Critical review: Can wideband energy reflectance be used in newborn hearing screening to detect transient middle ear dysfunction and to interpret screening results?

Laura Sangster
M.Cl.Sc (AUD) Candidate
University of Western Ontario: School of Communication Sciences and Disorders

Temporary conditions of the external ear canal and middle ear in neonates can affect sound transmission and lead to "fail" results in newborn hearing screening. The purpose of this review was to determine if wideband energy reflectance could be used as a screening measure to detect middle ear dysfunction in very young infants and to help interpret evoked otoacoustic emissions screening results. Four single group studies and one cohort study were found using a search of PubMed, EMBASE, and the reference lists of included articles. Critical evaluation suggested that while a test of wideband energy reflectance holds promise for improving the false positive rate of screening programs, inherent limitations exist in this body of literature. Further research is recommended to ensure that including this test in an early hearing screening battery would decrease false positives and be a valid and reliable addition to the battery.

Introduction

The position statement of the Joint Committee on Infant Hearing (JCIH; 2007) recommends that all infants under the age of one month receive hearing screening in order to provide early detection and intervention for children with hearing impairment. For the most part, newborn screening has been effective in the early identification of congenital hearing loss; however, there remains room for improvement (Hunter, Feeney, Lapsley Miller, Jeng, & Bohning, 2010). A significant issue in universal newborn hearing screening programs are "false positives," in which normally hearing ears do not produce a sufficient response to pass the screening test (Keefe, Zhao, Neely, Gorga, & Vohr, 2003b). It is a frequently held view that many of these referrals are the result of transient obstruction in the ear canal and/or middle ear (Keefe, Folsom, Gorga, Vohr, Bulen, & Norton, 2000). These relatively common conditions of the ear canal and middle ear (ME) following birth (e.g., fluid or mesenchyme in the ME space, vernix or debris in the ear canal) affect the forward and reverse sound transmission necessary to both elicit and measure the cochlear response in evoked otoacoustic emissions (EOAE) screening tests (Hunter, Bagger-Sjoback, & Lundberg, 2008; Hunter et al., 2010; Sanford et al., 2009). Reducing false positives is crucial in order to eliminate parental anxiety, to facilitate appropriate intervention and diagnostic follow-up, and to ensure cost effectiveness and credibility of screening programs (Hunter et al., 2010; Keefe et al., 2003b).

Suggested ways to address false positives are implementing a two-stage screening protocol (e.g., DPOAE followed by automated auditory brainstem response [ABR] if the infant fails the first-stage) or re-screening on an outpatient or community basis (Hunter et al., 2010; Hyde, 2010). However, these methods present issues in that: (1) short hospital stays are common after birth, which constrain the scheduling of two-stage screening prior to discharge (Hyde, 2010); (2) ABR screening tests are sensitive to ME problems as well (Hunter & Daly, 2010; Keefe et al., 2000); and (3) children are lost to follow-up between the failed screening and the re-screening stage (JCIH, 2007). As a result, a test of the sound transmission pathway (i.e., status of external ear and/or ME) in neonates may prove useful to detect transient conditions that are frequent contributors to failed results. In addition, such a test may help to interpret results by distinguishing failed newborn screening due to ME issues from failed screening results that are likely due to sensorineural hearing loss (Keefe et al., 2003b).

Current universal newborn hearing screening guidelines do not require assessment of ME function (Hunter & Daly, 2010). As well, there is no clinically accepted gold standard for testing ME status in neonates (Hunter et al., 2010; Keefe et al., 2000). Calibration issues prevent bone-conducted ABR measurement in neonates (Keefe et al., 2003b) and there are ethical concerns regarding the use of myringotomy to detect ME effusion in newborns (Shahnaz, 2008). In later stages of the program, in which infants with suspected hearing impairment are referred to audiologists for diagnostic testing, tympanometry with a 1-kHz probe is recommended for infants under 6 months (Ontario Infant Hearing Program, 2008). However, tympanometry may be problematic in very young infants because it requires pressurization, which can change the ear canal diameter of young infants who have flaccid canal walls (Keefe et al., 2000; Vander Werff, Prieve, & Georgantas, 2007). Also, there is uncertainty surrounding the sensitivity of 1-kHz tympanometry in neonates (Hunter & Margolis, 2010).

An alternative method to assess ME status in neonates is wideband energy reflectance (WBR). A probe is placed in the ear, similarly to tympanometry or OAE measurements, and a wideband chirp or click stimulus is delivered (Hunter & Margolis, 2010). Energy reflectance is the ratio of reflected energy at the probe's location to the energy delivered by the probe, and varies from values of 0 (all energy is absorbed) to 1 (all energy is reflected) (Sanford et al., 2009). Typically, reflectance is highest (i.e., closer to 1) for high and low frequencies but lowest (i.e., closer to 0) in the mid-frequency range (Vander Werff et al., 2007). WBR measures present certain advantages over tympanometry in infants, namely that pressurization is not required, it can test a wide range of frequencies, and the measurement can be made quickly using the same equipment as EOAEs (Vander Werff et al., 2007). It is less susceptible to environmental and infant noise than EOAEs (Keefe et al., 2000), and relatively insensitive to how deep the transducer is placed in the canal (Hunter et al., 2010). At present, wideband acoustic transfer functions have been measured in healthy newborns and infants in intensive care (e.g., Keefe et al., 2000), and used to predict ME status in children and adults (Keefe and Simmons, 2003; Piskorski, Keefe, Simmons, & Gorga, 1999), but it is not in widespread clinical use.

Objective

The purpose of this review was to critically analyze the available literature examining the use of WBR in newborn hearing screening to determine if it holds clinical value for detecting ME dysfunction in neonates and for interpreting absent EOAE responses.

Method

Search Strategy
PubMed and EMBASE were searched using the terms (reflectance) AND (neonate) AND (middle ear). The database search was limited to English articles with human subjects. In addition, the reference lists of included articles were examined for articles that met the selection criteria. One additional reference was found in a book chapter regarding pediatric ME assessment.

Selection Criteria
Studies were eligible for inclusion if the focus of the study was to determine the efficacy of wideband acoustic transfer functions as a screening or diagnostic tool (i.e., as opposed to studies focusing solely on establishing normative data for neonates). All types of study design were eligible for inclusion.

Data Collection
Application of the selection criteria resulted in four single group studies and one prospective cohort study. According to an experimental design decision tree (L. Archibald, personal communication, 2011), the cohort study provides a level 2c of evidence and the single group studies provide a level 3 of evidence.

Results

Single Group Study 1
Hunter et al. (2010) compared the test performance of WBR with 1-kHz tympanometry to assess its ability to predict distortion product OAE (DPOAE) results. Measurements were made on 324 healthy full-term neonates from two test sites during screening within 48 hours of birth. Receiver operating characteristic (ROC) analyses were used to determine which test provided better discriminability for pass or fail DPOAE results. Results from this analysis suggested that reflectance better discriminated between DPOAE pass and refer than 1 kHz tympanometry. Areas under the ROC (AROC) curve were 0.72 for 1-kHz tympanometry, 0.82 for reflectance at 1 kHz, and 0.90 for reflectance at 2 kHz. Frequency regions of the WBR that had the best discriminability involved 2 kHz, particularly, 1-2 kHz, 1-4 kHz, and 2 kHz.

Overall, this study supports the use of WBR to detect ME dysfunction in neonates because it predicts pass or refer result on a DPOAE test, a test that indirectly encodes information about the sound conduction pathway. However, some limitations exist. For example, the DPOAE module used in testing had different characteristics from commercially available products in its method of determining noise level and calibration, limiting the conclusions that can be drawn for a typical clinical setting. While statistical analyses were appropriate for the study design, the authors did not report confidence intervals for AROC analyses, which is insufficient statistical reporting.

Single Group Study 2
The purpose of Sanford et al. (2009) was to assess the ability of wideband acoustic transfer functions and 1-kHz tympanometry to predict the status of the sound conduction pathway in infants who passed or referred on a DPOAE test. On Day 1, DPOAE testing was conducted on 455 well-baby ears, followed by 1-kHz tympanometry and wideband acoustic transfer function measurements, which included ambient and pressurized energy absorbance (EA), acoustic admittance magnitude, and acoustic phase. On Day 2, infants who got a "refer" on DPOAE testing on Day 1 were re-screened, and the experimental test protocol was repeated. Data were analyzed using AROC curves, with DPOAE test outcomes as the comparison test to assess the performance of the experimental tests in classifying ears as pass or fail. Day 1 results revealed that ambient EA had an AROC of 0.86 (95% CI 0.80-0.89) whereas tympanometry had an AROC of 0.75 (95% CI 0.68-0.80). However, the only predictors exceeding chance performance on Day 2 were wideband measures, the highest AROC (0.74, 95% CI 0.51-0.89) being for ambient admittance.

In general, these findings met the authors' objective of determining which test had better performance but do not provide sufficient evidence for this review's objective of determining if WBR can be used to detect ME issues in neonates. Although absorbance and reflectance are related, it is unknown if they have different test performance. Importantly, even though the statistical procedures were sound, ambient pressure EA was not able to distinguish between pass and refer groups on Day 2, suggesting inadequate clinical utility. Further, this evidence should be interpreted with caution because not all children who were screened on Day 1 were screened again on Day 2. Hyde (2010) posits that to prospectively determine sensitivity all children who are screened must receive follow-up to determine their true hearing status.

Single Group Study 3

The basic purpose of Keefe et al.'s (2003b) retrospective study was to understand how ME function influences measures of cochlear function in neonates (i.e., EOAE and ABR) by evaluating acoustic admittance and reflectance (YR). Ears (*n* = 2766) included in the analysis had TEAOE, DPOAE, ABR, and YR measurements completed. Correlation, multivariate logistic regression, and AROC curve analyses were performed. Results indicated that high-frequency reflectance (2-4 kHz) had the highest correlation with EOAE level. Logistic regression analyses revealed that the most important factor in classifying DPOAE results was high-frequency reflectance. YR factors classified EOAE results with AROC curves ranging from 0.62-0.79. The odds ratio for high-frequency reflectance was 2.44 (95% CI 2.09-2.86) for classifying DPOAE results, indicating that ears with elevated reflectance from 2-4 kHz had a higher likelihood of having ME dysfunction.

Findings from this study are important in that they suggest that ME function has an impact on measures of cochlear function. This is evidenced in the relationship between EOAE levels and reflectance. The YR test was shown to predict whether or not a DPOAE response was present or absent, suggesting that this test is sensitive to ME dysfunction in neonates. Strengths of this study are the large sample size, the likelihood of representative data, and the randomization of the order of tests. In addition, the statistical manipulations are valid. However, in general, YR factors did not classify EOAE results with sufficiently high AROC values, leaving some question about their clinical utility.

Single group study 4

Building on the evidence in Keefe et al. (2003b), Keefe et al. (2003a) examined if YR could be added to a screening battery to improve the prediction of sensorineural hearing loss as later assessed by behavioural audiometry at 8-12 months. In a complex analysis, YR tests for ME dysfunction were developed on a sub-set of normally hearing ears (*n* = 1027) as demonstrated by VRA at 8-12 months. The authors subsequently evaluated tests for ME dysfunction on a population with unknown hearing status to evaluate test generalization (*n* = 1147). A logistic regression analysis was performed to see if EOAE and YR variables combined were better at predicting outcome than EOAE alone. For predicting sensorineural hearing loss, the multivariate model for DPOAE had higher likelihood than the univariate model, which suggested that including YR measurement improved the model's ability to predict SNHL with DPOAE measurements. For a test of ME dysfunction, high-frequency reflectance was the best predictor, with an AROC curve of 0.86 in classifying normally hearing ears that pass DPOAE and TEOAE as having normal ME function and those that fail both tests as having ME dysfunction. Fifty-one of 1027 ears failed this two-stage DPOAE and ABR screening and 40 ears had ME dysfunction, reducing unexplained failures (i.e., false positives) from 5% to 1.1%.

These findings imply that YR in addition to OAE could enhance the ability to predict hearing status, the main goal of any screening program. This lends itself to suggest that a test of ME function in neonates provides valuable information about hearing status and can help to decrease the number of false positives in a screening program. Strengths of this study are the relatively high AROC curve for high-frequency reflectance, the use of a gold standard test of hearing status, and the evaluation of ME dysfunction tests on a population with unknown hearing status to ensure that the test generalized to a sample that would be similar to the test's clinical use. Statistical analyses were appropriate given the design of the study.

Prospective cohort study

Vander Werff et al. (2007) aimed to study both the test-retest reliability of wideband reflectance in two groups of infants, as well as to determine if WBR distinguished infants who passed from those who failed TEOAE screening. They employed a repeated measures design, measuring WBR three times. The probe was left in place for the first two measurements and the probe was reinserted for the third measurement. Vander Werff and colleagues calculated test-retest differences for each infant (*n* = 127) across one-third octave frequency bands and calculated the mean and 90th percentile for test-retest differences by subject group and TEOAE result. They also compared WBR patterns between infants who passed and failed a TEOAE screening test. Mann-Whitney rank-sum tests were used to compare mean WBR values for each frequency band by subject group and TEOAE result.

Results for test-retest reliability indicated that mean test-retest differences were relatively small when the probe was kept in place, but were higher for all frequencies after reinsertion. In relating WBR to ME dysfunction, infants who failed the screening had significantly higher WBR from 630 Hz to 2000 Hz than infants who passed. The authors point out that research with older children (e.g., Piskorski et al., 1999) has suggested that this range is important for detecting ME dysfunction. As a result, this finding may indicate that infants who failed had more ME dysfunction than those who passed, possibly due to transient or permanent ME issues identified by WBR. While the statistical methods were valid, this study was limited in that testers knew the infants' screening result when performing WBR testing in the screening group, leaving room for detection bias. As well, although there was a significant difference between means, the authors report that the difference was somewhat minor for this sample. The sample also included older infants, which limits the findings' application to the neonatal population.

Discussion

Overall, these studies provide suggestive evidence that WBR could be used to detect ME dysfunction and to interpret screening results in neonates. Three weak-moderate studies (Hunter et al., 2010; Keefe et al., 2003b; Sanford et al., 2009) evaluated the ability of a wideband measure to predict DPOAE status, but the AROC curve findings and the wideband energy indicator evaluated were not consistent across studies. One strong study (Keefe et al., 2003a) evaluated the ability of WBR to predict hearing status as later assessed by behavioural audiometry. This study provided compelling evidence to suggest that including YR measurement could improve the ability to predict hearing status, and that a test of ME dysfunction could reduce false positives in screening. In contrast to these four studies, Vander Werff et al. (2007) did not evaluate WBR against a comparison test but investigated if WBR results were different between infants who passed and those who failed TEOAE screening, and evaluated its' test-retest reliability. This weak-moderate study indicated that WBR has adequate test-retest reliability and demonstrated that WBR specifies ME dysfunction in those who fail TEOAE screening. Unfortunately, the literature as a whole does not provide enough strong evidence at this time to merit clinical implementation of WBR in neonatal screening.

Evaluating test performance of a screening measure such as WBR presents a challenge because there is no clinically accepted gold standard of diagnosing conductive hearing loss in neonates. Most studies chose to circumvent this issue by using an EOAE as a comparison test. This test is appealing because it is already used in screening programs and it indirectly gives information on the forward and reverse transmission of sound through the ME (i.e., because it is sensitive to conditions in the ME that affect sound transmission; Keefe et al., 2003b). Therefore, most studies relate screening outcomes to a measure that provides a description of the sound conduction pathway (Sanford et al., 2009), but is not a gold standard of ME function. This represents a significant shortcoming in this literature, limiting its clinical applicability. However, the extent to which WBR correctly predicts that ears have ME dysfunction (pass or fail EOAE result) does provide an *estimate* of its efficacy. In the absence of an established gold standard this may be a realistic method to evaluate the potential usefulness of WBR in neonatal screening.

Additional limitations of this body of literature should also be considered. The same authors have completed a large portion of the work in this area (e.g., D. Keefe and L. Hunter), which introduces a potential limit to the generalizability of the findings. As well, differences exist between studies, limiting true comparison. Not all of the studies evaluated a single, separated wideband energy indicator (i.e., reflectance), and the studies used different compositions of infants in their samples (i.e., well-babies and infants in neonatal intensive care units) and different EOAE pass criteria. Also, the included studies employed different methods of calibration and coupling to the ear. For example, Hunter et al. (2010) used mainly rubber tips, whereas Vander Werff et al. (2007) found that rubber tips were too big and had less test-retest reliability. In general, a methodological weakness that prevents high quality evaluation of test performance in this population is the under-inclusion of infants with hearing loss (Hyde, 2010; Keefe et al., 2003a). Including a large number of these infants would be difficult to achieve because congenital hearing loss has a relatively low prevalence (Hyde, 2010). Combined, these limitations prevent definitive conclusions about the suitability of WBR to be drawn.

While conclusive interpretation is not possible, this literature does provide preliminary evidence regarding WBR's clinical utility. It remains possible that WBR may help to refine the screening process. Normative data for WBR measures in neonates is currently available (e.g., Hunter et al., 2010; Keefe et al., 2000; Shahnaz, 2008) to evaluate responses as normal or abnormal. As well, Hunter et al. (2010) and Sanford et al. (2009) give examples of the hypothetical clinical use of WBR. If an infant fails an initial screening and has abnormally high reflectance, the infant should be re-screened within a few hours or days because abnormal sound conduction may have played a role in the refer result. Conversely, if an infant fails the initial screening but has normal reflectance scores, the infant should receive an immediate referral for diagnostic testing. Thus, later hearing evaluation would still be needed to clarify and confirm hearing status, even with the use of a WBR test (Keefe et al., 2003a).

Conclusion and Clinical Implications

The potential use of WBR to detect ME dysfunction and interpret screening results in neonatal hearing screening is an appealing prospect, especially in light of the need to decrease false positives. In spite of this, inherent limitations in this body of literature prohibit making any definitive recommendations regarding the specific use of WBR at this time. While a possible protocol of WBR use is discussed above, there is not sufficient evidence currently to merit incorporating this measure into a screening test battery. That is not to say however, that WBR does not hold promise. Further research should employ longitudinal designs to evaluate test performance of WBR against a gold standard such as behavioural hearing assessment at 8-12 months (as demonstrated by Keefe et al., 2003a). As well, additional research should investigate the use of high-frequency reflectance values for clinical decision-making, their diagnostic importance being alluded to by the findings of Hunter et al. (2010) and Keefe et al. (2003b). Although there is not sufficient support for implementation of WBR at present, this test of ME function may enhance the identification of neonatal hearing impairment in future screening programs.

References

Hunter, L. L., Feeney, M. P., Lapsley Miller, J. A., Jeng, P. S., & Bohning, S. (2010). Wideband reflectance in newborns: Normative regions and relationship to hearing-screening results. *Ear Hear, 31*, 599-610. doi:10.1097/AUD.0b013e3181e40ca7

Hunter, L. L., & Daly, K. A. (2010). Conductive hearing loss in children: Otitis media with effusion and congenital impairments. In R. Seewald and A. M. Tharpe (Eds.), *Comprehensive Handbook of Pediatric Audiology* (pp. 157-173). San Diego: Plural Publishing.

Hunter, L. L., & Margolis, R. H. (2010). Middle ear measurement. In R. Seewald and A. M. Tharpe (Eds.), *Comprehensive Handbook of Pediatric Audiology* (pp. 365-388). San Diego: Plural Publishing.

Hyde, M. (2010). Principles and methods of population hearing screening in EHDI. In R. Seewald and A. M. Tharpe (Eds.), *Comprehensive Handbook of Pediatric Audiology* (pp. 283 337). San Diego: Plural Publishing.

Joint Committee on Infant Hearing (2007). Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics, 120*, 898-921. doi: 10.1542/peds.2007-2333

Keefe, D. H., Folsom, R. C., Gorga, M. P., Vohr, B. R., Bulen, J. C., & Norton, S. J. (2000). Identification of neonatal hearing impairment: Ear-canal measurements of acoustic admittance and reflectance in neonates. *Ear & Hearing, 21*, 443-461. doi: 10.1088/09524746/28/4/003

Keefe, D. H., Gorga, M. P., Neely, S. T., Zhao, F., & Vohr, B. R. (2003a). 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, 113*, 407-422. doi:10.1121/1.1523388

Keefe, D. H., & Simmons, J. L. (2003). Energy transmittance predicts conductive hearing loss in older children and adults. *J. Acoust. Soc. Am., 114*, 3217-3238. doi: 10.1121/1.1625931

Keefe, D. H., Zhao, F., Neely, S. T., Gorga, M. P., & Vohr, B. R. (2003b). Ear-canal acoustic admittance and reflectance effects in human neonates. I. Predictions of otoacoustic emission and auditory brainstem responses. *J Acoust Soc Am, 113*, 389-406. doi:10.1121/1.1523387

Ontario Infant Hearing Program (2008, January). Audiologic assessment protocol. Retrieved from http://www.mountsinai.on.ca/care/infant hearing-program/documents/IHPAudiologic AssessmentProtocol3.1FinalJan2008.pdf

Piskorski, P., Keefe, D. H., Simmons, J. L., & Gorga, M. P. (1999). Prediction of conductive hearing loss based on acoustic ear-canal response using a multivariate clinical decision theory. *J. Acoust. Soc. Am., 105*, 1750-1764. doi: 10.1121/1.426713

Sanford, C. A., Keefe, D. H., Liu, Y., Fitzpatrick, D., McCreery, R. W., Lewis, D. E., & Gorga, M. P. (2009). 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 Hear, 30*, 635-652. doi:10.1097/AUD.0b013e3181b61cdc

Shahnaz, N. (2008). Wideband reflectance in neonatal intensive care units. *J Am Acad Audiol, 19*, 419-429. doi: 10.3766/jaaa.19.5.4

Vander Werff, K. R., Prieve, B. A., & Georgantas, L. M. (2007). Test-retest reliability of wideband reflectance measures in infants under screening and diagnostic test conditions. *Ear Hear, 28*, 669-681. doi:10.1097/AUD.0b013e31812f71