Sound-Conduction Effects on DPOAE Screening Outcomes in Newborn Infants: Test Performance of Wideband Acoustic Transfer Functions and 1-kHz Tympanometry

Chris A. Sanford, Ph.D.[1], **Douglas H. Keefe, Ph.D.**[2], **Yi-Wen Liu, Ph.D.**[1], **Denis Fitzpatrick, Ph.D.**[1], **Ryan W. McCreery, M.S.**[1], **Dawna E. Lewis, Ph.D.**[1], and **Michael P. Gorga, Ph.D.**[1]

[1]Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska

[2]Sonicom, Inc., Omaha, Nebraska

Abstract

Objective—Universal newborn hearing screening (UNHS) test outcomes can be influenced by conditions affecting the sound-conduction pathway including ear-canal and/or middle-ear function. The purpose of this study was to evaluate the test performance of wideband (WB) acoustic transfer functions (ATFs) and 1-kHz tympanometry in terms of their ability to predict the status of the sound-conduction pathway for ears that passed or were referred in a UNHS program.

Design—A distortion-product otoacoustic emission (DPOAE) test was used to determine the UNHS status of 455 infant ears (375 passed and 80 referred). WB and 1-kHz tests were performed immediately following the infant s first DPOAE test (Day 1). Of the 80 infants referred on Day 1, 67 infants were evaluated again following a second UNHS DPOAE test the next day (Day 2). WB data were acquired under ambient and tympanometric (pressurized) ear-canal conditions. Clinical decision theory analysis was used to assess the test performance of WB and 1-kHz tests in terms of their ability to classify ears that passed or referred, using DPOAE UNHS test outcomes as the "gold standard". Specifically, for 1-kHz tympanometry, performance was assessed using previously published measurement criteria and for WB measurements, performance was assessed using a maximum-likelihood procedure.

Results—For measurements from Day 1, the highest area under the receiver operating characteristic (AROC) curve was 0.87 for an ambient WB test predictor. The highest AROC among several variables derived from 1-kHz tympanometry was 0.75. In general, ears that passed the DPOAE UNHS test had higher energy absorbance compared to those that referred, indicating that infants who passed the DPOAE UNHS had a more acoustically efficient conductive pathway.

Conclusions—Results showed that: 1) WB tests had better performance in classifying UNHS DPOAE outcomes than 1-kHz tympanometry, 2) WB tests provide data to suggest that many UNHS referrals are a consequence of transient conditions affecting the sound-conduction pathway, 3) WB data reveal changes in sound conduction during the first 2 days of life, and 4) WB measurements used in the present study are objective and quick, making these tests feasible for potential use in conjunction with UNHS programs.

Keywords

Newborn hearing screening; sound conduction; ear canal; middle ear; wideband acoustic transfer function; energy reflectance; energy absorbance; tympanometry; neonate

INTRODUCTION

With the growth of universal newborn hearing screening (UNHS), infants are routinely being evaluated at birth and more infants with hearing loss are being identified at earlier ages than ever before. Current hearing-screening guidelines of the Joint Committee on Infant Hearing (JCIH, 2007) endorse hearing screening for permanent, sensory and conductive hearing loss, in addition to timely and appropriate intervention and follow up through the establishment and development of early hearing detection and intervention (EHDI) programs. While there are many components in a successful EHDI program, the initial goal is to correctly classify an infant s hearing status so timely and appropriate intervention can take place, if needed.

While continued improvements in hearing-screening technology and protocols have helped move toward this goal, false-positive results (a referral on the hearing screening test when hearing is normal) continue to be an issue for UNHS programs. False-positive results add costs to screening and follow-up programs, generate parental anxiety, undermine confidence in EHDI programs, and delay classification of true hearing status (Kennedy, 1999; JCIH, 2007). A significant number of UNHS referrals are reportedly due to transient ear-canal and or middle-ear obstruction (Thornton, et al., 1993; Doyle, et al., 1997; Keefe, et al., 2000), most likely caused by normally-occurring conditions, such as vernix occluding the ear canal and/or, residual amniotic fluid, effusion or mesenchyme in the middle-ear space (Takahara, et al., 1986; Kok, et al., 1992; Chang, et al., 1993; Rosenfeld, 2004; Doyle, et al., 2004). Recently, Gravel et al. (2005) recommended additional research to assess the usefulness to EHDI programs of implementing a screening for middle-ear function at both the newborn and outpatient rescreening stages. The goal of the present work is to assess the test performance of wideband (WB) acoustic transfer functions (ATFs) and 1-kHz tympanometry in terms of their ability to classify the sound-conduction status (i.e., the functional status of the ear canal and/or middle ear based on acoustic responses measured in the ear canal) of ears that passed or referred in a UNHS program, with the long-term goal of developing an effective measure of ear-canal/middle-ear function which might result in improvements in UNHS and EHDI programs.

Auditory brainstem response (ABR) and otoacoustic emission (OAE) tests, the primary tests used in UNHS programs (Finitzo, et al., 1998; Norton, et al., 2000), are affected by the condition of the sound-conduction pathway (Stuart et al., 1994). Doyle et al. (1997) reported that UNHS pass rates in ears with occluding vernix were 66% for ABR, and only 38% for OAE testing. Lower OAE pass rates reflect the fact that OAE measurements are especially sensitive to ear-canal and/or middle-ear obstructions, because both the eliciting stimulus and its evoked cochlear response must travel through the ear canal and middle ear. Therefore, failure to pass either of these UNHS tests may be due to ear-canal and or middle-ear dysfunction, leaving questions regarding neural and/or cochlear status unanswered.

Assessment of middle-ear function is not currently part of the JCIH guidelines for UNHS (JCIH, 2007). Validated pneumatic otoscopy and bone-conducted ABR are impractical in a UNHS protocol and the variability of 226-Hz probe-tone admittance tympanometry in infants has led to conflicting interpretations of what tympanometric criteria define a normal infant middle ear, casting doubt on the clinical utility of these measures for newborns (Keith, 1975; Paradise, et al., 1976; Sprague, et al., 1985; Hunter & Margolis, 1992). Variability in infant relative to adult tympanograms may be influenced by a more compliant ear-canal wall, smaller ear canal and middle-ear space, and a more horizontal orientation of the tympanic membrane with respect to the axis of the ear canal (Saunders, et al., 1983; Ruah, et al., 1991; Ikui, et al., 1997). These features, and their interactions with the effects of air-pressure changes during the tympanometry test, are likely contributors to age-

dependent differences in acoustic-admittance characteristics. Holte et al. (1990) reported ear-canal diameter changes in infants as large as 70% as a consequence of pressurization during tympanometry, and Holte et al. (1991) observed tympanometric differences between adults and infants depending on pressure-sweep direction. In addition, Sanford and Feeney (2008) reported age-related, frequency-dependent effects of air pressure on WB acoustic ear-canal measurements.

Tympanometry using higher probe-tone frequencies (up to and including 1-kHz) is more sensitive to changes in external/middle-ear status in infants compared to 0.226-kHz tympanometry (Alaerts, et al., 2007; Calandruccio, et al., 2006). Some studies have reported normative tympanometric data and test performance of specific 1-kHz admittance criteria in predicting OAE screening results (Kei et al., 2003). Margolis et al. (2003) assessed 1-kHz tympanometric test performance in predicting whether a newborn infant passed or failed a DPOAE test. With the DPOAE test outcome serving as the "gold standard", the tympanometric test had a specificity of 91% but a sensitivity of only 50%. Baldwin (2006) compared admittance tympanometry results at 0.226, 0.678, and 1 kHz between infants (mean age of 10 weeks) grouped as having either normal or disordered middle-ear function, based on a combination of air- and bone-conduction ABR results and behavioral assessments. The tympanograms from these infants were organized using a traditional visual classification scheme (Lidén, 1969; Jerger, 1970), and an alternative method proposed by Marchant et al. (1986). The Marchant et al. method for classifying 1-kHz tympanograms provided the best results, with sensitivity of 0.99 and specificity of 0.89. Baldwin cautioned that these findings may not apply to neonates, as the mean chronological ages for the "normal" and "pathological" groups were 10.2 and 11.2 weeks, respectively.

WB ATFs, which include energy reflectance (ER) and acoustic admittance, provide a spectral analysis of ear-canal and middle-ear transfer functions measured in the ear canal (Keefe & Feeney, 2009). ER is defined as the ratio of energy reflected at the probe location within the ear canal to the incident energy delivered by the probe (see Appendix A for a list of frequently used abbreviations). If negligible sound energy is absorbed within the ear canal, then this ER equals the ratio of the energy reflected from the middle ear to the incident energy. ER ranges from 1, meaning all of the energy is reflected, to 0, meaning all of the energy is absorbed. The ER values obtained at ambient ear-canal pressure in normal-functioning adult ears are typically near 1.0 in the low frequencies, gradually decreasing to a minimum around 4 kHz, then increasing at higher frequencies (Stinson, 1990). Measurements in adult human ears of WB ATF and ER have been performed using systems with multiple microphones and/or multiple microphone locations in the ear canal to locations as close as 1–2 mm from the eardrum (Hudde, 1983; Lawton and Stinson, 1986; Stinson, 1990). These systems involved time-consuming measurements in which metal probe tubes were placed close to the eardrum, and thus were not intended for routine clinical use; furthermore, these systems might not be applicable in neonates, even under a research protocol. To overcome these limitations, a system to measure WB reflectance and other ATFs in a few seconds was developed that used a clinical insert probe at a single location within the ear canal (Keefe et al., 1992). This system used a calibration method developed to measure WB impedance in cat ears (Allen, 1986).

ER and other ATFs have been reported for normal-hearing adults (Keefe, et al., 1993; Voss & Allen, 1994; Feeney & Sanford, 2004; Shahnaz & Bork, 2006), children and infants (Keefe, et al., 1993; Vander Werff, et al., 2007; Hunter, Tubaugh, et al., 2008; Sanford & Feeney, 2008), healthy newborns (Keefe, et al., 2000; Abdala, et al., 2007; Keefe & Abdala, 2007) and newborns in a neonatal intensive care unit (Shahnaz, 2008). ER measurements are sensitive to middle-ear disorders including otitis media with effusion (Piskorski, et al., 1999) and otitis media in children with cleft palate (Hunter, Bagger-Sjöbäck et al., 2008); Feeney

et al. (2003) reported ER results for otitis media, otosclerosis, ossicular discontinuity and perforation of the tympanic membrane.

While WB ATF measurements have mainly been performed at ambient ear-canal pressure, WB ATFs have also been measured as a function of ear-canal air pressure, providing WB tympanometric data as a function of frequency *and* pressure (Keefe & Levi, 1996; Margolis, et al., 1999; Sanford & Feeney, 2008). Keefe and Simmons (2003) reported that WB tympanometry was more accurate in detecting a conductive hearing loss than ambient-pressure WB ATFs, and both WB ATF tests were more accurate than 226-Hz tympanometry. A WB system to measure ATFs at ambient pressure and using a tympanometry technique was described by Liu et al. (2008) and used to obtain data from normal-hearing adults. This WB tympanometry system, which is used in the present study, includes: 1) implementation of a pressure-sweep method, reducing the time to measure a tympanogram in an adult ear to less than 8 seconds, 2) automatic monitoring and control of air pressure in the ear canal, and 3) smoothing the WB tympanogram and re-sampling to a uniform grid of pressure and frequency.

Because WB ATFs provide information about middle-ear function in infants, they could potentially be a useful tool in UNHS and EHDI programs. For example, Keefe et al. (2003a) analyzed a subset of data obtained from a two-stage hearing-screening protocol, e.g., click-evoked (CE) OAE/ABR that produced a 5% false-positive rate. They reported significant correlations between two ER factors and CEOAE responses, suggesting that ER measures could be useful in interpreting CEOAE responses. Applying ER factors identified in Keefe et al. (2003a) to a predictive test of middle-ear dysfunction on the same subset of infant data, Keefe et al. (2003b) showed that inclusion of the ER test decreased the false-positive rate from 5% to 1% (based on knowledge of true hearing status assessed via behavioral audiometry obtained when the infants were between 9 and 13 months of age). They found similar results using a UNHS protocol based on distortion-product (DP) OAE and ABR tests. In addition, Vander Werff et al. (2007) reported that infants who did not pass an OAE re-screening had significantly higher ER from 0.63 to 2 kHz compared to infants who passed the OAE re-screening, suggesting the presence of some type of conductive dysfunction which may have contributed to the failure to pass the OAE screening.

This growing body of research suggests that WB ATF measurements may provide a test of the sound-conduction pathway in neonates, which may prove useful for interpreting ABR and OAE screening results and help improve the management and follow-up care of infants who do not pass their UNHS test. However, such a test would need to be objective if it is to be of potential use in UNHS programs due to the use of untrained screening personnel in many settings. The primary goal of the present study was to evaluate several types of WB ATF and 1-kHz admittance tympanometry tests in terms of their ability to predict the status of the sound-conduction pathway in infants less than 3 days old. A preliminary assessment of differences in WB ATF responses for infants who initially were referred based on a UNHS test then subsequently passed the following day was also performed to evaluate the extent to which changes in the conduction pathway may have influenced acoustic energy transfer in the first few days of life.

The term "sound-conduction pathway" is used in the present report for simplicity when referring to both ear-canal and middle-ear components. Determining a specific cause (e.g., ear-canal vs. middle-ear condition) of DPOAE refer outcomes was beyond the scope of the present study.

MATERIALS AND METHODS

Subjects and Testing Environment

The study enrolled 230 well-baby infants (100 female, 130 male). Institutional Review Board (IRB) approval was received prior to obtaining consent and gathering data. Following birth, informed consent was obtained from parents during their hospital stay. Due to hospital guidelines, consenting and data collection were limited to specific times of day. Therefore, not all infants born at the hospital were available for recruitment and the 230 infants enrolled represent a sampling of the well-baby population over a period of approximately 3 months. However, it is reasonable to assume that the sample whose data are summarized in this paper is representative of the larger sample of well babies. All measurements were performed in a procedure room adjacent to the well-baby nursery and in conjunction with the ongoing UNHS screening program at the hospital. UNHS and experimental (WB ATF and 1-kHz tympanometry) measurements were obtained from both ears of each infant. Clinical and research audiologists performed the UNHS and experimental tests, respectively, within a fixed two-hour block of time each day. 1-kHz and WB data were successfully acquired in all infants; in a cooperative infant, test time to acquire these data was less than one minute per ear. A total of 455 ears from the 230 infants were included in data analyses following exclusion of 5 ears due to data-file errors in the device used to measure 1-kHz tympanograms. Although not considered in data analyses, the composition of ethnicity and race for participants was obtained from parents in a format similar to that used by the National Institutes of Health (NIH), and the numbers of subjects were as follows: Hispanic or Latino (27), not Hispanic or Latino (193), not reported (10); American Indian or Native Alaskan (2), Black or African American (20), White (197), more than one race (1) and not reported (10).

DPOAE Screening Measurements and Subject Classification

DPOAE tests were performed using a Biologic AuDx Plus device and associated neonatal ear tips. DPOAE levels were measured for f_2 frequencies of 2, 3, 4 and 6 kHz with a frequency ratio (f_2/f_1) of approximately 1.22, and with fixed levels of 65 and 55 dB SPL for L_1 and L_2, respectively. A pass occurred if DPOAE criteria were met at 3 of the 4 f_2 frequencies. The pass criteria included a signal-to-noise ratio (SNR) of at least 6 dB and a DPOAE level of at least −6 dB SPL. Any ear not passing the DPOAE test was classified as a refer.

The UNHS consisted of an initial DPOAE test on all infants, in what is termed the Day-1 test. These infants also received the experimental tests immediately following the Day-1 DPOAE test. As many infants as possible of those who referred on either ear from the Day-1 DPOAE test were re-screened 24 hours later on Day 2 using the same DPOAE test protocol that was used on Day 1. The experimental protocol was also repeated in these Day-1 refer infants. Infants passing the DPOAE test on either Day-1 (DP1-Pass) or Day-2 (DP2-Pass) were classified as passing the UNHS. Infants who were referred based on the Day-2 DPOAE test (DP2-Refer), and those infants who were referred on the Day-1 DPOAE (DP1-Refer) test and left the hospital prior to further rescreening, were classified as refers on the UNHS. The study was designed to test a sufficient number of the Day-1 infants to determine how well the experimental tests classified ears, using the DPOAE test outcomes (pass or refer) as the "gold standard". In addition, analyses were performed on Day-2 data to identify trends in the test results of middle-ear assessment that might guide future research.

While a true gold standard for auditory conductive dysfunction might include behavioral measures, it was not possible to obtain reliable behavioral measurements in the newborn infants. Similarly, pneumo-otoscopic examinations were not possible because of the

unavailability of trained pediatric otolaryngologists at the time of data collection. DPOAE tests are in common use in UNHS programs, and the goal of the present study was to relate outcomes on screening measures to measures that describe the sound-conduction pathway. It is for these reasons that the results of the DPOAE screening were treated as the gold standard, despite its limitations as a test of middle-ear status.

The DP1-Pass group (N=375 ears) ranged in age from 9 to 58 hours with a mean age of 25.5 hours (SD=8.0). The DP1-Refer group (N=80 ears) ranged in age from 8 to 47 hours with a mean age of 22.2 hours (SD=8.0). Because of hospital policies regarding the time of day during which babies could be screened, some DP-1 refer infants were discharged prior to re-screening. Thirteen infants were discharged from the hospital before Day 2 tests could be completed and their eventual screening results were unknown. However, a subset of DP1-Refer ears (N=67) were retested on Day 2. The Day 2 pass (DP2-Pass) group (N=53 ears) ranged in age from 32 to 70 hours with a mean age of 45.3 hours (SD=7.1). The infants who did not pass the Day-2 screening (DP2-Refer; N=14 ears) ranged in age from 35 to 60 hours with a mean age of 45.7 hours (SD=7.5).

1-kHz Tympanometry Measurements

A Madsen Otoflex 100 middle-ear analyzer, with standard ear tips, was used to obtain 1-kHz admittance tympanograms. With the probe in place, a pressure sweep was initiated, sweeping from +200 to −400 daPa at a rate of 500 daPa/sec. At least two tympanograms were obtained from each ear. 1-kHz tympanometric data were saved in the Otoflex device, then exported and stored for future analysis. The presentation order of 1-kHz tympanometric and experimental WB measurements (described below) alternated with each subject to control for test-order effects, but no such effects were observed.

WB Measurements

Below is a general overview of WB ATF responses considered in the present study, the prototype research system, and procedures used to measure WB ATF responses, with a more detailed description given in Liu et al. (2008).

The prototype system consisted of a Windows-based computer, a soundcard (CardDeluxe, Digital Audio Labs) with 24-bit resolution, a pressure pump and controller system contained in an acoustic immittance instrument (modified by the manufacturer, Interacoustics, from its model AT235 device), and custom software in control of stimulus generation and data acquisition. A prototype probe assembly with two loudspeakers and one microphone (developed specifically for this project by Interacoustics) was used to present the stimuli to the infant s ear. Plastic tips were placed over the end of the probe, both for calibration purposes, and to provide acoustic and hermetic pressure seals once placed in the ear canal.

A calibration procedure was performed daily prior to data collection (see Keefe and Simmons, 2003 for a detailed discussion of calibration). The calibration procedure determined the source reflectance and incident sound pressure associated with the probe and its transducers based on acoustic measurements in two rigid-walled cylindrical calibration tubes (lengths 236.5 and 5.9 cm). Each tube was closed at one end and the probe was inserted into the opposite end. The inside diameter (0.476 cm) of each tube was smaller than that of an average adult ear canal (0.8 cm) although larger than that of an infant ear canal (0.3 cm) (Keefe and Abdala, 2007). The time to complete the calibration procedure was less than one minute. A calibration was judged acceptable on each day as long as the root-mean-squared reflectance error ΔR did not exceed 0.009 and the loss parameter χ was in the range from 1 to 1.09; definitions and properties of these variables are described in Keefe and Simmons. Any calibration that was not acceptable on any test date was repeated. Of the 78

days on which data were collected, the initial calibration of the day was acceptable on 77 days, while the calibration was repeated and subsequently judged acceptable on 1 day.

WB ATF measurements were obtained under both ambient and tympanometric conditions. For convenience in discussing these measurements, the letters a and t are used as prefixes to denote the tests conditions. (e.g., aEA and tEA correspond to ambient and tympanometric energy absorbance, respectively). The emphasis on EA in the present study, rather than energy reflectance (ER), was based on two properties: 1) an EA response typically has a single-peaked function, as in conventional admittance tympanometry, and 2) the frequencies at which the maxima of EA occur are frequencies at which the middle ear is most efficient in absorbing sound energy (Keefe & Simmons, 2003). In addition to aEA and tEA, WB ATFs with properties including acoustic admittance magnitude ($|Y|$) and phase ($Y\phi$) were also derived. Because each ATF response provides unique information, it is useful to examine multiple ATF variables. Therefore, WB ATF admittance and phase are reported for both ambient ($|aY|$ and $aY\phi$) and tympanometric ($|tY|$ and $tY\phi$) measurements.

Ambient WB ATF measurements were obtained by recording the acoustic response to clicks, presented at a rate of one click per 1024 samples (approximately every 46 ms). During testing, the operator was alerted via an on-screen prompt if the noise level was too high or if a leaky probe fit was likely based on the measured response. In previous measurements of infant ears, the signature of a leaky probe fit was a mass-like and resistive impedance at low frequencies, as would be associated with acoustic flow through a narrow aperture (Keefe et al., 2000). When such a leak occurred in infant ears, the equivalent volume was excessively negative. Keefe et al. proposed a criterion for detecting a leak might be based on an equivalent volume more negative than -1.15 cm^3 at low frequencies. The current system classified a probe fit as leaky if either the average equivalent volume was less than -1.15 cm^3 for frequencies between 0.5–1 kHz, or the minimum equivalent volume was less than -2.3 cm^3 at any 1/12th octave frequency between 0.5–1 kHz. If the probe fit was classified as leaky, the system prompted the operator to remove and re-seat the probe. Following artifact rejection (see Liu, et al., 2008), responses to a total of 16 clicks were averaged for each measurement, and WB ATFs were calculated for each individual response (Keefe & Simmons, 2003). Typical test time for data acquisition at ambient pressure was approximately 1 second.

Tympanometric WB ATF measurements, obtained using a swept-pressure paradigm, were typically measured with the same probe insertion as used in the ambient WB measurements. In the default descending sweep, pressure was swept from +220 daPa to −315 daPa at a nominal rate of 75 daPa/sec. Because a secondary aim of the present study was to assess the effects of pressure-sweep direction, an ascending (negative to positive) pressure sweep at the same rate and for the same pressure range was recorded in a subset of subjects (N=56). The acoustic response to a train of clicks, with an inter-click interval of 46 ms, was measured during the pressure sweep. At a sweep speed of 75 daPa/sec, the pressure changed approximately 3.5 daPa over this 46-ms interval. Approximately 180 responses were thus acquired across a pressure range of approximately 500 daPa. In addition to the noise and acoustic "leak" detection alerts discussed previously, prompts also alerted the operator if the pressure seal was compromised during tympanometric WB ATF measurements. An automatic artifact-rejection algorithm was used to analyze each click response. Responses were considered artifact if, during a 9-ms period encompassing each response, the sum of the pressure squared exceeded the 75th percentile + 3 times the inter-quartile range (IQR). If more than 10% of the responses were excluded based on this criterion, the operator was notified via an on-screen prompt and given the opportunity to save the data or acquire new data. Responses identified as artifact were automatically excluded from further analysis. Tympanometric WB ATFs were calculated for each individual response, just as for ambient

measurements. Typical test time to measure a tympanometric WB ATF was less than 10 seconds.

1-kHz Tympanometry Analyses

The 1-kHz tympanometric data provided by the Otoflex device included acoustic susceptance (B) and acoustic conductance (G) as a function of pressure. The pressure resolution was as fine as 1 daPa for pressure changes of 0 to ±75 daPa, and gradually increased to 7 daPa for pressures of $+200$ and -400 daPa. Custom software was used to derive the non-compensated admittance magnitude for these data, $\sqrt{G^2+B^2}$, and to display the resulting admittance magnitude ($|Y|$) tympanogram for analyses.

Tympanograms were then examined by two of the investigators who were blind to the DPOAE result associated with each tympanogram. Because more than one tympanogram was obtained from each ear, a single "best tympanogram", defined as the one with the lowest noise and least artifact, was chosen. Three 1-kHz test predictors, representing different ways of quantifying the admittance magnitude, were defined as in previous studies (Baldwin, 2006; Kei, et al., 2003; Margolis, et al., 2003). It should be noted that the 1-kHz tympanometer used in the present study was a different model from that used in previous studies. Differences in the 1-kHz tympanometry measurements may result from differences in manufacturer model, pump speed and algorithms used in different tympanometers to assure an adequate probe seal.

Software was created to automatically identify several points on each tympanogram that were used to construct the three test predictors. These points included 1) the largest value of admittance magnitude on the tympanogram tracing, or tympanometric peak pressure ($|Y|_{TPP}$), 2) the admittance magnitude at the most extreme positive pressure ($|Y|_{+}$), and 3) the admittance magnitude at the most extreme negative pressure ($|Y|_{-}$). A straight line connecting $|Y|_{+}$ and $|Y|_{-}$ was displayed (see Fig. 1) and termed the "pressure-slope line". In some cases, due to subject noise or movement during data collection, full pressure sweeps (extending to -400 daPa) were not completed. In these cases, $|Y|_{-}$ values were taken at the most negative pressure.

Tympanogram peak to positive and negative "tail" differences (the first two predictors) were defined as $\Delta Y_{Pos} = (|Y|_{TPP} - |Y|_{+})$, and $\Delta Y_{Neg} = (|Y|_{TPP} - |Y|_{-})$, respectively, similar to calculations used by others (Kei, et al., 2003; Margolis, et al., 2003). Calculation of the third 1-kHz predictor (ΔY_{Slope}) was performed by finding the admittance difference of $|Y|_{TPP}$ relative to a point on the pressure-slope line (after Baldwin, 2006); hence the term of ΔY_{Slope} (see example in Fig. 1). For a given tympanogram, where the peak of the tracing fell above the pressure-slope line (as shown in Fig. 1), the maximum deviation ($|Y|_{TPP}$) would be a positive value. However, consistent with the work of Baldwin, the point of maximum admittance deviation was also allowed to vary in a negative direction, (negative relative to the pressure-slope line). Hence, the $|Y|_{TPP}$ value used to calculate ΔY_{Slope} was not always positive.

Post-hoc review of each tympanogram was performed to adjust the points that were picked automatically, if necessary, in the event that points were erroneously selected by the software (e.g., a noise spike identified as a peak). With the exception of the post-hoc adjustments, all points from the admittance-magnitude tympanogram used to derive the 1-kHz tympanometric test variables were selected by the algorithm.

Ambient WB Analyses

An aEA response consisted of 60 data points (EA at $1/12^{th}$ octave frequencies from 0.250 to 8 kHz). The |aY| and aYϕ responses had similar representations to aEA over frequency. A single test measurement per subject, per test type (e.g., aEA, |aY| and aYϕ) was used for data analyses. To be suitable for clinical decision theory analyses, each aEA response was reduced to a single output variable. The statistical methods of likelihood estimation (Myung, 2003; Van Trees, 1967) were used to: 1) produce likelihood estimates that the selected ear response was from the pass or refer group, and 2) calculate the ratio of the likelihood that the test ear was selected from the refer group to the likelihood that the response was selected from the pass group. As further described in Appendix B, this procedure reduced the multivariate WB response to a univariate predictor in the form of a likelihood ratio. Such a likelihood ratio was calculated for each of the WB aEA, |aY| and aYϕ responses.

WB Tympanometry Analyses

Each tEA response was measured as a function of frequency over the same 60 frequencies as aEA and as a function of air pressure. The complete pressure-sampling grid included 101 pressures, spaced in 5-daPa increments, so that tEA was fully specified over 6060 unique pairs of frequencies and pressures. In an effort to identify the most useful frequency/pressure pairs, a strategy was implemented to investigate test performance when using all or a subset of these 6060 points. This method is described in the Day-1 Test Results section below. Whether the analysis of tEA was over the full range of the frequency-pressure set or some subset of values, a likelihood ratio for tEA was constructed as described in Appendix B. Likelihood ratios for |tY| and tYϕ were similarly constructed. This approach reduced the WB tympanometric response with as many as 6060 values to a univariate predictor. A single test measurement per subject, per test type (e.g., tEA, |tY| and tYϕ) was used for data analyses.

Because deviations in TPP may be related to deviations in the energy absorption properties of the ear canal and middle ear, it may be advantageous to measure ATFs at TPP rather than at ambient ear-canal pressure (Margolis et al., 1999). For this reason, it was of interest to compare estimates of TPP from 1-kHz tympanometry and from WB tEA. The WB TPP estimate consisted of frequency-averaged tEA calculated from 0.8 – 2 kHz. In contrast, Liu et al. (2008) defined a WB TPP in adult ears by averaging over frequencies from 0.38 to 2 kHz. Liu et al. suggested a lower limit of 0.38 kHz in adults because of concerns that noise might contaminate spectral responses below that frequency, and the upper limit is consistent with the assumption that effects related to eardrum mobility occur below 2 kHz (Puria, 2003). As shown in the Results section, tEA was similar in infant ears below 0.8 kHz that passed and referred on the UNHS exam, making measurements at these frequencies of little diagnostic value. As a result, the lowest frequency for averaging tEA in the present study was set to 0.8 kHz.

Measures of Test Performance

Clinical decision theory, based on the analysis of a receiver operating characteristic (ROC) curve, is a standard procedure to evaluate the accuracy of a univariate diagnostic test in which a dichotomous decision is made (Swets, 1988). A ROC curve is a plot of test sensitivity versus one minus test specificity (false-alarm rate). Overall test accuracy was assessed using the area under the ROC curve (AROC), which was calculated non-parametrically (Bamber, 1975). The point at which the sensitivity and specificity are most nearly equal, referred to as the point of symmetry (SYM) of the ROC curve (Pepe, 2003), is an alternative measure of performance with clinically relevant properties. Although not identical, SYM and AROC share the properties that values of 0.50 represent chance performance, and a larger SYM or AROC represents a more accurate test than another test with significantly smaller SYM or AROC, respectively.

One disadvantage of AROC is that AROC equally weights all points on the ROC curve in its calculation, including combinations of sensitivity and specificity that would be inappropriate in a screening test (e.g., choosing a refer criterion that successfully refers all impaired ears but with the unfortunate property of also referring most of the normal ears). In contrast, SYM lies within a range of sensitivity and specificity points of the ROC curve that might serve as plausible test criteria to consider, i.e., a plausible test would have acceptably low false-positive and false-negative rates. A second advantage of SYM compared to AROC is that SYM has a clinically relevant interpretation while AROC does not. AROC is equal to the probability that test results from a randomly selected pair of Refer and Pass results are correctly classified (Pepe, 2003). No clinician is ever asked to correctly classify a pair of ear tests, one selected from the Pass group and the other from the Refer group, but rather the clinician must judge the results on a test from a single subject. SYM summarizes test accuracy for the clinically meaningful case that high sensitivity and high specificity are equally desirable.

When used as a particular test criterion rather than as a summary measure of the ROC curve, a potential limitation of SYM is that it implicitly assumes that cost-benefit ratios for identifying normal and impaired cases are similar, which might not be the case in some clinical circumstances. In fact, for a given clinical situation (e.g., one in which the prevalence of hearing loss is high), a specific test criterion may be selected to minimize one error (such as the false-negative rate), knowing that the other error (false-positive rate) might be high. In the present study, both AROC and SYM were used to summarize test performance across all possible test criteria, recognizing that each measure has advantages and limitations.

The accuracy of 1-kHz tympanometry and WB tests was evaluated using the pass/refer status from the DPOAE UNHS tests on Day 1 and Day 2 as the "gold standard". Admittedly, DPOAE test outcomes are not ideal indicators of the status of the sound-transmission system; however, in the context of newborn screening, in which tests are performed within the first day or two of life, more direct gold-standard tests (comparisons of air and bone conduction thresholds or pneumo-otoscopic examinations) were not feasible. Furthermore, the purpose of the present study was to relate measures of sound transmission to outcomes on a newborn screening test. For these reasons, DPOAE test outcome was chosen as the "gold standard". Each of the three 1-kHz tympanometry test variables was univariate and each of the WB ambient and tympanometric tests provided a univariate output in the form of the likelihood ratio[i] (see Appendix B). Thus, each experimental test provided an output suitable for subsequent ROC analyses. To estimate the uncertainties in measuring AROC and SYM for each test, confidence intervals (CIs) were calculated non-parametrically using a bootstrap method. Specifically, 95% CIs for each test were estimated with 10,000 random samples with replacement of the test output (Efron & Tibshirani, 1986). This random sampling was performed on a per-ear basis, so that the CIs obtained for each test incorporated within-ear correlations across the test outputs.

The WB test outputs for the Day-1 groups were based on the likelihood ratios calculated for the Day-1 data sets, as defined in Eqs. (B1) and (B2). For the Day-2 groups, the likelihood-ratio predictor was developed on the Day-1 data set, which had a larger number of ears, and its performance was evaluated on the Day-2 data for each ear. This approach was implemented by using the mean and SD of the Day-1 pass and refer groups, as expressed in

---

[i]In particular, the rank order of each of the 1-kHz test variables was identical to the rank order of the likelihood ratio of the each of the 1-kHz test variables, so that the ROC curves and thus any of their summary measures (AROC or SYM) were identical whether the original test variable was used or its likelihood ratio.

Eq. (B1) of Appendix B, and then numerically evaluating the likelihood ratio using the measured response on Day 2, which is expressed as *x* in Eq. (B1).

RESULTS

Day-1 Test Results

The distributions of 1-kHz tympanometric variables (ΔY_{Pos}, ΔY_{Neg}, ΔY_{Slope}) for Day 1 are plotted in Figure 2A as box and whisker plots. The lower and upper horizontal lines of each box represent the 25th and 75th percentiles of responses that form the inter-quartile range (IQR), and the line within each box represents the 50th percentile (median). Upper and lower whiskers represent the smaller of either 1.5 x IQR or the upper or lower limit of the data, respectively (Tukey, 1977). Both ΔY_{Pos} and ΔY_{Neg} are non-negative, whereas ΔY_{Slope} was allowed to be negative, based on its definition described above. The ΔY_{Pos} predictors have the smallest IQRs, but the IQRs from pass and refer groups overlap for all three variables. The medians in the pass group for ΔY_{Pos} and ΔY_{Neg} (0.5 and 1.5 mmho, respectively) were less than those reported by Margolis et al. (2003) for 2–4 week old infants (1.0 and 1.7 mmho, respectively). The median ΔY_{Pos} was also less than the 1.1 mmho value reported by Kei et al. (2003) for newborn infants with a mean age of 3.29 days. Differences between the present study and previous work may be related to the type of equipment used or age differences in the sample, particularly so for the results in older infants reported by Margolis et al. (2003).

Results for Day-1 measurements of ambient WB responses as a function of frequency are displayed in Figure 3A, B, and C. In each panel, light and dark shaded areas represent IQRs for DP1-Pass and DP1-Refer groups, respectively, with data for a different ambient WB variable shown in each panel. The solid and dashed black lines represent the median of the DP1-Pass and DP1-Refer groups, respectively. DP1-Refer ears had lower aEA than DP1-Pass ears, with the best separation at 1.4–2.5 kHz (Fig. 3A). A similar frequency region of best separation was observed by Margolis et al. (2001) for differences in adult ER with changes in ear-canal pressure. The present results indicate that ears that passed the DPOAE-based UNHS test had proportionally more sound energy absorbed than did ears that were referred. The |aY| was larger for DP1-Pass ears than DP1-Refer ears, similar to aEA responses, with the best separation from 1–2 kHz (Fig. 3B). Contrary to the aEA results, there was more overlap between the IQRs of |aY| for frequencies above 4 kHz. The IQRs of aYφ had the most separation in two frequency regions (Fig. 3C). The aYφ was smaller in DP1-Refer group for frequencies from 0.75–1 kHz, and larger from 2.5–4.5 kHz.

Tympanometric WB measurements are summarized in Figure 4, which shows the median tEA for DP1-Pass (top) and DP1-Refer (bottom) groups plotted as two-dimensional functions of frequency and ear-canal pressure. tEA close to 1 is coded as red and tEA close to 0 is coded as blue. Overall, the median tEA for the DP1-Pass group exhibited higher EA (more yellow and red), relative to the DP1-Refer group, especially between 1 and 8 kHz. These findings indicated that infants who passed the DPOAE test had proportionally more sound energy absorbed than infants who did not pass the DPOAE test. To help interpret these measurements, it may be useful to think of a two-dimensional tEA response as a series of tympanograms obtained at discrete frequencies from 0.25 to 8 kHz. It would be possible, then, to extract from the WB tEA response in Fig. 4 a tEA defined over a sub-range of frequencies or at an individual frequency. The tEA at frequencies below 0.5 kHz exhibits a double-peaked pattern as a function of pressure for both Day-1 pass and refer groups, with peaks occurring just above and below approximately 0 daPa. These responses are consistent with the findings of Paradise et al. (1976) who reported multi-peaked 0.226-kHz tympanograms in young infants.

Preliminary analyses of tEA data revealed that 198 of 375 (53%) of DP1-Pass ears and 53 of 80 (66%) of DP1-Refer ears had more than 60 instances of negative tEA (60 ≈ 1% of the 6060 frequency-pressure points). The majority of these negative tEA responses occurred between 2 and 8 kHz for pressures between −5 and −300 daPa. Negative EA is unrealistic (i.e., in a passive system, the ratio of reflected to incident energy cannot exceed 1, so that a true EA value cannot be less than 0) and is considered evidence of measurement error. Such errors may have had a higher prevalence in infants due to maturational effects in the circumstances in which such artifact occurred (e.g., due to closure of the ear canal at negative pressures). Therefore, before likelihood ratios and test performance results were calculated for tEA measurements, negative tEA values (which are physically unrealistic) were all set to zero. It is unknown whether this modification to tEA introduced any bias to the ROC results, or whether the modification was equally likely to occur in both normal and refer ears.

In light of the presence of some negative tEA values and in an effort to investigate the potential for improvement in test performance using fewer tEA data points, an additional data-analysis strategy was implemented. To this end, two-dimensional tEA data were reduced to fewer frequency-pressure points to evaluate whether the test accuracy was improved relative to using all the data. To identify a region of the tEA frequency-pressure plane that would potentially be most useful in calculating this alternative WB predictor, the frequency-pressure points for tEA, |tY|, and tYϕ (dots, circles, and triangles, respectively, in Fig. 5) where the IQR of the DP1 Pass group did not overlap with the IQR of the DP1 Refer group were identified[ii]. Figure 5 identifies where, in the frequency-pressure plane, such separation between tympanometric WB responses for pass and refer groups occurred for each predictor. Such a subset of the frequency-pressure points might have importance for classifying an ear as pass or refer. With few exceptions, these points fell inside the region bounded by −100 to 30 daPa and 0.6 to 3 kHz (shaded area, Fig. 5). Data from frequency-pressure points in the entire shaded region, including those that are not marked with dots, circles or triangles in Fig. 5, were used to calculate an alternative set of univariate predictors equal to the likelihood ratio defined over the restricted range of pressure and frequency.

Pressure-sweep direction effects—The WB tympanometry results described in Day-1 test results were obtained using a descending pressure sweep. WB tympanometric data were also obtained using an ascending pressure sweep in a subset (*N*=56) of DP1-Pass ears. A comparison of tEA medians (Fig. 6) shows different tEA patterns depending on the direction of the pressure sweep. The median EA was smaller in the ascending sweep (Fig.6, top), especially at negative pressures. The descending-sweep median (Fig. 6, bottom) exhibits a double-peaked EA function below 0.5 kHz, whereas the ascending-sweep median revealed a single-peaked function in the lower frequencies. In the ascending-sweep median tEA, the slope of EA was steeper at pressures just below the TPP than in the descending-sweep response.

Comparison of 1-kHz and WB TPPs—The TPPs derived from 1-kHz admittance tympanometry and WB tympanometry tests using descending pressure sweeps were compared for Day-1 infants. Figure 7A and 7B shows distributions of WB TPP for DP1-Pass and DP1-Refer infants, respectively. The DP1-Pass distribution of WB TPP was Gaussian shaped, with a maximum at the 0–25 daPa interval, which, as expected, is close to ambient pressure. The mean ±1 SD of WB TPP for the pass group was 21 ±75 daPa. The DP1-Refer distribution of WB TPP was skewed towards positive pressures, with a mean ±1

[ii]In other words, either the 75th percentile of the DP1 Refer group was lower than the 25th percentile of the DP1 Pass group, or the 75th percentile of the DP1 Pass group was lower than the 25th percentile of the DP1 Refer group.

SD of WB TPP of 95 ±96 daPa. Figure 7C and 7D shows the distributions of 1-kHz TPP for DP1-Pass and DP1-Refer infants, respectively. The DP1-Pass distribution of 1-kHz TPP was similar to WB TPP with the exception of an increased proportion at the positive pressure extreme, which was also observed for the DP1-Refer distribution of 1-kHz TPP. The mean ±1 SD of 1-kHz TPP was 41 ±97 daPa for the DP1-Pass group, and 124 ±90 daPa for the DP1-Refer group.

Day-1 Test Performance

Examination of the data shown in Figs. 2 and 3 reveals overlap in the pass and refer distributions for both 1-kHz and WB measurements, respectively, but does not characterize the relative test performance. To provide this information, AROC and SYM were used to assess the test performance of each 1-kHz and WB test variable.

AROC and SYM for 1-kHz tympanometry tests (ΔY_{Pos}, ΔY_{Neg}, ΔY_{Slope}) on Day 1 are shown in the upper portion of the left column of Table I along with their 95% CIs. Each test exceeded chance performance because the lower bound of the 95% CI was larger than 0.5. ΔY_{Pos} had the highest AROC (0.75) and a SYM of 0.69 (i.e., ΔY_{Pos} correctly classified 69% of both pass and refer ears relative to the UNHS DPOAE test). Based on the overlap in their CIs, there were no significant differences in AROC or SYM between ΔY_{Pos}, ΔY_{Neg}, and ΔY_{Slope}. That is, the test performance for the 1-kHz tympanometric variables did not differ.

AROC, SYM and 95% CIs were also calculated for WB tests, aEA, |aY|, and aYɸ on Day 1, with results listed in the middle section of the left column of Table I. The test variable aYɸ had the largest AROC (0.87) and SYM (0.78). Like the results from the 1-kHz tympanometric tests, AROC and SYM exceeded chance performance. However, neither AROC nor SYM significantly varied with the type of ambient WB test.

AROC and SYM were also evaluated for WB tympanogram ATFs using all 6060 pressure-frequency points. AROC, SYM, and 95% CIs for tEA, |tY|, tYɸ on Day 1 are shown in the lower portion of the left column of Table I. AROC and SYM did not significantly vary with the type of the WB tympanogram test, with tYɸ having the best AROC (0.85). There was little difference in test performance between WB tests performed at ambient pressure or during pressure sweeps. Notably, both AROC and SYM for all WB tests were higher than any of the estimates of test performance for 1-kHz tympanometry regardless of whether they were conducted at ambient pressure or when pressure was varied.

For the WB tympanograms, AROC and SYM were also derived for the measured likelihood ratios that were calculated based on the subset of tEA frequency-pressure points defined by the separation in IQRs (as described above and shown in Fig. 5). Test performance (not shown) using the subset of tEA points was similar to performance when using all 6060 tEA frequency-pressure points (differences in AROC and SYM between tests were no greater than 0.03 and 0.04, respectively). Thus, there was no advantage to restricting the range of frequency-pressure points prior to calculating the univariate predictors.

The accuracies of WB ATFs and 1-kHz tympanometry were compared for those tests with the largest AROC for each test type (i.e., aYɸ for the WB ambient test, tYɸ for the WB tympanometric test, and ΔY_{Pos} for the 1-kHz test). Figure 8 shows the AROC and SYM results for these three conditions with error bars representing the 95% CIs. AROCs for both the WB ambient and tympanometric tests were significantly larger than AROCs for any of the 1-kHz tympanometric predictors. This indicates that WB ATF tests were more accurate at predicting when sound conduction influenced UNHS outcomes, compared to 1-kHz tympanometric tests. The SYMs for WB ATFs ranged from 0.75–0.78, which suggests that a

WB test may be able to identify ears with reduced sound conduction in a DPOAE-based UNHS program with sensitivity and specificity as high as 78%.

Day-2 Test Results

The main goal of the present study was to analyze test performance of Day-1 responses, in which the number of ears tested was sufficient to evaluate relative differences between WB measures and those associated with clinical techniques in which 1-kHz tympanometry was used. Nevertheless, analyses were also performed using Day-2 responses to examine the changes in auditory conductive function revealed by WB and 1-kHz tympanometric tests, even though fewer ears were tested. The interpretation of Day-2 data is constrained by the fact that the Day-2 refer group included only 14 ears. While 7 of these infants passed an outpatient, follow-up DPOAE screening, the screening outcomes for the remaining 7 infants were not known.

The distributions of 1-kHz tympanometry predictors on Day 2 (ΔY_{Pos}, ΔY_{Neg}, ΔY_{Slope}) are plotted in Fig. 2B as box and whisker plots following the convention used in Fig. 2A (Day-1 and Day-2 results are plotted next to each other for ease of comparison). The median $\Delta |Y|$ of the pass and refer groups differed by approximately 0.8 mmho for ΔY_{Neg}, and, in contrast to Day-1 results (Fig. 2A), also differed for ΔY_{Pos} and ΔY_{Slope}. Furthermore, the median in each of the refer groups of ΔY_{Pos} and ΔY_{Slope} was below the IQR of their respective pass group. A general trend is that the median and IQR were similar in the pass groups on Day 1 and Day 2.

Group results for Day-2 ambient WB ATFs are displayed in Fig. 3D, 3E, and 3F following the conventions used in Fig. 3A, 3B, and 3C (Day-1 and Day-2 results are plotted next to each other for ease of comparison). Unlike the Day-1 aEA results, the IQRs for Day-2 pass and refer group showed more overlap, lacking complete separation at any frequency. Results for Day-2 |aY| (Fig. 3E) were similar to those for Day 1, with the exception of higher |aY| for both pass and refer ears, especially above 4 kHz. Day-2 results for aYϕ (Fig. 3F) also showed a trend toward more IQR overlap for pass and refer groups in Day-2 ears, relative to Day-1 ears.

Comparison of aEA on Day 1 and Day 2

Differences in mean aEA were assessed for infants who were initially referred on Day 1 and then subsequently passed or were referred on Day 2. To evaluate trends in these differences, significance tests were performed at each frequency using a Welch *t* test with a Bonferroni correction (to adjust for tests at multiple frequencies) that allows for unequal variance in the two sample groups. Details of this test and its limitations in this application are described in Appendix C. Differences in mean aEA are plotted in Figs. 9A and 9B with error bars denoting the 95% CIs calculated by this *t* test.

Figure 9A shows the mean aEA difference for DP2-Refer ears relative to DP1-Refer ears. The mean aEA difference is positive at all but a narrow frequency range above 4 kHz, which is in the direction of an increased aEA for DP2-Refer ears relative to DP1-Refer ears. However, this trend was small compared to the 95% CIs and did not approach significance. Therefore, we interpret these data to mean that aEA in the Day-1 and Day-2 refer groups were approximately equal.

In contrast, comparisons of aEA between Day-2 passes and Day-1 refers appear to be more compelling. This comparison is shown in Fig. 9B. The mean aEA changed by as much as 0.33 for ears referred on Day 1 and passing on Day 2. Recall that EA can range from 0 (no energy absorbed by the middle ear) to 1.0 (all energy absorbed by the middle ear). Thus, a difference of 0.33 covers 1/3 of the total aEA range. The mean aEA differences exceeded

the 95% CI calculated using the *t* test over a broad frequency range from 0.84–4 kHz and 4.8–8 kHz, which suggests that the DP2-Pass ears absorbed significantly larger amounts of energy than they did one day earlier when they referred on the same test. These findings suggest that the transient sound-conduction effects, presumed to have been a contributing factor that led to an initial refer on Day 1, had begun to resolve by Day 2. This increased aEA for DP2-Pass ears is evidence of improved acoustic functioning in newborns over a 24-hour period.

Day-2 Test Performance

Following the conventions used to assess Day-1 results, test performance was assessed for 1-kHz tympanometric and WB tests for data collected on Day 2. In the case of the 1-kHz tympanometric predictors using the Day-2 responses, the AROC for each predictor did not differ significantly from chance performance (see upper portion of the right columns of Table I). Specifically, the 95% CIs associated with each variable included an AROC=0.50. Similarly, SYM did not exceed chance performance for any 1-kHz predictor. Thus, 1-kHz tympanometric data were unable to distinguish between Day-2 passes and Day-2 refers.

Evaluation of test performance for ambient WB tests revealed that |aY| and aYϕ performed above chance levels based on AROC at classifying the DPOAE status on Day 2, whereas aEA did not differ from chance performance (see middle portion of the right columns of Table I). For SYM, only |aY| exceeded chance performance. Each predictor had a smaller AROC and smaller SYM on Day 2 compared to Day 1. A pair-wise comparison showed no significant differences in AROC or in SYM among ambient WB tests. However, unlike 1-kHz tympanometry, some ambient WB tests were able to distinguish between pass and refer groups on Day 2.

AROC and SYM for Day-2 results are shown for tEA, |tY|, and tYϕ, in the lower portion of the right columns of Table I. Only tYϕ performed above chance for AROC, and none of the predictors performed above chance for SYM. Pair-wise comparisons in AROC and in SYM showed no significant differences among WB tympanometric tests. To compare the accuracy of WB ATFs and 1-kHz tympanometry for Day-2 results, two WB tests with the highest AROC (|aY| and tYϕ) were compared to a 1-kHz test with the highest AROC (ΔY_{Neg}). Figure 10 shows no significant differences for AROC and SYM for the WB tympanometric and 1-kHz tests, although |aY| performed above chance for both AROC and SYM.

Test performance for Day-2 data, evaluated using a subset of points defined by a separation in IQRs (re: discussion of Fig 5), revealed larger AROC values compared to test performance derived using all 6060 tEA frequency-pressure points. However, these differences were not significant, and AROC and SYM showed improvements only up to 0.04 and 0.07, respectively.

DISCUSSION

Test Performance

The primary focus of the present study was to assess the relative test performance of WB ATFs and 1-kHz admittance tympanometry in their ability to predict the status of the sound-conduction pathway in newborns being screened for hearing loss. Because a goal was to evaluate objective tests of sound conduction, which are suitable for a UNHS setting, the predictors for both WB and 1-kHz tympanometry were constructed with minimal subjective input, relying almost exclusively on automatic detection algorithms in both cases. The WB tests were more accurate than tests based on 1-kHz tympanometry in predicting effects of sound-conduction deficits in UNHS outcomes, demonstrating an advantage of WB

measurements over current tools that are used to assess the sound-conduction pathway early in life. Furthermore, WB tests showed greater energy absorbance in neonates who passed their DPOAE-based UNHS test, compared to infants who were referred on the primary screening test. It is often assumed that the majority of referrals from UNHS programs are the result of transient conditions of the external and/or middle ear. The results from this study provide evidence that the status of the sound-conduction pathway influences test performance based on DPOAE outcomes in a UNHS setting.

Compared to their test performance on Day 1, WB and 1-kHz predictors were less accurate in classifying Day-2 infants into pass and refer groups. Three of six WB predictors (|aY|, aYϕ, and tYϕ) exceeded chance performance for AROC on Day 2, while none of the 1-kHz predictors exceeded chance performance. Even so, the WB predictors resulted in performance only barely above the chance level. However, these results should be viewed with caution in light of the small number of ears in the DP2-Refer group (N=14); the insufficient statistical power of the Day-2 tests may have contributed to the resulting overlap in the CIs.

Piskorski et al. (1999) and Keefe and Simmons (2003) also showed improved accuracy in correctly predicting conductive hearing loss using WB ATF measurements compared to traditional 0.226-kHz immittance measurements. Specifically, Keefe and Simmons reported AROC values of 0.28, 0.72, and 0.94 for 0.226-kHz admittance tympanometry, WB ambient, and WB tympanometry measurements, respectively. Unlike the results from Keefe and Simmons, which support the use of tympanometric ATFs in combination with ambient ATF measurements with older children and adults, AROC for ambient and tympanometric WB measurements in the present study (for infants < 2 days old) were not significantly different. Differences in WB tympanogram test performance between Keefe and Simmons and the present study may be due to several factors. First, definitions of middle-ear status in the two studies were based on different gold standards. Keefe and Simmons (2003) assessed the ability of WB tests to predict a conductive hearing loss defined as an air-bone gap 20 dB, based on behavioral audiometric assessment. This traditional description of conductive hearing loss is not applicable in 1–2 day old neonates, which contributed to why the results from a DPOAE-based UNHS test were used as the gold standard. Second, Keefe and Simmons evaluated children (10 years) and adults. The differences in age are important in light of normative results showing maturational effects for ambient and tympanometric ATFs in infants and children (Keefe et al., 1993; Sanford & Feeney, 2008). Such maturational effects have also been described for single- and multi-frequency admittance tympanometry in young infants (Calandruccio, et al., 2006; Holte, et al., 1991; Hunter & Margolis, 1992).

Clinical Feasibility

The analysis of WB ATF data presented here suggest that algorithms can be developed to construct objective decision rules for classifying the status of the sound-conduction pathway of infant ears as "normal" or "impaired", which may be useful in the interpretation of results on primary UNHS tests. Specifically, application of the likelihood-ratio approach in the present study was an important tool in developing WB ATF test predictors, providing an efficient way of dealing with the large amount of data generated with ATF measurements. The bootstrap procedure enabled calculation of 95% CIs, which were helpful in assessing whether a test exceeded chance performance and to compare performance between tests. Analysis of tympanometric ATFs using likelihood ratios from a subset of WB tympanogram frequency-pressures points did not yield significantly different results in test performance. This result suggests that for the ATF tests evaluated in the present study, the criteria based on the likelihood ratio produced robust results, regardless of whether the entire range or the subset of tEA data were considered. Such WB criteria could be incorporated into an

instrument to provide an objective output indicative of the status of the sound-conduction pathway. This feature may be useful, given that UNHS tests are sometimes performed by untrained examiners. However, it should be noted that even though WB ATF tests produced the highest SYM values (0.78), additional research is need to determine if significant clinical benefit can be obtained with this level of test performance in a UNHS setting.

It is also useful to compare test results from other studies using similar fixed points on the ROC curve. For example, Margolis et al. (2003) assessed 1-kHz tympanometric test performance in predicting whether a newborn infant passed or failed a DPOAE test, and reported a specificity of 91% and a sensitivity of 50% (using peak-negative tail admittance values). The Margolis et al. results are compared in Fig. 11 with ROC data from Day 1 infants for the present study for the 1-kHz tympanometry peak-negative admittance predictor (ΔY_{Neg}), which is analogous to the predictor used by Margolis et al., and with the 1 kHz (ΔY_{Pos}) and WB ($aY\phi$, aEA) predictors with the highest AROC values. With a fixed specificity of 91% (see open-circle symbols in Fig. 11), the sensitivity of the ΔY_{Neg} predictor was 36%, which was lower than the 50% reported by Margolis et al. The results for both studies are similar in that the sensitivity was low using a peak-negative admittance predictor. For the additional predictors, with a fixed specificity of 91%, sensitivity was 2% for ΔY_{Pos}, 49% for aEA, and 65% for $aY\phi$. Conversely, with a fixed sensitivity of 90% (see open-diamond symbols in Fig. 11), the specificity was 18% for ΔY_{Neg}, 36% for ΔY_{Pos}, 57% for aEA, and 66% for $aY\phi$. Even though $aY\phi$ and aEA had the same SYM of 78% (open boxes on ROC curves in Fig. 11), the sensitivity at 90% specificity was larger for $aY\phi$ (65% vs. 49%).

This example shows effects of the real-world tradeoffs in choosing a particular pass-refer criterion for a particular test, but the results are consistent in demonstrating that the WB tests were more accurate than 1-kHz tympanometry. The $aY\phi$ test, which is the reference WB test in the present study, achieved a 90% sensitivity with a 66% specificity. This lower specificity would be interpreted alongside the fact that these ears passed the DPOAE UNHS exam, and therefore were likely to have conductive function within normal limits. A high sensitivity would appear to be advantageous in a test to detect conductive dysfunction in newborns. In comparison, the best 1-kHz tympanometry performance was 90% sensitivity with a specificity of only 36% (for ΔY_{Pos}).

The fact that a WB ATF test requires between 1 and 7 seconds to perform makes it clinically practical for potential use in UNHS programs. Given that both ambient and tympanometric measurements of ATFs performed equally well, a parsimonious approach might be one in which ambient-pressure tests are included. The fact that WB tympanometry was more accurate than ambient WB ATF measurements in predicting conductive hearing loss in children of age 10 years and older (i.e., Keefe and Simmons, 2003) shows that WB tympanometry provides information on sound-conduction status over and above an ambient WB test, although the age at which this transition in test performance occurs remains uncertain. More research is needed to compare the test performance of WB ambient and tympanometric measurements in infants recruited from neonatal intensive care units and at older ages, such as infants receiving follow-up screening and/or diagnostic tests because of their referral from an initial UNHS protocol.

WB Measurements in Relation to UNHS Test Outcomes

It has been widely assumed that some number of referrals occur as a result of transient conditions in the ear canal and/or middle ear. For example, a local (Omaha, NE) UNHS program has a referral rate of approximately 2%. Given a prevalence of permanent hearing loss of 1–3 per 1000 births, a referral rate of 2% is approximately 10 times greater than the prevalence. In an ideal situation, all infants who refer on their UNHS tests would return

shortly after hospital discharge for follow-up testing to determine whether permanent hearing loss exists. Unfortunately, this is not the case; many infants referred on the basis of a UNHS test do not return for follow-up testing in a timely fashion (Johnson, et al., 2005). Furthermore, recent data from the U.S. Centers for Disease Control (CDC) and Prevention shows almost half (46.3%) of infants born in 2006 who did not pass their final hearing screening as being lost to follow-up/lost to documentation (CDC, 2008). Because alteration in the status of the sound-conduction pathway can influence UNHS test outcomes, inclusion of a WB ATF test may improve the interpretation of OAE and ABR screening results (Keefe et al., 2003b), help differentiate screening outcomes due to conductive effects, and assist EHDI providers with screening and follow-up decisions. Spivak et al. (2008) reported that out of 1222 infants returning for follow-up hearing screenings, 37% had either transient or conductive hearing loss. Furthermore, they concluded that conductive hearing loss was a significant factor in infants being lost to follow up and/or not receiving appropriate amplification by 6 months of age, mainly due to delays in diagnosis. If more accurate information regarding the conductive status of an infants ear were known at birth, and during the timeframe over which follow-up and diagnostic testing is typically performed, this could provide hearing healthcare professionals with more accurate information and potentially reduce delays in intervention.

Given that (1) a high percentage of UNHS referrals are likely due to reduced conduction of sound by the ear canal/middle ear, (2) a significant number of referred infants do not return for follow-up testing in a timely manner, and (3) the resources that can be devoted to follow-up efforts may be limited, there may be potential value in determining which infants are likely to be in the greatest need of additional services. If an infant refers on their UNHS test but also shows evidence of *abnormal* sound conduction, this suggests, although it does not directly establish, that the external/middle-ear status may have played a role in the UNHS test outcome. Additional testing would need to be performed to rule out the possibility of a sensorineural hearing loss, but at least there would be data to suggest that the sound-conduction pathway may have contributed to the UNHS outcome. In contrast, consider the infant who refers on their primary UNHS test, but who also produces evidence that the sound-conduction pathway is functioning *within normal limits*. In this case, it is less likely that the referral is due to the status of the conduction pathway. While attempts should be made to re-test all infants who do not pass a UNHS test, it might be useful to provide more intensive follow-up efforts for these infants, given limited resources and the number of infants who do not return following a referral. This approach would be consistent with findings in Keefe (2007), in which analyses of data from Keefe et al. (2003b) showed that in ears referred by a two-stage DPOAE/ABR or CEOAE/ABR UNHS protocol, an ear that passed an ambient ATF test was twice as likely[iii] to have a sensorineural hearing loss as an ear that referred on an ambient ATF test. The simplest interpretation of this observation is that middle-ear status played a role in predicting subsequent UNHS test outcomes with

[iii]Sensorineural hearing loss (SNHL) was determined using visual reinforcement audiometry (VRA) at 8–12 months. Specifically, 51 ears were false positives on the DPOAE/ABR test (i.e., classified as normal hearing by VRA), and 12 ears were true positives on the DPOAE/ABR test (i.e., classified as SNHL by VRA). Of the false positives for SNHL, 11 were normal and 40 were referred on sound conduction status based on the ATF test. Of the true positives for SNHL, 5 were normal and 7 were referred on sound conduction status based on the ATF test. The probability of SNHL in the true-positive subjects who were normal on the sound conduction test was $5/(11+5) = 31\%$. The probability of SNHL in the false-positive subjects who referred on the sound conduction test was $7/(40+7) = 14\%$. The odds ratio for SNHL in the true-positive subjects relative to the false-positive subjects was sound conduction status in subjects with an ear that was normal on the sound conduction test was $31\%/14\% = 2.10$. This explains the statement in the text that an ear that passed the sound conduction test (i.e., the ambient ATF test in the cited study) was twice as likely to have a SNHL as an ear that referred on the sound conductive test. While these results are limited by a relatively small number of subjects, the primary factor driving the odds ratio larger than one is the fact that 40 of the normal-hearing ears had normal sound conduction while only 11 referred on sound conduction. As discussed in Keefe (2007), the odds ratio would tend to be increased (from 2.1 to perhaps 2.5) in a sample with a lower prevalence of SNHL than this test sample, whose prevalence was $12/(12+51) =19\%$ as a result of preferentially recruiting subjects likely to have SNHL in the VRA tests.

implications for follow-up efforts. Nevertheless, it remains the goal that every referral, regardless of whether there is evidence of middle-ear dysfunction or not, should return for a follow-up diagnostic exam to identify sensorineural hearing loss if UNHS is to achieve its primary goal of providing services to infants with hearing loss early in life.

Day-1 and Day-2 Comparisons

Comparison of Day-1 and Day-2 results in Figure 9B, showing trends toward increased energy absorbance, suggest that middle-ear effects, presumed to have been a contributing factor leading to a referral on Day 1, had begun to resolve by Day 2. These results are consistent with previous studies which have suggested that middle-ear and ear-canal factors, such as fluid and debris from birth, respectively, influence OAE and ABR UNHS outcomes (Chang, et al., 1993; Doyle, et al., 2000; McKinley, et al.; 1997; Thornton, et al., 1993). Additional studies of WB ATF measurements with larger numbers of "Day-2" infants are needed to extend the results reported in the present study.

Developmental Influences on WB ATF Measurements

As mentioned above, in Day-1 Test Results, adding to the complexity of the two-dimensional distribution of tEA data was the presence of negative EA across the tympanometric grid of air pressure and frequency. This effect was compared in groups of infant and adult ears by calculating the prevalence of WB tympanograms in each group for which tEA was negative for 1% or more of the grid points. The adult-group prevalence was based on data reported by Liu et al. (2008) using the same WB tympanometry system as was used in the present study, with a descending-pressure sweep over the same pressure range at the same nominal sweep rate. For infants, the prevalence of negative EA was 52% for the Day-1 pass group and 66% for the Day-1 refer group; the majority of these negative EA responses occurred between 2 and 8 kHz for pressures between −5 and −300 daPa. For the adults tested by Liu et al. (2008), the prevalence was only 5.4% in a sample of 92 adult ears. This prevalence is an order of magnitude smaller than the prevalence in infants.

Sanford and Feeney (2008) reported greater (in the direction of lower absorbance) negative-pressure-induced shifts in ATF responses for 4-week-old infants compared to adults. Although they did not report negative tEA data, Sanford and Feeney showed mean absorbance values at 4 kHz, at an ear-canal pressure of -200 daPa, that were as low as 0.3 (1 SD=0.29) and 0.6 (1 SD=0.08) for infants and adults, respectively. It is possible that the higher prevalence of negative EA values for infants in the present study, which would produce artifacts in acoustic admittance (in the form of negative conductance), may be attributed to the functional effects of an immature ear-canal wall.

Liu et al. (2008) commented on the presence of negative tEA at some grid points, suggesting that if a true value of EA were already close to 0, a small calibration error might shift EA to a negative value. Perhaps for infants in the present study, a calibration error also may have shifted EA to negative values, particularly for negative pressures. Liu et al. also suggested that errors might have been introduced if the transducer sensitivity were influenced by air-pressure change, although the probe was designed to achieve properties that were independent of air pressure over the tympanometric range. While calibration and/or transducer sensitivity cannot be excluded as contributing factors, there remains an order of magnitude difference in the prevalence of these negative EA values in newborns compared to adults. This fact is consistent with the presence of additional contributing factor(s) in the tympanometric measurements in newborns.

Model simulations based on temporal-bone data from a 22-day-old newborn have shown a 27–75% change in ear-canal volume over a pressure range of ± 300 daPa (Qi et al., 2006,

2008). Even larger relative volume changes may occur in younger infants, and the possibility exists for complete closure of the ear canal at negative pressures. For example, it is possible that a near closure of the ear canal may present an acoustical termination at the probe with characteristics that are outside the calibration range of the measurement system. Thus, the presence of negative EA values may be related to calibration issues resulting from changes in the ear canal that occur during the pressure sweep.

The differences in WB tympanograms for ascending vs. descending pressure-sweeps in the present study (Fig. 6) were similar to those reported by Holte et al. (1991) for multi-frequency admittance tympanograms (assessed at 6 frequencies from 0.226 to 0.9 kHz). The ascending/descending pressure-sweep results in infants differed from those observed in adults for both single-frequency admittance (ASHA, 1988) and WB tympanometry (Liu et al., 2008). While Liu et al. observed pressure-sweep direction effects in adults (shifts in TPP and morphological tympanogram changes); the effects were smaller than those observed in infants in the present study and in Holte et al. (1991). Although current models do not yet incorporate the temporal dynamics associated with positive versus negative sweep directions, the hysteresis effects that were large under negative-pressure conditions in the present study may be related to viscoelastic properties of the immature ear-canal wall, resulting in its collapse.. The present results support the hypothesis proposed by Holte et al. (1990, 1991) that ear-canal hysteresis contributes to developmental differences in tympanometric measurements of middle-ear function. Future research is needed on the effect of sweep polarity in infants as a function of age. In conjunction with developmental anatomical data and physical models incorporating dynamical hysteresis effects, such research may improve the understanding and clinical interpretation of tympanometric measurements in infants and young children.

Limitations and Summary

A limitation of the present study is that the extent to which these effects generalize to other populations of infants receiving UNHS tests is unknown. In particular, the likelihood ratio for each ear was calculated based on the multivariate WB ATF test outputs for that ear, and the means and standard deviations of the multivariate outputs for the pass and refer groups on Day 1. Thus, the manner of calculating the likelihood ratio was identical for the pass and refer groups, and did not explicitly rely on the gold standard for that ear. Nevertheless, it is possible that other UNHS populations may have different distributions of WB ATF responses than the Day-1 sample population in this study (i.e., graduates from intensive care nurseries). The data reported in the present experiment, however, are expected to be representative for the majority of neonates, who typically reside in well-baby nurseries. Still, it would be preferable to evaluate the uncertainty in test performance by repeating the experiment in independent groups of infants, but such a procedure exceeded the scope of this study.

The results from this study can be summarized as follows: (1) WB ATF tests outperformed tests based on 1-kHz tympanometry at predicting DPOAE-based UNHS outcomes, (2) WB measurements can be used to quickly and objectively assess the status of the newborn sound-conduction pathway, making them feasible for potential use in UNHS programs, (3) WB ATF tests provide data to suggest that many UNHS referrals are a consequence of transient deficits in the sound-conduction pathway and (4) WB ATF data reveal changes in energy transfer for the sound-conduction pathway during the first 2 days of life.

Acknowledgments

Sonicom, Inc., are involved in commercializing technology intended to improve the screening and diagnosis of middle-ear disorders.

References

American Speech-Language-Hearing Association. Working Group on Aural Acoustic-Immittance Measurements Committee on Audiologic Evaluation. Tympanometry. J Speech Hear Disord. 1988; 53:354–377. [PubMed: 3054320]

Abdala C, Keefe DH, Oba SI. Distortion product otoacoustic emission suppression tuning and acoustic admittance in human infants: Birth through 6 months. J Acoust Soc Am. 2007; 121:3617–3627. [PubMed: 17552713]

Alaerts J, Luts H, Wouters J. Evaluation of middle ear function in young children: Clinical guidelines for the use of 226- and 1,000-Hz tympanometry. Otol Neurotol. 2007; 28:727–732. [PubMed: 17948353]

Allen, JB. Measurement of eardrum acoustic impedance. In: Hall, JL.; Allen, JB.; Hubbard, A.; Neely, ST.; Tubis, A., editors. Peripheral Auditory Mechanisms. Springer-Verlag; New York: 1986. p. 44-51.

Baldwin M. Choice of probe tone and classification of trace patterns in tympanometry undertaken in early infancy. Int J Audiol. 2006; 45:417–27. [PubMed: 16938800]

Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. J Math Psychol. 1975; 12:387–415.

Calandruccio L, Fitzgerald TS, Prieve BA. Normative multifrequency tympanometry in infants and toddlers. J Am Acad Audiol. 2006; 17:470–80. [PubMed: 16927511]

Centers for Disease Control and Prevention [CDC Web site]. Final summary of 2006 national EHDI data (version 4). 2008 Oct. Retrieved April 15th, 2009 from http://www.cdc.gov/ncbddd/ehdi/documents/EHDI_Summ_2006_Web.pdf

Chang KW, Vohr BR, Norton SJ, et al. External and middle ear status related to evoked otoacoustic emission in neonates. Arch Otolaryngol Head Neck Surg. 1993; 119:276–282. [PubMed: 8435165]

Doyle KJ, Burggraaff B, Fujikawa S, et al. Neonatal hearing screening with otoscopy, auditory brain stem response, and otoacoustic emissions. Otolaryngol Head Neck Surg. 1997; 116:597–603. [PubMed: 9215369]

Doyle KJ, Kong YY, Strobel K, et al. Neonatal middle ear effusion predicts chronic otitis media with effusion. Otol Neurotol. 2004; 25:318–322. [PubMed: 15129112]

Doyle KJ, Rodgers P, Fujikawa S, et al. External and middle ear effects on infant hearing screening test results. Otolaryngol Head Neck Surg. 2000; 122:477–81. [PubMed: 10740164]

Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science. 1986; 1:54–77.

Feeney MP, Grant IL, Marryott LP. Wideband energy reflectance measurements in adults with middle-ear disorders. J Speech Lang Hear Res. 2003; 46:901–911. [PubMed: 12959468]

Feeney MP, Sanford CA. Age effects in the human middle ear: wideband acoustical measures. J Acoust Soc Am. 2004; 116:3546–58. [PubMed: 15658706]

Finitzo T, Albright K, O'Neal J. The newborn with hearing loss: Detection in the nursery. Pediatrics. 1998; 102:1452–1460. [PubMed: 9832584]

Holte L, Margolis RH, Cavanaugh RM Jr. Developmental changes in multifrequency tympanograms. Audiology. 1991; 30:1–24. [PubMed: 2059166]

Hudde H. Measurement of the eardrum impedance of human ears. J Acoust Soc Am. 1983; 73:242–247. [PubMed: 6826891]

Hunter LL, Bagger-Sjöbäck D, Lundberg M. Wideband reflectance associated with otitis media in infants and children with cleft palate. Intl J Audiology. 2008; 47:S57–S61.

Hunter LL, Margolis RH. Multifrequency tympanometry: Current clinical application. Am J Audiol. 1992; 1:33–43.

Hunter LL, Tubaugh L, Jackson A, et al. Wideband middle ear power measurement in infants and children. J Am Acad Audiol. 2008; 19:309–324. [PubMed: 18795470]

Ikui A, Sando I, Fujita S. Postnatal change in angle between the tympanic annulus and surrounding structures: Computer-aided three-dimensional reconstruction study. Ann Otol Rhinol Laryngol. 1997; 106:33–36. [PubMed: 9006358]

Jerger J. Clinical experience with impedance audiometry. Arch Otolaryngol. 1970; 92:311–24. [PubMed: 5455571]

Johnson JL, White KR, Widen JE, et al. A multi-site study to examine the efficacy of the otoacoustic emission/automated auditory brainstem response newborn hearing screening protocol: introduction and overview of the study. Am J Audiol. 2005; 14:178–85.

Joint Committee on Infant Hearing. Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. Pediatrics. 2007; 120:898–921. [PubMed: 17908777]

Keefe, DH. Influence of Middle-Ear Function and Pathology on Otoacoustic Emissions. In: Robinette, MS.; Glattke, TJ., editors. Otoacoustic Emissions: Clinical Applications. New York: Thieme Medical; 2007. p. 163-196.

Keefe DH, Abdala C. Theory of forward and reverse middle-ear transmission applied to otoacoustic emissions in infant and adult ears. J Acoust Soc Am. 2007; 121:978–993. [PubMed: 17348521]

Keefe DH, Bulen JC, Arehart KH, et al. Ear-canal impedance and reflection coefficient in human infants and adults. J Acoust Soc Am. 1993; 94:2617–2638. [PubMed: 8270739]

Keefe, DH.; Feeney, MP. Principles of acoustic immittance and acoustic transfer functions. In: Katz, J.; Burkhard, RF.; Medwetsky, L.; Hood, LJ., editors. Handbook of Clinical Audiology. New York: Lippincott, Williams & Wilkins; 2009. p. 125-156.

Keefe DH, Folsom RC, Gorga MP, et al. Identification of neonatal hearing impairment: ear-canal measurements of acoustic admittance and reflectance in neonates. Ear Hear. 2000; 21:443–461. [PubMed: 11059703]

Keefe DH, Gorga MP, Neely ST, et al. Ear-canal acoustic admittance and reflectance effects in human neonates. II. Predictions of middle-ear dysfunction and sensorineural hearing loss. J Acoust Soc Am. 2003b; 113:407–422. [PubMed: 12558278]

Keefe DH, Levi E. Maturation of the middle and external ears: acoustic power-based responses and reflectance tympanometry. Ear Hear. 1996; 17:361–373. [PubMed: 8909884]

Keefe DH, Ling R, Bulen JC. Method to measure acoustic impedance and reflection coefficient. J Acoust Soc Am. 1992; 91:470–485. [PubMed: 1737890]

Keefe DH, Simmons JL. Energy transmittance predicts conductive hearing loss in older children and adults. J Acoust Soc Am. 2003; 114:3217–3238. [PubMed: 14714804]

Keefe DH, Zhao F, Neely ST, et al. Ear-canal acoustic admittance and reflectance effects in human neonates. I. Predictions of otoacoustic emission and auditory brainstem responses. J Acoust Soc Am. 2003a; 113:389–406. [PubMed: 12558277]

Kei J, Allison-Levick J, Dockray J, et al. High-frequency (1000 Hz) tympanometry in normal neonates. J Am Acad Audiol. 2003; 14:20–28. [PubMed: 12833925]

Keith RW. Middle ear function in neonates. Arch Otolaryngol. 1975; 101:376–9. [PubMed: 1131105]

Kennedy CR. Controlled trial of universal neonatal screening for early identification of permanent childhood hearing impairment: coverage, positive predictive value, effect on mothers and incremental yield. Wessex universal neonatal screening trial guide. Acta Paediatr Suppl. 1999; 88:73–5. [PubMed: 10626585]

Kok MR, van Zanten GA, Brocaar MP. Growth of evoked otoacoustic emissions during the first days postpartum. A preliminary report. Audiology. 1992; 31:140–149. [PubMed: 1642565]

Lawton BW, Stinson MR. Standing wave patterns in the human ear canal used for estimation of acoustic energy reflectance at the eardrum. J Acoust Soc Am. 1986; 79:1003–1009. [PubMed: 3700855]

Lidén G. The scope and application of current audiometric tests. J Laryngol Otol. 1969; 83:507–20. [PubMed: 5785649]

Liu Y, Sanford CA, Ellison JC, et al. Wideband absorbance tympanometry using pressure sweeps: System development and results on adults with normal hearing. J Acoust Soc Am. 2008 (In Press).

Marchant CD, McMillan PM, Shurin PA, et al. Objective diagnosis of otitis media in early infancy by tympanometry and ipsilateral acoustic reflex thresholds. J Pediatr. 1986; 109:590–5. [PubMed: 3761072]

Margolis RH, Bass-Ringdahl S, Hanks WD, et al. Tympanometry in newborn infants--1 kHz norms. J Am Acad Audiol. 2003; 14:383–392. [PubMed: 14620612]

Margolis RH, Paul S, Saly GL, Schachern PA, et al. Wideband reflectance tympanometry in chinchillas and human. J Acoust Soc Am. 2001; 110:1453–64. [PubMed: 11572356]

Margolis RH, Saly GL, Keefe DH. Wideband reflectance tympanometry in normal adults. J Acoust Soc Am. 1999; 106:265–280. [PubMed: 10420621]

McKinley AM, Grose JH, Roush J. Multifrequency tympanometry and evoked otoacoustic emissions in neonates during the first 24 hours of life. J Am Acad Audiol. 1997; 8:218–23. [PubMed: 9188078]

Myung IJ. Tutorial on maximum likelihood estimation. J Math Psychol. 2003; 47:90–100.

Norton SJ, Widen JE, Gorga MP, et al. Identification of neonatal hearing impairment: Evaluation of TEOAE, DPOAE, and ABR test performance. Ear Hear. 2000; 21:508–528. [PubMed: 11059707]

Paradise JL, Smith CG, Bluestone CD. Tympanometric detection of middle ear effusion in infants and young children. Pediatrics. 1976; 58:198–210. [PubMed: 951134]

Pepe, MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford: Oxford University Press; 2003.

Piskorski P, Keefe DH, Simmons JL, et al. Prediction of conductive hearing loss based on acoustic ear-canal response using a multivariate clinical decision theory. J Acoust Soc Am. 1999; 105:1749–64. [PubMed: 10089599]

Puria S. Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions. J Acoust Soc Am. 2003; 113:2773–2789. [PubMed: 12765395]

Qi L, Funnell WR, Daniel SJ. A nonlinear finite-element model of the newborn middle ear. J Acoust Soc Am. 2008; 124:337–347. [PubMed: 18646981]

Qi L, Liu H, Lutfy J, et al. A nonlinear finite-element model of the newborn ear canal. J Acoust Soc Am. 2006; 120:3789–3798. [PubMed: 17225406]

Rosenfeld RM, Culpepper L, Doyle KJ, et al. Clinical practice guideline: Otitis media with effusion. Otolaryngol Head Neck Surg. 2004; 130:95–118.

Ruah CB, Schachern PA, Zelterman D, et al. Age-related morphologic changes in the human tympanic membrane. A light and electron microscopic study. Arch Otolaryngol Head Neck Surg. 1991; 117:627–634. [PubMed: 2036184]

Sanford CA, Feeney MP. Effects of maturation on tympanometric wideband acoustic transfer functions in human infants. J Acoust Soc Am. 2008; 124:2106–2122. [PubMed: 19062852]

Sanford, CA.; Fitzpatrick, D.; Keefe, DH., et al. Wideband energy reflectance predicts transient middle-ear dysfunction in neonates. American Auditory Society Annual Meeting; 2008.

Saunders, JC.; Kaltenback, JA.; Relkin, EM. The Structural and Functional Development of the Outer and Middle ear. In: Romand, R.; Romand, MR., editors. Development of Auditory and Vestibular Systems. New York: Academic; 1983. p. 3-25.

Shahnaz N. Wideband reflectance in neonatal intensive care units. J Am Acad Audiol. 2008; 19:419–429. [PubMed: 19256090]

Shahnaz N, Bork K. Wideband reflectance norms for Caucasian and Chinese young adults. Ear Hear. 2006; 27:774–88. [PubMed: 17086086]

Spivak L, Sokol H, Auerbach C, Gershkovich S. Newborn Hearing Screening Follow-up: Factors Affecting Hearing Aid Fitting by Six Months of Age. Am J Audiol. 2008 [Epub ahead of print].

Sprague BH, Wiley TL, Goldstein R. Tympanometric and acoustic-reflex studies in neonates. J Speech Hear Res. 1985; 28:265–72. [PubMed: 4010256]

Stinson MR. Revision of estimates of acoustic energy reflectance at the human eardrum. J Acoust Soc Am. 1990; 88:1773–1778. [PubMed: 2262633]

Stuart A, Yang EY, Green WB. Neonatal auditory brainstem response thresholds to air- and bone-conducted clicks: 0 to 96 hours postpartum. J Am Acad Audiol. 1994; 5:163–72. [PubMed: 8075412]

Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988; 240:1285–1293. [PubMed: 3287615]

Takahara T, Sando I, Hashida Y, et al. Mesenchyme remaining in human temporal bones. Otolaryngol Head Neck Surg. 1986; 95:349–57. [PubMed: 3108785]

Tukey, JW. Exploratory Data Analysis. Reading, MA: Addison-Wesley; 1977.

Thornton ARD, Kimm L, Kennedy CR, et al. External- and middle-ear factors affecting evoked otoacoustic emissions in neonates. Br J Audiol. 1993; 27:319–27. [PubMed: 8205077]

Van Trees, HL. Detection, Estimation, and Modulation Theory. New York: Wiley Interscience; 1967.

Vander Werff KR, Prieve BA, Georgantas LM. Test-retest reliability of wideband reflectance measures in infants under screening and diagnostic test conditions. Ear Hear. 2007; 28:669–681. [PubMed: 17804981]

Voss SE, Allen JB. Measurement of acoustic impedance and reflectance in the human ear canal. J Acoust Soc Am. 1994; 95:372–84. [PubMed: 8120248]

APPENDIX A

List of Frequently Used Abbreviations

ATF   Acoustic transfer function

DP1-Pass   Infant or group of infants who passed a DPOAE test on day 1

DP1-Refer Infant or group of infants who did not pass a DPOAE test on day 1

DP2-Pass   Infant or group of infants who passed a DPOAE test on day 2

DP2-Refer Infant or group of infants who did not pass a DPOAE test on day 2

EA   Energy absorbance (1-ER)

aEA   Energy absorbance obtained under ambient ear-canal conditions

tEA   Energy absorbance obtained under pressurized (tympanometric) ear-canal conditions

ER   Energy reflectance

$|Y|$ Admittance magnitude

$|Y|_+$ Admittance magnitude at the most extreme positive pressure on the 1 kHz tympanogram tracing

$|Y|_-$ Admittance magnitude at the most extreme negative pressure on the 1 kHz tympanogram tracing

$|Y|_{TPP}$ The largest admittance magnitude value (tympanometric peak pressure) on the 1 kHz tympanogram tracing

ΔY_{Neg} $|Y|_{TPP} - |Y|_-$

ΔY_{Pos} $|Y|_{TPP} - |Y|_+$

ΔY_{Slope} Admittance difference between $|Y|_{TPP}$ and a point on the 1-kHz tympanogram "pressure-slope line"

aY   Wideband admittance obtained under ambient ear-canal conditions

aYϕ   Wideband admittance phase obtained under ambient ear-canal conditions

tY   Wideband admittance obtained under pressurized (tympanometric) ear-canal conditions

tYϕ   Wideband admittance phase obtained under pressurized (tympanometric) ear-canal conditions

WB   Wideband denotes the bandwidth (0.226 – 8.0 kHz) of ATF stimuli and responses

APPENDIX B

Calculation of the Likelihood-Ratio Predictor

A WB response classifier was developed to produce likelihood estimates used to determine from which Day-1 group (DP1-Pass or DP1-Refer) a particular WB response was more likely to have come. Responses across the entire aEA bandwidth (0.226 to 8 kHz) were considered in the analyses (Van Trees, 1967), which produced larger relative weightings at frequencies where the overlap between the two distributions (pass and refer) was less. This enabled the calculation of a predictor which optimally combined multivariate responses into a univariate output with the potential for more accuracy than the response at any single frequency.

The means ($\mu_P(i)$, $\mu_R(i)$) and standard deviations ($\sigma_P(i)$, $\sigma_R(i)$) were calculated for the DP-Pass and DP-Refer groups, respectively, at the i-th frequency. The measurements at each frequency were assumed to be normally distributed and uncorrelated with the measurements at any other frequency. Consider a particular ambient WB measurement of x, with values $x(i)$ specified at N frequencies indexed by $i=1,2,...,N$. The likelihood that x was selected from the DP1-Refer group (L_R) was calculated as follows:

$$L_R = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_R^2(i)}} \exp\left(\frac{-(x(i)-\mu_R(i))^2}{2\sigma_R^2(i)}\right). \quad \text{(B1)}$$

The likelihood that x was selected from the DP1-Pass group (L_P) was calculated similarly. Whether the measurements are normally distributed and uncorrelated or not, the test-performance results provide an evaluation of the accuracy of this likelihood-predictor approach. Note that one change was needed when L_R and L_P were calculated for tEA data, in that the index i varied over all frequencies *and* pressures.

A likelihood ratio (ρ) for the WB measurement of x is defined as,

$$\rho = L_R/L_P. \quad \text{(B2)}$$

Likelihood ratios larger than 1 indicated a higher likelihood that the ear was from a refer group; likelihood ratios smaller than 1 indicated a higher likelihood that the ear was from a pass group. This is the univariate WB predictor described in the main text. This likelihood-ratio predictor was defined for each of the ambient-pressure WB responses (aEA response, |aY|, and aYϕ) and each WB tympanogram (tEA, |tY|, and tYϕ).

APPENDIX C

Comparison of aEA on Day 1 and Day 2

The difference between the means from a pair of sample populations of unequal sample size and possibly unequal population variance can be evaluated for statistical significance using a Welch *t* test. This *t*-test also provides the 95% CIs for the difference in the means. The test assumes that the samples are independently selected in the two sample populations.

Significance testing for the mean difference in aEA on Day 1 and Day 2 is more complicated. aEA was a multivariate predictor, defined at multiple frequencies with values that may not be independent. Assuming independence, a Bonferroni correction was applied, so that the α of 0.05 that would be used in the *t*-test of a univariate predictor was modified to an α of 0.05/60 based on the 60 frequencies in each aEA response. Moreover, aEA was a repeated measure for some ears present in both groups, but any particular ear in the Day-1 refer group was not necessarily present in either of the Day-2 groups (pass or refer). This situation arose either because not all infants in the Day 1 group were tested on Day 2, or the ear response was classified into the other Day-2 group. Thus, the *t*-test results were used in the present study as a guide to evaluate the role of maturation in aEA over a 24-hour period.

To the extent that the 95% CIs calculated by the Welch *t* test with Bonferroni correction were well separated from an aEA mean difference of 0, then there was probably a significant difference in the means of aEA in the Day-1 refer group compared to either of the Day-2 groups. An overlap in the 95% CIs with aEA=0 indicated no difference in the means of aEA on Day 1 and Day 2. Any mean differences of borderline significance on this *t* test would require a more detailed statistical analysis, but the presence of mean differences of only borderline significance would likely convey no practical clinical value.
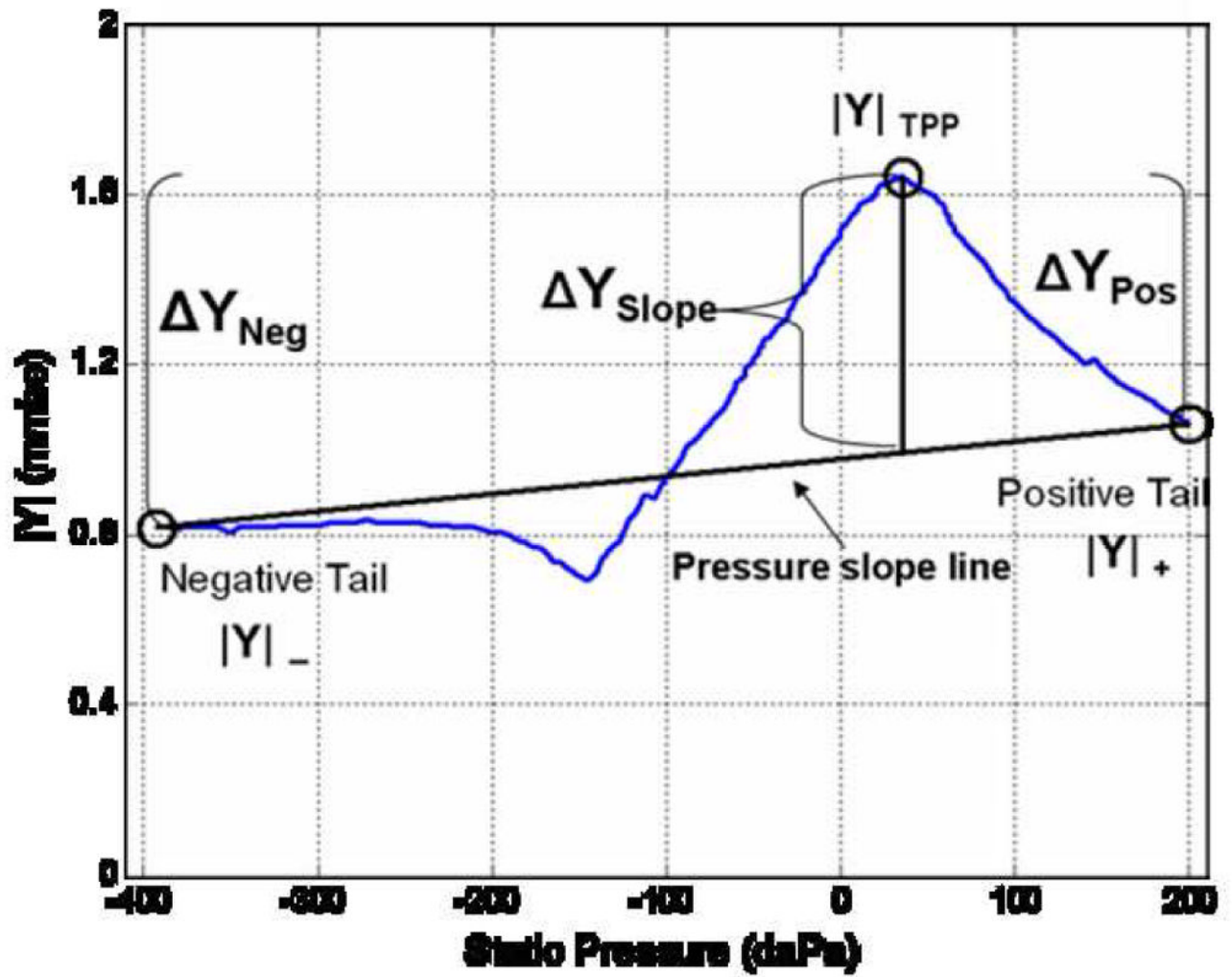
Fig. 1.
1-kHz |Y| tympanogram (single infant ear) obtained using a Madsen Otoflex 100 instrument and plotted using custom software. This example illustrates points on the tympanogram tracing used to derive 1-kHz tympanometric predictors (e.g., ΔY_{Neg}, ΔY_{Pos}, and ΔY_{Slope}).
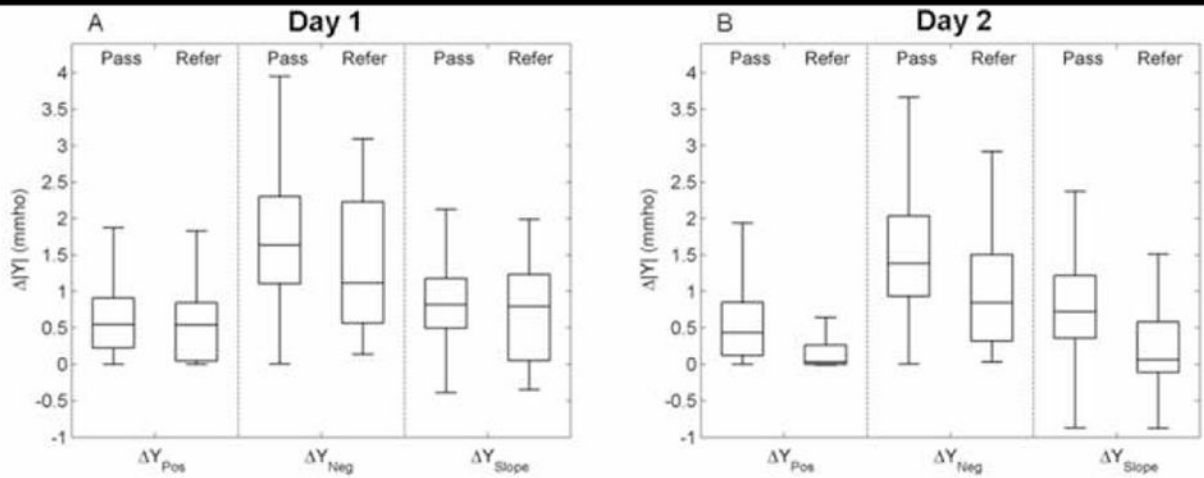
Fig. 2.
Box and whisker plots of average $\Delta|Y|$ from 1-kHz tympanograms for ΔY_{Neg}, ΔY_{Pos}, and ΔY_{Slope} variables for Day-1 (panel A) and Day-2 (panel B) infant pass and refer groups. For each box, the lower, middle, and upper horizontal lines represent the 25th, 50th, and 75th percentiles of the respective distributions. Upper and lower whiskers correspond to either 1.5 x the interquartile range (IQR) or the extremes of the data range.
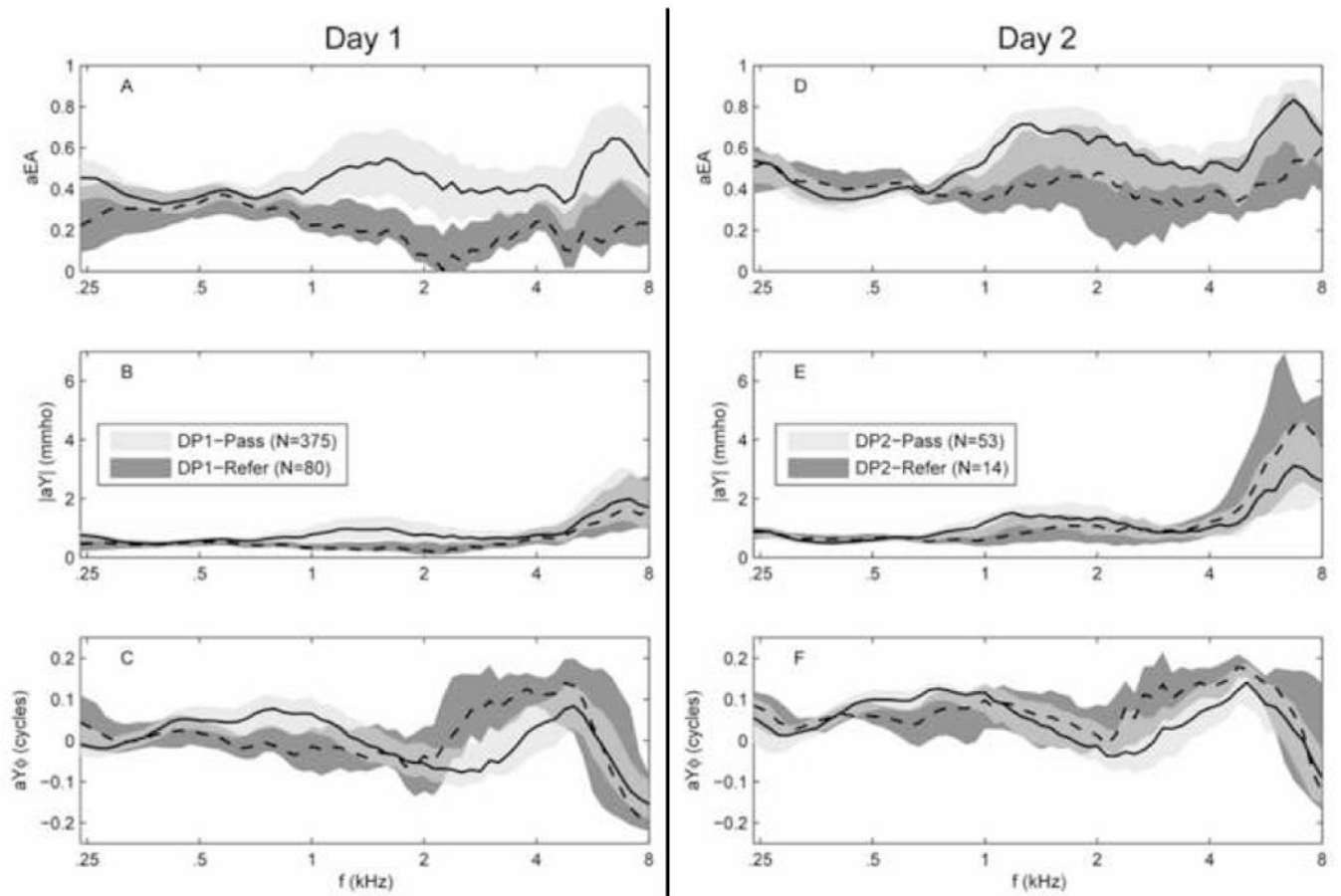
Fig. 3.
Group averaged Day-1 (panels A, B, and C) and Day 2 (panels D, E, and F) aEA, |aY|, and aYϕ for both pass and refer groups. Shaded areas denote the IQRs. (Pass = light gray shading; Refer = dark gray shading; where there is IQR overlap, the dark gray shading can been seen "behind" the light gray shading). Black lines, solid and dashed, represent the 50th percentile for the DP1-Pass and DP1-Refer groups respectively.
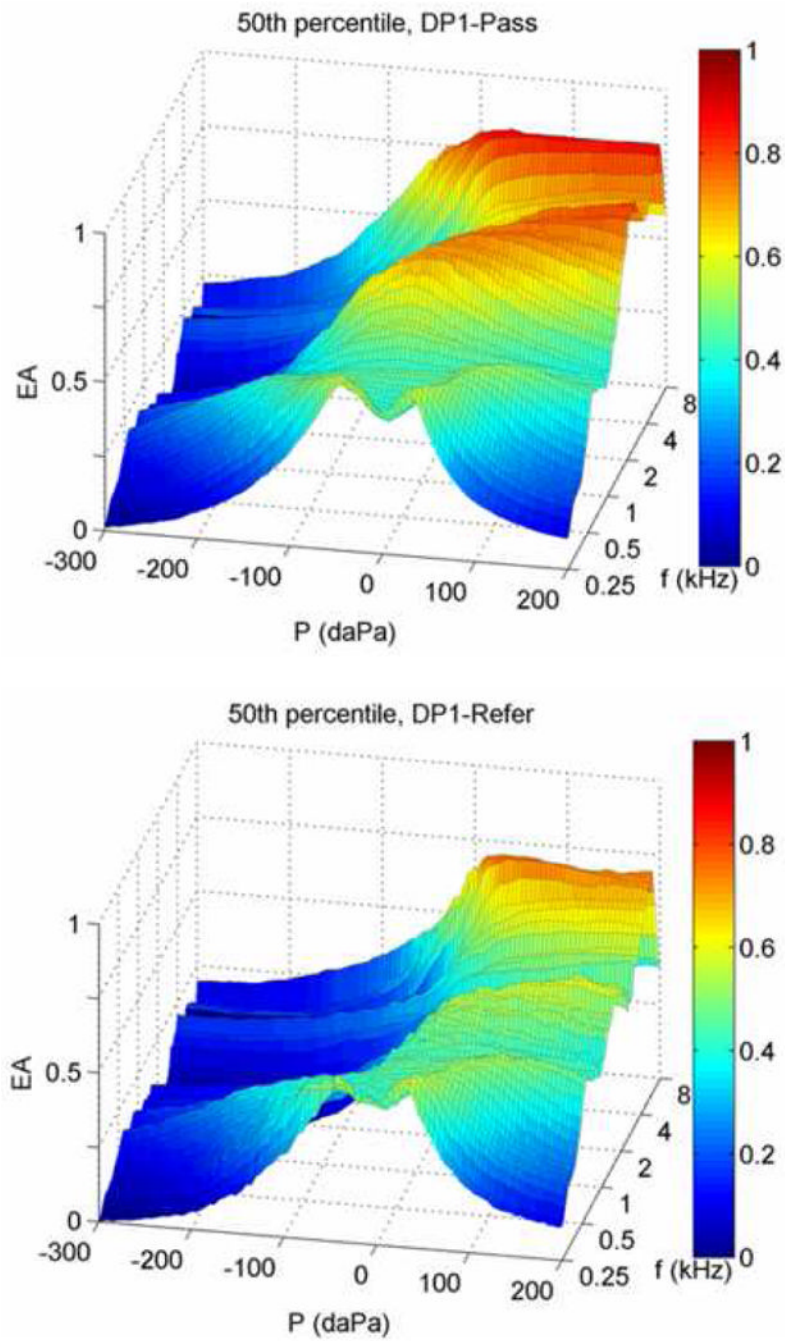
Fig. 4.
Group median tEA responses for DP1-Pass and DP1-Refer groups (top and bottom respectively). The 50th percentile from each group is plotted as EA as a joint function of ear-canal pressure P and frequency f.
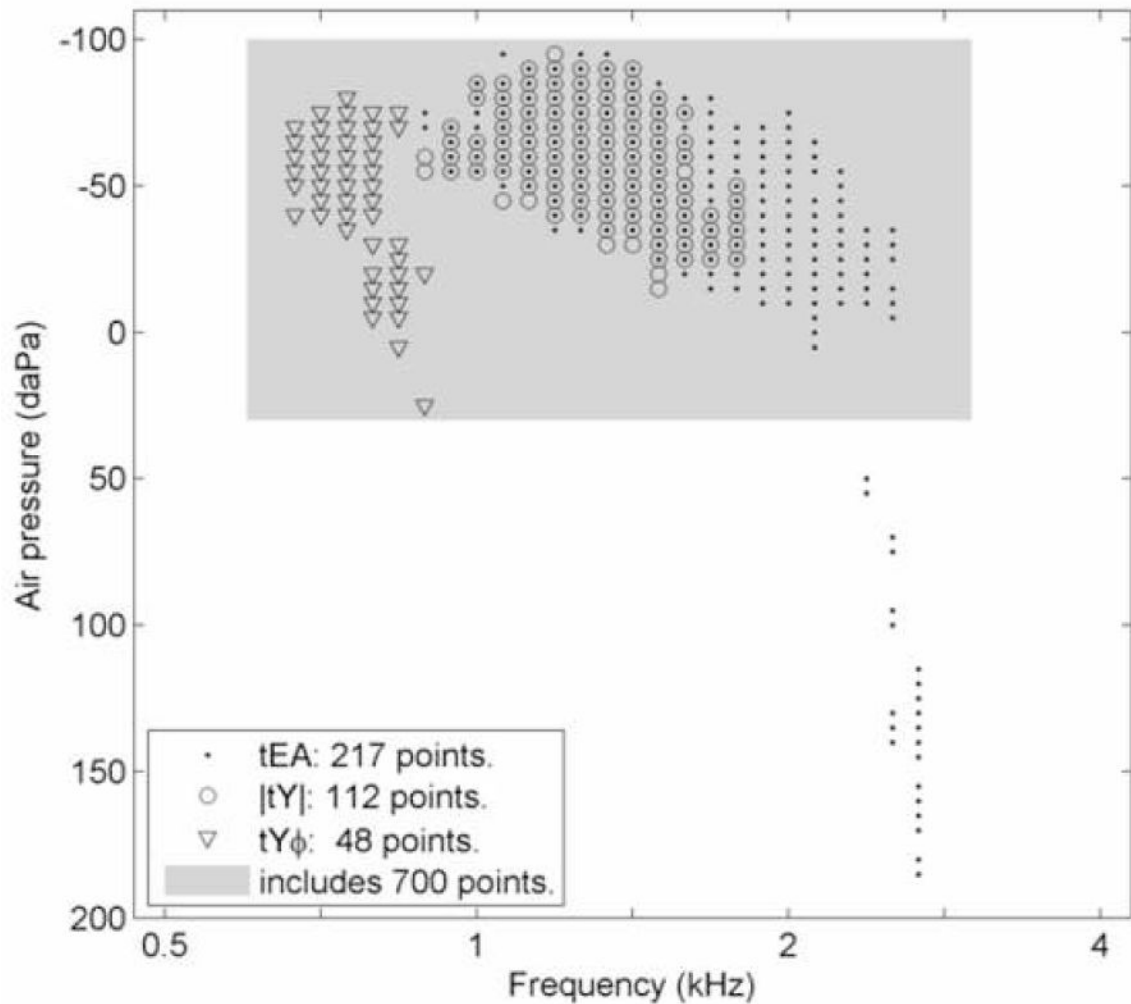
Fig. 5.
A one dimensional function representing an abbreviated frequency-pressure plane for WB tympanograms (−100 to 200 daPa and 0.5 to 4 kHz). Dots, circles, and triangles represent points where the 25th percentile of DP1-Pass ears was higher than 75th percentile of the DP1-Refer ears for tEA, |tY|, and tYφ tests, respectively. The shaded region represents the range of frequency-pressure points from which data were obtained (N=700 points) and which were included in the WB tympanogram likelihood ratio (ρ) tests.
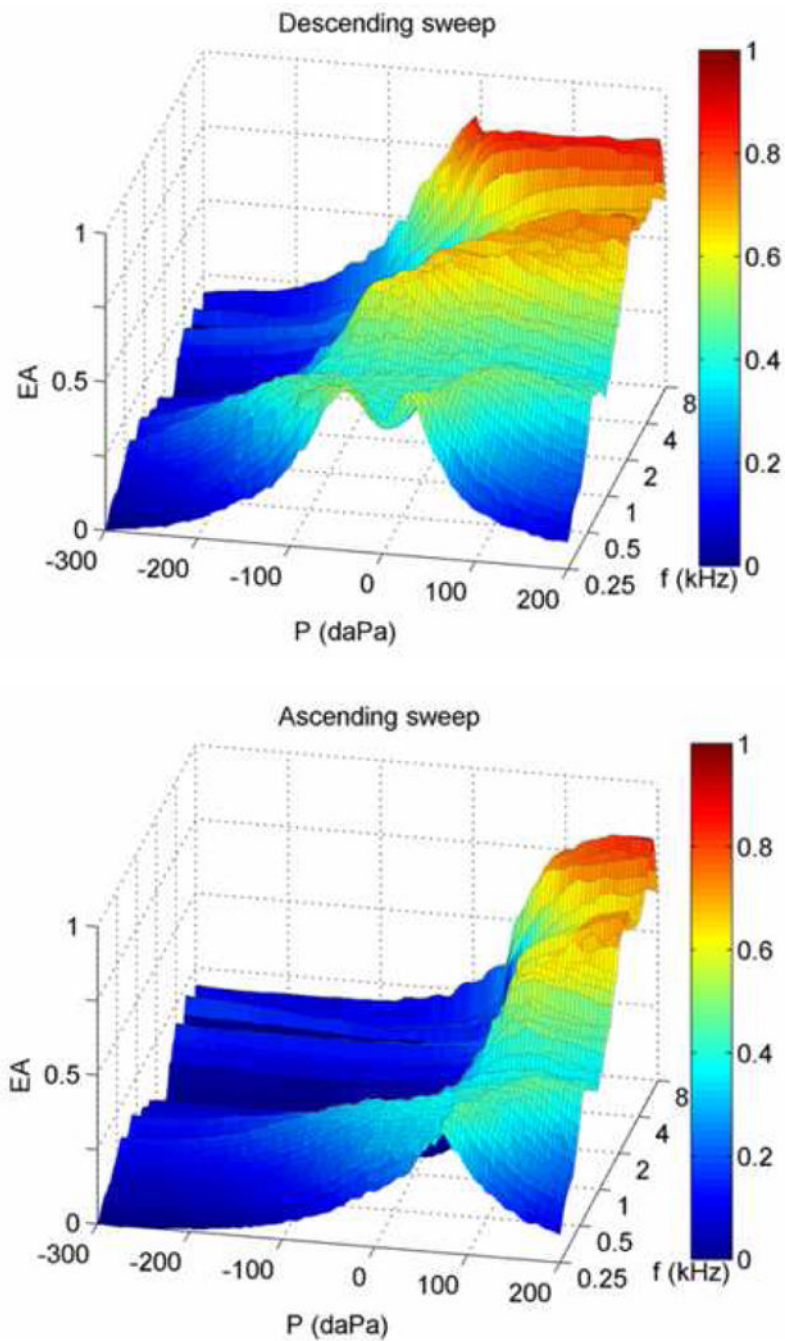
Fig. 6.
Group median WB energy absorbance tympanometry (tEA) responses for a subset of DP1-Pass ears (N=56). Top and bottom panels show descending (+ to −) and ascending (− to +) pressure sweeps respectively. The 50th percentile from each sweep direction is plotted as EA as a joint function of ear-canal pressure *P* and frequency *f*.
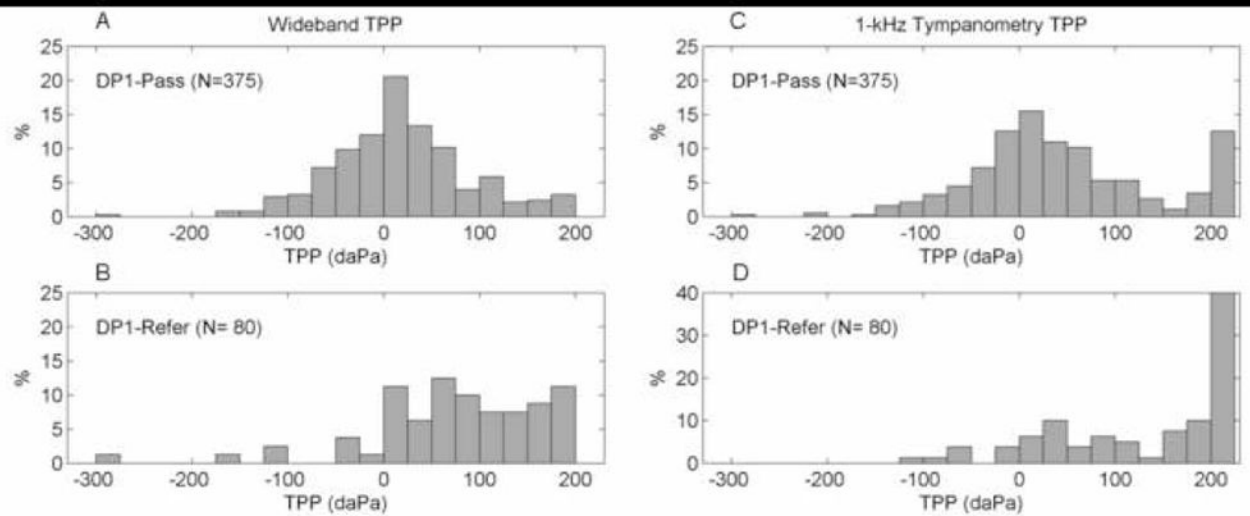
Fig. 7.
Distributions of WB (panels A and B) and 1 kHz (panels C and D) TPP for DP1-Pass and DP1-Refer groups. The vertical bars represent the proportion of subjects with a given TPP as a function of TPP intervals (width of 25 daPa).
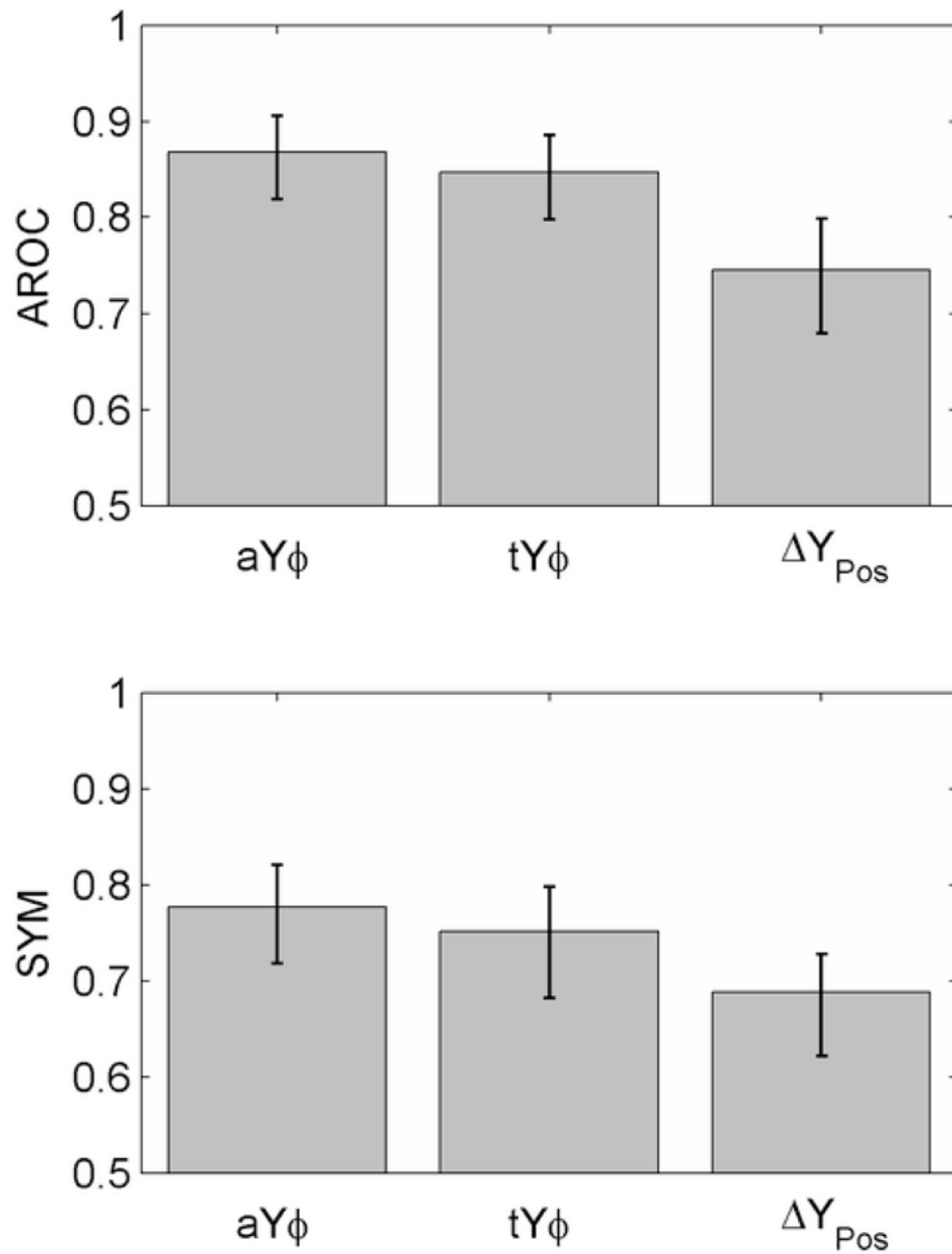
Fig. 8.
Day-1 test performance results for selected WB and 1-kHz predictors. Height of bars show AROC and SYM values (top and bottom panels, respectively). Error bars represent 95% CIs.
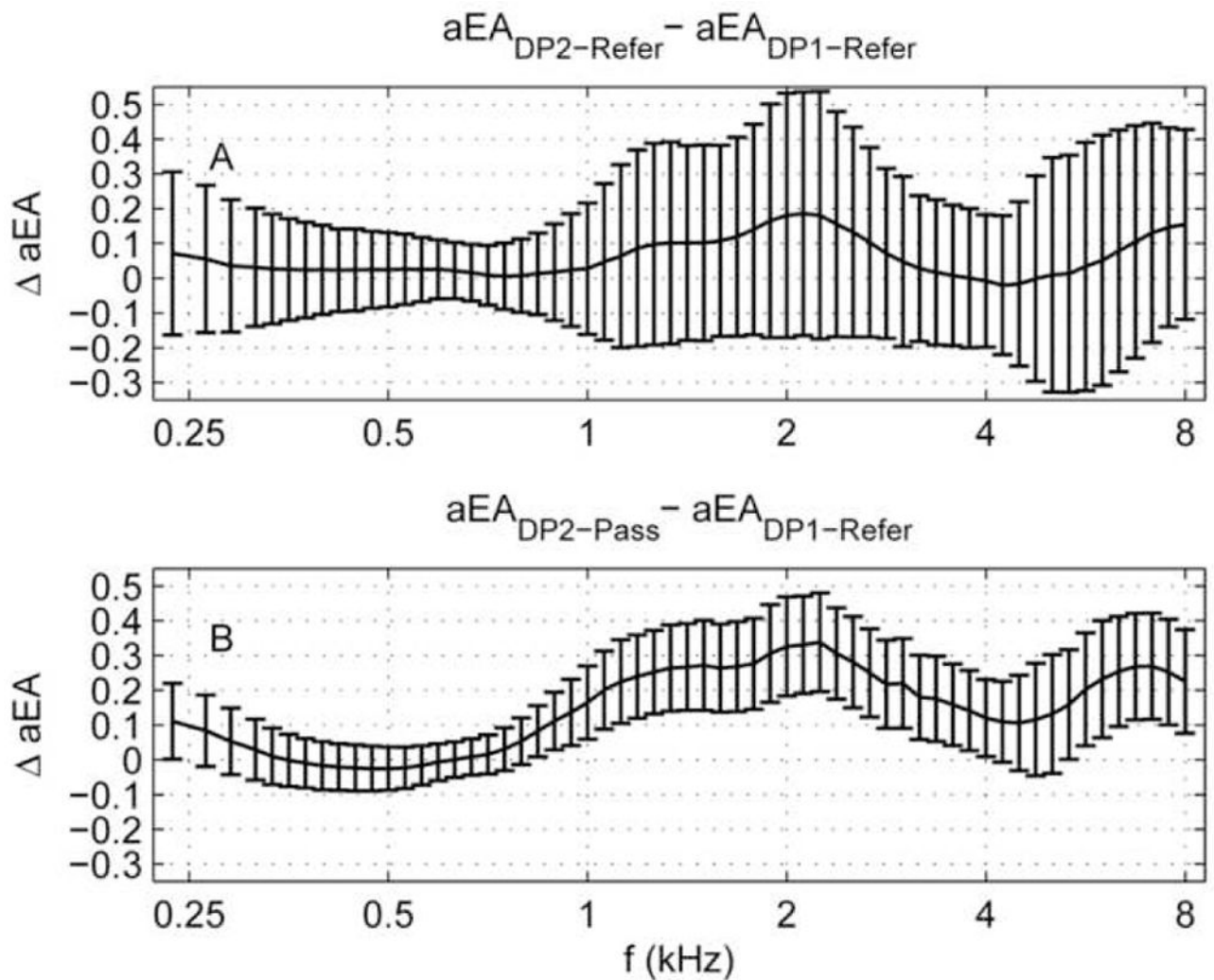
Fig. 9.
Panel A: Mean ΔaEA for the DP2-Refer group (N=14 ears) relative to DP1-Refer group
(N=80 ears), (i.e., it is the aEA shift in the refer groups on Day 2 relative to Day 1). Panel B:
Mean ΔaEA for the DP2-Pass group (N=53 ears) relative to DP1-Refer group (N=80 ears).
Error bars represent the 95% CI as calculated using a two-sided *t*-test for independent
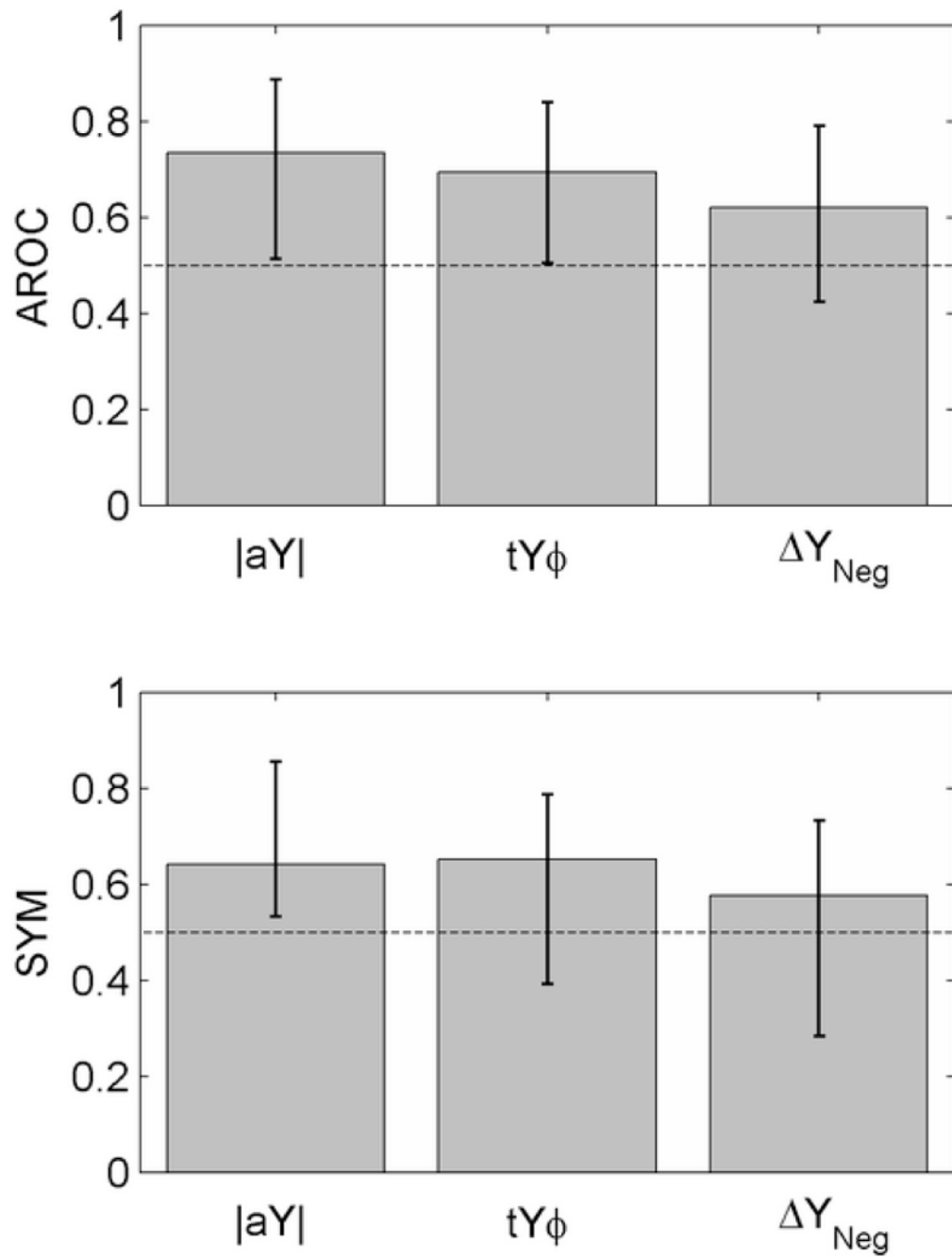samples not assuming equal population variances.

Fig. 10.
Day-2 test performance results for selected WB and 1-kHz predictors. Height of bars show AROC and SYM values (top and bottom panels, respectively). Error bars represent 95% CIs. The horizontal, dashed line represents an AROC or SYM of 50%, or chance performance.
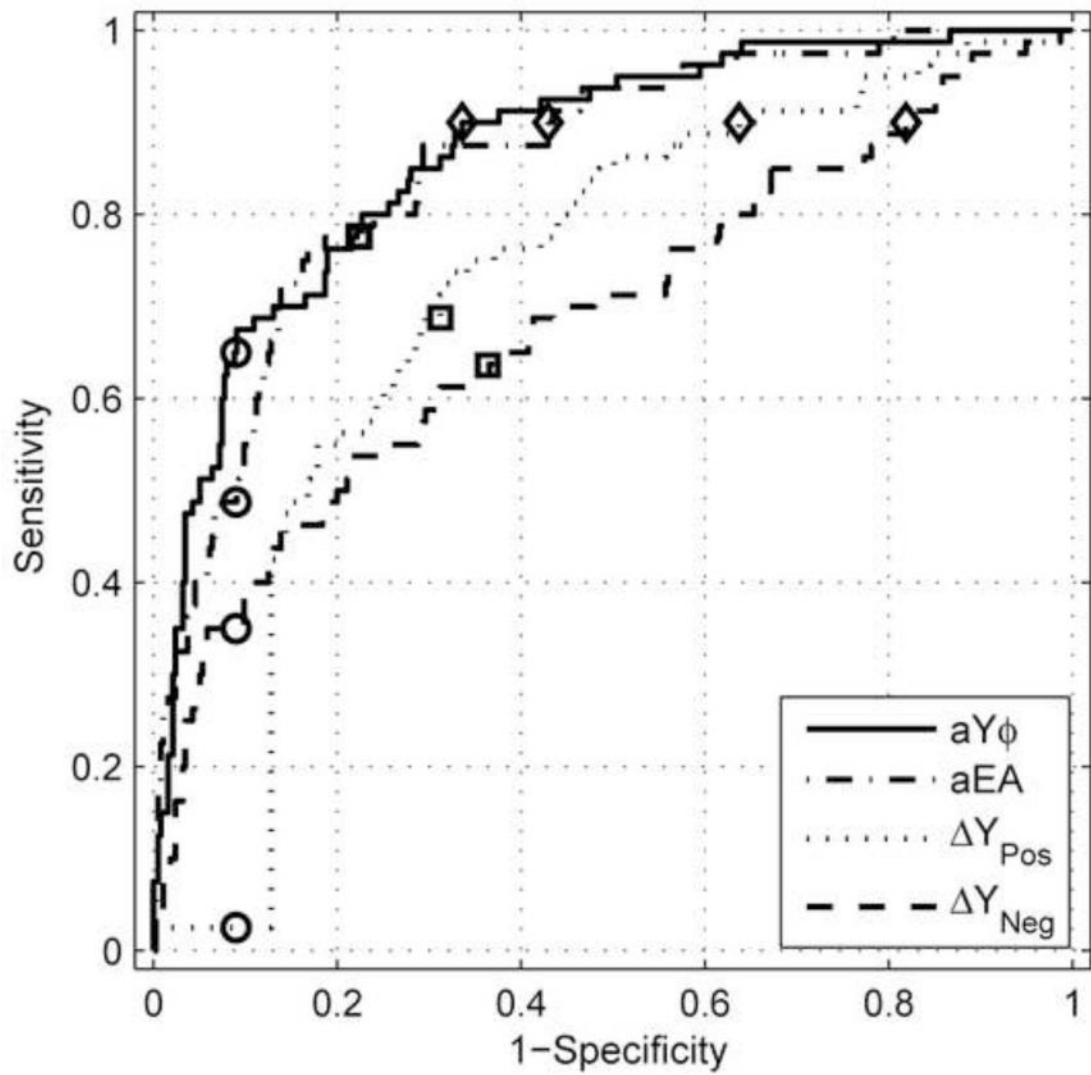
Fig. 11.
ROC curves plotted for two 1-kHz (ΔY_{Pos}, ΔY_{Neg}) and two WB ($aY\phi$, aEA) predictors. Squares represent points on an ROC curve at which the sensitivity and specificity are most nearly equal, referred to as the point of symmetry (SYM). Circles represent sensitivities at a fixed specificity of 91%. Diamonds represent specificities at a fixed sensitivity of 90%.

Table I

Performance of 1-kHz and WB test predictors in classifying DPOAE pass and refer test results for both Day-1 and Day-2 responses. Measures of test performance include AROC and SYM for each predictor, with 95% CI in parentheses.

Predictor	Day 1		Day 2	
	AROC (95% CI)	SYM (95% CI)	AROC (95% CI)	SYM (95% CI)
ΔY_{Pos}*	0.75 (0.68–0.80)	0.69 (0.62–0.73)	**0.54 (0.36–0.71)**	**0.50 (0.36–0.67)**
ΔY_{Neg}	0.69 (0.61–0.75)	0.64 (0.58–0.71)	**0.62 (0.42–0.79)**	**0.58 (0.28–0.73)**
ΔY_{Slope}	0.73 (0.67–0.79)	0.71 (0.62–0.77)	**0.52 (0.34–0.70)**	**0.50 (0.29–0.66)**
aEA	0.86 (0.80–0.89)	0.78 (0.70–0.83)	**0.67 (0.45–0.83)**	**0.64 (0.42–0.81)**
|aY|	0.86 (0.81–0.90)	0.76 (0.71–0.81)	0.74 (0.51–0.89)	0.64 (0.53–0.86)
aYϕ*	0.87 (0.82–0.91)	0.78 (0.72–0.82)	0.72 (0.53–0.86)	**0.64 (0.41–0.84)**
tEA	0.84 (0.78–0.88)	0.76 (0.70–0.82)	**0.64 (0.43–0.81)**	**0.64 (0.40–0.79)**
|tY|	0.82 (0.76–0.87)	0.75 (0.69–0.80)	**0.66 (0.45–0.82)**	**0.64 (0.36–0.80)**
tYϕ*	0.85 (0.80–0.89)	0.75 (0.68–0.80)	0.69 (0.50–0.84)	**0.65 (0.39–0.79)**

Asterisks (*) indicate predictors with the highest AROC on Day 1 and within a particular test class (i.e., 1-kHz, aATF or tATF). All Day-1 predictors exceeded chance performance (lower boundary of AROC and SYM CIs >0.50). Day-2 predictors *not* exceeding chance performance are identified with bold font.