



Published in final edited form as:

Ear Hear. 2010 October ; 31(5): 599–610. doi:10.1097/AUD.0b013e3181e40ca7.

Wideband reflectance in newborns: Normative regions and relationship to hearing screening results

Lisa L. Hunter, Ph.D.,

Audiology Division, Cincinnati Children's Hospital Medical Center

M. Patrick Feeney, Ph.D.,

Otolaryngology, Head and Neck Surgery, V. M. Bloedel Hearing Research Center, University of Washington

Judi A. Lapsley Miller, Ph.D.,

Mimosa Acoustics, Inc.

Patricia S. Jeng, Ph.D., and

Mimosa Acoustics, Inc.

Susie Bohning, Au.D.

University of Utah Hospital

Structured Abstract

Objectives—To develop normative data for wideband middle-ear reflectance in a newborn hearing-screening population, and to compare test performance with 1-kHz tympanometry for prediction of OAE screening outcome.

Design—Wideband middle-ear reflectance (using both tone and chirp stimuli over 0.2 to 6 kHz), 1-kHz tympanometry, and distortion-product otoacoustic emissions (DPs) were measured in 324 infants at two test sites. Ears were categorized into DP-pass and DP-refer groups.

Results—Normative reflectance values were defined over various frequency regions for both tone and chirp stimuli in ambient pressure conditions, and for reflectance area indices (RAIs) integrated over various frequency ranges. Receiver-operating-characteristic (ROC) analyses showed that reflectance provides the best discriminability of DP status in frequency ranges involving 2 kHz, and greater discriminability of DP status than 1-kHz tympanometry. Repeated-measures analyses of variance (ANOVA) established that (a) there were significant differences in reflectance as a function of DP status and frequency, but not sex or ear; (b) tone and chirp stimulus reflectance values are essentially indistinguishable, and (c) newborns from two geographic sites had similar reflectance patterns above 1-kHz. Birth type and weight did not contribute to differences in reflectance.

Conclusions—Referrals in OAE-based infant hearing screening were strongly associated with increased wideband reflectance, suggesting middle-ear dysfunction at birth. Reflectance improved significantly over the first 4 days after birth with normalization of middle-ear function. Reflectance scores can be achieved within seconds using the same equipment used for OAE screening. Newborns with high reflectance scores at Stage I screening should be rescreened within a few hours to a few days, as most middle-ear problems are transient and resolve spontaneously. If reflectance and OAE are not passed upon Stage II screening, referral to an otologist for ear examination is suggested along with diagnostic testing. Newborns with normal reflectance and a

refer result for the OAE screen should be referred immediately to an audiologist for diagnostic testing with threshold auditory brainstem response (ABR) due to higher risk for permanent hearing loss.

Keywords

Infant hearing screening; norms; middle-ear reflectance; tympanometry; otoacoustic emissions

Introduction

Newborn hearing screening programs have become routine and even mandated in many states and countries around the world, with over 95% of all infants now screened at birth in the United States and many developed countries with either otoacoustic emissions (OAE) or auditory brainstem response (ABR), as reported by the Joint Committee on Infant Hearing (2007). The primary goal of newborn hearing screening is to detect permanent hearing loss, to which OAE and ABR are sensitive (Finitzo et al. 1998; Norton et al. 2000), as early as possible, in order to initiate appropriate intervention before 6 months of age. However, these screening tests are also affected by transient outer-ear and middle-ear conditions that are often present at birth (Hall et al. 2004). The effectiveness of universal newborn screening has been evaluated in a meta-analysis of screening and outcome studies (Thompson et al., 2001). From this analysis, good evidence was found for the diagnostic accuracy of OAE and ABR testing, and for the benefit of screening for early identification and diagnosis of hearing problems.

While hearing screening has proven to be effective in the early diagnosis of congenital hearing loss, there is room for improvement in both screening and diagnostic procedures. In order to reduce referral rates, two-stage screening has become standard clinical practice in the U.S. In two-stage screening programs, infants who refer on the first screening by OAEs are tested again with either OAE or ABR, either in the birth hospital or at an outpatient clinic. The referral rate in newborn hearing-screening programs in full term, healthy newborns ranges from 6 to 10% in Stage I, and from 1.3 to 3.4% at Stage II (Thompson et al., 2001). A large, multi-site trial of newborn screening tests with validation by the gold standard of behavioral hearing testing at 8 to 12 months of age reported sensitivity of 80 to 94% for OAE and 84 to 90% for ABR, for moderate and profound hearing loss, respectively (Norton et al., 2000). A statewide study of nearly 70,000 newborns screened in eight New York hospitals over a 3-year period (Prieve et al. 2000) reported overall prevalence of hearing loss was approximately 2/1000, with 8/1000 identified from neonatal intensive care units and 1/1000 identified from well-baby nurseries. Positive predictive value for permanent hearing loss based on inpatient screening was only 4% but increased at outpatient rescreening to 22%. The best evidence of positive predictive value from a meta-analysis of high-quality screening studies is 6.7% , and the percentage of infants who pass newborn screening but are later found to have moderate to profound, bilateral permanent hearing loss (false negatives) is 6 to 15% (Thompson et al., 2001).

While the low incidence of congenital hearing loss is one reason for the low positive predictive value of newborn hearing screening, it is also apparent that many screening refers are not due to false-positives, but are actually true fails due to middle-ear dysfunction. Fluid or material in the newborn middle-ear space, such as amniotic fluid, mesenchyme or meconium has been reported in temporal bone studies (Northrop et al. 1986; Jaisinghani et al. 1999; Miura et al. 2008). These middle-ear contents increase the mass, stiffness, and resistance of the middle-ear space, and therefore alter impedance properties and the sound-conducting efficiency of the middle ear, which can affect hearing screening tests. In addition, there are developmental differences in the neonatal ear that affect impedance

(Keefe et al., 1993; Keefe et al., 2000). The middle-ear and mastoid air space is smaller, the tympanic membrane is thicker and more horizontal, and the outer ear canal is less rigid and tends to collapse more easily (Saunders et al. 1983; Eby and Nadol, 1986; Holte et al. 1991; Ruah et al. 1991; Ikui et al. 1997).

Transient middle-ear conditions can cause screening referrals despite normal cochlear function because forward stimulus transmission and reverse transmission of OAEs through the middle ear are altered. Evidence for improvement in middle-ear status after newborn hearing screening has been reported for air conduction ABR thresholds, which improve by 10 dB in the first 48 hours after birth, while bone conduction ABR thresholds change by less than 1 dB (Stuart et al. 1994). While middle-ear conditions may cause screening referrals, outer-ear conditions also may contribute to over-referral. In ears of newborns with occluding vernix, the pass rate was reported to be 66% for ABR but only 38% for OAE (Doyle et al. 1997).

Co-existing or transient conductive hearing loss (CHL) may delay definitive diagnosis due to uncertainty over diagnosis in the face of apparent middle-ear dysfunction. Vartiainen (2000) reported delayed diagnosis in 36% of infants with congenital or early onset sensorineural hearing loss (SNHL) due to coexistent middle-ear dysfunction. A study of infants referred to a pediatric hospital from newborn screening found that middle-ear effusion was identified in 65%, and ventilation tubes were required in 35% of those before definitive diagnosis could be made (Boone et al. 2005). Further, 79% of infants without effusion were found to have SNHL, while 11% of those with effusion had SNHL.

The prevalence of CHL appears to be higher than SNHL in infants. In the Rhode Island Hearing Assessment Project, 1850 newborns received a two-stage hearing-screening exam (Maxon et al. 1993), and of these, 37 infants were eventually diagnosed with persistent CHL due to otitis media, while 11 infants were diagnosed with SNHL. In the neonatal intensive care unit group, CHL prevalence was 36 per thousand (compared to 16 per thousand in the well-baby nursery), while SNHL prevalence was 23 per thousand (compared to 2.6 per thousand in the well-baby nursery). The prevalence rates did not include infants with transient CHL that resolved between the newborn screen and the later diagnostic test. Another study of diagnostic testing in 170 newborns referred from newborn screening reported that 5% had confirmed SNHL and 10% had CHL (Yang et al. 1993).

Middle-ear assessment is recommended for newborns referred from well-baby nurseries and neonatal intensive care units as part of the diagnostic assessment (Joint Committee on Infant Hearing 2007). Standard tympanometry using a 0.220 or 0.226 kHz probe tone is insensitive to middle-ear dysfunction in newborns (Paradise et al. 1976; Sprague et al. 1985; Marchant et al. 1986; Hunter and Margolis 1992). Other methods of middle-ear assessment, such as pneumatic otoscopy and bone-conducted ABR are useful, but require specialized expertise to perform and interpret in newborns. Tympanometry using a higher probe-tone frequency (e.g., 1 kHz) is recommended for diagnostic testing in infants less than four months old (Joint Committee on Infant Hearing 2007) since it is more sensitive to middle-ear dysfunction than tympanometry at 0.226 kHz (Margolis et al. 2003; Baldwin 2006; Calandrucchio et al. 2006; Alaerts et al. 2007). Some recent studies have reported sensitivity and specificity values for 1-kHz tympanometry using either bone conduction ABR or OAE as the comparison test, since there is no gold standard for diagnosis of middle ear fluid in newborns. Baldwin (2006) compared 1-kHz tympanometry to bone conduction ABR in infants 2 weeks to 19 weeks old with normal ABR (n=104) and with ABR suggestive of CHL (n=107). The tympanogram pass-fail criterion was based on the presence of any peak in admittance above baseline, with resultant sensitivity of 0.99 and specificity of 0.89 compared to the presence of an air-bone gap by ABR. Conversely, the sensitivity and

specificity of 1-kHz tympanometry has recently been compared to wideband reflectance tests in newborns and shown to produce poorer results (Sanford et al. 2009). Thus, there is not consensus about the validity of 1-kHz tympanometry in newborns, and there continues to be a need for effective validated tools for middle-ear assessment in the second stage of screening and in the diagnostic process after newborn hearing screening.

Wideband middle-ear reflectance is a better way to display wideband acoustic impedance measurements. Ear canal wideband acoustic impedance goes back to at least the 1920s, but was made popular in research with the development of the Zwislocki acoustic bridge (Grason-Stadler Model E8872A; Zwislocki, 1962). The transformation from acoustic impedance to reflectance provides a measure relatively insensitive to the transducer depth in the ear canal compared to admittance measurements (Voss et al., 2008). Power reflectance is defined as the ratio of retrograde to incident power. It is typically expressed as a percentage, and as a function of frequency from ~0.2 to ~6 kHz. It can provide a dramatic improvement over tympanometry, which is typically measured at only one or a few frequencies. Wideband power reflectance was first measured in human ears by Stinson et al. (1982). The first clinically practical device was developed for use with animals by Allen (1986), and adapted for measurement in humans by Keefe et al. (1992). Data were first reported for adult humans by Keefe et al. (1993) and Voss and Allen (1994), and for infants by Keefe et al. (1993).

In this study, we investigated an experimental system known as the HearID wideband middle-ear power analyzer (MEPA; e.g., Allen et al. 2005), manufactured by Mimoso Acoustics, Inc. (Champaign, IL). The MEPA system uses a broad-band chirp or tone-complex stimulus to measure reflectance, admittance, and impedance components of the middle-ear system. Stimulus delivery and measurement is via an acoustically-sealed speaker and microphone housed in a probe tip. Wideband measurements of middle-ear function provide an advantage over single-frequency immittance in that the entire frequency range important for hearing threshold measurement and speech perception is assessed using a broadband stimulus. This allows direct comparisons to other measurements made across a broad frequency range, such as for OAE, ABR and behavioral audiometry. Complex immittance measurements such as admittance phase and magnitude can be derived from the single wideband complex reflectance measurement, which provides better diagnostic accuracy than admittance magnitude alone.

Wideband reflectance measurements have been reported in children and healthy infants (e.g., Keefe et al. 1993; Vander Werff et al. 2007; Hunter et al. 2008; Sanford and Feeney 2008), and neonates (Keefe et al. 2000; Abdala et al. 2007; Keefe and Abdala 2007; Shahnaz 2008). Wideband reflectance is sensitive to middle-ear disorders including otitis media with effusion in infants and children, and has shown high test-retest reliability demonstrated by high inter-class correlations (Hunter et al. 2008). Keefe et al. (2003) analyzed wideband reflectance obtained from a two-stage newborn hearing screening protocol (OAE/ABR), which resulted in a 5% false-positive rate. Wideband reflectance measures demonstrated that 80% of the OAE screening referrals had abnormal responses, indicating evidence of middle-ear dysfunction (Keefe et al. 2003). Another study of OAE screening in infants showed significantly higher reflectance between 0.63 to 2 kHz in those infants who failed, compared to those who passed OAE screening (Vander Werff et al. 2007).

Summary of the current study

MEPA was measured in healthy full-term neonates to determine normal wideband reflectance characteristics, and to detect the presence of middle-ear dysfunction and its effects on newborn hearing screening. Test performance for reflectance was compared to 1-

kHz tympanometry using standard DPs as the comparison test. These tests were performed at Stage I screening in the hospital, typically within 48 hours after birth, for normal newborns. In this study, we obtained normative data for reflectance using a research prototype wideband instrument (HearID R4, Mimosa Acoustics, Champaign, IL) in two geographic locations that had diverse patient populations, and we tested repeatability in a subgroup of newborns.

Materials and Methods

The study was approved by the University of Utah, University of Washington and University of Illinois Human Subjects Committees. Infants enrolled in the study were born at the University of Utah Hospital (UUH) or University of Washington Medical Center (UWMC) well-baby nurseries. Parents were informed of the study and given the opportunity to enroll their child after birth but before discharge from the hospital. At the UUH site, there was preferential enrollment for babies who had failed their standard infant-hearing screen (using transient-evoked OAEs). This was done in an attempt to increase the chances of finding cases of CHL and SNHL. A consent form approved by the institutional review board was used to obtain parental permission.

Subjects

In total, 397 babies from the UUH site and 128 subjects from the UWMC site were screened and tested. A subset of 324 babies was selected for the analyses presented here (see Table 1). Babies were excluded from the analyses for incomplete data sets or because they were in an age range with an inadequate number for analysis. Thirty-five infants were also tested twice within a few hours to a few days (the first measurement was used for normative analyses, and the subsequent measurements used to consider temporal changes in reflectance).

A record was obtained of demographic factors (including gestational age, birth type, ethnicity, head circumference, birth weight, and Apgar scores) and factors known to be associated with congenital hearing loss (family history of hearing loss, craniofacial anomaly, syndromes associated with hearing loss, ototoxic medications, mechanical ventilation, and congenital infections). Risk factors were rare, as all babies were considered “well babies”. Demographic information was collected to help break down the norms, and also to help classify any baby that was found to subsequently have a SNHL.

Equipment & Stimuli

The HearID R4 system (Mimosa Acoustics, Inc., Champaign, Illinois) was used for the wideband MEPA and DP measurements. It consists of a laptop-hosted PC-card, connected to an ER-10C probe (Etymotic Research, Elk Grove Village, Illinois) with a probe-adaptor cable, and a calibration cavity set of four cavities. HearID R4 is a research tool and was under active development throughout the course of the study. It was designed to quickly and automatically guide the tester through a series of MEPA and DP measurements with a predetermined test sequence and protocols within the same test session without having to exit one measurement to start the second measurement. This reduced the testers' equipment interaction and allowed the set of measurements made within a test session to be automatically kept together in the patient database. The UUH site used versions 4.4.18 through 4.4.72 and UWMC used versions 4.4.28 through 4.4.72. The differences among the versions primarily affected the data recorded to file and did not affect the measurements themselves.

Other than some differences in the user interface, the MEPA module was essentially the same as Mimosa Acoustics' commercial MEPA product (HearID+MEPA R3), except that for each MEPA trial, two measurements were made – one for each probe speaker. The DP module, however, had some differences to Mimosa Acoustics' commercial DP products (DP2000 and HearID+DP R3). The biggest difference was that the noise level was estimated on-frequency (from the variance in the $2f_1 - f_2$ frequency bin, rather than the average of the six neighboring frequency bins). To shorten test time, the in-the ear calibration information for the DP test was derived from the preceding MEPA measurement, provided that the measurement had sufficient signal-to-noise ratio (SNR), otherwise the usual in-the-ear calibration was conducted. The in-the-ear calibration ensured the stimulus levels were delivered on target by taking individual ear canal geometry at the probe tip and probe fitting into account.

Ear-tips—The ER10C probe tubes were covered with a silicone rubber or foam tip to allow comfortable and stable insertion in the ear canal. The same ear-tip was used for both MEPA and DP measurements; however, in some cases, the probe was refitted in the ear canal between measurements if it appeared to have dislodged or the measurement was noisy. MEPA measurements depend upon the cross-sectional area of the ear canal in the plane of measurement. HearID estimated this area physically, based on the diameter of the ear tip used for the measurement. For the present study, four different tip sizes were used: ER10C-03 (4.3 mm), ER10C-04 (4.8 mm), ER10C-05 (5.3 mm), and ER10C-14B (child foam tip, 6 mm). The largest tip that provided a snug seal in the ear canal was selected to avoid acoustic leaks. The silicone rubber tips are relatively incompressible in an infant ear, but the ER10C-14B foam tip is compressed and then conforms to the size of the ear canal. Thus, an error may occur in the estimation of ear-canal area for a foam tip in these larger ears where it was used. However, Keefe et al. (1993) demonstrated that reflectance measurements were relatively insensitive to estimates of ear canal area with errors as high as 20% creating at most a 0.05 change in reflectance. The foam tip was rarely used, as the rubber tips were more appropriate in size for the newborn ear.

MEPA probe calibration—The probe was calibrated daily with HearID (and at any time when there were concerns about the probe status, e.g. if a probe tip was plugged with debris) using the Mimosa Acoustics 4-C Calibration Cavity Set (Voss and Allen 1994). The Thévenin equivalent parameters, source pressure and impedance were estimated from the pressure response of all four cavities.

MEPA in-the-ear calibration—Prior to each MEPA test session, an in-the-ear calibration using a 1-kHz tone was run to establish the overall level for outputting the chirp and tonal stimuli.

MEPA reflectance measurement—Within a test session, four MEPA measurements were made: one for each probe speaker and one for each stimulus type (chirp and tone). An in-the-ear calibration with a 1-kHz tone was used to adjust the overall level for the chirp stimulus to the desired level. The wideband chirp stimulus was output repeatedly at 60 dB SPL for 4 to 6.4 seconds (depending on stopping rules and noise levels), time-averaged, and analyzed within the 0.2 to 6 kHz frequency range. For the tone stimuli, the previous chirp measurement was used as the in-the-ear calibration to adjust the output level for each tone. The 11-tone series was output simultaneously (0.2, 0.3, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 3.0, 4.0, and 6.0 kHz) at 60 dB SPL for 1.6 to 6.4 seconds, terminating early if the measurement had a sufficiently low noise level (<0 dB SPL) and sufficiently high SNR (SNR >10 dB). These measurements from the subject's ear canal and the associated Thévenin equivalent

parameters from the probe calibration were used to calculate wideband reflectance, as described by Voss and Allen (1994).

DP measurement—DPs were measured at $f_2=2, 3, 4,$ and 6 kHz in descending order. The f_2/f_1 ratio was 1.2, and the intensity levels of the primary tones, f_1 and f_2 , were set at 65 and 55 dB SPL at each frequency. The primary tones were output for up to 8 seconds, terminating early if the DP had a sufficiently low noise level (< 0 dB SPL) and sufficiently high SNR ($\text{SNR} > 10$ dB). It was important to distinguish between ears that “referred” for technical reasons (e.g., high noise floor) and those that “referred” for low DP amplitude. Therefore, for *post hoc* analyses, the DP status at an individual frequency was considered (a) *present*, if the DP was -10 dB SPL and had a $\text{SNR} \geq 6$ dB, the noise level was < 0 dB SPL, and the stimulus levels were within ± 3 dB of target; (b) *absent*, if the DP was < -10 dB SPL, the noise level < 0 dB SPL, and the stimulus levels were within ± 3 dB of target; or (c) *status unknown/error*, typically due to high noise levels or off-target stimulus levels. The overall DP status for an ear was a “Pass” if at least 3 of the 4 frequencies were flagged as “Present”, “Refer” if at least 2 of the 4 frequencies were flagged as “Absent”, otherwise it was status “Unknown/Error” (the pass/refer criteria were based on Gorga et al. 2000). It was possible for an ear to achieve an overall Pass/Refer status even if there was an unknown/error status at one frequency, for a Pass, or two frequencies, for a Refer.

Tympanometry—A GN Otometrics Otoflex multi-frequency acoustic immittance instrument (GN Otometrics, Denmark) was used to obtain tympanometry. Volume calibration was performed for tympanometry daily. A 1-kHz probe tone was used for tympanometry. Pressure was changed from the positive to negative direction, and varied from $+200$ daPa to -400 daPa. The pressure rate of change was 400 daPa/sec. Admittance (Y) tympanograms were obtained and values for tympanometric peak pressure, and equivalent ear-canal volume was measured using the value at the negative tail to compensate for ear-canal volume. A cut-off criterion for normal admittance of 0.6 mmho was used as recommended by Margolis et al. (2003).

Procedures

Testing took an average of fifteen minutes per baby for all tests including tympanometry in both ears, but varied from approximately 5 to 30 minutes, depending on infant wakefulness and fussiness, as well as difficulties maintaining probe insertion in some ears. If the infant was fussy at the beginning of the test session, or if it became apparent that the infant was not quieting for adequate testing, the test was discontinued and attempted later. Infants were tested swaddled in a blanket in their bassinets, in a separate room from the nursery (UUH) or in parents’ single-occupancy recovery rooms (UWMC) to provide a quiet environment. They were tested after feeding while in natural sleep or in an awake and quiet state. Pacifiers were used if needed to sooth the infants.

Wideband MEPA, DP, and tympanometry tests were completed for both ears of each infant in this order. The most accessible ear was tested first. Testing was performed once the child was as quiet as possible. Tympanometry was done last, as in our experience, the pressurization, need for a tight seal, and louder probe tone as well as the stimulus for reflexes was more likely to awaken the infant than reflectance or DP measurements. This order of testing in pilot studies resulted in the best overall completion rate for the entire test battery.

Testers were trained in the typical response expected in low noise and adequate probe fit conditions for all three measures. If tracings were noisy or unrepeatable, or if low-frequency

responses were below 50% reflectance, they were instructed to repeat the test after improving the probe fit.

Data screening, cleaning, and reduction

There were the expected difficulties with keeping noise levels low while testing infants. Within the HearID system, it was possible to repeat tests that were noisy and testers often chose to do so, but sometimes it was not possible to quiet the infant sufficiently. In the measurement software, aside from the graphical feedback, there were no warnings to the tester as to whether levels were acceptable, so data were *post hoc* screened to remove tests with high noise and off-target stimulus levels.

Because each ear could have a different outcome, ears were treated as separate entities, and infants may have provided valid data for just one or for both ears. This maximized the available data, but analyses need to be interpreted knowing there is some correlation between measurements made in an infant's left ear and right ear.

The aim of data screening was to find for each ear one quality MEPA measurement (either a chirp-stimulus or a tone-stimulus or both) and one quality DP measurement that had been measured in the same test session (i.e., measured closely together in time, with, ideally, the same probe placement in the ear canal, however, this could not be guaranteed). There were often multiple measurements per ear, which could occur one of two ways: repeated within a specific measurement session and by rerunning the entire test session. Also, for MEPA measurements there were always at least two measurements: one for each channel in the probe. As a first pass through the dataset, test sessions were manually identified that had both a MEPA and a DP measurement and had no tester comments indicating a problem with the test. If there were no comments, the first test was chosen. Within a test session for DPs, the first DP measurement with an overall "Pass" result was automatically chosen, otherwise the first measurement with an overall "Reject" result was automatically chosen (after excluding tests with status "Unknown/Error"). Within a test session for MEPA, the best chirp-stimulus measurement and the best tone-stimulus measurement were automatically selected using an algorithm. Specifically, (a) the SNR below 1 kHz had to be greater than 10 dB for over half the tested frequencies, (b) reflectance for each channel within a measurement could not be separated by more than 5 percentage points for frequencies above 1 kHz, and (c) for measurements meeting these criteria, the measurement with the highest SNR between 1 to 6 kHz was chosen. After processing, out of range reflectance values (> 105%) were treated as missing. Allowing some reflectance values >100% meant noisy, but most-likely valid, high-reflectance measurements could be included.

The minimum dataset to be included in the subject count and overall demographics was (a) in at least one ear, either a tone or a chirp reflectance function (for the subset of tone frequencies only) with acceptable data (reflectance <105%) at 2 kHz and no more than one frequency with missing data between 0.4 and 6.0 kHz, and (b) a technically valid DP pass or DP refer result for the same test session and ear. Because tone-stimuli produce higher SNR with less noise, the dataset for reflectance using a tone stimulus was much larger than that for the chirp stimulus.

For tympanometry, multiple measurements were made if there was artifact or if the trace appeared to have ear canal collapse, and the test tip was changed as needed to obtain a tight seal. The best measurement (without artifactual notches or ear canal collapse and lowest noise) was selected for analysis. While the tympanometer provided automated measurements, all traces were manually measured to ensure that automated measurements for ear canal volume and static admittance were valid.

Results

Overview: Normative regions were defined across the wideband reflectance spectrum for both tone and chirp stimuli, and for integrated frequency ranges. Receiver-operating-characteristic (ROC) analyses showed that reflectance provided the best discriminability of DP-status in frequency ranges involving 2 kHz, and greater discriminability of DP status than 1-kHz tympanometry. Further analyses showed that reflectance decreased (improved) with age. Birth type and weight did not contribute to differences in reflectance. Before conducting the main analyses, key independent variables were identified with repeated-measures analyses of variance (ANOVA), which demonstrated (a) that tone and chirp stimuli produced equivalent reflectance values, (b) that data from the two sites were equivalent for frequencies above 1-kHz, and (c) that DP status and frequency, but not sex or ear, determined differences in reflectance.

Chirp and tone stimuli produce essentially the same reflectance value

A repeated-measures ANOVA (stimulus type: tone, chirp; 9 frequencies: 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 3.0, 4.0, and 6.0 kHz), for 331 ears, showed a significant main effect for frequency ($F_{8,2640} = 214.93$, $p < .05$), but not stimulus type ($F_{1,330} = 0.46$, not significant). There was a significant stimulus type by frequency interaction ($F_{8,2640} = 5.49$, $p < .05$). A Bonferroni *post-hoc t*-test ($p < .05/9 = 0.006$ for 9 comparisons) showed two significant same-frequency differences in reflectance, but the actual numerical differences were minor, and probably not clinically significant. The tone stimulus was 1.5 percentage points higher at 0.4 kHz and 1.3 percentage points lower at 2.0 kHz. All other differences were less than 1 percentage point. Because there was little-to-no difference between the chirp and tone stimuli, data for just the tone stimuli were used for most subsequent analyses as the data set was much larger for tone stimuli (see Table 1).

Reflectance across test site is different only below 1 kHz

To establish if there were differences in reflectance due to test site, a repeated-measures ANOVA (9 frequencies: 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 3.0, 4.0, and 6.0 kHz) on the tone-stimulus dataset, with site as a categorical factor (site: UUH, UWMC), was run on a subset of 99 ears (42 from UWMC and 57 from UUH) for which various demographic and methodological factors were controlled. Specifically: only ear-tip size 04, only DP-pass status (due to different pass/refer rates), and age between 24 to 48 hours (UWMC babies tended to be older at testing). There was a significant main effect for site ($F_{2,604,542} = 4.81$, $p < .05$) and for frequency ($F_{8,776} = 136.21$, $p < .05$), and a significant Site \times Frequency interaction ($F_{8,776} = 9.61$, $p < .05$). Bonferroni *post hoc t*-tests showed that the significant on-frequency differences between sites were limited to 0.4 and 0.6 kHz; however, the trend for lower reflectance at the UWMC site extended to 0.8 kHz too. Thus, subsequent analyses that used the combined dataset were restricted to 1 to 6 kHz.

Reflectance differs for DP-status and frequency, but not sex or ear

A repeated-measures ANOVA (frequency: 1.0, 1.5, 2.0, 3.0, 4.0, and 6.0 kHz) on the tone-stimulus dataset for 479 ears, with three categorical factors (DP-status: pass, refer; sex: female, male; ear: left, right), showed significant main effects for all factors and numerous interactions (see Table 2). Bonferroni *post hoc t*-tests were used to examine 28 same-frequency interactions of interest (family-wise significance level: $p < .05/28 = 0.002$). Across-frequency interactions were not of interest, but were expected given that reflectance is known to vary as a function of frequency.

Sex was not a significant factor in any three or four-way interaction, but it was a significant factor in the two-way Sex \times DP-status interaction. There are four comparisons of interest –

whether there is a difference in reflectance between males and females for pass and refer DP-status (separately) and whether there is a difference in reflectance between DP-status for males and females (separately). The across-status/across-sex interactions were not of interest. For females and males separately, there were significant differences for DP-status (reflectance values for females with DP-pass status were 28.4 percentage points lower than for females with DP-refer status; reflectance values for males with DP-pass status were 21.4 percentage points lower than for males with DP-refer status). For DP-status separately, there were no significant differences found between males and females. Therefore, DP-status was a significant factor, but not sex.

Ear was not a significant factor in the four-way interaction, but it was for one three-way interaction: Frequency \times Ear \times DP-status, and one two-way interaction: Ear \times DP-status. In the three-way interaction, the comparisons of interest were ears at the same frequencies, separately for DP-pass and DP-refer, then DP-status for the same frequencies, separately for each ear – 24 comparisons in total. In the breakdown of DP-status by ear for each frequency, reflectance for right ears with DP-pass status were significantly lower than right ears with DP-refer status. Likewise, reflectance for left ears with DP-pass status were significantly lower than left ears with DP-refer status. In the breakdown of DP-status by ear for each frequency, for the DP-pass group, there was no difference in reflectance between left and right ears. For the DP-refer group, there was a trend for left ears to have lower reflectance than right ears, especially at 6 kHz, but there were no significant differences. Therefore, ear was not a significant factor. DP-status was a significant factor within ear and across frequency, but because ear was not a factor, the remaining interaction of interest was DP-status by frequency.

Frequency \times DP-status (see Figure 1): From the *post hoc* analyses done so far, it was established that there were significant differences in reflectance at every frequency between the DP-pass and DP-refer groups. Going to the main effect for DP-status, there was an overall significant difference in reflectance of 25 percentage points (DP-pass = 39.9%, DP-refer = 64.8%). The biggest difference in reflectance was at 2 kHz, with the DP-pass group 34.7 percentage points lower than the DP-refer group.

Because ear, sex, stimulus type and site factors did not contribute significantly to any same-frequency interactions, they were dropped from subsequent analyses including development of normative regions.

Normative reflectance regions for predicting DP-status

The normative reflectance region for normal newborns was defined using the DP-pass group, because the presence of a DP depends on efficient stimulus transfer and reverse transmission of the OAE through the middle-ear system. The use of DP to define normative ranges is not intended to imply a gold standard comparison, as DP may be absent for other reasons, such as cochlear dysfunction, and DP may be present despite minor middle-ear dysfunction. However, it provides the best comparison test available in newborns without resorting to invasive procedures, such as myringotomy, that carry risk and are not ethical in otherwise healthy newborns. Tympanometry was not chosen for the comparison test, because a primary goal of the study was to compare effectiveness of tympanometry against wideband reflectance for prediction of newborn hearing screen status. Percentile rankings for reflectance were calculated at each frequency, separately for the tone-stimulus and the chirp-stimulus datasets. There were also enough ears to define a reflectance region for newborns with a DP-refer result. Figure 2 shows the distributions for 0, 10, 50, 90, and 100th percentiles for both DP-pass and DP-refer groups for the chirp-stimulus dataset.

The DP-pass group tended to have much lower reflectance than the DP-refer group; however, these two distributions overlapped, introducing a region of uncertainty. In Figure 3 the distributions from Figure 2 were used to plot three regions for the chirp-stimulus data: pass, ambiguous, and retest. The pass region was defined by the 0th percentile of the DP-pass group and the 10th percentile of the DP-refer group. The retest region was defined by the 100th percentile of the DP-refer group and the 90th percentile of the DP-pass group. The ambiguous region was therefore defined by the 90th percentile of the DP-pass group and the 10th percentile of the DP-refer group. Overlaid on the plot in Figure 3 are the equivalent percentiles for the tone-stimulus dataset at the six test frequencies. The tone-stimulus percentiles are similar, but not identical to the chirp-stimulus percentiles, partly because they are based on data from more ears.

Reflectance Area Index (RAI) summarizes reflectance as a function of frequency

In many clinical situations, it may be difficult to judge middle-ear status if the reflectance function falls across more than one normative region for different frequency regions. Furthermore, some frequency regions may prove more diagnostic than others. An alternative approach is to use a Reflectance Area Index (RAI) where the reflectance values are integrated (averaged) over a specified frequency range. A RAI can be applied to both the continuous chirp-stimulus reflectance function and the discrete tone-stimulus function (individual points). RAI has the same unit (percentage) as reflectance. Figure 4 shows RAI normative regions for various frequency ranges, defined similarly to the reflectance norms, using the chirp-stimulus dataset (N = 347).

Table 3 provides the percentiles underlying the normative regions for the tone-stimulus and RAI frequency ranges.

Frequency regions involving 2 kHz best discriminate between the DP-pass and DP-refer groups

ROC curves show to what extent two distributions overlap (in this case DP-pass and DP-refer). The further apart the distributions, the higher the ROC curve and the greater the area under the ROC curve.¹ Areas (see Table 4) were calculated for the reflectance distributions underlying the tone-stimulus norms in Figure 3 and for the eight RAI norms from Figure 4. All areas were significantly different from chance (Bamber 1975).² Regions involving 2 kHz provided the greatest discriminability, particularly 1 to 2 kHz, 1 to 4 kHz, and 2 kHz. The two ROC areas for the RAI at 2 kHz – one for each stimulus type – showed equal discriminability (i.e., ROC area = 0.9), even though the chirp-stimulus RAI was based on data from just one frequency bin in the chirp-stimulus spectrum. Including the frequencies below 1 kHz did not change discriminability, as the ROC area for the wideband frequency range of 0.2 to 6 kHz was identical to the ROC area for 1 to 6 kHz.

Reflectance discriminates between DP-pass and DP-refer better than 1-kHz tympanometry

An important clinical question is how reflectance compares with other middle-ear tests in terms of predicting DP outcomes in the newborn population. In Figure 5, ROC curves for 1-kHz tympanometry (Y1) were directly compared with ROC curves for tone-stimulus reflectance at 1 kHz (same frequency for direct comparison) and 2 kHz (best reflectance frequency for discriminating DP-pass from DP-refer) in the same ears for the subgroup of newborns with both valid tympanometry and reflectance measurements (DP-pass 318 ears; DP-refer 127 ears). Flat tympanograms were assigned a Y1 value of zero. The

¹The area under the ROC curve is based on the separation of the distributions themselves, not the separation of their means (Lapsley Miller et al. 2002).

²Chance performance is defined by the negative diagonal on the ROC curve and by an ROC area of 0.5.

corresponding areas under the ROC curves were 0.72 for Y1, 0.82 for reflectance at 1 kHz, and 0.90 for reflectance at 2 kHz. These three ROC areas were also significantly different from one another using Hanley and McNeil's (1983) method for comparing ROC areas derived from the same cases (Y1 vs. 1-kHz Reflectance: $z = 3.15$; Y1 vs. 2-kHz Reflectance: $z = 5.02$; 1-kHz Reflectance vs. 2-kHz Reflectance: $z = 2.18$).

DP-pass rate increases with age due to decreasing reflectance over time

Newborns in this study were all younger than 5 days old at testing, and most were tested within the first 24 to 48 hours after birth. It is of clinical interest to know if reflectance changes over time during this short time period, and whether there is an optimal time for infant hearing screening in the newborn population. Ears were categorized into ten age-groups, which were of uneven length in an attempt to achieve an even spread of counts per cell (see Figure 6 for counts). The DP-pass rate, as a percentage, increased over time (Figure 6).

The increase in the DP-pass rate can be at least partly ascribed to a decrease in reflectance over time. Figure 7 shows reflectance at 2 kHz (for the tone-stimulus dataset) plotted as a function of the age categories defined above.

Reflectance norms were not generated for the various age groups for a variety of reasons. The number of ears was too few in each age band, especially for the DP-refer group; age at testing was not randomized; there were potential confounds and selection biases when looking by age due to differences in test procedures between the sites (UWMC newborns were tested at a slightly older age; UUH newborns had more refers due to preferential enrollment for newborns who had received a refer on the hospital hearing screening test); and because caesarian-delivered babies tended to be tested at an older age.

A subset of newborns at the UUH site (54 ears, 35 babies) received a second round of testing in order to examine test-retest differences using the same test protocol, a few hours to a few days later. This allowed a second look at age effects, this time using a more powerful within-subjects design, albeit with a smaller number of ears. Using the best tone-stimulus frequency of 2 kHz, reflectance at the second test was on average 23 percentage points lower than at the first test ($t = 8.7$, $df = 53$, $p < .05$).

No differences in reflectance due to birth type or to size of newborn

There was no difference in reflectance due to birth type (vaginal or caesarian), when controlled for age at testing.³ There was no difference in reflectance due to the size of the baby (specifically, the outer head circumference and weight), although the N was small for many cells.

Discussion

The results of this study provide wideband reflectance normative regions and reflectance area indices (RAI) for newborns with normal hearing (who received a *pass* result for the OAE screen) and for newborns with abnormal results (who received a *refer* result) for both tone and chirp reflectance stimuli and a variety of frequencies/frequency ranges.

³The UUH site already had an informal procedure in place whereby caesarian-delivered babies were often tested at an older age as anecdotal evidence indicated this group had a higher referral rate. In addition, the hospital-stay duration was generally longer for caesarian-delivered babies compared with vaginal births.

Reflectance is predictive of DP pass/refer status

Wideband reflectance was significantly different in ears that referred on the DP test compared to ears that passed the DP test, and showed strong test performance for prediction of DP status. At 2 kHz and frequency regions involving 2 kHz, wideband reflectance differentiated between ears that subsequently got a DP pass or refer result, with areas under the ROC curves of 0.83 to 0.90 (Table 4). The best performance was 0.90 for tone and chirp stimuli at and around 2 kHz. These results are similar to those reported recently by Sanford et al. (2009), using a different experimental wideband reflectance system in a group of 230 well babies (of these, 80 ears did not pass DP), who reported an ROC area of 0.87 for an ambient wideband test. Evidence for middle-ear dysfunction as a primary cause of DP referrals is mainly due to this strong relationship, as wideband reflectance is purely a measure of passive middle-ear energy transfer unrelated to cochlear function. Additionally, as reflectance improved over a matter of days in newborns tested twice, OAE pass rates also improved, demonstrating a strong relationship between changes in DP status and reflectance.

Strengths and weaknesses of the current study

While similar results have been recently reported using a different wideband reflectance system (Sanford et al., 2009), it is important to compare normative data and test validity in different instruments and in different patient populations with ethnic diversity. The current study has several strengths.

Nearly 500 ears were included in the analysis of DP prediction, with a large number in the DP-refer group ($n = 141$, ~28%) due to the study design that preferentially enrolled ears that referred on initial transient-evoked OAE screening. Thus, the statistical analyses were well-powered to detect middle-ear dysfunction and to construct receiver-operator characteristics of test performance.

Two MEPA stimulus types, tone and chirp, were used in each test session with the same probe-fit, allowing direct comparison. In ears producing usable data, these stimuli provided essentially identical results, so either stimulus type may be used to assess middle-ear function in newborns. Similar results were reported in older infants and children tested in an outpatient clinic by Hunter et al. (2008). An advantage of the tone stimulus is the ability to achieve higher signal-to-noise ratio more efficiently, which in this study produced more usable data in comparison with the chirp stimulus, albeit for a select set of frequencies rather than the full wideband range. However, the overall results suggest that for newborns, a brief screen with a 2-kHz tone could quickly ascertain if the middle ear is able to transmit OAE stimuli and responses.

A comparison with a different middle-ear test, 1-kHz tympanometry, showed superior prediction of OAE status for wideband reflectance than did tympanometry, which had only an area under the ROC curve of 0.72, compared with 0.90 for reflectance (Figure 5). Similar results have recently been reported by Sanford et al. (2009), who reported an area under the curve of .75 using the same 1-kHz-immittance instrument (Madsen Otoflex).

Measurements at two different sites using the same equipment showed equivalent results above 1 kHz. Differences in results between the two sites occurred below 1 kHz, which may be due to differences in technique of probe insertion depth, or tightness of probe seal. This has been reported previously (Hunter et al. 2008), whereby leakage at low frequencies can occur with shallow or loose probe insertion. Probe insertion technique did not affect measurements in the higher frequencies (2 to 6 kHz), where measurements at the two sites were equivalent. This is also the frequency range that is most effective for prediction of OAE status in newborns. The lower frequency range could be important for assessment of other types of disorders, such as tympanic membrane perforations (as shown by Feeney et

al. 2003), so probe depth and seal may be an important issue when low frequency information is desired to assess other types of middle-ear disorders.

Some limitations of this study are also important to consider. Test order was kept constant with MEPA done first, followed by DP and then tympanometry. Because test order was not counter-balanced, this could potentially affect results if the probe location moved between tests. For example, a minor difference was found in chirp versus sine-stimulus results for MEPA and since the chirp was always done first, an order effect cannot be ruled out. However, major effects due to probe location should not have occurred, as stimulus levels were measured *in situ* for MEPA and DP and any insufficient tests were excluded from the analysis. A different probe was used for tympanometry and seal integrity was tested prior to measurements.

Analysis of age effects demonstrates that testing within the first 20 hours after birth is not ideal, since the DP pass rate is only about 50%. The pass rate at age 21 to 30 hours increases to around 75%, and the highest pass rate occurs around 36 hours at 85%. Caution should be exercised in interpreting these age-related data, as many of the age groups had a small number, age at testing was not randomly assigned, and the study design at UUH intentionally enrolled babies who did not pass transient-evoked OAE screening so the overall refer rate is higher than typical for a normal newborn nursery population. Despite these limitations, these data agree well with previous reports of OAE screening in newborn nurseries, which have shown that refer rates are higher in the first 24 hours, due to middle-ear issues (White et al. 1993; Roberts et al. 1995). Future studies that collect normative data should aim to break down by age by randomly assigning babies to be tested at particular times and/or repeat testing them over time.

Clinical feasibility of wideband reflectance testing along with OAE testing

Wideband reflectance using ambient pressure is a feasible screening test for middle-ear function in newborns, with a brief measurement time. The addition of reflectance to the OAE screening test usually took 10 seconds per ear in a sleeping infant once the probe was positioned. In the majority of babies, the same probe position was adequate for reflectance and OAE measurements. If the infant had a collapsing canal or if an acoustic leak was observed, repositioning of the probe was needed. In a few infants, many reattempts to position the probe were necessary. In general, if probe and noise issues were present, all 3 tests were affected, but reflectance was affected less by noise compared to OAE testing. In fussy babies, reflectance often could still be completed even when OAE and tympanometry could not, because reflectance does not require a noise floor as low as that needed for OAEs to get an adequate measurement, and reflectance at ambient pressure does not require a hermetic seal as for tympanometry. When leaks occurred, reflectance was affected in frequencies below 1 kHz, but above 1 kHz was relatively unaffected.

In summary, wideband middle-ear power analysis is better able than 1-kHz tympanometry to detect those newborns who are more likely to obtain a refer result on an OAE-based hearing screen due to middle-ear dysfunction. Detecting such cases could reduce unnecessary referrals for audiometric work-ups, as most middle-ear problems are transient, and affected ears will most likely pass screening once the middle ear normalizes.

Newborns with high reflectance scores at Stage I screening should be rescreened within a few hours to a few days, as most middle-ear problems are transient and resolve spontaneously. If reflectance and OAE are not passed upon Stage II screening, referral to an otologist for ear examination should be considered in combination with diagnostic testing. If reflectance is not normal at the time of threshold ABR, bone conduction testing is recommended. Newborns with normal reflectance and a refer result for the OAE screen

would logically be referred immediately to an audiologist for diagnostic testing with threshold ABR due to higher risk for permanent hearing loss. In any case, diagnostic testing should not be delayed due to poor middle ear status since underlying sensorineural hearing loss may occur in combination with middle ear dysfunction.

Acknowledgments

Financial support for this study was provided by National Institute of Deafness and Other Communication Disorders (Grant #: R44 DC006554). Many thanks to Adrienne Jackson, Mary Leigh Horn, and Shellie Sacco who collected data in the newborn nursery at UUH, and Jessica Day who collected data in the newborn nursery at UWMC. Thanks are due to the nursing and neonatology staff at UUH and UWMC for their assistance during the study. Thanks to Paul Boege and Pierre Parent who did the software coding. Thanks to Jont Allen, Vit Drga, Linton Miller, and the anonymous reviewers who made many helpful suggestions. Portions of this study were presented at the Lake Como Early Hearing Detection and Intervention Conference in 2007, Audiology Now! in 2007, and the 33rd Annual Midwinter Research Meeting of the Association for Research in Otolaryngology in 2010.

References

- 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(6):3617–3627. [PubMed: 17552713]
- Alaerts J, Lutz 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(6):727–732. [PubMed: 17948353]
- Allen JB, Jeng PS, Levitt H. Evaluation of human middle ear function via an acoustic power assessment. *J Rehabil Res Dev.* 2005; 42(4 Suppl 2):63–78. [PubMed: 16470465]
- Baldwin M. Choice of probe tone and classification of trace patterns in tympanometry undertaken in early infancy. *Int J Audiol.* 2006; 45(7):417–427. [PubMed: 16938800]
- Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psych.* 1975; 12:387–415.
- Boone RT, Bower CM, Martin PF. Failed newborn hearing screens as presentation for otitis media with effusion in the newborn population. *Int J Pediatr Otorhinolaryngol.* 2005; 69(3):393–397. [PubMed: 15733600]
- Calandruccio L, Fitzgerald TS, Prieve BA. Normative multifrequency tympanometry in infants and toddlers. *J Am Acad Audiol.* 2006; 17(7):470–480. [PubMed: 16927511]
- 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(6 Pt 1):597–603. [PubMed: 9215369]
- Eby TL, Nadol JB Jr. Postnatal growth of the human temporal bone. Implications for cochlear implants in children. *Ann Otol Rhinol Laryngol.* 1986; 95(4 Pt 1):356–364. [PubMed: 3755577]
- Feeney MP, Grant IL, Marrayott LP. Wideband energy reflectance measurements in adults with middle-ear disorders. *J Speech Lang Hear Res.* 2003; 46(4):901–911. [PubMed: 12959468]
- Finitzo T, Albright K, O'Neal J. The newborn with hearing loss: detection in the nursery. *Pediatrics.* 1998; 102(6):1452–1460. [PubMed: 9832584]
- Gorga MP, Norton SJ, Sininger YS, et al. Identification of neonatal hearing impairment: distortion product otoacoustic emissions during the perinatal period. *Ear Hear.* 2000; 21(5):400–424. [PubMed: 11059701]
- Hall JW, Smith SD 3rd, Popelka GR. Newborn hearing screening with combined otoacoustic emissions and auditory brainstem responses. *J Am Acad Audiol.* 2004; 15(6):414–425. [PubMed: 15341223]
- Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* 1983; 148(3):839–843. [PubMed: 6878708]
- Holte L, Margolis RH, Cavanaugh RM Jr. Developmental changes in multifrequency tympanograms. *Audiology.* 1991; 30(1):1–24. [PubMed: 2059166]

- Hunter LL, Margolis RH. Multifrequency Tympanometry: Current Clinical Application. *Am J Audiol.* 1992; 1(3):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(4):309–324. [PubMed: 18795470]
- Ikui A, Sando I, Sudo M, et al. Postnatal change in angle between the tympanic annulus and surrounding structures. Computer-aided three-dimensional reconstruction study. *Ann Otol Rhinol Laryngol.* 1997; 106(1):33–36. [PubMed: 9006358]
- Jaisinghani VJ, Paparella MM, Schachern PA, et al. Residual mesenchyme persisting into adulthood. *Am J Otolaryngol.* 1999; 20(6):363–370. [PubMed: 10609480]
- Joint Committee on Infant Hearing. Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics.* 2007; 120(4):898–921. [PubMed: 17908777]
- 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(2):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(5):2617–2638. [PubMed: 8270739]
- 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(5):443–461. [PubMed: 11059703]
- Keefe DH, Gorga MP, Neely ST, et al. Ear-canal acoustic admittance and reflectance measurements in human neonates. II. Predictions of middle-ear in dysfunction and sensorineural hearing loss. *J Acoust Soc Am.* 2003; 113(1):407–422. [PubMed: 12558278]
- Keefe DH, Ling R, Bulen JC. Method to measure acoustic impedance and reflection coefficient. *J Acoust Soc Am.* 1992; 91(1):470–485. [PubMed: 1737890]
- Lapsley Miller JA, Scurfield BK, Drga V, et al. Nonparametric relationships between single-interval and two-interval forced-choice tasks in the theory of signal detectability. *J Math Psych.* 2002; 46(4):383–417.
- 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(4):590–595. [PubMed: 3761072]
- Margolis RH, Bass-Ringdahl S, Hanks WD, et al. Tympanometry in newborn infants - 1 kHz norms. *J Am Acad Audiol.* 2003; 14(7):383–392. [PubMed: 14620612]
- Maxon AB, White KR, Vohr BR, et al. Using transient evoked otoacoustic emissions for neonatal hearing screening. *Br J Audiol.* 1993; 27(2):149–153. [PubMed: 8220282]
- Miura T, Suzuki C, Otani I, et al. Marrow-tympanum connections in fetuses and infants. *Nippon Jibiinkoka Gakkai Kaiho.* 2008; 111(1):14–20. [PubMed: 18260300]
- Northrop C, Piza J, Eavey RD. Histological observations of amniotic fluid cellular content in the ear of neonates and infants. *Int J Pediatr Otorhinolaryngol.* 1986; 11(2):113–127. [PubMed: 2427466]
- Norton SJ, Gorga MP, Widen JE, et al. Identification of neonatal hearing impairment: a multicenter investigation. *Ear Hear.* 2000; 21(5):348–356. [PubMed: 11059697]
- Paradise JL, Smith CG, Bluestone CD. Tympanometric detection of middle ear effusion in infants and young children. *Pediatrics.* 1976; 58(2):198–210. [PubMed: 951134]
- Prieve B, Dalzell L, Berg A, et al. The New York State universal newborn hearing screening demonstration project: outpatient outcome measures. *Ear Hear.* 2000; 21(2):104–117. [PubMed: 10777018]
- Roberts DG, Johnson CE, Carlin SA, et al. Resolution of middle ear effusion in newborns. *Arch Pediatr Adolesc Med.* 1995; 149(8):873–877. [PubMed: 7633540]
- 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(6):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(4):2106–2122. [PubMed: 19062852]

- Sanford CA, Keefe DH, Liu Y-W, et al. Sound-Conduction Effects on Distortion-Product Otoacoustic Emission Screening Outcomes in Newborn Infants: Test Performance of Wideband Acoustic Transfer Functions and 1-kHz Tympanometry. *Ear & Hearing*. 2009; 30(6):635–652. [PubMed: 19701089]
- Saunders, JC.; Kaltenbach, 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*. NY: Academic; 1983. p. 3-25.
- Shahnaz N. Wideband reflectance in neonatal intensive care units. *J Am Acad Audiol*. 2008; 19(5): 419–429. [PubMed: 19256090]
- Sprague BH, Wiley TL, Goldstein R. Tympanometric and acoustic-reflex studies in neonates. *J Speech Hear Res*. 1985; 28(2):265–272. [PubMed: 4010256]
- Stinson MR, Shaw EA, Lawton BW. Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal. *J Acoust Soc Am*. 1982; 72:766–73. [PubMed: 7130535]
- 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(3):163–172. [PubMed: 8075412]
- Thompson DC, McPhillips H, Davis RL, Lieu TL, Homer CJ, Helfand M. Universal newborn hearing screening: summary of evidence. *JAMA*. 2001; 286(16):2000–2010. [PubMed: 11667937]
- 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(5):669–681. [PubMed: 17804981]
- Vartiainen E. Otitis media with effusion in children with congenital or early-onset hearing impairment. *J Otolaryngol*. 2000; 29(4):221–223. [PubMed: 11003073]
- Voss SE, Allen JB. Measurement of acoustic impedance and reflectance in the human ear canal. *J Acoust Soc Am*. 1994; 95(1):372–384. [PubMed: 8120248]
- Voss SE, Horton NJ, Woodbury RR, Sheffield KN. Sources of variability in reflectance measurements on normal cadaver ears. *Ear Hear*. 2008; 29(4):651–665. [PubMed: 18600136]
- White KR, Vohr BR, Behrens TR. Universal newborn hearing screening using transient-evoked otoacoustic emissions: Results of the Rhode Island Hearing Assessment Project. *Seminars in Hearing*. 1993; 14:18–29.
- Yang EY, Stuart A, Mencher GT, et al. Auditory brain stem responses to air- and bone-conducted clicks in the audiological assessment of at-risk infants. *Ear Hear*. 1993; 14(3):175–182. [PubMed: 8344474]
- Zwislocki JJ. Analysis of the middle-ear function. Part I: Input Impedance. *J Acoust Soc Am*. 1962; 34(8 (2)):1514–1523.

DP–status x Frequency Interaction

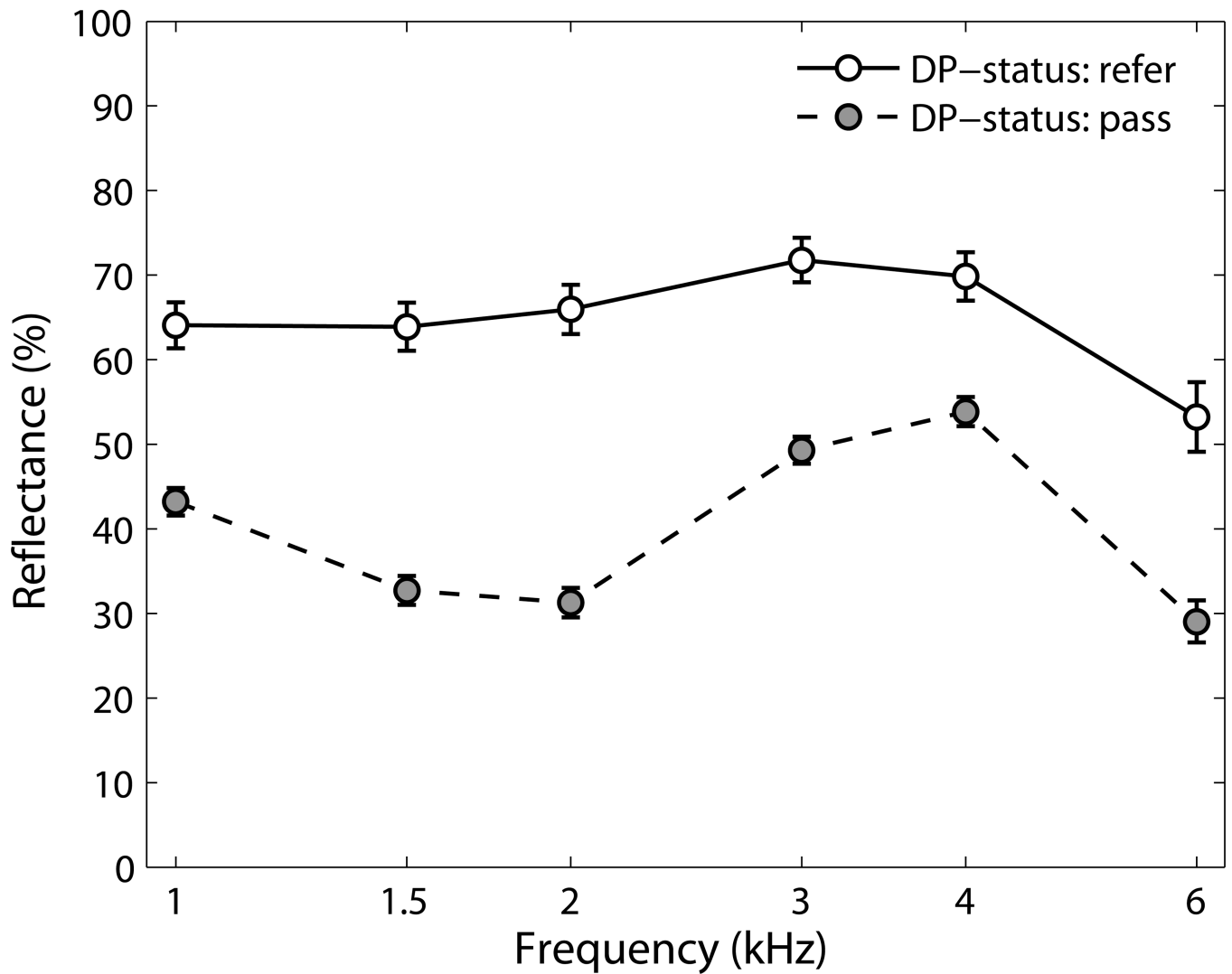


Figure 1. Frequency × DP-status interaction for the 479 ears with tone-stimulus reflectance results. Ears that achieved a DP pass result had significantly lower reflectance than ears that were referred. Error bars are 95% confidence intervals.

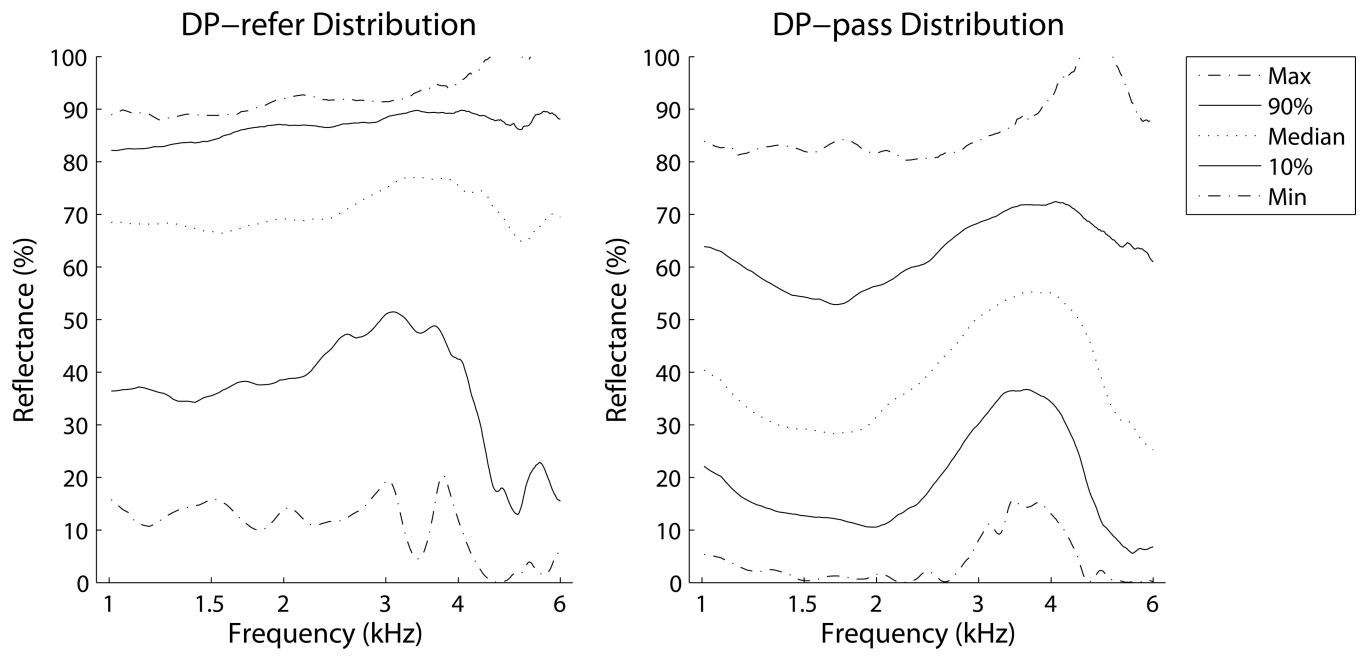


Figure 2. Reflectance distributions for chirp-stimuli by frequency for DP status groups: pass (right) and refer (left). Reflectance for the DP-refer group is much higher than for the DP-pass group.

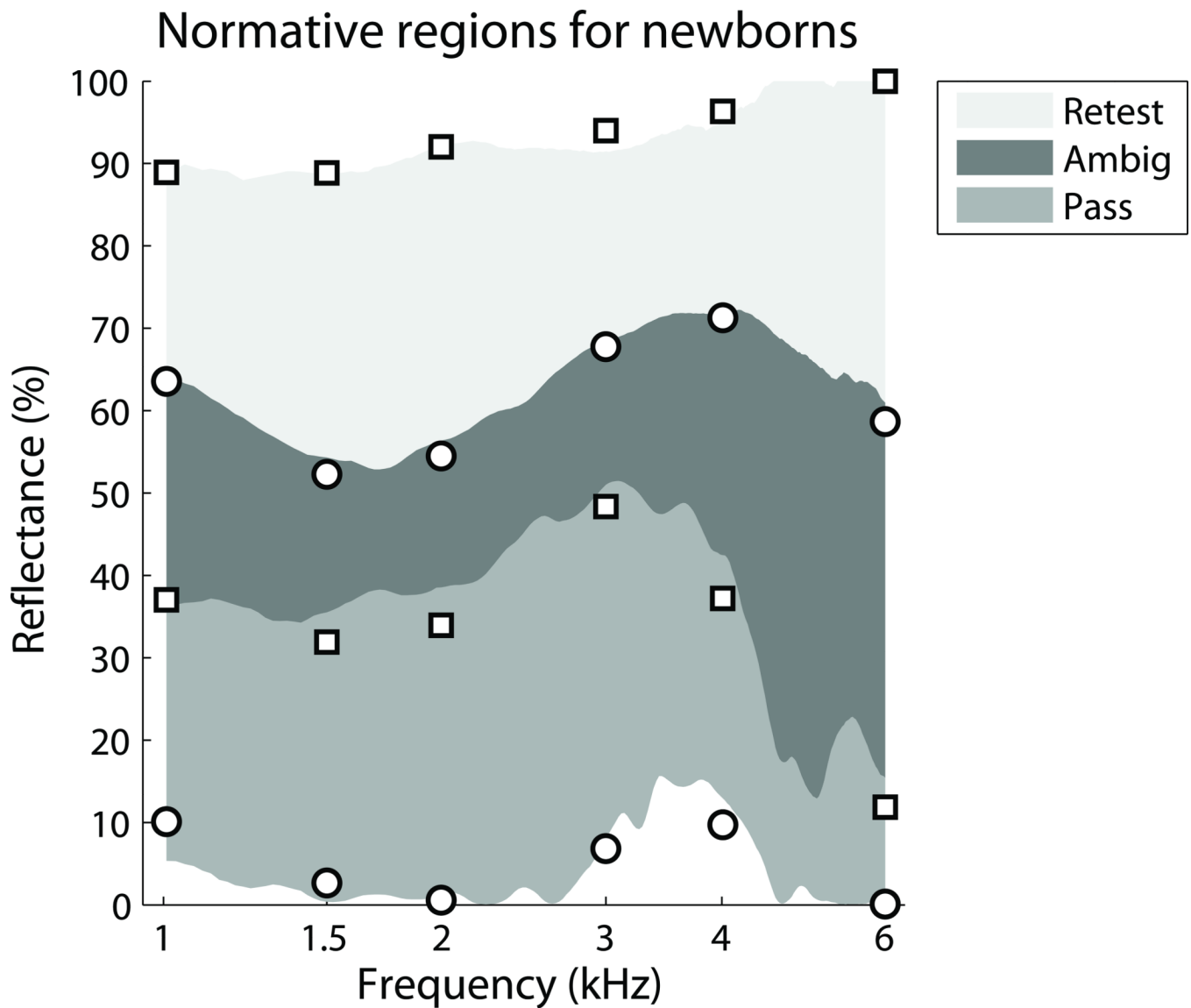


Figure 3. Reflectance norms for newborns, showing Retest, Ambiguous (Ambig.), and Pass regions. The solid gray regions are defined from Figure 2 and are based on the chirp-stimulus dataset. The circle and square symbols overlaying these regions are the norms from the larger tone-stimulus dataset, which only provided reflectance values at discrete frequencies.

Norms for Reflectance Area Indices

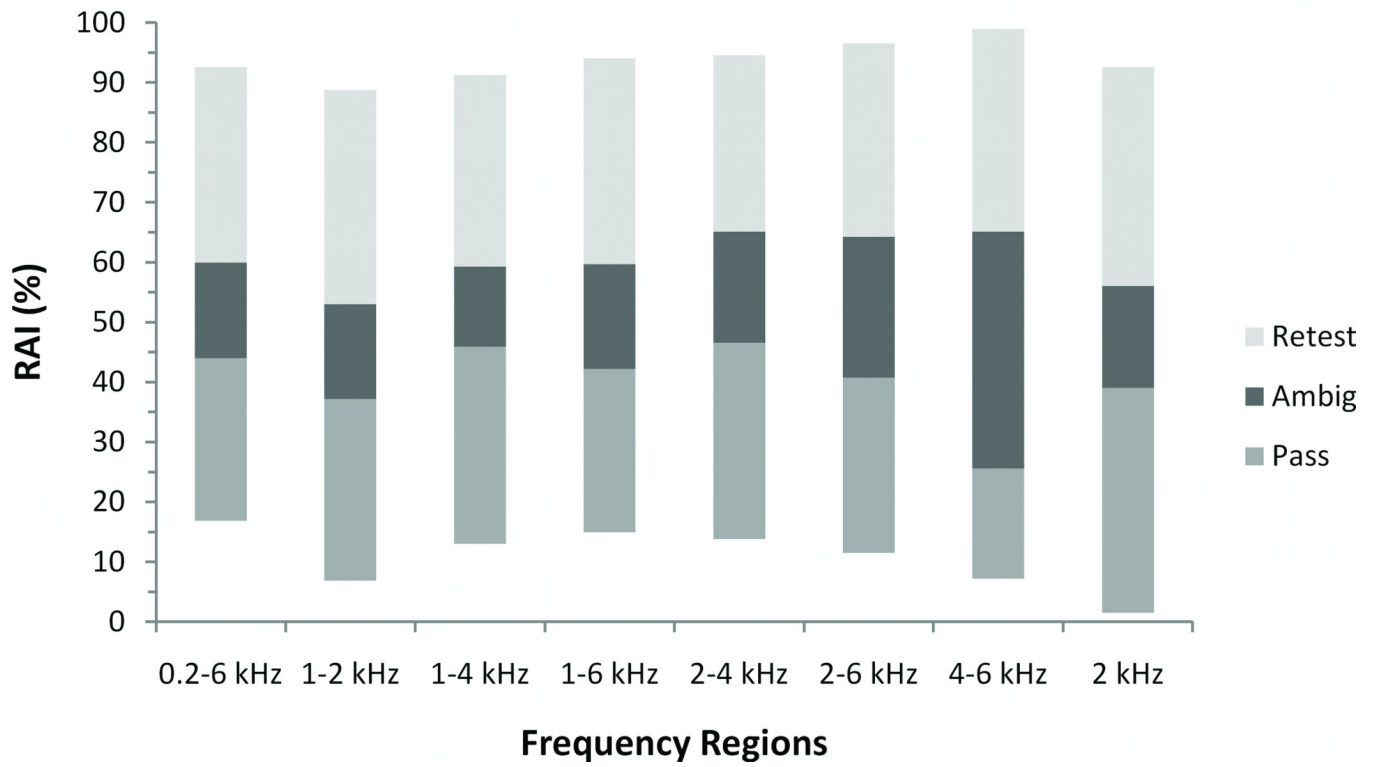


Figure 4. Normative regions for reflectance area indices (RAIs) for various frequency ranges, based on the chirp-stimulus dataset.

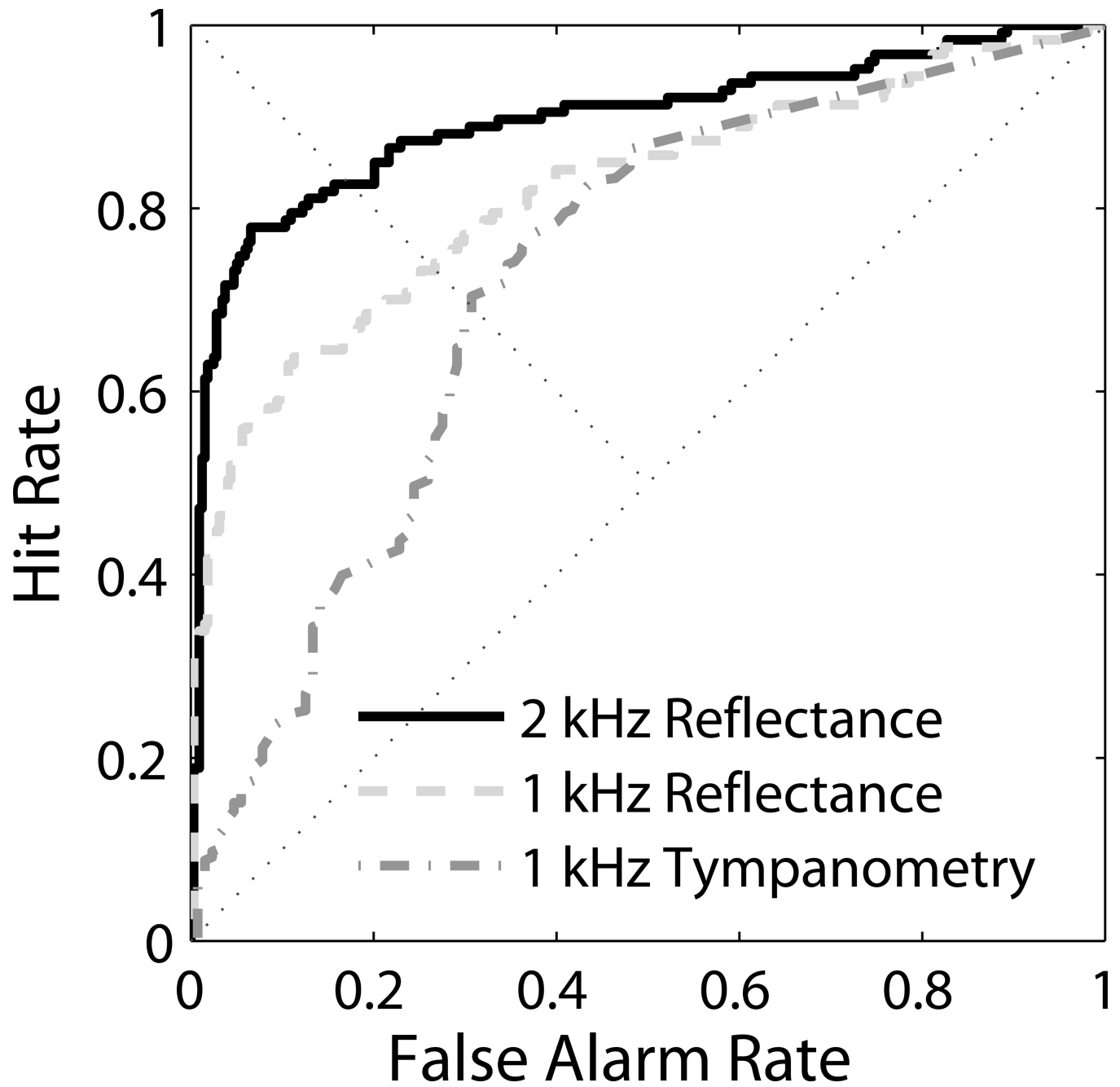


Figure 5. ROC curves comparing Reflectance at 1 and 2 kHz with Tympanometry at 1 kHz.

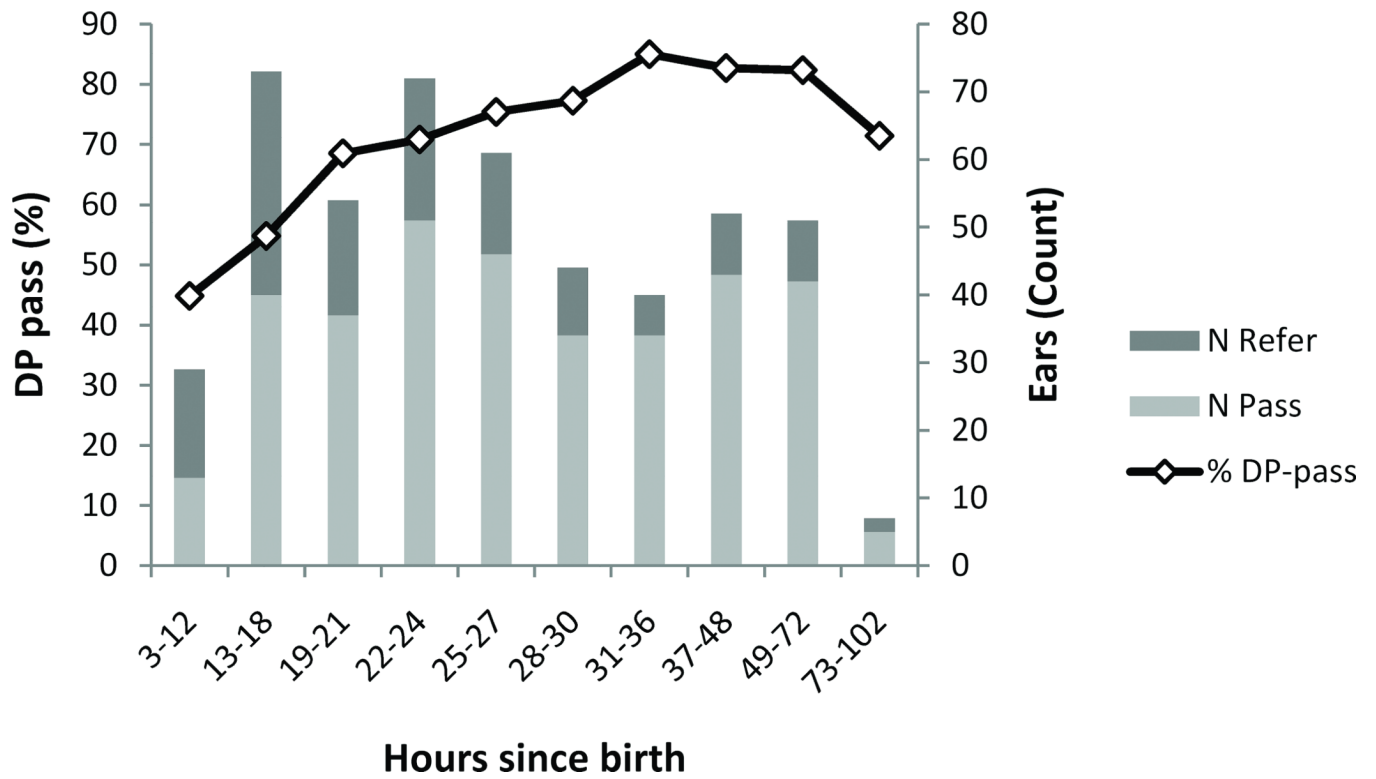


Figure 6. Percentage of DP-pass ears by hours since birth and the number of DP-pass and DP-refer ears. The DP-pass rate improved over time.

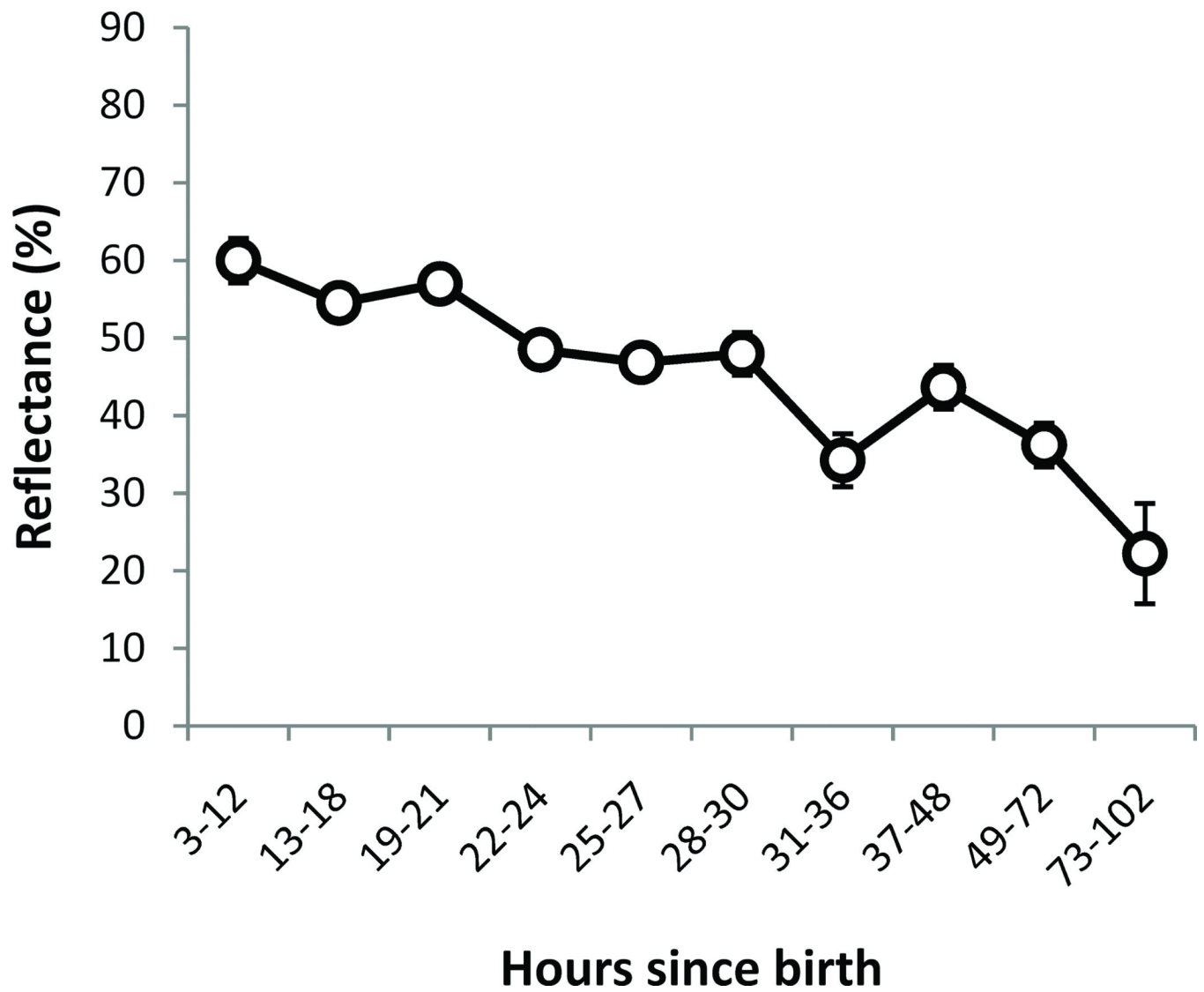


Figure 7. Reflectance (for tone-stimulus) at 2 kHz as a function of age-group at testing. For number of ears see Figure 6. Error bars are ± 1 standard error. Note that the age-group bands were of varying durations.

Table 1

Number of babies and ears from each test site, broken down by ear, sex, age at first test, number of ears contributing to the MEPA and DP datasets, birth weight, head circumference, birth type, race/ethnicity (as reported by the parents).

		Site		
		UUH & UWMC	UUH	UWMC
Babies (Count)		324	262	62
Ears (Count)	Total	493	395	98
	Left ear only	67	50	17
	Right ear only	88	79	9
	Both ears	169	133	36
	Left ears total	236	183	53
	Right ears total	257	212	45
Sex (Count)	Females	151	120	31
	Males	173	142	31
Age (Hours)	Range	3 to 102	3 to 96	19 to 102
	Mean	29	26	40.6
	SD	15.5	13.9	16.9
	Median	25	24	38.5
MEPA (Count)	Ears for 2kHz tone	483	385	98
	Ears for chirp	347	261	86
	Both tone and chirp	337	251	86
DP (Count)	Pass status	352	263	89
	Refer status	141	132	9
Birth Weight (g)	Mean	3295	3311	3227
	SD	458	434	541
Head Circumference (cm)	Mean	34	34	35
	SD	1.8	1.8	1.6
Birth Type – Vaginal	Total	304	267	37
Birth Type – C-section	Total	182	128	54
Birth Type – Other	Total	7	0	7
Ethnicity – African American	Total	7	3	4
Ethnicity – American Indian or Alaskan Native	Total	4	4	0
Ethnicity – Asian	Total	4	1	3
Ethnicity – Hispanic/Latino	Total	149	148	1
Ethnicity – Pacific Islander	Total	4	4	0
Ethnicity – White, Non-Hispanic	Total	143	126	17
Ethnicity – Undeclared	Total	182	109	73

Table 2Breakdown of ANOVA effects (ns = not significant at $p < .05$).

Effect	dof	F	p
Ear	1	8.68	<.05
Sex	1	5.46	<.05
DP-status	1	357.65	<.05
Ear × Sex	1	0.59	ns
Ear × DP-status	1	4.04	<.05
Sex × DP-status	1	7.00	<.05
Ear × Sex × DP-status	1	2.35	ns
Error	471		
Frequency	5	135.04	<.05
Frequency × Ear	5	1.19	ns
Frequency × Sex	5	0.92	ns
Frequency × DP-status	5	25.41	<.05
Frequency × Ear × Sex	5	1.10	ns
Frequency × Ear × DP-status	5	2.83	<.05
Frequency × Sex × DP-status	5	1.10	ns
Frequency × Ear × Sex × DP-status	5	1.20	ns
Error	2355		

Table 3

Reflectance and reflectance area index (RAI) values for the 0, 10, 90, and 100th percentiles for the DP-pass and DP-refer groups. These percentiles were used to define the normative pass, ambiguous, and retest reflectance regions for the tone-stimulus frequencies illustrated in Figure 3 and the frequency ranges in Figure 4. The pass region was defined by the 0th percentile for DP-pass and the 10th percentile for DP-refer; the ambiguous region was defined by the 10th percentile for DP-refer and the 90th percentile for DP-pass, and the retest region was defined by the 90th percentile for DP-pass and the 100th percentile for DP-refer.

Frequency/ Frequency Range (kHz)	Stimulus type	DP-status	Percentile			
			0	10	90	100
1.0	Tone	Pass	10.1	24.8	63.6	84.7
1.5	Tone	Pass	2.6	15.3	52.3	87.6
2.0	Tone	Pass	0.6	10.7	54.5	87.9
3.0	Tone	Pass	6.8	30.4	67.7	92.2
4.0	Tone	Pass	9.7	35.2	71.3	94.2
6.0	Tone	Pass	0.1	4.3	58.6	90.5
1.0	Tone	Refer	16.0	37.0	82.4	88.9
1.5	Tone	Refer	10.5	31.9	84.3	88.9
2.0	Tone	Refer	10.8	33.9	86.0	92.0
3.0	Tone	Refer	19.3	48.3	88.7	94.0
4.0	Tone	Refer	11.7	37.2	88.9	96.3
6.0	Tone	Refer	0.0	11.8	88.5	100.0
0.2 to 6 kHz	Chirp	Pass	16.9	28.8	59.9	84.5
1 to 2 kHz	Chirp	Pass	6.8	17.2	53.0	82.3
1 to 4 kHz	Chirp	Pass	13.0	27.7	59.3	84.4
1 to 6 kHz	Chirp	Pass	14.9	25.3	59.7	86.1
2 to 4 kHz	Chirp	Pass	13.8	29.0	65.1	87.5
2 to 6 kHz	Chirp	Pass	11.5	25.4	64.2	88.5
4 to 6 kHz	Chirp	Pass	7.2	18.2	65.1	95.2
2 kHz	Chirp	Pass	1.5	10.1	56.0	84.5
0.2 to 6 kHz	Chirp	Refer	31.7	44.0	86.1	92.6
1 to 2 kHz	Chirp	Refer	17.2	37.2	84.6	88.7

Frequency/ Frequency Range (kHz)	Stimulus type	DP-status	Percentile			
			0	10	90	100
1 to 4 kHz	Chirp	Refer	28.8	45.9	87.5	91.3
1 to 6 kHz	Chirp	Refer	26.4	42.2	87.3	94.1
2 to 4 kHz	Chirp	Refer	19.6	46.6	89.1	94.6
2 to 6 kHz	Chirp	Refer	21.2	40.7	87.8	96.6
4 to 6 kHz	Chirp	Refer	13.9	25.6	88.9	98.9
2 kHz	Chirp	Refer	14.0	39.1	87.1	92.6

Table 4

Area under the receiver-operating characteristic (ROC) curves based on the DP-pass and DP-refer reflectance and reflectance area index (RAI) distributions, for various frequencies and frequency regions, based on both the tone-stimulus and chirp-stimulus datasets.

Frequency/Frequency Range (kHz)	Stimulus type	Area under ROC curve
1	Tone	0.82
1.5	Tone	0.88
2	Tone	0.90
3	Tone	0.84
4	Tone	0.76
6	Tone	0.74
Wideband (0.2 to 6)	Chirp	0.86
1 to 2	Chirp	0.90
1 to 4	Chirp	0.90
1 to 6	Chirp	0.86
2 to 4	Chirp	0.87
2 to 6	Chirp	0.83
4 to 6	Chirp	0.78
2	Chirp	0.90