Otology & Neurotology
39:S30–S42 © 2018, Otology & Neurotology, Inc.

Assessment of Hearing During the Early Years of the American Otological Society

Matthew B. Fitzgerald and Robert K. Jackler

Department of Otolaryngology—Head and Neck Surgery, Stanford Ear Institute, Stanford University, Palo Alto, California

Objective: To describe the manner in which hearing was evaluated in American Otological Practice during the late 19th and early 20th centuries before introduction of the electric audiometer.
Methods: Primary sources were the Transactions of the American Otological Society and American textbooks, especially those authored by Presidents of the Society.
Results: In the era before electric audiometry multiple methods were used for evaluating the thresholds of different frequencies. Tuning forks were important for lower frequencies, whisper, and speech for mid-frequencies, and Galton's whistle and Konig's rod evaluated high frequencies. Hearing threshold was often recorded as in terms of duration of a sound, or distance from the source, rather than intensity. Hearing ability was often recorded a fraction, for example, with the distance a watch tick could be heard over the distance of a normal hearing individual. A variety of devices, such as Politzer's Acoumeter, attempted to deliver sound in a calibrated manner, thus enhancing the accuracy and reproducibility of test results.
Conclusion: The early years of the American Otological Society were marked by a number of ingenious efforts to standardize hearing assessment despite the technical limitations. These efforts facilitated the development of the audiometer, and continue to influence clinical practice even today. **Key Words:** American Otological Society—Hearing assessment—History of otology.

Otol Neurotol **39:**S30–S42, 2018.

The diagnosis and management of hearing disorders has a long history, predating the onset of the American Otological Society (AOS) by centuries. For example, Hippocrates (460–337 BC) is widely regarded as ''The Father of Medicine,'' due to his introduction of key concepts such as the power of observation, the importance of the case history, and for developing the ethical code that underlies many facets of medicine even today (e.g., the Hippocratic Oath). Less known, however, was that he was among the first to investigate hearing disorders (1–3). While his belief that hearing loss was related to the direction of winds or weather changes have not held up to modern scrutiny, his reports that hearing loss is often associated with tinnitus or skull-based trauma reverberate into today's medical practice as part of our modern case history.

Over 2000 years later, the AOS was created, and has played a significant role in the diagnosis and management of hearing disorders over the last 150 years. In modern otological practice, assessment of hearing is a routine and crucial part of patient care. In the early days of the AOS, however, obtaining an accurate measure of hearing was a challenging endeavor, and virtually impossible in many respects. For example, in 1877, Charles Burnett (AOS President 1884–1885) wrote that ''No precise standard of normal hearing has ever been defined. The normal ear hears all sounds that fall on it; but it cannot be said, a priori, where good hearing patients and defective hearing begins, for in many senses these are relative terms (4).'' Similarly, J. S. Prout (AOS President 1886–1889) noted that accuracy of hearing assessment would remain challenging until ''an instrument can be made which shall always produce uniform tones.'' Until the advent of the audiometer, Prout's comment proved largely prescient. Nonetheless, several methods were used to estimate hearing with remarkable degrees of ingenuity; the principles of some of these approaches underlie clinical practice even today. The purpose of this manuscript is to highlight techniques used to assess hearing before the advent of the audiometer, which irrevocably changed hearing assessment for the better shortly after entering into widespread use.

METHODS

The primary resource for determining hearing testing in American Otological Practice during the early years of the AOS was the Transactions of the AOS over its initial decades (5). Additional sources include the otology textbooks and paper

Address correspondence and reprint requests to Matthew B. Fitzgerald, Ph.D., Department of Otolaryngology—Head and Neck Surgery, Stanford Ear Institute, Stanford University, 2452 Watson Court, Suite 1700, Palo Alto, CA 94303; E-mail: fitzmb@stanford.edu
The authors disclose no conflicts of interest.
DOI: 10.1097/MAO.0000000000001759

authored by the founder generation of the AOS including those of Roosa (5), Blake (6), Buck (7), and Burnett (8) as well as publications by others describing hearing test methods in the late 19th and early 20th century America.

The Voice Test

In 1887, An AOS Committee on ''The examination of the power of hearing'' chaired by H. Knapp concluded that: ''The human voice is generally acknowledged to be the most important test of hearing (9).'' This statement is consistent with the idea that perhaps the most common measurement of hearing used in the first 25 years of the AOS was the ''voice test'' or the ''whisper test,'' in which the human voice is used to infer the hearing status of the patient. Variants of this test are used in current audiologic practice with measurement of the speech reception threshold, which is widely used to cross-check pure-tone thresholds. Remarkably, the implementation of ''the voice test'' changed little during the early years of the AOS. In 1869, Anton von Troltsch recommended, ''. . .you must make a closer examination, by testing the power of hearing the voice and conversation. While one ear is being examined to this, the other should be closed by the finger of the patient, and you should speak slowly and distinctly, at first in a whisper. . . You must guard against deception, by seeing that the patient does not practice the habit of watching the mouth of the speaker... thus you will often be informed by a patient. . . that he hears much worse by twilight and at night in bed, than when it is light around him (10).'' In 1882, Winslow recommended, ''It is best to stand a few feet away from the patient upon the side of the ear to be tested, so that he cannot see the lips move, then ask him questions in a low voice. If he cannot hear, address him in a medium tone, and if he is still unable to hear what is said, raise the voice to even a shout if necessary. There are varying degrees of hearing for each tone, but low, medium, and high will be sufficiently exact for all practical purposes (11).'' Thirty years later, Barr provided a similar set of instructions, ''The patient and physician stand at opposite ends of the room, the ear to be examined turned towards the physician. The opposite ear is closed firmly by a finger to the meatus. Standing thus sideways to the physician, the patient cannot see his lips, and the element of lip-reading is eliminated. The physician now repeats the words or numbers which he chooses to employ, the patient having been instructed to repeat after him. If the patient cannot hear, or hesitates, or calls the word out incorrectly, the physician at once moves nearer and repeats the experiment, but using different words, but those having as nearly as possible the same sound values. The distance between patient and physician is thus reduced until one is reached at which the words are repeated promptly and correctly (12).'' The same author noted that differences in pitch, timbre, volume, etc., of different voices make it impossible to determine an exact level of hearing, but reported consensus that conversational speech can be heard at 60 to 70 feet.

Also noteworthy during this time was the awareness that some speech sounds may be audible, while others are inaudible. This point was illustrated in 1877 by Burnett, who articulated, ''The distance at which separate vowels can be heard has not yet been established, but they are endowed with the greatest strength of tone, being heard and understood at a distance at which all the consonants are inaudible (4).'' In his manual, Burnett subsequently provided distances at which various consonants could be heard, noting that ''H is the weakest of all consonants when not followed by a vowel. It is lost at a distance of a few paces. . .. Next in strength is B, Ba being heard further than Ha (4),'' and so on. Such comments are similar in spirit to the early studies of acoustic phonetics, which began in earnest during a similar time frame as to the beginning of the AOS, and were later reiterated by Politzer (13).

Because it was widely understood that the voice can vary tremendously between different individuals, some physicians attempted to standardize presentation of speech of these early attempts, the phonograph was perhaps the most widely used. In 1904, Bentley proposed, ''Instead of employing directly the voice of the investigator, and instead of relying upon acoustic and organic conditions which vary from experimenter to experimenter and from place to place, it proposes to use permanent phonographic records, which can be copied an indefinite number of times and can be reproduced independently of local conditions (14).'' Similarly, in 1890 Fiske noted, ''to sum up briefly we need a method of testing the hearing which shall 1, make use of human speech; 2, which shall be accurate and independent of the examiner; 3, which shall make a record capable of interpretation and use by other aurists (15).'' Fiske proposed using the ''phonometer'' developed by Lucae which would enable a recording of the assessment; this would allow for a record of each appointment, which could then be shared with other physicians as needed. Ultimately, however, the cost of the device, and difficulties with reliability meant that widespread use of the phonometer never occurred. The principles of standardized speech materials, presentation levels, and recording of the responses, however, reverberate through audiologic practice even today.

The Watch Test

In addition to the voice test, one of the most widely used measures of hearing assessment during the early years of the AOS was the ''watch test'' (Fig. 1 (16)). Indeed, it was often stated that, ''Thus far, the ticking of the watch has been found to afford the

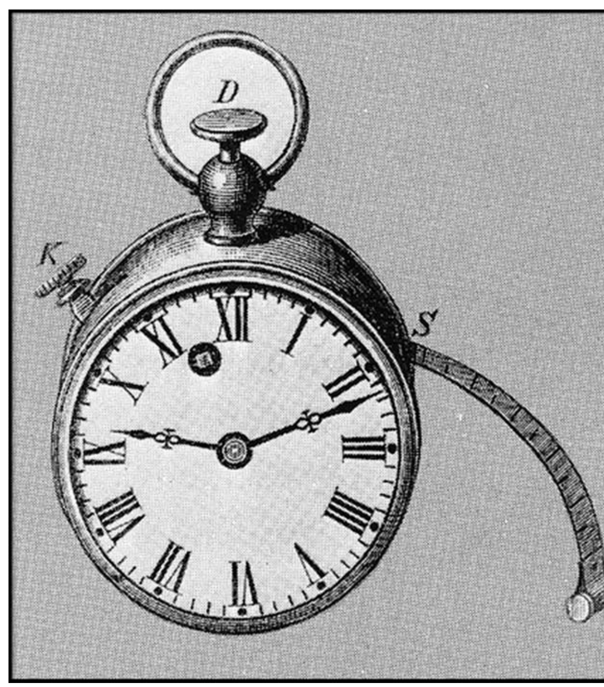

FIG. 1. A specialized watch for use in hearing testing from Bing 1890 (16). Hearing ability was recorded as the distance at which the watch tick could be heard. Note the attached tape measure used for this purpose.

Table showing the Disproportion between the Power of Hearing the Tick of a Watch and the Human Voice.[1]

No.	Sex and age.	Hearing distance for the watch.	Hearing distance for conversation, the patient being with the back to the speaker.
1	Female, 17.	R. $\frac{\text{laid}}{40}$, L. $\frac{0}{40}$.	Words spoken loudly at 10 feet with difficulty.
		R. $\frac{4}{40}$, L. $\frac{0}{40}$.	Loud conversation at 20 feet.
2	Male, 45.	R. $\frac{0}{40}$, L. $\frac{8}{40}$.	Voice at 30 feet; cannot tell the direction from which sound comes.
3	Female, 28.	R. $\frac{7}{40}$, L. $\frac{4}{40}$.	Conversation at 20 feet.
4	Male, 56.	R. laid, L. laid.	Conversation at 20 feet.
5	Male, 62.	R. pressed, L. pressed.	Loud conversation at 20 feet.
6	Female, 23.	R. $\frac{5}{40}$, L. $\frac{3}{40}$.	Loud conversation at 6 feet.
7	Male, 9½.	R. $\frac{1}{40}$, L. $\frac{3}{40}$.	Loud conversation at 30 feet.
8	Male, 16.	R. $\frac{\text{laid}}{40}$, L. $\frac{\text{laid}}{40}$.	Conversation at 20 feet.
9	Male, 18.	R. $\frac{40}{40}$, L. $\frac{30}{40}$.	Conversation at 12 feet.
		R. $\frac{40}{40}$, L. $\frac{40}{40}$.	Conversation at 30 feet.
10	Female, 15.	R. $\frac{4}{40}$, L. $\frac{8}{40}$.	Conversation at 20 feet.
11	Male, 19.	R. $\frac{5}{40}$, L. $\frac{10}{40}$.	Conversation at 20 feet.
12	Female, 29.	R. $\frac{\text{laid}}{40}$, L. $\frac{\text{laid}}{40}$.	Conversation at 20 feet.
13	Male, 40.	H. D. R. $\frac{1/2}{40}$, L. $\frac{\text{laid}}{40}$.	Ordinary conversation with great ease at 30 feet.
14	Female, 25.	R. $\frac{12}{40}$, L. $\frac{6}{40}$.	Ordinary conversation with difficulty at 20 feet.
15	Male, 32.	R. $\frac{0}{40}$, L. $\frac{1\frac{1}{2}}{40}$.	Conversation at 16 feet.
16	Male, 15.	R. $\frac{6}{40}$, L. mastoid.	Conversation at 20 feet.
17	Male, 41.	R. $\frac{8}{40}$, L. $\frac{0}{40}$.	Conversation with ease at 40 feet.

FIG. 2. Table of hearing ability from Roosa 1885 (18) for watch tick compared with spoken voice both expressed in terms of distance from the sound source. The fraction 4/40 refers to perception of the watch tick in inches from the ear (4) over the distance with which a tick was heard for a normal ear (40).

best practical means of testing the capacity of the ear for distinguishing delicate sounds (7).'' The basic premise of this approach was to determine whether a patient could detect the ticking of a watch, and if so, then to determine the distance at which the patient could no longer hear the watch. A detailed and widely cited utilization of this approach was described in 1872 by Prout (17) (Fig. 2 (18)). In his report, he recommended the use of distance to estimate hearing acuity in much the same manner as the Snellen chart is used in the visual system. In his system, hearing acuity was recorded as a fraction. ''The numerator of which is the distance at which the particular sound is heard, the denominator the distance at which it should be heard by an ear of good average hearing power. This denominator must vary according to the sonofactor used, and should generally be expressed in inches (17).'' Thus, 12/36 would indicate that the ticking of a watch was heard at 12 inches, when it should have been heard at 36 inches. According to Prout, one advantage of using fractional distances was its potential applicability to any signal, whether a watch or a whispered voice. In retrospect, it is interesting to consider the use of distance to assess hearing acuity given that the American Otological Society initially began as an offshoot of the American Ophthalmological Society (see Jackler et al., elsewhere in this issue), and visual acuity as a function of distance is a key aspect of the testing of vision.

During the early years of the AOS, several recommendations were given to physicians to increase the accuracy of their measures or the diagnostic power of the watch test. For example, it was generally accepted that ''the distance at which the watch used is heard by the normal ear should be known by the examiner (11).'' Internal consistency in the testing approach was also reported to be a key step, as ''it makes considerable difference whether one hangs the watch by the finger, or holds it in the palm of the hand with the whole hand as a resonator (11).'' The watch was also used to assess hearing via bone conduction, ''The watch may be placed on the vertex or the forehead to determine roughly the condition of the middle ear and auditory nerve. . . If the watch is not heard when applied thus, it is pretty sure evidence that there is disease of the labyrinth or nerve (11).'' Finally, use of a stopwatch was widely recommended as well; the rationale behind this recommendation was that with a stopwatch, the ticking can be stopped or started, and in this way false positives (e.g., reporting hearing the watch when no ticking is present). In other words, use of the stop watch was a ''means of finding out whether the patient really hears the sound of the watch, or whether he thinks he does because he knows a watch is being held before his ear.'' This approach was reported to be particularly useful with children who ''as a rule, give erroneous statements as to their ability to hear a watch (11).''

While the watch was widely used, its limitations were evident from the beginning. First and foremost, watches regularly differed with regard to the intensity and pitch of the ticking; for obvious reasons, this meant that the replicability of hearing tests across institutions was virtually nonexistent. Such concerns were articulated effectively by Albert Buck (AOS President 1879–80) in 1880. ‘‘If measurements of the hearing distance could be universally made with some standard source of fixed intensity, the necessity for recording our measurements in fractions (Prout’s method) would be done away with; it would be sufficient to merely state the actual distance measured, and every physician who was familiar with such tests would appreciate at once the degree of impairment of the hearing reported (7).’’ Another significant limitation was the relationship between hearing a watch and the ability of the patient to communicate with others. Such concerns were noted as early as 1853, ‘‘The degree of hearing with a watch is sometimes deceptive; some patients who cannot hear a watch, or even a clock, will hear the voice even in a low tone (19).’’ Such concerns were repeatedly articulated in different textbooks of Otology, ‘‘The watch alone does not afford a sufficient means of determining the amount of hearing examined, because the distance at which it can be heard does not always stand in proper proportion to the power of understanding conversation (11).’’ Nonetheless, use of the watch to assess hearing status continued until widespread adoption of the audiometer too place. For example, more than 50 years later, general guidelines were provided to physicians as to its use, ‘‘Naturally, this sound (the watch tick) varies considerably in intensity with the size, form, thickness of covers, etc. of different watches. Taking, however, a man’s watch of average size, its tick will be heard by the normal ear of a young adult... at a distance of 40 to 50 inches... As age advances, the hearing distance for the watch is gradually diminished... (20).’’

POLITZER’S ACOUMETER

As noted by Buck and many others, there was an understanding that accurate assessment of hearing would require a signal of a given intensity which could be reliably delivered. Early attempts in this regard were often classified as ‘‘Mechanical Acuity Meters.’’ Among the earliest of such devices was reported by Wolke in 1802. His device was comprised of a pendulum-like hammer that could be dropped onto a wooden board approximately 1.5 m high. The height of the pendulum swing could be varied, and by doing so, different intensities of sound could be produced. This sort of device was improved upon approximately 20 years later by Itard with the development of the ‘‘accumeter (21).’’ In this device, a ring of copper was used as the sound source; the ring was suspended by a string, and struck by a ball at the end of a pendulum. The strength of the strike, and thus the intensity of the signal, depended on the height from which the ball was dropped. This instrument was widely used in the early half of the 1800s, as Itard was the Director of the Paris Institute for the Deaf. [See Feldman’s History of Audiology for a more detailed summary of these early mechanical acuity meters (1)].

The most well known and widely used of these devices was the acoumeter developed by the legendary otolaryngologist Adam Politzer of Vienna in 1877 during the beginning stages of the AOS (Figs. 3 and 4) (11,22). One key advancement of Politzer’s acoumeter was that it was hand-held between the middle finger and the thumb. When the middle finger was depressed, it would raise a small mallet; when released, the mallet would fall to strike a small iron cylinder. The primary advantage of this approach was that the mallet was always dropped at a constant height, unlike the aforementioned devices in which the height was generally estimated. Hearing was then measured at known distances at which individuals with normal hearing could detect the sound of the mallet. This provided for a more consistent measure of hearing assessment than the widely used ‘‘watch test.’’ A key advantage of Politzer’s acoumeter was that, by attaching a small metal disk to the acoumeter, bone-conduction hearing could also be measured using this device.

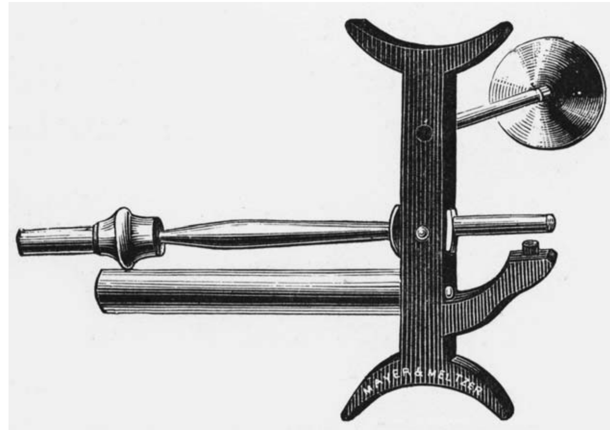

FIG. 3. Politzer’s Acoumeter from Love 1904 (22). Sound was generated by a small mallet struck by a metal rod with calibrated force not as readily obtained with tuning forks. An attached metal disk was used for bone conduction.

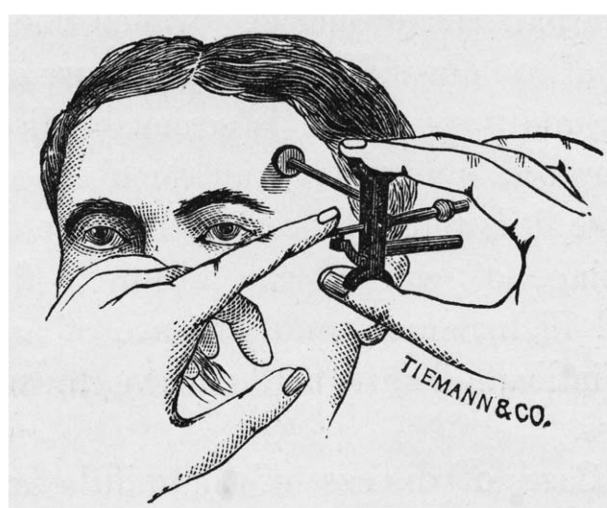

FIG. 4. Illustration of Politzer’s Acoumeter in use for bone conduction from Winslow 1882 (11).

While Politzer's acoumeter resolved some of the concerns surrounding hearing testing at the time, it also presented with a number of limitations. For example, as with the watch test, the relationship between hearing acuity measured with the acoumeter and the ability to understand speech was poorly understood at best. More problematic for some physicians was the fact that early acoumeters were ''being nothing more than loud watches (1),'' and Politzer's acoumeter was plagued by a similar issue. Politzer himself noted that ''The acuteness of hearing for the acoumeter, or for the watch, frequently shows marked differences...'' with an average normal hearing distance for Politzer's acoumeter being 15 m (13). Ultimately though, the factors that may have hindered greater acceptance of Politzer's acoumeter were described succinctly by Buck in 1880, ''Politzer's idea in producing the ''acoumeter'' undoubtedly was to furnish a standard test of hearing. Unfortunately, in its present shape this instrument costs too much, is likely to get out of order too easily, and cannot be manipulated with comfort (7).'' Others held similar views, which persisted for over 30 years, ''Use the stop watch with a fairly sharp tick; this will take the place of the Politzer acoumeter, which can be discarded (12).''

TUNING FORKS

The use of tuning forks to evaluate hearing began early in the 19th century (23–25) They were originally developed to assist in tuning musical instruments. By the late 19th century, they their use had become routine, but not necessarily universal. In an 1887 position statement in the AOS Transactions titled ''The examination of the power of hearing, and how to record its results'' opined that tuning forks should be part of the standard hearing evaluation: *''They should, in every case of impairment of hearing, be used as regularly as the watch and voice tests. (9)''* By contrast, in his 1880 textbook, Buck in his chapter on ''Test of the Hearing-Power'' did not even mention tuning forks, emphasizing instead perception of the spoken voice and watch ticking (7).

There was a wide diversity of tuning fork design (Fig. 5 (22)). Typically, forks were available in C-tones one octave apart: 64, 128, 265, 1024, and 2048 Hertz. To

FIG. 5. A collection of tuning forks and whistles used in clinical otology from Love 1904 (22).

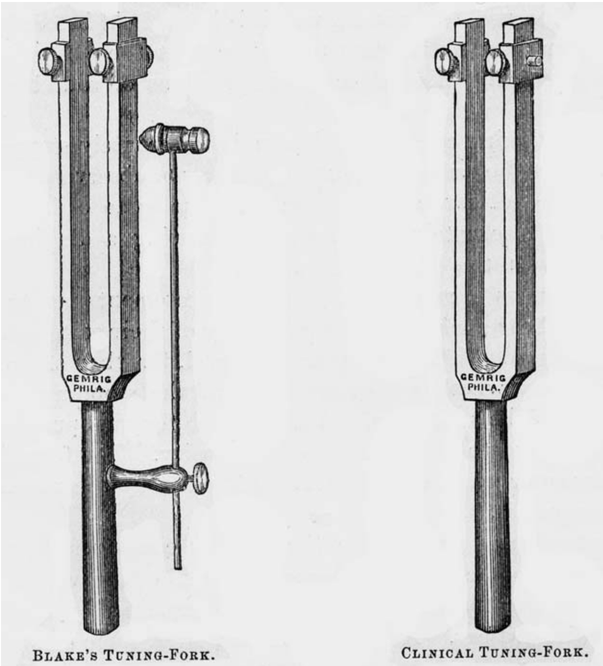

FIG. 6. Modified tuning forks from Burnett 1877 (4). The Blake modified tuning fork had an attached hammer in an effort to calibrate the strike force. The adjustable weights at the end of the tuning fork served two purposes: dampening overtones and adjusting the pitch of the fork's ring.

mitigate the potential for excessive strike force to generate overtones, some had small attached hammers to help calibrate the amount of force to the tine. Burnett in 1877 lauded: ''A very beautiful instrument is the tuning fork devised by Dr. C. J. Blake, in which the force setting in vibration is obtained by means of a steel hammer padded with rubber. The handle of the hammer is adjustable at any point along its length, but which means the blow can be weakened or strengthened s desired (4)'' (Fig. 6 (4)). Charles H. Burnett was AOS President 1884–5 while Blake served in this role 1877–78. Other tuning fork designs had clamps attached to dampen overtones, but these tended to shorten the vibration period. Forks with an attached weight which could be slid along the tine allowed tests multiple frequencies without the need to carry a large supply of individual frequency forks. Others had resonating chambers to enhance the sound for patients with severe losses (Fig. 7 (26)). Tuning forks with a rubber tube attached were facilitated comparison of the physician's hearing with that of the patient (Fig. 8 (27)).

Today only two tests introduced in the mid-19th century, Weber (1845) and Rinne (1855), remain in widespread use. Over 20 different tests were in use during the late 19th and early 20th centuries before the introduction of the electric audiometer. In 1887, Knapp and his co-authors emphasized the central importance of the Rinne test: ''Rinne's method, gentlemen, is the most expeditious and practically the most important (9).'' In

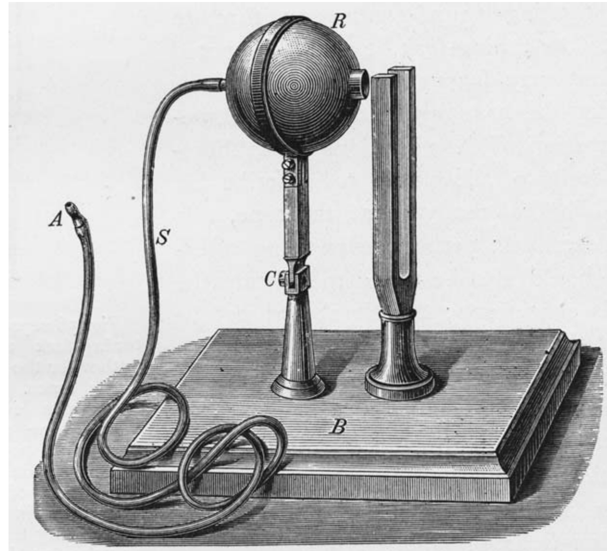

FIG. 7. Tuning fork with resonating chamber from Gruber 1890 to enhance audibility for those with severe hearing losses (26).

1881, DB St. John Roosa (AOS President 1874–76) explained the use of the Rinne test: ''If the vibrating tuning-fork be heard better on the mastoid than when placed in front of the meatus, there is disease predominantly of the middle ear (28).'' Roosa also explained the Weber test: ''If one ear be normal as to the hearing power, and the other abnormal, and a vibrating tuning-fork ''C'' be placed upon the vertex or the teeth, if its sound be intensified in the ear whose hearing power is diminished, there is disease of the external or middle ear, but no lesion of the labyrinth or nerve (28).''

Other tuning fork tests were developed to discriminate sensory from conductive losses. In the Bing test, a tuning fork is placed on the mastoid and when it is no longer

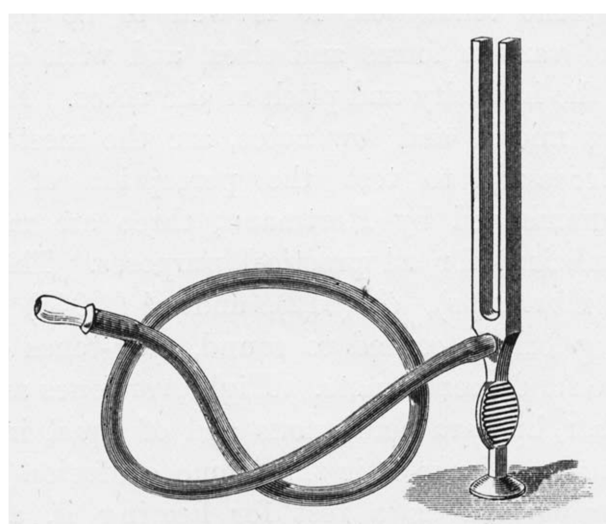

FIG. 8. Tuning fork with a listening tube by Hovell 1894 allowing comparison of the examiner's perception with the patient's during bone conduction testing (27).

heard the patient's meatus is occluded with a finger. In sensory losses, perception of the tone returned whereas in conductive losses it did not. The Gelle Test was intended to evaluate severe hearing loss for stapes fixation. With the fork on the mastoid compressed air was delivered to the ear canal via a Politzer bag. If the stapes was mobile, it was compressed inward thus diminishing hearing. If fixed, the added pressure did not alter hearing.

While tuning forks are now used principally to discern sensory from conductive loss, in the 19th century they were also a primary means of assessing hearing ability at different frequencies. Use of the tuning fork for threshold testing is somewhat of a lost art today. For threshold testing, tuning forks were especially important for the lower frequencies, speech for the mid-frequencies, while Galton's whistle and Konig's rod evaluated the high frequencies (6). In Schwabach's Test the duration by which the tuning fork is heard when applied to the cranial bones and compared with the duration of a patient of similar age with normal hearing. Measuring the duration of hearing with a variety of tuning forks, struck in a consistent manner, could provide an estimate of threshold not dissimilar to an audiogram (Fig. 9 (29)). Criticism of the Schwabach test was that it was laborious and time consuming and required repetition at each frequency to enhance accuracy. An early form of audiometer consisted of a rotating turret of tuning forks of various frequencies struck in a calibrated manner with a hammer connected to a stop watch (30). The tuning fork audiometer charted the number of seconds perceived at each frequency.

An awareness that the diagnostic reliability of tuning forks is imperfect was recognized in the 19th century. Striking the fork with an excessive force results in overtones at higher frequencies that intended (31). The difficulty in assessing one conduction at low frequencies due to vibrotactile perception was understood: ''In testing bone conduction for lower tones, it is difficult to determine

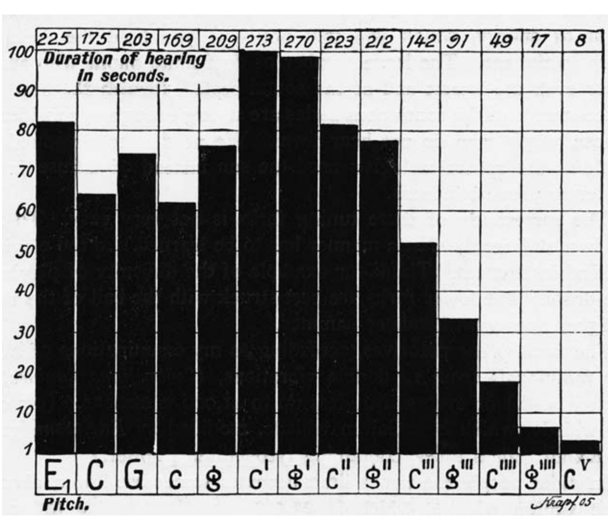

FIG. 9. An "audiogram" from Bezold 1908 created by charting the duration which the patient heard a tuning fork as its vibrations abated (29).

whether the patient hears the fork or feels the jar transmitted to the head. Some patients can differentiate between the two sensations, while others admit that they cannot be sure whether they feel the vibrations or hear the sound (31).'' In 1887, Theobold noted regarding the Weber test: ''It is by no means an uncommon experience with me, when testing with the tuning-fork, that when I place it on the vertex it is heard louder, we will say, in the right ear. Then I will strike it again and place it on the forehead, and it will be heard louder in the left ear. This observation has given me less confidence in the tuning-fork as a differential test between middle ear and labyrinthine troubles than I before had (9).'' Regarding the Rinne test, the nomenclature that designates a negative test to be abnormal and a positive test normal has been a source of confusion since the test was first described. Many contemporary otologists use the terms AC > BC (air conduction > bone conduction) or BC > AC to avoid confusion. In his 1902 contribution titled ''Sources of error in functional tests of hearing'' A.H. Andrews descried: ''In the Schwabach test there are two objections to forks which can be heard longer that the time mentioned: 1. In making repeated tests in order to secure accuracy, much valuable time is lost waiting for the fork to run down. 2. Repeated tests with forks which vibrate along time are apt to wear out the patient's attention, so that after a few trials his replies are found to be uncertain (31).''

In the 21st century, tuning forks are hardly a quaint anachronism and remain relevant in contemporary ontological practice. Their use is both art and science, with results varying, and the clinician needs to exercise judgement in interpreting results. In today's practice, tuning forks are an important check of the audiogram in cases of apparent conductive losses. Insufficient masking can make a deaf ear appear to have a conductive hearing loss, with the potential for misdiagnosis leading to improper therapeutic intervention. Use of the Rinné and Weber tests can clarify this situation. James Sheehy (1926–2006) of the House Group routinely inquired of his neurotology fellows about whether or not they completed the ''DFTF test.'' New fellows soon became initiated in his meaning: ''don't forget tuning forks.''

OTHER METHODS FOR ASSESSING TONAL HEARING: THE GALTON WHISTLE, KONIG RODS, AND SCHULZ'S MONOCHORD

One widely known limitation of hearing assessment in the 19th century was the inability to reliably test hearing for higher frequencies. The importance of the use of high-frequency tonal stimuli was articulated clearly by Blake in 1879, when he wrote ''that the upper limit of audibility of high musical tones by the normal ear being taken as the standard, any considerable deviation from this standard, within certain limits, may be taken as evidence of an abnormal condition – a. of the sound-transmitting apparatus of the middle ear; b. of the sound-transmitting structures of the labyrinth; c. of the auditory nerve and the ultimate organ of perception (6).'' To address the

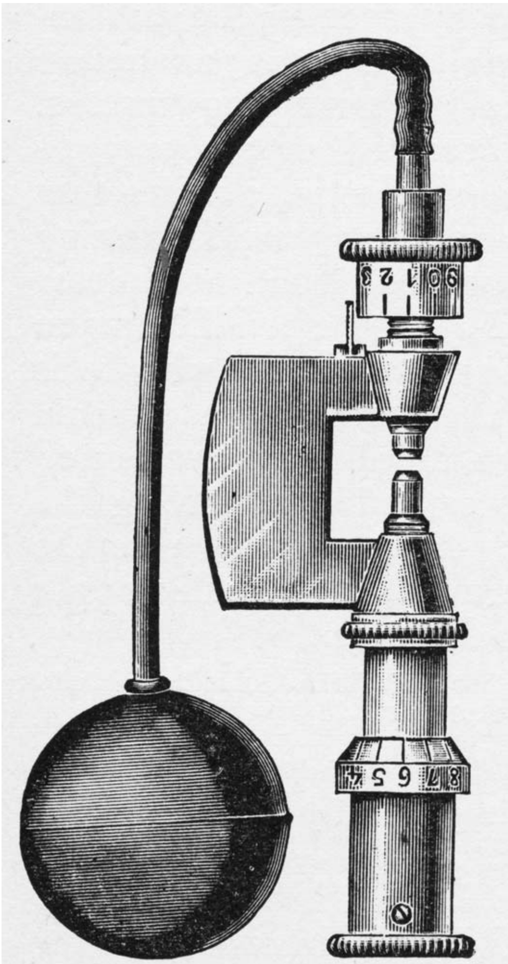

FIG. 10. Galton whistle from Bruhl 1906 produced a variety of high pitched sounds by varying its aperture.

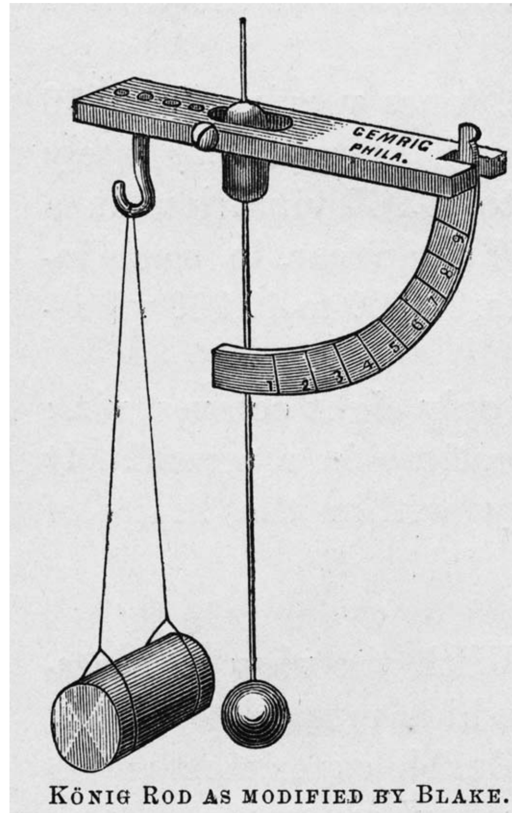

FIG. 11. Koning's rod used in testing high frequencies from Burnett 1877 (4). Intensity was determined by the height to which the ball bearing was lifted before release.

limitation of tools for assessing high-frequency hearing, Sir Frances Galton (1822–1911) invented the ''Galton Whistle'' in 1876 (Fig. 10). This device consists of a small whistle, which has an obturator controlled by a slider. By varying the aperture, the frequency produced by the whistle can be varied. Some variants of the Galton Whistle produced sounds ranging from 5 to 42 kHz. Galton successfully used this device to estimate the hearing acuity in both humans and animals; much of this work is described in his 1883 book, ''Inquiries into Human Faculty and Its Development.'' Most notably, through the use of this device, Galton estimated that the upper limit of normal human hearing was approximately 18 kHz, and that the ability to hear high frequencies deteriorated with age. Thus, Galton's research provided some of the earliest characterization of presbycusis in humans. After the audiometer entered widespread use in the 20th century, Galton's whistle continued to be used to assess hearing in animals. Notably, use of this device continues even today, but is widely referred to as the ''dog whistle.''

Konig's rods were similar in principle to tuning forks, and like the Galton whistle, were used to assess high-frequency hearing sensitivity (Fig. 11 (4)). The Konig rods consisted of steel cylinders suspended by cords that produced high-frequency tones when struck by a small hammer. While similar in their use to Galton's whistle, they differed in some key respects. For example, ''the intensity of the tone of a Konig's rod diminishes regularly from the moment that it is set in vibration, while the intensity of the tone of the whistle evidently can be maintained. The auditory impression produced by the latter is therefore proportionately greater, and of two tones of the same pitch, sounded at the same distance, by Konig's rod and a whistle, the latter will be more distinctly heard (6).'' While Blake favored the use of the Konig rods, they ultimately fell out of favor because of the factors described above, leading some to conclude that they ''provide notes of constant pitch, but with variable intensity. They are inconvenient, and not of general utility (32).''

Schulz' monochord was another tool developed to deliver high-frequency tonal stimuli to the patient, and to determine the highest frequency that could be heard by a given individual. This consisted of a metal wire akin to those used in string instruments. When vibrated, the string would elicit a high-frequency tone, and the patient would indicate whether the tone was heard. The monochord was not as widely adopted, but its adherents noted

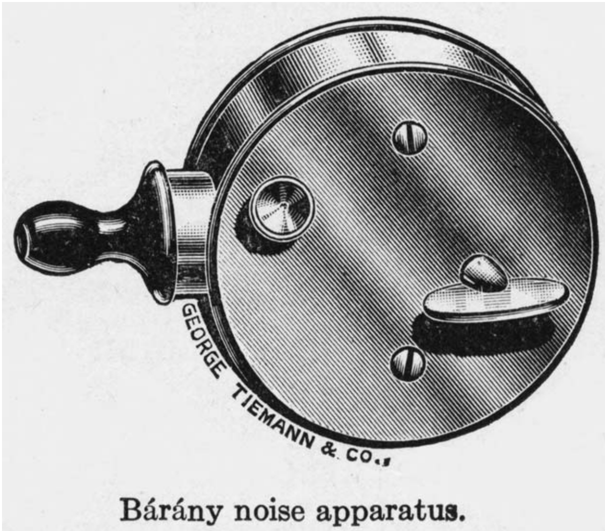

FIG. 12. Barany noise apparatus also known as Barany noise Box from Gorham Bacon (AOS President 1891–4) 1918 (33). The device, still in use today in many centers, is used to mask the better ear in unilateral or asymmetrical hearing loss.

some advantages over the Galton Whistle or Konig's Rods. Notably, ''the limits (of high-frequency hearing) when tested by the whistle is lower... a finding which may be due to the whistle giving less intensity of sound at these high pitches (32).''

EARLY ATTEMPTS AT MASKING: THE BARANY NOISE BOX

Another widely known constraint on hearing assessment in the early years of the AOS was the inability to reliably test hearing in one ear without the contralateral ear contributing in some capacity. Most attempts involved closure of the contralateral ear canal by some means; basic forms of plugging the ear were widely used in the various voice tests. However, it was clear that such approaches were likely insufficient to achieve their desired goal, particularly when trying to identify or rule out unilateral deafness. Thus, in 1908, Barany introduced his ''noise box'' or ''noise apparatus,'' as it became widely known (Fig. 12 (33)). To use this device, it would first be wound up similar to that of a watch. It was then inserted into the ear to be masked, and when turned on, would create a loud buzzing sound while the examiner speaks or shouts into the contralateral ear. If the patient failed to respond, the ear was considered ''Barany deaf.'' Around this time, other approaches were developed with the intent of achieving the same goal as described in Feldman's History of Audiology (1). For example, Voss blew compressed air into the ear to be masked (1908–1909), while Luc advocated caloric irrigation (1910). Lucae (1908) and Davidson (1910) both attempted to mask one ear through use of an electrical vibrator. Nonetheless, the Barany noise box was likely the most widely used of these approaches, although some physicians noted limitations of this device, ''In my experience this apparatus has proved useful, but it has seemed sometimes to so confuse the individual as to prevent an accurate test of the ear under examination, the ear hearing both the noise machine and the fork or voice, as the case may be, but both with less accuracy (12).'' Such comments are interesting as they reflect the experiences of many modern-day audiologists, either with regard to over-masking, or with ''central masking'' (e.g., decrease in ipsilateral hearing threshold when the noise is presented to the contralateral ear, presumably due to a central mechanism).

TESTING OF CHILDREN

With the limited (by today's standards) tools for hearing assessment, it is perhaps no surprise that even fewer options were available for ''hard-to-test'' populations such as children. For example, the limited reports on hearing assessment in children generally noted that children are often unreliable in their responses, and that caution should be taken when assessing hearing in this population. Representative comments are found in the chapter from Barr, who stated that, ''we contend with two principal difficulties: the unwillingness or the inability of the child to answer correctly... The little patients tire easily... Prolonged examinations of children under 10 years are apt to be unsatisfactory... Many children cannot be accurately tested until the third or fourth year in school... (12).''

One unique report during the early years of the AOS, however, came from Harold Walker in 1907 (34). He reported data on 289 children who were tested in their school in a quiet room. The session began with examining the eardrum, nose and throat. Then, ''the hearing was tested by a whispered voice which could be heard by the average normal ear at a distance of twenty-five feet, and a spoken voice with thirty-five feet as the normal limit. Numbers from one to one hundred, words, and short sentences were used, and the distance at which the child could repeat what was heard was recorded (34).'' Prout's ratio approach was then used to determine the hearing ratio, and these results were compared with the presence of adenoids, abnormal otoscopic results, and ''the general facial expression.'' Using this technique, 2/3 of children were reported to have normal hearing, with 23% having ''hypertrophied turbinates,'' 21% showing ''chronic supparation of the middle ear,'' and so on. Finally, he reported what may be perhaps the first relationship between hearing acuity and academic performance, as ''of the pupils marked with the grade of 'excellent' 17% showed diminished hearing. Of those marked 'good', 20% showed diminished hearing. Of those marked 'fair' 30% showed diminished hearing. Of those marked 'poor' 42% showed diminished hearing... (34).'' Such work was prescient in many regards, as the relationship between untreated hearing loss and decreased academic outcomes has been replicated on numerous occasions.

TESTS TO DETECT MALINGERERS

Malingering was well known and tests were devised to reveal it. According to Kerrison in 1922: ''Pretended deafness is said to be comparatively common in countries where army service is compulsory (20).'' He went on to remark that: ''In America it is met with chiefly in the case of imposters seeking indemnity on account of pretended injury to one or both ears (20).'' Regarding the identification of malingerers, several approaches were used, some of which continue to be used today in various forms. Among the first tests, and one that continues to be widely used is the Stenger test. This test is based on the Stenger principle, which stipulates that when a signal of two intensities is presented to two ears of similar hearing, the patient will only report hearing in the ear that receives the more intense signal. This test can be performed with two tuning forks held at the same distance from each ear. In the case of feigned unilateral deafness, as one tuning fork is brought closer to the ''impaired'' ear, the malingerer will report hearing nothing. However, given that the other tuning fork has not moved, and remains audible, the physician can then determine that the patient is malingering. Feigned bilateral deafness was reported to be more challenging to identify however, with sample approaches being ''to wake the patient from his sleep by a moderately loud call (13)'' or ''by making disparaging remarks about him in the presence of a third party, one may be able to determine by changes in his facial expression his ability to hear the conversational voice. Usually, however, the pretense of complete bilateral deafness is too difficult to maintain... (35).''

THE ADVENT OF THE AUDIOMETER

Perhaps the most significant development in the assessment of hearing was the audiometer. This device not only revolutionized hearing assessment, but the practice of Otology, and paved the way for the birth of Audiology to come in subsequent years. Shortly after Alexander Graham Bell invented the telephone in 1876, the electric audiometer began to be developed, and these efforts were led by A. Hartmann and D.E. Hughes. In 1878, Hartmann developed an instrument for hearing assessment in which electric current was used to vibrate a tuning fork, and the resulting signal was then passed through a telephone receiver (36). In 1879, Hughes developed what he termed an ''electric sonometer,'' which also used electric current to vibrate a tuning fork. In his device, the electric current could be increased or decreased by sliding a movable induction coil, and by this process hearing acuity could be assessed (37). Both devices were limited, however, by several factors, including that ''Different fundamental tones can be secured only by installing forks of different pitch. This tends to make the apparatus complicated, unstable, cumbersome, and difficult to standardize (38).'' These attempts were followed by several others until the first commercially available audiometer was patented in the United States in 1914. This audiometer was the Western Electric 1A, which was limited by its size and prohibitive price. However, it was followed closely by the Western Electric 2A in 1923, which was considerably smaller and designed for clinical use. This device rapidly gained acceptance by many otologists. The history of these audiometers, and their predecessors were described in great detail by CC Bunch in 1941 (38), and by Feldmann in 1970 (1), and will not be discussed in greater detail here.

Of historical interest are the thoughts of some AOS members during the advent of the audiometer. For example, in 1930, Keeler stated, ''The greatest value of the audiometer is the possibility of a uniform standard of measuring hearing loss which it presents. At present, every otologist has his own method of testing, and of estimating the loss of hearing in the subjects of aural impairment whom he examines. There is no uniform standard, and the examiner in California whose patient travels to New York cannot send his records to his colleague on the Atlantic coast into whose hands the patient goes, with any certainty that they will coincide with the records and standards of the New York otologist (39).'' Similarly, when describing the existing test battery (e.g., voice, watch, and tuning forks), Clarke noted that ''these tests form the backbone of our functional diagnosis, and I believe that the lack of otological progress in the last thirty years is largely due to their inherent inaccuracy (32).'' Ultimately, the development and widespread adoption of the audiometer led to a significant change in hearing assessment, and virtually obviated the previous forms of hearing assessment other than tuning forks.

EARLY ATTEMPTS TO PLOT HEARING—AND WHY IS THE AUDIOGRAM UPSIDE DOWN?

In modern otology practice, the audiogram is virtually ubiquitous. However, in the early years of the AOS, attempts to plot hearing ability varied tremendously depending on the approach utilized. For example, in 1885, Dr. Hartmann created the ''Auditory Chart'' to record results from tuning-fork testing. This chart indicated the length of time that a given tuning fork could be heard; seven tuning forks ranging from 64 to 4096 Hz were included in this graph. To facilitate interpretation, Hartmann even provided ''norms'' for the duration that each tuning fork could be heard via air and bone conduction (1). Similar tables existed to report hearing for the voice, or the watch tick. Hartmann's normative data for tuning forks were eventually called into question. Nonetheless, the desire to have true normative data for hearing persisted, and eventually led to the creation of the audiogram.

While the history of the audiogram itself extends far beyond the first sesquicentennial of the AOS, it may be of interest for the AOS members to know why the audiogram is plotted ''upside down,'' with regard to hearing thresholds. A more detailed accounting of how this came to pass was provided by Dr. James Jerger in 2013 (40), and is well worth reading. An abridged version of his article is as follows.

Dr. Edmund Fowler (AOS President 1930), a legendary otologist from the first half of the 20th century, came to work closely with Dr. Harvey Fletcher and RL Wegel. Fletcher was one of the early pioneers in the field of speech and hearing sciences, while Wegel was a physicist who worked predominately with telephones. Fletcher and Wegel designed the first commercially available audiometer in the United States, the Western Electric Model 1-A; this device was subsequently used in the practice of Dr. Fowler. The question then became how to represent the data obtained from the audiometer.

In 1922, Wegel (41) published research demonstrating the range between audibility and the sensation of ''feeling.'' From data of this sort, Fowler derived that, when intensity was plotted in a logarithmic manner, hearing could be plotted in terms of ''sensation units'' relative to normal hearing. In this manner, for each frequency the number of ''sensation units'' could be determined. Then, based on the intensity required to obtain the threshold of audibility, one could determine the percentage loss of sensation units. Thus, this approach gave the physician and the patient the ''percentage of hearing remaining'' at a given frequency. Notably, early attempts to plot this graphically had 100% at the top, and 0% at the bottom; in other words, better hearing was depicted at the top, and worse hearing at the bottom of the graph, akin to what we see in today's audiogram. Fowler favored such an approach, as he thought that a percentage of remaining hearing at a given frequency made for an excellent counseling tool with patients. Based on comments from patients even today who ask questions such as ''What percentage of hearing loss do I have?'', many audiologists and physicians might concur that such an approach would be useful!

However, Fletcher was a physicist, and argued that a more accurate representation of hearing should convey the units of hearing loss (e.g., pressure levels needed to elicit a response) rather than a percentage. His early presentations on representing hearing levels in this way plotted these pressure levels in a conventional manner (e.g., more intense signals toward the top of the graph, rather than at the bottom). In today's clinical practice, many audiologists fitting hearing aids would agree that this would be a logical way to plot hearing thresholds, because plotting SPL as a function of frequency is precisely how hearing aids are fit today!

Eventually, Fletcher convinced Fowler to abandon the ''percentage of hearing loss'' approach to plotting hearing thresholds. Upon doing so, however, he surprisingly did not change the scale along the *y* axis. Rather, he simply shifted the ''percentage loss'' to ''sensation units'' and left the zero line at the top of the graph, while renumbering the *y* axis so that increasing amounts of hearing loss were lower. Ultimately, this had the effect of ensuring that the audiogram would forever be ''upside down,'' with the inverted *y* axis to which we have become accustomed. Eventually, the concept of sensation units was modified to a decibel notation, which was later converted to the ''dB HL'' (Hearing Level in dB) that we know today, and the audiogram has since remained unchanged for decades.

Dr. Jerger wisely notes that either Fowler's original suggestion of ''percentage loss'' as a function of frequency, or Fletcher's revision which plotted intensity for hearing threshold in a conventional manner (e.g., more intense signals at the top of the graph) would be preferable to the current plotting of the audiogram. First and perhaps most important, both would preserve traditional plotting of data in which larger values are at the top of the graph, and smaller values at the bottom. Moreover, Fowler's ''percentage loss'' approach has great counseling utility for the layperson, while the plotting of SPL as a function of frequency would align hearing thresholds with procedures for fitting hearing aids. The latter approach would also have counseling benefits during the fitting process itself, as it would help both audiologists and patients avoid the ''mental gymnastics'' sometimes necessary to convert from the existing dB HL graph to a traditional plot of sound pressure as a function of frequency. Given its ubiquity of the audiogram in today's practice, it is highly unlikely to ever be changed, but it is interesting to consider the possibilities had different decisions been made by Fowler and Fletcher over 80 years ago.

DISCUSSION

Mark Twain was a contemporary of many of the early AOS members, and in a letter to Helen Keller, he wrote, ''. . .all ideas are second-hand. . .'' In his own autobiography, he expanded on this concept stating, ''There is no such thing as a new idea. It is impossible. We simply take a lot of old ideas and put them in a sort of mental kaleidoscope. We give them a turn and they make new and curious combinations. We keep on turning and making new combinations indefinitely; but they are the same old pieces of colored glass that have been in use through all the ages.'' When one considers how hearing assessment has evolved since the first 25 years of the AOS, one could readily agree with Twain that ''there are no new ideas.'' The otologists of that time knew the limitations of their chosen approaches, whether the watch, the voice, or the tuning fork. Leaders of the founder generation of the AOS including Roosa, Buck, Burnett, and Blake each knew that a reproducible signal with true normative data was required to obtain a truly accurate measure of hearing, and they strove to create those norms using the best available tools at that time. Many of the lessons learned from their efforts are still in use today. Ultimately, they were limited by the available technology, and not by their ideas. Rapid technical advancements in the first half of the 20th century ultimately led to significant changes in how hearing was measured. At the same time, however, the goals of those assessments remained unchanged from those of the 19th century, which are reliable methods for determining hearing acuity via air and bone-conduction, and assessing the ability of the patient to communicate with others.

In many regards, hearing assessment has changed little in the last 60 years. Air- and bone-conduction thresholds, along with word-recognition in quiet, make up the majority of audiologic evaluations, just as they did shortly after widespread implementation of the audiometer. While useful, one could argue that new advancements in technology could again yield to sizable revisions in clinical practice, and frankly, that such revisions are long overdue. One likely shift will involve the desire to better understand the ability of patients to communicate in their everyday environment. Such advancements could come from the simple introduction of speech-in-noise assessment as part of the basic audiologic test battery. However, given improvements in signal processing, such measures may be needlessly simple. For example, it is possible to digitally create virtual auditory environments; this could enable the testing of patients in increasingly realistic environments in an extremely controlled manner. Such assessments could occur within the clinic, or even outside the clinic with improvements in data logging and auditory environment recognition algorithms in both hearing aids, cochlear implants, or perhaps even smartphones.

Such potential advancements in patient assessment are potentially compelling, but it is also possible that there may be a period of time in which hearing assessment outside the sound booth could become as fragmented as the varied approaches used in the 19th century. For example, there are a myriad of smartphone and tablet apps from hundreds of sources, all purporting to provide some measure of hearing acuity. These devices use different approaches, often with little normative data, and unreliable equipment due to different types of headphones in different acoustic environments. Moreover, the proliferation of such apps may well increase given the deregulation and seismic changes about to take place within the hearing aid and ''hearables'' marketplace. Taken together, audiologists and physicians may need to engage in the realm of hearing assessments not only within, but outside of the sound booth, to provide some order in what looks to be an increasingly chaotic marketplace. Ultimately, our current approaches may seem primitive relative to those in use 150 years from now, particularly with regard to those procedures implemented outside of the physician's or audiologists' office. However, Twain's ''kaleidoscope of ideas'' that will underlie the new testing procedures are likely to remain unchanged, in much the same way that the principles underlying our current testing procedures echo those from 150 years ago at the beginning of the AOS.

REFERENCES

1. Feldmann H. *A History of Audiology: A Comprehensive Report and Bibliography from the Earliest Beginnings to the Present*. Beltone Institute for Hearing Research; 1970.
2. Psifidis A. The prehistory of audiology and otology. 2006;2:41–46.
3. Manolidis L. Otolaryngology through the works of Hippocrates. *ORL J Otorhinolaryngol Relat Spec* 2002;64:152–6.
4. Burnett. Charles H. *The Ear: Its Anatomy, Physiology, and Diseases. A Practical Treatise for the Use of Medical Students and Practitioners*. Philadelphia, PA: Henry C. Lea; 1877 .
5. Roosa D. *A Practical Treatise on the Disease of the Ear Including the Anatomy of the Organ*. William Wood & Company; 1874.
6. Blake C. Audibility of high musical notes. 1879;1:267–281.
7. Buck A. *Diagnosis and Treatment of Ear Diseases*. William Wood & Company; 1880.
8. Burnett C. *The Ear; Its Anatomy, Physiology, and Diseases. A Practical Treatise for the Use of Medical Students and Practitioners*. Henry C. Lea; 1877.
9. Knapp H, Prout J, Roosa D. The examination of the power of hearing, and how to record its results. *Trans Am Otol Soc* 1887; 20: 13–27.
10. Von Troltsch A. *Treatise on the Diseases of the Ear. Including the Anatomy of the Organ*. New York: William Wood and Company; 1869.
11. Winslow. William H. *The Human Ear and Its Diseases: A Practical Treatise upon the Examination, Recognition, and Treatment of Affections of the Ear and Associate Parts*. New York: Boericke & Tafel; 1882.
12. Barr T. *Transactions of the Ninth International Otological Congress (Boston, August 12–17, 1912)*. Boston, MA: Williams and Wilkins; 1912.
13. Politzer A, Ballin MJ. *Textbook of the Diseases of the Ear for Students and Practitioners*. London: Balliere, Tindall and Cox; 1926.
14. Bentley IM. Standard Tests of Audition. *Science* 1904;19:959–61.
15. Fiske G. The phonograph in testing hearing. *J Am Med Assoc* 1890;24:854–6.
16. Bing A. *Verlesungen über Ohrenheilikunde Braunmüller*. Wien: Verlag von W. Braunmuller; 1890.
17. Prout JS. A new and uniform method of recording the hearing power in otological practice: a new and uniform method of recording the hearing power in otological practice. 1872;9: 141–142.
18. Roosa D. *A Practical Treatise on Diseases of the Ear, Including a Sketch of Aural Anatomy and Physiology*. 6th ed. 1885.
19. Wilde WR. *Practical Observations on Aural Surgery of Diseases of the Ear*. Blanchard & Lea; 1853.
20. Kerrison P. *Diseases of the Ear*. JB Lippincott; 1922.
21. Itard J. *Taite Des Maladies Le l'oreille et de l'audition*. Meguignon Marvis; 1821.
22. Love J. *Diseases of the Ear: For Practitioners and Students of Medicine*. John Wright & Co; 1904.
23. Ng M, Jackler RK. Early history of tuning-fork tests. *Am J Otol* 1993;14:100–5.
24. Johnson EW. Tuning forks to audiometers and back again. *Laryngoscope* 1970;80:49–68.
25. Tschiassny K. Tuning forks; a historical review. *Ann Otol Rhinol Laryngol* 1946;55:423–30.
26. Gruber J. *A Textbook of the Diseases of the Ear*. D. Appleton and Co; 1890.
27. Hovell T. *A Treatise on the Diseases of the Ear*. Churchill; 1894.
28. Roosa D. The tuning fork in the differential diagnosis of diseases in the middle ear. 1881;15:464–470.
29. Bezold F, Siebenmann F. *Textbook of Otology (Lehrbuch Der Ohrenheikunde) (English Translation)*. EH Colegrove; 1908.
30. Roth A. The acoustimeter. 1934;20:575–575.
31. Andrews A. Sources of error in functional tests of hearing. 1902;12:249–253.
32. Clarke T. On hearing tests. 1929:361–376.
33. Bacon G. *A Manual of Otology*. Lea & Febiger; 1918.
34. Walker D. Aural and nasal examinations in school children. 1907;155:709–710.
35. Callahan JF. Hearing test to detect malingering. 1918;179: 236–239.

36. Hartmann A. Eine neune Methode der Hörprüfung mit Hülfe elektrischer Ströme. *Arch F Physiol* 1878;155.
37. Hughes DE. On an induction-currents balance and experimental researches made therewith. *Proc Roy Soc* 1879;29:56.
38. Bunch C. The development of the audiometer. 1941;51:1100–1118.
39. Keeler JC. *Modern Otology*. F.A. Davis Company; 1930.
40. Jerger J. Why the audiogram is upside-down. *Int J Audiol* 2013;52:146–50.
41. Wegel R. Physical examination of hearing and binaural aids for the deaf. 1922;8:155–160.