Estimating Audiometric Thresholds Using Auditory Steady-State Responses

Terence W. Picton*  
Andrew Dimitrijevic*  
Maria-Cecilia Perez-Abalo†  
Patricia Van Roon*

Abstract

Human auditory steady-state responses (ASSRs) were recorded using stimulus rates of 78–95 Hz in normal young subjects, in elderly subjects with relatively normal hearing, and in elderly subjects with sensorineural hearing impairment. Amplitude-intensity functions calculated relative to actual sensory thresholds (sensation level or SL) showed that amplitudes increased as stimulus intensity increased. In the hearing-impaired subjects this increase was more rapid at intensities just above threshold ("electrophysiological recruitment") than at higher intensities where the increase was similar to that seen in normal subjects. The thresholds in dB SL for recognizing an ASSR and the intersubject variability of these thresholds decreased with increasing recording time and were lower in the hearing impaired compared to the normal subjects. After 9.8 minutes of recording, the average ASSR thresholds (and standard deviations) were 12.6 ± 8.7 in the normal subjects, 12.4 ± 11.9 dB in the normal elderly, and 3.6 ± 13.5 dB SL in the hearing-impaired subjects.

Key Words: Auditory steady-state responses, objective audiometry, physiological thresholds, signal-to-noise ratio

Abbreviations: AD = analog-digital; ASSR = auditory steady state response; DA = digital-analog; EEG = electroencephalogram; MASTER = multiple auditory steady-state response; PTA = pure tone average (for 500, 1000, 2000, and 4000 Hz); SL = Sensation Level; SNR = signal-to-noise ratio

Sumario

Se registraron respuestas auditivas de estado estable (ASSR) en humanos, usando una tasa de estímulos de 78-95 Hz en sujetos jóvenes con audición normal, en sujetos ancianos con una audición relativamente normal, y en sujetos ancianos con un trastorno auditivo sensorineural. Las funciones de amplitud-intensidad calculadas en relación con los umbrales auditivos reales (nivel de sensación o SL) mostraron que las amplitudes aumentan conforme se incrementa la intensidad. En los sujetos hipoacúsicos este incremento fue más rápido en intensidades justo por encima del umbral ("reclutamiento electrofisiológico") que a intensidades mayores, donde el incremento era similar al visto en los sujetos normales. Los umbrales en dB SL para reconocer una ASSR, y la variabilidad inter-sujetos de estos umbrales, disminuyó conforme aumentó el tiempo de registro y fue menor en los sujetos hipoacúsicos, comparada con los normales. Luego de 9.8 minutos de registro, los umbrales promedio de las ASSR (y sus desviaciones estándar) fue de 12.6 ± 8.7 en los sujetos normales, 12.4 ± 11.9 dB en los ancianos normales y 3.6 ± 13.5 dB SL en los sujetos hipoacúsicos.
Recording human auditory steady-state responses at different intensities can provide an objective assessment of audiometric thresholds (reviewed in Picton et al., 2003). Present practice records the responses to tones that are amplitude modulated, or amplitude and frequency modulated, at rates of 70–110 Hz. At these rates the responses are not attenuated by sleep (Cohen et al., 1991; Picton et al., 2003) and can be recorded in young infants (Rickards et al., 1994; Lins et al., 1996; Savio et al., 2001; Cone-Wesson et al., 2002a, 2002b, 2002c; Luts et al., 2004).

The accuracy with which behavioral thresholds can be estimated from the physiological responses depends on how the responses are recorded. The simplest technique is to estimate behavioral thresholds as equivalent to the threshold for recognizing the physiological response as significantly different from noise. In general such estimated thresholds will be higher than the actual behavioral thresholds by amounts that vary between 5 and 40 dB (Rickards et al., 1994; Lins et al., 1996; Herdman and Stapells, 2001, 2003; Perez-Abalo et al., 2001; Dimitrijevic et al., 2002; Rance and Briggs, 2002). The difference between estimated and actual thresholds varies with several parameters, among the most important of which are the frequency of the sound, the degree of hearing loss, the age of the subjects, and the duration of the recording.

Our contention is that these parameters basically depend on the signal-to-noise ratio (SNR) of the recording. The amplitude of a response varies with the carrier frequency, responses at frequencies 1000 and 2000 Hz being larger than responses to lower and higher audiometric frequencies (see John et al., 2001b). Physiological thresholds are closer to the behavioral thresholds for patients with a sensorineural hearing loss than for normal subjects (e.g., Dimitrijevic et al., 2002; Rance and Briggs, 2002). When a sensorineural hearing loss is present, the amplitude of the response likely increases more rapidly than normal in the first 10–20 dB above threshold, making the response easier to recognize near threshold. This presumed amplitude-intensity function has not been clearly demonstrated, although Dimitrijevic et al. (2002) reported that for stimuli of equal hearing level (HL) the response in patients with sensorineural hearing loss is larger than in normal subjects.

Age can affect the amplitude of both the response and the noise. In infants and in elderly subjects the noise levels are often higher than in young adults, and in infants the responses are smaller (Lins et al., 1996; Savio et al., 2001; John et al., 2004). When the duration of the recording is longer, the residual noise of the electroencephalogram (EEG) is less (John and Picton, 2000; Luts et al., 2004), making small near-threshold responses easier to recognize.

In order to evaluate this idea, we recorded ASSRs (auditory steady-state responses) in normal subjects and in hearing-impaired subjects at different intensities above behavioral threshold and for different durations. Our hypotheses were that the physiological thresholds would be closer to behavioral thresholds and less variable for the frequencies 1000 and 2000 Hz compared to higher or lower frequencies, for the hearing-impaired subjects compared to the normal subjects, and for longer compared to shorter durations of recording. In the hearing-impaired subjects, we recorded at 5 dB steps in order to delineate any abnormality of the amplitude-intensity function.

METHODS

Subjects

We examined three separate groups of subjects. Ten normal young subjects (mean age 25 years, range 19–31, 5 male) were recruited from laboratory personnel. All had thresholds equal to or less than 20 dB HL at
their pure-tone averages (PTA) across the frequencies 500, 1000, 2000, and 4000 Hz averaged 3 with a range of 1 to 9 dB. Ten hearing-impaired subjects (mean age 78 years, range 64–86, 4 male) replied to a notice requesting subjects posted in the audiology clinic at Sunnybrook and Woman’s College Health Sciences Centre, or volunteered through our subject pool. These subjects were selected to have a moderate or severe sensorineural hearing loss. Their average PTA was 51 with a range of 34–74 dB. Since these subjects were older than our normal subjects, we also examined a group of 10 elderly subjects (mean age 69 years, range 61–77, 2 male) from our volunteer pool. These subjects had normal hearing or a mild high-frequency sensorineural loss (untreated with hearing aids). Their average PTA was 15 with a range of 2 to 36 dB. Table 1 shows the average behavioral thresholds at each audiometric frequency for each group of subjects.

### Stimuli

Stimuli were created and presented using the multiple auditory steady-state response (MASTER) system (Lins and Picton, 1995; John et al, 1998; John and Picton, 2000; see also www.mastersystem.ca). Eight stimuli were presented simultaneously, four to each ear. Carrier frequencies were 500, 1000, 2000, and 4000 Hz in each ear. The stimuli were modulated using exponential modulation with an exponent of 2 (John et al, 2002a). For the main experiments, the stimuli were modulated at frequencies between 78 and 95 Hz, with each carrier frequency associated with a unique modulation frequency. The normal young subjects were also tested using modulation frequencies between 110 and 125 Hz. (The rationale of this part of the experiment was to determine whether the 78–95 Hz rates were indeed optimal or whether responses might be more readily recognized at faster rates where the noise levels were lower). Table 2 shows the exact modulation frequencies. The modulation frequencies were intercalated so that the same carrier frequencies in each ear were adjacent and increased with increasing carrier frequency beginning with the right ear. The stimuli were digital-to-analog (DA) converted at a rate of 32 kHz and routed to a Grason Stadler Model 16 audiometer where they were amplified to a calibration intensity, attenuated to achieve the desired intensity levels, and presented using Etymotic-3A insert earphones. Although the MASTER system simultaneously presented four stimuli to each ear, each stimulus was separately calibrated (accuracy to within 2 dB) using a Bruel and Kjaer DB 0138 coupler to equivalent intensity measured in root-mean-square sound pressure level (SPL). The combined stimulus was 7 dB more intense than the individual stimuli.

In the normal subjects (both young and elderly) the stimuli were presented at levels of 60 dB SPL and decreased in intensity by 10 dB steps until no responses were recognized as significant or until 10 dB SPL. Normal subjects were examined on two separate days, one for the 78–95 Hz modulation rates and one for the 110–125 Hz rates. In the hearing-impaired subjects, the intensity began at 80 dB SPL and decreased

### Table 1. Behavioral Thresholds (mean and range in dB HL)

<table>
<thead>
<tr>
<th>Group</th>
<th>Frequency (Hz)</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young</td>
<td>500</td>
<td>6(-5 to 20)</td>
<td>4(-5 to 15)</td>
<td>4(0 to 10)</td>
<td>-2(-10 to 5)</td>
</tr>
<tr>
<td>Elderly</td>
<td>1000</td>
<td>15(5 to 20)</td>
<td>11(0 to 30)</td>
<td>12(0 to 30)</td>
<td>22(0 to 55)</td>
</tr>
<tr>
<td>Hearing Impaired</td>
<td>2000</td>
<td>39(15 to 65)</td>
<td>48(5 to 65)</td>
<td>57(35 to 80)</td>
<td>60(35 to 95)</td>
</tr>
<tr>
<td>4000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Stimulus Modulation Frequencies (Hz)

<table>
<thead>
<tr>
<th>Ear</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier</td>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td>78–95 Hz</td>
<td>80.1</td>
<td>85.0</td>
</tr>
<tr>
<td>108–125 Hz</td>
<td>110.4</td>
<td>115.2</td>
</tr>
</tbody>
</table>
in 5 dB steps. This required two separate recording sessions.

**Recordings**

Auditory steady-state responses were recorded from an electrode placed at the vertex, using a neck electrode as a reference and an electrode placed on the left clavicle as ground. Subjects slept or drowsed in a reclining chair, located within a darkened sound-attenuated chamber during the recordings. The EEG was collected using a filter bandpass of 1 to 300 Hz, a gain of 50,000, and an analog-digital (AD) rate of 1000 Hz. The EEG was continuously monitored to ensure that the subject was relaxed, that is, showing little fast muscle activity. Individual EEG epochs of 1024 points each (1.024 seconds) were rejected if they contained any value that exceeded ±80 µV. Sixteen individual data epochs of 1024 points each were collected and linked together into sweeps lasting 16.384 seconds. As each sweep was completed it was combined with a running average sweep using weighted averaging based on the activity in the signal between 70 and 110 Hz (John et al, 2001a). The averaged sweep was analyzed using a Fast Fourier transform (FFT). The amplitude of the steady-state response to a given carrier frequency was equal to the amplitude at the frequency of modulation in the resulting spectrum. This amplitude was compared to the amplitudes in adjacent regions of the spectrum (60 frequencies above and 60 frequencies below the response frequency) using an F-ratio with 2 and 240 degrees of freedom (Lins et al, 1995). A response was judged as significantly different from residual EEG noise at a criterion of p < 0.05. For each experimental condition of our experiments, 36 sweeps, or approximately 10 minutes of data was collected. All data were stored to disk and then analyzed offline after 1, 3, 6, 12, 24, and 36 sweeps.

**Analyses**

Two decisions were necessary before evaluating the amplitude-intensity functions. The first decision was whether to base the intensity on the normal thresholds for each carrier frequency (hearing level or HL) or on each subject’s own threshold (sensation level or SL). Since our intent was to show the response increased more rapidly from threshold in subjects with a hearing impairment, we decided to use SL. In order to convert our intensities from SPL to SL, we used the SPL of the stimulus minus the SPL of the tone used for the audiometric evaluation. In order to allow us to combine data across subjects these differences were rounded to the nearest 5 dB. However, in the normal subjects, where the intensities were changed in 10 dB steps, this would give missing data at many intensities. This approach gave informative graphs about the effect of hearing loss on the responses but made it impossible to conduct statistical analyses. We therefore used SPL to measure the effects of ear, intensity, frequency and age. This analysis used a four way ANOVA with repeated measures within the factors ear, intensity and frequency. A second analysis compared the young and elderly with the elderly shifted 10 dB to compensate for the average change in the PTA. In this analysis the response of the elderly subjects at 60 dB SPL was compared to the response of the young subjects at 50 dB SPL, and this was repeated down to 20 dB in the elderly compared to 10 dB in the young. A third analysis compared the responses in the young subjects across the two different modulation frequencies (78–95 versus 108–125 Hz). Geisser Greenhouse criteria were used when appropriate to assess the repeated measures effects.

The second decision was how to assess amplitudes when the response was not significant. One way is just to use whatever amplitude was measured. In this approach the amplitude decreases with decreasing intensity and then flattens off at the noise level of the recording for responses that are below the physiological threshold. The second approach is to consider the amplitude of a nonsignificant response as zero. This may be inappropriate when the noise level is particularly high for one recording but is generally still preferable to taking the noise level as the amplitude of the response. We opted for the second approach.

The noise level of the recordings was determined as the mean amplitude of the response in the ±60 bins of the spectrum adjacent to the response bin (±3.7 Hz). An ANOVA compared groups with repeated measures for carrier-frequency, ear, and intensity. Since there were incomplete data
at the lower intensities for the hearing-impaired subjects, this only involved levels of 50, 60 and 70 dB SPL.

Physiological thresholds were determined using two rules. The first was that the threshold was the lowest intensity at which a significant response was detected, provided that the response was insignificant at no more than one higher intensity. The second was that if no threshold was determined according to the first rule then the threshold was set at 10 dB above the highest intensity used to elicit the response. Since the step size for intensity in the hearing-impaired patients was smaller than in the other subjects, we calculated the thresholds in these subjects using 10 dB step sizes (either 5-15-25 ... or 10-20-30 ...) when making comparisons across groups. The residual EEG-noise levels in the recordings were lower for the young subjects compared to the other two subject groups, and this could alter any comparison of threshold estimation. We therefore decided to compare physiological thresholds for the hearing-impaired subjects with those of the normal elderly. The difference between physiological and behavioral thresholds were thus evaluated using an ANOVA with two subject groups (normal elderly, hearing impaired) and repeated measures for different times of recording (or number of sweeps averaged), ear, and carrier frequencies. A further ANOVA compared the thresholds for the different modulation frequencies in the young subjects.

There is no readily available method to test for changes in intersubject variability in a multifactorial ANOVA. In order to get some sense of changes in this variability, we selected data for the responses at 2000 Hz, combined data across the two ears, and estimated Levene’s statistic to evaluate whether there was a significant change in variance between the data obtained after averaging 6 sweeps (1.6 minutes) and after 36 sweeps (9.8 minutes). This analysis was performed for selected intensities for the amplitude and noise measurements and for the estimated physiological thresholds.

**RESULTS**

**Amplitude-Intensity Functions**

Typical responses are shown in Figures 1 and 2. Figure 1 shows ASSRs recorded at intensities ranging from 20 to 70 dB SPL from a single young subject at two different modulation bands, 78–95 Hz, and 108–125 Hz after the full recording time (36 sweeps, 10 minutes) at each intensity. Figure 2 shows ASSRs recorded at intensities ranging from 20 to 70 dB SPL from a single elderly subject with normal hearing sensitivity after 6 sweeps (left side) and after 36 sweeps (right side). The mean amplitudes for the responses across the groups are shown in Figure 3. The initial ANOVA (comparing responses in the young and elderly groups at the same SPL—the data on the left side of Figure 3) showed a main effect of group (F = 9.0; df = 1,18; p < 0.01) and intensity (F = 111.3; df = 5,90; p < 0.001) and a group by intensity interaction (F = 4.3; df = 5,90; p < 0.01). The amplitude was larger in the young than the elderly subjects, and for higher intensity compared to lower intensity, with the interaction being caused by the group effect being larger at higher intensities. The 500 Hz response was smaller than the others, and this effect was more prominent at low intensity (frequency and intensity interaction F = 7.4; df = 15,270; p < 0.001). Since we found no effects of ear or interactions with ear, the responses were collapsed across ears in Figure 3. When the responses in the young were compared with the elderly response to stimuli 10 dB lower, the group effect and group interaction with intensity was no longer significant, but the frequency effect and frequency by intensity interaction remained. In the young subjects, the amplitudes of the responses were much smaller at 108–125 Hz compared to 78–95 Hz (graphs in the upper half of Figure 3) (F = 26.0; df = 1,9; p < 0.001), especially at higher intensity (intensity by rate interaction F = 11.3, df = 5,45; p < 0.01).

**Results in Hearing-Impaired Subjects**

Responses for two different hearing-impaired subjects are shown in Figure 4. These responses were those obtained after averaging 36 sweeps. The average amplitudes in the hearing-impaired subjects are shown in the bottom right of Figure 3. The main findings were that the responses were smaller at lower intensities, which were subthreshold, and that this difference was greater for the higher carrier frequencies, where the thresholds were higher. At the higher intensities where the
Figure 1. ASSRs in a normal young subject. On the left are plotted the responses to the ASSRs using stimuli modulated in the frequency range of 78–95 Hz, and on the right are the responses for 108–125 Hz. The responses were averaged over 36 sweeps. Dotted lines separate the results at the different carrier frequencies. The responses identified as significant are indicated by the triangles. The filled triangles indicate the responses for right ear stimuli, and the open triangles are for left ear stimuli. The carrier frequencies increase with increasing modulation frequency (Table 2), and thus the identified responses beginning on the left are to 500 Hz in the right ear, 500 Hz in the left ear, 1000 Hz in the right ear, etc. The responses at 78 and 80 Hz (for 500 Hz carriers in the right and left ears, respectively) are not recognized at 40 dB SPL but then become recognizable at 30 dB. According to our rules, the thresholds would have been 30 dB SPL. The responses at the faster modulation frequencies are smaller. However, since the noise level is also smaller, the responses are recognized down to similar thresholds.

Figure 2. Effects of averaging. This shows the response in a normal elderly subject, as shown after 6 sweeps and 36 sweeps were averaged. The residual EEG-noise levels decrease with increased averaging. This makes it easier to recognize responses at low intensities, thus leading to lower physiological thresholds. Physiological threshold levels (determined according to the rules described in the text) are shown with asterisks. The responses indicated with “?” are likely false positives. The response at 1000 Hz in the left ear at 20 dB SPL is likely a false positive but according to our rules is judged suprathreshold. Even granting this value, the physiological thresholds are on average 15 dB lower after averaging more sweeps.
stimuli were above threshold, however, the amplitudes of the responses were similar to those of the normal elderly (e.g., compare data at 80 dB SPL in the hearing impaired to the data at 70 dB SPL in the elderly).

**Sensation Levels**

In order to compare the data across the different subjects, we readjusted the intensities to SL by subtracting the behavioral thresholds from the actual physical intensities. Sample data from individual subjects (one normal hearing and one hearing impaired) are plotted in Figure 5. The mean data (for each carrier frequency and for each of the three subject groups) are plotted in Figure 6. The graph in the upper right of Figure 6 collapses the data across carrier frequencies and plots all three groups together. Since the behavioral thresholds were obtained in 5 dB steps and since the ASSRs were only obtained at 10 dB steps in the normal young and elderly subjects, it was not possible to evaluate these data using an ANOVA model. The mean data were plotted only if four or more values contributed to the mean.

**Residual EEG Noise**

As can be seen in Figure 2, the level of the residual noise in the recording—the amplitudes at bins in the spectrum that did not contain responses—decreased with an increase in the number of sweeps averaged. The ANOVA showed a significant main effect of group ($F = 5.5; \text{df} = 2,27; p < 0.01$) with the young subjects having lower levels than the other two groups. The noise levels were significantly lower for lower intensities, for higher numbers of sweeps, for higher carrier frequencies, and for the left ear as opposed to the right ear. Graphs showing the effects of number of sweeps, intensity, and group are shown in Figure 7. The effects of frequency and ear were too small to be easily visible on such graphs. The changes in the variability of the noise from averaging 6 sweeps to averaging 36 sweeps are shown in Table 3 together with the changes in the response amplitude. We present only the data for 2000 Hz at selected intensities.

**Thresholds**

The physiological thresholds decreased with an increase in the number of sweeps averaged. This is illustrated for actual ASSRs in Figure 2. The mean difference between the physiological thresholds and the behavioral thresholds (i.e., the physiological thresholds expressed in dB SL) are graphed in Figure 8, which shows the effects of carrier...
Figure 4. ASSRs in two hearing-impaired subjects. The behavioral thresholds (in dB SPL) are given at the top of the graph. The responses were averaged over 36 sweeps. For the subject on the left of the figure, the physiological threshold at 78 Hz (500 Hz to right ear) is 60 dB, which is 25 dB higher than the behavioral threshold at 35 dB SPL. The physiological threshold at 95 Hz (4000 Hz to right ear) is 60 dB, which is the same as the behavioral threshold. Note that the amplitude scales are different since the responses were much larger in the subject whose ASSRs are plotted on the right. (The general noise levels of the residual EEG were similar for the two subjects. The response at 30 dB to 2000 Hz in the left ear (identified with "?")) is a false positive response.

Figure 5. ASSR amplitudes relative to sensation level (SL). This figure shows the response amplitudes (78–95 Hz) of two subjects (one ear each) in relation to intensity as plotted in dB SL. The young subject on the left has normal hearing (thresholds in dB HL are given at the top of the graph), and the amplitudes increase with increasing intensity at approximately the same rate regardless of intensity or frequency. The subject on the right has a hearing impairment. The amplitudes increase rapidly immediately above threshold and then continue to increase at a lesser rate with increasing intensity.
frequency in the three groups of subjects and the overall effects of subject group in the upper right graph. The ANOVA (performed only using the 10 dB stepped data for hearing impaired) showed a main effect of group (F = 19.7; df = 1, 18; p < 0.001) with the hearing impaired showing lower thresholds than the normal elderly subjects. The within-subject analyses showed the anticipated main effect of number of sweeps (F = 123.2; df = 5, 90; p < 0.001) and an interaction between subject group and number of sweeps (F = 6.9; df = 5, 90; p < 0.01). This interaction was caused by the threshold decrease with increasing number of sweeps being less for the hearing impaired than for the other subjects. A main effect of frequency (F = 16.3; df = 3, 54; p < 0.001) was related mainly to the threshold at 500 Hz being higher than at the other frequencies. A second ANOVA compared the physiological thresholds in the young subjects across the different modulation rates (78–95 versus 108–125 Hz) showed no significant difference between the modulation rates.

Table 4 shows the results of our analysis of the intersubject variability in threshold estimation. The standard deviations of threshold estimations are tabulated after 6

Table 3. Variability of Response Amplitude and Residual EEG Noise (nV)

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Condition</th>
<th>6 Sweeps Mean</th>
<th>6 Sweeps SD</th>
<th>36 Sweeps Mean</th>
<th>36 Sweeps SD</th>
<th>Probability*</th>
<th>Levene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>Young 60 dB SPL</td>
<td>75.6</td>
<td>21.1</td>
<td>72.9</td>
<td>20.6</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Elderly 70 dB SPL</td>
<td>60.5</td>
<td>31.3</td>
<td>58.2</td>
<td>18.1</td>
<td>ns</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>Hearing Impaired 80 dB SPL</td>
<td>57.3</td>
<td>27.6</td>
<td>53.8</td>
<td>23.8</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Noise</td>
<td>Young 60 dB SPL</td>
<td>17.5</td>
<td>4.1</td>
<td>8.1</td>
<td>2.6</td>
<td>&lt;0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Elderly 70 dB SPL</td>
<td>30.6</td>
<td>9.5</td>
<td>11.1</td>
<td>3.3</td>
<td>&lt;0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Hearing Impaired 80 dB SPL</td>
<td>33.3</td>
<td>10.8</td>
<td>12.6</td>
<td>3.7</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* Probability of the null hypothesis. The F-test evaluates the difference between the means, and the Levene test evaluates the difference between the variances. Only values below 0.10 are tabulated.
and 36 sweeps. In addition to the decreasing mean threshold, the intersubject variability of the physiological threshold decreased with an increasing number of sweeps. This was more often significant for the normal-hearing subjects (young and elderly) than for the hearing-impaired subjects.

### DISCUSSION

#### Prefatory Comments

The results clearly demonstrated that the physiological thresholds were further away from behavioral thresholds for 500 Hz carriers than for the 1, 2, or 4 kHz carriers. We had hypothesized that the thresholds for the 4 kHz carrier would also be elevated, but this was not the case. Physiological thresholds

<table>
<thead>
<tr>
<th>Group</th>
<th>Frequency (Hz)</th>
<th>6 Sweeps</th>
<th>36 Sweeps</th>
<th>Probability*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Young</td>
<td>500</td>
<td>34.8</td>
<td>15.9</td>
<td>21.3</td>
</tr>
<tr>
<td>78–95 Hz</td>
<td>1000</td>
<td>16.3</td>
<td>8.3</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>18.5</td>
<td>8.9</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>4000</td>
<td>23.0</td>
<td>15.2</td>
<td>13.0</td>
</tr>
<tr>
<td>Young</td>
<td>500</td>
<td>36.3</td>
<td>13.7</td>
<td>25.8</td>
</tr>
<tr>
<td>108–125 Hz</td>
<td>1000</td>
<td>22.8</td>
<td>13.2</td>
<td>13.8</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>23.0</td>
<td>13.9</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>4000</td>
<td>22.0</td>
<td>10.4</td>
<td>13.5</td>
</tr>
<tr>
<td>Elderly</td>
<td>500</td>
<td>39.8</td>
<td>16.8</td>
<td>21.3</td>
</tr>
<tr>
<td>78–95 Hz</td>
<td>1000</td>
<td>28.8</td>
<td>16.8</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>28.8</td>
<td>14.2</td>
<td>11.3</td>
</tr>
<tr>
<td></td>
<td>4000</td>
<td>33.0</td>
<td>18.7</td>
<td>11.5</td>
</tr>
<tr>
<td>Hearing</td>
<td>500</td>
<td>25.3</td>
<td>19.8</td>
<td>10.8</td>
</tr>
<tr>
<td>Impaired</td>
<td>1000</td>
<td>8.5</td>
<td>17.6</td>
<td>-4.0</td>
</tr>
<tr>
<td>78–95 Hz</td>
<td>2000</td>
<td>16.0</td>
<td>14.6</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>4000</td>
<td>15.3</td>
<td>16.7</td>
<td>5.3</td>
</tr>
</tbody>
</table>

*Probability of the null hypothesis. The F-test evaluates the difference between the means, and the Levene test evaluates the difference between the variances. Only values below 0.10 are tabulated.
were closer to behavioral thresholds for the hearing-impaired subjects. Both the differences between the physiological and the behavioral thresholds and the variability of this measurement decreased as the duration of the recording increased. We shall discuss these findings in the context of objective audiometry.

Accurately estimating behavioral thresholds from physiological measurements is the goal of objective audiometry. The usual approach is to measure the lowest intensity at which a physiological response can be detected and then to use this physiological threshold to estimate the behavioral threshold. The estimation can involve the simple subtraction of the normal mean difference between physiological and behavioral thresholds, or a more complicated regression of the physiological threshold to the behavioral threshold, that is, subtracting a value that varies with the intensity level of the physiological threshold (Rance et al, 1995; Rance and Briggs, 2002).

In both cases, the accuracy of estimating behavioral thresholds depends upon the variability of the physiological threshold from one subject to the next (or from one recording to the next). The mean difference (provided this is not too large) can be subtracted away, but one is left with the variability. A measurement of threshold that averages at 0 dB but has a standard deviation of 20 dB is much less accurate than a measurement that averages at 10 dB but has a standard deviation of 10 dB. The first technique will give individual thresholds that vary from 40 dB below to 40 dB above (the range of ±2 standard deviations that normally accounts for 95% of the values), whereas the second technique, once the mean is subtracted, gives individual thresholds that vary from 20 dB below to 20 dB above. The intersubject variability in threshold estimation depends upon the amplitude and variability of the near-threshold ASSR and the amplitude and variability of the residual EEG noise level of the recording. The noise level of the recording will depend upon the baseline level of EEG and muscle noise in the recording and the

Figure 8. Estimation of behavioral thresholds. These graphs show how increasing the number of sweeps decreases the difference between the physiological threshold and the behavioral threshold, that is, the physiological thresholds expressed in dB SL. The upper right graph shows the mean data across all frequencies for the three subject groups. Included in these data are the results for the young subjects using rates of 108–125 Hz. The two sets of data for the hearing-impaired subjects are derived from 10 dB steps starting at 70 dB or 65 dB.
time spent in reducing that noise (by averaging or by increasing the duration over which the frequency transform is calculated).

Using some suprathreshold measurement to estimate thresholds would have the advantage that physiological responses at these levels are larger and more reliably measured. However, the estimation of thresholds from suprathreshold data is confounded by the possibility of recruitment. For example, using the slope of the amplitude-intensity function to extrapolate back to threshold requires that the slope be predictable at all levels of intensity and threshold. In a sensorineural hearing impairment, the effect of intensity on the amplitude of a physiological response is often nonlinear. Both the amplitudes and the slopes of the amplitude-intensity function can be normal at intensities more than 10 dB above the elevated threshold.

**Signals in Noise**

Whether or not a response can be recognized depends upon its amplitude and the amplitude of the background EEG noise. The amplitude of a response generally increases with increasing intensity (see Figure 3). In subjects with hearing impairment, the response is not present at intensities below threshold but increases more rapidly than normal above threshold until it reaches normal levels (diagrams in the left column of Figure 9, deriving from the mean data shown in Figure 6). Dimitrijevic et al (2002) also showed that the ASSR amplitudes were larger in the hearing impaired than in normal subjects for stimuli of equivalent SL. We shall refer to this phenomenon as "physiological recruitment." The term "recruitment" originally referred to the abnormally rapid growth in perceived loudness rather than the change in the amplitude of a physiological response. Physiological recruitment is likely associated with the cerebral processes that lead to perceptual recruitment. The amplitude of the ASSR at a given intensity varies from subject to subject. The recorded amplitude depends on many parameters, the most important of which are the amount of synchronized current in the generators, the orientation of these generators relative to the recording electrodes, and the impedance of the volume conductor (brain, skull, skin).

At stimulus rates of 78–95 Hz, the amplitudes can span a range of about nine-fold, for example, from 30 to 270 nV at 70 dB SPL in normal young subjects. This variability is greater at 500 Hz than at other frequencies. This variability makes it possible that responses will not be recognized in some subjects even at 70 dB SPL unless the noise levels are reduced significantly below 30 nV. The variability is diagrammatically plotted in Figure 9 using approximately ±1 standard deviations.

The residual EEG noise level in a recording also varies from subject to subject. This variability is for the most part independent of the variability of the ASSR amplitude. It varies mainly with the level of muscle activity in the recording, being much lower in relaxed or sleeping subjects than in tense subjects. We found that the noise levels were significantly higher in older subjects. This is likely related to the fact that the young subjects were able to sleep through the testing whereas the older subjects found it difficult to sleep. In addition, we found that the noise level decreased with decreasing stimulus intensity. This may have been in part related to subjects being less able to sleep when the stimuli were louder. In addition, since our protocol started out at high intensities and then decreased the intensity to determine threshold, lower-intensity stimuli would occur later in the recording session as the subject became more relaxed or fell asleep. The amount of noise in the recorded spectrum decreases with increasing frequency (John and Picton, 2000; Picton et al, 2003). The way we presented our stimuli (Table 2) therefore caused the noise to be lower for stimuli with higher carrier frequencies and for left-ear compared to right-ear stimuli.

The main determinant of the residual EEG-noise level, however, is the duration of the recording. The amplitude of the residual noise decreases by the square root of the number of sweeps (Figure 7). A longer time spent in the recording will give less background EEG noise. This will lead to easier recognition of the responses and greater precision in the threshold estimation. It will also cause less absolute variability in the background noise levels, which will reduce the variability in threshold estimation. Luts et al (2004) have recently stressed the effects of recording time on the estimation of
thresholds in young babies.

Our measurement of the amplitude of the noise is the root-mean-square value of the activity in 120 bins adjacent to the signal bin. We based the Fourier transform on a 16.4 second sweep, which resulted in bin widths of 0.061 Hz. If the sweep had been shorter, the bin width would have been broader, and the noise amplitude measured within the bin would have been higher. However, if the shorter sweep was averaged more (i.e., the time for the recording was the same) prior to the transform, the noise would be reduced in a way similar to increasing the sweep duration (see John et al, 1998, for further discussion).

The measured amplitude of the response is affected by the level of residual EEG noise in the recording. This noise can either add to or subtract from the actual amplitude, but the overall tendency is to cause a larger measurement. The Appendix (together with Figure 10) demonstrates how the amplitude is overestimated by a factor that varies with the ratio of the measured amplitude to the noise. This effect is evident in the results shown in Table 3. As the number of sweeps increases from 6 to 36 the average noise levels of the recording decreases by a factor approximately equal to the square root of 6 (2.45). For example, the noise level in the response of the hearing-impaired subjects decreases from 33.3 to 12.6 nV (an actual factor of 2.64). As the noise level decreases, the measured amplitude also decreases by a small amount. To continue with the example, the amplitude decreases from 57.3 to 53.8 nV. If we use the calculations described in the Appendix, an unbiased estimate of the amplitude would be 51.3 nV after an average of 6 sweeps and 52.5 nV after 36 sweeps. The amplitudes reported in this paper have not been corrected using these procedures. Such corrections would have made the slopes of the amplitude-intensity functions a little steeper at the lower intensities.

The ASSRs recorded at rates of 108–125 Hz in the young subjects were about one-half the size of the responses recorded at rates of 78–95 Hz. However, since the residual EEG noise at these frequencies was also lower, the responses could still be recognized at intensities near threshold. The variability and the level of the threshold estimates were slightly but not significantly higher at the higher rates. We therefore decided not to
pursue these responses in the elderly subjects. Another factor was our recent finding that the decline in the response at higher frequencies is greater in elderly than in young subjects (Purcell et al, 2004).

**Threshold Estimation**

The main goal of objective audiometry is to estimate frequency-specific thresholds in an individual subject. The accuracy of this estimation is measured by the range of values obtained in a group of similar subjects. Provided that the mean difference between the physiological and behavioral thresholds is not too large, that is, less than 30 dB, it can be subtracted away from the physiological threshold to give an estimated behavioral threshold. If the difference is large, errors start to occur because the expected range of hearing loss cannot be fully covered. Also, the variability of threshold estimation often varies with the mean size of the difference between physiological and behavioral thresholds. Once the mean difference is subtracted, the variability of the difference becomes the variability of threshold estimation. The two main determinants of this variability are the variability in the amplitude of the response near threshold and the variability of the noise level. This is illustrated in Figure 9.

Physiological thresholds are measured closer to behavioral thresholds in subjects with sensorineural hearing impairment than in normal subjects (Figure 8 and in Table 4). After 6 sweeps we found the average difference to be 23.2 and 27.6 dB in the young and elderly subjects with normal hearing, respectively, and 16.3 dB in the elderly hearing impaired. After 36 sweeps, the values were 12.6 ± 8.7, 12.4 ± 11.9 dB and 3.6 ± 13.5 dB. The values obtained after 36 sweeps are of the same order as those reported by others: Dimitirijevic et al (2002) reported an average difference of 9.0 dB in normal subjects and 7.8 dB in the hearing impaired (averaged data from their Table 3); Herdman and Stapells (2001; 2003) reported average differences of 11.5 and 8.8 dB. The differences in these papers are less than reported here, perhaps because their responses were recorded over longer periods (discussed in a subsequent paragraph). Perez-Abalo et al (2001) recording over a shorter period (up to 4.5 minutes per intensity) found average differences of 12.0 and 7.5 dB for the normal and the hearing-impaired subjects.

The likely explanation for these differences is physiological recruitment occurring in patients with sensorineural hearing loss. This can be seen in the diagrams in the middle and right columns of Figure 9. The amplitude-intensity functions, which were plotted in dB HL in the left column, are now plotted in dB SL. The approximate level of noise after recording for 6 or 36 sweeps is then plotted together with its standard deviation to show the intersubject variability. The average noise levels have been multiplied by 1.73 to give the p < 0.05 upper limits of the noise beyond which a response would be considered significant (this is the square root of the F value for degrees of freedom 2 and infinity). The point where the mean level of this noise crosses the mean level of the amplitude-intensity function (indicated with arrows) gives the mean estimated threshold. The threshold in dB SL is lower for the hearing-impaired subjects.

The Melbourne group has provided extensive data showing that the regression of the behavioral threshold on the physiological threshold has a slope of greater than one (Rance et al, 1995; Cone-Wesson, 2002a; Rance and Briggs, 2002; Rance and Rickards, 2002). (Note that some of the earlier formulations plot physiological rather than the behavioral threshold on the y-axis and the slope is thus less than 1 rather than greater than 1). The physiological thresholds are close to the same as the behavioral thresholds near 90 dB and about 30 dB higher when the behavioral thresholds are near 0 dB HL giving regression slopes near 1.3.

The regression slopes will, however, depend on the amount of time spent in the recording. If a longer time is spent, the physiological thresholds become closer to the behavioral thresholds and this change is greater for normal-hearing subjects than for hearing-impaired subjects (see Figure 8 and Table 4). These effects are shown diagrammatically in the graphs of the middle and right columns of Figure 9. Some ballpark effects on the regression can also be derived in the data of Table 4. Given the average behavioral thresholds in our hearing-impaired subjects and normal elderly subjects of 51 and 15 dB, the slopes of a regression for our data can be roughly estimated as the difference between these values (36 dB).
divided by the difference in the estimated thresholds—54.7 (51 plus the 3.7 dB, which is the average of the physiological thresholds in our hearing impaired) less 27.9 (15 plus 12.9)—which gives 1.34. For the data obtained after 6 sweeps, the approximate slope of the regression would be 1.45. Another variable that might affect the regression equations is the amount of noise in the recording. We found that the noise levels in normal younger subjects were less than in older subjects. This could move the data for the younger subjects closer to the diagonal, making the slope closer to 1.

Regressions are typically derived from the threshold estimates of subjects with normal hearing and with various degrees of sensorineural hearing loss. The regressions are not appropriate for patients with conductive hearing loss. These patients have no recruitment and therefore will show similar amplitude-intensity functions to normal subjects except for a shift to the right (to higher intensities). The differences between physiological and behavioral thresholds will be similar to those obtained in normal subjects.

Concluding Comments

The accuracy of threshold estimation depends on the variability of threshold estimation rather than on any mean difference between physiological and behavioral thresholds. The results of these experiments demonstrate several ways to improve the accuracy of estimating behavioral thresholds from the ASSRs. The main factor is the level of background EEG noise in the recording. This noise can be reduced by having the subject relax or fall asleep, and by spending a longer time in the recording. However, time is limited in the clinical situation. If we need to assess thresholds within about an hour (e.g., 10–15 minute recordings at 5 or 6 intensity levels), we will typically wind up with thresholds that have a standard deviation of about 10 dB. This is similar to the variability obtained using tone ABRs (Stapells, 2000). This means that some subjects will have thresholds that are 20 dB higher, and some will have thresholds that are 20 dB lower (provided that the mean difference has been subtracted).

To the audiologist fitting a hearing aid, this degree of variability—far from the normal ±5 dB of the behavioral audiogram—is worrisome. It is better than no information but far from optimal. Several procedures might improve the estimation. First, we could measure the size of the response at high intensities and use this to tell us what the amplitude at low intensities might be. In a subject with a small response at 80 dB HL, we might decide to average for longer periods of time to ensure that his or her inherently small response is properly recognizable at low intensities. Second, phase-weighting might help to facilitate the recognition of a response at low intensities (Picton et al, 2001). Third, since reduction of the background EEG noise is the most important factor determining the threshold variability, sedation might be needed for the test. Finally, accuracy might be improved by using some Bayesian approach to estimating the whole audiogram rather than thresholds at individual audiometric frequencies (e.g., as used in behavioral audiometry by Eilers et al, 1993).

Acknowledgment. This research was supported by the Canadian Institutes of Health Research, the Rotman Research Institute, and James Knowles. Malcom Binns provided statistical advice.

REFERENCES


Herdman AT, Stapells DR. (2001) Thresholds determined using the monotonic and dichotic multiple


---

**Figure 10.** Overestimation of amplitude of a two-dimensional signal measured in noise. The left side of the figure shows a sample signal recorded in noise. The signal is represented by the arrow. The noise is represented by the circle, which plots its standard deviation. The right side of the figure graphs the amount of overestimation. The relationship can be fitted with a double exponential curve such that the overestimation factor equals \(1 + 0.965e^{-1.348X} + 0.078e^{-0.285X}\) where “X” is the ratio between measured signal amplitude and noise amplitude.
APPENDIX: OVERESTIMATION OF SIGNAL AMPLITUDE

The measurement of a two-dimensional signal recorded in noise is a combination of the actual signal (arrow in the diagram at the left of Figure 10) and noise (represented by the circle surrounding the point of the arrowhead). The noise can be considered a normally distributed two-dimensional variable with equal variance in both dimensions. The circle in the diagram represents the standard deviation. Since the distribution of the measured values is symmetrical around the line of the signal phase, the measured phase is an unbiased estimate of the signal phase. However, the measured amplitude (distributed in and around the standard deviation circle) more overestimates (dark shading) than underestimates (light shading) the actual amplitude. This will make the average measurements of the signal amplitude larger than the actual amplitude by a factor that varies with how large the measured amplitude is relative to the noise. This factor was derived from stochastic modeling and plotted in the graph at the right. The range of the graph is not extended below a ratio of 1.25 since the response would not be considered significant at lower levels, nor beyond 10 since the differences are then too small to be meaningful. The straight lines in the graph show how to calculate an unbiased estimate of the signal amplitude. If the measured amplitude of the response is 50 nV and the residual background EEG is 10 nV (F value of 25), an unbiased estimate of the amplitude is 50 nV divided by the compensation factor for a ratio of 5:50/1.02 or 49 nV.