Similarities and Differences in Distortion-Product Otoacoustic Emissions Among Four FDA-Approved Devices

Terandalur K. Parthasarathy
Barbara Klostermann

Abstract

Similarities and differences in distortion-product otoacoustic emissions (DPOAEs) among four Food and Drug Administration (FDA) approved devices were assessed quantitatively. DPOAEs were recorded from 42 adult subjects (84 ears) ranging in age from 19 to 40 years. All subjects had hearing thresholds of 20 dB HL or better for the test frequencies from 0.25 to 8.0 kHz and normal acoustic immittance findings. DPOAEs were collected in a quiet non-sound-treated room. DPOAE measurement parameters included an f\textsubscript{2}/f\textsubscript{1} ratio of -1.2, with f\textsubscript{1} at 65 dB SPL (L\textsubscript{1}) and f\textsubscript{2} at 55 dB SPL (L\textsubscript{2}). There were no significant differences in the mean emissions levels among the four devices. This investigation showed that validity criteria, pass criteria, and strategies for DPOAE measurements interact to produce varying pass and refer results. However, when DPOAEs are obtained with consistent validity criteria, pass criteria, and strategies for measurement, the results are remarkably consistent.

Key Words: Distortion-product otoacoustic emissions, otoacoustic emissions, outer hair cells, signal-to-noise ratio, universal hearing screening

Abbreviations: ANOVA = analysis of variance, CNS = central nervous system, DPOAEs = distortion-product otoacoustic emissions, FDA = Food and Drug Administration, L = level, OAEs = otoacoustic emissions, SNR = signal-to-noise ratio

Otoacoustic emissions (OAEs) have been recommended for use in universal newborn hearing screening by the Joint Committee on Infant Hearing (2000) and the National Institutes of Health (1993) consensus panel. OAEs are well suited for newborn screening for several reasons. First, patient preparation time is minimal, and the responses are independent of behavioral state (Stach and Santilli, 1998). Second, OAEs are not affected by neuromaturation of the central nervous system (CNS); thus, OAEs can be obtained in newborns independently of gestational age (Popelka et al, 1995, 1998; Abdala, 1996). Finally, OAE screening can be carried out quickly and efficiently by minimally trained personnel (Stach and Santilli, 1998). The clinical advantages of using OAEs for newborn hearing screening may be compromised, however, because OAE levels vary significantly in relation to the test signal parameters that have not been standardized. Further, the criteria for what constitutes a pass result have also not been standardized and differ across devices. In addition, there is a wide range of signal-processing schemes used to separate the tonal distortion-product otoacoustic emission (DPOAE) signal from the ambient noise.

Hornsby et al (1996) published test results comparing DPOAEs for five Food and Drug Administration (FDA)-approved devices (i.e., Virtual 330, Etymotic Research/Mimosa Acoustic CubDis, Madsen Celesta, Grason Stadler 60, and Biologic Scout). Distinct and statistically significant differences in mean emission levels existed among the five devices, particularly among the higher frequencies (5 and 6 kHz). The differences in the results for the high frequencies have been attributed to ear canal acoustics resulting in standing wave effects in the outer ear canal and crosstalk between the sound source and probe microphone (Siegel, 1994, 1995).
Vinck et al (1996) conducted a normative study of DPOAEs using the ILO 92 system. Analysis of the DPOAEs showed distinct, measurable emitted responses for all of the 101 normal ears. DPOAEs displayed a bimodal distribution with maximum levels at 1.44 and 4.561 kHz. In addition, there was a decrease in the emission levels occurring at about 2 to 3 kHz for all stimulus intensity levels (70-80 dB SPL in 5-dB steps) of stimulation. This finding, in part, could be a result of resonance phenomenon encountered during the reverse transmission of the vibratory energy through the middle ear (Lonsbury-Martin et al, 1990) or perhaps through multiple pathways (Popelka et al, 1993).

Gorga et al (2000) measured the DPOAE and noise floor in newborn babies at f2 frequencies of 1.0, 1.5, 2.0, 3.0, and 4.0 kHz and for primary levels of (L1/L2) of 65/50 dB SPL and 75/75 dB SPL. DPOAEs in the vast majority of subjects were unaffected by subject group (neonatal intensive care unit vs well-baby nursery), test environment, and state of arousal. The signal-to-noise ratio (SNR) of the DPOAEs was significantly affected by frequency. Poor SNR was obtained for f2 frequencies below 1.5 kHz, and the SNR increased systematically as the f2 frequency increased. Furthermore, their results provided clear evidence that reliable DPOAEs in newborn neonates can be recorded by focusing on higher frequencies—2.0, 3.0, and 4.0 kHz—and that babies can be reliably tested in the arousal states that are commonly encountered in the newborn nurseries.

The proliferation of newborn hearing screening programs throughout the world has resulted in the development of numerous devices to perform quick, efficient, and cost-effective screening. A variety of instruments are now FDA-approved for the measurement of DPOAEs. However, these devices use different pass criteria, different test frequencies, and different noise reduction strategies. The tonal DPOAE at 2f1−f2 for these test stimuli is typically around 5 dB SPL, well below ambient noise levels, even in the quietest settings. As a result, signal averaging is used to reduce the ambient noise. Some devices quantify the noise level as the level in an adjacent frequency region, although what constitutes the adjacent frequency region differs substantially across devices. Other devices quantify the noise specifically at the DPOAE frequency. The amount of signal averaging also differs across devices. Signal averaging proceeds either until the level of the noise falls a specified value below the DPOAE (at least 3 dB but to as much as 12 dB lower) or for a specified averaging time or number of samples that also differs across devices. In certain cases, a valid DPOAE may go undetected either because the noise level was inadequately defined or simply because there was insufficient signal averaging.

Therefore, the purpose of this article was to assess the differences and similarities across four FDA-approved DPOAE screening devices under controlled settings. A secondary goal was to provide an understanding of the effect that a lack of standardization can have on screening results.

METHOD

Subjects

Forty-two subjects (38 females and 4 males) between the ages of 19 and 40 years (mean = 21.5, SD = 2.1) volunteered to participate under protocols approved by an Institutional Review Board (IRB). Subjects were recruited from our graduate program and/or were subjects involved in ongoing research projects. All subjects met the following inclusion criteria: (1) physically and neurologically normal based on subject interview; (2) negative family history of hearing loss and neurologic problems; (3) negative history of ototoxic drug use, excessive noise exposure, middle ear disease, and metabolic diseases associated with hearing loss; (4) external auditory canals free from obstruction based on an otoscopic examination; (5) normal hearing sensitivity at or better than 20 dB HL for both ears by air conduction at octave intervals from 0.25 to 8.0 kHz and by bone conduction at octave intervals from 0.25 to 4.0 kHz; (6) no gap between air- and bone-conduction thresholds greater than 10 dB at any measured frequency; (7) speech reception thresholds consistent with the pure-tone thresholds; (8) speech recognition scores using the Central Institute for the Deaf Word List 22 W-22 word list at 30 dB SL better than 90 percent for both ears; and (9) normal acoustic immittance results for both ears defined as a peak pressure between -100 and +50 daPa and static values between 0.4 and 1.8 mL.

Equipment

Four FDA-approved DPOAE systems were used. Three were hand-held, battery-operated devices designed specifically for screening: the Audioscreener (Everest Biomedical Instruments), the AuDx (Biologic Corporation), and the EroScan (Etymotic Corporation). The fourth
was a conventional desktop diagnostic device, the ILO 92 system (Otodyamics), which has often been used for the collection of previously published data.

Procedure

DPOAEs were measured in response to the primary tones \( f_1 \) and \( f_2 \), with \( f_1 \) equal to 65 dB SPL and \( f_2 \) equal to 55 dB SPL for all OAE measurements. The primary frequency ratio \( f_1/f_2 \) was held constant at ~1.2, and the DPOAE frequency of \( 2f_1-f_2 \) was measured in all cases. The frequency series under investigation ranged from 2.0 through 6.0 kHz in 1.0-kHz steps, although the number of frequencies differed across devices. The \( f_1 \) frequency was set to 2.0, 3.0, 4.0, 5.0, and 6.0 kHz for the Audioscreener and the ILO 92 system and to the default settings of 2.0, 3.0, 4.0, and 5.0 kHz for the AuDx and 2.0, 3.0, and 4.0 kHz for the EroScan. No attempt was made to obtain data below 2.0 kHz because the measurement of DPOAEs at the lower frequencies is often unreliable (Gorga et al., 2000). Following audiometric and middle ear assessments, the cubic \( (2f_1-f_2) \) DPOAEs for each device were obtained and repeated in both ears of each subject, resulting in 16 frequency series each, or a total of 672 frequency series. All DPOAE measurements were made in a quiet room adjacent to the clinic. None of the DPOAE data reported in this research were collected in a sound-treated room to simulate the screening environment. Each of the subjects was seated in the same quiet environment with minimal amounts of background noise. The default signal averaging scheme for each device was used in all cases. After the probe tip was inserted into the ear, the DPOAEs and associated noise floors were measured. All DPOAE and associated noise floors were placed into a single database to form the basic raw data. All pass criteria were then applied to all of the data in the database rather than relying on a pass result provided by the individual devices. A stopwatch was used to measure the amount of time to adequately couple the probe to the ear and the amount of time for the measurement itself. The probe was left in the ear to retest for each ear and for each device.

RESULTS

The mean and standard deviation for the time required to couple the probe to the ear are shown in Figure 1. These times were just under 10 seconds for the screening devices and slightly longer, at just over 20 seconds, for the diagnostic device.

The mean amount of time required for each device to obtain the measures for a series of frequencies is shown in Figure 2, along with 1 SD. Because the devices differed in the number of frequency regions measured in a series, from three to six frequency regions, the times were normalized to a series of four frequency regions for each device. The mean processing times were

![Figure 1](image1.png)  
**Figure 1** The mean (+1 SD) amount of time to couple the probe to the ear for each device.

![Figure 2](image2.png)  
**Figure 2** The mean (+1 SD) amount of time for each device to obtain the measures for a series of frequency regions. Because the devices differed in the number of frequency regions measured in the series ranged from three to six, the times were normalized to a series of four frequency regions for each device.
under 17 seconds for the screening devices and just over 27 seconds for the diagnostic device.

The mean grouped DPOAE levels for each device were plotted as a function of the \( f_2 \) frequency in Figure 3 along with ± 1 SD. Also plotted are neonate data from Gorga et al (2000). The data points at each frequency have been shifted slightly on the frequency axis to allow visualization of individual data points that overlapped.

![Figure 3](https://example.com/figure3.png)

**Figure 3** Mean DPOAE levels at \( 2f_1-f_2 \) (± 1 SD) for each device as a function of the \( f_2 \) frequency. Also plotted are neonate data from Gorga et al, 2000. The data points at each frequency have been shifted slightly on the frequency axis to allow visualization of individual data points that overlapped.

Statistical analysis of variance (ANOVA) at each frequency showed that there was no significant difference in the mean DPOAE levels between ears (\( p < .001 \)) or between the two measures (\( p < .001 \)), allowing data from the four conditions to be grouped at each frequency for each device. Furthermore, ANOVA at each frequency showed that there was no significant difference in the mean grouped DPOAE levels among the four devices (\( p < .001 \)).

The distributions of the noise measures were plotted at each frequency for each device and fitted with a normal curve. These distributions were quite similar across devices and fell into two categories. Figure 4 shows two representative examples of the distributions of the noise levels obtained at two different test frequencies for a single device (Audio screener). The distribution in the left panel is centered at around 0 dB SPL, is symmetric, and follows a normal distribution. The distribution in the right panel has a peak at around –8 dB SPL but is clearly not normally distributed or symmetric because of the large number of measures at –20 dB SPL. This asymmetry is attributable only to the fact that the lower level of noise measures was limited—20 dB SPL, a value low enough to obviate any additional signal averaging.
averaging. Although these factors prevented any summary of noise floor data across devices, or conditions, all of the noise floor measures at each frequency for each device for the individual devices fell into either of the two categories represented by the examples in Figure 4.

The two most common methods of establishing pass criteria were evaluated: one based on the absolute level of a valid OAE (level method) and one based only on the presence or absence of a valid OAE (presence method). Each of these methods was evaluated further regarding the number of frequencies and the number of measures that were necessary for the ear to have been given a pass result. The two methods are illustrated in Figure 5. Each panel shows the levels of the two stimuli, f₁ and f₂, the level of the OAE, 2f₁−f₂, and the level of the noise in dB SPL. The upper left panel illustrates a typical OAE response with no signal averaging. The OAE cannot be detected because it is well below the noise level. The upper right panel illustrates a typical OAE response with signal averaging. The OAE can now be detected because the signal averaging greatly reduced the noise level, and the absolute value of the OAE (level) and the SNR for the response can now be determined. Once the OAE is detectable, additional signal averaging will not affect the level of the OAE but will greatly affect the SNR. The lower left panel illustrates a pass result for the level method. A valid OAE (the SNR exceeded or equaled the criterion level of 6 dB) exceeded or equaled a criterion OAE level (−5 dB SPL). The lower right panel illustrates a pass result for the presence method. A valid OAE (the SNR exceeded or equaled the criterion level of 6 dB) exceeded or equaled a criterion SNR level (the same SNR of 6 dB). Note that the absolute level of the OAE is not considered in the presence method.

The criterion for a valid OAE across published studies has varied greatly from an SNR as low as 3 dB (many studies) to as high as 12 dB (Popelka et al, 1993). An OAE result that fails to meet this criterion is treated differently across devices. For diagnostic devices, such results are not marked as valid or invalid, leaving it up to the clinician to decide on an individual basis. For screening devices, an OAE is first marked as valid or invalid based on a criterion SNR. Invalid OAE results are marked as “noisy” in some devices but as “refer” for other devices. Depending on the screening program protocol, when a “noisy” result occurs, the operator may immediately perform a retest, may actively reduce the amount of noise in the environment and then perform a retest, or may simply extend the amount of signal averaging. In all cases, these strategies interact with the settings of the devices and can cause changes in the referral rates, changes that are likely independent of the devices. To demonstrate this effect, the number of valid OAE results for all of the data was computed first with a SNR criterion set to 6 dB and then relaxed to 5 dB. Figure 6 shows the percent increase in the number of valid OAE responses for each of the devices at the individual frequencies. Without regard to processing time, in practical terms, a 1-dB change in the SNR criterion for a valid OAE can increase the overall number of valid OAE responses from 2 to 5 percent for most devices. In our database, where only normal ears were tested, this can translate directly to a substantial reduction in the corresponding refer rate. A screening program can halve a 10 percent refer rate to a 5% refer rate by simply changing the criterion SNR by 1 dB.

The number of valid OAE results can also be affected by whether a retest is allowed in the screening protocol. The data in Figure 3
show clearly that even in normal ears under optimal testing conditions, an OAE level will fall below ~ -5 dB SPL for any single measure a small proportion of the time. The data in Figure 4 also show clearly that the noise level for a single OAE measure can exceed 0 dB SPL a small proportion of the time. These two conditions result in a number of measures that will be considered invalid based on statistical sampling alone for both the level and the presence methods. There are two approaches to dealing with this problem. The first is to perform a single retest and accept the better of the two measures. This is a good strategy when considering that the total test time is so short, as illustrated in Figure 2. A second approach is to provide the option of systematically discarding a measure at a single frequency. This is accomplished by counting as valid measures only those from a subset of the total number of frequencies measured for the ear, three of four, for example. This is also a good strategy, not only in view of the minimal time it takes to increase the number of frequencies but also because the function relating OAE level and frequency in normal ears can have areas of limited response (Kemp et al, 1990).

Several pass criteria were applied to all data in the database. First, the validity of the DPOAEs was ascertained by specifying a criterion SNR of 6 dB for the response to be considered valid for each measure in the database. This value was sufficiently high to avoid interacting with the stopping criteria across devices, allowing an appropriate consideration of all data in the database, regardless of the device. Second, the best of two consecutive measures was considered rather than each individual measure to reduce the number of invalid responses attributable only to statistical variability. Once these validity criterion were met, several additional pass criteria were evaluated.

For the level method, each valid OAE had to equal or exceed a criterion level to receive a pass result. Criterion levels were selected based on the values used in several devices and are given in Table 1. For the presence method, each valid OAE was considered a pass. The absolute level of the OAE was not considered. For both the level and presence methods, additional pass

### Table 1  Criterion Levels for a Pass Result at Each Frequency

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>2000</th>
<th>3000</th>
<th>4000</th>
<th>5000</th>
<th>6000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion (dB SPL)</td>
<td>-7</td>
<td>-8</td>
<td>-5</td>
<td>-7</td>
<td>-7</td>
</tr>
</tbody>
</table>

Figure 6  Percent increase in the number of valid OAE responses by changing the SNR from 6 to 5 dB.

Figure 7  Increase in the pass rate for both the level and the presence methods for the four devices when using three frequencies.
Figure 8  Increase in the pass rate for both the level and the presence methods for the four devices when using four frequencies.

Figure 9  Increase in the pass rate for both the level and the presence methods for the four devices when using a pass result at any three of the four frequencies tested.

Figure 10  Absolute refer rate for both the level and the presence methods for the four devices when using one repeat measure and four frequencies but with pass criteria applied to only three of the four frequencies.

Criteria were evaluated to determine if the ear received a pass result. These criteria were based on the number of measures at individual frequencies that produced a pass result. For the ear to be considered a pass, a pass result was required at all three frequencies of 2.0, 3.0, and 4.0 kHz, at all four frequencies of 2.0, 3.0, 4.0, and 5.0 kHz, or at any three of four frequencies of 2.0, 3.0, 4.0, and 5.0 kHz for a single test.

Figures 7 through 9 show the increase in the pass rate for both the level and the presence methods for the four devices when selecting the best of two consecutive frequency sequences. Figure 7 shows the increase in the pass rate when using four frequencies. A pass result at all four frequencies was necessary for the ear to be considered a pass. Figure 8 shows the increase in the pass rate when using three frequencies. A pass result at all three frequencies was necessary for the ear to be considered a pass. Figure 9 shows the increase in the pass rate when using only three of four frequencies. A pass result at any three of four frequencies was necessary for the ear to be considered a pass. Only those devices that were set to measure at least four frequencies were evaluated.

The data in the previous figures relate only to changes in pass rates in relation to changes in pass criteria. However, because the database contained data only from normal ears, an absolute refer rate can also be computed. Figure 10 shows the refer rate for both the level and the presence methods for the four devices when using the best of two consecutive measures at any individual frequency and with pass criteria applied to only three of the four frequencies.

**DISCUSSION**

On average, as shown in Figures 1 and 2, the desktop device required slightly more time to couple the probe to the ear and to make the measures. However, there were no practical differences in these times across the screening devices. It is clear from this data analysis that validity criteria, pass criteria, and strategies
for measurement interact to produce varying pass and refer results. However, when data from the devices are collected with consistent validity criteria, pass criteria, and strategies for measurement, the results are remarkably consistent. The mean data in Figure 3 are so close at some frequencies that the data points completely overlapped. The mean data for adults are also extremely close to similar DPOAE data in adults (Hornsby et al, 1996) and more recently published data for neonates (Gorga et al, 2000), even though the stimulus parameters differed slightly. The fairly low variability in the data was also remarkable. The standard deviations were nearly identical across devices and across frequencies and were quite small.

The normal variability in the data, although relatively small, can cause some problems for pass criteria based on the absolute level of the DPOAE. The values for −1 SD are already very close to the pass criterion levels in Table 1. If −2 SD are considered, a value that represents only ~95 percent of the measures at a frequency, many measures fail to reach the pass criterion. It is readily apparent that a single DPOAE measure at any single frequency with any individual device will fail to produce a pass result in a normal ear every time.

Normal variability in the OAE data is less of a problem for criteria based only on the presence or absence of an OAE because the absolute level of the DPOAE is not considered. However, normal variability in the noise data can still be a problem for criteria based only on the presence or absence of an OAE because of variability in the absolute levels of the noise. In either case, pass criteria based only on the presence or absence of an OAE with no consideration of the absolute level of the OAE can result in an increase in the number of hearing-impaired subjects who receive a pass.

Regarding pass criteria, the data in Figures 7 through 9 suggest rather strongly that a single repeat measure will substantially increase the pass rate in normal ears regardless of the presence or absence of an OAE because of variability in the absolute levels of the noise. In either case, pass criteria based only on the presence or absence of an OAE with no consideration of the absolute level of the OAE can result in an increase in the number of hearing-impaired subjects who receive a pass.

If a pass result is not obtained on the first test, the testing is finished. If a pass result is not obtained on the first test, a repeat test should be conducted, even without removing the probe. If a pass result is obtained on the second test, the testing is finished. If a pass result is not obtained on the second test, a repeat test should be conducted but at a later time to allow any patient-related conditions to resolve, such as debris in the external ear canal. Further repeat testing should not be performed for the same reason that a single repeat test should be performed.

The noise levels for each measure across devices cannot easily be compared because of different, yet equally valid signal averaging schemes. In general, the linear averaging methods used in these devices result in noise levels that continue to decrease the longer the signal averaging occurs. Therefore, mean noise levels will differ substantially for devices that stop averaging when a criterion noise level is reached compared with those devices that average for a specific duration. Another complication is the fact that most of the devices place a lower limit on the lowest noise level they are capable of measuring, a value that may differ across devices. Finally, the devices all differ in how the noise is measured, using noise measures at various locations adjacent to the OAE or directly at the OAE frequency.

The data in Figure 10 suggest that a single repeat measure with a pass result based on three of four frequencies will result in a very low refer rate in normal ears regardless of the method (level or presence).

CONCLUSIONS

1. Measurement performance is associated more with stimulus and signal averaging considerations and should be separated from pass/refer performance, which is associated more with pass criteria differences.
2. Each of the four devices will produce similar OAE measurements if all measurement parameters have been set to be equivalent.
3. Each of the four devices will produce similar pass and refer rates if pass refer criteria are set to be equivalent.
4. An OAE measure can be considered valid with a criterion SNR of 6 dB, although other values may also be valid.
5. A second measure performed immediately after obtaining a refer result without removing the probe will substantially reduce the false-positive responses owing to normal variability.
6. Pass criteria based on pass results for any three of four frequencies will result in refer rates in normal-hearing ears well under 2 percent when using the presence method.
7. Pass criteria based on pass results for any three of four frequencies will result in refer
tates in normal-hearing ears well under 3 percent when using the level method.
8. The present research should be replicated
in newborn neonates under different test
environments and states of arousal.

Acknowledgment. The authors wish to thank Dr. Roanne
K. Karzon and three anonymous reviewers for their helpful
suggestions on an earlier version of this article.

REFERENCES

Abdala C. (1996). Distortion product otoacoustic emis-
sion (2f, f) amplitude as a function of frequency ratio
and primary tone level separation in human adults and

Gorga MP, Norton SJ, Siniger YS, Cone-Wesson B, Folsom
of neonatal hearing impairment: distortion product oto-
acoustic emissions during perinatal period. Ear Hear
21:400–424.

Harris F, Probst R. (1997). Otoacoustic emissions and
audiometric outcomes. In: Robinette MS, Glattke TJ, ed.
Otoacoustic Emissions: Clinical Applications. New York:
Thieme Medical, 151–180.

for five FDA-approved distortion product OAE system.

Joint Committee on Infant Hearing. (2000). Principles
and guidelines for early hearing detection and interven-


Lonsbury-Martin BL, McCoy MJ, Whitehead ML, Martin
GR. (1993). Clinical testing of distortion-product otoa-
coustic emissions. Ear Hear 1:11–22.

National Institutes of Health. (1993). Early identifica-
tion of hearing impairment in infants and young children.
Consensus statement. Washington, DC: Public Health
Service.

Popelka GR, Osterhammel PA, Nielsen LH, Rasmussen
emissions with primary-tone level in humans. Hear Res
71:12–22.

of the 2f,-f, distortion product otoacoustic emission for
low level stimuli in human neonates. Ear Hear
16:159–165.

of noise sources that influence distortion product otoa-
coustic emission measurements in human neonates. Ear

Siegel JH. (1994). Ear canal standing waves and high-
frequency sound calibration using otoacoustic emission


Vinck MB, Del Vel ED, Xu ZM, Van Cauwenberge PB.
(1996). Distortion product otoacoustic emissions: a nor-