2014

Intrasubject Variability in Power Reflectance

Defne Abur
Smith College

Susan E. Voss
Smith College, svoss@smith.edu

Nicholas J. Horton
Amherst College

J Am Acad Audiol 25:441–448 (2014)

Intrasubject Variability in Power Reflectance

DOI: 10.3766/jaaa.25.5.3

Defne Abur*
Nicholas J. Horton†
Susan E. Voss*

Abstract

Background: Power reflectance measurements are an active area of research related to the development of noninvasive middle-ear assessment methods. There are limited data related to test-retest measures of power reflectance.

Purpose: This study investigates test-retest features of power reflectance, including comparisons of intrasubject versus intersubject variability and how ear-canal measurement location affects measurements.

Research Design: Repeated measurements of power reflectance were made at about weekly intervals. The subjects returned for four to eight sessions. Measurements were made at three ear-canal locations: a deep insertion depth (with a foam plug flush at the entrance to the ear canal) and both 3 and 6 mm more lateral to this deep insertion.

Study Sample: Repeated measurements on seven subjects are reported. All subjects were female, between 19 and 22 yr old, and enrolled at an undergraduate women's college.

Data Collection and Analysis: Measurements on both the right and left ears were made at three ear-canal locations during each of four to eight measurement sessions. Random-effects regression models were used for the analysis to account for repeated measures within subjects. The mean power reflectance for each position over all sessions was calculated for each subject.

Results: The comparison of power reflectance from the left and right ears of an individual subject varied greatly over the seven subjects; the difference between the power reflectance measured on the left and that measured on the right was compared at 248 frequencies, and depending on the subject, the percentage of tested frequencies for which the left and right ears differed significantly ranged from 10% to 93% (some with left values greater than right values and others with the opposite pattern). Although the individual subjects showed left-right differences, the overall population generally did not show significant differences between the left and right ears. The mean power reflectance for each measurement position over all sessions depended on the location of the probe in the ear for frequencies of less than 1000 Hz. The standard deviation between subjects' mean power reflectance after controlling for ear (left or right) was found to be greater than the standard deviation within the individual subject's mean power reflectance. The intrasubject standard deviation in power reflectance was smallest at the deepest insertion depths.

Conclusions: All subjects had differences in power reflectance between their left and right ears at some frequencies; the percentage of frequencies at which differences occurred varied greatly across subjects. The intrasubject standard deviations were smallest for the deepest probe insertion depths, suggesting clinical measurements should be made with as deep an insertion as practically possible to minimize variability. This deep insertion will reduce both acoustic leaks and the effect of low-frequency ear-canal losses. The within-subject standard deviations were about half the magnitude of the overall standard deviations, quantifying the extent of intrasubject versus intersubject variability.

Key Words: Middle ear, reflectance, test-retest, intrasubject variability

Wideband acoustic immittance measures in the ear canal are a potential noninvasive diagnostic test for conductive hearing loss and the causes of conductive hearing loss (Stinson et al, 1982; Keefe et al, 1993; Keefe and Levi, 1996; Vander Werff et al, 2007; Shahnaz et al, 2009; Voss and Allen,

*Picker Engineering Program, Ford Hall, Smith College, Northampton, MA; †Department of Mathematics and Statistics, Amherst College, Amherst, MA

Susan E. Voss, Picker Engineering Program, Ford Hall, Smith College, Northampton, MA; Phone: (413) 585-7008; Fax: (413) 585-7001; E-mail: svoss@smith.edu

This work was supported by the National Science Foundation through both a CAREER award and the Four College Biomath Consortium.

1994). One factor limiting the development of such diagnostic tests is the need to describe variability in test-retest data or intrasubject variability. To date, limited test-retest data are available in the literature.

A handful of studies have investigated the test-retest reliability of power reflectance data, but most of these only include a total of two measurements per subject. Both Vander Werff et al (2007) and Werner et al (2010) showed with two unique measurements per subject that adults tend to have smaller test-retest variability than infants. Werner et al (2010) also provide substantial discussion and analysis of their test-retest data, including evidence that general frequency trends of power reflectance measurements across frequency are consistent on repeated measurement. More recently, Rosowski et al (2012) made a total of four repeated measurements (weekly) on seven subjects; their results show standard deviations within subjects that are similar to population standard deviations at the lowest frequencies (<500 Hz), and at higher frequencies, the standard deviations within subjects are one half to two thirds of the population standard deviations. Rosowski et al (2012) concluded that "a significant fraction of the differences seen between individuals may be explained by intra-subject variations in reflectance measurements."

The goal of the present work is to add to the growing literature regarding test-retest measurements and to quantify a few specific issues related to understanding intrasubject variability. Specifically, we examine differences between the left and right ears of subjects, effects of the probe position within the ear canal, and test-retest variability in repeated measurements over periods of several weeks.

METHODS

Subjects

Power reflectance was measured in the ear canals of 10 female subjects (age range, 19-22 years). Measurements were approved by the Smith College Institutional Review Board, and each subject provided written consent. Female subjects were selected because the study was done at an undergraduate women's college and the available subjects were female. Seven subjects are included in the analyses presented here because three of them (subjects 2, 5, and 9) did not return for four or more measurement sessions.

Before participating in the study, each subject's ears were checked for excessive ear wax, and subjects underwent an audiogram at 500, 1000, 2000, and 4000 Hz to ensure normal hearing; all seven subjects had hearing levels of 10 dB or less at these four frequencies. Additionally, all subjects had normal tympanograms at every measurement session.

Measurement Methods

Measurements were made during four to eight unique measurement sessions for each of the seven subjects. For five of the subjects (subjects 1, 3, 5, 6, and 7), measurements were made at about weekly intervals. For one subject (subject 8), five measurements were made over a period of 2 wk. For one subject (subject 10), the first three measurements were made at weekly intervals, with a fourth measurement made just less than 5 mo later; in this case there are no obvious differences in the measurements made weekly and the one made several months later.

At each session, tympanic peak pressure was measured by means of tympanometry (in both ears) with a Zodiac 901 middle-ear analyzer; tympanometry was used to ensure continued normal middle-ear function between sessions and to control for potential changes in middle-ear pressure. Next, pressure measurements in the ear canal were made with the Mimosa Acoustics' HearID Auditory Diagnostic System Version 4.0; this experimental system allowed two consecutive measurements on two distinct channels. Power reflectance was calculated from the Thevenin equivalent (or calibration) measured for the system.

All measurements were made with the Etymotic 14A yellow foam Eartip coupled to the Mimosa system's Etymotic ER10c transducer. Ear-canal pressure measurements (and subsequent power reflectance calculations) were made in both the left and right ears at three locations. For each measurement session, two lines were drawn with marker on the yellow foam tip; one line was 3 mm from the lateral end, and the other line was 6 mm from the lateral end. One of two experimenters inserted the yellow foam tip to a position judged to have an insertion depth exactly flush with the entrance to the ear canal. The first measurement was made at this position 1 and is referred to as the deep insertion. Next, the yellow foam tip was pulled away from the tympanic membrane so that the first line 3 mm from the end of the tip was at the entrance to the ear canal; a second measurement was made at this position 2, which was 3 mm lateral to position 1. Next, the yellow foam tip was pulled farther away from the tympanic membrane so that the second line 6 mm from the end of the tip was at the entrance to the ear canal; a third measurement was made at this position 3, which was 6 mm lateral to position 1.

Data Selection Criteria

At each session, all subjects' ears were within 25 daPa of zero tympanic peak pressure, except for one measurement on subject 8 that was −45 daPa; no measurements were omitted because of tympanic peak pressure deviations from zero. For all analyses, all subject data were used, with the exception of measurements from two sessions that were consistent with an acoustic leak (subject 3, session 5; subject 10, session 4), as assessed by the low-frequency impedance phase having an angle approaching or greater than zero (Keefe et al, 2000).

Statistical Analysis

For each subject, a test was done to determine whether the power reflectance differed between the left and right ears of that subject. For each subject, a 95% confidence interval was computed for the difference between each power reflectance measurement made on the left minus the right ear. These confidence intervals were computed by using a permutation test with the MATLAB function "bootci" (MATLAB version 7.12.0.635); the number of permutations was 10,000 iterations ("NBOOT"), and all other function inputs were the default values. The interpretation of these computations is that when the 95% confidence interval contains the number zero, then there is not strong evidence for a difference between the left and right ears, but when the confidence interval does not contain zero for a substantial number of neighboring frequencies, then there is likely a difference between the left and right ears. We test 248 frequencies for each subject, and we report the percentage of frequencies that have a 95% confidence interval that includes zero.

Additionally, we tested whether differences exist between the left and right ears across the entire population. For this test, a linear mixed-effects model (Laird and Ware, 1982) was used to analyze the significance between the left and right ears. For each measurement location (positions 1, 2, and 3), the ear (left or right) was treated as a fixed effect, and the subject was treated as a random effect. The p values were calculated at each of the 248 measurement frequencies.

A separate linear mixed-effects model (Laird and Ware, 1982) was used for each position and each frequency to understand the variability in power reflectance between subjects and a given subject's test-retest variability. These models assessed differences in power reflectance, as well as variability between and within subjects' ears at a fixed position. Each ear (left or right) was treated as a random effect, with fixed-effects terms for session.

RESULTS

Overview of Measurements

Power reflectance measurements made over four to eight unique sessions, at three locations, and on seven subjects (2 channels measured during each session) are shown in Figure 1. All subjects show power reflectance that approaches 1 at the lowest frequency (200 Hz) and decreases with increasing frequency to a local minimum between 700 and 2000 Hz. Above this first (lowest frequency) minimum, measurements show a fine structure that is generally novel to the individual subject. At the highest frequencies (4000–6000 Hz), power reflectance generally increases toward 1 in all subjects.

Left versus Right

The individual measurements from Figure 1 show that there are systematic differences between the left and right ears for some subjects. At position 1, subject 4 has a minimum at a lower frequency for the left ear than for the right ear. Similarly, the low-frequency minimum occurs at a lower frequency for the right ear of subject 6 than for the left ear. These differences are consistent across all measurements in these ears. In contrast, the extrema for both subjects 8 and 10 are more consistent between the left and right ears.

The right-most column of Figure 1 plots the 95% confidence intervals at position 1 for the differences in power reflectance between the left and right ears. The results appear different for each of the seven subjects. Subjects 3, 7, and 10 have relatively smaller difference between their left and right ears, with 95% confidence intervals that generally hover around zero and include zero at a majority of frequencies. Three of the other subjects (subjects 1, 4, and 6) have more systematic differences between their left and right ears, with the majority of frequencies not having confidence intervals that include zero. The seventh subject (subject 8) falls somewhere in between, with about half of the frequencies including zero in the 95% confidence interval. In summary, the results show some differences between the left and right ears, and the extent and direction of these differences depends on the subject.

Figure 2 plots the p values from the linear mixed-effects models developed for each frequency to test whether there are differences between the left and right ears of the entire population. The models treated ear (left or right) as a fixed effect and subject as a random effect. The p values for each measurement location show narrow bands of frequencies at which the ear effect might be significant, but for most frequencies, the p value was not significant at the 0.001 level. There are fewer significant frequencies at position 1 than at positions 2 and 3. All three positions appear to have a significant difference near 1000 Hz.

Measurement Location

The mean power reflectance measurements at each of the three locations are plotted in the upper part of Figure 3; specifically, these are the mean values calculated from the means from each of the seven individual

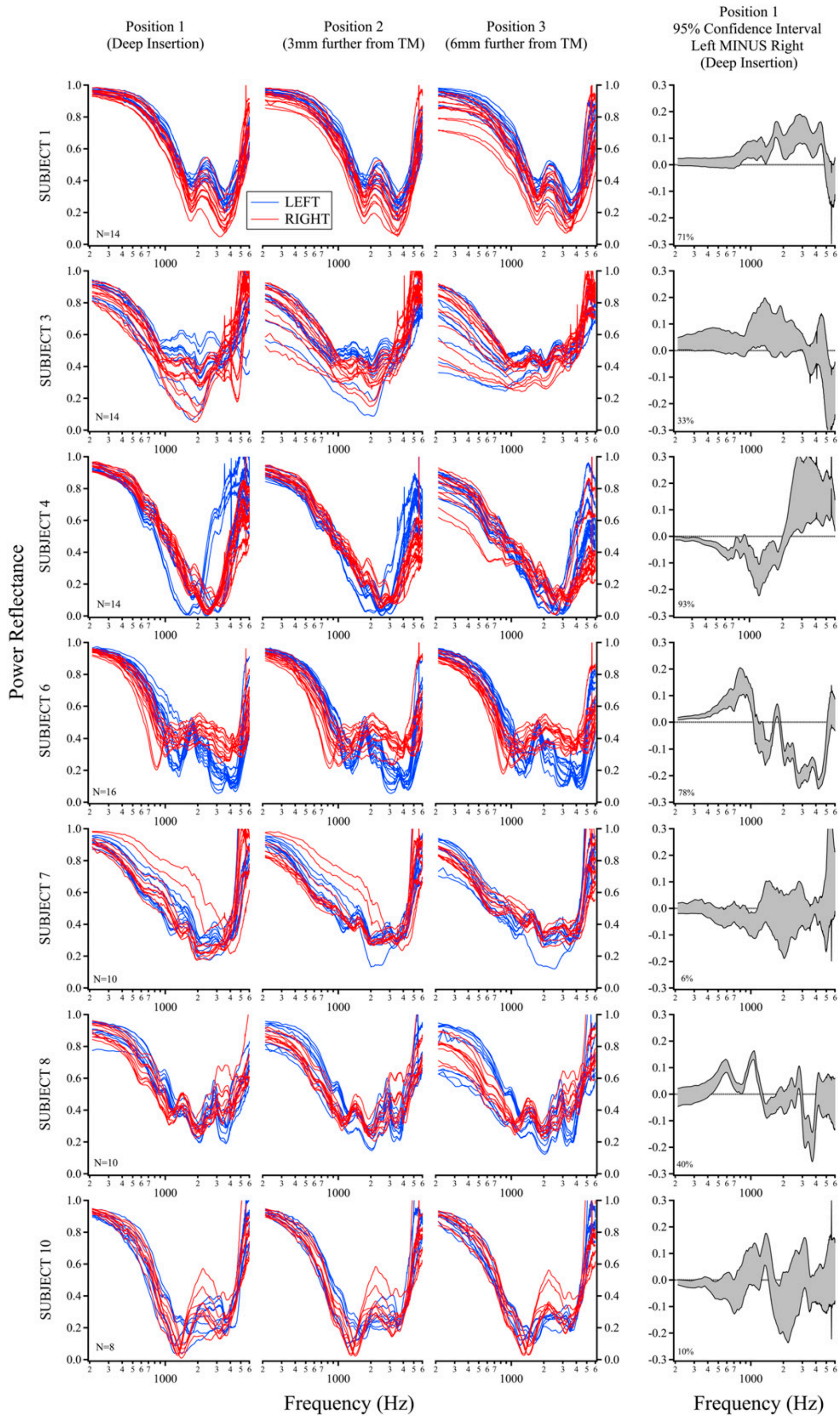

Figure 1. *Left*, Repeated measurements of power reflectance on the left (blue) and right (red) ears of seven subjects. Measurements were made at three measurement positions, with the farthest insertion (position 1) made first, followed by positions 2 and 3. *Right*, Computed 95% confidence intervals to test the null hypotheses that the difference in the power reflectance between the left and right ears (of a given subject) is zero. These calculations are associated with the measurements at position 1 only. The percentage indicated in the lower left corner indicates the percentage of frequencies for which the confidence interval does not include zero. Note the frequencies are plotted on a logarithmic scale but their actual spacing for the measurements is linear, and therefore as frequency increases, the density of data points on these plots also increases. *TM*, Tympanic membrane.

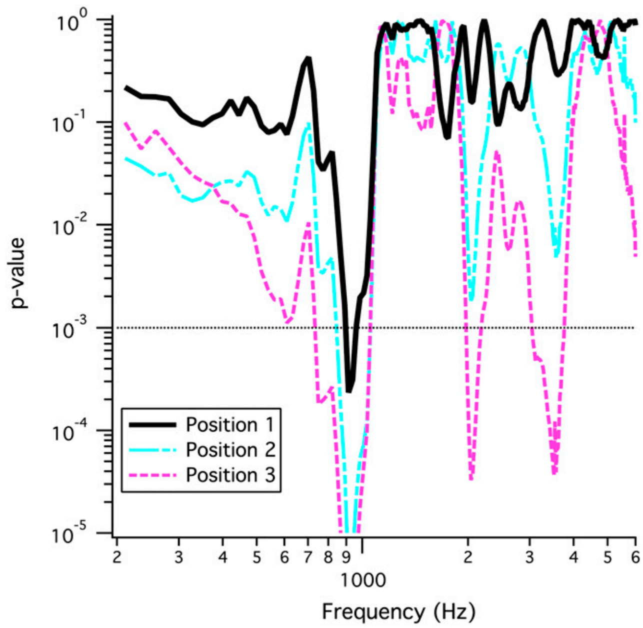

Figure 2. Computed *p* values from the linear mixed-effects model used to examine differences between all left and right ears (population) at each measurement location.

subjects. At less than about 1000 Hz, the mean power reflectance at each position (left and right ears) shows a systematic decrease as the measurement location moves from position 1 to position 3. At greater than about 1000 Hz, the power reflectance does not appear to be affected by the measurement location. Individual plots from each subject show the same trends and are not reproduced here. Of the seven subjects, there is not a single example that does not fit the trends shown by these mean plots; in two of the ears the differences between positions 1 and 2 are nearly indistinguishable, but there is no example that violates the trends demonstrated here.

Intrasubject and Intersubject Variability

A summary of two measures of intrasubject variability is plotted in the lower part of Figure 3. First, the standard deviations of the mean power reflectance calculated at each position are plotted (solid lines). The trends in the standard deviations calculated across all subjects are the same as the individual measurements; at the lower frequencies ($<$1000 Hz), the standard deviations increase systematically as the measurement location moves from position 1 to position 3. At greater than 1000 Hz, there is more variability in any sort of trend across subjects, and the standard deviations appear far less dependent on measurement location.

The second measure of intrasubject variability is the mean of the differences between the measurements at position 1 and position 2 and the differences between position 1 and position 3. These calculations show the variability introduced by measurement location: at frequencies of less than about 1000 Hz, power reflectance varies by up to about 0.1, when the deep insertion (position 1) is compared with the most lateral insertion (position 3). Again, the trends for every subject match these mean trends for frequencies of less than 1000 Hz, at which the differences are largest.

At greater than 1000 Hz, the standard deviations of the mean power reflectance measurements (intersubject variability) are generally larger than the variability that results from measurement location alone. At the lower frequencies (less than about 1000 Hz), the measurement location is an important part of the comparison: the differences in measurements between position 1 and position 3 are generally larger than the standard deviation of the population mean.

To compare intrasubject variability with intersubject variability, all measurements on both ears (left and right) were combined, and a linear mixed-effects model was used to calculate the within-subject and between-subject standard deviations in power reflectance (Figure 4). Results from position 1 are reported, in which the entire overall (population) standard deviation (calculated from all subjects and all measurements) is compared with the within- and between-subject standard deviations. The overall standard deviation is the standard deviation for position 1 (Figure 3, lower plots), but here, the left and right ears have been combined. The overall standard deviation is the largest of the three standard deviations, and the variability between subjects is consistently higher than the variability within subjects.

DISCUSSION

Overview of Measurements

The general features of the power reflectance measurements reported here are consistent with other reports from normal adult ears within the literature (Keefe et al, 1993; Voss and Allen, 1994; Shahnaz and Bork, 2006; Rosowski et al, 2012). Specifically, the power reflectance approaches 1 at the lower frequencies (a few hundred Hertz), decreases with increasing frequency to a minimum that is in the range of 500 to 2000 Hz, has a fine structure dependent on the subject at the middle frequencies (1000 to 4000 Hz), and increases toward 1 at greater than about 4000 Hz.

Left versus Right

All of the left and right ear pairs showed differences in power reflectance at some of the frequencies (Figure 1), but the extent of the differences depended largely on the individual subject. Three of the seven subjects had a majority of frequencies with no detectable differences

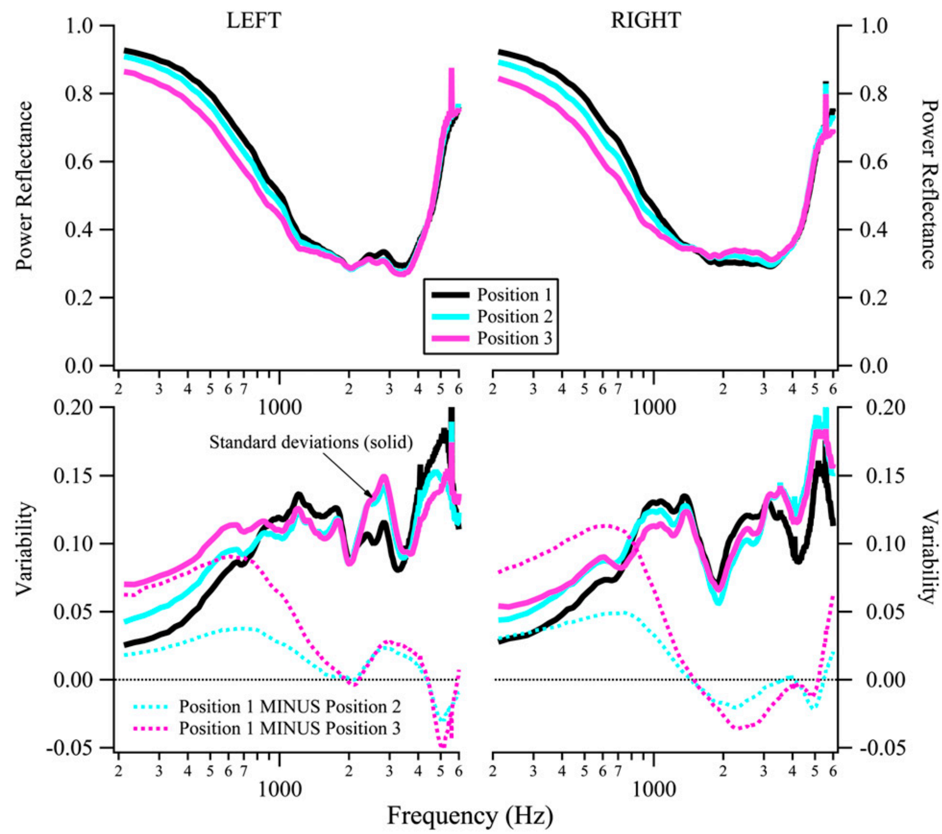

Figure 3. *Top*, Mean of power reflectance measurements for left (left column) and right (right column) ears of all seven subjects for measurement position 1 (black), position 2 (cyan), and position 3 (magenta). *Bottom*, Solid lines indicate standard deviations of the mean power reflectance measurements by position; these are the standard deviations of the mean measurements plotted above. Dashed lines indicate mean differences in power reflectances calculated between position 1 and position 2 (cyan) and position 1 and position 3 (magenta).

between the left and right ears, three of the seven subjects had the majority of frequencies with significant differences between the left and right ears, and one subject fell in the middle of these groups. It is likely that differences exist more often than not between the left and right ears when the lowest frequency minimum in power reflectance occurs at a different frequency in the left and right ear. Although the general pattern of the power reflectance might appear similar between the left and right ears, the computed difference will be affected by shifts in frequency of power reflectance extrema.

Across all ears, the mixed-effects model identifies a narrow frequency range from about 800 to about 1000 Hz for which the left and right ears could be different (Figure 2). At the same time, the significance of the difference is the least prominent for the deepest insertion depth (measurement location at position 1). At position 1, the significance of the difference between the left and right ears is certainly open to interpretation. Here the relatively large differences between the left and right ears of subjects 4 and 6 play a substantial role in the model results because there are only seven subjects. Measurements on additional subjects would be needed to draw definite conclusions as to whether there are true left-right differences in a larger population.

Measurement Location

The measurement location within the ear canal (assessed at three positions here) affects power reflectance in two systematic ways. First, at less than 1000 Hz, the magnitude of the power reflectance decreases systematically for frequencies of less than 1000 Hz, with a total average change in the measured power reflectance across the 6-mm measurement location span approaching about 0.1. At greater than 1000 Hz, there are no clear systematic effects from the three measurement locations tested here. These results are consistent with the work of Voss et al (2008). At the most lateral parts of the ear canal, losses in the ear canal lead to low-frequency reductions in power reflectance; the longer the ear canal, the more loss and, consequently, the more absorbance of sound energy along the canal walls, leading to a lower reflectance. Using measurements of power reflectance in cadaveric ear canals, Voss et al (2008) also suggest that measurements made at deeper insertion depths, where the measurement

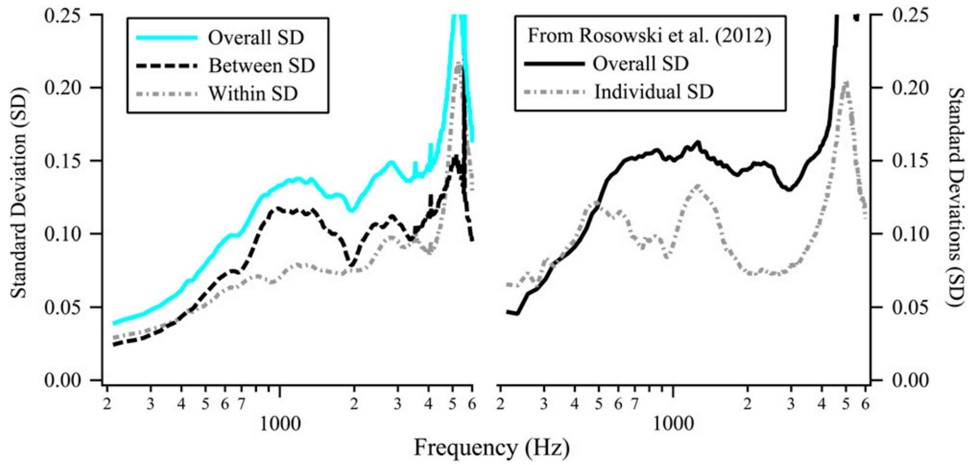

Figure 4. *Left*, All quantities are for the measurement location at position 1 (deep insertion). Results from the linear mixed-effects model for the overall (population) standard deviation, within-subject standard deviation, and between-subject standard deviation are shown. *Right*, Similar data from Rosowski et al (2012) in which four repeated power reflectance measurements were made on each of seven subjects. The solid black line is the overall standard deviation, and the dashed gray line is the mean standard deviation within subjects.

location is within the bony part of the ear canal and not the cartilaginous part, show much smaller changes with measurement location.

A second effect of measurement location involves the variability of the measurement itself. Figure 3 shows that the standard deviations associated with repeated measurements at the deepest measurement location (position 1) are generally smaller than those at the two more lateral positions. Again, this is most prominent at frequencies of less than 1000 Hz. The consistency of the measurements from session to session can also be appreciated visually by inspection (Figure 1, left), where some of the subjects clearly have more variable measurements at position 3 than at position 1. Thus it appears that the deeper insertion depth of the probe leads to more consistent measurements. These observations might be explained by a combination of at least two factors: (1) the deeper insertion depth likely leads to improved acoustic seals with fewer effects of undetected but perhaps small acoustic leaks and (2) the deeper insertion depth likely is less affected by acoustic losses along the ear canal, which add additional variability when the probe is not placed at the identical location across multiple measurement sessions.

Intrasubject and Intersubject Variability

Figure 4 shows that between-subject variability is larger than within-subject variability for position 1. This is also the case for the other two measurement locations (results not plotted). The trends of these results are not surprising: Variability within subjects would be expected to be smaller than variability between subjects because anatomic differences across measurement sessions are not expected, especially because tympanometry was performed to ensure that changes in middle-ear pressure did not occur. At the same time, the results do show that the within-subject variability is a substantial fraction of both the between-subject variability and the overall standard deviation. A reasonable summary of these results is that the within-subject variation is about one half of the overall standard deviation.

Figure 4 shows similar results presented by Rosowski et al (2012) of test-retest variability. The results of Rosowski et al (2012) have a slightly larger overall standard deviation and individual standard deviation than our results, but the two sets of data and analyses are consistent with one another.

Clinical Applications

The results of this work are relevant to a few aspects of the development of power reflectance measurements for the clinic. First, it could be helpful to understand typical differences and similarities between the left and right ears of a given subject. For example, if the power reflectances from the left and right ears are similar, then differences in the power reflectance between the two ears of a given subject might provide useful clinical information when interpreting power reflectance measurements. Second, the results here suggest that the probe insertion depth should be as deep as possible to minimize variability in intrasubject power reflectance measurements. This deep insertion depth could both reduce acoustic leaks and reduce the effects of low-frequency ear-canal losses.

Acknowledgments. The authors thank Erika Miguel for assistance in data collection related to this work as a STRIDE scholar at Smith College.

REFERENCES

Keefe DH, Bulen JC, Arehart KH, Burns EM. (1993) Ear-canal impedance and reflection coefficient in human infants and adults. *J Acoust Soc Am* 94(5):2617–2638.

Keefe DH, Folsom RC, Gorga MP, Vohr BR, Bulen JC, Norton SJ. (2000) Identification of neonatal hearing impairment: ear-canal measurements of acoustic admittance and reflectance in neonates. *Ear Hear* 21(5):443–461.

Keefe DH, Levi EC. (1996) Maturation of the middle and external ears: acoustic power-based responses and reflectance tympanometry. *Ear Hear* 17(5):361–373.

Laird NM, Ware JH. (1982) Random-effects models for longitudinal data. *Biometrics* 38(4):963–974.

Rosowski JJ, Nakajima HH, Hamade MA, et al. (2012) Ear-canal reflectance, umbo velocity, and tympanometry in normal-hearing adults. *Ear Hear* 33(1):19–34.

Shahnaz N, Bork K. (2006) Wideband reflectance norms for Caucasian and Chinese young adults. *Ear Hear* 27(6):774–788.

Shahnaz N, Bork K, Polka L, Longridge N, Bell D, Westerberg BD. (2009) Energy reflectance and tympanometry in normal and otosclerotic ears. *Ear Hear* 30(2):219–233.

Stinson MR, Shaw EA, Lawton BW. (1982) Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal. *J Acoust Soc Am* 72(3): 766–773.

Vander Werff KR, Prieve BA, Georgantas LM. (2007) Test-retest reliability of wideband reflectance measures in infants under screening and diagnostic test conditions. *Ear Hear* 28(5):669–681.

Voss SE, Allen JB. (1994) Measurement of acoustic impedance and reflectance in the human ear canal. *J Acoust Soc Am* 95(1): 372–384.

Voss SE, Horton NJ, Woodbury RR, Sheffield KN. (2008) Sources of variability in reflectance measurements on normal cadaver ears. *Ear Hear* 29(4):651–665.

Werner LA, Levi EC, Keefe DH. (2010) Ear-canal wideband acoustic transfer functions of adults and two- to nine-month-old infants. *Ear Hear* 31(5):587–598.