Automated Audiometry: Progress or Pariah?

Robert H. Margolis
Professor
Department of Otolaryngology
University of Minnesota

Introduction:

The chapter titled "Automatic Audiometry" in Jerger's 1963 *Modern Developments in Audiology* began with….

> *The number of audiometric examinations made today has grown to such a magnitude that it is only natural that some of the techniques of measurement should become automated. (p. 30)*

and continues in the next paragraph with:

> *… it appears only natural that those features of audiometry which can be automated will be, and the audiologist will find himself [sic] fully occupied with the task of analyzing and interpreting the data. The routine work can be done by the machine. (p. 31)*

Automation has been incorporated in many audiologic tests including tympanometry, auditory brainstem response, otoacoustic emissions, and various electroacoustic tests. However, clinical methods for pure-tone and speech audiometry have remained largely unchanged over several decades.

My first audiology professor, Dr. Kenneth Berger, told us in 1966 that every audiology test in use at that time would be replaced within ten years. Ken was partially right. Tone decay, SISI, ABLB, and Bekesy audiometry have indeed given way to ABR, OAE, and MRI, although it took a little more than ten years!

But the pure-tone audiogram and speech-recognition tests remain the core of the basic hearing evaluation. Most other audiologic tests are done only when we can't do those basic tests, or when supplemental tests are desired to derive a differential diagnosis.

The fact that our most important tests have not changed since the advent of the electronic audiometer in the 1920's, can be viewed as a strength and a weakness. Pure-tone and speech audiometry remain the core of audiologic evaluation, due to the power and significance of their results as they relate to diagnosis and treatment. As one of our colleagues pointed out, few tests in medicine have the diagnostic power of the audiogram (David Zapala, unpublished communication). The endurance of these tests as core components of the audiologic evaluation speaks to their importance in the evaluation of communicative function and medical diagnosis.

However, as often happens in medicine and health related professions, technology has progressed faster than clinical science. As audiometers evolved from vacuum tubes to transisters to integrated circuits to microprocessors, "functionality" of the audiometer has remained basically the same. In essence, the microprocessors that control audiometers in 2005 are used to provide the same stimuli provided by the earliest electronic audiometers – almost 100 years ago! Although there have been many important technological advances in modern audiometers, the computer power within our current instrumentsis capable of doing much more.

Although there have been tremendous improvements in diagnostic and rehabilitative audiology, the lack of progress in our most fundamental tests should be a concern to all of us. If the lack of progress was accompanied by solid research which demonstrated that our methods could not be improved upon, we would have a solid basis for doing what we do, and confidence that we have the best methods.

Unfortunately, the evidence supporting our methodology does not justify the lack of progress.

A Look Back:

Although non-electric devices for hearing testing have been in existence for centuries, we can trace current audiometric instruments and methods to the electric devices of the late 19th century, when audiometers were in development in England, Germany, and the U.S.A. The first commercial device was likely developed by David Edward Hughes (1830–1900) and was manufactured by the T. Hawksley company of London. The device employed the electrically-driven tuning fork, which was developed by Hermann von Helmholtz (1821-1894) in Germany. It was von Helmholtz who formulated the theory of auditory frequency channels that was, and remains, the basis for testing with frequency-specific stimuli. Three years after Alexander Graham Bell (1847-1922) invented the telephone in 1876, he drew plans for an audiometer, which he presumably built early in 1880.

Modern audiometers can be considered direct descendants of the first vacuum tube devices of the early 20th century. The vacuum tube, invented by Lee de Forest (1873-1961) in 1906 (who later won an Academy Award for bringing sound to motion pictures), gave rise to the oscillator circuit that was used to produce sinusoidal electrical waves. The first vacuum tube audiometer was probably developed by Schwarz in Germany in 1919. In the U.S.A., Western Electric (the manufacturing affiliate of the Bell Telephone Laboratories) developed the first commercial vacuum tube audiometers in the 1920s – the 1A and 2A audiometers. It was the introduction of these audiometers that gave rise to clinical hearing testing, primarily in otology offices.

The first published report on clinical pure-tone audiometry is probably the 1924 article by Isaac H. Jones (an otologist) and Vern O. Knudsen (a physicist). They discussed the need for standard hearing tests with frequency-specific stimuli which could be used to compare results obtained in different offices. Before the advent of the audiometer, they lamented, there were "almost as many methods of recording functional tests of hearing as there are otologists" (p. 674).

Jones and Knudsen recommended a set of test frequencies for air and bone conduction testing at which thresholds are measured by reducing "the loudness of the tone … to the point of minimal audibility." Then "the reading of the dial is noted, and compared with the normal given on a calibration sheet." Knudsen (professor, dean, and vice chancellor at UCLA after whom the physics building is named) published a brief text in 1937 describing behavioral audiometry a bit more thoroughly, but without the detail typically associated with experimental methods. Interestingly, Knudsen was concerned about the lack of formal training of professionals who conducted hearing tests, and not anticipating that an entirely new hearing profession would emerge with high professional standards, he wrote…

> *If the medical profession does not assume this responsibility, and assume it without delay, there will come to be a new type of practitioner, "the audiometrist," who will operate outside the control of medical standards. (p. 3)*

As audiometers came into greater use, clinicians developed their own threshold measurement procedures. A significant step toward standardization occurred when Walter Hughson, an otologist, and Harold D. Westlake, a speech pathologist, published a "Manual for Program Outline for Rehabilitation of Aural Casualties Both Military and Civilian." Their classic article, published in the Transactions of the American Academy of Ophthalmology and Otolaryngology five months before D-Day, was only slightly more detailed than the Jones and Knudsen report. Test frequencies, attenuation step sizes, bone conduction testing, and masking were discussed in general terms. In view of the lack of detail in the article, it seems a stretch to refer to modern audiometric testing as the "Hughson-Westlake method".

At the same time when audiometers and clinical test methods were emerging, a new branch of experimental psychology, "psychophysics," drew attention as a powerful approach for understanding the brain mechanisms which underlie behavioral responses to physical stimuli. Early psychophysical methods such as the Method of Limits, the Method of Constant Stimuli, and the Method of Adjustment were used to quanitify relations between behavior and stimulus dimensions. Though controversial at first, these methods became essential research tools decades before electrophysiology and imaging techniques could probe the nervous system. An important advance in auditory psychophysics occurred when the Theory of Signal Detectability (TSD) originally formulated to optimize the electronic detection of radar signals for military applications, was applied to human detection of auditory stimuli. TSD provides a framework for describing and understanding detection of sound in a wide range of conditions and has been used to study almost every aspect of hearing. Had our audiometric methods been based on the scientific underpinnings of auditory psychophysics (i.e., "psychoacoustics"), we would do things quite differently today.

As computers entered psychoacoustics laboratories, it was quickly recognized that rules governing psychophysical methods used for auditory threshold measurement were easily stated in algorithm form to provide automated and sensitive tests of auditory sensitivity. These methods have been in use for more than four decades. They have been subjected to scientific scrutiny and peer review and are widely accepted as accurate and repeatable methods for measuring auditory thresholds.

With rare exceptions, the psychoacoustics literature addressing auditory sensitivity and the audiology literature dealing with pure-tone audiometry have not crossed paths!

There is much we can learn from the scientific literature on auditory sensitivity. When I began my doctoral work at the University of Iowa in 1970, I was assigned to the laboratory of Dr. Arnold Small. Dr. Small was the first psychoacoustician to serve on the faculty of an academic audiology and speech pathology department. We measured auditory thresholds to 1-dB accuracy using an automated, adaptive, threshold measurement method. Similar methods were used in laboratories throughout the U.S.A. and Europe to study hearing. Nevertheless, clinical audiometry continued along the path set by Hughson and Westlake, without significant influence from psychoacousticians.

One exception to the separate paths of psychoacoustics and clinical audiology was the landmark text written by Dr. Ira Hirsh more than 40 years ago, *The Measurement of Hearing*. Hirsh's descriptions of research and clinical methods of threshold estimation, were the most thorough of the day. His description of clinical pure-tone audiometry is similar to that of Hughson and Westlake.

Probably the most scientific discussion of clinical audiometry was the classic 1959 article by Carhart and Jerger. They recommended the standard use of the Hughson-Westlake procedure and specified a few variables that were left uncertain in the original report. Even with their further specification, many procedural variables (for example; masking and management of false positive responses) were left to the discretion of the tester.

In 1978 The American National Standards Institute (ANSI) published the American National Standard Methods for Manual Pure-Tone Threshold Audiometry (ANSI S3.21-1978). The standard specified many variables inherent in threshold testing, utilizing the familiar "10-down 5-up" method of Hughson and Westlake. It provides an excellent guideline for manual audiometry. But many variables were not specified. The number of ascending and descending stimulus sequences is not specified, nor is there a governing rule or guideline. There is no rule for testing interoctave frequencies. Masking in air-conduction testing is described in two sentences, with no operational rules governing masker levels or validation of masked thresholds.

Even with the lack of specificity of the standard, some of the recommended methodological variables are routinely ignored by clinical audiologists. Early in the training and experience of audiologists, clinicians begin to take shortcuts which are not governed by formal rules or principles. Many audiologists become highly skilled at obtaining accurate audiograms despite the lack of standardization

However, the lack of standardization results in a level of uncertainty of which, I believe, we are all at least somewhat aware. It is not uncommon for audiograms to be repeated in one clinic based on a lack of confidence in results recently obtained in another clinic. When hearing testing is needed for research purposes, such as studies of effects of treatments on hearing, it is difficult to justify test methods so devoid of standardization.

One could argue that in view of the lack of detail in the standard, or any other authoritative document, it is not accurate to say there is a standard method for pure-tone audiometry.

Among the variables not standardized in clinical practice are:

1. Stimulus starting levels.
2. Hearing level step size.
3. Number of response reversals defining a threshold.
4. Number of descending and ascending stimulus sequences.
5. Masking levels.
6. Stimulus frequencies.
7. Patient instructions.
8. Management of false positive responses.

Perhaps the unstated value of audiograms obtained by experienced audiologists is the expert knowledge used to obtain accurate results and identify potential errors. But skills vary considerably, even among the most experienced audiologists, and this leads to inconsistent quality. Skilled audiologists make observations of patient behavior which impacts diagnosis and treatment including false positive responses, delays in response latency, inconsistency throughout the evaluation, and test-retest reliability, all of which contribute to the audiologist's judgment of the accuracy of the audiogram and the need for further testing -- but these factors are not formally incorporated in audiometric methods.

Personal Perspective:

In 1984 when I was on the faculty at Syracuse University, I visited the private practice of one of the founding members of our profession, past ASHA president, Alan Feldman. Dr. Feldman had retired from the State University of New York Medical Center where he served as Director of the Communication Disorder Unit for many years and established a practice in partnership with an ENT group. During our visit, I observed Dr. Feldman testing a patient with the automated pure-tone Bekesy tracking method. While the audiogram was being recorded, he attended to other things. Although Bekesy audiometry had existed for more than 30 years, it had been used primarily as a "special" diagnostic test. That was the first and only time I saw Bekesy audiometry used for routine threshold audiometry in a clinical setting.

In 1997, I began performing more routine hearing evaluations than I had for many years. Beginning with my early clinical practice with Ken Berger in the 1960's, I always enjoyed the process

of obtaining data to understand the patient's hearing, formulating a management plan, and counseling the patient. As a result of my re-involvement with basic audiometric clinical testing, I became increasingly aware of two aspects of the clinical process which we could significantly improve; informational counseling and automation.

1. Informational Counseling

Although I thought I had developed and could competently deliver to the patient(s) a clear, informative explanation of their patient's hearing loss, the effects of hearing loss on communication, and their many options for management, I began to realize that most of what I told the patient was soon forgotten.

I presumed "patient recall" had probably been studied so I reviewed the literature. In 2004, I published a paper titled "In One Ear and Out the Other – What Patients Remember" the first version of which was published on www.audiologyonline.com. I learned that other professions had learned and documented that patients forget half of what is told to them by medical professionals by the time they get to the elevator (http://www.audiologyonline.com/articles/arc_disp.asp?id=548).

This led to the development of the "*Understanding…"* series of patient education materials to provide important information in a format that can be understood, taken home, discussed with family members, reviewed at later times and used to track progression or remission. These materials can be viewed at www.audiologyincorporated.com and purchased from www.oaktreeproducts.com .

2. Automation

The second aspect of the basic audiometric evaluation which I thought could be improved upon was the use of manually methods for -administering tests which were amenable to automation.

Most patients I've seen are perfectly capable of following instructions necessary to be tested by automated methods, such as those used in psychoacoustics labs for many decades. Through the use of automated clinical tests, limited professional time could be more effectively used in activities that require the skills of a doctoral professional. Automating audiologic tests has the additional benefits of standardizing our methods, and formally incorporating the quality indicators employed by our most skilled clinicians. The incorporation of quality indicators is a critical feature if a skilled clinician is not administering the test. Patient behaviors that are related to the quality of test results can be tracked by computers in real time, quantified, studied and exploited as objective quality indicators, providing perhaps the first significant progress in routine pure-tone and speech audiometric procedures in many decades.

An Adaptive Method for Testing Auditory Sensitivity

Beginning in 1999, I began developing an automated pure-tone audiometric procedure intended for patients that can be tested by routine behavioral audiometry. Quality indicators are an essential

feature of the procedure. In 2002, U.S. patent 6,496,585 was awarded for the method of automatically obtaining a pure-tone audiogram with quality indicators.

To obtain funding for development and evaluation of the method, I applied for a Small Business Technology Transfer (STTR) grant from the National Institute of Deafness and Other Communication Disorders (NIDCD), a branch of the National Institutes of Health. This grant program supports the development and commercialization of products with health benefits. A Phase I grant was awarded in 2001 to develop a prototype and demonstrate feasibility. The results were incorporated into a Phase II application to conduct a multi-center clinical trial. That grant was funded in 2002, and a clinical trial was begun at three research institutions – The University of Minnesota, the University of Utah, and the James H. Quillen VA Medical Center in Tennessee. The trial is still under way and preliminary results are under analysis.

The method is designed for use by audiologists using modern audiometers which can be controlled by computer. After the earphones and bone vibrator are placed on the patient, verbal instructions are provided and the patient responds using a touch screen. The computer determines threshold, quantifies behaviors used to calculate quality indicators, archives the data including a complete history of each stimulus and response, and produces a report.

An Automated Hearing Test for Children

A version of the automated method has been developed for children (5-12 years old) who can be tested by behavioral audiometry. The testing method and quality indicators are identical to the adult version. The primary differences are the instructions and visual reinforcements. The instructions are presented in a training module that determines if the patient understands and responds appropriately. Visual reinforcements are provided after each threshold determination.

Automated Tests of Word Recognition

Automating clinical speech-recognition tests present special challenges. Our routine methods for obtaining speech-recognition thresholds and word-recognition scores are open-set methods that require patients to repeat the word they believe they heard.

Automated speech-recognition (ASR) systems have been in development for many years and are used in research and commercial applications. The fact that in clinical speech-recognition tests there is only one correct response to each test item and the correct response is known, makes the task of scoring the response a relatively easy one for ASR technology. However, the diversity of speech characteristics and dialects of our patients presents a challenge.

In 2003 NIDCD funded my Phase I application to develop automated clinical speech-recognition tests. The ASR software is based on the CSLU Toolkit developed at the Center for Speech and Language Understanding at the Oregon Health Sciences University. The Phase I project indicated a strong possibility that speech audiometric tests could be automated for many patients. This approach has

the added benefit that testing can be performed in any language and the tester need not speak the language of the patient, reducing (but not eliminating) dependence on foreign-language interpreters.

A Phase II application for further development and a clinical trial of automated speech audiometric tests has been submitted and is currently being considered by NIDCD.

The Otogram™

In 2003, Tympany Inc. introduced *The Otogram*™**,** an automated pure-tone and speech audiometer. Pure-tone audiometry is performed by an automated Hughson-Westlake procedure. Masking noise is automatically presented to the non-test ear when the computer determines masking is needed. Speech-recognition thresholds and word-recognition scores are obtained by a closed-set, picture-pointing task.

In December of 2004, Sonic Innovations Inc. announced their acquisition of Tympany Inc. and the formation of the Hearing Health Network (HHN). HHN seeks to establish a provider network that includes physicians who utilize the Otogram™ in their offices, and "hearing health professionals" to whom they would refer patients for hearing care. Hearing health professionals may be licensed audiologists, hearing instrument dispensers or otolaryngologists. Sonic Innovations stated there was no requirement that hearing health professionals dispense Sonic Innovations products.

The HHN website (hhnusa.com) provides data comparing results obtained by the Otogram™ with results obtained by an audiologist. The posted results indicate very good agreement. To my knowledge there are no published, peer-reviewed research articles comparing Otogram™ results to manual audiometry.

The Future of Automated Audiometry in Audiology

Although automated hearing testing is not new, its re-introduction into the clinical arena is seen as both progress and a threat. Over the next few years, automated testing will probably find its proper place in audiology practices and other venues where hearing tests are performed.

In an unpleasant meeting I had a few years ago with a hospital medical director about why our clinic was losing money, she suggested we could improve efficiency by automating some of our tests. Several well-educated patients commented to me after hearing evaluations that some tests I had just performed manually could be automated. When knowledgeable "outsiders" view what we do, we do not always appear to take full advantage of the benefits of technology.

Pure-tone audiometry is governed by a set of rules that can accurately and efficiently be stated in algorithm form and administered by a computer with fewer errorsthan human testers. Computers never forget rules. Even the most skilled clinicians make mistakes, which computers never make. Of course, human errors are offset by the power of clinical skill, cross checking, and clinical judgment. When audiometry is conducted by machine rather than a skilled clinician, the procedure lacks of the expert knowledge of the audiologist whose clinical observations are used to identify and solve problems related

to test accuracy. While this loss is arguably acceptable for a screening test, it is a disadvantage in the clinic where a diverse mix of patients require flexible procedures to obtain accurate results and skilled observation to identify problems.

It is the philosophy of both AMTAS and the Otogram™ that most patients can be tested accurately with automated tests, freeing audiologists' to work with the more difficult cases.

This begs the question, how are "difficult cases" identified? Some can be readily identified by a triage process, an intake questionnaire, or from past knowledge of the individual patient. Others may be identifiable from quantitative quality indicators. Some quality indicators can be performed more capably by computer than by human testers. Computers can quantitatively measure performance variables such as response time, test-retest reliability, false-positive responses, and response consistency and compare the measures to norms in real time, providing an instantaneous assessment of quality.

Therein lies the real power of automated testing. It is not difficult to program a computer to test thresholds; that has been done for decades. If we deploy computers in our clinics to measure the quality and validity of test results we will have the first real advance in basic audiometry since the Battle of the Bulge.

Many audiologists are understandably concerned about the emergence of automated audiometry. These concerns are related to test accuracy, effects on clinical income, the loss of patients to primary care physicians and others, over-utilization of audiometry billing codes and its effect on reimbursement levels, loss of jobs, and a threat to audiology's claim as the professionals who are most capable of assessing hearing.

These issues clearly need to be considered carefully and cautiously as we discuss the proper implementation of automated tests.

I believe our profession should embrace, develop and implement automated testing in our practices. We, in audiology, are in the process of an historic transformation to a doctoral profession. This transformation is occurring because our leaders have successfully argued that audiologists need doctoral-level skills and training to care for our patients in the manner they deserve. Doctoral-level healthcare providers should not spend their time performing routine tasks that, for most of our patients, are amenable to automation. When I go to my primary care physician, he does not measure my height and weight, draw the blood, or run the X-Ray machine. Like audiometry, those tasks do not usually require the skill level of a doctor. We should be "fully occupied with the task of analyzing and interpreting the data" and caring for our patients

My opinion is that if our profession does not embrace automated testing and drive the process of progress, other professions will do it for us.