Missing Waves

In general, missing waves can be dealt with by using longer paths to generate expectations: instead of generating an expectation for only the the first observable event, additional expectations can be generated for the second and subsequent observable events. This is already done in the presence of ventricular premature complexes, and is an integral part of the "hot nodes" algorithm [9]. The conclusion that a wave is in fact missing is reached when subsequent events match the expectations (generated from the longer path) that correspond to that conclusion.

Additionally, missing P waves are inferred by a least-common-denominator algorithm that reevaluates intervals between P waves, QRS complexes, and P-QRS pairs in a "look-back" step (see below) after each observed P wave and QRS complex.

Extra Waves

This is an uncommmon problem because the inexperienced user tends to miss cardiac events rather than to overinterpret them. By contrast, automatic systems err in the direction of overinterpretation. We have considered three different approaches that can be used for extra waves. The simplest approach is to try to disqualify the suspected beat based on the model by showing, for example, that it falls on the absolute refractory period of its alleged chamber of origin. Secondly, the current models can be suspended, rather than discarded, until further events show the suspected false wave to be genuine or artifactual. This requires a "look-back" procedure after all hypotheses have been suspended that compares the set of hypotheses that existed before with the set that exists after the confusing event. Further, it requires the assumption (by Occam's razor) that the rhythm did not change during the confusing event. Finally, the proposed system can request the user to re-examine the original signal for other evidence of artifact. So far, we have avoided implementing any of these assumptions. Instead, we have preferred to have the system fail to reach a conclusion and then to query the user.

Mispositioned Waves

Mispositioned waves that fall outside a given expectation window constitute two errors: a missing wave, since the expectation was falsely not met, and an extra wave, since the observed event appears elsewhere in the input data stream. We handle this error as a combination of the other two errors.

Mislabeled Waves

The problem of mislabeled waves is more difficult. Four approaches can be used for mislabeled waves. Filters on the input data stream, which look for obvious incompatibilities, such as overlapping waves, can also look for physiologically impossible combinations such as a T wave without a preceding QRS complex, two sequential QRS complexes without an intervening T wave (assuming visible T waves, and usually resulting from mislabeling of the true T wave as a second QRS), impossibly wide P waves (usually resulting from mislabeling of a wide QRS or a T wave as a P wave), and so forth. The second and third are similar to those proposed for dealing with extra waves: try to work around the problem after flagging it as a source of potential error or allow the analysis to fail, then to ask the user to reconsider the wave annotations in the region in which it failed.

The fourth approach is to relax the criteria for certain Abnormal Finding rules and then to rely on strict criteria for the Transition rules that maintain their diagnoses. This approach can be used only with the special diagnoses that must be invoked almost indiscriminantly (see below). For example, the conventional diagnosis of multifocal atrial tachycardia requires the finding of three abnormally shaped P waves. One of the Abnormal Finding rules for this diagnosis requires only two abnormally shaped P waves and one normal P wave. It carries the "tentative" confidence label, so that the corresponding diagnosis is dropped if subsequent events refute it.

Errors that Terminate the Analysis

The program architecture can be modified to allow it to work around otherwise fatal errors. Ordinarily, the top-level algorithm stops processing the input specimen when an observed event fails to match any of the expectations of any of the models it considers valid based on the prior observed event. For certain rhythms, processing is allowed to continue even though one of the valid models fails to predict a subsequent observed event. These rhythms are those that must be invoked almost indiscriminately, such as atrioventricular dissociation, mul-

tifocal atrial tachycardia, and persistent junctional reentrant tachycardia. The models for these diagnoses are simply discarded if their expectations are not fulfilled at any point in the analysis.

We have considered adding rules intended to specifically recognize and, if possible, to characterize possible erroneous inputs, such as alleged QRS complexes not followed by T waves, in a record in which T waves are readily evident; and alleged premature P waves that do not change the timing of the subsequent normal P waves. Similarly, tape speed error masquerading as change in heart rate (interval from one QRS to the next QRS) in Holter monitor recordings can be detected by a parallel change in the PR interval, QRS width, and QT interval of each cardiac cycle in the abnormal region. At this time, however, we have elected to allow the analysis to fail and to inform the user of the input events that caused the failure. The alternative is to allow the system to modify the input on the basis of incorrect assumptions that the inexperienced user may fail to correct.

DISCUSSION

The performance of decision support systems is critically dependant on the combination of the accuracy of the input data and the ability of the systems to recognize and react appropriately to errors in the input. This is a general problem in the design of expert systems for real clinical applications. In this paper, we examined one such system in detail.

We have described the types of errors that can be expected from medically sophisticated users who are relatively inexperienced in the interpretation of cardiac rhythms on the ECG. We believe that our description is a valid generalization and is not idiosyncratic to one particular cardiologist. We are not aware of prior published work on this topic, however. We have also described compensatory mechanisms we have considered, some of which we have implemented.

Below we describe additional compensatory mechanisms that are not implemented because they have substantial disadvantages at this stage of the development of EINTHOVEN.

1. Potentially fatal errors could be noted and then bypassed by re-invoking the Abnormal Finding rules. This would bypass any error because one Abnormal Finding rule, which invokes atrioventricular dissociation, is always satisfied. The problem with this modification is to distinguish inadequacies in the knowledge bases and/or in the reasoning mechanisms, from artifact in the input. This modification would allow the possibility of ignoring a dangerous but correct interpretation in favor of an artifactual but incorrect interpretation. This modification might be safe if EINTHOVEN were developed to greater clinical reliability.
2. Meta-Abnormal Finding rules to guess what errors might have been made.

   The current paradigm for EINTHOVEN is a single level, event-by-event analysis in which future events are not considered. We could implement a second level of analysis that would be invoked after the first level fails to interpret its findings and must resort to one of its fall-back strategies. This second level would be invoked after one or more subsequent events have been analyzed and found consistent. It would take a bird's eye view of all of the events in a defined window of interest to mimic the global view of the physician who scans the entire rhythm strip prior to beginning interpretation. Note that this level would not violate the real-time model underlying EINTHOVEN because it would be retrospective and would not access *future* events.

   This type of "look-back" analysis is already implemented in EINTHOVEN in two areas: (1) to remove diagnoses that must be raised and can be excluded only in retrospect, such as second-degree AV nodal block in a case in which the PR interval prolongs progressively but no P wave fails to conduct to the ventricles, and (2) to detect intermittent failure to find all instances of regularly occurring P waves, which may be somewhat more visible in some parts of the record than in others.

   Features of the events in the time window of interest would be determined and made available as predicates in meta-Abnormal Finding rules. These predicates would be more powerful than those in the first-level, ordinary Abnormal Finding rules because they would include features of the events that follow, as well as of those that precede, the confusing events. This meta-level analysis would then be able to:

   (a) to encode regional patterns, such as Marriott's famous rules that give the differential diagnosis of group beating, of pauses, and so forth [13]. Recognition of these patterns would allow them to conclude that confusing events may be truly physiologic.
   (b) to determine whether the rhythm "marches through" the confusing events as if they were not there. This finding is most consistent with the interpretation that the confusing events are artifactual.
   (c) to measure variations in signal features that should not vary, such as the durations of events that the user classifies as having the same shape class. If the user measures P wave and QRS complex durations sufficiently inaccurately, the expectations will need to be made more flexible.
   (d) to detect intermittent failure to find all instances of regularly occurring non-P wave events.

   Because the meta-Abnormal Finding rules would be able to reinterpret events that have already been interpreted, they would be able to split individual hypotheses into several hypotheses, each representing one possible interpretation of a confusing event, and then to re-run the first-level AF rules on each of the new hypotheses starting with

the reinterpreted event and continuing until the end of the analysis window. We have elected to defer implementation of this level of sophistication until EINTHOVEN is further developed in its current form.

3. Interaction with the user.

   In some cases of erroneous input, it may be necessary to suggest that the user reexamine the input annotations. For example, a narrow complex tachycardia at a rate of 150 beats per minute has a good chance of being atrial flutter even if the user identified only one "P" wave per QRS complex. In this case, as when teaching students and residents, it is appropriate to suggest that the tracing be reexamined for a second atrial wave, perhaps adjoining the QRS complex. Another example would be an interpolated narrow-complex QRS. While an interpolated, premature wide-complex QRS with morphology indicating ventricular origin is quite reasonable and is consistent with a subsequent on-time narrow QRS complex, it is unlikely that the supraventricular location of an ectopic focus that could cause a narrow QRS complex would also fail to affect the next atrial event.

CONCLUSIONS

Decision support systems for clinicians must recognize and deal adequately with errors made by users who are, by definition, not domain experts. We have explored this issue in the context of EINTHOVEN, a system under development for automatic interpretation of the electrocardiographic rhythm. The design goal for EINTHOVEN relevant to this paper is to improve the robustness of its performance when errors in its input are due to human error in annotation by hand of cardiac waves on the electrocardiogram. A secondary design goal is to develop compensatory mechanisms that would also be useful for dealing with errors made by automated rhythm analysis systems.

We have described the classes of errors that are made by medically sophisticated health care professionals who are inexperienced in the interpretation of the cardiac rhythm in the electrocardiogram. We have described how some of these are recognized and compensated in the EINTHOVEN decision support system. The pitfalls of attempting to compensate for others of these are also described. These results may be interesting as an example of the practical problems that arise in the design of real-world expert systems

References

1. D.A. Tong, V. Gajjala, and L.E. Widman. Results of developing a client/server system for remote diagnosis of cardiac arrhythmias. In *Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care*, pages 601–605. Hanley & Belfus, Philadelphia, 1995.
2. D.A. Tong, V. Gajjala, and L.E. Widman. Demonstration of a client/server system for remote diagnosis of cardiac arrhythmias. In *Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care*, page 998. Hanley & Belfus, Philadelphia, 1995.
3. L.E. Widman and D.A. Tong. EINTHOVEN on the World Wide Web: A tool for analysis of cardiac arrhythmias. In *Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care*, page 968. Hanley & Belfus, Philadelphia, 1995.
4. E.S. Berner, J.I. Kennedy, G. Blackwell, and J.B. Box. Use of computer-generated ECG reports by residents and faculty. In *Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care*, page 953. Hanley & Belfus, Philadelphia, 1995.
5. L.L. Leape. Error in medicine. *Journal of the American Medical Association*, 272(23):1851–1857, 1994.
6. Ramon Felciano. Human error: Designing for error in medical information systems. *http://www-camis.stanford.edu/people/felciano/-hci/humanerror/humanerrortalk.html*, 1996.
7. D.A. Tong, K. Beckman, and L.E. Widman. Model-based rhythm analysis of the ECG: Evaluation of a prototype implementation. *Journal of Electrocardiology*, 26(suppl):182–193, 1993.
8. L.E. Widman and D.A. Tong. Computer-based interpretation of the ECG rhythm: Validation of the EINTHOVEN system with three classes of clinical dysrhythmias. *Journal of the American College of Cardiology*, 27(2 Suppl A):66A, 1996.
9. L.E. Widman and D.A. Tong. An algorithm for complete enumeration of the mechanisms of supraventricular tachycardias that use multiple Atrioventricular, AV Nodal, and/or Mahaim pathways. *Artificial Intelligence in Medicine*, 7(4):361–386, 1995.
10. B.G. Buchanan and E.H. Shortliffe. *Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison Wesley, Reading, MA, 1984.
11. L.E. Widman. Representation method for dynamic causal knowledge using semi-quantitative simulation. In *Proceedings of the Fifth World Conference on Medical Informatics MEDINFO86*, pages 180–184. North Holland, 1986.
12. L.E. Widman. A model-based approach to the diagnosis of the cardiac arrhythmias. *Artificial Intelligence in Medicine*, 4:1–19, 1992.
13. G.S. Wagner. *Marriott's Practical Electrocardiography.* Williams & Wilkins, Baltimore, ninth edition, 1994.
14. D.A. Tong. *Model-Based Rhythm Analysis of the Electrocardiogram.* PhD thesis, Louisiana Tech University, 1992.