Taylor & Francis
Taylor & Francis Group

Original Article

International Journal of Audiology 2005; 44:244–253

Heleen Luts
Jan Wouters

Lab. Exp. ORL, Katholieke Universiteit Leuven, Leuven, Belgium

Key Words
Auditory steady-state responses
Objective audiometry
Hearing threshold

Comparison of MASTER and AUDERA for measurement of auditory steady-state responses

Comparación de MASTER y AUDERA para la medición de las respuestas auditivas de estado estable

Abstract
Two approaches to assess auditory steady-state responses (ASSR) are compared under similar test conditions: a monaural single-frequency technique with a detection method based on phase coherence (AUDERA), and a binaural multiple-frequency technique using the F-test (MASTER). ASSR thresholds at four frequencies were assessed with both methods in both ears of ten normal-hearing and ten hearing-impaired adult subjects, within a test duration of one hour. The test-retest reliability and the influence of prolonging the test duration are assessed. For the total subject group the multiple-frequency technique outperforms the single-frequency technique. In hearing-impaired subjects, however, both techniques perform equally well. Hearing thresholds can be estimated with a standard error of the estimate between 7 and 12 dB dependent on frequency. About 55% of the estimates are within 5 dB of the behavioral hearing threshold, and 94% within 15 dB. Prolonging the test duration improves the performance of both techniques.

Sumario
Se compararon dos enfoques para evaluar las respuestas auditivas de estado estable (ASSR) en condiciones similares de prueba: la técnica monoaural unifrecuencial con un método de detección basado en la coherencia de fase (AUDERA) y una técnica binaural multifrecuencial usando el F-test (MASTER). Se determinaron los umbrales ASSR en cuatro frecuencias con ambos métodos en los dos oídos de diez normo-oyentes y diez hipoacúsicos adultos durante una hora, así como la exactitud de la prueba repetida y la influencia de laprolongación de la prueba. En todos los casos la técnica multifrecuencial superó a la unifrecuencial. No obstante, en los hipoacúsicos, ambas técnicas resultaron iguales. Los umbrales auditivos pueden estimarse con un error estandar estimado entre 7 y 12 dB dependiendo de la frecuencia. Alrededor del 55% de las estimaciones estuvieron en un rango de 5 dB de los umbrales comportamentalesy 94% en un rango de 15 dB. La prolongación de la prueba mejora el rendimiento de ambas técnicas.

Over the past decades, the need for objective audiometric techniques in clinical practice has increased. This is partly the result of the growing target population for objective techniques after the world-wide introduction of hearing screening in newborns. The technique that is applied most in this young population is the click-evoked auditory brainstem response (ABR), because of the short test duration. For an efficient fitting of hearing aids, however, hearing threshold estimates at different octave frequencies are required. Tone-burst-evoked ABR and auditory steady-state responses (ASSR) can provide frequency-specific hearing threshold estimates. ASSRs are the periodic electrical responses of the brain to auditory stimuli presented at a rate fast enough to cause an overlap of successive responses (Maiste & Picton, 1989; Stapells et al, 1984). These potentials can be elicited by amplitude- and/or frequency-modulated pure tones. Responses evoked by stimuli modulated at 80 Hz can reliably be recorded in children (Aoyagi et al, 1993; Aoyagi et al, 1994; Cohen et al, 1991; Rance et al, 1995; Rickards et al, 1994) and are not affected by sleep or sedation (Cohen et al, 1991; Plourde & Picton, 1990).

The ASSR technique has several potential advantages over tone-burst-evoked ABR. Firstly, test duration can be shorter. Tone-burst-evoked ABR is time-consuming since results for each audiometric frequency will take the same amount of time as that for one click-evoked ABR (Stapells & Oates, 1997). Secondly, because of the continuous nature of the stimuli used to elicit ASSRs the maximum output level is less restricted compared to tone-burst-evoked ABR. And finally, the response detection in the frequency domain based on statistical tests assures that the ASSRs are detected objectively.

Several studies have been carried out to investigate how accurate behavioral hearing thresholds can be predicted by means of the 80 Hz ASSR. In general, two major approaches have been thoroughly investigated: firstly, a monaural single-frequency approach, with short recording times and a response detection method based on phase coherence (Aoyagi et al, 1999; Cohen et al, 1991; Rance et al, 1995; Rance et al, 1998; Rance & Briggs, 2002; Rickards et al, 1994); and secondly, a binaural multiple-frequency technique with long recording times and a response detection based on an F-test (Dimitrijevic et al, 2002; Herdman & Stapells, 2001; Lins & Picton, 1995; Luts et al, 2004; Luts & Wouters, 2004; Perez-Abalo et al, 2001). Besides the difference in number of stimuli simultaneously presented and the response detection paradigm, variations in subject group, test environment and total test duration will also affect the results of these studies. Therefore, it is difficult to determine whether both

ISSN 1499-2027 print / ISSN 1708-8186 online
DOI: 10.1080/14992020500057780

Received: April 23, 2004
Accepted: September 16, 2004
Heleen Luts
Lab. Exp. ORL, K.U.Leuven, Kapucijnenvoer 33, B-3000 Leuven, Belgium
E-mail: Heleen.Luts@uz.kuleuven.ac.be

approaches produce similar results or whether one technique is more accurate and/or faster than the other. Little research has been done directly comparing the two methods.

In this study, the single-frequency and multiple-frequency approach were compared in the same subjects and in a quiet test environment. Firstly, the ability of both techniques to predict hearing thresholds by means of ASSR over a wide intensity range within a total test time of 1 hour was evaluated. Secondly, the influence of prolonging the test duration was assessed. Thirdly, the test-retest reliability of both techniques was examined.

Methods

The software for the multiple-frequency ASSR recordings, the MASTER (Multiple Auditory STEady-state Responses), was developed by and based on the research of John and Picton (2000) at the Rotman Research Institute, University of Toronto. For the single-frequency approach a GSI AUDERA device of Grason-Stadler was used. The equipment manufactured by ERA Systems, Ltd., based on research at the Department of Otolaryngology, The University of Melbourne, served as a prototype for the AUDERA device. For the sake of simplicity in what follows, both systems will be referred to as the MASTER and the AUDERA system. Parameter settings for both systems were selected, as far as possible, as proposed by the developer and/or manufacturer. As a consequence, there are some differences in testing parameters between both systems (see Table 1).

The MASTER system

Eight stimuli were simultaneously presented, four to each ear. A clinical audiometer, Madsen Orbiter 922, was used to control the overall stimulus intensity of the left and right channel. The eight separate signals were calibrated using a Brüel & Kjaer Sound Level Meter 2260 in combination with an artificial ear 4152 and a 2-cc coupler DB0138.

Silver-silver chloride electrodes were connected to a Stanford Research Systems SR560 amplifier. A conductive paste was used to keep the electrodes in place and to obtain inter-electrode impedances of less than 5 kOhms, in a few cases impedances were between 5 and 10 kOhms. The electroencephalogram (EEG) was amplified 50000 times and band-pass filtered between 30 and 300 Hz (6 dB/octave). The MASTER software (John & Picton, 2000) was used to generate the stimuli and record the electrical responses. In the digitizing step, an AD conversion rate of 1000 or 1250 Hz was used. Data were recorded in epochs that contain 1024 data points. Artifact rejection limits were set in such a way that about 5 to 10% of the epochs was rejected, in order to eliminate potentials due to muscle or movement artifacts. This corresponded to limits between 15 and 20 μV. Epochs were linked together to form a sweep, which lasted for 16.38 seconds. For each stimulus intensity, 8 to 48 EEG recording sweeps were averaged. Response waveforms were added in the time domain and the result was submitted to FFT analysis. The level of significance of the responses was monitored after each sweep. The probability that the amplitude of the signal is within the distribution of the noise amplitudes of 120 neighboring frequency bins (approximately 3.7 Hz on both sides) was evaluated using F-ratio statistics (John & Picton, 2000).

The AUDERA system

Individual stimuli were presented monaurally. The eight separate signals were calibrated using a Brüel & Kjaer Sound Level Meter 2260 in combination with an artificial ear 4152 and a 2-cc coupler DB0138. The EEG was monitored using silver-silver chloride electrodes. Impedances were less than 5 kOhms, with the exception of a few cases where impedances were between 5 and 10 kOhms. The system incorporates a noise threshold warning that depends on the modulation frequency in use and that is defined in the test protocol. Two different test protocols were used. If possible the protocol ' >10 years ASLEEP' was applied, which used high modulation rates and a low noise criterion. In case the EEG noise level was too high and recordings were all characterized as 'noise', the protocol ' >10 years AWAKE' was selected, which included a low modulation rate and a high noise threshold. Three types of results can occur.

Table 1. Comparison of test parameters for the MASTER and AUDERA techniques

	MASTER	*AUDERA*
Software version	MASTER 1d	SSEP GSI version 2
Stimulus presentation	Multiple-frequency, binaural	Single-frequency, monaural
AM/FM	100% AM, 20% FM	100% AM, 20% FM
Modulation frequencies	Left 82, 90, 98, 106 Hz for 500, 1000, 2000, 4000 Hz	Test protocol ' > 10 years ASLEEP': Left and right 74, 81, 88, 95 Hz for 500, 1000, 2000, 4000 Hz
	Right 86, 94, 102, 110 Hz for 500, 1000, 2000, 4000 Hz	Test protocol ' >10 years AWAKE': Left and right 46 Hz for all carriers
Transducers	Insert earphones (ER-3A)	Insert earphones (GSI TIP-50)
Calibration	dBSPL	dBHL
Maximum level	100 dBSPL	100 dBSPL (or 94.5, 100, 97 and 94.5 dBHL)
Starting level	50 dBSPL for NH and 70 dBSPL for HI	50 dBHL for NH and 70 dBHL for HI
Trial duration	2.5–15 minutes	40–107 seconds
Response detection	F-test ($p < 0.05$)	Phase-coherence ($p < 0.01$)
Electrode position	Vertex, inion, Pz (common)	Left and right mastoids, high and low forehead (common)

AM, amplitude-modulation; FM, frequency-modulation; NH, normal-hearing; HI, hearing-impaired. The amount of frequency modulation is defined as the difference between the maximum and minimum frequencies divided by the carrier frequency.

A *noise result* occurs when no response is found after 64 samples and when the EEG exceeds the noise threshold limit. A *random result* occurs when no response is found and the EEG does not exceed the noise threshold level. If a significant response is found, a *phase-locked result* occurs, regardless of the noise level.

The presence or absence of a response was determined automatically using a detection criterion which looked for non-random phase behavior. This was equivalent to the phase coherence technique described by Jerger et al (1986) and Stapells et al (1987), and as further described in Cohen et al (1991) and Rance et al (1995). Calculations are performed on each EEG sample. Up to 64 samples are analyzed for each trial, with a trial being defined as a tone frequency-intensity combination. In each EEG sample, the magnitude and phase of the EEG activity corresponding to the tone modulation frequency is quantified. A phase-locked or random response is determined on the basis of statistical analysis. The analysis algorithm will automatically halt stimulation and data sampling when the probability level $p < 0.01$ is achieved or after a maximum of 64 samples.

Subjects

Ten normal-hearing (NH) and ten hearing-impaired (HI) volunteers participated in the study. The NH subjects varied in age between 21 and 28 years. Hearing thresholds were better than or equal to 20 dBSPL at all octave frequencies between 500 and 4000 Hz. Mean Pure Tone Average (PTA) was 7 dBSPL. In the HI group, the ages varied between 18 to 73 years. These subjects had cochlear hearing loss and were selected from the patient population at the ENT-Department of the K.U.Leuven University Hospital. Hearing thresholds ranged from 10 to more than 110 dBSPL. The mean PTA was 63 dBSPL.

Experimental design

All experiments were carried out in a double-walled soundproof room with Faraday-cage. MASTER and AUDERA thresholds were obtained in separate test sessions. In the beginning of each test session, behavioral hearing thresholds (BHT) were determined in the same experimental conditions as for the ASSR, for MASTER and AUDERA separately, with insert phones and modulated sinusoids, at 5 dB accuracy with the Hughson-Westlake method. The later obtained ASSR thresholds are compared to the corresponding BHTs. After this behavioral test, the objective ASSR recordings were started. The subject was asked to lie down on a bed with eyes closed and to relax or sleep. Lights were switched off. ASSR thresholds were assessed for 500, 1000, 2000, and 4000 Hz in 10 dB steps. Recordings were started at an intensity level of 50 dBSPL/dBHL for the NH group and 70 dBSPL/dBHL for the HI group. The maximum presentation level was 100 dBSPL. A threshold was defined as the lowest intensity level at which a response was judged to be significant or when a phase-locked response was obtained. 'Noise' epochs or trials were excluded from all further evaluation. In total, 160 ASSR thresholds were compared to the corresponding BHTs for MASTER as well as for AUDERA. Difference scores (ASSR – BHT) are presented, in order to cancel out differences in calibration and in BHTs between test sessions.

Thresholds were calculated after different lengths of total test duration. For the MASTER system, recordings were carried out in multiples of 8 sweeps. If responses for all stimuli were significant after a multiple of 8 sweeps, recordings were stopped. If significance was not reached after averaging 48 sweeps a no-response was assumed for any one frequency. It was calculated what the thresholds for the different stimuli would be after recordings of maximum 16, 24, 32, 40, or 48 sweeps. A recording of 8 sweeps lasted a minimal 2.2 minutes, but taking into account the rejected epochs, this resulted in approximately 2.5 minutes. The maximum duration of a recording was 48 sweeps, which corresponded to about 15 minutes or more in case of excessive noise levels. Maximums of 32 and 16 sweeps were recorded at 90 and 100 dBSPL respectively, to avoid over-stimulation. Total test duration was calculated for different lengths of the individual recordings.

AUDERA thresholds were defined after short and long test sessions. In a short test session, only one EEG-recording with a good noise level was made for each tone frequency-intensity combination, resulting in a random or phase-locked response. In a long test session, recordings just below the threshold level obtained in the short test (thus with a random result) were repeated and in case of phase-locking intensity was lowered until a new threshold was defined. Thresholds for the long test were thus equal to or better than the thresholds obtained in the short test. As well for MASTER as for AUDERA, the total test duration does not include the time needed to prepare the patient, to place the electrodes and to find a comfortable position of the patient. Noisy measurements or rejected epochs during the threshold seeking procedure are included in the calculation of the total test duration. This is to assure a realistic estimate of the test duration.

Besides the comparison of ASSR and behavioral thresholds the test-retest reliability was assessed for both techniques. MASTER and AUDERA thresholds (and the corresponding behavioral thresholds) were retested in extra test sessions in three NH and three HI subjects in the exact same way as described above.

Statistical analyses

Statistical analyses were carried out with the 10.0 SPSS software. In a first step, the test duration was calculated for the different test procedures and the procedure with the most comparable total test duration for MASTER and AUDERA was sought. For this test procedure, difference scores (ASSR – BHT) were calculated. For both techniques the linear relationship between ASSR thresholds and BHTs was evaluated using Pearson correlation coefficients and linear regression analyses. Hearing threshold estimates will not be based on previously published data, since procedural and environmental variations may have an influence on the results. Predictions will be made based on the data of this study only. Frequency-specific regression equations will be used, because carrier frequency affects the ASSR response (John et al, 2002). The standard error of the estimate indicates how large the typical error is in predicting Y from X. It is the standard deviation of the expected values for the dependent variable. The R-squared describes the amount of variance of the dependent variable (behavioral threshold) that is accounted for by the independent variable (ASSR threshold).

Paired samples t-tests were applied to compare MASTER and AUDERA concerning difference scores (ASSR – BHT) and test duration. Correlation coefficients for the relationship between the ASSR and BHTs in HI subjects were compared between AUDERA and MASTER with the Fisher's z_r transformation.

The influence of test duration was evaluated for the two techniques separately, by comparing the difference scores that resulted from the different test procedures.

The differences between the test and retest results were compared with a paired samples t-test. Test-retest reliability was assessed by calculating the variability of difference scores, measured as the within-subjects standard deviation of the difference scores (σ_w) with the formula

$$\sigma_w = \sqrt{\frac{1}{2}\sum_{i=1}^{n}\frac{(x_{i1} - x_{i2})^2}{n}}$$

where x_{i1} is the ith difference score (ASSR – BHT) of the first ASSR session, x_{i2} is the ith difference score of the second ASSR session (the retest) and n is the total number of difference scores to compare. For six subjects and eight threshold comparisons per subject, n is 48.

Results

BHTs measured with the MASTER and the AUDERA set-up for NH and HI subjects are reported in Figure 1. For a clear comparison AUDERA thresholds are corrected to dBSPL. Both measures are highly correlated ($r \geq 0.96$ for all frequencies). The average within-subject differences are within 2 dB for 500 and 1000 Hz and within 1 dB for 2000 and 4000 Hz. The overall within-subject difference is 0 dB with a standard deviation of 7 dB.

The average total test duration for the MASTER test with trials of maximum 16, 24, 32, 40 and 48 sweeps and for the long and short AUDERA test is given in Table 2. For AUDERA the test duration is shorter for the HI compared to the NH. In contrast, test duration with the MASTER system is considerably longer for the HI. However, none of these differences is statistically significant. For both approaches, the variation is larger in the HI group. The longer test duration and greater variability in HI subjects with the MASTER technique is a consequence of the multiple-frequency approach which takes more time in sloping audiogram configurations compared to the flat audiograms of the NH subjects, since a higher number of intensity steps needs to be tested.

Because a short test duration is important towards the clinical applicability of a technique, the short AUDERA test will be compared to the best matching MASTER condition. In the HI group, the correspondence is best for the 16 sweeps condition and in the NH group for the 24 sweeps condition. However, because it is not advisable to average results of different test protocols, the mean test duration of the total subject group was taken into account. The MASTER 24 sweeps condition corresponds best for the total subject group. The mean within-subject difference between the short AUDERA test and the MASTER 24 sweeps is 6 min, but this difference is statistically not significant ($p = 0.174$).

Short test duration

Each AUDERA test session was started with the ASLEEP protocol. In 8 NH and only 3 HI subjects, ASSR recordings could be carried out at the higher modulation rates. At least 4 recordings (a few minutes), which were characterized as noise, were carried out before switching over to the AWAKE protocol. These recordings were not included in the calculation of the total test duration in contrast with other noise measurements that were included.

On average, 35 trials were carried out for the AUDERA test to define 4 thresholds in both ears. This corresponds to 4 to 5 trials per frequency. MASTER thresholds were all defined at modulation rates between 80 and 110 Hz. A test duration of 45 to 55 minutes corresponds to 6 or 7 intensity steps lasting for 24 sweeps or approximately 7.5 minutes each.

For both systems, 160 pairs of BHTs and ASSR thresholds were compared. In instances where BHTs or ASSR responses were absent at the audiometric limits, these comparisons were not included in further data analysis. For both approaches, both BHTs and ASSR responses were absent in 3 cases. In 5 instances for AUDERA and in 9 instances for MASTER, BHTs were

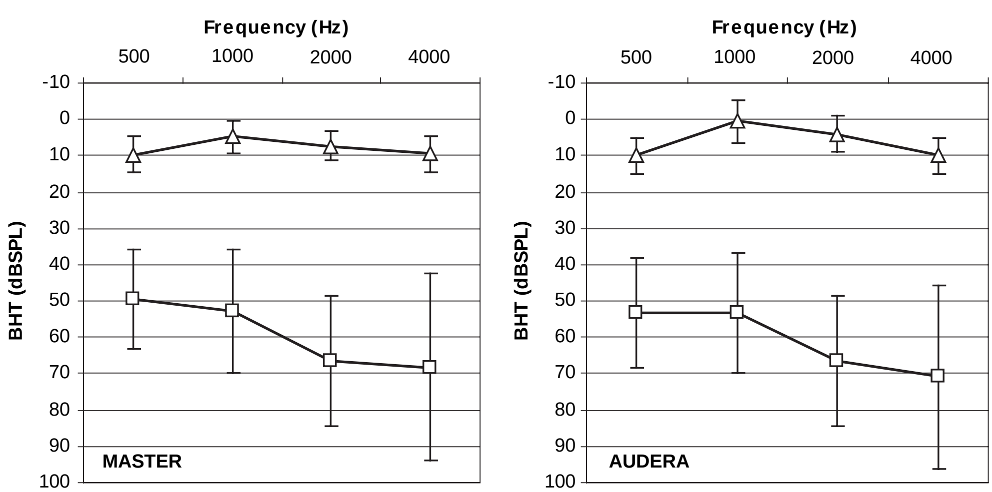

Figure 1. Average behavioral hearing thresholds and standard deviations in dBSPL for the normal-hearing and hearing-impaired subjects, measured in the beginning of the test session with the MASTER and the AUDERA set-up.

Table 2. Average total test duration, in minutes, for the MASTER test with trials of maximum 16, 24, 32, 40 or 48 sweeps and for the short and long AUDERA test. For the total subject group the short AUDERA test corresponds best with the 24 sweeps MASTER

	MASTER					*AUDERA*	
	16 sweeps	*24 sweeps*	*32 sweeps*	*40 sweeps*	*48 sweeps*	*short*	*long*
NH	32±7	45±10	58±13	69±17	79±19	46±12	61±10
HI	39±9	55±13	70±17	82±20	95±24	42±16	55±19
Total	36±9	50±12	64±16	76±19	87±22	44±14	58±15

present, but ASSR responses were absent. In these cases the BHTs were on average 86 dBSPL for AUDERA (range 69–96 dBSPL) and 88 dBSPL for MASTER (range 70–100 dBSPL). The elimination of these few cases with missing data (about 6% of the total) was not considered to have a significant effect on the overall findings.

The mean differences between the measured ASSR thresholds and the corresponding BHTs are reported in Table 3. The relationship between BHTs and ASSR thresholds for 20 NH and 20 HI ears is shown in Figure 2. The data have been fitted with regression lines. Regression equations are shown in Table 4, together with the correlation coefficient calculated at each of the carrier frequencies, R-squared and the standard error of the estimate. An overall correlation coefficient of 0.93 and 0.77 is obtained with the MASTER and AUDERA approach respectively. The formulae in Table 4 can be used to predict BHTs from ASSR thresholds. The distribution of the absolute behavioral prediction errors, obtained by subtracting the predicted BHT from the actual BHT, are given in Table 5.

For the total subject group, the raw difference scores of MASTER are on average 10±18 dB lower than those of AUDERA. This difference is highly significant (paired samples t-test, $p \leq 0.001$). The difference between the correlation coefficients was evaluated using the Fisher z_r transformation. Probabilities for the differences were 0.014, ≤ 0.001, 0.016, and 0.123 for 500, 1000, 2000, and 4000 Hz respectively, and $p \leq 0.001$ for the difference between the overall correlation coefficients. For the total subject group the MASTER technique outperforms the AUDERA technique.

By visual inspection of the regression plots and based on the results in Table 3, it is clear that there is a discrepancy between the AUDERA results of the NH and the HI group. For all frequencies the difference scores are significantly higher in the NH group than in the HI group (independent samples t-test, p always < 0.01). Moreover, variability of the difference scores is higher and data points are more scattered in the NH group. This deteriorates the results of the total subject group for the AUDERA approach. For the MASTER technique difference scores are not significantly different between NH and HI subjects for all frequencies (p always > 0.05).

Hearing-impaired subject group

Raw difference scores between ASSR thresholds and BHTs for the HI group, as given in Table 3, are not significantly different between MASTER and AUDERA (paired samples t-test, mean difference is 0±11 dB, $p = 0.957$).

Table 6 shows the regression equations based on the data of the HI group only. Frequency-specific correlations for AUDERA are all equal than, or slightly higher than, those for MASTER. The largest difference is apparent at 500 Hz. The data range at this frequency is rather small, which affects the correlation coefficient, especially for the MASTER data. The difference between the correlation coefficients was evaluated using the Fisher z_r transformation. Probabilities for the differences were 0.124, 0.745, 0.952, and 1.000 for 500, 1000, 2000 and 4000 Hz respectively. The overall correlation coefficient between the estimated thresholds and the real BHTs is 0.88 and 0.90 for MASTER and AUDERA respectively. These are also not significantly different ($p = 0.682$).

The behavioral prediction errors range from −18 to 34 dB for AUDERA and from −18 to 25 dB for MASTER, the standard error of the estimate is 8.4 dB for AUDERA and 8.6 dB for MASTER. The distribution of the absolute behavioral prediction errors in the HI group is given in Table 5, and is very similar for both techniques. More than 50% of the predicted hearing thresholds is within 5 dB of the real hearing threshold and 94% is within 15 dB. The mean absolute behavioral prediction error is 6±6 dB for both techniques.

Increase of test duration

MASTER thresholds were calculated for different maximum numbers of collected sweeps per intensity. Mean difference

Table 3. Mean difference scores and standard deviations (in dB) of the ASSR threshold (measured at 10 dB precision) and the corresponding behavioral threshold (measured at 5 dB precision) for 20 normal-hearing (NH) and 20 hearing-impaired (HI) ears, after a test duration of on average 50 and 44 minutes for MASTER and AUDERA respectively

		500 Hz	*1000 Hz*	*2000 kHz*	*4000 Hz*	*Total*
MASTER	NH	24±11	17±9	14±7	21±11	19±10
	HI	17±12	12±8	17±8	19±12	16±10
	Total	21±12	14±8	16±7	21±11	18±10
AUDERA	NH	48±21	40±21	33±10	30±20	38±20
	HI	20±8	14±7	13±7	14±13	15±9
	Total	34±21	27±20	24±13	23±19	27±19

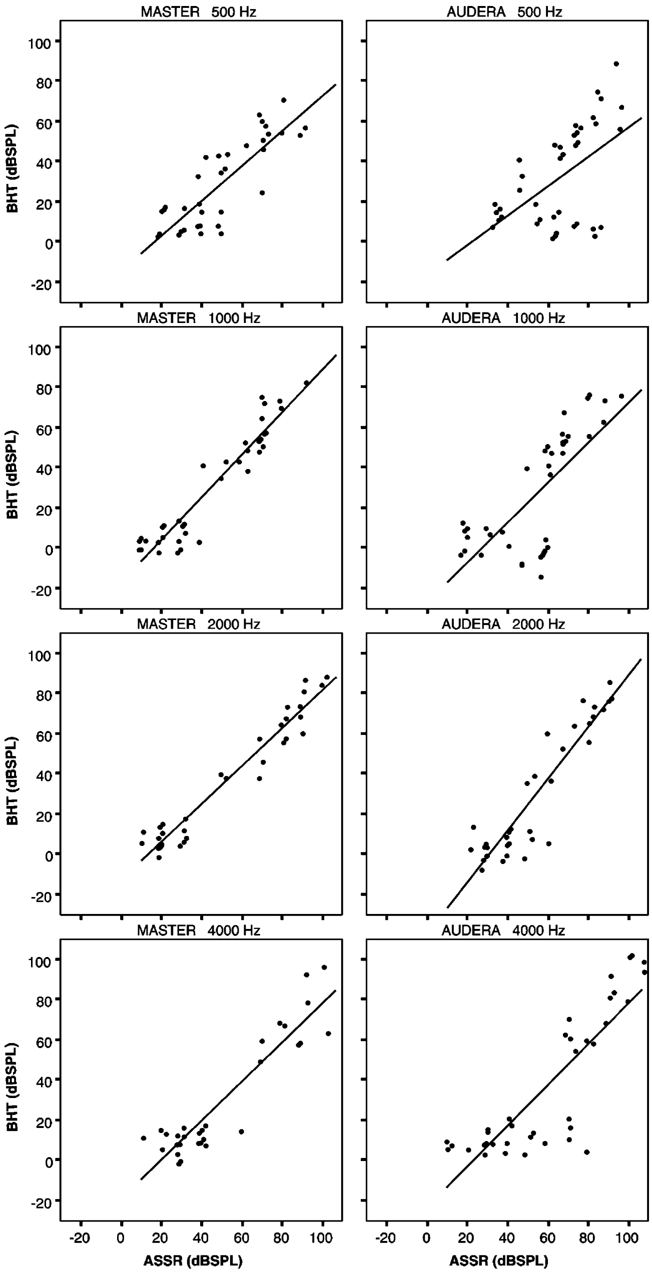

Figure 2. Scatter plots of the ASSR thresholds after test durations of on average 50 and 44 minutes as a function of the corresponding behaioral hearing thresholds BHT for 20 normal-hearing and 20 hearing-impaired ears per frequency. Points are jittered for 5% to show oerlapping of the data. Behaioral hearing thresholds and ASSR thresholds determined with the AUDERA set-up are corrected to dBSPL.

scores and standard deviations gradually decrease from 21 $\pm$ 10 dB after 16 sweeps to 15 $\pm$ 9 dB after 48 sweeps for the NH group and from 18 $\pm$ 11 dB to 13 $\pm$ 9 dB for the HI group. It was sometimes very difficult to prolong the AUDERA test, especially in the HI group, because noise levels started to increase and exceeded the noise criterion. Thus recordings were characterized as 'noisy'. In the NH group, however, this increase in test duration had a rather large effect, difference scores and standard deviations decrease. Averaged over the four frequencies, the mean difference scores decrease from 38 $\pm$ 20 dB to 33 $\pm$ 17 dB in the NH group and for 15 $\pm$ 9 dB to 14 $\pm$ 8 dB in the HI group.

Mean difference scores per frequency for different test protocols, separated for the NH and HI subject group, are depicted in Figure 3. The graph clearly shows the difference between MASTER and AUDERA for the NH subjects and the similar results obtained in the HI group.

Test-retest reliability

Test-retest reliability was assessed for AUDERA and MASTER. Difference scores (ASSR – BHT) of the test and retest were compared with a paired samples t-test. Table 7 shows the mean differences between the test and retest results and the associated standard deviations. As a reference, BHTs of the test and retest were compared. None of the differences were significant.

To define the reliability, the variability defined as the within-subjects standard deviation of the difference scores was calculated (see also Table 7). The reliability is higher for MASTER. However, for AUDERA there is a big difference between both subject groups. The variability is 13 dB in the NH group in contrast with 8 dB in the HI group. Moreover, there is a large effect of test duration, especially in the NH group. The variability decreases to 9 dB and 7 dB for the NH and HI group respectively. For MASTER, the reliability is comparable between the subject groups and also between the different lengths of recordings.

Discussion

In this study, a single-frequency and multiple-frequency ASSR approach were compared in similar test conditions. For the total subject group, which consisted of ten normal-hearing and ten hearing-impaired subjects, the MASTER approach predicted behavioral hearing thresholds with more accuracy than the AUDERA technique in the same amount of testing time. MASTER could predict hearing-thresholds in NH and HI subjects with a similar accuracy. For AUDERA, however, results were very different for both groups. Performance was better in HI subjects. When comparing the MASTER and AUDERA technique for HI subjects only, similar results were obtained. BHTs could be predicted based on the ASSR thresholds with a standard error of the estimate of 7 to 12 dB dependent on frequency. About 55% of the estimations was within 5 dB of the real behavioral threshold, 94% was within 15 dB. Clearly, the composition of the subject group had a big influence on ASSR results obtained with the AUDERA set-up. Outcomes obtained in NH subjects cannot always be used to predict performance in HI subjects, although this is often done in ASSR research. Additionally caution will have to be taken to use adult data to predict performance in young children.

Estimating hearing thresholds

Previous studies have investigated multiple-frequency ASSR thresholds in normal-hearing (Dimitrijevic et al, 2002; Herdman & Stapells, 2001; Perez-Abalo et al, 2001) and hearing-impaired adults (Dimitrijevic et al, 2002). Reported variability of the results (standard deviations of the difference scores) is comparable to the current data. The difference scores, however, are lower than in the present study. The lower difference scores in

Table 4. Comparison of the 24 sweeps MASTER and the short AUDERA test for the total subject group. Frequency-specific regression equations, Pearson correlation coefficients, R-squared and the standard error of the estimate are given

	Frequency (Hz)	*Regression equation*	*Correlation*	*R-squared*	*Std. error of the estimate (dB)*
MASTER	500	BHT = −14.29+0.87 * ASSR	0.83	0.69	12
	1000	BHT = −16.45+1.05 * ASSR	0.95	0.91	8
	2000	BHT = −12.63+0.94 * ASSR	0.97	0.95	7
	4000	BHT = −19.33+0.97 * ASSR	0.93	0.86	11
AUDERA	500	BHT = −17.21+0.73 * ASSR	0.54	0.29	21
	1000	BHT = −26.74+0.99 * ASSR	0.72	0.52	20
	2000	BHT = −39.28+1.28 * ASSR	0.92	0.85	12
	4000	BHT = −22.85+1.01 * ASSR	0.85	0.71	19

Perez-Abalo et al (2001) are probably the result of elevated behavioral hearing thresholds that are the consequence of the high environmental noise levels. In that study, behavioral hearing thresholds of normal-hearing subjects are on average 16 dB higher than in the current study. In the studies of Dimitrijevic et al (2002) and Herdman & Stapells (2001), test duration was considerably longer, which will have influenced the difference scores. The correlation coefficients in this study are very similar to those reported in Dimitrijevic et al (2002). The longer test duration apparently particularly affects the difference score, and to a smaller extent the variability of the data. This is also seen in the current study. Extending the test duration from 16 to 48 sweeps per trial in this study has changed the difference scores and standard deviations from 21±10 dB to 15±9 dB. For the multiple-frequency technique, difference scores are elevated for 500 and 4000 Hz. This was also reported in previous studies (Dimitrijevic et al, 2002; Herdman & Stapells, 2001).

As can be deduced from the studies of Rance and colleagues (Rance et al, 1995; Rance et al, 1998; Rance & Briggs, 2002), single-frequency ASSR results can vary substantially, dependent on the degree of hearing loss. Composition of the subject group will thus be a determining factor. Rance & Briggs (2002) compared behavioral hearing thresholds and ASSR thresholds in subjects with moderate to profound hearing loss. Correlation coefficients ranged from 0.81 to 0.93. This agrees well with the correlations found in the HI subjects of the present study. For the total subject group of normal-hearing and hearing-impaired subjects, correlation coefficients are lower in the present study. In Rance et al (1995), however, subjects with hearing thresholds ranging from normal to profound were tested and correlations ranged from 0.97 to 0.99. This is due to the relatively limited number of subjects with hearing thresholds below 10 dBHL in contrast to the normal-hearing group tested in the current study. Particularly in this group, ASSR thresholds measured with the AUDERA are extremely variable. Moreover, the very wide range of ASSR levels in Rance et al (1995), from approximately 20 to 120 dBHL, has also a positive effect on the correlation coefficient.

Table 5. Distribution of the absolute differences between the predicted hearing thresholds and the behavioral hearing thresholds for the total subject group (NH+HI) and for the hearing-impaired group alone (HI). Behavioral hearing thresholds are predicted based on the formulae in Table 4 for the total subject group and Table 6 for the hearing-impaired group

	NH+HI		*HI*	
	MASTER	*AUDERA*	*MASTER*	*AUDERA*
≤5dB	41%	25%	56%	54%
≤10dB	70%	42%	78%	81%
≤15dB	90%	59%	94%	94%
≤20dB	97%	72%	97%	99%
>25 dB	0%	19%	0%	1%

Different parameter settings

In this study, the total test duration of both techniques was kept approximately equal. Since the multiple-frequency technique is estimated to be two to three times faster than a single-frequency approach using the same MASTER system (John et al, 2002) and MASTER and AUDERA perform equally well in HI subjects within the same test duration, AUDERA appears to be relatively faster. For NH subjects, however, AUDERA performs worse. These differences between MASTER and AUDERA may, in part, be caused by different parameter settings. In general, the manufacturer's advice was followed as much as possible, since this was considered to be the optimal way to use the device and this is how it will be used by most clinicians who purchase the device. In this way, the total MASTER approach was compared to the total AUDERA approach. Besides the number of signals simultaneously presented, the techniques compared in this study also differ on other parameters, such as stimulus levels, modulation rate, electrode montage, response detection algorithm, and test duration per tone frequency-intensity combination.

Firstly, the AUDERA set-up was calibrated in dBHL by the manufacturer and the MASTER set-up was calibrated in dBSPL. Consequently, stimulus levels for both systems were not equivalent. However, according to the ISO 389-2 for insert phones, dBHL levels are within 0 to 5.5 dB of the dBSPL levels for frequencies between 500 and 4000 Hz. Moreover, ASSR thresholds were always compared to behavioral hearing thresholds in the same units (dBHL or dBSPL) and difference scores were calculated. In this way, issues related to calibration and to differences in hearing level at the time of testing were eliminated. Secondly, for AUDERA, an alternative test protocol was used in cases of excessive noise levels, which included a slower modulation rate and a higher noise criterion. This disparity in modulation rates complicates the interpretation of the results,

Table 6. Comparison of the 24 sweeps MASTER and the short AUDERA test for the hearing-impaired subject group. Frequency-specific regression equations, Pearson correlation coefficients, R-squared and the standard error of the estimate are given

	Frequency (Hz)	*Regression equation*	*Correlation*	*R-squared*	*Std. error of the estimate (dB)*
MASTER	500	BHT = −16.43+0.48 * ASSR	0.64	0.41	9
	1000	BHT = −17.12+1.08 * ASSR	0.89	0.80	8
	2000	BHT = −17.06+1.00 * ASSR	0.89	0.79	8
	4000	BHT = −27.17+1.10 * ASSR	0.88	0.78	12
AUDERA	500	BHT = −8.82+0.84 * ASSR	0.86	0.74	8
	1000	BHT = −7.39+0.90 * ASSR	0.91	0.83	7
	2000	BHT = −15.96+1.04 * ASSR	0.90	0.80	8
	4000	BHT = −28.10+1.20 * ASSR	0.88	0.78	12

since the modulation rate has an influence on the activated intracerebral generator (Herdman et al, 2002), on the size of the response. and effect of sleep or drowsiness (Cohen et al, 1991). However, the comparison of 40-Hz or 80-Hz modulation frequencies is beyond the scope of this study. The 40-Hz modulation rate was only used in case of noise levels that exceeded the noise criterion and this would bias the results of the comparison. Thirdly, the electrode montage was different for MASTER and AUDERA. In ASSR research, electrode positions typically used in case of monaural stimulation include the mastoid position, and in case of binaural stimulation electrodes are placed on the midline. According to van der Reijden and colleagues (2001) a significantly larger SNR was found for the Cz-inion derivation compared to the Cz-ipsilateral mastoid derivation for the 90-Hz modulation rate. Fourthly, the response detection method was different for both approaches. According to Picton et al (2001), the difference between detection protocols based on both phase and amplitude and phase alone is small, so this cannot explain the big difference between MASTER and AUDERA in the NH group. And finally, a very important factor to explain this difference is the test duration for each tone-frequency combination that was relatively larger for the MASTER approach. In HI ears, ASSRs above threshold are relatively larger compared to NH ears because of recruitment and thus faster to detect. Detecting ASSRs at threshold level in NH subjects, however, might require larger EEG samples and thus a longer test duration.

Advantages and disadvantages

Both techniques to determine hearing thresholds have shown strengths and weaknesses in functionality. The main differences are related to high EEG noise levels, sloping audiogram configurations, and testing subjects with small or no hearing-impairment.

High EEG noise levels

In case of restless patients and high EEG noise levels, the MASTER approach is most advantageous since recordings can be prolonged in order to decrease the noise level and increase the signal-to-noise ratio. In this way, it is always possible to carry out the recordings at modulation frequencies of 80 Hz, also in restless patients. The AUDERA technique allows repeating trials that are too noisy, but this does not influence the noise level or the quality of the measurement. This difference is not related to

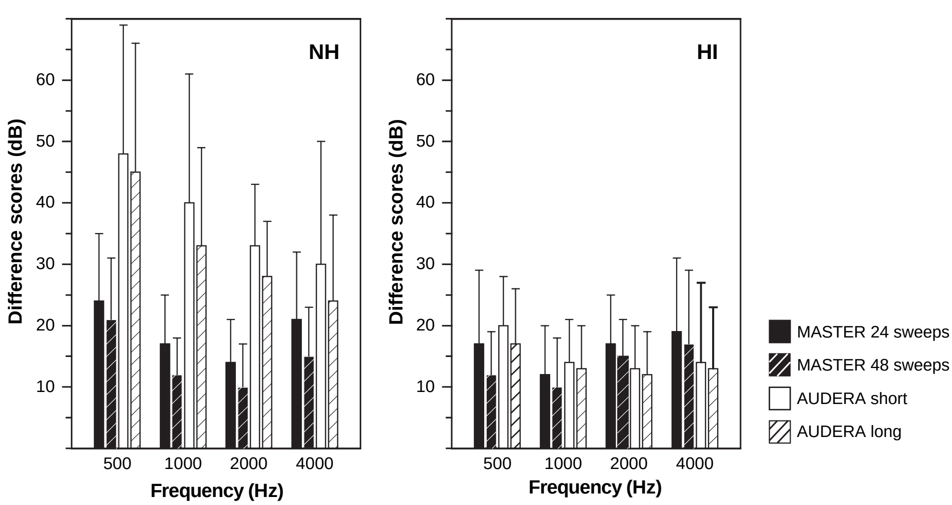

Figure 3. The influence of test duration. Bars show the mean difference between the ASSR threshold and the corresponding behavioral hearing threshold. Error bars show one standard deviation of the mean. MASTER is represented in black, AUDERA in white. Solid bars represent the 24 sweeps MASTER test and the short AUDERA test, which have a comparable test duration. Shaded bars represent the 48 sweeps MASTER test of on average 87 minutes and the long AUDERA test of approximately 60 minutes.

Table 7. Comparison of the test and retest of the 24 sweeps MASTER and the short AUDERA test. Difference scores (ASSR-BHT) of the test and retest are compared with a paired samples t-test. Variability of the difference scores is given

	Mean difference (dB)	*Sig. (2-tailed)*	*Variability (dB)*
ASSR MASTER	-1 ± 9	0.489	6
ASSR AUDERA	0 ± 16	0.901	11
BHT MASTER	1 ± 4	0.243	3
BHT AUDERA	0 ± 6	0.726	4

the single- or multiple-frequency approach, but rather to the way it is implemented in the software.

In this study, carried out in a double-walled soundproof Faraday cage room, it was often not possible to record at 80 Hz with AUDERA in hearing-impaired subjects (in 7 out of the 10 subjects) because the noise level exceeded the default noise criterion, even in subjects that were asleep. In these cases the 46 Hz modulation rate was applied. This could be a substantial problem when testing sleeping children since the detection of 40 Hz responses have been found to be affected by sleep (Cohen et al, 1991) and inconsistent in young children (Aoyagi et al, 1993; Kraus et al, 1985; Maurizi et al, 1990; Stapells et al, 1988). The noise criterion seemed too strict for this subject group. It might be advisable to adjust the noise criterion, although this could deteriorate the results.

Sloping audiogram configuration

An advantage of the single-frequency approach is that audiogram configuration has no influence on test duration. With a multiple-frequency approach, test duration is considerably lengthened in case of sloping audiogram configurations, when the intensity of frequencies in the stimulus cannot be adjusted separately. Firstly, it is often required to record the maximum number of sweeps per intensity in hearing-impaired subjects, since often one or more responses do not reach significance. And secondly, more recordings will have to be registered since the intensity range that has to be tested will be broader. Alternatively, by presenting different stimuli simultaneously, each frequency will be presented at more intensity steps than strictly needed, which can serve as an additional verification of the thresholds. This might positively effect the reliability.

Subjects with small or no hearing-impairment

Both techniques can accurately predict hearing thresholds in hearing-impaired adult subjects. In normal-hearing subjects, however, the AUDERA technique comes short. Since ASSR assessments are in the first place designed to use in hearing-impaired children, as a follow-up to general neonatal hearing screening, this might not seem to be a problem. However, false positive referrals have to be traced, and therefore it is important that also normal hearing can be assessed. Moreover the lower modulation frequencies that often had to be applied in this study are not appropriate to test children. In patients with low EEG noise levels, where the hearing loss was first diagnosed with e.g. ABR and normal hearing can be ruled out, the AUDERA can be used to accurately predict the audiogram.

Conclusions

Both approaches, MASTER and AUDERA, as implemented in commercial products, make it possible to accurately predict frequency-specific hearing thresholds in hearing-impaired adult patients within a clinically acceptable test duration. Hearing thresholds can be predicted with a standard error of the estimate between 7 and 12 dB dependent on frequency. For MASTER and AUDERA test-retest accuracy is high and performance is improved by prolonging the test duration. The AUDERA is less suited for testing subjects with normal hearing or limited hearing loss. Moreover, EEG noise levels often exceed the noise criterion of the 80 Hz test protocol of AUDERA. This could be a problem when testing sleeping children. The composition of the subject group has a big influence on the results of ASSR studies.

Acknowledgements

We gratefully acknowledge Dr. P. Lemkens of Ziekenhuis Oost-Limburg in Genk for putting part of the equipment at our disposal. We thank I. Van Frausum for her help with the experimental part of this study and K. De Voecht for selecting the hearing-impaired patients. This work is supported by the Fund for Scientific Research – Flanders (Belgium), through the projects 3.0168.95, G.0233.01 and G.0504.04.

References

Aoyagi, M., Kiren, T., Furuse, H., Fuse, T., Suzuki, Y., et al. 1994. Pure-tone threshold prediction by 80-Hz amplitude-modulation following response. *Acta Otolaryngol Suppl*, 511, 7–14.

Aoyagi, M., Kiren, T., Kim, Y., Suzuki, Y., Fuse, T., et al. 1993. Optimal modulation frequency for amplitude-modulation following response in young children during sleep. *Hear Res*, 65, 253–261.

Aoyagi, M., Suzuki, Y., Yokota, M., Furuse, H., Watanabe, T., et al. 1999. Reliability of 80-Hz amplitude-modulation-following response detected by phase coherence. *Audiol Neurootol*, 4, 28–37.

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.

Dimitrijevic, A., John, M.S., Van Roon, P., Purcell, D.W., Adamonis, J., et al. 2002. Estimating the audiogram using multiple auditory steady-state responses. *J Am Acad Audiol*, 13, 205–224.

Herdman, A.T., Lins, O., Van Roon, P., Stapells, D.R., Scherg, M., et al. 2002. Intracerebral sources of human auditory steady-state responses. *Brain Topography*, 15, 69–86.

Herdman, A.T. & Stapells, D.R. 2001. Thresholds determined using the monotic and dichotic multiple auditory steady-state response technique in normal-hearing subjects. *Scand Audiol*, 30, 41–49.

Jerger, J., Chmiel, R., Frost, J.D. & Coker, N. 1986. Effect of Sleep on the Auditory Steady-State Evoked-Potential. *Ear Hear*, 7, 240–245.

John, M.S. & Picton, T.W. 2000. MASTER: a Windows program for recording multiple auditory steady-state responses. *Comput Methods Programs Biomed*, 61, 125–150.

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

Kraus, N., Smith, D.I., Reed, N.L., Stein, L.K. & Cartee, C. 1985. Auditory Middle Latency Responses in Children - Effects of Age and Diagnostic Category. *Electroencephalogr Clin Neurophysiol*, 62, 343–351.

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

Luts, H., Desloovere, C., Kumar, A., Vandermeersch, E. & Wouters, J. 2004. Objective assessment of frequency-specific hearing thresholds in babies. *Int J Pediatr Otorhinolaryngol*, 68, 915–926.

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.
Maiste, A. & Picton, T. 1989. Human auditory evoked potentials to frequency-modulated tones. *Ear Hear*, 10, 153–160.
Maurizi, M., Almadori, G., Paludetti, G., Ottaviani, F., Rosignoli, M., et al. 1990. 40-Hz steady-state responses in newborns and in children. *Audiology*, 29, 322–328.
Perez-Abalo, M.C., Savio, G., Torres, A., Martin, V., Rodriguez, E., et al. 2001. Steady state responses to multiple amplitude-modulated tones: an optimized method to test frequency-specific thresholds in hearing- impaired children and normal-hearing subjects. *Ear Hear*, 22, 200–211.
Picton, T.W., Dimitrijevic, A., John, M.S. & Van Roon, P. 2001. The use of phase in the detection of auditory steady-state responses. *Clin Neurophysiol*, 112, 1698–1711.
Plourde, G. & Picton, T.W. 1990. Human auditory steady-state response during general anesthesia. *Anesth Analg*, 71, 460–468.
Rance, G. & Briggs, R.J. 2002. Assessment of hearing in infants with moderate to profound impairment: the Melbourne experience with auditory steady-state evoked potential testing. *Ann Otol Rhinol Laryngol Suppl*, 189, 22–28.
Rance, G., Dowell, R.C., Rickards, F.W., Beer, D.E. & Clark, G.M. 1998. Steady-state evoked potential and behavioral hearing thresholds in a group of children with absent click-evoked auditory brain stem response. *Ear Hear*, 19, 48–61.
Rance, G., Rickards, F.W., Cohen, L.T., De Vidi, S. & Clark, G.M. 1995. The automated prediction of hearing thresholds in sleeping subjects using auditory steady-state evoked potentials. *Ear Hear*, 16, 499–507.
Rickards, F.W., Tan, L.E., Cohen, L.T., Wilson, O.J., Drew, J.H., et al. 1994. Auditory steady-state evoked potential in newborns. *Br J Audiol*, 28, 327–337.
Stapells, D.R., Galambos, R., Costello, J.A. & Makeig, S. 1988. Inconsistency of auditory middle latency and steady-state responses in infants. *Electroencephalogr Clin Neurophysiol*, 71, 289–295.
Stapells, D.R., Linden, D., Suffield, J.B., Hamel, G. & Picton, T.W. 1984. Human auditory steady state potentials. *Ear Hear*, 5, 105–113.
Stapells, D.R., Makeig, S. & Galambos, R. 1987. Auditory Steady-State Responses - Threshold Prediction Using Phase Coherence. *Electroencephalogr Clin Neurophysiol*, 67, 260–270.
Stapells, D.R. & Oates, P. 1997. Estimation of the pure-tone audiogram by the auditory brainstem response: a review. *Audiol Neurootol*, 2, 257–280.
van der Reijden, C.S., Mens, L.H. & Snik, A.F. 2001. Comparing signal to noise ratios of amplitude modulation following responses from four EEG derivations in awake normally hearing adults. *Audiology*, 40, 202–207.