Phase Stability of Auditory Steady State Responses in Newborn Infants

Jong Min Choi,[1] David W. Purcell,[1] and M. Sasha John[2,3]

Objectives: This study examined the phases of auditory steady state responses (ASSRs) evoked by exponentially amplitude-modulated (AM^2) tones in 44 newborn infants (within 3 days of birth) and in 15 older infants (within 3 to 15 wks of birth). Our hypothesis was that the phases of the ASSRs would show orderly changes with modulation rate/carrier frequency and that this stability could be used with phase-biasing statistical techniques to augment response detection.

Design: Multiple ASSRs were recorded to four modulated tonal carriers with intensities of 50 dB SPL, which were combined and presented simultaneously. The carriers of 0.5, 1, 2, and 4 kHz were modulated at rates between 78 and 95 Hz. Recordings lasted 12.3 mins. Data were analyzed offline with particular attention to phase and its possible exploitation in response detection using a phase-weighted *t* test (PWT). Population normative phase values were compared with self-normative values. The latter uses phase estimates from ASSRs that are detected at an earlier time to estimate expected phases of ASSRs, which have not yet been detected. This was implemented as an interstimulus phase-weighted *t* test (iPWT). A secondary analysis compared using fixed test durations where data were evaluated once at the end of the recording with variable test durations where data were evaluated after every sweep.

Results: Average phases were not statistically different between the newborn and older infants. The mean ASSR phases across both infant groups were 10°, 36°, 83°, and 110° in the left ear and 78°, 97°, 135°, and 138° in the right ear for the four modulated carriers, respectively. Of a total of 172 detected ASSRs across the four carriers, 63% (109/172), 84% (144/172), and 99% (170/172) of the phase values fell within ±30°, ±45°, and ±90° of the population mean values, respectively. Self-normative phase values were slightly closer to actual measured phases, than population normative values. Compared with the *F* test, with a fixed duration, the iPWT technique did slightly better (71.7% versus 77.1% detected). Compared with the *F* test, with variable test duration, test time was reduced using the iPWT technique for normal and weighted averaging by 4 and 2.9 sweeps (66 and 48 secs), respectively, while false-positive rates were maintained. Compared with tests that relied on the *F*-ratio and a fixed time of 12.3 mins, using variable test times and the iPWT approach resulted in a halving of test time, while slightly improving comparable ASSR detection rates (66.7% versus 72.5%). An inter-ear average phase difference of 52° was found, which was not accounted for by modulation rates used for left/right ears. Converting phase to latency yielded similar results to prior studies.

Conclusions: The phase responses of ASSRs evoked by AM^2 tones are stable in newborn and young infants. When using the multiple auditory steady state response (MASTER) technique, it is possible to employ phase-biasing methods to reduce test time and increase detection rates. Using self-normative intrastimulus phase difference values provides better estimated phases than average population phases for purposes of response detection.

(Ear & Hearing 2011;32;1–●)

[1]National Centre for Audiology, The University of Western Ontario, Ontario; [2]Rotman Research Institute, Baycrest, Toronto; and [3]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada.

INTRODUCTION

Human auditory steady state responses (ASSRs) evoked by amplitude-modulated (AM) tones with modulation frequencies between 70 and 110 Hz have been used to assess hearing in infants using either single stimuli (Cohen et al. 1991; Rickards et al. 1994) or multiple simultaneous stimuli (Lins and Picton 1995; John et al. 1998). These responses can provide audiometric information without requiring a patient to respond behaviorally to sounds. This is clinically important when evaluating the hearing of infants, young children, cognitively impaired adults, and patients who may have a functional hearing loss. The smaller ASSRs of infants, compared with adults, serve as an obstacle with respect to performing robust frequency-specific testing of hearing in clinical settings.

The reported detectability of frequency-specific ASSRs has varied across different studies and especially so in infants (John et al. 2004, see Table 1; Van Maanen & Stapells 2009). This variability is due to factors such as the age of the infants tested, the level and frequency of the stimuli, the acoustic noise levels in the recording environment, and the duration of the recording. It has been clearly shown that the response becomes more detectable with increasing age (Savio et al. 2001; John et al. 2004; Luts et al. 2004; Rance & Tomlin 2006). As in adults, response detection also varies with carrier frequency: a 2-kHz stimulus usually produces more robust ASSRs, which are detectable at lower stimulus levels than, for example, stimuli in the 500 Hz range. Since the background electrical noise in the recording decreases with averaging, distinguishing a response from noise becomes easier as the test duration is increased (Picton et al. 2003; Luts & Wouters 2004; Luts et al. 2004). For example, the thresholds in the study by Rance and Rickards (2002) were elevated compared with that in the study by Lins et al. (1996), who used longer recording durations. Lins et al. (1996) found that, on average, ASSRs in the first few months of life are between one-third and one-half the size of the adult response and the corresponding physiological threshold estimates are 10 to 15 dB higher. Methods that increase the detection of these smaller ASSRs without simply relying upon longer test times would be useful.

Since the speed and accuracy of estimating frequency-specific thresholds using ASSRs depends on response amplitude, increasing that amplitude allows improved ASSR detection. In young infants, one can try to increase the size of the response using specially designed stimuli. John et al. (2004) examined the use of amplitude-modulated (AM), mixed-modulated (MM), and exponentially modulated (AM^2) tones in infants and found enhanced responses for the latter two stimuli. Others have used new types of stimuli, such as multicarrier and chirp stimuli, to evoke larger responses in adults (Stürzebecher et al. 2006; Elberling et al. 2007).

Alternatively, one can improve ASSR test performance by decreasing the noise levels of the recording. One method of

0196/0202/11/3202-0001/0 • Ear & Hearing • Copyright © 2011 by Lippincott Williams & Wilkins • Printed in the U.S.A.

decreasing noise is to use weighted averaging rather than standard averaging (Luetkenhoener et al. 1985; John et al. 2001). Another approach is to use methods that reward ASSR amplitudes (when the responses have expected characteristics) while punishing noise.

Although the ASSRs are two-dimensional and can be described in terms of their amplitude and phase, most statistical techniques only rely on either amplitude, such as the *F* test (Schuster 1898; Zurek 1992), or phase, such as phase-coherence techniques (Rayleigh 1880; Stapells et al., 1987). It should be noted that all tests of amplitude implicitly are affected by phase variance, as larger phase variance will decrease response amplitude (John & Purcell 2008). Although tests that combine phase and amplitude information, such as the T^2 test (Hotelling 1931; Picton et al. 1987; Victor & Mast 1991) or magnitude squared coherence (Dobie & Wilson 1989, 1994a), are theoretically more powerful than phase measurements alone (Dobie & Wilson 1993), these different tests have been shown to be about equally effective in detecting responses when real data are used (Picton et al. 1987; Valdes et al. 1997).

Picton et al. (2001) demonstrated that phase biasing could be used to increase the speed of ASSR detection in adults. The idea of phase biasing is that if the phase of the response is likely to be a particular value, one can bias the detection procedures toward recognizing responses as significant when phases are close to this expected value. Early attempts (Dobie & Wilson 1994b; Lins et al. 1996) used cosine or cosine-squared functions to favorably weight responses with phases within 90° of an expected phase. Although these techniques improved the detection rate, they required empirical adjustment of the statistical decision criteria to prevent too many false-positive detections of noise in the absence of a response. The phase-biasing technique explored by Picton circumvented this need for empirical adjustment by modifying the *F* test (Picton et al. 2001). The estimates at the stimulus modulation frequency and adjacent noise frequencies were projected onto an expected phase. The projected amplitude at the stimulus modulation frequency was then compared with the distribution of the projected noise. This new "phase-weighted *t* test" was simple to implement and did not require setting an empirical decision criterion.

The phase-weighted *t* test has not been applied to infant data. John et al. (2004) briefly evaluated the phases of the steady state responses in newborns, and for AM^2 stimuli the variability was found to be similar to adult values, for the same stimuli, reported by Picton et al. (2001; see Table 6 in the 2004 study compared with Table 1 in the 2001 study). In addition to simply using population phase values to phase bias data during a test, it was also hypothesized that intrasubject phase values could be used for phase biasing. For example, the expected phases could be derived based on the phases of responses, which were tested at higher intensities during initial conditions of the hearing test. Alternatively, intrasubject phase values could be derived using the phases of responses that were detected earlier during a test at a particular intensity level, when using simultaneously presented modulated carriers. These values can then be used for the remainder of the recording (at that intensity) to derive expected phase for the stimuli for which ASSRs had not yet been detected.

This study evaluated the characteristics and stability of phase data, for both newborn and young infants, previously recorded in the study by John et al. (2004). Both absolute phases for each response and interstimulus phase differences were calculated. The use of these phase measurements in biasing detection of ASSRs was then evaluated. Three main objectives were undertaken. The first was to thoroughly characterize the phase distribution across our infant sample. The second was to assess whether population absolute phase or within-subject interstimulus phase relationships would be more stable. The third was to evaluate phase biasing to determine whether this could improve performance by increasing response detection rate and/or shortening measurement time in ASSR-based testing of infant populations.

METHODS

Subjects

Data from 59 young infants were used for the analysis. There were two age groups including a "newborn" group of 44 subjects, who were tested within 74 hours after birth, and an "older" group of 15 subjects, who were tested between 3 and 15 wks after birth. Fifty-three infants were tested using monaural stimuli (27 left ears; 26 right ears), and six subjects were tested using binaural stimuli. Four data sets were removed (three monaural and one binaural) because they failed a noise criterion requiring the average noise amplitude to be <40 nV at the end of the recording. The resulting dataset included 50 subjects tested monaurally and 5 subjects tested binaurally (60 ears total; 29 left, 31 right).

ASSR Stimuli, Recording, and Phase Measurement

Although John et al. (2004) assessed ASSRs from AM, MM, and AM^2 stimuli, this study evaluated only response data evoked by AM^2 stimuli. An initial review indicated that this dataset had the largest number of significant responses and the smallest circular standard deviation (CSD) of ASSR phase values for both 500 and 4000 Hz responses, with comparable CSDs for 1000 and 2000 Hz responses. Overall, CSDs were 41°, 35°, and 37° for AM, MM, and AM^2, respectively, across all carriers. AM^2 stimuli are amplitude-modulated pure tones where the modulation is a sinusoid raised to the exponent of two (John et al. 2002). The modulating frequencies were 80.08, 84.96, 89.84, and 94.73 Hz in the left ear and 78.13, 83.01, 86.91, and 91.80 Hz in the right ear for the carrier frequencies of 500, 1000, 2000, and 4000 Hz, respectively. All recordings had 45 sweeps of 16.384 secs (with each sweep having 16 data epochs of 1.024 secs), leading to a total test time of 12.3 mins. The stimuli were fixed at 50 dB SPL for each of the four carrier frequencies, which were combined into a single stimulus of 56 dB SPL. Since the cochlea will separate the four carrier frequencies onto different tonotopic regions, the functional SPL may be regarded, clinically, as 50 dB SPL. Stimulus creation and data recording were performed using a modified version of the multiple auditory steady state response (MASTER) system implemented on Bio-logic Navigator-Pro System. For further measurement details, see John et al. (2004). The raw recorded data were stored so that offline analysis could subsequently be carried out.

Unless otherwise noted, the phase values reported here are "onset phases," which have been measured relative to sine onset and which have been adjusted to compensate for nonbiological delays such as filter delays produced by signal ampli-

fication. When discussing ASSR latencies, we will calculate "phase delay" as −360° onset phase (John & Picton 2000).

Normal Versus Weighted Averaging

The analysis of the data was accomplished offline using MATLAB (MathWorks, Natick, MA). During the offline testing, virtual ASSR tests were accomplished by reanalyzing the raw EEG data that were loaded into the analysis routine in the same order that occurred when infants were actually tested. Effects of normal compared with weighted averaging were assessed for both *F* test and phase-weighted *t* test approaches evaluated in this study (John et al. 2001). With weighted averaging, there is no need to explicitly employ a rejection threshold to exclude data with large amounts of noise, as is commonly done in other types of electrophysiological measurement such as the auditory brain stem response.

F Test Versus Phase-Weighted *t* Test

The *F* test is a detection statistic that employs a power ratio between the Fourier component of a modulation frequency (where the ASSR signal occurs) and those of its 120 neighboring frequencies located 60 bins above and below the modulating frequency (the "noise"). The width of each frequency bin was 0.061 Hz (i.e., 1/16.384), with 60 bins spanning 3.71 Hz. The noise estimate was therefore derived using a total range of 7.42 Hz. The *F*-ratio is evaluated as *F*-distributed with 2 and 240 degrees of freedom (Zurek 1992; John & Picton 2000).

Picton et al. (2001) introduced the phase-weighted *t* test (PWT) as a new method for detecting the ASSR. The idea of the PWT is to use prior knowledge of the anticipated ASSR phase to "reward" the estimate of the ASSR for having a phase that is similar to the expected phase while also "punishing" noise bins for having phase values that are different from the expected phase. A population-based expected phase can be computed as the circular mean of the phase obtained from a sample group of subjects using specific stimulus and recording parameters. Circular mean and circular SD are necessary to address the circularity of phase and calculate correctly that, for example, the mean phase of 355° and 5° is 0° rather than 180° (see Zar 1999, p. 599–604). In practice, the data for the *F* test are transformed into that used by the phase-weighted *t* test by projecting the amplitudes for the ASSR signal and the 120 adjacent noise frequencies onto the expected phase. Projected amplitudes are obtained by multiplying the amplitude of a particular frequency bin by the cosine of the difference between the estimated phase for that FFT bin and the expected phase of the ASSR. The significance of the ratio between the projected amplitude and the projected noise is then assessed using a *t* test with 119 degrees of freedom:

$$t = \frac{A_S \cos(\theta_E - \theta_S)}{\mathrm{std}(A_{N_i} \cos(\theta_E - \theta_{N_i}))} \quad (1)$$

where θ_E is the expected phase of the ASSR response, A_S and θ_S are the amplitude and phase actually observed in the recorded ASSR, and A_{Ni} and θ_{Ni} are the amplitude and phase of each of the 120 noise bins. The term "std" means standard deviation, which gives the root mean square-projected amplitude. This projection produces two desired effects. When the ASSR has a phase that is close to the expected phase, the amplitude is "rewarded" by being multiplied by a coefficient that is near 1, while noise amplitudes are multiplied by coefficients that decrease from 1 to 0 as the phase of the energy in the noise bin moves away from the phase at which the ASSR is expected to occur.

Population-Based Expected Phase Versus Interstimulus Expected Phase

Phase biasing can be implemented using various strategies for deriving the expected phase values. One option is to use "population-based expected phase" where the expected phase is the (circular) mean phase of a sample of an age-matched population tested with the same stimuli and testing environment (e.g., intensity, transducer type, and filter settings). Alternatively, "interstimulus expected phase" is the expected phase difference between ASSRs to two specific stimuli within the same individual and is based on the average difference found for the population. This type of relative phase ("self-normative") strategy may rely on, for example, using the same stimulus presented at two or more different intensities or using two or more stimuli (e.g., modulated carrier frequencies), which are simultaneously presented at the same intensity. Since our dataset only had ASSRs recorded at one intensity, we explored using a phase-biasing test based on expected interstimulus differences. In this case, after at least one ASSR elicited by a particular modulated carrier in the multiple-stimulus complex became significant, the expected phases of the ASSRs to other modulated carriers that were simultaneously presented were estimated. These estimated expected phases were then used to detect the ASSRs that had not yet become significant.

In the case where an individual's phases were dissimilar to population-based expected absolute phase values (e.g., an individual's phase values to each modulated carrier were consistently shifted by 50° relative to the population average values), phase differences between ASSRs evoked by different modulated carriers could still be consistent to those found for the population. We therefore expected that the interstimulus phase-weighted *t* test (iPWT) would show similar, if not better, performance compared with the population-based phase-weighted *t* test (PWT) using absolute phase.

To derive population-based expected phases for left and right ears, ASSR data for the four AM^2 carriers presented to each ear were analyzed and the phases of the detected ASSRs were used to calculate the population-based expected phases for each modulated carrier frequency (total of eight for the two ears). Population-based expected phases were calculated using only ASSRs that were significant, assessed using the *F* test and weighted averaging, at the end of the 45th sweep of the recording. We relied on the phases at the 45th sweep for calculating population expected phase since it is here that the SNR is often the highest. Higher SNRs are preferred since these provide phase estimates least affected by noise contamination. Expected phases must be calculated separately for the left and right ears because different modulation frequencies are used in each ear for a given carrier (so that the evoked ASSRs can be differentiated). Furthermore, unlike adult ASSRs, phases for the left and right ears appear to be distinct in newborns in manners that may not be solely attributed to modulation frequency differences between the two ears. This will be addressed further in the Discussion section.

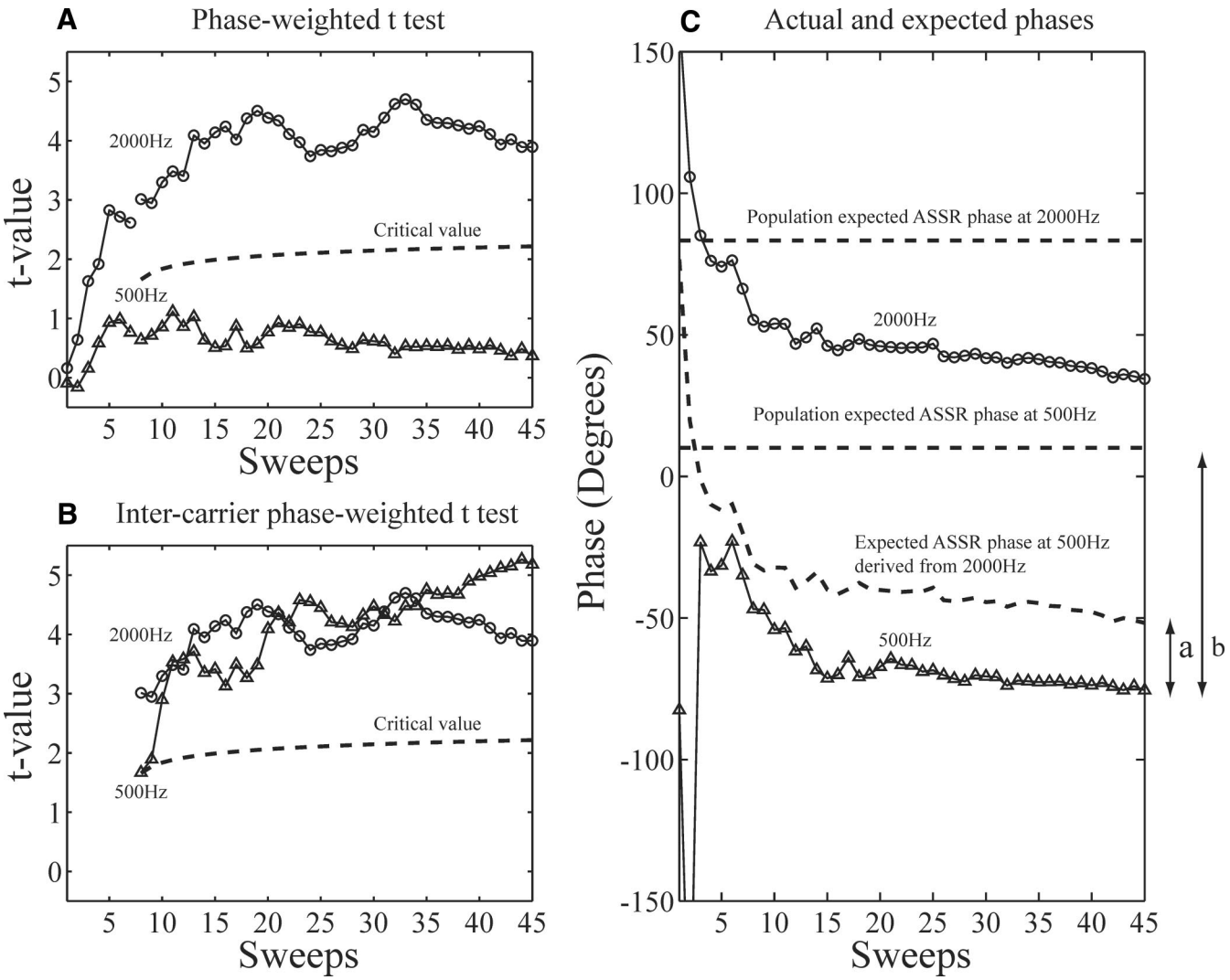

Fig. 1. Phase biasing with population expected phases and intercarrier expected phases. The *t* values of the PWT are shown in panel A, where triangles and circles denote the *t* value for the 500 Hz and 2000 Hz ASSRs, respectively. The dashed line shows the critical *t* value required for statistical significance at $p < 0.05$, determined using an adjusted Bonferroni correction (which will be discussed later in the article). In panels A and B, *t* tests did not begin until the 8th sweep. For the PWT in panel A, the *t* value was never statistically significant at 500 Hz. The *t* values for the iPWT are shown in panel B. With the iPWT test, the *t* values at 500 Hz were significant. The population expected phase values and actual phases at 500 and 2000 Hz are shown in panel C. The indication "a" in panel C is the phase difference between actual phases at 500 Hz and expected phases calculated from actual phases at 2000 Hz. The indication "b" is the phase difference between actual phases at 500 Hz and absolute population expected phases at 500 Hz. The length of "a" is much shorter than that of "b", showing that iPWT would favor response detection over PWT.

To employ the iPWT, ASSR testing began by evaluating the sweeps near the beginning of the data records using a PWT strategy. After the first ASSR was detected, the subject's expected phases for the other modulated carriers were calculated using anticipated interstimulus differences, and the iPWT test was then applied for the remainder of the dataset. For example, if the first ASSR detected was to the 2000 Hz modulated carrier at the 11th sweep, the expected phases for the other carrier frequencies were set using both the phase of the ASSR elicited by the 2000 Hz carrier and anticipated interstimulus phase differences. These derived expected phases were then used with iPWT testing for sweeps 12 to 45.

The iPWT could have been implemented in several manners. For example, if the first detection happened at both 2000 and 4000 Hz, then either detected response phase could serve to calculate the expected phase for 500 Hz. We tested all possibilities, but the detection differences were minor (results differed by only one or two detections), and a McNemar test (Zar 1999, p. 169–175; McNemar 1947) showed no significant difference in detection rates. However, since algorithms must follow rules, priority was given to 2k, 4k, 1k, and then 500 Hz, in that order, when simultaneous detections occurred. Figure 1 shows an example in which the iPWT detection was more effective than PWT.

Detection Paradigm

During MASTER-based testing, data sweeps are consecutively added to a running average sweep, which is used for ASSR calculations (other ASSR-based testing methods can use alternative data analysis techniques). In one strategy, ASSR

testing proceeds for a fixed number of sweeps and ASSR detection occurs a single time after the last sweep. Although simple to implement, test times will often be much longer than necessary. Alternatively, after each sweep is acquired, it is possible to evaluate ASSR detection. However, repeated testing will lead to a higher false-positive rate unless a compensatory method is employed (Luts et al. 2004; Stürzebecher et al. 2005).

False-positive rates can be controlled by requiring that the statistic being evaluated exceeds a critical value (e.g., 0.05) for N consecutive sweeps, where as when N becomes larger, the test becomes more conservative. Based on the previous work by Luts et al. (2008), we evaluated using two values of N and compared these results to simply testing at the completion of the last sweep (the 45th sweep in each subject's dataset). These approaches are referred to here as "Consecutive 4" rule, "Consecutive 8" rule, and the "last sweep" rule. Luts et al. (2008) noted that initial sweeps of a recording often contained noise that caused false alarms. This is likely due to the greater influence of relatively noisy sweeps when there are few sweeps in the average. Similar to the prior study, the testing strategy used here did not initiate response evaluation until after the first eight sweeps were "acquired." By applying both a "wait for the first eight sweeps" and a "Consecutive" rule, the ASSRs were not detected before the 8 + N − 1 sweep of a recording, where N was set to 4 or 8. Accordingly, detection did not occur until at least 3.5 mins of data had been submitted to the detection algorithm. With the simple "last sweep" rule, the ASSR was detected if the statistic at the last sweep (i.e., the 45th) was over the critical value for $p = 0.05$. We will use the abbreviation "C4-min8" to mean a "Consecutive 4" rule was applied after eight initial sweeps were collected. This expression is the same as the Luts et al. (2008) expression "V4-min8" in their Table 1. Similarly, "C8-min8" refers to our "Consecutive 8" rule as applied after eight initial sweeps were collected.

Multiple Endpoint Correction Strategies

It is sensible to assume that ASSRs persist across consecutive sweeps. A real ASSR should consistently remain detectable in the absence of a change in background noise (e.g., myogenic artifact), which may obscure the response. However, as shown by Luts' group, requiring response detection to span across even eight consecutive sweeps can still produce an unacceptable rate of false alarms since noise, mimicking an ASSR, may also persist across multiple consecutive tests. Managing false-positive detections by further increasing the number of consecutive sweeps relied on by a detection criterion (e.g., requiring 12 sweeps) would increase test time and could decrease test sensitivity, as both false-positive and real responses might not meet this criterion, especially when testing at lower intensities and in infants. To overcome this issue, in addition to relying on a consecutive sweep strategy, we increased the conservatism of our criteria by adjusting the critical value of our detection statistic according to a correction factor.

A simple Bonferroni correction has been shown to be too conservative for data that are correlated (Stürzebecher et al. 2005). An adjusted Bonferroni correction (ABC) was modeled upon consecutive sweeps in a running average sweep being correlated. An ABC was calculated based on the work by Sankoh et al. (1997, p. 2534) as follows:

$$\alpha_N = 1 - (1 - \alpha)^{1/N^{(1-r)}} \quad (2)$$

where α_N is the adjusted critical alpha level at the Nth comparison to maintain a desired false-positive rate during multiple comparisons (i.e., an overall false-positive rate of 0.05), α is the original nominal critical level that would be used for one comparison, N is the number of tests made so far (i.e., for the 10th sweep, N would equal 3 [10 − 8 + 1] since, in this example, testing occurred only after the 8th, 9th, and 10th sweeps, due to waiting at least eight sweeps), and r is a correction factor that was fixed as 0.65. This value is based on correlations of ASSR amplitudes at each sweep. While John and Purcell (2008) suggested using 0.75 for adult data, we found a correction factor of 0.65 to more appropriately maintain the expected false-positive rate of our infant data. The ABC approach caused the probability associated with the detection criterion to decrease from $p = 0.050$ at the eighth sweep to $p = 0.014$ ($N = 38$, $r = 0.65$) at the 45th sweep. The traditional Bonferroni correction would have required a p value of 0.0013 (0.05/[45 − 8 + 1]) or less at the 45th sweep. The ABC was used in the "Consecutive 4" and "Consecutive 8" strategies but not for the "last sweep" strategy where the responses were only evaluated after all data were included in the average sweep.

Statistical Analysis

Similar to previous studies that have evaluated multiple detection strategies (Picton et al. 2001; Luts et al. 2008), we employed the McNemar test to compare detection rates among the F, PWT, and iPWT tests. Two-by-two contingency tables were made, and the χ^2 statistic was calculated. In addition, repeated-measures analyses of variance (ANOVAs) and two-tailed paired t tests were applied to compare mean test times.

RESULTS

Figure 2 shows individual onset phases and population expected phases as polar plots. Individual phase values were measured after the 45th sweep from ASSRs that reached significance using weighted averaging and the F test. Data for the older and younger age groups were combined since an age group (older versus younger) × ear (left versus right) × carrier frequency (500, 1000, 2000, versus 4000 Hz) ANOVA failed to find a main effect for age ($F = 0.89$; df = 1,156; $p = 0.35$). Of the 29 left ears tested, the number of significant responses were 13, 23, 26, and 25 for carrier frequencies of 500, 1000, 2000, and 4000 Hz, respectively. Similarly for 31 right ears tested, there were 15, 19, 28, and 23 significant responses. The ANOVA showed a significant effect of carrier frequency ($F = 146.26$; df = 3,156; $p < 0.01$) and ear ($F = 97.95$; df = 1,156; $p < 0.01$).

Population-based expected phases and their SDs (given in parentheses, see Fig. 2) were computed as circular means and circular SDs (Zar 1999, p. 604, Eq. 26.21). In both ears, as carrier and modulation frequency increased, the expected phases moved in a counterclockwise direction (i.e., phase

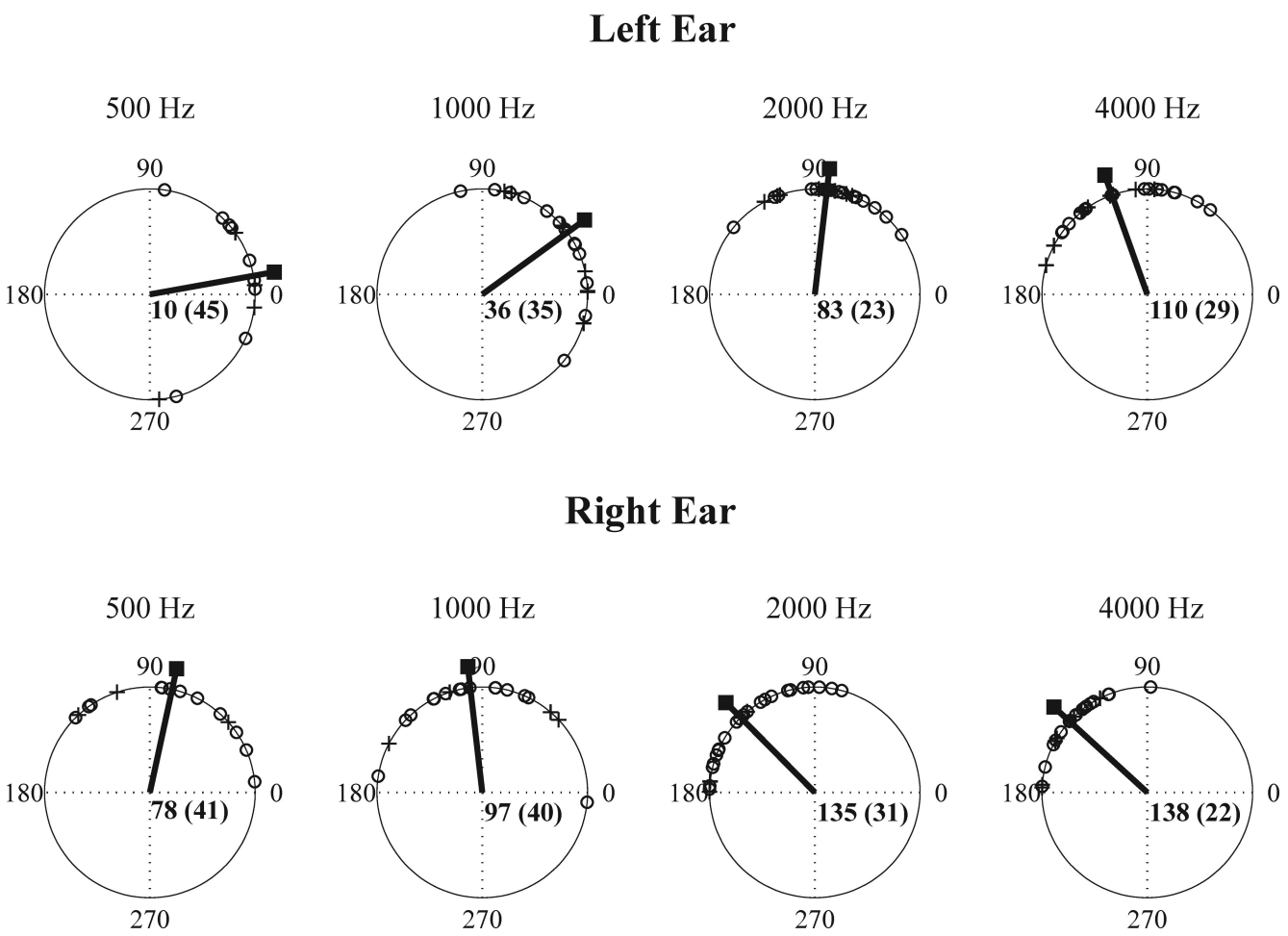

Fig. 2. Population expected phases and distributions of individual phases. Polar plots for the left and right ears are shown on the upper and lower row, respectively. Carrier frequency and modulation rate increase from left to right. Open circles represent newborn ASSR phases and the plus (+) symbols reflect response phases of the older group. The bold lines (terminated with squares) represent population-based expected phases. The circular mean and SD (in parentheses) is shown in the lower right quadrant of each plot.

increased). As discussed earlier, since latency is calculated by subtracting phase values from 360°, to obtain phase delay, increasing phase is actually shorter latency. Differences in expected phases between left and right ears for each carrier frequency ranged between 28° and 68°, with an average of 52°. Only a small proportion of the 52° (less than 10°) would be expected due to the different modulation frequencies used for corresponding carriers presented to the two ears. For example, for a 10-msec delay, the phase of the response generated to a 78 Hz modulated carrier would be 281° (i.e., = 10/[1000/78] × 360), whereas for 80 Hz, this same delay would be 288°. The circular SDs generally decreased as carrier and modulation frequency increased.

Figure 3 shows how densely the subjects' phases were distributed. Individual phase values are segregated into three groups based on differences between the expected and observed phases. Of a total of 172 detected ASSRs across the four carriers, 63% (109/172), 84% (144/172), and 99% (170/172) of the phase values were within ±30°, ±45°, and ±90°, respectively. Accordingly, the majority of the ASSRs were within ±45° and almost all were within ±90° of the population expected phases.

Intercarrier-Based Expected Phase

Figure 4 replots the same phase values of Figure 2, to illustrate the relationships between phase values of different modulated carrier frequencies as a function of ear. This figure highlights the moderate stability of the relative ordering of phase across the modulated carriers within individual subjects. Diamonds representing the phases of ASSRs evoked by 4000 Hz tend to be on top, followed by circles (2000 Hz), then squares (1000 Hz), with triangles (500 Hz) on the bottom for each ear. The right ears tend to have slightly larger absolute phases, which shift the spatial distribution toward the top of the graph.

Figure 4 also illustrates that while some ears had significant responses for all four modulated carriers (e.g., ear number 20), only one ASSR was detected in three ears (6%) (ears 37, 44, and 60). To compute normative values for interstimulus phase differences, only ears that had two or more detected ASSRs were used.

Table 1 shows the six possible interstimulus phase differences that were used in the iPWT tests. These values were calculated from individual data shown in Figure 4. It is noteworthy that most of the SD estimates were <45° with

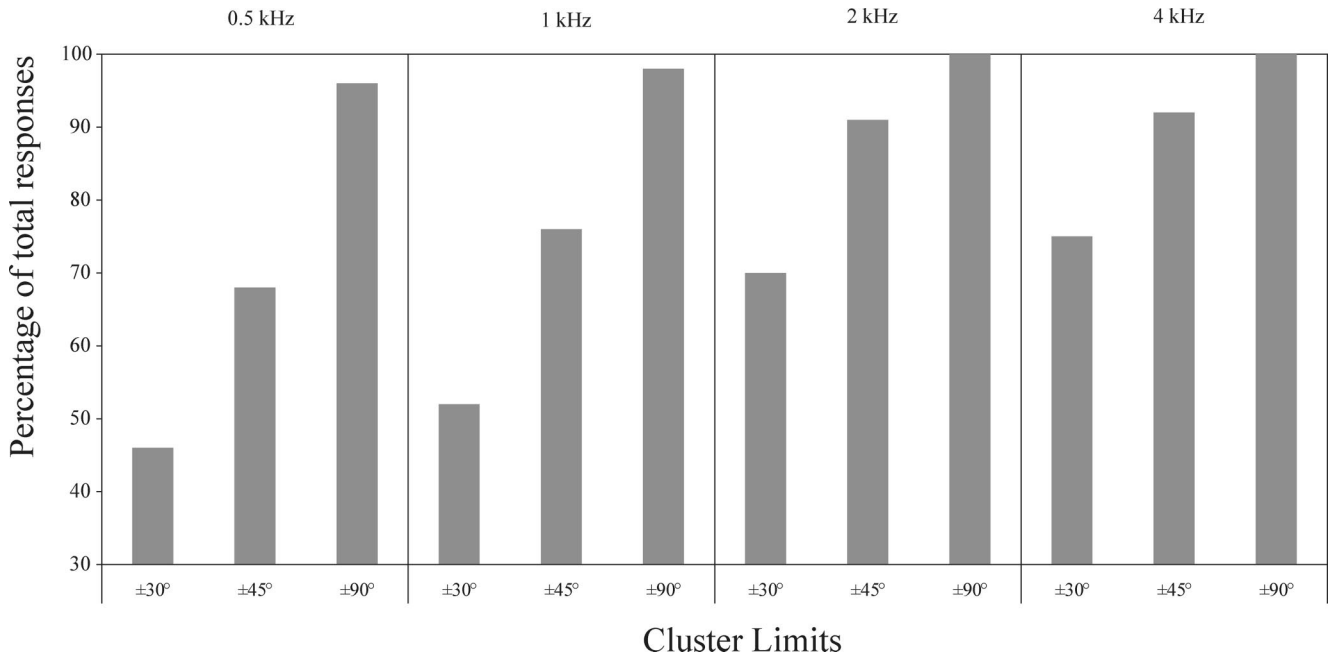

Fig. 3. Differences between expected and individual phases. The bars show how closely individual phase values were clustered to the expected values for each carrier frequency. Three categories are used to show the percentage of detected responses, which had a phase difference between population expected phase and individual phases—within 30°, 45°, and 90°.

similar range of variance to those seen for population-based absolute phases in Figure 2.

Comparing Performance of *F* Test, PWT, and iPWT

Comparison of Detection and False-Positive Rates • Table 2 shows the performance using "C4-min8", "C8-min8", and the "last sweep" testing strategies. Weighted averaging was clearly more effective than standard averaging both for increasing detection rates and decreasing false-positive rates. While the detection rate at the 45th sweep was numerically the highest among the three paradigms, the iPWT incorporated into the C4-min8 detection paradigm approximated the detection performance of the 45th sweep strategy using the traditional *F* test (72.5 versus 71.7).

Detection rate was evaluated from 240 possible detections (60 ears × 4 modulated carriers), whereas false-positive rate was evaluated from 7200 frequencies (240 × 30 control frequencies, as is reviewed further in the Discussion section) at which no ASSR would occur. The false-positive rates for the "C4-min8" strategies were just slightly >5% with standard averaging and slightly <5% when weighted averaging was used. Phase-weighted detection rates were consistently higher than the traditional *F*-ratio for both standard and weighted averaging and using both variable and fixed testing strategies. When reviewed as a function of carrier frequency, the iPWT increased detection performance over the *F*-ratio more for the 500 and 4000 Hz ASSRs than the 1000 and 2000 Hz ASSRs (data not shown). Predictably, the relatively conservative "C8-min8" yielded false-positive rates well below 5%, but detection rates consequentially suffered. Failure to rely on ABC caused the false-positive rates to be >10% with "C4-min8" and >6% using the stricter "C8-min8" criterion. Increasing the sequential requirement to a minimum of eight sweeps was not quite sufficient to control false-positive rates with weighted averaging when ABC was not implemented.

Comparisons of Total Detections and Relative Test Speed • Table 3 compares total detections and test speed using the *F* test, PWT, and iPWT strategies. Test speed refers to which test detected an ASSR first during sequential testing as new sweeps of EEG were made available. Test speed measures were computed for each ASSR and did not constrain the evaluation by requiring that ASSRs to all modulated carriers of a complex stimulus, which was presented to a particular ear, be detected. Only results for the "C4-min8" paradigm are shown since C8-min8 did not perform as well, as indicated in Table 2. For weighted averaging, the *F* test detected 160 (e.g., *F* versus PWT row: Both 146 + Only A 14) ASSRs while the PWT and iPWT detected 162 (e.g., *F* versus PWT row: Both 146 + Only B 16) and 174 (e.g., *F* versus iPWT row: Both 153 + Only B 21) responses out of the total possible 240 ASSRs. This means that the *F*, PWT, and iPWT approaches detected 66.7%, 67.5%, and 72.5%, respectively, as was reported in Table 2. Comparing *F* versus iPWT, among 153 ASSRs that were detected by both tests, the iPWT was faster than the *F* test in 68 cases compared with 19 ASSRs where the opposite was true (i.e., iPWT was faster 3.6 times more often). McNemar tests showed that the iPWT detected statistically significantly more responses and was significantly faster than the *F* test. The iPWT also detected significantly more responses than the PWT.

Comparisons of Test Time • Although test speed was reported in Table 3 under the category "faster," this information does not allow for an understanding of "how much faster." The difference between the two testing techniques may be one sweep or many. Accordingly, we also computed test time for ASSRs, which were detected by either test (i.e., all detections = Both A and B + Only A + Only B).

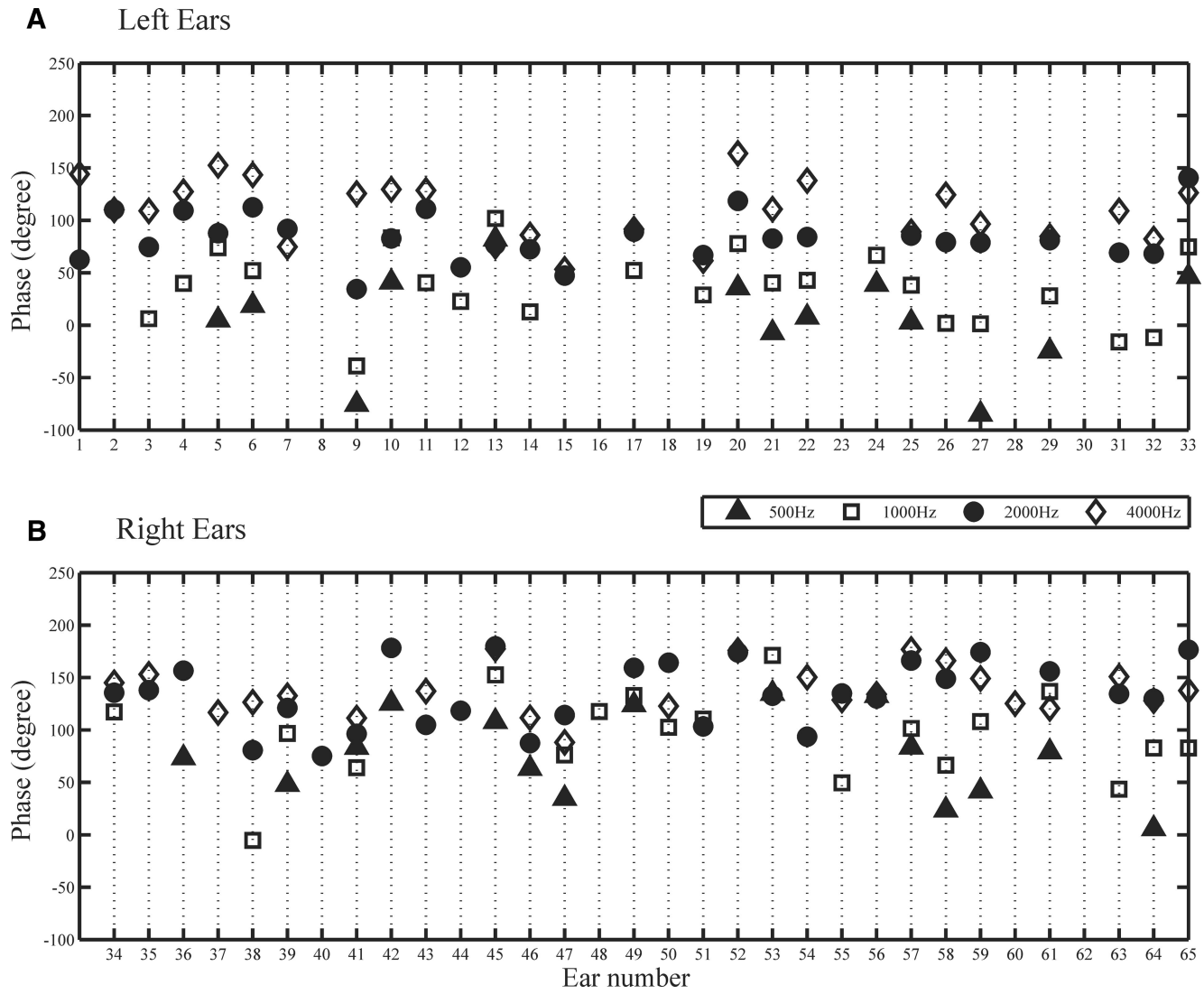

Fig. 4. Individual phases across carrier frequency. Individual phases for each carrier frequency presented to the left (panel A) and right (panel B) ears are shown. Only six individuals were tested binaurally and comprise ears 28–33 and 60–65. The data are fairly well ordered, with 500 Hz triangles tending to reside near the bottom of each distribution for a particular ear and 4000 Hz diamonds appearing near the top area of each graph. Only phases for ASSRs that were detected are shown.

Table 4 shows the mean test time using C4-min8 and the ABC strategy. The mean test time of iPWT was slightly, but significantly, faster than that of the *F* test for both averaging methods (two-tailed paired *t* tests, $p < 0.01$). Compared with the *F* test, using the iPWT reduced test times for normal and weighted averaging by 4 and 2.9 sweeps (66 and 47.6 secs), respectively, while false-positive rates were maintained (see Table 2). The mean test time of iPWT was not significantly faster than that of PWT for either averaging method.

TABLE 1. Intercarrier frequency phase differences

	0.5–1 kHz	0.5–2 kHz	0.5–4 kHz	1–2 kHz	1–4 kHz	2–4 kHz
Left						
No. responses	13	12	12	22	21	25
Mean	42.5°	86.2°	119.9°	51.5°	78.4°	24.1°
CSD	17.3°	37.2°	49.0°	26.8°	37.5°	27.9°
Right						
No. responses	11	15	11	18	15	21
Mean	39.0°	64.8°	71.7°	46.6°	51.1°	4.0°
CSD	26.0°	43.7°	41.4°	35.7°	38.6°	25.7°

The data for each value were computed using the number of responses, detected for carrier pairs, out of a total possible 29 left ears and 31 right ears for each carrier frequency.

TABLE 2. Detection and false-positive rates

		C4-min8		C8-min8		45th	
Averaging	Test	DR	FPR	DR	FPR	DR	FPR
Standard	F	34.6 (45.0)	5.0 (12.2)	27.5 (37.9)	2.7 (7.5)	34.2	4.7
	PWT	36.3 (51.3)	5.2 (11.0)	30.8 (43.8)	2.8 (7.6)	38.3	5.0
	iPWT	36.7 (52.1)	5.2 (11.4)	32.1 (43.3)	2.9 (7.4)	37.1	5.0
Weighted	F	66.7 (73.8)	4.9 (11.0)	60.8 (68.8)	2.3 (6.6)	71.7	4.6
	PWT	67.5 (75.8)	4.8 (10.7)	64.6 (72.9)	2.4 (6.4)	75.4	4.8
	iPWT	72.5 (76.3)	4.9 (10.9)	67.5 (75.0)	2.4 (6.3)	77.1	4.9

Detection and false-positive rates are given (in percentage) for standard and weighted averaging with ABC. The values in parenthesis were calculated without ABC. DR, detection rate as %; FPR, false-positive rate as %; F, F test; PWT, population phase-weighted t test; iPWT, intercarrier phase-weighted t test.

DISCUSSION

This study illustrates that individual infant ASSR phases were orderly and had a relatively narrow distribution around the circular population mean of each modulated carrier. The circular means served as population-based expected phases and were used for the PWT detection paradigm. In addition, we evaluated a new phase-weighted *t* test termed the iPWT, which relied on population-based interstimulus phase differences to provide estimates of expected phase. Phase-biasing methods showed modest, statistically significant improvements in detection rates and test speed when using variable length testing strategies, compared with the *F* test. Use of an ABC strategy allowed successful management of false-positive detection rates when ASSR tests relied on repeated testing designs.

Regularity of Observed Phases

The regularity of population expected phases has been previously reported (Picton et al. 2001, Table 1; John et al. 2004, Table 6; Alaerts et al. 2010, Table 1). Using the same stimuli and intensity as the present study for the right ear, Picton et al. (2001) found mean onset phases in normal-hearing adults to be 83°, 100°, 127°, and 120°, respectively. The CSDs for the adults of that study were comparable to those of our infant population (their Table 1 gives 60°, 37°, 28°, and 39° compared with our values in Figure 2, which, averaged across both ears, were 43°, 38°, 27°, and 26°).

Alaerts et al. (2010) reported phase delay values for infants using MM (with AM of 100% and FM of 20%). Their onset phases were −41°, −10°, 36°, and 39° at 50 dB SPL for 500, 1000, 2000, and 4000 Hz, respectively (personal communication). This progression is what one might expect due to traveling wave delays and was similar to the present study. The CSDs of their corresponding results were larger than ours (e.g., on the order of 20° to 30° larger). This may have been due to combining left and right ear data by Alaerts et al. Even for identical stimuli, phase values (and corresponding means and variances) can be influenced by stimulus intensity, age group of the subjects, and methods used to assess phase values both in individual recordings and across subjects.

Figure 2 shows that there are phase differences between left and right ears. We do not know the reasons for these differences. A small part of the difference is due to the different modulation rates used for the same carrier in the two different ears. This contribution would, however, produce less than 10° difference in the corresponding ASSR phases.

Individual ASSR phases for each carrier frequency were generally distributed relatively near the circular mean of the population. Although phase biasing has not yet been implemented by commercial systems, the fact that almost all (170/172) of the detected ASSRs were within ±90° of the population mean is noteworthy. This suggests that in clinical audiology if any ASSR has a phase that is outside of this range (or a slightly less conservative range of ±110°), then it should only be accepted as a real ASSR with caution, as there is a higher risk of a spurious detection (i.e., false positive).

ASSR Latencies

Phase can be converted into phase delay by subtracting onset phases from 360°. One expects increasing phase delay with decreasing carrier frequency, due to increasing traveling wave delays to more apical cochlear regions. Alaerts et al. (2010) reported phase delays of 24.3, 22.3, 19.4, and 18 msecs

TABLE 3. Comparisons of test speeds and detection rates

Averaging	Test	Neither	Both	Only A	Only B	A Faster	B Faster	Equal
Standard	F vs. PWT*	143	73	10	14	9	38	26
	F vs. iPWT*	143	74	9	14	8	37	29
	PWT vs. iPWT	144	79	8	9	9	10	60
Weighted	F vs. PWT*	64	146	14	16	15	67	64
	F vs. iPWT*†	59	153	7	21	19	68	66
	PWT vs. iPWT†	59	155	7	19	23	27	105

The table compares performance of F test, PWT, and iPWT detection methods using C4-min8 with ABC. "Neither" indicates that neither Test A nor Test B detected a response. "Both" indicates both Test A and B detected a response. "Equal" indicates times to detect ASSRs in Test A and B were the same. For "A faster," "B faster," or "Equal" comparisons, only data in which the ASSRs were detected by both tests was evaluated. For example, in the first row, strategy A was faster 9 times, B was faster 38 times, and these were equal 26 times, which sums to 73 total detections (i.e., the sum total of ASSRs which were detected by both strategies, as listed in the "Both" column).

**McNemar test—Detection speed significantly different at $p < 0.0001$.*

†McNemar test—Detection rate significantly different at $p < 0.05$.

TABLE 4. Mean test times

Averaging	Test	A or B	A	B
Standard	F vs. PWT*	97	24.4 ± 13.1	21.1 ± 12.0
	F vs. iPWT†	97	24.4 ± 13.1	20.4 ± 11.9
	PWT vs. iPWT	96	20.8 ± 11.9	20.2 ± 11.7
Weighted	F vs. PWT	176	21.4 ± 11.9	19.9 ± 11.2
	F vs. iPWT†	181	22.0 ± 12.4	19.1 ± 10.9
	PWT vs. iPWT	181	20.6 ± 11.8	19.1 ± 10.9

Mean test times using C4-min8 with ABC are shown for standard and weighted averaging. Results are reported as number of 16.384 second sweeps. For each subject, each individual ASSR for testing method A or B was only evaluated if the ASSR was present for at least one of these two methods. Column A or B provides the total number of ASSRs detected using either method A or B, with a maximum number of cases being 240 (60 ears × 4 carrier frequencies). There were three possibilities. If an ASSR was present using both methods A and B, then the number of sweeps for each method were simply compared. If an ASSR was detected as present using testing method A but not method B, then the number of sweeps for method B was set to 45 (since the testing would have had to continue until the end of the test period to determine the absence of the ASSR). If neither method detected an ASSR, then that data was not included in the analysis.
*Two-tailed paired t test: *p < 0.05; †p < 0.01.*

for 500, 1000, 2000, and 4000 Hz, respectively. Our phase delay results, similarly fitted with two preceding cycles, yielded latencies of 24.6, 22.4, 19.7, and 17.9 msecs for the left ear and 22.9, 20.9, 18.7, and 17.6 msecs for the right ear. Our right ear latencies averaged 1.2 msecs shorter than our left ears. This may reflect subtle lateralization differences in maturation of the auditory system. Sininger and Cone-Wesson (2006) reported asymmetry in neonates, with right ear having slightly faster ABRs than left ear for waves III and V. This finding is in line with ASSRs evoked by stimuli presented to the right ears having larger phases (shorter phase delay) than those evoked by left ear stimuli, although our ASSR latency differences are much larger than those reported for ABR. Left/right asymmetries have been found to exist at the level of the cochlea as well (Sininger & Cone-Wesson 2004). While these modest ear asymmetries have been reported in newborns, the phase and group delay latency differences reported here are more substantial than would be expected from these earlier reports.

John et al. (2004) reported onset phase values for AM^2 stimuli in their Table 6. The phases of AM^2 reported in John et al. (2004) were different to those reported here although the two studies used the same dataset. There were two reasons for this difference. The first was due to the instrumentation used to collect the data, which was a customized version of the Bio-Logic MASTER platform. During this study, using calibration waveforms, we discovered that the noninverting (active) and inverting (reference) leads of the custom instrument were reversed, causing the collected waveforms to be shifted by 180°. This montage caused the onset phase of the prior study to decrease with increasing carrier frequency (a trend that is backward to that which is normally found in adults). Second, since John et al. (2004) did not focus on phase and latency measurements, the filtering delays introduced by the instrumentation (−32.4°, −122.6°, −212.8°, and 57.0° for left ear modulation envelopes and −35.7°, −128.0°, −201.5°, and 66.1° for right ear modulation envelopes) were neither evaluated nor compensated for in that dataset. Since the current study adjusted for these factors, the conversion of phase to latency produced much more sensible results.

Sequential Testing Versus Fixed Recording Times

Luts et al. (2008) explored the use of both fixed and variable recording strategies. One variable test length option they explored used weighted averaging and was termed "V4-min8". This was identical to our "C4-min8" condition. This strategy resulted in an unacceptable false-positive rate of 8.9%. They then derived an adjusted critical p value of 0.0256, instead of 0.05, to achieve an acceptable error rate of 5% (see Table 2 in their study). Similarly, our ABC strategy, which maintained a false-positive rate of just below 5%, caused the critical p value to decrease from an initial value of 0.05 to 0.014 by the end of the recording. The average critical p value over the entire recording was 0.036, which is very close to the 0.0256 correction in the article by Luts et al. Accordingly, both studies found that C4-min8 can be used to save test time and still maintain detection rates, when weighted averaging and a similar alpha correction were applied. Compared with tests that relied on the F-ratio and a fixed recording time of 12.3 mins (45 sweeps), using a variable test time, ABC and the iPWT approach resulted in a halving of test time (from 45 sweeps to 19.1 sweeps, see Table 4) and comparable ASSR detection rates (71.7% versus 72.5%, for F-ratio fixed and iPWT variable, respectively; see Table 2).

Unlike the study by Luts et al. (2008), increasing the detection criteria from four to eight consecutive sweeps did not sufficiently maintain false-positive rates when ABC was not used. We found a false-positive rate of 6.6% without ABC using C8-min8 when recording for up to a maximum of 45 sweeps. Using a shorter maximum recording time of 32 sweeps, Luts et al. (2008) reported a false-positive rate of 4.4% using a similar consecutive rule (their V8 with no minimum number of sweeps). Taken together, these studies suggest that requiring eight consecutive significant detections as a stopping rule can be employed cautiously but may not be sufficient to control false-positive rates for recording times longer than 32 sweeps (i.e., tests that include more than 32 "looks" at the data). In addition, both studies rejected "noisy" patients with recordings with final noise levels near 10 nV (in the study by Luts et al., using an inion electrode) and 40 nV (current study, using a nape electrode). In patient recordings with more noise, larger numbers of false positives may emerge at the beginning of the recording.

Luts et al. (2008) proposed that a fixed recording length was the best option for ASSR detection in infants and adults with currently available equipment. This was based on two considerations. The first is that most current commercial systems do not allow for independent intensity control of the different carriers of a stimulus, automatic application of a consecutive rule, or alpha adjustment related to multiple testing (let alone all three). The second consideration was that variable test length did not save much measurement time or influence threshold estimation much. In the current study, only one intensity level was tested and so we were not able to infer the influence of test strategy on threshold estimation. However, there was a considerable decrease in test time. This may have been in part due to our use of longer maximum test durations (than used by Luts et al.), which may be required for ASSR detection during testing of young infant populations, since their responses are relatively small. Taken together, the evidence from these two studies suggests that if commercial systems implemented both independent stimulus control and alpha

correction, then test times could be reduced significantly (by 27% in the study by Luts et al. and by 58% in the current study) while increasing detection rates and maintaining false-positive rates.

Phase Biasing

The phase-weighted *t* test has not been previously applied to infant data using either PWT or iPWT techniques. The PWT and iPWT performed better than the *F* test with infant data for both test speed and detection rate. The current study furthers the conclusions of Picton et al. (2001) who showed, in adults, that prior knowledge about phase can improve detection while maintaining a 5% false-positive rate. We found that the new iPWT strategy performed better than the PWT.

Detection and False-Positive Rates

The calculation of detection rate for a given analysis method is simple: the number of ASSRs detected is divided by the number of ears tested multiplied by 4, due to the presentation of four modulated carriers to each ear. In contrast, there are at least two methods that can be used for determining false-positive rates. One option is to use data where no stimulus was presented, while another approach is to evaluate "control" frequencies where no ASSR should occur (i.e., using frequencies that are different from the stimulus modulation rates). In this study, we used 30 control frequencies, whereas John et al. (2001) and Luts et al. (2008) used 8. Use of 30 control frequencies may have caused the false-positive rate reported here to be a little more reliable than if we had used 8. The 30 control frequencies included 15 frequencies on each side of a stimulus modulation rate. This was the maximum number we could select to conveniently avoid ASSRs evoked by the neighboring stimulus modulation rates.

Test Time—Independent Carrier Frequency Intensity Control

Our comparison of testing time among the detection methods was done with only one intensity level, 50 dB SPL. Tables 3 and 4 are computed on individual ASSRs rather than basing the analysis on the clusters of four ASSRs. In many real testing situations, ASSRs are evoked by a stimulus with fixed carrier levels that cannot be independently adjusted. For example, if the responses were found on the 17th sweep for 1000, 2000, and 4000 Hz carriers, that audiologist must wait until the 500 Hz carrier is detected before lowering stimulus levels. With an instrument able to independently control stimulus levels, the carriers whose responses were detected on the 17th sweep could proceed immediately after detection to collect data at a lower stimulus level. At the same time, data collection could continue uninterrupted for the 500 Hz carrier until the ASSR is either detected or until a maximum limit related to time or noise is reached. This approach could improve the proportion of time spent recording near the threshold for each carrier frequency (where the ASSRs would have the smallest amplitude). John et al. (2002) showed that in adults the intercarrier level differences of 10 or 20 dB can be used without significant interaction between carrier frequency components.

Our variable length results model those which would be obtained using clinical instruments that exploit independent intensity control and enable stimulus parameters to be adjusted after each ASSR is detected. An instrument capable of independently controlling the intensity of each carrier frequency would be able to exploit a faster detection method in a multi-intensity search for the ASSR thresholds.

Phase at Lower Intensity

The stimulus intensity employed here was fixed at 50 dB SPL, which is somewhat higher than much of the intensity range at which hearing testing may occur. However, Van Maanen and Stapells (2009) suggest that normal air conduction ASSR "screening" levels for infants would be 50, 45, 40, and 40 dB HL for 500, 1000, 2000, and 4000 Hz carriers, respectively. In addition to frequency-specific threshold type tests, investigators have reported various ASSR screening paradigms (Cone-Wesson et al. 2002; Açikgöz et al., 2006; Savio et al. 2006), which could be improved using phase biasing since speed is an issue. For optimal performance, the expected phases used in PWT and iPWT should be changed based on stimulus level. John and Purcell (2008) showed a preliminary figure where onset phase changed in a counterclockwise direction with increasing intensity. John and Picton (2000) reported, in adults, a latency change of 0.06 msec/dB of change in stimulus level, and Alaerts et al. (2010) reported 1 msec difference between 40 and 50 dB (0.1 msec/dB). Averaging these two estimates and assuming 0.8 msec/10 dB change equates to between 22° (e.g., 0.8/[1000/75] × 360) and 29°/10 dB change, for modulation rates of 75 and 100 Hz, respectively. The relative phase differences between carriers may be roughly maintained at different intensity levels, but future studies should evaluate phase versus intensity functions in infants.

Limitations

This study generated its expected phases from the same population that was used to evaluate the phase-biasing techniques. This could serve to increase the benefit determined for phase biasing over that which might be found if the expected phase values were evaluated on a different out-of-sample infant population. When the sample of phase values reported here was randomly resampled nine times using half of all available values (16 left and 16 right ears), the distribution of circular means that resulted had a SD of only 1.6° to 9.5°. This would have had relatively little effect on the results reported here either for the expected phases obtained from the population means or for the performance of the phase-biasing techniques, which subsequently used these values. Regardless, the expected phases reported here should be evaluated with other infant data sets to provide an understanding of the replicability of the results reported here. We did not examine these techniques on infants with hearing loss, who may produce ASSRs with different phase values due to factors such as broader tuning curves.

CONCLUSIONS

Infant ASSR phases are sufficiently stable to enable benefits of phase-biasing detection strategies. ASSR detection algorithms using interstimulus expected phases are promising. The iPWT was modestly faster than the *F* and PWT tests and should be considered for use with infants after expected phases are derived for several stimulus levels to ensure that the results reported here are generalizable. Variable test time strategies, including phase biasing, which use compensatory strategies to address consecutive testing, can save substantial test time

compared with using fixed duration recordings, while also maintaining desirable detection and false-positive rates. Commercial instruments should provide features that enable audiologists to capitalize on these findings.

ACKNOWLEDGMENTS

The authors thank Patricia Muir, Patricia Van Roon, and David Brown for assisting with the data collection reported in John et al. (2004), upon which the current article is based. The authors also thank Drs. J. Alaerts, H. Luts, and J. Wouters, who kindly provided data from their recent article so that our two studies could be more usefully compared.

This work was supported by a Proof-Of-Principal Grant, the Canadian Institutes of Health Research and the (Canadian) Natural Sciences and Engineering Research Council, The Ontario Ministry of Research and Innovation through the Ontario Research Fund and by grants from the Advanced Biometric Research Center (ABRC), and the National Research Foundation (NRF) of KOREA.

Address for correspondence: Sasha John, PhD, Rotman Research Institute, 3560 Bathurst Street, Toronto, ON M6A-2E1, E-mail: sjohn@rotman-baycrest.on.ca.

Received November 9, 2010; accepted January 17, 2011.

REFERENCES

Açikgöz, N., Ozdamar, O., Delgado, R. E., et al. (2006). Audiometric threshold screening method using envelope detection filters of intensity ramping click auditory steady-state responses. *Conf Proc IEEE Eng Med Biol Soc*, *1*, 4983–4986.

Alaerts, J., Luts, H., Van Dun, B., et al. (2010). Latencies of auditory steady-state responses recorded in early infancy. *Audiol Neurotol*, *15*, 116–127.

Cohen, L. T., Rickards, F. W., Clark, G. M. (1991). A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans. *J Acoust Soc Am*, *90*, 2467–2479.

Cone-Wesson, B., Parker, J., Swiderski, N., et al. (2002). The auditory steady-state response: Full-term and premature neonates. *J Am Acad Audiol*, *13*, 260–269.

Dobie, R. A., & Wilson, M. J. (1989). Analysis of auditory evoked potentials by magnitude-squared coherence. *Ear Hear*, *10*, 2–13.

Dobie, R. A., & Wilson, M. J. (1993). Objective detection in the frequency domain. *Electroencephalogr Clin Neurophysiol*, *88*, 516–524.

Dobie, R. A., & Wilson, M. J. (1994a). Objective detection of 40 Hz auditory evoked potentials: Phase coherence vs. magnitude-squared coherence. *Electroencephalogr Clin Neurophysiol*, *92*, 405–413.

Dobie, R. A., & Wilson, M. J. (1994b). Phase weighting: A method to improve objective detection of steady-state evoked potentials. *Hear Res*, *79*, 94–98.

Elberling, C., Don, M., Cebulla, M., et al. (2007). Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay. *J Acoust Soc Am*, *122*, 2772–2785.

Hotelling, H. (1931). The generalization of Student's ratio. *Ann Math Statist*, *2*, 360–378.

John, M. S., Brown, D. K., Muir, P. J., et al. (2004). Recording auditory steady-state responses in young infants. *Ear Hear*, *25*, 539–553.

John, M. S., Dimitrijevic, A., Picton, T. W. (2001). Weighted averaging of steady-state responses. *Clin Neurophysiol*, *112*, 555–562.

John, M. S., Lins, O. G., Boucher, B. L., et al. (1998). Multiple auditory steady-state responses (MASTER): Stimulus and recording parameters. *Audiology*, *37*, 59–82.

John, M. S., & Picton, T. W. (2000). Human auditory steady-state responses to amplitude-modulated tones: Phase and latency measurements. *Hear Res*, *141*, 57–79.

John, M. S., Purcell, D. W., Dimitrijevic, A., et al. (2002). Advantages and caveats when recording steady-state responses to multiple simultaneous stimuli. *J Am Acad Audiol*, *13*, 246–259.

John, M. S., & Purcell D. W. (2008). Introduction to Technical Principles of Auditory Steady-State Response Testing. In G. Rance (Ed). *The Auditory Steady-State Response: Generation, Recording, and Clinical Application* (pp. 11–53). San Diego, CA: Plural Publishing.

Lins, O. G., & Picton, T. W. (1995). Auditory steady-state responses to multiple simultaneous stimuli. *Electroencephalogr Clin Neurophysiol*, *96*, 420–432.

Lins, O. G., Picton, T. W., Boucher, B. L., et al. (1996). Frequency-specific audiometry using steady-state responses. *Ear Hear*, *17*, 81–96.

Lutkenhoner, B., Hoke, M., Pantev, C. (1985). Possibilities and limitations of weighted averaging. *Biol Cybern*, *52*, 409–416.

Luts, H., Desloovere, C., Kumar, A., et al. (2004). Objective assessment of frequency-specific hearing thresholds in babies. *Int J Pediatr Otorhinolaryngol*, *68*, 915–926.

Luts, H., Van Dun, B., Alaerts, J., et al. (2008). The influence of the detection paradigm in recording auditory steady-state responses. *Ear Hear*, *29*, 638–650.

Luts, H., & Wouters, J. (2004). Hearing assessment by recording multiple auditory steady-state responses: The influence of test duration. *Int J Audiol*, *43*, 471–478.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*, 153–157.

Picton, T. W., Dimitrijevic, A., John, M. S., et al. (2001). The use of phase in the detection of auditory steady-state responses. *Clin Neurophysiol*, *112*, 1698–1711.

Picton, T. W., John, M. S., Purcell, D. W., et al. (2003). Human auditory steady-state responses: The effects of recording technique and state of arousal. *Anesth Analg*, *97*, 1396–1402.

Picton, T. W., Vajsar, J., Rodriguez, R., et al. (1987). Reliability estimates for steady-state evoked potentials. *Electroencephalogr Clin Neurophysiol*, *68*, 119–131.

Rance, G., & Rickards, F. (2002). Prediction of hearing threshold in infants using auditory steady-state evoked potentials. *J Am Acad Audiol*, *13*, 236–245.

Rance, G., & Tomlin, D. (2006). Maturation of auditory steady-state responses in normal babies. *Ear Hear*, *27*, 20–29.

Rayleigh, L. (1880). On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philos Mag*, *10*, 73–78.

Rickards, F. W., Tan, L. E., Cohen, L. T., et al. (1994). Auditory steady-state evoked potential in newborns. *Br J Audiol*, *28*, 327–337.

Sankoh, A. J., Huque, M. F., Dubey, S. D. (1997). Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med*, *16*, 2529–2542.

Savio, G., Cárdenas, J., Pérez Abalo, M., et al. (2001). The low and high frequency auditory steady state responses mature at different rates. *Audiol Neurootol*, *6*, 279–287.

Savio, G., Perez-Abalo, M. C., Gaya, J., et al. (2006). Test accuracy and prognostic validity of multiple auditory steady state responses for targeted hearing screening. *Int J Audiol*, *45*, 109–120.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestr Magnet Atmos Electr, 3,* 13–41.

Sininger, Y. S., & Cone-Wesson, B. (2004). Asymmetric cochlear processing mimics hemispheric specialization. *Science*, *305*, 1581.

Sininger, Y. S., & Cone-Wesson, B. (2006). Lateral asymmetry in the ABR of neonates: Evidence and mechanisms. *Hear Res*, *212*, 203–211.

Stapells, D. R, Makeig, S., Galambos, R. (1987) Auditory steady-state responses: Threshold prediction using phase coherence. *Electroencephalogr Clin Neurophysiol*, *67*, 260–270.

Stürzebecher, E., Cebulla, M., Elberling, C. (2005). Automated auditory response detection: Statistical problems with repeated testing. *Int J Audiol*, *44*, 110–117.

Stürzebecher, E., Cebulla, M., Elberling, C., et al. (2006). New efficient stimuli for evoking frequency-specific auditory steady-state responses. *J Am Acad Audiol*, *17*, 448–461.

Valdes, J. L., Perez-Abalo, M. C., Martin, V., et al. (1997). Comparison of statistical indicators for the automatic detection of 80 Hz auditory steady state responses. *Ear Hear*, *18*, 420–429.

Van Maanen, A., & Stapells, D. R. (2009). Normal multiple auditory steady-state response thresholds to air-conducted stimuli in infants. *J Am Acad Audiol*, *20*, 196–207.

Victor, J. D., & Mast, J. (1991). A new statistic for steady-state evoked potentials. *Electroencephalogr Clin Neurophysiol*, *78*, 378–388.

Zar, J. H. (1999). *Biostatistical Analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Zurek, P. M. (1992). Detectability of transient and sinusoidal otoacoustic emissions. *Ear Hear*, *13*, 307–310.