GUEST EDITORIAL

In Search of a Perfect Partnership

JANICE M. JENKINS

From the Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, Michigan

Introduction

The journal *PACE* is perhaps singular in that it is a scientific journal that publicly proclaims the marriage of engineering and medicine. The audience consists of both engineers and physicians and manuscripts contain joint authors from both fields. It is in the interest of both groups that we communicate with a minimum of misunderstanding and adopt a mutual appreciation of each other's terminology and methodology. It is for this reason that I wish to call your attention, gently I hope, to a variety of gaffes, omissions, and outright errors in engineering methodology that frequently creep into the medical literature.

My field is computer engineering with a specialty in digital signal processing. The focus of this signal processing happens to be electrocardiographic signals (a happy accident), which has engendered a close relationship with many colleagues in medicine. I work in close collaboration with cardiologists in order to ensure medical oversight and validation of the *computerization* of this all important signal.

During the past decade, with the availability and low cost of the personal computer, you my medical colleagues have discovered the wonder of my favorite tool. You've used it for word processing (a modern miracle, we'll all agree), you've plugged your data into statistical packages (with a trust I need to chide you about), and now more recently, you've even dared to enter my world of signal processing of the electrocardiogram itself (with an intrepidity that amazes engineers who wouldn't dream of attempting it without serious formal training).

I'm genuinely delighted with your newfound discovery of my computer world. It's certainly an improvement over your earlier reluctance to trust any computer to invade your medical purview. I'm reminded of my efforts to introduce a computerized ECG system to the section of cardiology at a major medical school in 1979. The resistance was incredible. It ranged from "No computer can ever interpret an ECG as well as a human," to "I refuse to deal with this new contraption which is just a fad." How times have changed. We've seen a complete about-face in the acceptance of computer-read ECGs such that the medical profession now mostly takes them for granted. (Did you even bother to get out your calipers to verify the computer measurements the last time you overread an ECG?)

Your acceptance of the computer as a significant part of modern medical science has, in fact, changed the relationship of physicians and engineers. (We used to be looked at as technicians standing on the sidelines as minor players; now you avidly recruit us as collaborators in your clinical and research tasks.) Your adoption of our electronic toy is a pleasure, but the sometimes incautious use of it raises concerns that we need to voice. In some cases, physicians are using computers in the practice of medicine or in medical research with little or no real understanding of computer engineering. The medical literature is replete with carelessness and errors (in an engineering sense) that should never have made it past the reviewers. I list below four of my favorite pet peeves (i.e., violations of engineering methodology) in the hope of raising your consciousness and perhaps easing our future communication.

Address for reprints: Janice M. Jenkins, Ph.D., Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122. Fax: (313) 763-1503.

Received August 12, 1992; accepted August 12, 1992.

Misuse of Computer Terminology

A particular vexation of mine is the misuse by noncomputer specialists of the term *real-time*. Computer engineering has a scientific vocabulary as precise as that of medicine. The term real-time has a specific meaning. Its use by the medical community has distorted it to such an extent that a return to purity may be impossible. But let me try.

Real-time refers to computer execution that is simultaneous with the process under analysis. It's typically used to describe a sampling system that monitors a continuous external process and effects an immediate response or modification of the process under observation. If the modification part is present, the system is a control system with feedback. The clue to real-time execution is the capability of uninterrupted monitoring with *simultaneous* computer processing. This requires that the sampling of the external process continues unabated and analysis of the sampled signal *is completed within the intersample period.* Real-time monitoring was the basis for our initial space launches. The trajectory of the space vehicle could be determined on a millisecond-by-millisecond basis and immediately corrected. Without instant correction (feedback) the error would have grown and a delayed response would have lagged hopelessly beyond correction. Think of coronary care monitoring as a true real-time system. It continuously follows all patients in the unit and gives an instantaneous report when an alarm condition occurs. If instead, you have a program that captures data for a limited amount of time and subsequently processes it, that is NOT real-time analysis, no matter how fast your computer completes the task.

A manuscript published a few years ago described a real-time program for analysis of electrophysiology signals. I believe that the term real-time even appeared in the title. The procedure went as follows: signal channels were sampled and stored in a file; data were subsequently retrieved and displayed on the monitor; cursors were applied to identify waveforms; 5 to 8 seconds elapsed before analysis was complete. Any one of these four conditions alone would violate the definition of real-time. The presence of all four in the face of the claim of real-time execution indicates a serious misunderstanding of the meaning of the term.

A similar mistake is made by the frequent use of *on-line* as a synonym for real-time. The true definition of on-line is "in direct communication with the computer" and usually refers to printers and terminals that are switched to on-line status. Real-time is of its very nature on-line, so if one describes an on-line analysis system, the word on-line in this context has the connotation of real-time. Unfortunately, that's often misused as well, for the same reasons as above. Those of us who develop true real-time programs (a nontrivial task) find ourselves in a box when trying to compare our software with those who claim real-time improperly. Remember, it's always easier to process data after the fact (off-line) because the algorithm can search forward as well as backwards, await subsequent information before committing to a diagnosis, or perform multiple passes or iterations through the data. None of these techniques is possible in real-time analysis.

Blind Trust in Computer Results

In the training of programmers/engineers in the art of diagnostic software, the most important consideration is the validation technique. As an example, students in my laboratory are frankly amazed when (upon their announcement that their program "works") I tediously hand measure every waveform and interval with calipers from strip chart recordings, and insist that the program must duplicate my measurements. We invariably find errors in this initial stage, and the students learn an important rule. *I don't trust anyone's computer program until I have exercised it under a variety of conditions and found all possible errors.* How I wish I could instill the same skepticism in my medical colleagues.

More and more often I have seen my medical associates purchase commercial software and naively believe the results without proper validation, or hire a programmer and believe the computer results to be gospel. When a student of mine writes a Fast Fourier Transform for spectral analysis (or uses a canned program), I inquire how he or she knows if it works. (I usually get a blank stare.) I suggest a very simple but necessary test. Use a function generator and apply a signal for which you know the result, for instance, a sine wave of a particular frequency. A look of wonder replaces the blank stare: yet they hadn't even thought of that.

I urge you, my medical colleagues, to observe this same caveat: don't trust anyone's software, or your own for that matter, until it has been properly validated. Test it on simulated data for which you *know the correct outcome* before you trust it on a real problem. This should be done for statistical packages as well. For instance, if you wish to calculate a correlation coefficient, put in two sets of monotonically increasing numbers. That should yield a correlation of 1. Two sets of numbers in which one is monotonically increasing and the other decreasing should give a −1. Two sets of random numbers should produce a value between −1 and +1. That test is so easy, you can check it on your fancy calculator as well. But, surely, test your programmer's version.

Pattern Recognition: Its Use and Misuse

Many of our present research papers fall into the category of pattern recognition or pattern classification. Any time you're separating samples into two (or more) classes by some measurement(s), such as distinguishing sinus rhythm from ventricular tachycardia, you are doing pattern recognition. This field, like software engineering, is a well-defined discipline with rules and conventions. If you intend to engage in this discipline, take the time to learn the rules of the road.

The rules are simple and straightforward. When developing a classification system, one starts with a *training set* of data. The software is fine-tuned through numerous iterations to produce optimal classification on that set. Once the algorithm has been finalized, a true validation of its performance must be done on a separate and independent *test set*. And, if you intend to publish the performance measures of your algorithm, it's not kosher to make any more changes or further tweak the software during the testing process. An engineering journal would insist that authors observe this rule. Yet in my review of papers for medical journals, I find little or no observance of this most basic requirement. Almost invariably, manuscripts that come from medical investigators present results of a developmental computer system on *the data from which it has been developed*. Of course the software will produce exemplary results on the training set; the true test is whether your software will work on other data. Only then can you present your performance ratings with any confidence.

Is it legitimate to report work-in-progress on a promising algorithm that is still undergoing modification? Sure. Just make certain that you properly describe which stage you are in. For honesty's sake, tell your audience that you're still tweaking your program and that results are given for a training set only. (The results are heavily biased, you know.) Pattern recognition is a solid, statistically-based discipline, and only test set results are valid for extrapolation to the population as a whole. (This presumes that you have observed other cogent statistical criteria as well [sample size, independence, appropriate choice of statistical test], and that too is often amiss. But I shall address subject in a later editorial, if so permitted.)

Signal Characteristics and Their Importance

This subject (so near and dear to my heart) deals with investigations in which signal processing plays a role. Digital signal processing, like pattern recognition, is a major discipline within electrical engineering. Any judgment about the quality of the process and the validity of the results relies heavily on knowledge of the hardware used to acquire the signal (amplifiers, tape recorders, analog-to-digital converters), the manner of signal conditioning (filter and gain settings), the type of transducer (catheter electrode, patch electrode, unipolar vs bipolar configuration, electrode spacing and surface area, electrode location, etc.), the computer components used, the type of software (programming language, home-grown or commercially purchased), and the details of program execution (real-time or on-line vs off-line, interactive vs completely automatic, etc.). Without providing these "specs," there is no way that individual investigators can make a sensible comparison of performance of one technique versus another. An anecdote can perhaps best make my point.

We wished to investigate the robustness of some of our morphological algorithms in the presence of increased heart rate. Rapid atrial overdrive pacing, infusion of epinephrine, and infusion of isoproterenol were the three interventions used. In attempting to compare our peak-to-peak amplitude measurements with findings of two other published studies (in which results were in disagreement), we discovered that a straightforward comparison was impossible. Our signals were recorded with wideband filter settings (0.5 or 1 to 500 Hz) and the others were recorded with passbands of 30 to 100 Hz. The characteristics of these two passbands are distinctly different. The high pass filter (low frequency cutoff of 30 Hz) is essentially a differentiator and will dramatically distort the raw signal. It will take a monophasic waveform and transform it into a biphasic one, a biphasic into a multiphasic. It will exaggerate the high frequencies and reduce the baseline to a rock-solid isoelectric line. The low pass portion (100 Hz high frequency cutoff) will attenuate the amplitude of the waveform and again distort the true measurement. If you're skeptical, run an ECG signal through a wideband setting (0.04 to 500 Hz is selectable on some systems, 0.5 to 500 Hz on others) and then observe the effect of changing the low frequency cutoff to 30 Hz. Repeat the experiment for just the 100 Hz high frequency cutoff. You could use a high frequency cutoff of 2,500 Hz for your original choice but it's immaterial because you won't see any perceptible amplitude difference between the 500 and 2,500 Hz setting. There's no appreciable signal content above 500 Hz so there's nothing to attenuate. (Can that statement be trusted? If I were you, I'd do that experiment too, just to be safe, and to check my veracity.)

I realize that 30 to 100 Hz filtering is commonplace in almost all electrophysiology laboratories but I urge you to remember why. It's because the electrophysiologist prefers the *display* characteristics of that signal. It eliminates extraneous noise and baseline wander, and keeps the signal on scale. But if you wish to take careful measurements of a signal and report it as scientific data, it behooves you to use the raw signal itself, not some adulterated version of it. Only then can we "signal processors" communicate in the same language.

These are other mechanisms that impose a sometimes unintended filtering effect. For instance, in frequency modulated (FM) magnetic tape recording, the tape speed determines the bandwidth of the signal. FM recorders that comply with interrange instrumentation group (IRIG) standards have a frequency response *in the normal bandwidth mode* of DC to 312 Hz at a tape speed of 15/16 inches per second (ips). As tape speed doubles so does frequency response, such that 1 ⅞ ips gives recording fidelity in the bandwidth DC to 625 Hz. A judicious choice of tape speed is imperative if you wish to ensure faithful reproduction of the original signal. Strip chart recorders and other mechanical devices also have upper limits on frequencies that they can reproduce. If measurements are being reported from signals recorded on such devices, a knowledge of the frequency response is critical to proper interpretation of the data. In order to make comparisons between two studies, we must speak the same language, we must use device settings that provide for fidelity in signal acquisition, and we must be complete in our description of all pieces of hardware and their characteristics; otherwise the comparisons are meaningless.

Summary

There's no question that our partnership is a natural, given the technological emphasis in modern medicine. My comments are not intended as criticism, but offered in the spirit of broadening our mutual understanding. I believe that open communication will enhance our continued collaboration in this interdisciplinary effort and yield significant advances in medical science.