Validity of the Ipsilateral Acoustic Reflex as a Screening Parameter

Janet P. Sells*
Raymond M. Hurley†
C. Robin Morehouse†
Joyce E. Douglas*

Abstract

Historically, the ipsilateral acoustic reflex (IAR) has played a prominent role in middle ear screening protocols. However, in 1990, the American Speech-Language-Hearing Association’s revised guidelines, “Screening for Hearing Impairment and Middle Ear Disorders,” deleted the IAR as a screening parameter. This decision was based on studies that demonstrated that the inclusion of the IAR resulted in a high false-positive rate without adding to the hit rate and thus was a major contributor to over-referral. However, other investigations demonstrated the converse, that the IAR is a sensitive and contributory measure in a screening protocol. Our data suggest that the conflicting results are due to different operating systems in the middle ear analyzers that were used to elicit the IAR. Based on our findings, an optimal IAR screening protocol would use a pulsed elicitation immittance system (multiplexing circuit) and a maximum intensity level of 105 dB HL. Further, if the IAR screening were to be carried out using an automated (programmed) immittance system with an equivalent volume change criteria for IAR detection, the equivalent volume change of ≥0.03 ml is recommended.

Key Words: Acoustic reflex, hearing screening, immittance measures, ipsilateral acoustic reflex, middle ear screening

Abbreviations: ASHA = American Speech-Language-Hearing Association, GSI-33 = Grason-Stadler Inc, Model 33 Middle Ear Analyzer, IAR = ipsilateral acoustic reflex, MEE = middle ear effusion

Historically, the ipsilateral acoustic reflex (IAR) has played a prominent role in middle ear screening protocols. However, the American Speech-Language-Hearing Association’s revised guidelines, “Screening for Hearing Impairment and Middle Ear Disorders” (ASHA, 1990), deleted the IAR as a screening parameter. This decision was based on studies that demonstrated that the inclusion of the IAR resulted in a high false-positive rate without adding to the hit rate and thus was a major contributor to over-referral. However, other investigations demonstrated the converse, that the IAR is a sensitive and contributory measure in a screening protocol. Our data suggest that the conflicting results are due to different operating systems in the middle ear analyzers that were used to elicit the IAR. Based on our findings, an optimal IAR screening protocol would use a pulsed elicitation immittance system (multiplexing circuit) and a maximum intensity level of 105 dB HL. Further, if the IAR screening were to be carried out using an automated (programmed) immittance system with an equivalent volume change criteria for IAR detection, the equivalent volume change of ≥0.03 ml is recommended.

1Division of Audiology/Hearing Conservation, Naval Hospital, Newport, Rhode Island; 2Department of Communication Disorders, Louisiana State University Medical Center, New Orleans, Louisiana

Reprint requests: Raymond M. Hurley, Department of Communication Disorders, LSU Medical Center, 1900 Gravier Street, New Orleans, LA 70112-2234; Tel: (504) 568-4340; Fax: (504) 568-4352; E-mail: rhurle@sumc.edu

Over-referral rate. To assess ASHA’s decision, Roush et al (1992) compared the previous ASHA middle ear screening protocol (ASHA, 1979) that included an IAR elicited at 105 dB SPL to the revised 1990 protocol that excluded the IAR and found that the revised protocol resulted in a slight decrease in the hit rate, but also in a large decrease in the false-positive rate over the 1979 protocol. The IAR was cited as the primary reason for the 1979 protocol’s poorer performance.

The belief that the IAR should be excluded from a middle ear screening protocol is not, however, universally supported. Silman et al (1992) modified the traditional ASHA procedure by increasing the maximum stimulus intensity level at 1000 Hz to 110 dB HL (117 dB SPL) and demonstrated significantly improved hit and false-positive rates. They maintained that the ASHA protocol failed to achieve the desired outcome because the 105 dB SPL stimulus level was not high enough to allow for physiologic maturation of the IAR response and to adjust for
Validity of the Ipsilateral Acoustic Reflex/Sells et al

nonpathological eardrums and/or middle ear variability. Further, Silman et al (1992) recommended that the IAR be assessed using an immittance instrument with a “multiplexing circuit.”

Nozza et al (1992) examined the effectiveness of the IAR in identifying the presence of middle ear effusion (MEE) in a group of children scheduled for myringotomies and in a group of children not suspected of having MEE. In the presurgery group, the IAR was found to be the most sensitive single measure to determine the presence of MEE; however, in the group of children not suspected of having MEE, the IAR was reported to be either not obtainable or absent in 28 percent of the ears that were otoscopically confirmed to be free of MEE. The reason for this discrepancy in the usefulness of the IAR between the two subject groups was not made clear; however, the authors urged caution when using the IAR in screening protocols.

The conflicting data from one study to another or within a study may be attributed to the operating system within the middle ear analyzer that was used to elicit the IAR. From an instrumentation perspective, the IAR has been plagued by artifact problems that are caused by the probe tone and the reflex activator tone being in the same ear. The Micro Audiometrics Earscan automatic tympanometer used by Roush et al (1992) controls for artifacts by limiting the maximum stimulus output to 105 dB SPL (personal communication, Micro Audiometrics, 1995). The Grason-Stadler 1723 used by Silman et al (1992) measures admittance changes in between rapidly pulsed presentations of the reflex activator tone, a multiplexing circuit that minimizes artifacts and allows for higher intensity presentation levels. The Grason-Stadler Middle Ear Analyzer Version 1 (GSI-33) used by Nozza et al (1992) is equipped with both a “Diagnostic” IAR mode that uses multiplexing circuitry and a “Screen” IAR mode that measures admittance changes with the probe tone and the stimulus tone on simultaneously. The Screen mode uses an equivalent volume change criterion of $\geq 0.05$ ml for a positive response (Grason-Stadler, 1989). Thus, the findings of IAR studies may be biased by the operating system used to elicit the IAR rather than the IAR’s failure to be a usable screening parameter. Therefore, it is reasonable to ask if all IAR measurements are equal or if the validity of the IAR is contingent upon the program within the instrument with which the IAR is elicited, as suggested by Silman et al (1992).

The present investigation used the GSI-33 (Version 2.0), which has (1) a Diagnostic IAR mode with multiplexing circuitry that measures admittance changes in between pulsed presentations of the stimulus tone and allows the user to control the stimulus intensity level to determine reflex threshold; and (2) a programmed Screen IAR mode that generates a tympanogram and then measures admittance changes with the probe tone and the stimulus tone on simultaneously using an equivalent volume change criterion of $\geq 0.05$ ml for IAR detection. This investigation addressed several questions: first, under the same conditions, do the GSI-33’s two modes of IAR elicitation yield different results; second, is the screening IAR response repeatable; and third, if the two operation modes produce different results, what are possible explanations for these findings?

**METHOD**

**Subjects**

A total of 363 ears were tested. Subjects were from 3 to 70 years of age and had hearing threshold levels $\leq 70$ dB HL from 250 Hz to 4000 Hz. The study was initiated immediately after the GSI-33 was calibrated by a GSI representative. Daily calibrations were conducted in addition to monitoring for wax collection in the probe. Total test time per patient was approximately 90 seconds.

**Procedure**

The following protocol was used to obtain the measures. First, the ear canal was sealed with the appropriate probe tip. Once the seal was obtained, the probe assembly was not touched again. The GSI-33 was set in the Screen mode and the first programmed screening run was recorded. The GSI-33’s default program screening stimulus is a 1800-msec 1000-Hz tone that steps in intensity from 85 to 95 to 105 dB HL (600 msec per step). A second screening run was ini-
tiated by touching the "Return" key followed by
the "Screen" key. Next, the unit was switched to
the Diagnostic mode. The IAR stimulus was
presented at 85, 95, and 105 dB HL using the
default programmed stimulus, which is a 1500-
msec sequence of pulsed 1000-Hz tone bursts
(multiplexing circuit). The Diagnostic mode run
was then repeated. Finally, the unit was switched
back to the Screen mode and a third screening run was recorded. The data were displayed on
the GSI-33's strip chart printout. The presence
or absence of the Screen mode IAR was indicated
by a "YES" or "NR" on the printout. The intensity
level and the amount of equivalent volume change for the Screen mode response are not dis-
played on the printout; however, both are
recorded for the Diagnostic mode.

RESULTS

Three experienced audiologists independently
determined the presence of the IAR in the
Diagnostic mode using the parameters of reflex
pattern, reflex growth, repeatability on the sec-
dond run, and the amount of equivalent volume change. The initial agreement among the audi-
ologists was ≥98 percent. All disagreements con-
cerned the intensity level at which the IAR was
elicted, not the presence or absence of an IAR,
and were resolved through mutual review.
Ninety-one percent of the ears were judged to
have exhibited an IAR and 9 percent of the ears
were judged not to have exhibited an IAR (Fig.
1a). Fifty percent of the ears exhibited a reflex
on both runs at the 85 dB HL level, 81 percent
of the ears exhibited a reflex on both runs at the
95 dB HL level, and 91 percent of the ears met
the criterion at the 105 dB HL level.

In the Screen mode, the IAR was found to
be present on all three runs 53 percent of the
time, to be absent on all three runs 20 percent
of the time, and to be a combination of present
and absent responses on the three runs (mixed responses) 27 percent of the time (see Fig. 1b).
For the three Screen mode runs, 64 percent of
the ears had a response on the first run, 67 per-
cent of the ears had a response on the second run, and 68 percent of the ears had a response on the third run. The response outcome on the three
Screen mode runs was not significantly differ-
ent ($\chi^2 = 1.81, p > .05$). Further, the overall relation-
ship between the intensity level at which the
Diagnostic mode IAR was reliably elicited and the IAR response pattern on the three screen-
ing runs was not significant ($r = .36, p > .05$).

Using the Diagnostic mode IAR as the "gold
standard," the IAR data were analyzed to deter-
mine how often the two elicitation modes were
in disagreement. These data are displayed in a
$2 \times 2$ matrix format in Figure 2. The first run

![Figure 1](a) The results of the two "Diagnostic" mode IAR runs that were scored by the three audiologists; (b) the IAR response patterns for the three repetitions of the "Screen" mode.

![Figure 2](2 × 2 matrix with the presence or absence of a "Diagnostic" mode IAR constituting the columns and the presence or absence of the first run "Screen" mode IAR constituting the rows. "Diagnostic" mode results are in disagreement with the first run "Screen" mode results 31 percent of the time.)
Validity of the Ipsilateral Acoustic Reflex/Sells et al

Screen mode disagreed with the Diagnostic mode 30 percent of the time, which was significant ($\chi^2 = 25.8, p < .001$). Ninety-four percent of the ears with conflicting results between the two modes had present Diagnostic mode IAR and an absent Screen mode IAR. Another way to view these data is to determine the extent to which subjects could be separated based on the presence or absence of a Screen mode IAR using the first screening run data. In other words, how effective would the Screen mode be in identifying individuals with an absent acoustic reflex? The hit rate or true positive rate (sensitivity) of the Screen mode IAR was 78 percent and the false-positive rate (1 – false-positive rate = specificity) or potential over-referral rate was 31 percent.

As the Screen mode uses a preset equivalent volume change of $\geq 0.05 \text{ ml}$ as the sole criterion for the presence of an IAR, the two Diagnostic mode IAR runs were analyzed to determine the percentage of ears that reached the $\geq 0.05 \text{ ml}$ equivalent volume change criterion at 105 dB HL. The results showed that 77 percent of the ears met the criterion on both Diagnostic mode runs, 16 percent did not meet the criterion on either run, and 7 percent met the criterion on only one of the two runs. The test–retest reliability for the equivalent volume change parameter was $r = .77$ for the two Diagnostic mode runs at the intensity level at which an IAR first occurred.

To analyze further the use of a preset equivalent volume change criterion, the pass and fail rates for volume changes of $\geq 0.02 \text{ ml}$, $\geq 0.03 \text{ ml}$, $\geq 0.04 \text{ ml}$, and $\geq 0.05 \text{ ml}$ at the intensity level at which an IAR was first detected and at 105 dB HL were calculated using data from the first Diagnostic mode run. These data are displayed in Table 1. Note that as the amount of equivalent volume change decreases from $\geq 0.05 \text{ ml}$ to $\geq 0.02 \text{ ml}$, the number of passes increases and the number of fails decreases. The risk of a small equivalent volume change criterion is an unacceptable false-negative rate, whereas the risk of a large equivalent volume change criterion is an unacceptable false-positive rate. Note, however, that at the $\geq 0.02 \text{ ml}$ equivalent volume change criterion the pass and fail rates are similar to the Diagnostic mode pass and fail rates of 91 percent and 9 percent.

**DISCUSSION AND CONCLUSIONS**

This investigation demonstrated that the GSI-33’s two modes of IAR elicitation yield significantly different results 30 percent of the time and, further, that the Screen mode IAR response is inconsistent on three repetitions 27 percent of the time. Thus, all IAR measurements are not equal and the validity of an IAR is contingent upon the operating system with which the IAR is elicited.

Based on these data, Roush et al (1992) might have shown a wide range of results if they had used a variety of middle ear analyzers to determine the effectiveness of the 1979 ASHA protocol versus the 1990 ASHA protocol. Also, these data help explain the differences Nozza et al (1992) found in the utility of the IAR between their two subject groups. If the group who were not suspected of having MEE have been assessed using the Diagnostic mode of the GSI-33 instead of the Screen mode, different conclusions regarding the use of the IAR in screening protocols might have been drawn. Interestingly, Nozza et al’s (1992) presurgery subject group and Silman et al’s (1992) subject group were both tested using a similar IAR elicitation system and both studies showed the IAR to be highly sensitive in detecting MEE.

The reason for the GSI-33’s IAR elicitation modes producing significantly different results is twofold. First, the Screen mode uses a simultaneous presentation of the stimulus tone and probe tone, which does not control for artifacts as effectively as the pulsed presentation used in the Diagnostic mode. Additionally, artifacts are most likely the cause of the demonstrated inconsistency in the Screen mode results. Second, the Screen mode uses a relatively large preset equivalent volume change criterion to determine the presence of a response. The $\geq 0.05 \text{ ml}$ criterion ensures that a positive response is truly indicative of a present reflex; however, it also ensures a high false-positive rate. Based on
our findings, an optimal IAR screening protocol would use a pulsed elicitation immittance system (multiplexing circuit) and a maximum intensity level of 105 dB HL. Further, if the IAR screening were to be carried out using an automated (programmed) immittance system with an equivalent volume change criterion for IAR detection, the equivalent volume change of ≥0.03 ml is recommended. In short, the assertion of Silman et al. (1992) that an immittance screening protocol should include an IAR measured with a pulsed elicitation system (multiplexing circuit) is supported.

In summary, although the GSI-33's Diagnostic and Screen modes attempt to elicit the same event, they produce significantly different results, with the additional factor of the Screen mode results being inconsistent on three consecutive runs over a quarter of the time. The primary reason for these findings is that the two modes use different reflex elicitation systems and different criteria to determine the presence of a response. The validity of an IAR measurement and the utility of the IAR in an immittance protocol is clearly dependent upon the operating system used to measure and quantify the response.

Acknowledgment. This paper is based on a poster presentation at the Annual Convention of the American Academy of Audiology, Dallas, TX, 1996. We are grateful to Shlomo Silman, Ph.D. and two anonymous reviewers for their valuable comments.

REFERENCES


