Intrasubject Test–Retest Reliability in Tympanic Electrocochleography

Debra L. Park*
John A. Ferraro*

Abstract

To help establish the clinical reliability of electrocochleography (ECochG), we defined normal limits of variability for the summating potential (SP)/auditory nerve action potential (AP) amplitude ratio across multiple recording sessions from the same subjects. ECochG was recorded bilaterally from the tympanic membranes of 10 normal-hearing adults at six sessions separated by 2- to 14-day intervals. The measures of variability examined included overall range, overall average standard deviation, and mean average difference. Examiner improvement over the six sessions also was examined. Results from the variability data indicate that the reliability of tympanic ECochG can be determined by how many of the measures of variability fall within the normal limits established in this study. No significant examiner improvement was seen between the first three and last three recording sessions.

Key Words: Electrocochleography, reliability, variability

Abbreviations: ABR = auditory brainstem response, AP = action potential, dB SL = decibels referenced to sensation level, ECochG = electrocochleography, ECochGm = electrocochleogram, MD/ELH = Meniere's disease/endolymphatic hydrops, N1 = first peak of the AP, SP = summating potential, TM = tympanic membrane

The importance of electrocochleography (ECochG) as a tool in the diagnosis, assessment, and monitoring of Meniere's disease/endolymphatic hydrops (MD/ELH) is well documented (Coats, 1981; Kumagami et al, 1982; Ferraro et al, 1985; Staller, 1986; Ruth, 1993). Conventional interpretation of the electrocochleogram (ECochGm) for this purpose generally involves measurement of the cochlear summating potential (SP) and auditory nerve action potential (AP) components to derive the SP/AP amplitude ratio. Although the cause remains unclear, an enlarged SP/AP amplitude ratio is considered to be a positive finding for ELH regardless of the recording approach used to obtain the ECochGm (i.e., extratympanic/tympanic, transtympanic) (Cullen et al, 1972; Ruth et al, 1988; Ruth and Lambert, 1989). Given the fluctuating nature of the symptoms associated with MD/ELH, it is not unusual for patients to undergo successive ECochG recordings during the course of their medical/surgical management. To contribute meaningful information under these conditions, ECochG must prove to be a reliable test. Other tests of auditory electrophysiology have been tested for reliability. Auditory brainstem response testing using different recording configurations was shown to have good inter- and intrajudge reliability (Pratt et al, 1995). However, information from the literature regarding the intrasubject test–retest reliability of ECochG for the different recording approaches currently being used is limited. In the two studies we found, Mori et al (1981) reported excellent intertest reliability using a silver-ball, ear canal electrode, whereas Roland et al (1993) observed a high degree of intrasubject variability with the gold-foil "TIPtrode" electrode. To our knowledge, no data are available regarding the test–retest reliability of ECochG recorded from the tympanic membrane (TM), the approach favored in our laboratory (see Ferraro and Ferguson, 1989).

To help validate the clinical utility of tympanic ECochG, the present study examined intrasubject test–retest reliability in normal-hearing subjects over multiple recording
Our intent was to provide normative data from which clinical assessments could be based.

**METHOD**

Tympanic ECochG was recorded bilaterally from 10 adult subjects with normal hearing and no recent history of otologic pathology. To ensure normalcy prior to each ECochG session, bilateral behavioral hearing tests were given in which thresholds were equal to or better than 25 dB HL for all audiometric frequencies and bilateral tympanograms were measured to screen for abnormal middle ear function.

The TM electrode ("tymptrode") employed was a modification of the Stypulkowski and Staller (1987) electrode described in detail by Ferraro and Ferguson (1989). Specific information regarding construction and placement of the tymptrode can be found in Ferraro (1997). A new tymptrode was constructed prior to each session. Placement of the electrode was confirmed using otoscopic examination and EEG monitoring. The tymptrode was then held in place by the foam plug of the insert transducer (EAR-Tone 3A), which delivered the acoustic stimulus.

Electrode configuration was TM(+-) to contralateral earlobe(-) with ground at the forehead. Acoustic stimuli were broad-band clicks, provided by 100-μsec electrical pulses, presented in alternating polarity. Stimulus repetition rate was 11.3/sec. Intensity was 80 dB SL, referenced to behavioral thresholds to the clicks as measured at each session.

All recordings were made using a Nicolet Pathfinder II auditory evoked potential unit. Sensitivity of the analog-to-digital converter was ±50 μV with a filter bandpass of 5 to 3000 Hz (6 dB/octave slope). A 10-msec window was used to average 1000 samples.

Both ears of each subject were tested six times. The session intervals were between 2 and 14 days. To avoid bias, the examiner did not review any previous test results from the study until all of the data were collected. When two repeatable waveforms were obtained, they were averaged into one waveform, which was used for measurement. The two repeatable waveforms were always obtained within two or three runs. Amplitudes of the SP and AP were defined as the number of microvolts between the positive peak and the negative trough (Fig. 1). Measurements were verified blindly by an independent observer with ECochG expertise (JAF).

Data were collected for the components commonly assessed during clinical ECochG: AP amplitude, SP amplitude, N1 latency, and SP/AP amplitude ratio. For this manuscript, our results focus on the SP/AP amplitude ratio, which is the most clinically popular measurement for diagnosis of MD/ELH.

The three measures of variability of the SP/AP amplitude ratio were calculated as follows: (1) "overall mean"—the mean was calculated across the sessions for each subject. Then an average of the individual means was computed for overall mean; (2) "overall average standard deviation"—the standard deviation was calculated across the sessions for each subject. The mean of these individual standard deviations was computed as the overall average standard deviation; (3) "mean average difference"—(as described by Roland et al, 1993)—data for left and right ears were not collapsed. The minimum was subtracted from every other data entry across the session for each subject. These differences were then averaged to achieve the average difference for that subject's ear. The mean of these average differences was calculated across subjects as the mean average difference.

To determine if examiner improvement (effect of practice) occurred during the six sessions, the data for each parameter were divided between sessions 1 to 3 and sessions 4 to 6. These data were compared for significant differences using t-tests.
RESULTS

Figure 1 displays an example of one subject's tracings across sessions. The averaged waveform from each session is shown. As can be seen, the waveforms appear consistent among the six different sessions.

SP/AP amplitude ratios from both ears were measured. A two-factor (ear, session) repeated-measures analysis of variance was performed for the dependent variable, the SP/AP ratio. Results indicated no significant main effects of ear or session and no significant interactions between factors ($F[1, 5] = 1.024, p > .05$). Therefore, data for right and left ears were collapsed for each subject.

Of the 10 subjects tested, one subject's data are excluded from all analyses. This decision was based on one SP/AP ratio from the subject that fell more than three standard deviations outside of the mean. While the abnormal ratio occurred only once in the left ear, it may be indicative of pathology. However, this subject was asymptomatic at the time of testing and continues to be asymptomatic. Irrespective of this subject's possible pathology, including these data would skew the entire database inappropriately.

Examiner Improvement

Three aspects of the SP/AP ratio data were analyzed to assess examiner improvement: overall mean, overall average standard deviation, and average range as displayed in Figure 2. As can be seen, data from the first three sessions did not differ significantly from the last three sessions for any of the three aspects examined ($F[8, 8] = 1.07, p > .05$). Thus, using these measures, no examiner improvement was observed across the six sessions in administration of tympanic ECochG.

Within-Subject Reliability

The next analysis concerns defining the amount of variability tympanic ECochG exhibits within an individual over time. Usually, standard deviations are used when examining variability. However, standard deviations would not be the most accurate measure of variability if the database is not normally distributed. A Kolmogorov-Smirnov test with Lilliefors correction was used to analyze the SP/AP ratio distribution. The result was significant ($p < .05$), indicating the distribution was not normal. Therefore, variability was analyzed by examining both overall average standard deviations and average ranges. A third measure of variability, mean average difference, also was analyzed to make direct comparisons with variabilities reported by Roland et al (1993).

Figure 3 displays the data for the three measures of variability. The choice of the best of these measures of the variability in a given subject's responses is somewhat arbitrary. As expected and seen in Figure 3, the average range covers the broadest spectrum of datapoints. The average range from the current study is 0.36. These data suggest that a patient's SP/AP ratio, for example, may be 0.05 on one visit and 0.41 the next. The overall average standard deviation (0.10) and the mean average difference (0.12)
cover considerably fewer datapoints than average range and are negligibly different from each other. In a clinical setting, these data would be derived from the average of values obtained at several visits. No statistical analyses were performed on these data. Instead, these descriptive data provide examples of standards to which subjects can be compared for normal variability of tympanic ECochG from one test date to the next.

**Cumulative Distribution**

Figure 4 displays a cumulative distribution curve constructed of SP/AP ratios to further apply the measures of intrasubject variability. The curve represents a sample of normal (non-pathologic) ECochGms from which generalizations can be made about the population of subjects with normal ECochGms.

The normal cumulative distribution curve was computed from a percentile histogram. The data used to compute this curve include 75 subjects, 9 of whom are from the present study. The mean SP/AP ratio across sessions was used to represent each of these nine subjects' normal ratio. The remaining 66 datapoints were obtained from previous studies conducted in our laboratory in which the SP/AP ratios were obtained using similar methods and the same type of electrode as employed in the present study.

From the cumulative distribution curve, percentiles can be used to determine amplitude ratio values below which a given proportion of individuals will fall. To obtain a derived SP/AP ratio for a given percentile, the cumulative distribution was best fit with a line that is defined by the equation shown in Figure 4. In this equation, X represents the SP/AP ratio and Y represents the percentile in the cumulative probability curve. Thus, for a given percentile, the upper limit for a normal SP/AP ratio can be determined by solving for X. Using this cumulative distribution curve, the 95th percentile yields an SP/AP ratio of 0.43. Table 1 shows the upper limit for various SP/AP percentiles. These data can be used to compare future ECochG results with a sample of normal subjects.

### DISCUSSION

Tympanic ECochG has been reported as an effective method of clinically recording an ECochGm (Cullen et al, 1972; Ruth and Lambert, 1989). When compared to tracings using a TIPtrode, tracings using a tymptrode yield larger amplitudes (Ruth et al, 1988). The larger the amplitudes of the ECochG components, the eas-
ier it is for the clinician to interpret the waveform. Regardless of the method used, however, reliability of ECochG must be established.

The term “reliability” is defined in our study as the amount of variability across sessions within a subject. Clinically, this variability should aid in deciding whether an ECochGm represents pathology or the normal range of change for a patient.

**Examiner Improvement**

Before within-subject variability was addressed, the effect of practice was examined. We found no significant difference between the first three and the last three test sessions for the SP/AP ratio. In other words, the examiner’s ability to record ECochG did not change over the course of six sessions.

The lack of “effect of practice” is interesting. The amount of experience performing ECochG prior to data collection is an obvious factor in any effect of practice. If a clinician has 10 years of experience performing ECochG frequently, an effect of practice over six sessions would not be expected. However, if the clinician has little experience, an effect of practice may occur. The researcher (DLP) collecting the data for this experiment had performed approximately 16 tympanic ECochGms in a 6-month period prior to collecting the data for the present study. Approximately 4 months separated the last “practice” ECochGm and the first ECochGm for this study.

The learning curve for electrocochleography using the tymptrode has not been established. Given the researcher’s amount of experience, some conclusions can be reached concerning tympanic ECochG’s effectiveness. Namely, the results of the examiner improvement section of this study indicate that audiologists do not need extensive experience to obtain consistent results. With proper training, the skills to perform tympanic ECochG can be acquired in a relatively short period. This should encourage more audiologists to perform tympanic ECochG clinically.

**Intrasubject Variability**

Variability was assessed using three measurements: (1) overall average standard deviation, (2) average range, and (3) mean average difference.

The mean average difference was calculated so that direct comparisons of the present data could be made with Roland et al (1993). Roland et al, using the TIPtrode electrode, report large variations in the SP/AP ratio even though the values fall within normal limits. Roland et al’s mean average difference for the SP/AP ratio is four times smaller than our mean average difference using the tymptrode. Roland et al’s average range (reported as “mean greatest difference”) is 2.4 times smaller than our average range. These researchers report their average range (0.15) is indicative of “large variations” in the SP/AP ratio of normal subjects.

Using the tymptrode results in larger variations in the SP/AP ratio than using the TIPtrode. It is suspected that electrode placement is the cause for this difference. A TIPtrode covers more surface area than a tymptrode, which likely leads to less variable recordings. However, amplitudes are clearly larger using the tymptrode. We consider large amplitudes to be a greater advantage to the clinician than the smaller variable differences that the TIPtrode produces.

The other two methods of calculating variability can be used in the following manner. It is clinically popular to use the 95th percentile (approximately two standard deviations above the mean) as a cut-off for normalcy. Using this premise, we will apply the measure of variability, overall average standard deviation.

Suppose a patient who is suspect for ELH comes into a clinic for an ECochGm on Monday. The patient is not experiencing symptoms at the time of testing. The electrocochleogram reveals an SP/AP ratio of 0.20, which falls within normal limits. Since the patient is suspected to be pathologic, a second examination is scheduled. On Friday, the patient returns and is experiencing symptoms. The SP/AP ratio is 0.55. The data from this study can be used to determine if this change is indicative of pathology or of normal variability for an individual. For example, the patient’s SP/AP ratio varied 0.35. Using the average range (0.36), the patient’s change is considered within normal variability. However, using the overall average standard deviation (±2 standard deviations) of 0.20, the patient’s highest ratio could only have been 0.40 to remain within normal variability. Since this creates a discrepancy, a third element is considered: the normal cumulative distribution curve (see Table 1). According to this curve, 95 percent of the population will have a normal SP/AP ratio of equal to or less than 0.43. Since the patient’s ratio is 0.55, this third element also defines him/her as pathologic. Therefore, of the three elements
considered, two indicate pathology for this patient.

As with any diagnostic test in audiology, the entire picture needs to be considered. In our example, two of the three elements indicated pathology. The patient also reported symptoms of pathology. It is advisable to incorporate all of these factors when interpreting the ECochGm. We agree with Roland et al. (1993) and Ferraro et al. (1985) in noting that it is also advisable to repeat testing on different occasions for a patient with labile symptoms.

Our findings indicate that tympanic ECochG is a reliable test that can be used in the assessment and reassessment of normal-hearing subjects' SP/AP amplitude ratios. Clinically, these data can be further applied to patients suspected of having MD or ELH during diagnosis and throughout treatment.

Acknowledgment. We would like to acknowledge Dr. Mark Chertoff and Dr. Dianne Durham for statistical analysis suggestions and assistance in interpretation of the data.

REFERENCES


