**Naval Submarine Medical Research Laboratory**

**NSMRL TECHNICAL REPORT #TR1231 February 26, 2004**

# STIMULUS-FREQUENCY OTOACOUSTIC EMISSIONS

## VALIDITY & RELIABILITY OF SFOAES IMPLEMENTED ON MIMOSA ACOUSTICS SFOAE MEASUREMENT SYSTEM V2.1.18

by

**Judi A. Lapsley Miller, Paul Boege, Lynne Marshall, Christopher A. Shera, & Patricia S. Jeng**

Released by:
CAPT G. A. Higgins, MSC, USN
Commanding Officer
NavSubMedRschLab

Approved for public release; distribution unlimited.

# STIMULUS-FREQUENCY OTOACOUSTIC EMISSIONS

## VALIDITY & RELIABILITY OF SFOAES IMPLEMENTED ON MIMOSA ACOUSTICS SFOAE MEASUREMENT SYSTEM V2.1.18

**by**

**Judi A. Lapsley Miller, Paul Boege, Lynne Marshall, Christopher A. Shera, & Patricia S. Jeng**

Approved and Released by:

CAPT G. A. HIGGINS, MSC, USN
Commanding Officer
NavSubMedRschLab

Approved for public release; distribution unlimited.

# SUMMARY

The Problem
The validity and reliability of stimulus-frequency otoacoustic emissions was unknown.

The Findings
The validity and reliability of stimulus-frequency otoacoustic emissions was determined for the Mimosa Acoustics SFOAE Measurement System v2.1.18.

Application
Stimulus-frequency otoacoustic emissions are sounds made by healthy inner ears to tonal stimulation. They are thought to hold the potential for diagnosis and monitoring of noise-induced hearing loss in military hearing-conservation programs. By establishing the validity and reliability of the measurement, changes in SFOAEs due to noise can be detected from artifact and test-retest variability.



Administrative Information. “This work was conducted under Work Unit #50309, entitled “Evoked Otoacoustic Emissions in Military Hearing Conservation Programs.” The opinions or assertions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the Department of the Navy, Department of Defense, or the United States Government. This research has been conducted in compliance with all applicable Federal Regulations governing the Protection of Human Subjects in Research.” This Technical Report was approved on March 15, 2004, and designated as NSMRL Technical Report #TR1231.

# STIMULUS-FREQUENCY OTOACOUSTIC EMISSIONS

## VALIDITY & RELIABILITY OF SFOAES IMPLEMENTED ON MIMOSA ACOUSTICS SFOAE MEASUREMENT SYSTEM V2.1.18

BY JUDI A. LAPSLEY MILLER, PAUL BOEGE, LYNNE MARSHALL, CHRISTOPHER A. SHERA, & PATRICIA S. JENG

## ABSTRACT

The validity and reliability of stimulus-frequency otoacoustic emissions (SFOAEs) were evaluated with the Mimosa Acoustics SFOAE measurement system (v. 2.1.18). This system is an implementation of Shera's method (Shera & Guinan, 1999). SFOAEs were measured in clusters of 5 frequencies centered on the frequencies 1, 1.4, 2, 2.8, 4, and 5.6 kHz (probe stimulus level 40 dB SPL, suppressor stimulus level 55 dB SPL). On average, low-level artifactual responses (1.2 dB) were seen in hearing-impaired ears (N = 7) that should not produce SFOAEs. However, the use of an SFOAE-to-noise (SNR) criterion of 9 dB eliminated 99.5% of artifactual responses in the hearing-impaired ears. SFOAE group delay was used to validate SFOAE measurements – no measurements in hearing-impaired ears showed group delays that were consistent with SFOAE theory. In normal-hearing ears (N = 30), group delays based on 3 to 5 SFOAEs measured within a 50 Hz band were deemed valid if the correlation for the unwrapped phase as a function of frequency was less than -.94 and the least-squares regression error term less than 2 ms. These criteria accurately distinguished between measurements made in hearing-impaired and normal-hearing ears. If group-delay information was available, an SNR criterion of only 0 dB was enough to ensure group-delay validity and SFOAE-amplitude validity within a cluster. For single SFOAE measurements, an SNR criterion of around 9 to 10 dB (which is much lower than the 15 dB recommended by Shera) was shown to be sufficient for validity. Test-retest reliability in the normal-hearing ears (ears = 61, participants = 33) measured using the standard error of measurement ranged from 0.7 to 2.3 dB, which is comparable to reliability for other OAE types. Minor enhancements to the measurement system could improve validity and reliability further.

## ACKNOWLEDGEMENTS

Thanks to Linda Westhusin, Linda Hughes, Chris Duplessis, Rebecca Christian, Keith Wolgemuth, and Cindy Burgess-Russotti for help with data collection. This research was supported the US Navy Bureau of Medicine and Surgery and the US Army Medical Research and Materiel Command. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the United States Government. This research has been conducted in compliance with all applicable Federal Regulations governing the protection of human subjects in research.

## INTRODUCTION

The Mimosa Acoustics SFOAE measurement system is an implementation of the method developed by Shera (Shera & Guinan, 1999) for measuring stimulus-frequency otoacoustic emissions (SFOAEs). SFOAEs are sounds produced by the inner ear when stimulated with a tone (known as the probe stimulus). The SFOAE is produced at the same frequency as the probe stimulus making it difficult to directly measure the SFOAE. In Shera's method, a suppressor stimulus is also presented at a nearby frequency. The effect of the suppressor is to eliminate the production of the SFOAE. By cycling the suppressor on and off while the probe is on continuously, the vector difference in sound pressure measured during the probe-alone and probe+suppressor intervals reveals the SFOAE.

SFOAEs are considered to be the most frequency-specific OAE measurement. They are also thought to consist primarily of reflection-source (place-fixed) OAEs, rather than distortion-source (wave-fixed) OAEs in Shera's taxonomy of OAE production (Shera & Guinan, 1999). Reflection-source OAEs are of interest for use in hearing-conservation programs, because they are thought to be more vulnerable to damage to the inner ear from sound overexposure.

To date, the use of SFOAE measurements has been restricted to research laboratories where the emphasis is on theory, not application. There is little to no information about the validity and reliability of SFOAEs in the literature that could be used to develop the measurement technique for clinical usage. This report is a start towards that goal. Here, we consider whether the current SFOAE implementation produces valid and reliable SFOAE measurements.

Validity is concerned with establishing that the SFOAE measurements consist primarily of an evoked stimulus-frequency otoacoustic emission, and are not contaminated with stimulus waveform energy, cross talk, noise, or other measurement artifacts. Reliability is concerned with the repeatability of the measurements over time.

Dealing with artifact and noise is particularly important for hearing conservation, because the people in hearing-conservation programs often have low OAE levels (unlike in infant screening programs). In such populations, low-level artifactual responses may be confused for real SFOAEs and high noise levels may hide real SFOAEs. Furthermore, one aim in hearing conservation is to detect small changes in emission level. Noise and artifact both hinder such measurements. Once identified, artifacts may in some circumstances be reduced or removed with signal processing, a redesign of the stimulus presentation method, or using stricter criteria to define what constitutes an acceptable measurement.

If we want to reliably detect small changes in SFOAEs due to sound overexposure in individuals, test-retest variability must be minimized. Otherwise changes in SFOAEs may be attributed to sound overexposure when in fact they are due to other factors. For reliability to be improved, the sources of the variability need to be identified and reduced or removed. Some causes could be noisy calibration, contamination from deterministic or correlated noise during the measurement (which is not necessarily removed with artifact and noise rejection procedures), and variations in actual stimulus levels. The effect of test-retest variability may also be reduced

by using stricter criteria as to what constitutes an acceptable measurement, but at the expense of losing data.

We cannot easily determine in normal-hearing ears whether an OAE measurement is contaminated with artifact. Some artifacts can be detected in an artificial ear (a mechanical coupler, which simulates the acoustic properties of an average ear, but which does not produce OAEs). However, some artifacts may be apparent only when the measurement is made in a human ear (Heller & Marshall, unpublished data). Mimosa Acoustics has established the validity of the SFOAE implementation in an artificial ear (the B&K 4157 acoustic coupler); the results are reported in the technical documentation. Here, two data sets are used to evaluate validity and reliability of SFOAEs in human ears: one set comes from 7 hearing-impaired ears and the other set from 33 normal-hearing ears. Both data sets were collected using the same SFOAE protocol.

The initial analysis to determine validity considers the SFOAE-to-noise ratio (SNR) at each measurement frequency in the hearing-impaired ears. The responses for these ears should be indistinguishable, on average, from the noise level. If the response is consistently above the noise, then artifact will be assumed to be present and the causes investigated. Statistically, there will sometimes seem to be an SFOAE present, even in an acoustic coupler. Therefore, the range of SFOAE responses will also be considered.

Then, we investigate if the SFOAE group delay can be used to validate an SFOAE measurement. Group delay is defined as the negation of the first derivative (the slope) of the unwrapped phase as a function of probe-stimulus frequency (the phase spectrum). Its unit is milliseconds (when calculated like Shera & Guinan, 1999). If the SFOAE phase spectrum is mapped out in fine detail (in a normal ear), then it will have a steep negative slope because the auditory filter introduces a delay—the sharper the filter, the longer the delay (Shera & Guinan, 2003; Shera, Guinan, & Oxenham, 2002). If the SFOAE spectrum is mapped out in an acoustic cavity (an artificial ear), the phase spectrum is expected to be flat, because changes in phase should either occur randomly due to noise or be constant if there is some artifactual response. In a hearing-impaired ear, where there is little-to-no filtering, it is expected that the phase response should resemble an acoustic cavity.

If the phase spectrum is sampled too sparsely, it may not be possible to uniquely unwrap the phase (due to aliasing). However, it takes approximately an hour to obtain the entire spectrum over the frequency range of interest. Instead, we sample the phase spectrum in clusters of 5 closely-spaced frequencies centered on key frequencies (e.g., 2 points above and 2 below center frequencies of 1, 1.4, 2, 2.8, 4, and 5.6 kHz) giving 5-frequency clusters.[1,2] The unwrapped phase slope of each 5-frequency cluster is an estimate of group delay at the center frequency of the

[1] The SFOAE system only outputs frequencies corresponding to the frequency resolution (defined by the FFT size and sampling rate). The maximum resolution of the system is 11.7 Hz. It is difficult to optimally sample in small frequency ranges, therefore sampling is done around the center frequencies in multiples of the frequency resolution, e.g., every 11.7 Hz, every 23 Hz, etc.

[2] Shera suggested using a power-law sampling scheme to optimally sample the phase spectrum (for example: $f^{0.65}$), but this is not currently an option in the system.

cluster. Because group delay is calculated from a slope, it cannot be estimated from just one datum. Instead, at least three and preferably more points are needed to get an accurate estimate.

Reliability is then assessed to establish if repeated measurements produce the same SFOAE amplitudes and phases. There will always be some variability in repeated measurements, but the key is whether this variability is smaller than the size of the changes in SFOAEs due to the independent variables of interest. When evaluating OAE types for suitability in hearing-conservation programs, estimates of test-retest reliability indicate at an early stage of development whether the test will be sufficiently reliable to be able to detect small changes in OAEs. Even with good validity, if reliability is poor, then the OAE test will be of little use in hearing conservation.

In hearing conservation, one scenario of interest is determining whether a change in an SFOAE between two tests in an *individual* ear is significantly different from chance. This scenario occurs when monitoring SFOAEs for signs of changes due to sound overexposure (where the first test is a pre-sound-exposure baseline test and the second and subsequent tests are post-sound-exposure). Another scenario involves measuring the strength of the auditory efferent system (thought to be a measure of resistance to noise-induced hearing loss). One method is to measure the SFOAE with a contralateral efferent elicitor and compare it to the SFOAE without a contralateral elicitor. Again, this measurement is in individual ears.

Ideally, the variability seen in each individual ear at baseline would be used to determine significance, but in most settings there is only time to take one baseline measurement and not repeated measures. Instead, the standard error of measurement[3] ($SE_{MEAS}$), of a non-noise-exposed, normal-hearing, control *group* can be used as a criterion to say whether or not an *individual* ear has had a significant shift in SFOAE amplitude or group delay (Lapsley Miller & Marshall, 2001; Lapsley Miller, Marshall, & Heller, 2003; Marshall, Heller, Westhusin, & Lapsley Miller, 2000). That is, the size of the shift must be greater than the variability seen in the control group to be flagged as significant (where a shift is defined as the arithmetic difference between the two tests).

Here, SFOAEs are evaluated using (a) participants with at least one ear with moderate to profound hearing loss, which should not show any response to SFOAE stimulation, and (b) participants with normal-hearing ears, to ensure that valid, reliable SFOAEs were produced in response to the stimulation and that the waveform characteristics of the responses were consistent with SFOAE theory and with data collected elsewhere.

[3] The $SE_{MEAS}$ is a repeated-measures standard deviation that can be used to specify the magnitude of a statistically significant change within an individual (Ghiselli, 1964). It is defined as $SE_{MEAS} = \sqrt{\frac{1}{2}\left(s_1^2 + s_2^2\right)\left(1 - r\right)}$ where $s_1^2$ and $s_2^2$ are the variances of test 1 and test 2 respectively and $r$ is the correlation between test 1 and test 2.

# METHOD

## PARTICIPANTS

Eight people with hearing impairments in at least one ear (5 male and 3 female participants, aged 14 to 80 years) and 33 people with normal hearing in at least one ear were enrolled (61 ears, 16 female and 17 male participants, aged 14 to 57 years).

## SCREENING

For hearing-impaired participants, the etiology of all hearing losses was consistent with sensorineural hearing loss (excluding Meniere's disease). The hearing losses were at least 50 dB HL at the frequencies 0.5, 1, 2, 3, 4, 6, and 8 kHz. It is highly unlikely that there would be any OAEs present in ears with hearing loss to this degree across *all* of these frequencies.[4] If both the participant's ears met the criteria, then the most impaired ear was chosen for testing. See Appendix A for individual participant summaries.

For normal-hearing participants, hearing levels were less than or equal to 15 dB HL from 0.5 to 8 kHz.[5] If both ears met the screening criteria, then both ears were tested.[6]

An otoscopic examination was done to ensure ear canals were clear (no cerumen), and that there were no external-ear pathologies. Cerumen removal was done where necessary before any OAE testing. Middle-ear pressures were required to be between ±50 daPa for OAE testing.

All hearing-impaired ears were tested for the absence of distortion-product OAEs (DPOAEs). Ears with this type and degree of hearing loss do not have DPOAEs (absence defined as no DPOAE level 10 dB above noise). Parameters used were $L_1/L_2$=65/55 dB SPL and $f_2/f_1$ =1.21 (Stover, Gorga, Neely, & Montoya, 1996). $f_2$ ranged from 1 to 4 kHz with 10 points/octave for a total of 21 test points. The Mimosa Acoustics DP2000 system was used for testing.

## SFOAE METHOD

Participants were evaluated with both transient-evoked OAEs (TEOAEs) and SFOAEs, but only the results for SFOAEs are reported here. Hearing-impaired participants received the OAE test battery once in the test ear. Normal-hearing participants received the test battery in both ears

[4] Dorn et al (2001), showed some hearing-impaired participants with OAEs. These participants had hearing levels of at least 50 dB HL at the test frequency, but not necessarily across a wider frequency range as we require.

[5] One participant had an HL of 20 dB HL at 2 kHz, but was normal at all other frequencies. She was allowed to continue due to high availability and the fact she had normal OAEs.

[6] To make validity analyses on normal-hearing ears commensurate with the hearing-impaired data, one ear was arbitrarily chosen and only the first test used. For reliability analyses, data from both ears were used.

(if both ears passed screening), and then the test battery was repeated a few minutes later in the same test session after refitting the probe. They also received a spontaneous OAE (SOAE) test.[7]

The probe/suppressor levels were 40/55 dB SPL. The suppressor frequency was 47 Hz higher than the probe frequency. These are close to the values that Shera successfully uses (Shera & Guinan, 1999, 2003). Thirty probe frequencies were tested in six 5-frequency clusters. The six clusters were half-octave spaced at: 1, 1.4, 2, 2.8, 4, and 5.6 kHz. The five frequencies in each cluster were spaced at 11.7 Hz intervals centered on the half-octave frequencies (the sampling rate was 48 kHz and FFT size was 4096 points).

The two stimulus channels were each calibrated with a chirp (50 averages were taken). Each stimulus buffer consisted of two probe-alone and two probe+suppressor intervals (2 sub-averages). The stimulus buffer was presented 16 times (16 averages), and the results of the 2x16=32 averages were analyzed as described by Shera (Shera & Guinan, 1999). Noise-rejection level started at 36 dB SPL, but was often adjusted up by the examiner.

## RESULTS

### HEARING IMPAIRED EARS: SFOAE VALIDITY

First, the data were examined for measurement and testing problems. Second, the averaged data were evaluated to find responses consistently above the noise level. Then, individual cases were considered where there appeared to be an "SFOAE" present, as a function of SNR criterion. Finally, the group delay associated with suspicious data points was examined.

The calibrations for each measurement were examined for problems. The calibration for Participant 1038 was very much lower than the rest and was most likely due to the wrong probe-interface cable (PIC) being used (see Figure 1). **Data for Participant 1038 were not considered further.** Calibrations for Participants 1035 and 1037 were also not ideal, with spectra tending to be low below 3 kHz and high between 3 to 7 kHz (see Figure 1 and summary plots in the Appendix). The two channels for Participant 1035 also tended to show some separation. The data collected using these calibrations were not eliminated, but interpretation of the results for these participants was more cautious.

The data were examined at each measurement frequency, for each participant, for the following problems. Data points meeting these criteria were rejected and were not included in the analyses. Data points where the measured stimulus levels were more than ±3 dB from target levels were rejected (probe stimulus level 40 dB SPL; suppressor stimulus level 55 dB SPL). The probe levels were checked for both the probe-alone interval and probe+suppressor interval; the suppressor level was checked for just the probe+suppressor interval. The probe and suppressor levels were both off-target for Participant 1029 at 4031 Hz. This may have occurred due to probe slippage. The difference between the probe and suppressor stimulus levels was checked to ensure it was within 15 dB ±3 dB (inclusive). All data passed this test.

[7] SOAE results were not considered in these analyses.

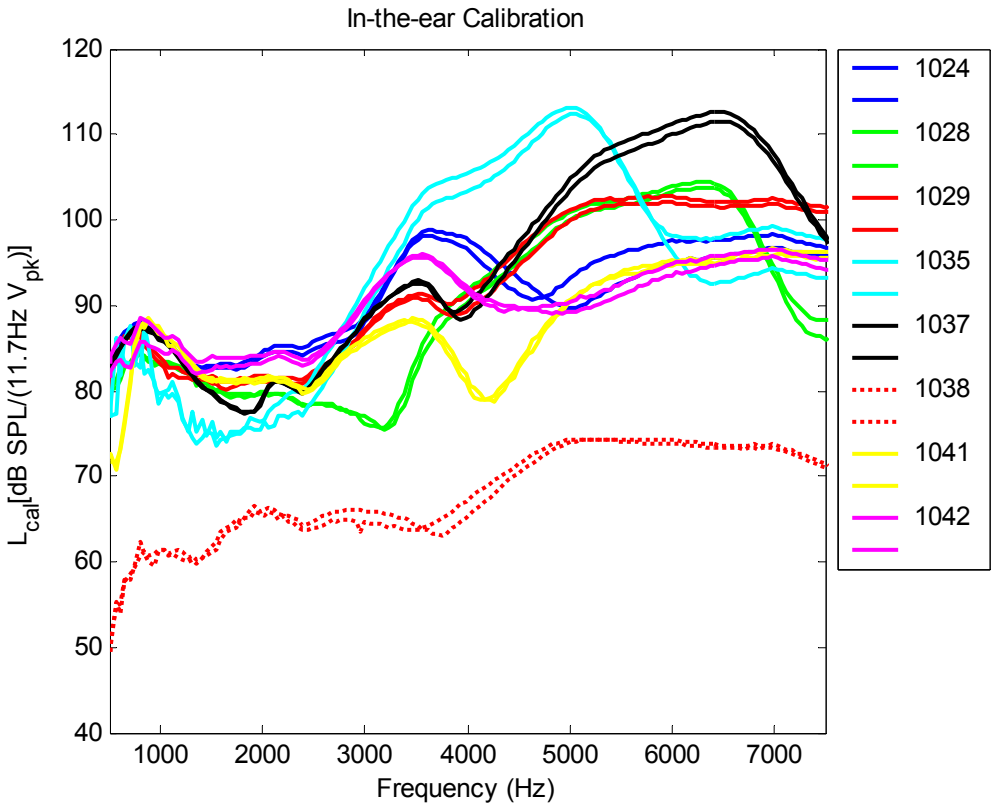



*Figure 1 In-the-ear calibrations for the 8 hearing-impaired participants. Same color pairs indicate each channel of the ER10C probe for each participant. Participants 1035 and 1037 did not have good calibrations – the channels are separated and the spectrum is very peaked. Participant 1038 was probably run with the wrong probe-interface cable.*

The total noise level associated with each SFOAE measurement was checked to ensure it was no greater than -10 dB SPL. This criterion is rather lax to account for the tendency of hearing-impaired participants to have noisier measurements. All but one participant had from 1 to 13 test frequencies affected by high noise levels. Most high noise levels affected the 1 kHz 5-frequency cluster and to a lesser extent, the 5.6 kHz cluster. Participant 1035 had 13 noisy measurements. Noise rejection was set very high for these measurements (57 dB SPL), indicating the tester needed to adjust the rejection level up by about 20 dB to obtain any measurements. This participant was severely hearing-impaired in both ears, making the monitoring of his own body and movement noises difficult. Noise at the lower frequencies was probably partly due to the increased probe amplification needed at these frequencies (see calibration plot in Figure 1) and partly due to the participant's internal body or movement noise. In all, 27 data points were affected (13%, after dropping the one bad datum for Participant 1029).

The resulting data set constituted 182 good-quality measurements (i.e., calibration was acceptable, stimulus levels were within target level, the difference between stimulus levels was within target, and the noise was relatively low level).

AVERAGE SFOAE & NOISE LEVELS

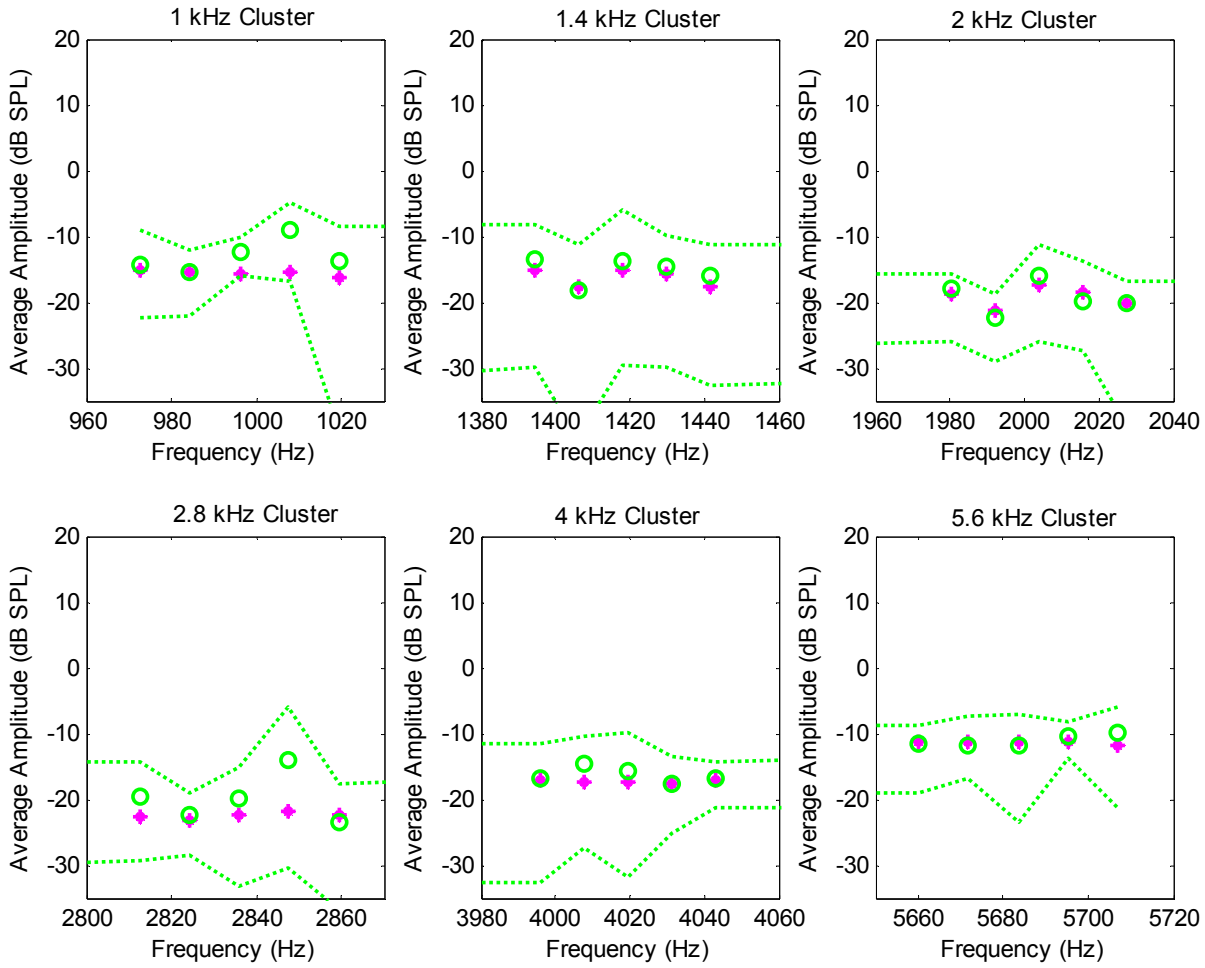



*Figure 2 Average amplitude of SFOAE measurement and associated noise level in hearing-impaired ears (N=7) for each frequency cluster. Only measurements with stimulus levels on target and low noise levels were included. Green circles are "SFOAE" responses, magenta asterixes are total noise level. Dotted lines indicate the averaged maximum and minimum responses.*

One approach to establishing SFOAE validity is to look at the average SFOAE response. In a hearing-impaired ear, it would be expected that the average SFOAE response would be indistinguishable from the noise. Here, there were only up to 7 ears contributing to the average, so the results should only be considered indicative. There were not enough data points to look at the distribution of noise and SFOAE amplitude.

Figure 2 shows the averaged responses for all measurements meeting the stimulus-level and low-noise criteria. The average SFOAE response overall was slightly above the noise (by 1.2 dB SNR). Also shown are the maximum and minimum responses, indicating that statistically there will sometimes be responses much higher than the noise, even if on average there is little to no difference. The overall average SNR across all frequencies was 1.2 dB (when high noise levels were excluded) and 0.7 dB (when high noise levels were included). The average (over

participant) maximum SNR (over any frequency) was 15.8 dB (at 2.8 kHz). This gives a ballpark estimate of the range over which the SNR may vary.

Excluding all measurements with SNR≤16 dB would ensure clean data. This is consistent with Mimosa Acoustics' coupler measurements and with the recommended 15 dB SNR criterion used by Shera (Shera & Guinan, 2003). However, potentially good data could also be eliminated. What is the effect of lowering the SNR criterion on the number of random/artifactual SFOAEs? Can the cluster's group delay be used to validate SFOAEs with lower SNRs?

## THE EFFECT OF THE SNR CRITERION ON VALIDITY

The validity of responses with mid-to-high SNRs from hearing-impaired ears was evaluated firstly using an SNR criterion of 12 dB. This criterion was 3 dB lower than the criterion used by Shera for determining the presence of SFOAEs in normal-hearing ears (Shera & Guinan, 2003). This laxer criterion was to make sure any measurements even close to being considered a normal SFOAE were caught. By considering these data points along with the associated group delay, any problem with spurious or systematic artifactual responses was identified.

Only Participant 1035 had a measurement where SNR exceeded 12 dB: SNR=16.6 dB at 2.8 kHz (which also exceeded the stricter 15 dB SNR criterion). As cautioned earlier, the data for this participant was potentially affected by a poor calibration and high noise levels; however, at this frequency, the noise level was -22 dB SPL. With only one point exceeding the SNR criterion, there was no basis to calculate group delay to further validate the measurement. This point constituted only 0.5% of the available data.

In case the SNR criterion was too strict, it was lowered to 9 dB. Participant 1035 showed 2 measurements with SNR exceeding 9 dB, but they were in different 5-frequency clusters. Lowering the SNR criterion further to 6 dB, thirteen frequencies (7%) now exceeded the SNR criterion, spread across all but one participant. However, there were still no cases where there were enough responses with high SNRs within the same cluster to calculate group delay. Lowering the criterion further to 3 dB SNR produced three clusters where group delay could be calculated.

Rather than evaluate these group delays (phase slope) directly, linearity was assessed instead. It would be expected that the phase spectrum for a true SFOAE would be generally monotonically decreasing over the entire spectrum and close to linear over smaller frequency ranges. On these assumptions, the phase slope should be well-fitted by a straight line, over a small frequency range, because of how phase systematically changes with frequency (see Shera & Guinan, 1999 and page 14 of this document for analysis of group delay in normal-hearing ears). On the other hand, a chance ordering of phase would be less likely to fall exactly on a line. It is more difficult to assess the validity of group delay directly, because group delay varies with frequency and with individual ears. In other words, to establish validity, the quality of the group-delay estimate should be considered, and then whether the group delay make sense given the frequency and the ear.

The typical way to assess linearity is to calculate the correlation between the two variables: here, between unwrapped phase and frequency. If the underlying process is truly linear (and not noisy), then the correlation would be expected to be very high. Another way to assess the quality of the group-delay estimate is to assess the standard deviation of the slope estimate (after a least-squares regression). Here, we do both. Looking ahead to normal hearing ears (see the next section), the vast majority of the group delay estimates had correlations <-.94 and group-delay standard deviation <2 ms. For the 3 group-delay estimates (from the hearing-impaired ears using an SNR criterion of 3 dB, from above), none meet both the typical correlation and group-delay standard deviation seen in normal-hearing ears. In fact, one had a positive, not negative, correlation.

To fully show the relationship between SNR criterion and the percentage of "SFOAE" responses and group delays in the hearing-impaired ears, Figure 3 shows the effect of SNR criterion as it is varied from -10 dB to 20 dB SNR.

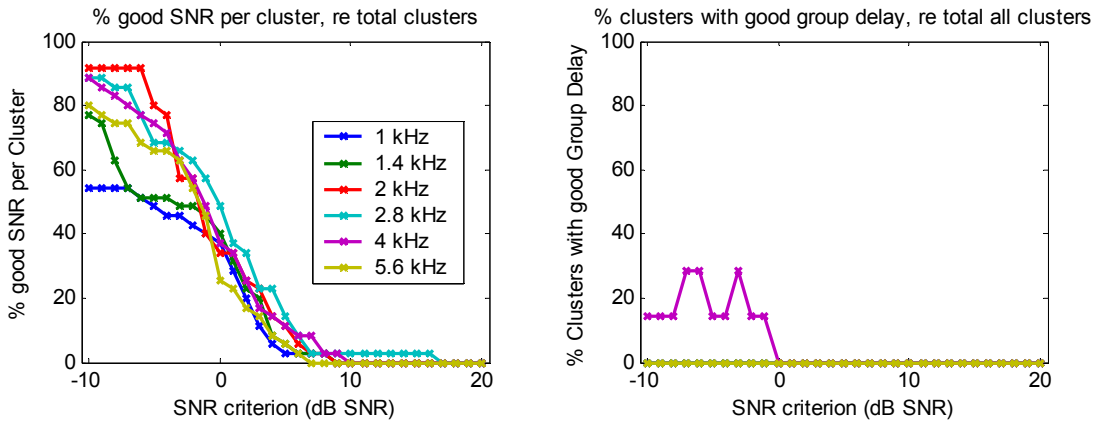



*Figure 3 Effect of SNR criterion on the percentage "SFOAE" responses measurable in each 5-frequency cluster for the 7 hearing-impaired ears (left), and the percentage clusters with measurable group delay (right). Compare with Figure 5 for normal-hearing ears.*

The left panel shows how the percentage of test frequencies in each 5-frequency cluster exceeding the SNR criterion increased as SNR criterion decreased. The SNR criterion had to decrease below 10 dB before there were many measurements meeting the criterion. This indicated that an SNR criterion of 9 to 10 dB (rather than Shera's 15 dB) would be adequate to ensure most random/artifactual "SFOAE" responses were removed.

Furthermore, the right panel shows the percentage of clusters with measurable group delay as a function of SNR criterion. Measurable group delay here means those estimates with high negative correlation and small error (see criteria developed from normal-hearing ears in the next section). This indicated that an SNR criterion of 0 dB would not introduce erroneous group-delay estimates, even if it meant more random/artifactual measurements were included. For the handful of measurable group delays at 4 kHz, most were not consistent with measures taken in normal-hearing ears, because the actual delays were too long to be consistent with theory.

Based on hearing-impaired data alone, the SNR criterion for a valid SFOAE response would be 9 to 10 dB, if SNR was the only criterion. If a group-delay criterion was also used, the SNR criterion for a valid SFOAE response could be as low as 0 dB. These criteria will be investigated further using SFOAE measurements from the normal-hearing group.

## NORMAL-HEARING EARS: SFOAE VALIDITY

First, the data were checked for measurement problems. Second, the number of measurable SFOAEs per 5-frequency cluster was counted as SNR was systematically changed. Then, group delay was calculated, again as SNR was systematically varied.

The initial data set consisted of 900 measurements (30 participants, 6 clusters, 5 frequencies per cluster), or 150 measurements per cluster. For comparison with the hearing-impaired ears, one ear was randomly chosen for analysis, giving a total of 30 ears.

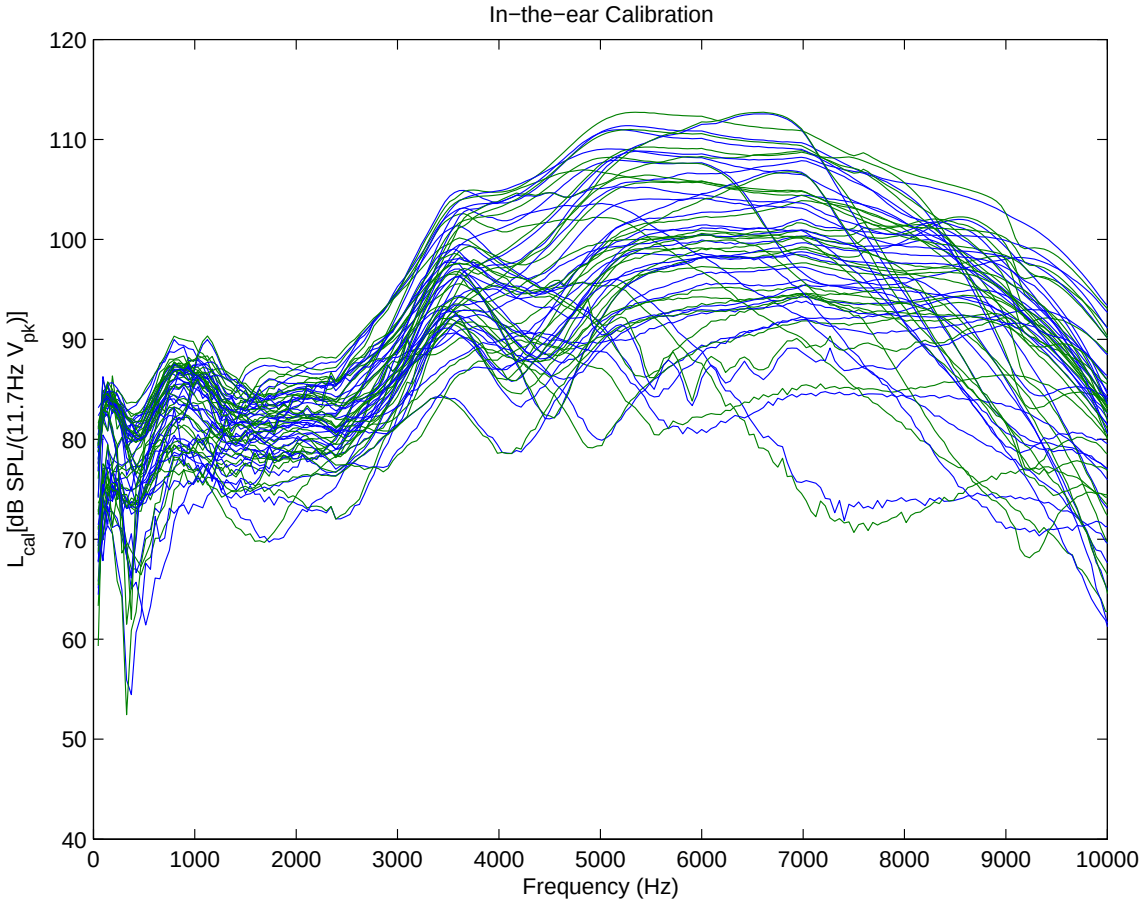



*Figure 4 In-the-ear calibrations for the 30 normal-hearing participants. Blue lines indicate Channel 1 and green lines indicate Channel 2 in the ER10C probe.*

The data were assessed for the quality of the measurement using the same criteria as in Part One (checking correct PIC was used, checking stimulus levels, checking for high noise). Calibrations indicated no problem with incorrect PIC being used (see Figure 4), but again many calibrations were not ideal with there being differences between the channels.

Probe and/or suppressor stimulus levels were off-target for 4 participants (Participant 1009: 3 frequencies, Participant 1019: 10 frequencies, Participant 1044: 15 frequencies, and Participant 1049: 2 frequencies), affecting a total of 30 test frequencies (3.3% out of all possible test frequencies). In most cases, both the probe and suppressor levels were affected. The difference between probe and suppressor stimulus levels did not exceed 15 dB ± 3 dB at any frequency.

Eighteen participants had some noise levels exceeding -10 dB SPL (the same noise-level criterion was used here to be consistent with the analyses in Part One). Overall, 43 test frequencies were affected, mostly those in the 1 and 5.6 kHz clusters (6 frequencies were also affected by off-target stimulus levels). By cluster, there were 129 (86%), 144 (96%), 144 (96%), 150 (100%), 145 (97%), and 117 (78%) measurements with stimulus levels on target and low noise for the 1, 1.4, 2, 2.8, 4, and 5.6 kHz clusters, respectively.

All but one participant had some frequencies with SFOAEs lower than 15 dB SNR (see below for more detailed analysis). The exception was Participant 1026, who also had many large spontaneous OAEs. Participant 1033 had no SFOAEs above 15 dB SNR. However, both his hearing and his DPOAEs were normal, so there was no cause to remove this data set.[8]

## AVERAGE SFOAE & NOISE LEVELS

Figure 5 shows the average SFOAE response and noise levels. The SFOAEs were consistently above the noise for all frequencies, in contrast to Figure 2 for hearing-impaired ears. The average SNR across ears and frequencies was 24 dB with a maximum of 42 dB. This was in stark contrast to the hearing-impaired ears.

## PERCENT MEASURABLE

Figure 6 (top left) shows the percentage of measurements with good SNR, low noise, and stimulus levels on target within each cluster (averaged over participants) as a function of SNR criterion. The percentage was calculated relative to the total number of good quality measurements. Accordingly, as the SNR criterion decreases, the percentage does not necessarily tend to 100% (this is most noticeable at 1 and 5.6 kHz). As the SNR criterion increased, there were fewer frequencies per cluster that met the SNR criterion. The best frequencies (most measurable) were those in the 1.4 and 2.8 kHz clusters. The worst by far was the 5.6 kHz cluster. The differences among clusters were not systematic.[9]

---

[8] For those curious, his other ear also showed low to no SFOAEs. TEOAEs were present only at low levels around 1 kHz. Furthermore, this participant was related to a bilaterally hearing-impaired participant with a cochlear implant.

[9] The low/no SFOAEs for Participant 1033 accounts only for a maximum of 3.3% of the data loss at each frequency.

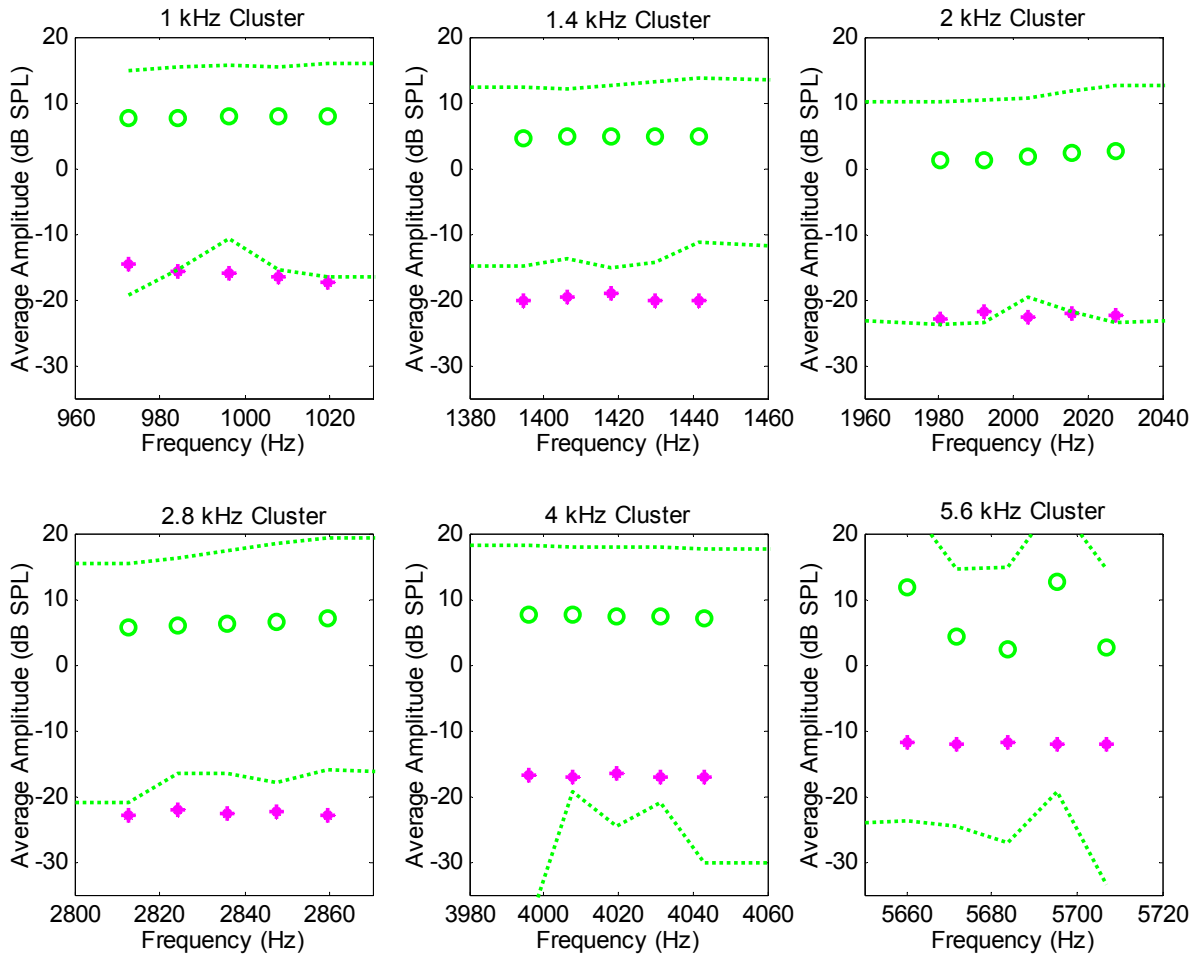



*Figure 5 Average amplitude of SFOAE measurement and associated noise levels for the 30 normal-hearing ears. Only measurements with stimulus levels on target and low noise were included. Green circles are SFOAE responses, magenta asterixes are average noise levels, and the dashed lines show the maximum and minimum SFOAE responses.*

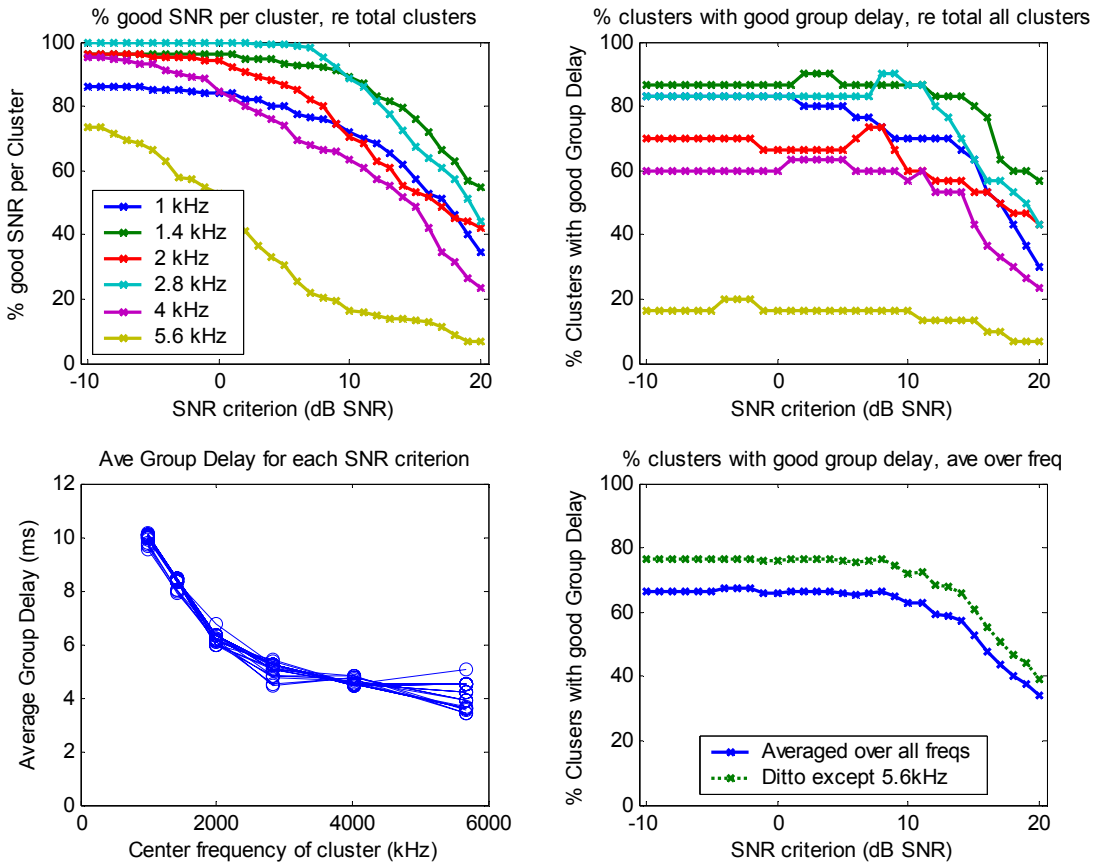



*Figure 6 Effect of SNR criterion on percentage "SFOAE" responses measurable in each cluster for the 30 normal-hearing ears (top left), the percentage clusters with measurable group delay, (where at least 3 frequencies with good SNR were required in the cluster) (top right) and then averaged over all frequency clusters (blue) or averaged over all frequency clusters except 5.6 kHz (green) (bottom right), and average group delay for each SNR criterion as a function of frequency (bottom left). Compare with Figure 3 for the hearing-impaired ears.*

## GROUP DELAY

Potentially good data is thrown away when a high SNR criterion is used. What is the effect of lowering the SNR criterion on the validity of these SFOAE measurements? This can be assessed by seeing how the quality of the group delay estimate changes as the SNR criterion is varied. First, an example using the recommended 15 dB SNR criterion is examined, and then the SNR criterion is systematically varied from -10 to 20 dB.

How is quality determined? First, consider the group delay data produced with an SNR criterion of 15 dB (where as before, only 3 out of 5 frequencies in a cluster need to meet good measurement and SNR criteria to calculate group delay). There were 98 out of 180 clusters (6 clusters x 30 ears; 54%) with at least 3 frequencies that met these criteria. Two clusters had negative group delay (positive slope). What does it mean in a normal-hearing ear to have a

cluster with *negative* group delay and good SNR? Does it have validity? Theoretically, negative group delays would only be expected in regions such as spectral notches where the SFOAE *amplitude* is changing rapidly as a function of frequency. This was the case with one of the clusters in question.[10] However, regions with notches are not good places to be making reliable SFOAE measurements, even if they have validity. From an empirical standpoint, the numbers themselves were not compelling with correlations of .71 and .79, whereas the majority of other group-delay estimates had correlations of <-.94. It seems reasonable to eliminate any negative group delays with correlations this low.

For the remaining 96 measurements, 67% of the correlations were equal to -1.0 and 99% were less than -.94 (the distribution of correlation took the form of a negative exponential). It is expected from a theoretical perspective that over small frequency ranges the unwrapped phase slope should be linear as a function of frequency. It would seem to be the case here. In hearing-impaired ears, the largest correlation was -.9. So a criterion somewhere between -.9 and -.94 should maximally discriminate between the two groups. To ensure the least overlap with the hearing-impaired ears, a criterion for correlation of less than -.94 was used (which excluded only 1 out of the 96 group delays from the normal-hearing ears).

Another measure of quality is group-delay error, which is the standard deviation of the slope estimate from the linear regression. In the current example, there were no errors greater than or equal to 2 ms. In the larger data set there were a handful of cases greater than or equal to 2 ms, with the overall distribution taking the form of a negative exponential. A criterion of 2 ms seems reasonable, given the underlying distribution.

It appears that an SNR of 15 dB ensures good quality group-delay estimates, and it follows that the vast majority of SFOAEs contributing to these group delays are valid.

Is a criterion of SNR>15 dB too strict? Could it be lowered and quality measurements (validated by group delay criteria) still obtained? The top right plot in Figure 6 shows the percentage of clusters (averaged across participants) that had good group delays, relative to the total number of available clusters. As the SNR criterion was systematically reduced from 20 dB to -10 dB (reading the functions from right to left), the percentage of good group delays increased until it reached an asymptote around 10 dB (the bottom right plot shows the same data averaged across frequency). Although the available data increased as the SNR criterion was decreased (top left plot), the number of clusters with good group delays did not increase once the SNR criterion decreased to around 10 dB. This was presumably because although more clusters became available, the group-delay error also increased.[11,12]

---

[10] Looking more closely at the data, the SFOAE amplitude of one cluster varied considerably, and only 3 of the 5 points met the criteria for inclusion in the calculation. It is not clear why the other cluster resulted in a negative delay; all five frequencies had good SNR, were good measurements, and there were no SOAEs nearby.

[11] The group-delay error and SNR criteria traded off against each other. When the group-delay error criterion was reduced, the asymptote pushed out to higher SNRs (e.g., for group-delay error < 1 ms, the asymptote was around 14 dB SNR), showing that only the higher SNRs were contributing to the average group delay. Therefore, the ultimate answer depends on how much error is tolerable in the group delay estimate versus how much data will be eliminated from the analysis.

What about the actual group delay measurements? Did the actual group delay change as the SFOAEs with lower SNR were included? The lower left plot in Figure 6 shows that the average group delay estimates were relatively stable as a function of SNR criteria, except for around 5.6 kHz (where there were many rejected data points).

It appears that the correlation and group-delay error criteria were strong enough on their own to validate the SFOAE *cluster*. The inclusion of more low-SNR SFOAEs did not increase the correlation, but also did not decrease the correlation either.

## EFFECT OF SNR CRITERION ON GOOD FREQUENCIES PER CLUSTER

As the SNR criterion was increased there were more clusters with only 3 or 4 (out of 5) test frequencies meeting the SNR criterion. There were also altogether more clusters meeting all the small group-delay error and strong negative correlation criteria as SNR decreased. Therefore, the quality of the good group delays themselves did not decrease and there were more good clusters available. Figure 7, 8, and 9 illustrate this effect as SNR was varied.

Figure 7 shows the percentage of clusters with 3, 4, or 5 good measurements per 5-frequency cluster (the number needed to be able to estimate group delay). Here, good measurements refer to on-target stimulus levels, low noise, and exceeding the SNR criterion. Various SNR criteria were examined, including -99, 0, 5, 10, 12.5, 15, and 20 dB. The -99 dB criterion was equivalent to not using an SNR criterion at all for this data set, and was included for comparison. As expected, as the SNR criterion increased, there were more clusters with 1 or 2 data points not meeting the criterion.

Group delay was then calculated for all the clusters above with at least 3 good measurements per cluster, for all the SNR criteria above. Figure 8 shows the percentage of group delays that did not meet the correlation or group-delay error criteria, as a function of the number of good measurements per 5-frequency cluster and SNR criterion. There were more “bad” group delays for lower SNR criteria and for clusters with only 3 or 4 good measurements.

Finally, Figure 9 provides a *count* of the group delays that met the good group delay criteria, broken down by the number of good measurements per 5-frequency cluster and the SNR criterion. The total number of good group delays increased as the SNR criterion decreased, despite there being more “bad” group delays (see Figure 8). This is because there were simply more data available for the lower SNR criteria.

---

[12] Group-delay error is partially affected by the number of points going into the calculation. There are more clusters-of-three with large error than clusters-of-five e.g., for the 15 dB SNR criterion, 6/94 had 3 points and error < 0.5 ms (6.4%), and 7/23 had 3 points and error > 0.5 ms (30%).

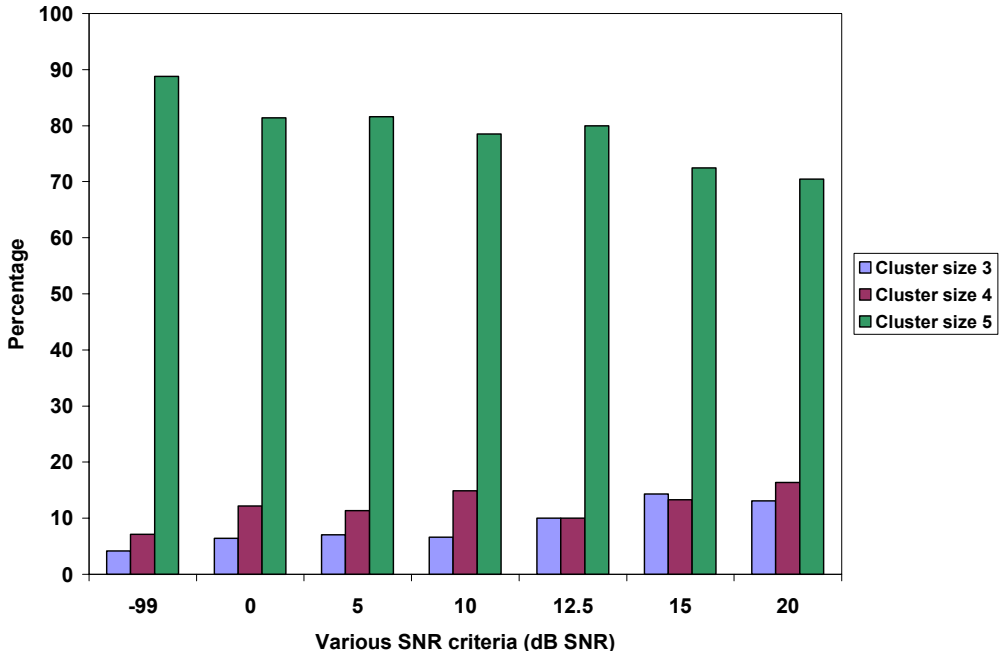



*Figure 7 Percentage of clusters with 3 to 5 good measurements per 5-frequency cluster. As the SNR criterion increased, there were more clusters with 1 or 2 data points not meeting the criterion.*

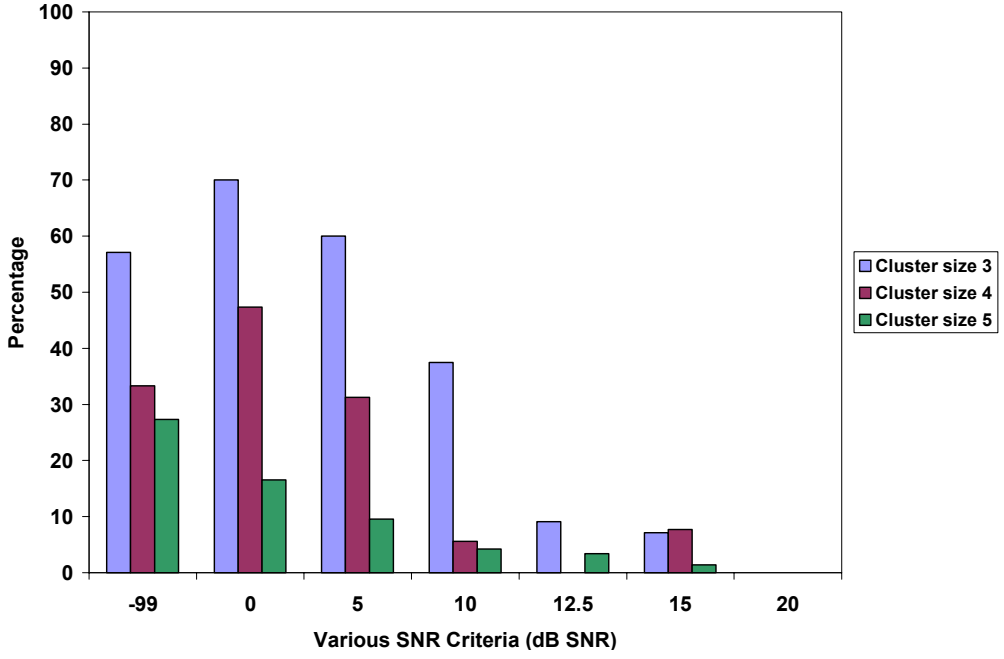



*Figure 8 Percentage of group delays with low correlation or large group-delay error, as a function of the number of good measurements per 5-frequency cluster and SNR criterion. There were more "bad" group delays for lower SNR criteria and for clusters with only 3 or 4 good measurements.*

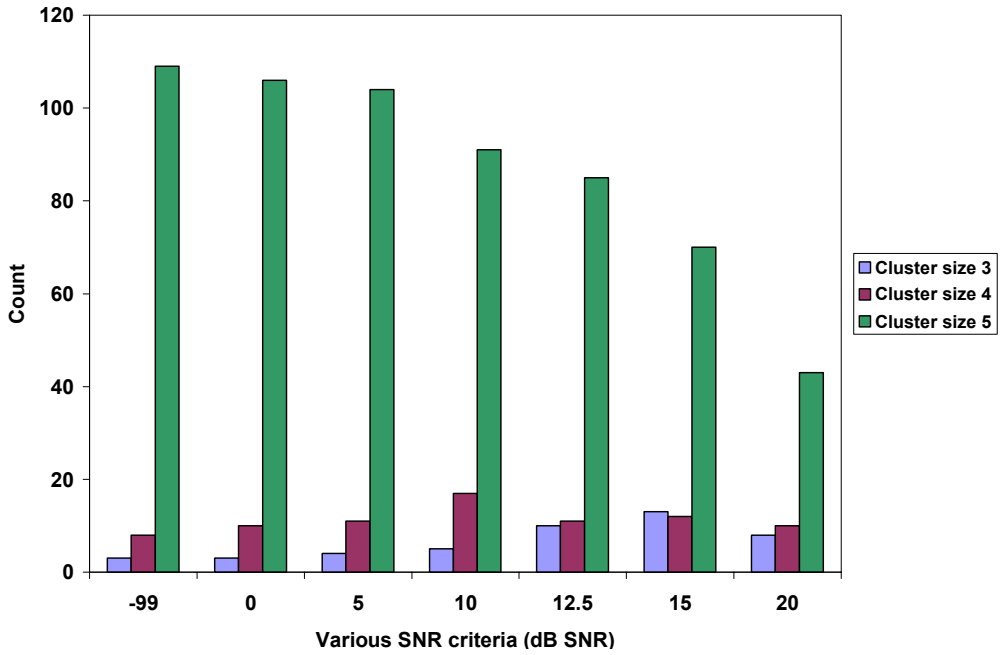



*Figure 9 Count of good group delay by the number of good measurements per 5-frequency cluster and SNR criterion. The total number of good group delays increased as the SNR criterion decreased, despite there being more bad group delays (see Figure 8). This was because there were simply more data points available for the lower SNR criteria.*

## VALIDATING SINGLE-FREQUENCY SFOAE MEASUREMENTS

How can a single frequency SFOAE be validated if there is no group delay information? When analyzing a single frequency, an SNR criterion of at least 15 to 18 dB should ensure no spurious responses are recorded as SFOAEs. A less strict criterion could safely be set as low as 9 dB, because this gave only 0.5% spurious responses in hearing-impaired ears.

## NORMAL-HEARING EARS: SFOAE TEST-RETEST RELIABILITY

The data from the normal-hearing group (61 ears, 33 participants) were used to calculate the SFOAE standard errors of measurement. This group was given two SFOAE tests, with the probe replaced between measurements (to simulate what happens when repeated measures are taken across sessions). Most participants had both ears tested.

The data were examined for problems in a similar way as for the validity analyses, which ensured that each datum going into the analyses had good stimulus levels and low noise. A range of SNR criteria were examined. The data from the first repeat were paired with the corresponding data from the second repeat (matched by subject, ear, and test frequency). To be included in the analysis, the two data points in each pair were required to be good-quality measurements with good SNR. If either point was bad, then both were dropped. Each ear was not required to have a complete set of good measurements at all frequencies. Table 1 shows how many measurement pairs (N) met these criteria (out of a maximum possible of 305 pairs).

## POOLED SFOAE AMPLITUDE RELIABILITY

The $SE_{MEAS}$ for the frequency region of each cluster were calculated by pooling the single-frequency data pairs in each cluster, rather than calculating separate $SE_{MEAS}$ for each of the five test frequencies in each cluster (see Table 1). This increased the number of frequency pairs (N) going into each calculation. The resulting $SE_{MEAS}$ is not for the average SFOAE amplitude in the cluster, it is a single-frequency SFOAE $SE_{MEAS}$ covering the frequencies in the ~50 Hz range around the cluster's center frequency.

The $SE_{MEAS}$ were calculated for SNR criteria of 10, 12.5, and 15 dB. Reliability tended to increase with an increasing SNR criterion; however, the amount of data meeting the criterion decreased. Reliability was also lower at 2 and 2.8 kHz compared with the other frequency clusters, though it was not apparent why.

## AVERAGED SFOAE AMPLITUDE RELIABILITY (WITHIN A CLUSTER)

One way to increase reliability is to average SFOAE amplitude across frequencies. For the current data, it would make most sense to average within each cluster. Because the frequencies in each cluster were so close together, frequency specificity was not lost in the process. The same data set was used from the pooled SFOAE calculation (above), but there was an additional requirement that there were at least three frequencies with good SNR per cluster before averaging occurred (see Table 2). No data in the 5.6 kHz cluster met this requirement.

In all but one case (2 kHz, 15 dB SNR criterion), the $SE_{MEAS}$ for the averaged SFOAE amplitude was about 0.2 dB smaller than the corresponding $SE_{MEAS}$ for the single-frequency SFOAE amplitudes in the same cluster (from Table 1).

## SFOAE GROUP-DELAY RELIABILITY

Reliability can also be evaluated for group delay. Group delay was calculated as before and the same criteria applied (error <2 ms and correlation <-.94). These additional requirements slightly decreased the available data from that used to calculate the average SFOAE $SE_{MEAS}$ in Table 2. Table 3 shows the mean group delay for each test and the resulting $SE_{MEAS}$. Overall, group-delay reliability was on the order of 0.5 ms and did not appear to be particularly dependent on SNR criterion.

Outliers were investigated to see if there was a consistent reason for poor reliability. For group delay, there were no obvious patterns, but some poor reliability may have occurred when the test frequency fell in a region where the calibration spectrum was in a dip (these dips were not deep enough to cause the system to estimate the appropriate stimulus level, which it attempts to do in areas such as in a standing wave notch). A cursory examination was made to see if there was any association between presence/absence of spontaneous OAEs and poor reliability, but none was found.

*Table 1 $SE_{MEAS}$ for single frequencies in each cluster for three SNR criteria. Shown is the center frequency of the cluster, the number of frequency pairs going into the calculation (N), the mean SFOAE amplitude from the first test (R1) and second test (R2), respectively, and the resulting $SE_{MEAS}$.*

| Center Freq (kHz) | N | Mean R1 (dB SPL) | Mean R2 (dB SPL) | $SE_{MEAS}$ (dB SPL) |
|---|---|---|---|---|
| SNR>10 dB | | | | |
| 1 | 197 | 5.72 | 5.83 | 1.13 |
| 1.4 | 230 | 3.89 | 3.97 | 1.16 |
| 2 | 194 | 1.04 | 1.43 | 1.89 |
| 2.8 | 224 | 1.81 | 2.25 | 2.29 |
| 4 | 167 | 7.25 | 7.23 | 1.60 |
| 5.6 | 47 | 8.28 | 7.85 | 1.36 |
| SNR>12.5 dB | | | | |
| 1 | 182 | 6.16 | 6.24 | 1.01 |
| 1.4 | 203 | 4.68 | 4.73 | 1.10 |
| 2 | 174 | 1.75 | 2.16 | 1.83 |
| 2.8 | 193 | 2.77 | 3.26 | 2.32 |
| 4 | 145 | 8.15 | 8.12 | 1.54 |
| 5.6 | 40 | 9.18 | 8.65 | 1.36 |
| SNR>15 dB | | | | |
| 1 | 154 | 6.96 | 7.00 | 0.96 |
| 1.4 | 179 | 5.40 | 5.48 | 1.02 |
| 2 | 149 | 2.72 | 2.95 | 1.59 |
| 2.8 | 147 | 4.46 | 4.98 | 2.02 |
| 4 | 115 | 9.54 | 9.53 | 1.19 |
| 5.6 | 28 | 10.82 | 10.41 | 1.13 |

*Table 2. $SE_{MEAS}$ for the averaged SFOAE amplitude in each cluster (at least 3 frequencies with good SNR per cluster). Shown is the center frequency of the cluster, the number of frequency pairs going into the calculation (N), the mean SFOAE amplitude from the first test (R1) and second test (R2), respectively, and the resulting $SE_{MEAS}$.*

| Center Freq (kHz) | N | Mean R1 (dB SPL) | Mean R2 (dB SPL) | $SE_{MEAS}$ (dB SPL) |
|---|---|---|---|---|
| SNR>10 dB | | | | |
| 1 | 41 | 4.15 | 4.40 | 0.81 |
| 1.4 | 51 | 1.43 | 1.71 | 0.98 |
| 2 | 39 | -1.26 | -1.02 | 1.66 |
| 2.8 | 47 | -2.01 | -1.75 | 2.08 |
| 4 | 32 | 4.49 | 4.32 | 1.25 |
| SNR>12.5 dB | | | | |
| 1 | 36 | 4.93 | 5.08 | 0.73 |
| 1.4 | 40 | 3.42 | 3.58 | 1.00 |
| 2 | 34 | -0.14 | 0.14 | 1.75 |
| 2.8 | 39 | -0.49 | -0.20 | 2.28 |
| 4 | 29 | 5.43 | 5.20 | 1.24 |
| SNR>15 dB | | | | |
| 1 | 30 | 6.01 | 6.08 | 0.71 |
| 1.4 | 37 | 3.98 | 4.31 | 0.91 |
| 2 | 30 | 1.22 | 1.44 | 1.73 |
| 2.8 | 29 | 1.51 | 2.02 | 1.62 |
| 4 | 22 | 7.35 | 7.39 | 0.83 |

*Table 3 $SE_{MEAS}$ for SFOAE group delay (ms) for 3 SNR criteria. Shown is the center frequency of the cluster, the number of frequency pairs going into the calculation (N), the mean SFOAE group delay from the first test (GD1 and second test (GD2), respectively, and the resulting $SE_{MEAS}$ s.*

| Center Freq (kHz) | N | Mean GD1 (ms) | Mean GD2 (ms) | $SE_{MEAS}$ (ms) |
|---|---|---|---|---|
| SNR>10 dB | | | | |
| 1 | 38 | 9.90 | 9.84 | 0.56 |
| 1.4 | 50 | 8.69 | 8.49 | 0.78 |
| 2 | 36 | 6.48 | 6.45 | 0.42 |
| 2.8 | 38 | 5.31 | 5.29 | 0.58 |
| 4 | 28 | 4.47 | 4.57 | 0.32 |
| SNR>12.5 dB | | | | |
| 1 | 34 | 9.87 | 9.82 | 0.50 |
| 1.4 | 40 | 8.54 | 8.43 | 0.83 |
| 2 | 32 | 6.34 | 6.25 | 0.39 |
| 2.8 | 33 | 5.18 | 5.20 | 0.57 |
| 4 | 25 | 4.55 | 4.71 | 0.29 |
| SNR>15 dB | | | | |
| 1 | 29 | 9.86 | 9.66 | 0.44 |
| 1.4 | 37 | 8.60 | 8.48 | 0.79 |
| 2 | 28 | 6.40 | 6.30 | 0.35 |
| 2.8 | 25 | 5.03 | 5.02 | 0.36 |
| 4 | 22 | 4.57 | 4.81 | 0.28 |

### EFFECT ON RELIABILITY OF NUMBER OF GOOD MEASUREMENTS PER CLUSTER

As seen earlier, as the SNR criterion was increased there were more clusters with only 3 or 4 good measurements, because more and more SFOAEs did not meet the SNR criterion. Did this affect reliability? The $SE_{MEAS}$ actually decreased as SNR increased, even though there were more clusters with only 3 or 4 good measurements. However, there were fewer data for higher SNR criteria; in other words, a quality versus quantity trade-off. The group-delay criteria were developed with the intention of distinguishing "real" SFOAEs from artifact; they were not developed with reliability in mind. The criteria could be made stricter if it was important to improve reliability further, but more data would be lost.

## DISCUSSION

There is an indication of low-level artifactual "SFOAE" responses in the seven hearing-impaired ears. On average, the response in hearing-impaired ears was around 1.2 dB SNR. Various SNR criteria were used to define the presence/absence of an SFOAE. The lower the SNR criterion, the more data from hearing-impaired ears was flagged as having an "SFOAE" present. However, even when reducing the SNR criterion down to 0 dB, no group delay estimates came close to those typically seen in normal-hearing ears, implying that the "SFOAEs" were not true SFOAEs.

Without group delay information, we cannot tell if a measurement with high SNR at a single frequency was due to an artifact or due to a real SFOAE. However, here, even with an SNR criterion of 9 dB, only 1 datum had a high SNR (0.5%), so there is little cause for concern. The caveat of course is that only seven hearing-impaired ears were tested, and the data were somewhat noisy.

In normal-hearing ears, the majority of the data passing the good-quality measurement criteria (stimulus levels on target and low noise) showed high SNRs and very linear group delays, with slope consistent with SFOAE theory. The exception was in the 5.6 kHz cluster, where there was much missing and rejected data. It is recommended that these frequencies around 5.6 kHz are not used in general testing because the percentage of usable data does not seem to warrant the time taken to make the measurements.

The unwrapped phase slope (used to calculate group delay) in normal-hearing people was very negatively linear when sampled over a small frequency region (cluster). This was not the case for hearing-impaired ears. This finding leads to the recommendation that when analyzing a *cluster* containing at least three measurement points, an SNR criterion >0 dB can be used so long as strict correlation and group-delay error criteria are also used; however, SFOAEs with SNR <10 dB are unlikely to add to the quality of the group-delay estimate. When analyzing a single frequency, an SNR criterion of at least 9 dB is required. The criterion needs to be stricter because there is no group-delay information to help validate the measurement. An SNR criterion of >9 dB gave only 0.5% spurious responses in hearing-impaired ears.

Shera also uses information about the SFOAE amplitude to validate measurements. In particular he has noticed poor estimates of group delay in regions where SFOAE amplitude is rapidly changing, because of how the amplitude change affects the phase. In the current data set, this has not been assessed. However, by merely assessing the quality of the group-delay estimate, all spurious measurements in hearing-impaired ears were accounted for.

Reliability (using the standard error of measurement) was assessed for single frequencies, averaged frequencies (for each frequency cluster), and group delay (for each frequency cluster). The $SE_{MEAS}$ for an SFOAE measured at a single frequency can be compared with the $SE_{MEAS}$ for the equivalent DPOAE measurement. For example, the $SE_{MEAS}$ for DPOAEs varied from 1.6 to 4.1 dB data in a field study looking at noise-induced changes in sailors from aircraft carriers (Marshall, Lapsley Miller, & Hughes, 2003). Another example is from a laboratory study on 14 subjects, where the across-session $SE_{MEAS}$ for DPOAEs was 0.97 to 1.73 dB (Marshall & Heller, 2003). Here, within a similar frequency range, the SFOAE $SE_{MEAS}$ varied from 0.7 to 2.3 dB. It is difficult to draw a direct comparison because the measurement methodology, equipment, and data "cleaning" criteria were different, but the indication is that reliability is comparable, and perhaps better, despite the SFOAE being a more difficult measurement to make.

Reliability appeared to be good and also to be comparable with reliability for other OAE types. Whether reliability is good enough for SFOAEs to be used in hearing conservation depends on the size of the noise-induced changes in SFOAEs to be detected and as such is the topic of further study. Further enhancements to the SFOAE measurement should improve reliability further.

A recent finding indicates that in ears with cochlear implants, on-frequency SFOAE noise/distortion is higher and not predictable from off-frequency adjacent bands (Schairer, Fitzpatrick, & Keefe, 2003). It is also higher than that seen in a B&K4157 coupler. However, it becomes apparent only above a probe stimulus level of 40 dB SPL (higher than used here). The Mimosa Acoustics system estimates the on-frequency noise by averaging the noise in the six adjacent frequency bins, so Schairer's finding cannot be investigated with this data set. In addition, Keefe's SFOAE method is quite different to Shera's method, so it is unknown to what extent this phenomenon is methodology dependent. Regardless, the SNR and phase-slope criteria used in the current analysis should still identify spurious SFOAE measurements. However, the high SNR criterion means many good SFOAE measurements possibly are discounted.

## REFERENCES


Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., & Gorga, M. P. (2001). Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears. *J Acoust Soc Am, 110*(6), 3119-3131.

Ghiselli, E. E. (1964). *Theory of Psychological Measurement* (First ed.). NY: McGraw-Hill.

Lapsley Miller, J. A., & Marshall, L. (2001). Monitoring the effects of noise with otoacoustic emissions. *Seminars in Hearing, 22*(4), 393-403.

Lapsley Miller, J. A., Marshall, L., & Heller, L. M. (2003). *A longitudinal study of changes in evoked otoacoustic emissions and pure-tone thresholds as measured in a hearing-conservation program* (Accepted for publication).

Marshall, L., & Heller, L. M. (2003). *A comparison of transient-evoked and distortion-product otoacoustic emissions following short-duration noise exposure in humans*. Unpublished manuscript.

Marshall, L., Heller, L. M., Westhusin, L. J., & Lapsley Miller, J. A. (2000). *EOAEs in NITTS and NIPTS cases from a hearing conservation program.* Paper presented at the Military Audiology Association Meeting, Norfolk, VA, 2 Feb 2000.

Marshall, L., Lapsley Miller, J. A., & Hughes, L. M. (2003). Evoked otoacoustic emissions and hearing before and after a six-month deployment on an aircraft carrier. Unpublished data.

Schairer, K. S., Fitzpatrick, D., & Keefe, D. H. (2003). Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears. *J Acoust Soc Am, 114*(2), 944-966.

Shera, C. A., & Guinan, J. J., Jr. (1999). Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs. *The Journal of the Acoustical Society of America, 105*(2 Pt 1), 782-798.

Shera, C. A., & Guinan, J. J., Jr. (2003). Stimulus-frequency-emission group delay: a test of coherent reflection filtering and a window on cochlear tuning. *J Acoust Soc Am, 113*(5), 2762-2772.

Shera, C. A., Guinan, J. J., Jr., & Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci U S A, 99*(5), 3318-3323.

Stover, L. J., Gorga, M. P., Neely, S. T., & Montoya, D. (1996). Toward optimizing the clinical utility of distortion product otoacoustic emission measurements. *Journal of the Acoustical Society of America, 100*(2 Pt 1), 956-967.

# APPENDIX A: INDIVIDUAL PARTICIPANT SUMMARIES (HEARING-IMPAIRED)

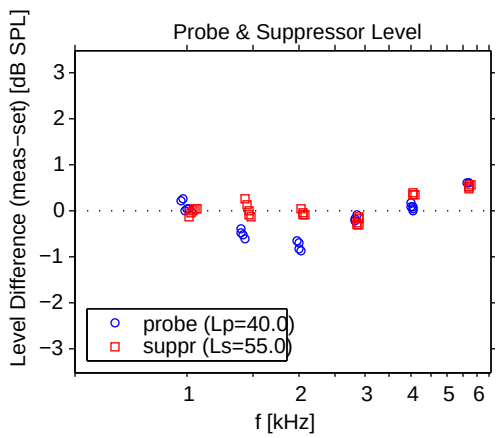
Printouts of the Participant Summaries for the hearing-impaired participants.

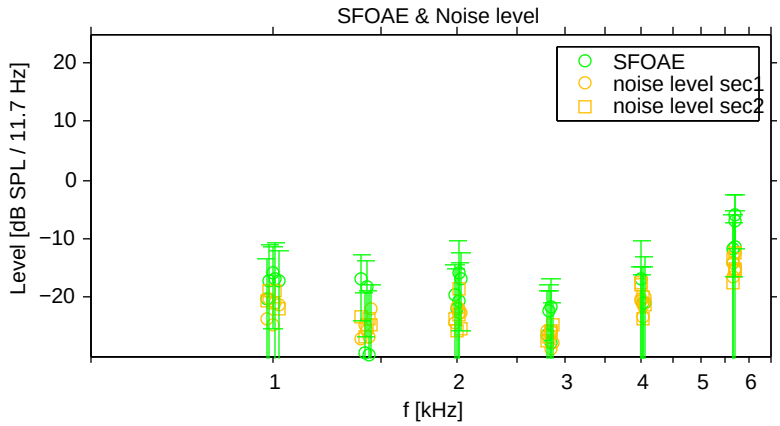
- Participant 1024: flat, moderately severe sensorineural hearing loss through 2.0 kHz, abruptly falling to severe-profound at higher frequencies.
- Participant 1028: profound sensorineural hearing loss in right ear (cochlear implant in right ear, also hearing-impaired in left ear).
- Participant 1029: profound sensorineural hearing loss in right ear (cochlear implant in right ear, also hearing-impaired in left ear).
- Participant 1035: bilateral hearing loss. Left ear chosen for testing. Audiogram was gradually sloping, from moderate to profound, consistent with sensorineural hearing loss.
- Participant 1037: bilateral hearing loss; right ear chosen for testing. Audiogram showed profound sensorineural hearing loss.
- Participant 1038: profound sensorineural hearing loss in right ear from acoustic neuroma surgery; normal hearing in left ear.
- Participant 1041: hearing impaired in both ears; right ear chosen for testing. Audiogram was sharply sloping, from moderately severe to profound, consistent with sensorineural hearing loss.
- Participant 1042: profound sensorineural hearing loss in left ear from acoustic neuroma surgery; normal hearing in right ear.

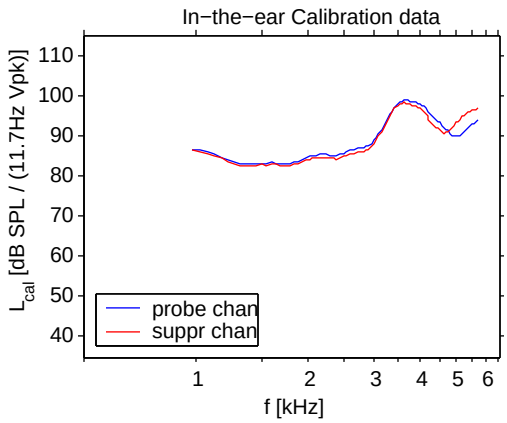
# SFOAE Test Summary

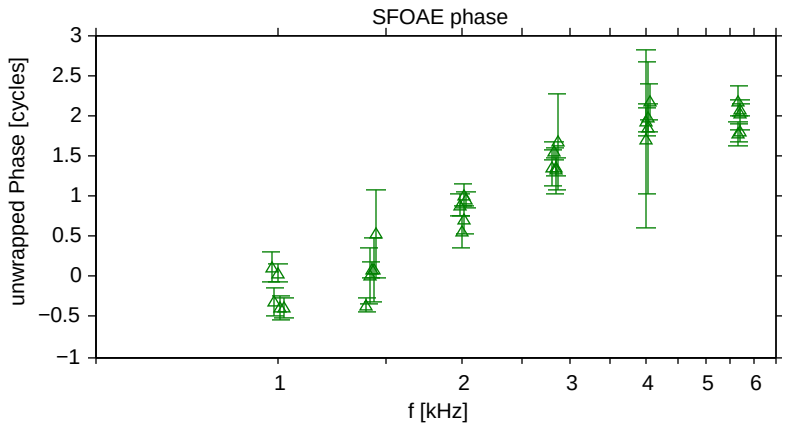
| | | | |
|---|---|---|---|
| Patient ID: | 1024 | | |
| Patient name: | XXX | Test ear: | L |
| Date of birth: | 1951 | Clinician: | NSMRL |
| Date of test: | 2003−08−04 4:57:55 PM | System ID: | none |
| File Name: | C:\SFOAE.Data\1024_XXX_L_sfoae_20030804_165755.mat | | |

Parameters:
Start frequency: 1000 Hz; End frequency: 5686 Hz; Number of freq: 6;
Freq.Difference: −46.9 Hz; Number of fine−freq: 5; Fine−freq. step: 11.7 Hz;
Probe level: 40 dB SPL (rms); Suppressor level: 55 dB SPL (rms); Number of averages: 16;
Number of subaverages: 2; Number of calib averages: 50; Multlevel factor: 2.0;
FFT size: 4096; Sample rate: 48 kHz; Max number of rejections: 64;
Rejection criterion: 37 dB SPL; Probe ID: ER10C #1210




Comments:

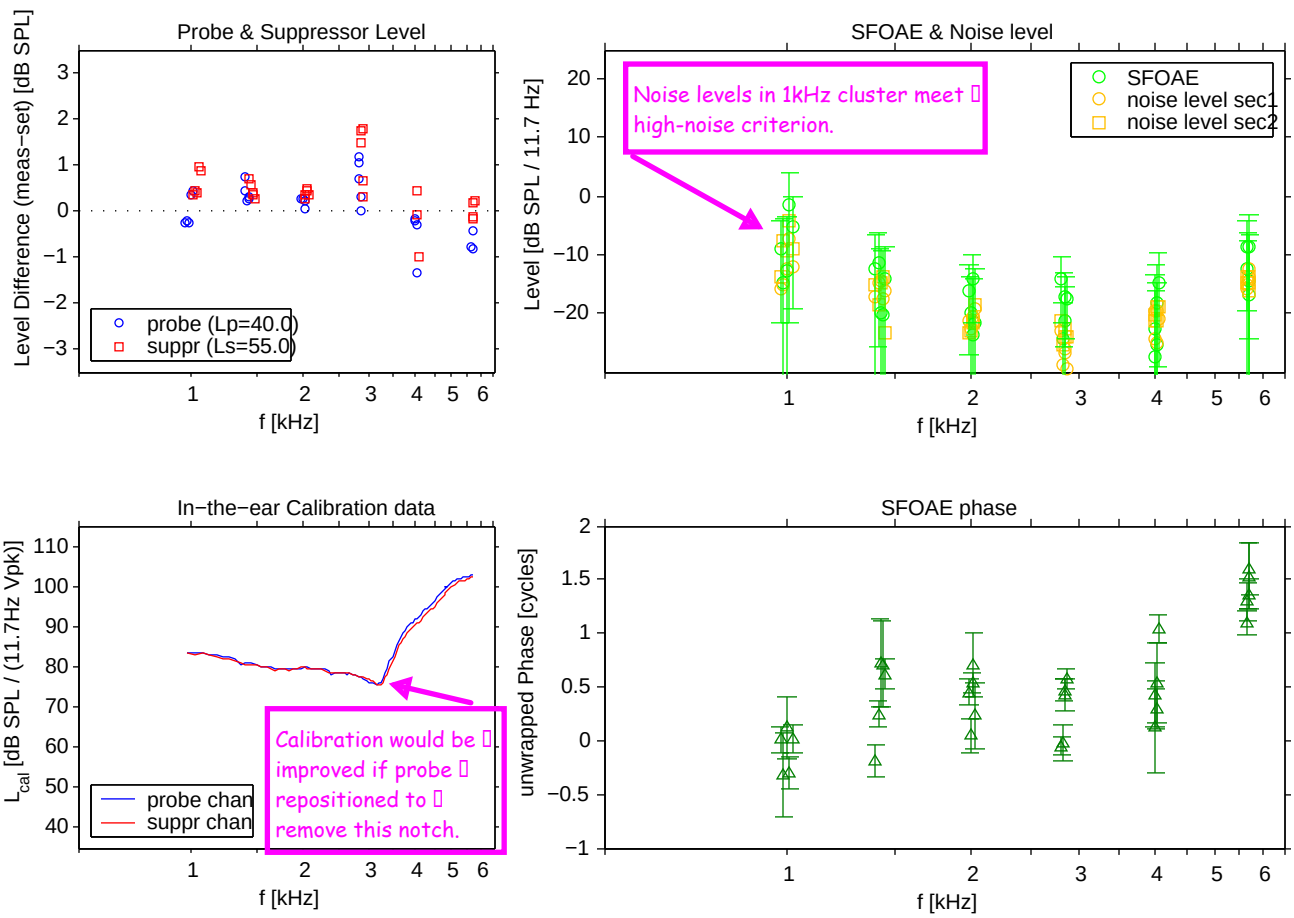
SF2003 version 2.1.18
NSMRL
Groton
860−694−2557

*SFOAE Measurement System − Copyright Mimosa Acoustics*

# SFOAE Test Summary

| | | | |
|---|---|---|---|
| Patient ID: | 1028 | | |
| Patient name: | XXX | Test ear: | R |
| Date of birth: | 1930 | Clinician: | NSMRL |
| Date of test: | 2003−08−05 3:03:00 PM | System ID: | none |
| File Name: | C:\SFOAE.Data\1028_XXX_R_sfoae_20030805_150300.mat | | |

Parameters:
Start frequency: 1000 Hz; End frequency: 5686 Hz; Number of freq: 6;
Freq.Difference: −46.9 Hz; Number of fine−freq: 5; Fine−freq. step: 11.7 Hz;
Probe level: 40 dB SPL (rms); Suppressor level: 55 dB SPL (rms); Number of averages: 16;
Number of subaverages: 2; Number of calib averages: 50; Multlevel factor: 2.0;
FFT size: 4096; Sample rate: 48 kHz; Max number of rejections: 64;
Rejection criterion: 44 dB SPL; Probe ID: ER10C #1210




Comments:

SF2003 version 2.1.18
NSMRL
Groton
860−694−2557

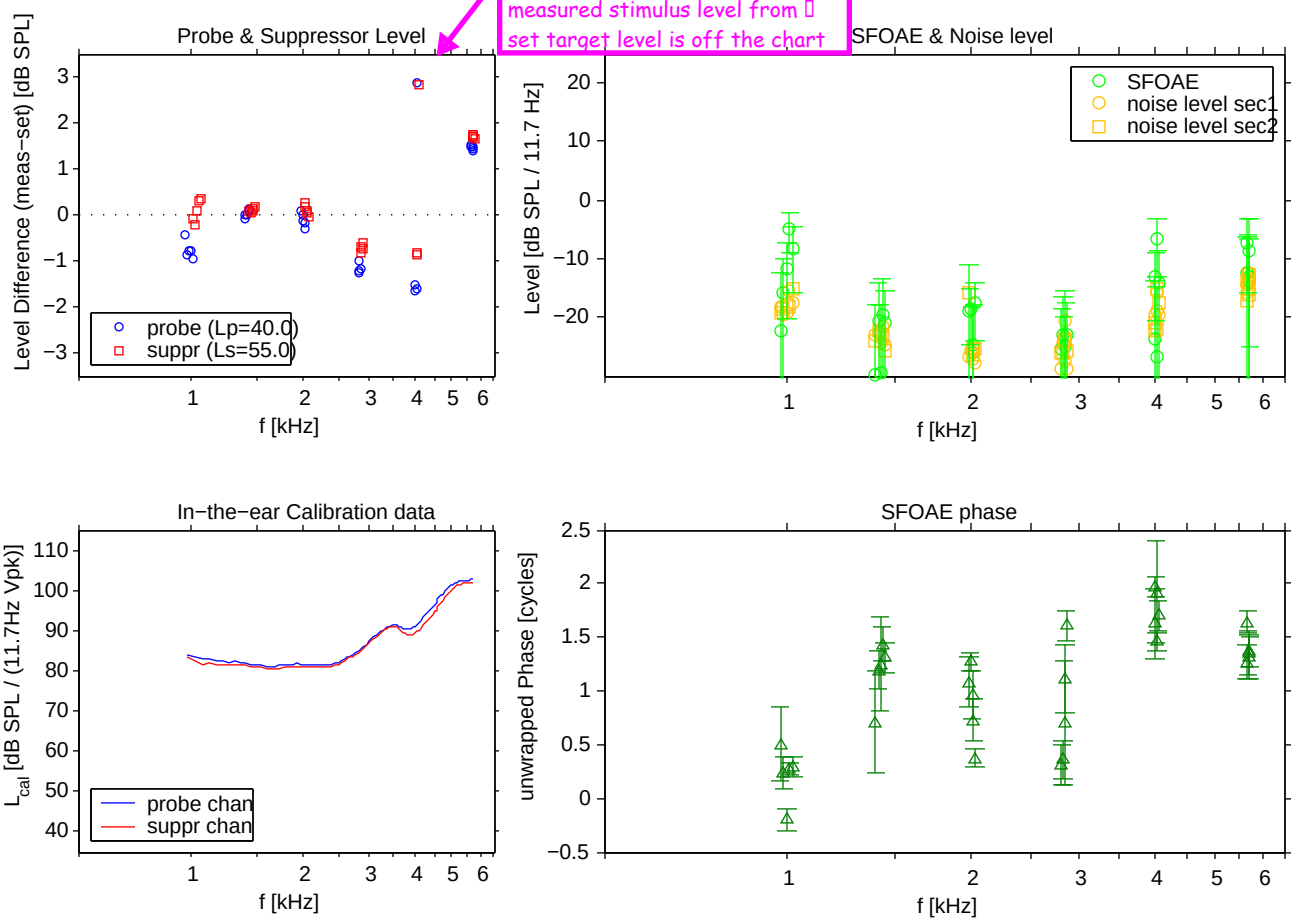
*SFOAE Measurement System − Copyright Mimosa Acoustics*

# SFOAE Test Summary

| | | | |
|---|---|---|---|
| Patient ID: | 1029 | | |
| Patient name: | XXX | Test ear: | R |
| Date of birth: | 1989 | Clinician: | NSMRL |
| Date of test: | 2003−08−06 10:45:42 AM | System ID: | none |
| File Name: | C:\SFOAE.Data\1029_XXX_R_sfoae_20030806_104542.mat | | |

Parameters:
Start frequency: 1000 Hz; End frequency: 5686 Hz; Number of freq: 6;
Freq.Difference: −46.9 Hz; Number of fine−freq: 5; Fine−freq. step: 11.7 Hz;
Probe level: 40 dB SPL (rms); Suppressor level: 55 dB SPL (rms); Number of averages: 16;
Number of subaverages: 2; Number of calib averages: 50; Multlevel factor: 2.0;
FFT size: 4096; Sample rate: 48 kHz; Max number of rejections: 64;
Rejection criterion: 37 dB SPL; Probe ID: ER10C #1210

4.031 kHz: difference of
measured stimulus level from
set target level is off the chart




Comments:

SF2003 version 2.1.18
NSMRL
Groton
860−694−2557

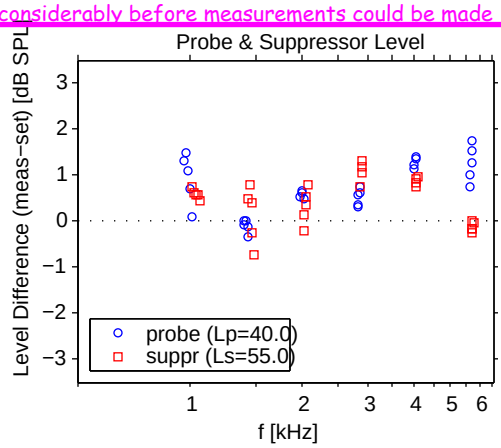
*SFOAE Measurement System − Copyright Mimosa Acoustics*

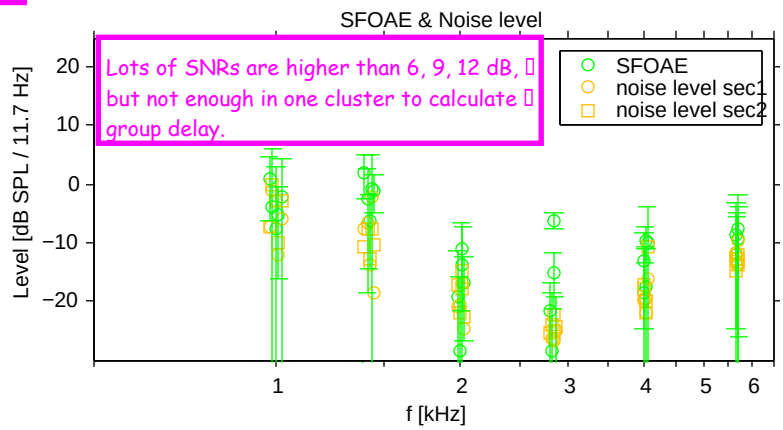
# SFOAE Test Summary

| | | | |
|---|---|---|---|
| Patient ID: | 1035 | | |
| Patient name: | XXX | Test ear: | L |
| Date of birth: | 1923 | Clinician: | NSMRL |
| Date of test: | 2003−08−06 5:33:23 PM | System ID: | none |
| File Name: | C:\SFOAE.Data\1035_XXX_L_sfoae_20030806_173323.mat | | |

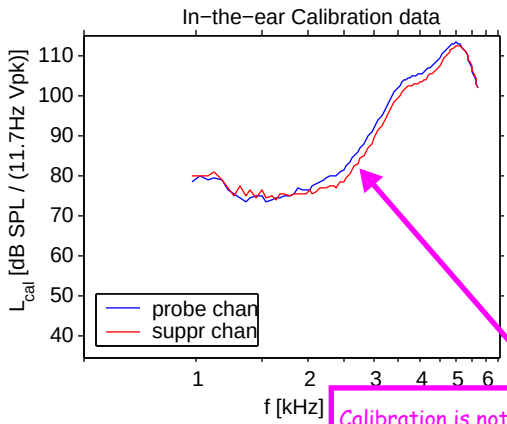
Parameters:
Start frequency: 1000 Hz; End frequency: 5686 Hz; Number of freq: 6;
Freq.Difference: −46.9 Hz; Number of fine−freq: 5; Fine−freq. step: 11.7 Hz;
Probe level: 40 dB SPL (rms); Suppressor level: 55 dB SPL (rms); Number of averages: 16;
Number of subaverages: 2; Number of calib averages: 50; Multlevel factor: 2.0;
FFT size: 4096; Sample rate: 48 kHz; Max number of rejections: 64;
Rejection criterion: 57 dB SPL; Probe ID: ER10C #1210

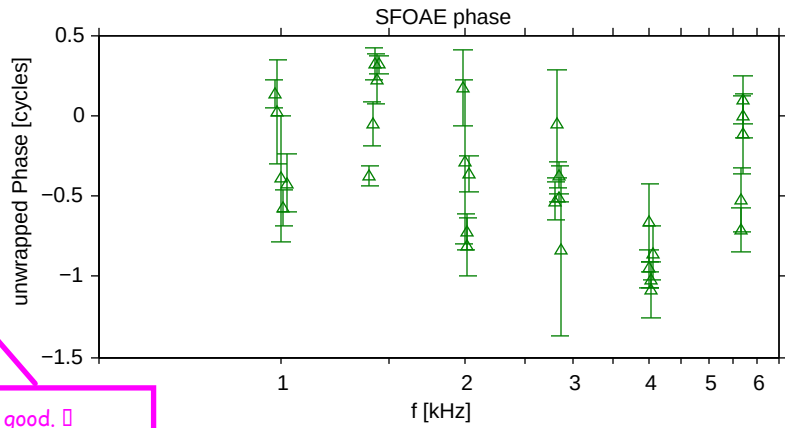
Noise rejection criterion had to be increased considerably before measurements could be made




Lots of SNRs are higher than 6, 9, 12 dB, but not enough in one cluster to calculate group delay.




Calibration is not good. Channels are separated and spectrum is far from flat and smooth. Jitter indicates noise.

Comments:

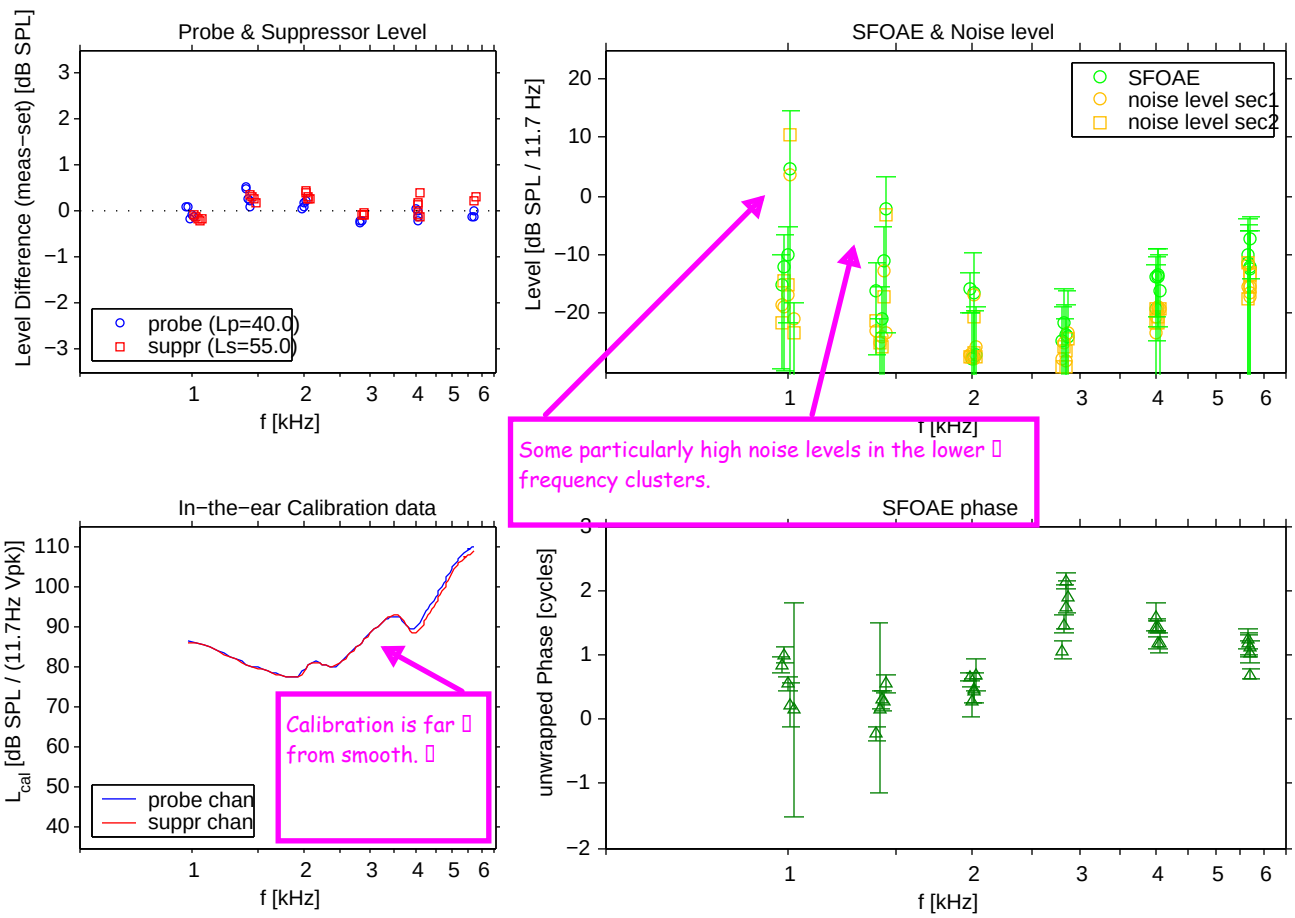
SF2003 version 2.1.18
NSMRL
Groton
860−694−2557

*SFOAE Measurement System − Copyright Mimosa Acoustics*

# SFOAE Test Summary

| | | | |
|---|---|---|---|
| Patient ID: | 1037 | | |
| Patient name: | XXX | Test ear: | R |
| Date of birth: | 1938 | Clinician: | NSMRL |
| Date of test: | 2003−08−07 2:14:34 PM | System ID: | none |
| File Name: | C:\SFOAE.Data\1037_XXX_R_sfoae_20030807_141434.mat | | |

Parameters:
Start frequency: 1000 Hz; End frequency: 5686 Hz; Number of freq: 6;
Freq.Difference: −46.9 Hz; Number of fine−freq: 5; Fine−freq. step: 11.7 Hz;
Probe level: 40 dB SPL (rms); Suppressor level: 55 dB SPL (rms); Number of averages: 16;
Number of subaverages: 2; Number of calib averages: 50; Multlevel factor: 2.0;
FFT size: 4096; Sample rate: 48 kHz; Max number of rejections: 64;
Rejection criterion: 47 dB SPL; Probe ID: ER10C #1210




Comments:

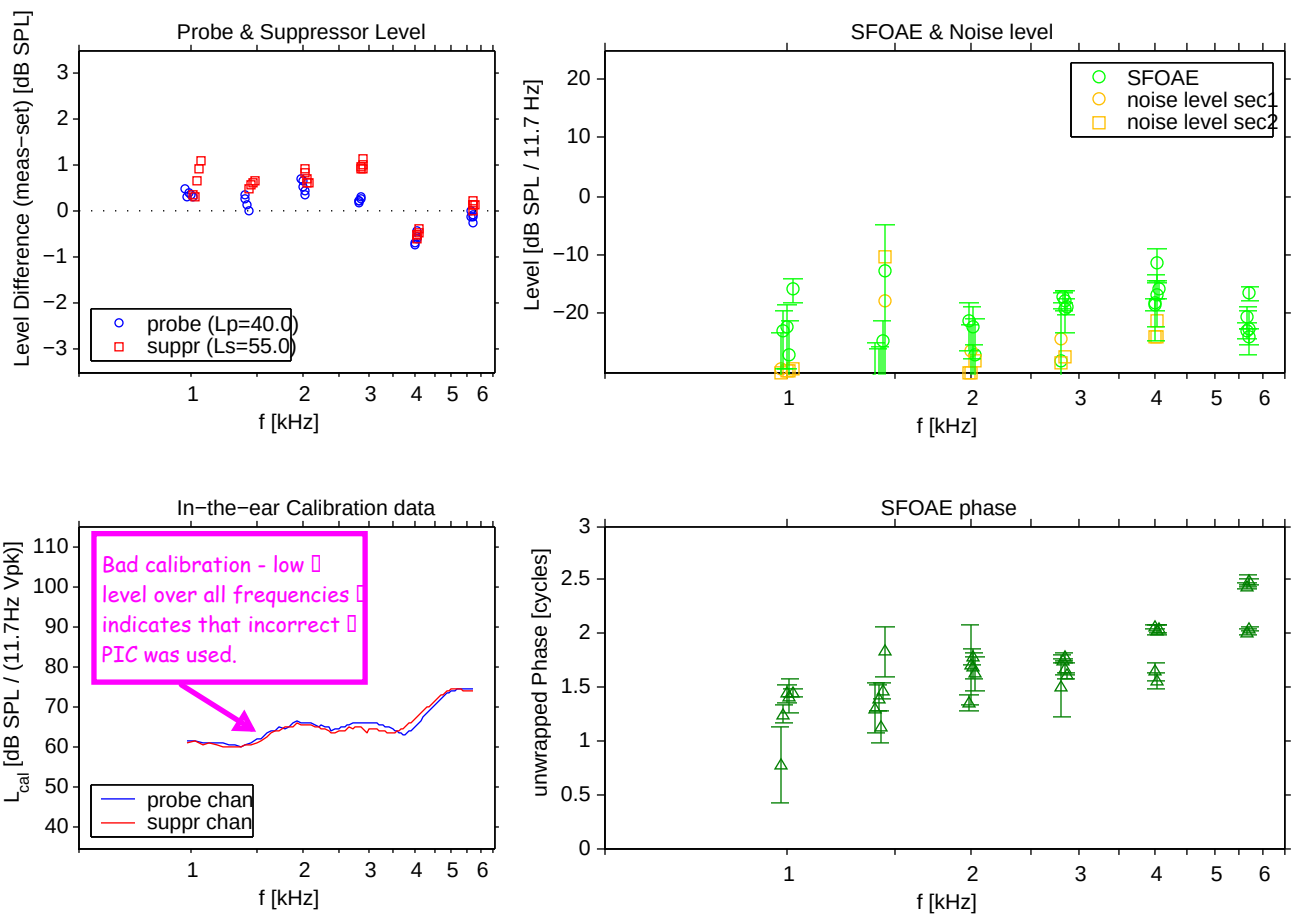
SF2003 version 2.1.18
NSMRL
Groton
860−694−2557

*SFOAE Measurement System − Copyright Mimosa Acoustics*

# SFOAE Test Summary

| | | | |
|---|---|---|---|
| Patient ID: | 1038 | | |
| Patient name: | XXX | Test ear: | L |
| Date of birth: | 1953 | Clinician: | NSMRL |
| Date of test: | 2003−08−07 4:07:21 PM | System ID: | none |
| File Name: | C:\SFOAE.Data\1038_XXX_L_sfoae_20030807_160721.mat | | |

Parameters:
Start frequency: 1000 Hz; End frequency: 5686 Hz; Number of freq: 6;
Freq.Difference: −46.9 Hz; Number of fine−freq: 5; Fine−freq. step: 11.7 Hz;
Probe level: 40 dB SPL (rms); Suppressor level: 55 dB SPL (rms); Number of averages: 16;
Number of subaverages: 2; Number of calib averages: 50; Multlevel factor: 2.0;
FFT size: 4096; Sample rate: 48 kHz; Max number of rejections: 64;
Rejection criterion: 37 dB SPL; Probe ID: ER10C #1210




Comments:
Incorrect PIC was used for this measurement.

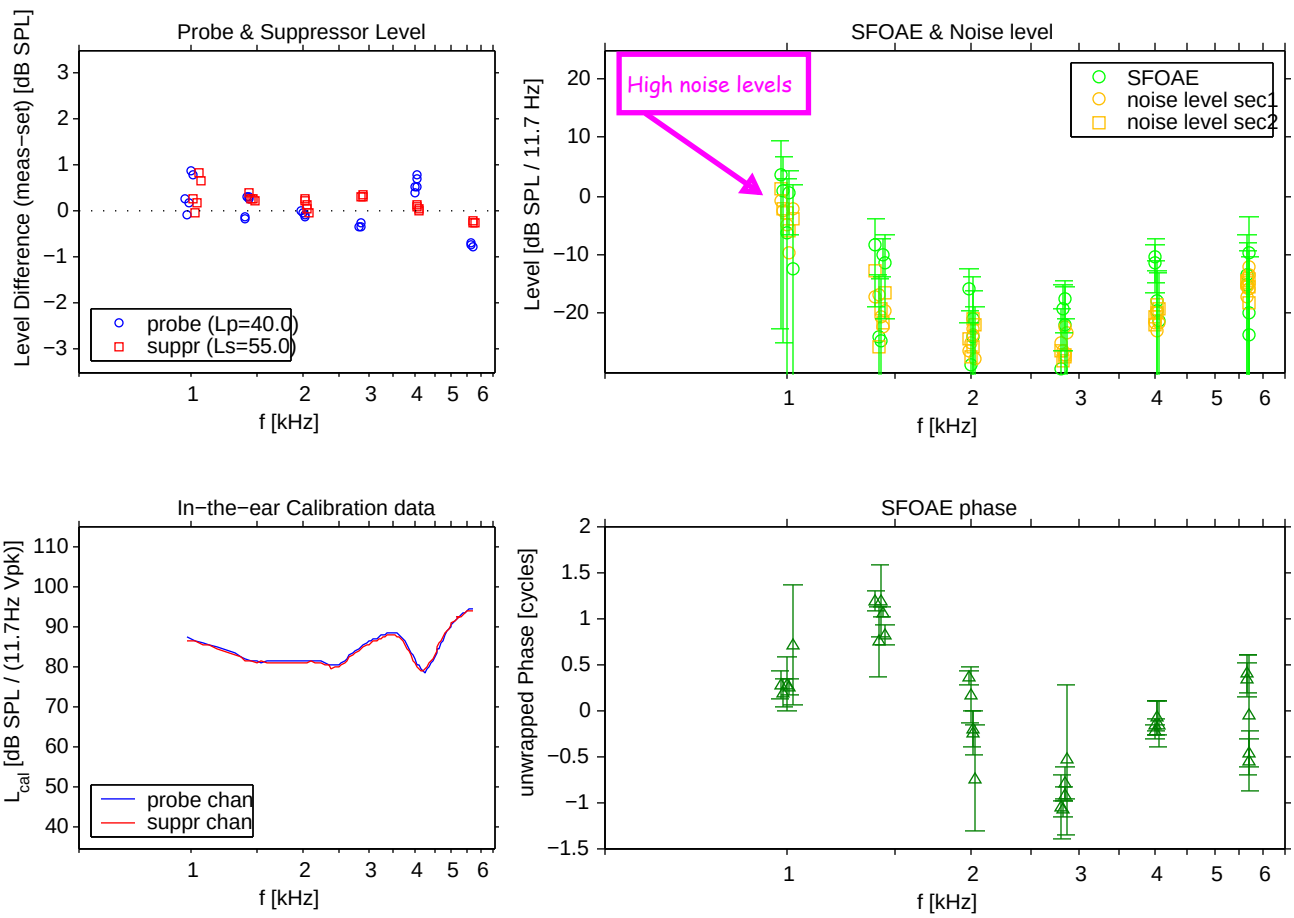
SF2003 version 2.1.18
NSMRL
Groton
860−694−2557

*SFOAE Measurement System − Copyright Mimosa Acoustics*

# SFOAE Test Summary

| | | | |
|---|---|---|---|
| Patient ID: | 1041 | | |
| Patient name: | XXX | Test ear: | R |
| Date of birth: | ??? | Clinician: | NSMRL |
| Date of test: | 2003−08−08 3:11:48 PM | System ID: | none |
| File Name: | C:\SFOAE.Data\1041_XXX_R_sfoae_20030808_151148.mat | | |

Parameters:
Start frequency: 1000 Hz; End frequency: 5686 Hz; Number of freq: 6;
Freq.Difference: −46.9 Hz; Number of fine−freq: 5; Fine−freq. step: 11.7 Hz;
Probe level: 40 dB SPL (rms); Suppressor level: 55 dB SPL (rms); Number of averages: 16;
Number of subaverages: 2; Number of calib averages: 50; Multlevel factor: 2.0;
FFT size: 4096; Sample rate: 48 kHz; Max number of rejections: 64;
Rejection criterion: 37 dB SPL; Probe ID: ER10C #1210




Comments:

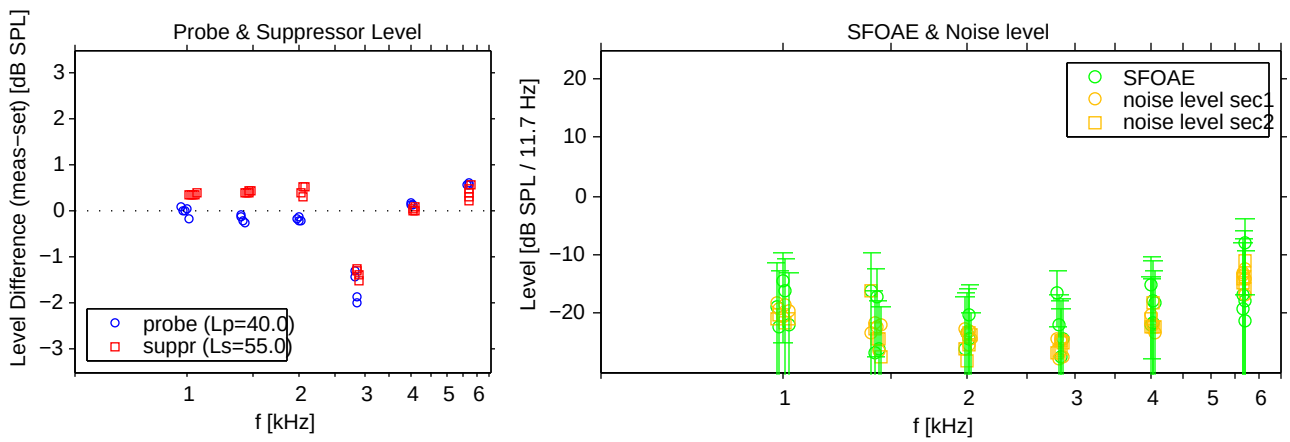
SF2003 version 2.1.18
NSMRL
Groton
860−694−2557

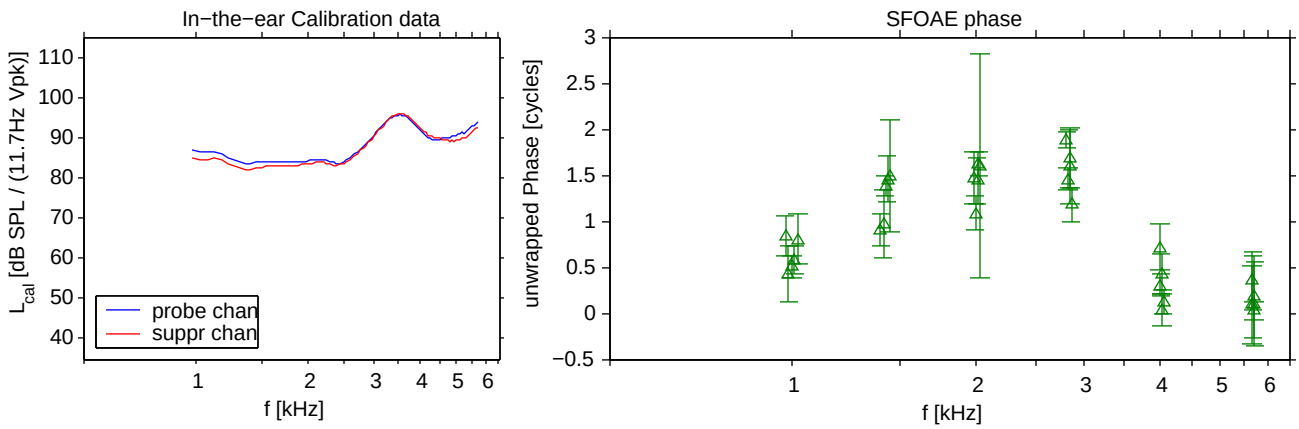
*SFOAE Measurement System − Copyright Mimosa Acoustics*

# SFOAE Test Summary

| | | | |
|---|---|---|---|
| Patient ID: | 1042 | | |
| Patient name: | XXX | Test ear: | L |
| Date of birth: | 1941 | Clinician: | NSMRL |
| Date of test: | 2003−08−13 3:05:05 PM | System ID: | none |
| File Name: | C:\SFOAE.Data\1042_XXX_L_sfoae_20030813_150505.mat | | |

Parameters:
Start frequency: 1000 Hz; End frequency: 5686 Hz; Number of freq: 6;
Freq.Difference: −46.9 Hz; Number of fine−freq: 5; Fine−freq. step: 11.7 Hz;
Probe level: 40 dB SPL (rms); Suppressor level: 55 dB SPL (rms); Number of averages: 16;
Number of subaverages: 2; Number of calib averages: 50; Multlevel factor: 2.0;
FFT size: 4096; Sample rate: 48 kHz; Max number of rejections: 64;
Rejection criterion: 37 dB SPL; Probe ID: ER10C #1210




Comments:

SF2003 version 2.1.18
NSMRL
Groton
860−694−2557

*SFOAE Measurement System − Copyright Mimosa Acoustics*

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| 09-26-2004 | Final | 2001 - 2004 |

**4. TITLE AND SUBTITLE**

STIMULUS-FREQUENCY OTOACOUSTIC EMISSIONS
VALIDITY & RELIABILITY OF SFOAES IMPLEMENTED ON
MIMOSA ACOUSTICS SFOAE MEASUREMENT SYSTEM V2.1.18

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

1) Judi A. Lapsley Miller, Ph.D.,
2) Paul Boege, Ph.D.,
3) Lynne Marshall, Ph.D.,
4) Christopher A. Shera, Ph.D., and
5) Patricia S. Jeng, Ph.D.

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

50309

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

NAVSUBMEDRSCHLAB
Box 900
Groton, CT 06349-5900

**8. PERFORMING ORGANIZATION REPORT NUMBER**

NSMRL Technical Report #TR1231

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution unlimited.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

The validity and reliability of stimulus-frequency otoacoustic emissions (SFOAEs) were evaluated and determined with the Mimosa Acoustics SFOAE measurement system (v.2.1.18). Stimulus-frequency otoacoustic emissions are sounds made by healthy inner ears to tonal stimulation. They are thought to hold the potential for diagnosis and monitoring of noise-induced hearing loss in military hearing-conservation programs. By establishing the validity and reliability of the measurement, changes in SFOAEs due to noise can be detected from artifact and test-retest variability. WU# 50309 – Evoked Otoacoustic Emissions in Military Hearing Conservation Programs

**15. SUBJECT TERMS**

otoacoustic, otoacoustic emissions, stimulus-frequency, stimulus-frequency otoacoustic emissions, hearing conservation

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | M. Fitzgerald, Publications Manager |
| U | U | U | | 37 | 19b. TELEPHONE NUMBER *(Include area code)* (860) 694-2442 |