Objective Detection of Auditory Steady-State Responses: Comparison of One-Sample and q-Sample Tests

Mario Cebulla*
Ekkehard Stürzebecher†
Claus Elberling‡

Abstract

Auditory steady-state responses (ASSR) are expected to be useful for the objective, frequency-specific assessment of hearing thresholds in small children. To detect ASSR close to the hearing threshold, a powerful statistical test has to be applied. At present, so-called one-sample tests are used. These tests only evaluate the phase, or the phase and amplitude, of the first harmonic, that is, the fundamental frequency. It is shown that higher harmonics with significant amplitudes are also contained in the ASSR spectrum. For this reason, statistical tests that only consider the first harmonic ignore a significant portion of the available information. The use of a q-sample test, which, in addition to the fundamental frequency, also includes higher harmonics in the detection leads to a better detection performance. The evaluation of test performance uses both detection rate and detection time.

Key Words: Amplitude-modulation following responses, auditory steady-state responses, comparison of statistical tests, statistical response detection

Sumario

Se supone que las respuestas auditivas de estado estable (ASSR) son útiles para la evaluación objetiva y con especificidad frecuencial de los umbrales auditivos en niños pequeños. Para detectar las ASSR cercanas al umbral de audición, una prueba estadística poderosa debe aplicarse. En el momento presente, se utilizan las llamadas pruebas de una muestra. Estas pruebas evalúan solo la fase, o la fase y la amplitud del primer armónico; esto es, la frecuencia fundamental. Se ha demostrado que armónicos más altos con amplitudes significativas están contenidos en el espectro de las ASSR. Por esta razón, las pruebas estadísticas que sólo consideran el primer armónico ignoran una porción significativa de la información disponible. El uso de una prueba de muestra “q”, la cuál, además de la frecuencia fundamental incluye armónicos superiores en la detección, conduce a mejores desempeños de detección. La evaluación del desempeño de la prueba utiliza tanto una tasa de detección como un tiempo de detección.

Palabras Clave: Respuestas de seguimiento de amplitud/modulación, respuestas auditivas de estado estable, comparación de pruebas estadísticas, detección de respuesta estadística

*ENT Clinic, Faculty of Medicine, University of Würzburg, Germany; †ENT Clinic, Faculty of Medicine, Johann Wolfgang Goethe-University, Frankfurt am Main, Germany; ‡Oticon, ‘Eriksholm’, Snekkersten, Denmark

Dr. Mario Cebulla, Universitätsklinikum Würzburg, Hals-Nasen-Ohrenklinik, Josef-Schneider-Str. 11, D-97080 Würzburg, Germany; Phone: ++49 931 20121743; Fax: ++49 931 20121369; E-mail: mario.cebulla@mail.uni-wuerzburg.de

Basically, there are two types of auditory steady-state responses (ASSR): ASSR can be recorded either in response to continuous tones that are amplitude or/and frequency modulated (Picton et al, 1987a; Cohen et al, 1991; Aoyagi et al, 1994; Lins et al, 1995; Lins and Picton, 1995; Rance et al, 1995; John et al, 1998) or in response to transient stimuli (Galambos et al, 1981; Stapells et al, 1984; Lenarz et al, 1986; Stürzebecher et al, 2003).

ASSR to amplitude-modulated tones are also referred to as amplitude modulation following responses (AMFR). ASSR are expected to be useful for the objective frequency-specific estimation of hearing thresholds in small children (Cone-Wesson et al, 2002; Rance and Rickards, 2002; Vander Werff et al, 2002). The response to the sinusoidally modulated tone is itself quasi-sinusoidal, that is, the main power of the response is represented in the frequency domain by a spectral component at the modulation frequency (MF)—called the "fundamental frequency."

For the objective ASSR detection in the frequency domain, mainly so-called one-sample tests are used. These tests only consider the spectral component of the fundamental frequency of the response (first harmonic) corresponding to the modulation frequency. Most often only the phase information of the first harmonic is used (Picton et al, 1987b; Stapells et al, 1987; Champlin, 1992; Aoyagi et al, 1993; Dobie and Wilson, 1994; Valdes et al, 1997; Cebulla et al, Wernecke, 2001). A well-known one-sample test is the Rayleigh test (Mardia, 1972), also called phase coherence test (PC) (Picton et al, 1987b). The Rayleigh test regards only the phase information. There is a close relationship between PC and the component synchrony measure (CSM) proposed by Fridman et al, 1984 (i.e., CSM = PC^2). Moore (1980) created a modification of the Rayleigh test, which in addition to the phase angles also takes into account the spectral amplitude information in form of the ranks of the amplitudes. Dobie and Wilson (1989) introduced the magnitude-squared coherence test (MSC) that considers spectral phases and amplitudes too. Another one-sample test that is applied in the widely used MASTER system (John et al, 1998) is the F test. This test procedure calculates an F ratio of the power in the signal frequency component of the first harmonic to the mean power in n adjacent noise components.

The performance of any test that uses the phase angles and the spectral amplitudes of the first harmonic can be expected to be higher than that of a test that uses information about the phases only.

It is known that the response waveform is not exactly sinusoidal (Tucci et al, 1990; Lins et al, 1995).

As a first order of approximation, a standard ABR evoked by a click or a tone burst can be described by a sawtooth-shaped waveform. The ASSR to amplitude-modulated stimuli recorded at high stimulation rates (in the range of 70–100/sec) and at levels down to about 20 dB nHL produces similar waveforms that are peaked and nonsinusoidal (see, e.g., Lins et al, 1995, figures 5 and 7). Such waveforms will contain energy not only at the fundamental frequency (corresponding to the inverse of the repetition rate or to the modulation frequency) but also at a significant number of harmonic frequencies or overtones (see, e.g., the discussion by Lins et al, 1995).

Therefore, the use of a test that, in addition to the first harmonic, includes higher harmonics in the detection seems to be more favorable. In the present report, this kind of statistical test is referred to as a q-sample test.

Since the ASSR amplitude is rather low, response detection close to the behavioral threshold is difficult. For this reason, stimuli that in comparison to sinusoidal modulation of a single frequency carrier excite a wider range on the basilar membrane (i.e., stimuli with envelopes containing more rapid changes) were introduced (Stürzebecher et al, 2001; John et al, 2002). Because the envelope of these stimuli clearly deviates from the sinus function, it can be expected that a greater part of the response power will be distributed to the higher harmonics than would normally be the case with sinusoidally modulated single-tone stimuli. Using these other stimuli, a q-sample test would be expected to be even more effective than a one-sample test.

In a previous article (Cebulla et al, 2001), we applied four different one-sample tests to a large sample of ASSR data and found the modified Rayleigh test (Moore, 1980) and the magnitude-squared coherence test (MSC) (Dobie and Wilson, 1989) to be the most powerful tests among the four tests.

In addition to the spectral phase, both

tests also included the spectral amplitudes—in the case of the modified Rayleigh test in form of the ranks of the amplitudes.

In another article (Stürzebecher et al, 1999a), the q-sample uniform scores test (Mardia, 1972) and a modified version of the q-sample uniform scores test (Stürzebecher et al, 1999b) were compared on the basis of real ABR data and data from Monte Carlo simulations. Mardia's q-sample uniform scores test only works with the ranks of the phases. The modified version of this test also included the ranks of the amplitudes. This new version was clearly superior to Mardia's original test. The fact that the ranks of measured variables are used instead of the measured variables themselves is advantageous insofar as the statistical test is a nonparametric one. However, the disadvantage is a loss of information caused by ranking.

For this reason, we proposed a further modification of Mardia's q-sample uniform scores test (Stürzebecher, 2001). This modification incorporates the actual phase values instead of their ranks. Table 1 lists the specified modifications and additional possible modifications, both for the Rayleigh test and the q-sample uniform scores test, in such a way that the level of information used for the decision "Response present/not present" increases as one moves down the table.

Additionally, the F test was also included in the present comparison.

On the basis of two large ASSR data samples, the aim of this study is to clarify which type of statistical test, a one-sample test or a q-sample test, is more efficient for ASSR detection.

In addition, this study also investigates whether test performance indeed increases with increasing information or whether it is more advantageous to use a nonparametric test with ranked data. The data of the first sample consist of responses from stimulation with sinusoidally modulated tones, whereas the data of the second sample consist of responses from stimulation with a nonsinusoidal envelope.

METHODS

Test Subjects and Recording Details

All data used in the present investigation stem from other experiments aimed at developing efficient stimuli for ASSR recording (Stürzebecher et al, 2001). In these experiments, a number of different stimuli were applied to normal-hearing and hearing-impaired subjects. This is the reason why the data pool consists of a mixture of ASSRs evoked by different stimuli.

Fifty-seven male and female adults, aged between 20 and 64 years, participated in the experiments mentioned above. Forty-six subjects had normal hearing with thresholds of 10 dB HL or better at 500–4000 Hz. Eleven subjects had a sensorineural hearing loss in the range from 30 dB HL to 65 dB HL for at least one of the four frequencies 500, 1000, 2000, and 4000 Hz (the reason for restricting the hearing loss was that the maximum stimulus level of the research equipment was limited to 95 dB HL). The subjects reclined comfortably on an examination couch in a soundproof and electrically shielded room. They were asked to relax and, if possible, to sleep during the examination. All stimuli

Table 1. Listing of the One-Sample Tests and q-Sample Tests Used in the Present Investigation

	used information		
	spectral phase	spectral amplitude	number of harmonics
Rayleigh test - RT	yes	no	1
mod. Rayleigh test V1 (Moore) - MRTV1	yes	ranks	1
mod. Rayleigh test V2 - MRTV2	yes	yes	1
F-test	no	yes	1
q-sample test (Mardia) - QST	ranks	no	6
mod. q-sample test V1 - MQSTV1	yes	no	6
mod. q-sample test V2 - MQSTV2	ranks	ranks	6
mod. q-sample test V3 - MQSTV3	yes	ranks	6
mod. q-sample test V4 - MQSTV4	yes	yes	6

Note: The table indicates in what form each test uses the information of the spectral phases and amplitudes.

were tested separately using a modulation frequency (MF) of 90 Hz. Whenever possible, all stimuli were applied with the carrier frequencies of 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz. A stimulus level of 30 dB nHL was chosen for the normal-hearing subjects and 30 dB SL for the hearing-impaired subjects.

The normal-hearing subjects took part in up to four recording sessions, each lasting less than one hour. Since not all hearing-impaired subjects were available for more than one recording session, it was not possible to obtain recordings with all stimuli from all of these subjects. Furthermore, not all four frequencies could be tested in all hearing-impaired subjects, because of the magnitude of the individual hearing loss.

From the large data sample of 1484 ASSR recordings, each lasting 102.4 seconds, two subsamples were drawn (the recordings obtained with the 500 Hz stimulus were omitted, see below under the subheading "Test Application and Choice of Critical Test Values"):

Subsample 1

The subsample 1 consists of 206 recordings in response to the usual stimulus for ASSR recording (one-carrier stimulus, amplitude modulated (AM1), amplitude modulation depth 100%).

Subsample 2

The subsample 2 consists of 432 recordings in response to multiple-carrier stimuli (Stürzebecher et al, 2001). A multiple-carrier stimulus consists of several amplitude-modulated carriers. All the carriers of a stimulus were modulated with the same MF. The frequency difference between consecutive carriers was one or two times that of the MF. Amplitude modulation depth was 100%. A multiple-carrier stimulus activates a wider region on the basilar membrane than the usual ASSR stimulus. Therefore, the response amplitude is somewhat higher. Four different multiple-carrier stimuli were applied:

AM2MF2: Two amplitude-modulated carriers; distance between the two carriers 2xMF.

AM3MF1: Three amplitude-modulated carriers; distance between consecutive carriers 1xMF.

AM3MF2: Three amplitude-modulated carriers; distance between consecutive carriers 2xMF.

AM5MF1: Five amplitude-modulated carriers; distance between consecutive carriers 1xMF.

Figure 1 shows the waveforms and frequency spectra of the usual one-carrier stimulus (AM1) in comparison to a multiple-carrier stimulus (AM3MF2). These data were used for the present assessment of statistical test performance. It can be assumed that differences in the test results are primarily related to differences in test effectiveness, since all the tests were applied to the same set of data.

Statistical Tests

The tests included in the investigations are summarized in Table 1. As already stated above, the particular sequence of the tests has been chosen in such a way that the tests incorporate increasingly more information for the response detection when moving from top to bottom. The first four tests are one-sample tests. The Rayleigh test only uses the spectral phases. Moore (1980) suggested a modification (V1) whereby, in addition to the phases, spectral amplitudes are also taken into account in the form of their ranks. We made a further modification (V2) that includes the amplitudes themselves, instead of the ranks, because ranking is associated with a loss of information. The F test neglects the phases but in addition to the amplitude of the spectral response component (first harmonic), the test uses the information about the amplitude of the background noise estimated from the adjacent spectral noise components.

The additional five tests in Table 1 are q-sample tests. The q-sample uniform-scores test (Mardia, 1972) only works with the ranked spectral phases. The possibility of gradually upgrading the test was also applied here. The version V1 continues to be restricted to incorporate the phases; however, instead of the ranks, we introduced the unranked phases. The version V2 includes the amplitudes, but both phases and amplitudes are ranked; Version V3 uses the unranked phases but the ranked amplitudes. Finally, version V4 uses all the available information from the unranked phases and amplitudes.

To our surprise, the tests with unranked amplitudes originally showed a loss of test performance compared to those where the spectral amplitudes were ranked. A weighting of the spectral amplitudes by the mean amplitude of the adjacent spectral noise components (30 noise components on each side of the response component) led to an increase in test performance. A small but insignificant increase in performance could also be observed for the tests with ranked amplitudes when the amplitudes were weighted before ranking.

Offline Data Processing

Each data segment of 102.4 seconds was divided into 100 time-epochs with a length of 1.024 seconds. For further calculations, the single epochs of all data sets were transformed into the frequency domain offline by a Fourier transformation. The resulting frequency resolution of the calculated spectra was about 0.976 Hz. Since it was ensured that each epoch contained only complete response periods, there was no splatter of response energy to the side frequencies (apart from that induced by the application of a rectangular windowing function). The resulting frequency spectra were stored on hard disc.

For the F test, each data segment of 102.4 seconds was divided into 25 time-epochs with a length of 4.096 seconds. The resulting frequency resolution of the calculated spectra was about 0.244 Hz. The noise was estimated from 60 spectral noise components (30 on each side of the response component).

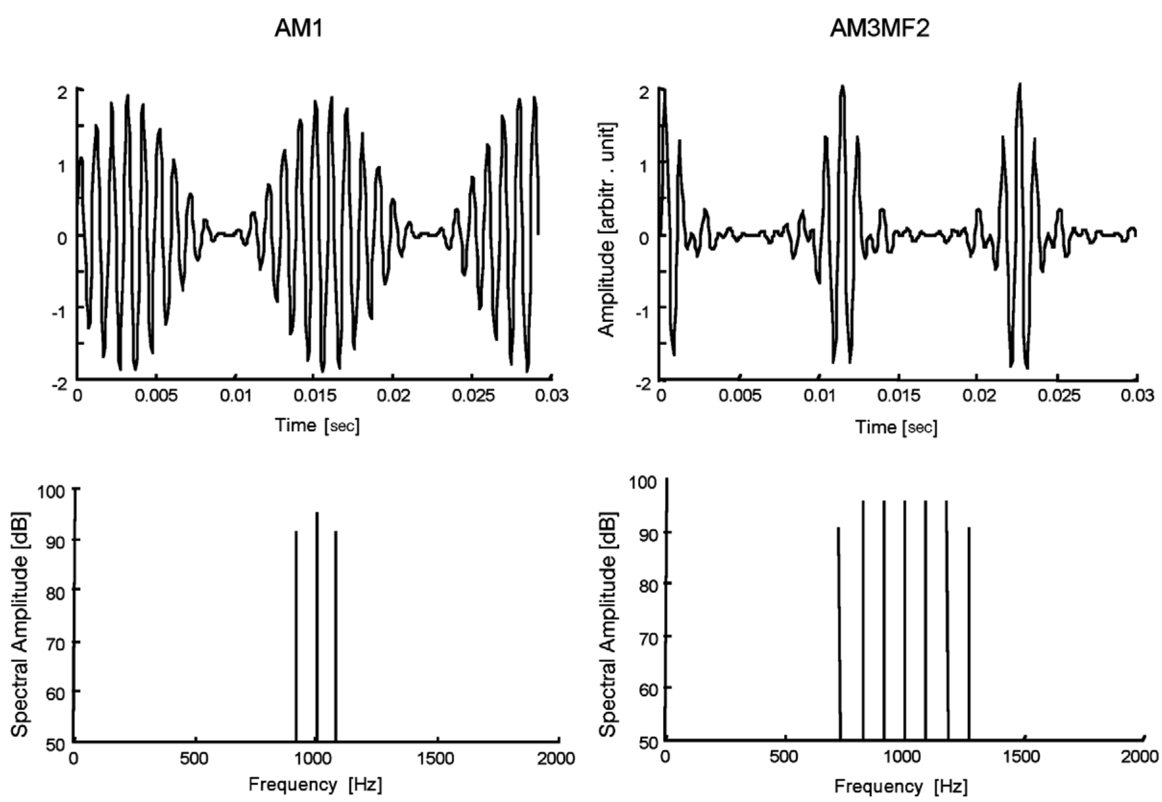

Figure 1. Time function and frequency spectrum of the stimulus (AM1) usually used to evoke an AMFR and of a multiple-carrier stimulus (AM3MF2).

Test Application and Choice of Critical Test Values

In the case of the q-sample tests, one has to decide how many harmonics should be included. Preliminary investigations have shown for frequency specific stimuli that the use of more than six harmonics has no practical benefit as the higher harmonics do not really contribute to the response detection because of their lower amplitude. For this reason, the q-sample tests were based on six harmonics. However, even six harmonics may be too many in the case of response detection at 500 Hz. Normally, the electrical stimulus artifact does not play a significant role because the carrier frequency of the stimulus is at quite a distance from the spectral components of the response. However, if a multicarrier stimulus of 500 Hz is used, the lower-frequency spectral components of the stimulus artifact overlap with the higher harmonics of the response at 500 Hz. This is the reason why the two subsamples do not contain any responses to 500 Hz stimuli. For the detection of 500 Hz response, a one-sample test has to be used.

A statistical test can either be applied to a sample with a predetermined fixed sample size or in a sequential manner (step-by-step enlargement of the number of epochs and testing at each step). In a practical application, a sequential test procedure will be more time efficient than using a fixed sample size because data collection can be stopped as soon as a response is detected. For this reason, the present study simulates online recording and testing in a sequential order. In the first step, the test was applied to the first ten epochs. Then the next epoch was added to the sample and the test was carried out again. In this way, the number of epochs was enlarged incrementally until a response had been detected or all 100 epochs (91 steps) had been included.

In our opinion, a significance level of $p = 0.01$ ($\alpha = 1$ %) is appropriate for hearing threshold assessment. Using the sequential test procedure described above, one has to consider the fact that multiple testing increases the selected significance level that, in turn, results in a higher proportion of false rejections of the null hypothesis (Lütkenhöner, 1991). One way to overcome this problem is to use a significance level corrected according to Bonferroni´s rule (Hochberg and Tamhane, 1987). However, this rule is too conservative for dependent data; this means that the level of significance α' corresponding to Bonferroni´s rule is lower than necessary in order to retain the predetermined level of significance α. For this reason, we have developed a procedure to assess the null hypothesis for the multiple testing of dependent data (Stürzebecher et al, 2005); this new procedure enables us to ascertain the critical test values for the sequential testing of dependent or partly dependent samples. The critical test value for $\alpha = 1\%$ had been determined according to this procedure for each of the statistical tests specified in Table 1 on the basis of 58,400 noise data sets drawn from the noise components of the 1484 ASSR data recordings.

Comparison of Test Performance

The response detection performance of the investigated tests was compared on the basis of the ASSR detection rate and the detection time. The statistical tests were applied to the spectra of the 206 and 432 data sets of the two subsamples in the sequential manner described above. The detection time was calculated from the number of epochs necessary to reach response detection. In the case that no response could be detected with 100 (25) epochs, the maximum time of 104 seconds was used. Since the detection time is not normally distributed, the median of the distribution is used to characterize the detection performance of a statistical test. Additionally to detection rate and detection time, the overall performance of the different tests is characterized by a performance index (PI), which is calculated by PI = detection rate/detection time.

For the significance testing of the differences between the detection rates within and between the subsamples, McNemars χ^2 test was applied. The differences between the detection times were tested by the Wilcoxon matched pairs signed rank test (both tests: Siegel, 1956).

RESULTS AND DISCUSSION

For the two subsamples (subsample 1: ASSR recorded with usual single-carrier stimulus; subsample 2: ASSR recorded with multiple-carrier stimuli), a one-sample test (modified Rayleigh test V2) was applied successively to the first harmonic and the five higher harmonics of the ASSR spectra to demonstrate that apart from the fundamental frequency also useful information is contained in the higher harmonics—both in the case of the usual single-carrier stimulus and in the case of multicarrier stimuli. The obtained detection rates are presented in Figure 2.

Based on these results, the following conclusions can be made:

1. The ASSR spectrum contains higher harmonics, both in the case of the usual single-carrier stimulus and in the case of multicarrier stimuli. The application of a one-sample test to the higher harmonics leads to surprisingly high detection rates (e.g., for the second harmonic—53.5/32.1%). Therefore, the application of a q-sample test, which also incorporates the response information carried by the higher harmonics, can be expected to result in a higher detection rate in comparison to a one-sample test, which uses the first harmonic only.
2. For all six harmonics investigated, the response detection rate is higher for the responses recorded with multiple-carrier stimuli than for those recorded with the usual single-carrier stimulus (same stimulus level of 30 dB above threshold for both groups). For most of the higher harmonics, the difference between the detection rates of both groups is larger than in the case of the fundamental frequency. These results confirm that responses from multiple-carrier stimuli are of higher amplitude than those obtained with the usual amplitude-modulated single-carrier stimulus. As expected, the results also demonstrate the stronger contribution from higher harmonics to responses evoked by the multiple-carrier stimuli than to responses evoked by the usual single-carrier stimulus.
3. The inclusion of up to six harmonics seems advantageous when detecting responses to multiple-carrier stimuli.

The results of the comparison of the

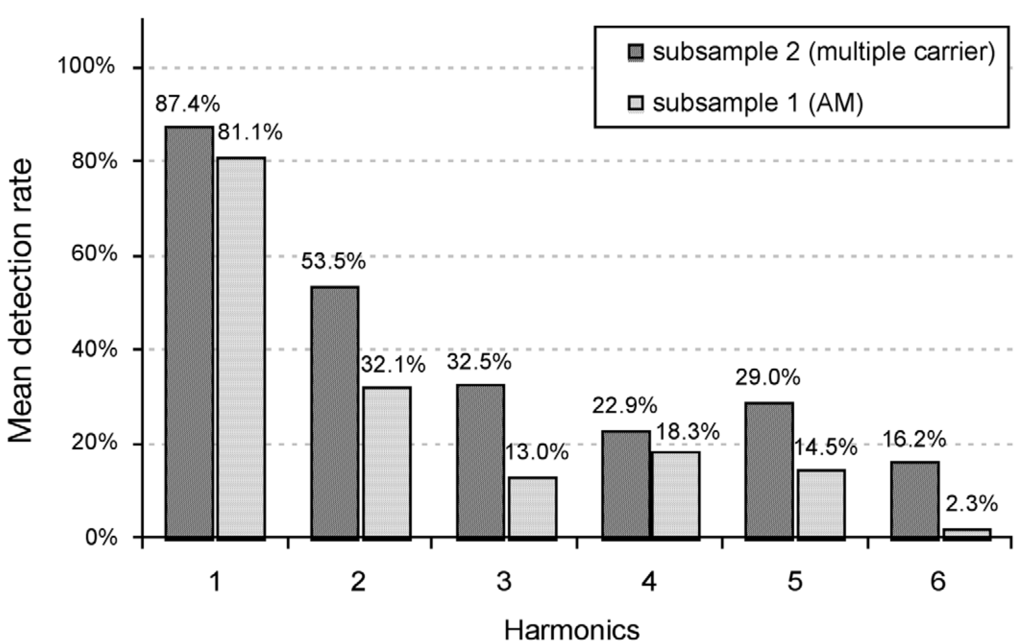

Figure 2. Response detection rate in the case of the application of a one-sample test (MRTV1) at each of the first six harmonics of the spectrum of the AMFR of the two subsamples. The AMFR of the first subsample (n = 206) were recorded with the usual amplitude-modulated one-carrier stimuli, and the AMFR of the second subsample (n = 432) with multiple-carrier stimuli.

performance of the different one-sample and q-sample tests are summarized in Table 2. Here the detection rate and the median detection time that resulted from the application of different statistical tests are listed together with the performance index PI. For all tests, the detection rate is significantly higher and the median detection time is significantly shorter with the multiple-carrier stimuli than with the usual AM stimuli. This advantage of the multiple-carrier stimuli is displayed very well by the performance index PI, which is always clearly larger for the multiple-carrier stimuli compared to the usual AM stimuli.

The results also demonstrate the expected order of precedence of performance for the one-sample tests and the q-sample tests. The more information about the response that is included, the higher the detection rate.

Detection rates of 74.8/83.7% are achieved with the Rayleigh test, which only makes use of phase information. The modification V1 (suggested by Moore, 1980) also includes spectral amplitudes in the form of their ranks. This leads to higher detection rates (79.6/85.8%). We could achieve a further increase in the detection rates up to 81.1/87.4% by introducing the spectral amplitudes themselves, instead of their ranks.

The detection performance of the F test (detection rates 79.6/85.1%) is similar to that of the modified Rayleigh test (proposed by Moore, 1980—MRTV1: 79.6/85.8%). As far as we are aware, Moore's modification of the Rayleigh test (MRTV1) has not been considered previously, and our modification of the Rayleigh test (MRTV2) is published here for the first time. Therefore, the F test and the Rayleigh test (RT) are those tests that most often have been used for response detection in the frequency domain. However, as shown here, with respect to detection rate, both tests were outperformed by the new one-sample test MRTV2. The difference between the detection rates of the F test and that of MRTV2 for subsample 2 is significant ($p = 0.032$). The same applies for the difference in the detection rate of MRTV2 and MRTV1 ($p = 0.026$).

Considering the overall performance characterized by the performance index PI, the rank of the performance of the F test among the one-sample tests differs for the usual AM stimuli and the multiple-carrier stimuli. Whereas for the usual AM stimuli the F test ranks below MRTV2, the F test is performing best among the one-sample tests for the multiple-carrier stimuli.

Comparing the detection performance of the one-sample tests and the q-sample tests, one has to consider those pairs that make use of the same level of information. That is, one has to compare the detection rates of the following pairs:

RT (74.8/83.7%) – MQSTV1 (78.6/82.3%)

MRTV1 (79.6/85.8%) – MQSTV3 (83.5/90.2%) ($p = 0.009/<0.001$)

MRTV2 (81.1/87.4%) – MQSTV4 (86.4/92.1%) ($p = 0.0007/<0.001$)

The first pair shows no clear difference in performance. However, for the two other pairs, significantly higher detection rates are

Table 2. Detection Rate, Median Detection Time and Performance Index (PI) in the Application of the Three One-Sample Tests and the Five q-Sample Tests to the AMFR Subsample 1 (n = 206) and Subsample 2 (n = 432)

		usual AM stimuli (n = 206)			multiple-carrier stimuli (n = 432)		
test	number of harmonics	detection rate [%]	median detection time [sec]	PI	detection rate [%]	median detection time [sec]	PI
RT	1	74.8	58	1.29	83.7	36	2.33
MRTV1	1	79.6	49	1.62	85.8	35	2.45
MRTV2	1	81.1	47	1.73	87.4	31	2.82
F-test	1	79.6	48	1.66	85.1	28	3.04
QST	6	76.2	58	1.31	77.4	40	1.94
MQSTV1	6	78.6	55	1.43	82.3	42	1.96
MQSTV2	6	79.1	46	1.72	85.1	34	2.50
MQSTV3	6	83.5	43	1.94	90.2	28	3.22
MQSTV4	6	86.4	42	2.06	92.1	28	3.29

demonstrated for the q-sample tests compared to the one-sample tests.

The amount of reduction in detection time and increase in detection rate are both dependent on the level of response information that is included by the one-sample and q-sample tests. However, looking at the results from the above third pair of tests, MRTV2 (47/31 sec) – MQSTV4 (42/28 sec), the reduction of the detection time by using a q-sample test instead of a one-sample test is not as impressive as the increase in the detection rate. Nevertheless, the difference between the detection times of both tests is significant with both subsamples (p = 0.0014/0.0002). More striking is the reduction in detection time when MQSTV4 is used to detect responses to the multiple-carrier stimuli (median: 28 sec) instead of the usual AM stimuli (median: 42 sec). The difference is highly significant (p = 0.001).

In some commercially available equipment, the usual AM stimuli are used for evoking the frequency-specific ASSR, and the simple Rayleigh test (RT) is applied to detect the responses. With the corresponding subsample used here, this leads to a detection rate of 74.8% and a median detection time of 58 sec. With the multiple-carrier stimuli and the most powerful q-sample test (MQSTV4), one gets a significantly higher detection rate (92.1%) and almost a halving of the detection time.

The cumulative frequency distributions shown in Figure 3 illustrate the advantage of using a q-sample test for the detection of responses evoked by multiple-carrier stimuli compared to applying the Rayleigh test to detect responses evoked by the usual amplitude-modulated tone. Apart from the higher detection rate, the figure shows, for example, that of all the 432 responses from subsample 2, 66% will be detected by MQSTV4 within 40 seconds. For detecting the same percentage of responses from subsample 1 by RT, one needs more than 80 seconds!

Even the cumulative distribution curve for MQSTV4 does not reach a 100% detection after 100 seconds, because only 398 responses of the 432 were detected. Obviously, the chosen maximum time of 100 seconds is too short for detecting all responses to a stimulus of 30 dB above threshold. In this sample, about 7.9% will be missed due to a low SNR. A longer maximum time is necessary for response detection at and below 30 dB re. threshold. Since a longer maximum test time can prolong the overall test time considerably, an optimization of the test time should be introduced by an intelligent controlling of the termination of the test run depending, for instance, on the trend of the test values in the course of the sequential testing.

For all tests reported here, the error

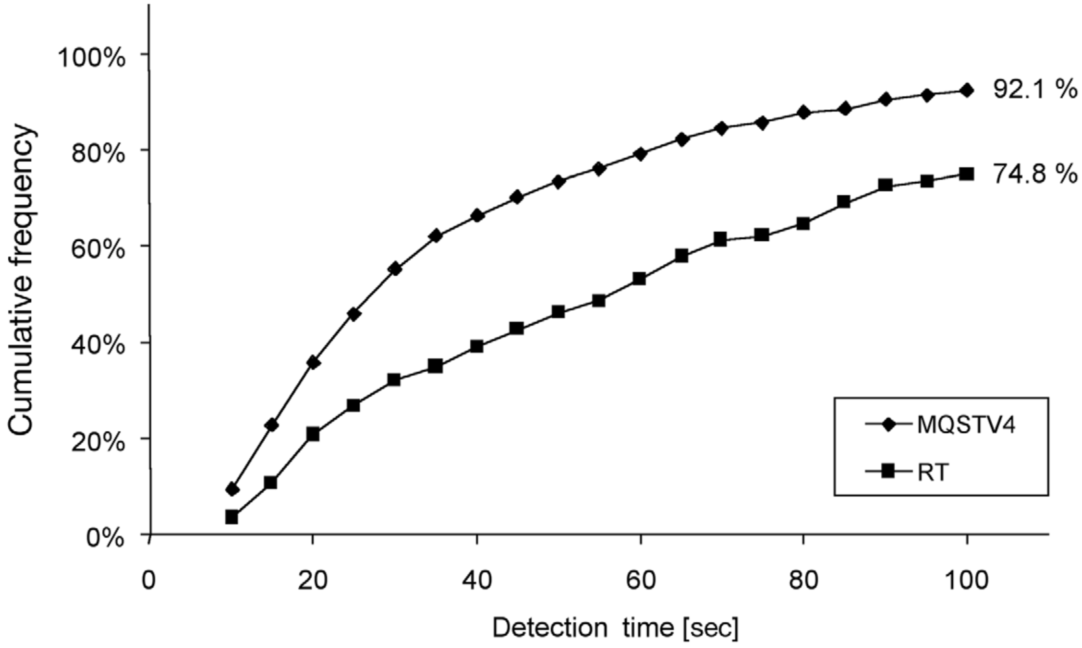

Figure 3. Cumulative frequency distribution of the detection times for the highest performing one-sample test (MRTV2, subsample 1) and the highest performing q-sample test (MQSTV4, subsample 2).

probability was $\alpha = 1\%$. This means that out of 100 test runs with no true response, the test will falsely detect a response in one case. For threshold assessment, this error rate is acceptable. In commercially available devices, one can also test with an error probability of $\alpha = 5\%$. This will of course increase the detection rate and decrease the detection time. With MQSTV4, for instance, there would be an increase in the detection rate from 92.1% to 95.6% and a decrease of the median detection time from 28 sec to 21 sec. However, in this situation and in the cases of no true response we thus have to accept a 5% false detection rate.

GENERAL DISCUSSION

A statistical test with the highest possible test power should be used for response detection, in addition to an optimized stimulus for generating the ASSR with the highest possible signal-to-noise ratio. In an earlier publication (Cebulla et al, 2001), we have shown that the modification of the Rayleigh test published by Moore (1980) (here MRTV1) was the most efficient among the one-sample tests, closely followed by the MSC test (Dobie and Wilson, 1994). In the first part of the present study, it was shown that the performance of both the one-sample and the q-sample tests improves with the amount of information that is included in the response detection.

Thus, a further modification of the Rayleigh test presented here (MRTV2), which uses unranked (but weighted) amplitudes, instead of the ranks of the spectral amplitudes as introduced by Moore (1980), leads to a further increase in the performance of the Rayleigh test. This test is a very good tool for response detection when restricted to the evaluation of the first harmonic.

According to the results from the present study, the ASSR is also represented by several relevant higher harmonics in the frequency domain. This is particularly the case if a multiple-carrier stimulus is used whose envelope deviates significantly from a sinusoidal shape. The higher detection rates of the responses to multiple-carrier stimuli demonstrated in Figures 2 and 3 and Table 2 indicate that with the multiple-carrier stimuli, the spectral SNR of the response components (harmonics) is larger than in the case of the usual amplitude-modulated single-carrier stimulus. As demonstrated here, the use of all available information (spectral amplitude and phase) from higher harmonics in addition to the first harmonic in a q-sample test leads to significantly higher detection rates and shorter detection times in comparison to a one-sample test, which uses the same information of the first harmonic only.

Even a well-performing statistical test together with efficient stimuli may not be sufficient for detecting a response close to the subjective hearing threshold if the following is not taken into account: we know from ABR recordings that a significant prerequisite for detecting a response close to the threshold is that the amplitude of the background noise caused by the spontaneous EEG of the patient is sufficiently low. For this reason, the patient must be comfortable, relaxed, and, if possible, asleep. This applies in particular to the ASSR, which has an even lower amplitude than that of the standard ABR. For this reason, a low amplitude of the spontaneous EEG is especially important for the detection of ASSR close to the threshold. Therefore, threshold assessment in babies and small children should preferably be carried out during natural sleep or in a condition of light sedation.

REFERENCES

Aoyagi M, Fuse T, Suzuki T, Kim Y, Koike Y. (1993) An application of phase spectral analysis to amplitude-modulation following response. *Acta Otolaryngol (Stockh) Suppl* 504:82–88.

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

Cebulla M, Stürzebecher E, Wernecke KD. (2001) Objective detection of the amplitude-modulation following response (AMFR). *Audiology* 40:245–252.

Champlin CA. (1992) Methods for detecting auditory steady-state potentials recorded from humans. *Hear Res* 58:63–69.

Cohen LT, Rickards FW, Clark GM. (1991) A comparison of steady-state evoked potentials to modulated tones in awake and sleeping subjects. *J Acoust Soc Am* 90:2467–2479.

Cone-Wesson B, Rickards F, Poulis C, Parker J, Tan L, Pollard J. (2002) The auditory steady-state response: clinical observations and applications in infants and children. *J Am Acad Audiol* 13:270–282.

Dobie RA, Wilson MJ. (1989) Analysis of auditory evoked potentials by magnitude-squared coherence. *Ear Hear* 10:2–13.

Dobie RA, Wilson MJ. (1994) Objective detection of 40 Hz auditory evoked potentials: phase coherence vs magnitude-squared coherence. *Electroencephalogr Clin Neurophysiol* 92:405–413.

Fridman J, Zapulla R, Bergelson M, Greenblatt E, Malis L, Morrell F, Hoeppner T. (1984) Application of phase spectral analysis for brain stem auditory evoked potential detection in normal subjects and patients with posterior fossa tumors. *Audiology* 23:99–113.

Galambos R, Makeig S, Talmachoff PJ. (1981) A 40-Hz auditory potential recorded in the human scalp. *Proc Natl Acad Sci* USA 78:2643–2647.

Hochberg Y, Tamhane AC. (1987) Multiple comparison procedures. New York: Wiley and Sons.

John MS, Dimitrijevic A, Picton TW. (2002) Auditory steady-state responses to exponential modulation envelopes. *Ear Hear* 23:106–117.

John MS, Lins OG, Boucher BL, Picton TW. (1998) Multiple auditory steady-state responses (MASTER): stimulus and recording parameters. *Audiology* 37:59–82.

Lenarz T, Gülzow J, Grözinger M, Hoth S. (1986) Clinical evaluation of 40-Hz middle-latency responses in adults: frequency specific threshold estimation and suprathreshold amplitude characteristics. *Otorhinolaryngol* 48:541–558.

Lins OG, Picton TW. (1995) Auditory steady-state responses to multiple simultaneous stimuli. *Electroencephalogr Clin Neurophysiol* 96:420–432.

Lins OG, Picton PE, Picton TW, Champagne SC, Durieux-Smith A. (1995) Auditory steady-state responses to tones amplitude-modulated at 80-110 Hz. *J Acoust Soc Am* 97:3051–3063.

Lütkenhöner B. (1991) Theoretical considerations on the detection of evoked responses by means of the Rayleigh test. *Acta Otolaryngol (Stockh) Suppl* 491:52–60.

Mardia KV. (1972) *Statistics of Directional Data*. London: Academic Press.

Moore BR. (1980) A modification of the Rayleigh test for vector data. *Biometrika* 67:175–180.

Picton TW, Skinner CR, Champagne SC, Kellett AJC, Maiste AC. (1987a) Potentials evoked by sinusoidal modulation of the amplitude or frequency of a tone. *J Acoust Soc Am* 82:165–178.

Picton TW, Vajsar J, Rodriguez R, Campbell KB. (1987b) 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 potentials. *J Am Acad Audiol* 13:236–245.

Rance G, Rickards FW, Cohen LT, De Vidi S, Clark GM. (1995) The automated prediction of hearing thresholds in sleeping subjects using auditory steady-state evoked potentials. *Ear Hear* 16:499–507.

Siegel S. (1956) *Nonparametric Statistics for the Behavioral Sciences*. London: McGraw-Hill.

Stapells DR, Linden D, Suffield JB, Hamel G, Picton TW. (1984) Human auditory steady-state potentials. *Ear Hear* 5:105–113.

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

Stürzebecher E. (2001) Method for hearing screening of newborns by means of steady-state response evoked with high click rate. European Patent application PCT/EP02/06262.

Stürzebecher E, Cebulla M, Baag M, Thie R. (1999b) Verfahren zur frequenzspezifischen Hörschwellenbestimmung mittels der Amplitude-Modulation Following Response (AMFR). European Patent application 199 54 666.5.

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, Neumann K. (2003) Click-evoked ABR at high stimulus repetition rates for newborn hearing screening. *Int J Audiol* 42:59–70.

Stürzebecher E, Cebulla M, Pschirrer, U. (2001) Efficient stimuli for recording of the amplitude-modulation following response. *Audiology* 40:63–68.

Stürzebecher E, Cebulla M, Wernecke KD. (1999a) Objective response detection in the frequency domain: comparison of several q-sample tests. *Audiol Neurootol* 4:2–11.

Tucci DL, Wilson MJ, Dobie RA. (1990) Coherence analysis of scalp responses to amplitude-modulated tones. *Acta Otolaryngol* (Stockh) 109:195–201.

Valdes JL, Perez-Abalo MC, Martin V, Savio G, Sierra C, Rodriguez E, Lins O. (1997) Comparison of statistical indicators for the automatic detection of 80 Hz auditory steady state responses. *Ear Hear* 18:420–429.

Vander Werff KR, Brown CJ, Gienapp BA, Schmidt Clay KM. (2002) Comparison of auditory steady-state response and auditory brainstem response thresholds in children. *J Am Acad Audiol* 13:227–235.