

Original Article

Comparing statistics for objective detection of transient and steady-state evoked responses in newborns

Eleina Mijares, María Cecilia Pérez Abalo, Didiesdle Herrera, Agustin Lage & Mayrim Vega

Cuban Neurosciences Center, Havana, Cuba

Abstract

Objective: Receiver operating characteristic (ROC) methodology was used to compare the performance of different statistics in the automatic detection of multiple auditory steady-state responses (MSSR) and click auditory brainstem responses (cABR). **Study sample:** Thirty-five healthy newborns tested within the first two weeks of birth. **Design:** In each case cABR and MSSR (0.5 and 2 kHz) were recorded with and without acoustic stimulation, using AUDIX equipment. With this test sample of recordings, ROC curves were evaluated separately for each statistic evaluated: (1) the standard deviation ratio (SDR) and the correlation coefficient ratio (CCR) for the cABR; (2) the Hotelling T2 (HT2) and circular T2 (CT2) for the MSSR. **Results:** All objective detection methods performed well (areas under ROC (AUC) > 0.9). The MSSR statistics showed significantly larger AUCs at both frequencies (HT2: 0.98 and 1; CT2: 0.96 and 0.99) than the cABR measures (SDR: 0.91 and CCR: 0.92). The HT2 hits rate was the highest (97–100% at 0.5 and 2 kHz) for fixed false alarms rates of both 10 and 20%. **Conclusions:** This superiority of performance of T2-like statistics, reflecting inherent advantages of MSSR analysis for automation, warrants serious consideration for further development of newborn screening technology.

Key Words: Auditory brainstem responses; steady state responses; ROC curves; hearing screening; signal detection methods; objective detection

Over the last decade a diversity of rapidly evolving automated screening devices has become available, facilitating worldwide expansion of the universal neonatal hearing screening program. Although this equipment using otoacoustic emissions (OAE) and/or click auditory brainstem responses (cABR) has been proven reasonably effective in practice (Mehl & Thomson, 1998; Vohr et al, 1998; Connolly et al, 2005) some improvements in methods and technology can be beneficial.

The accuracy of any automated screening device depends, among other factors, on the performance in the statistical methods used for objective response detection and its ability to discriminate the signal embedded in noise. Most automated screening technology nowadays uses transient responses such as cABR or OAE. Because of the complex morphology of these signals, it is not easy to quantify and to differentiate the response from the underlying noise using the existing time domain statistical methods. These limitations could be a biasing factor on the performance of available OAE or cABR automated equipments.

Over the past decade, the multiple auditory steady state responses (MSSR) have received increasing interest as a potential method for frequency specific screening (Savio et al, 2006; Savio & Pérez-Abalo,

2007). Among other advantages of this technique ease of automation is important. Since the MSSR are continuous periodic signals, they are represented in the frequency domain as distinct spectral peaks (at the modulation frequency) and can be described with only two parameters (amplitude and phase). Also, because the signals can be deterministically predicted (at specific spectral components), the residual noise can be easily differentiated and quantified. There is a diversity of statistical indicators calculated in the frequency domain which have been demonstrated to be effective for objective detection of MSSR (Picton et al, 1987; Victor & Mast, 1991; Valdés et al, 1997). For this reason, all MSSR technology currently available incorporates frequency domain statistical methods to detect the signals.

Despite the potential advantages for automation of MSSR testing over transient signals, such as cABR, few studies have directly compared the performance of the statistical methods used. In these investigations the receiver operating characteristic (ROC) methodology has been used to compare the performance of different quantitative indicators in cABR (Valdés et al, 1987; Dobie & Wilson, 1993; Sturzebecher & Cebulla, 1997; Sturzebecher et al, 1999; Cebulla et al, 2000; Norton et al, 2000; Savio et al, 2007) or MSSR

Abbreviations

AUC	Area under the ROC curve
cABR	Click auditory brainstem responses
CCR	Correlation coefficient ratio
CSM	Coherence synchrony measure
CT2	Circular T2
HT2	Hotelling T2
MSSR	Multiple auditory steady-state responses
OAE	Otoacoustic emissions
PC	Phase coherence
ROC	Receiver operating characteristics
SDR	Standard deviation ratio

(Dobie & Wilson, 1994 a,b; Valdés et al, 1997; Cebulla et al, 2001; Savio et al, 2007). There is only one previous report by Savio et al (2007) that compares the performance of both cABR and MSSR statistics with this methodology. The measures compared were those inbuilt in the AUDIX system: (1) the circular T2 (CT2) (Victor & Mast, 1991) and the Hotelling T2 (HT2) (Valdés et al, 1997) for MSSR signals; (2) the standard deviation ratio (SDR) and the correlation coefficient (CCR) for cABR (Picton et al, 1983; Valdés et al, 1987). Their findings showed some advantages for MSSR statistics over those used for cABR detection. However, the sample was obtained in a long term follow up clinical study which was not designed for this purpose (Savio et al, 2006). As a result the true positives were scarce and unevenly distributed. Also, the gold standard was the actual diagnosis of the children made at the end of the follow up (3–4 years old). Thus possible inter-current events between the first and second testing that may affect the results, especially in the true negative condition, cannot be ruled out.

The aim of this study is to confirm and expand to the context of neonatal hearing screening previous findings on the advantages of the frequency domain statistics for objective detection of MSSR (HT2 and CT2) versus those applicable to cABR (SDR and CCR). For this purpose, ROCs were evaluated in a test sample of recordings evenly distributed between conditions (with and without auditory response) obtained in healthy newborns.

Methods

Subjects

Thirty-five healthy newborns without risk factors were tested within the first two weeks of life. Ethical approval for the study was obtained through the Ethics Committee at the Cuban Neurosciences Center. A sample of 139 recordings (69 recordings with stimulus and 70 without stimulus) was collected for each type of response (cABR and MSSR) and ROC curves were evaluated. The cABR and MSSR recordings were obtained in the same babies during a single testing session.

Study design

Each baby underwent testing only if the parents consented after a detailed explanation of the procedures for a threshold seeking standard cABR test which was used as the gold standard for normal hearing. Threshold was identified by an experienced clinical neurophysiologist and only those babies with an identifiable peak V at 30 dB nHL were included in the study.

The cABR and the MSSR screening tests were subsequently administered in random order. Recordings without stimulation were also obtained for both screening tests with the earphones kept in place but disconnected from the equipment.

Acoustic stimuli

Brief clicks and multiple amplitude modulated tones were used as stimuli. All stimuli were delivered monaurally through insert earphones. Thus, both ears were tested, but independently. The clicks were generated with a rectangular pulse of 0.1 ms and presented at a repetition rate of 0.017 kHz. The multiple frequency stimuli consisted of a combination of two continuous carrier tones (0.5 and 2 kHz, mixed electronically) modulated in amplitude (95% depth) at rates of 0.111 and 0.115 kHz respectively. The intensity was expressed in dB nHL with reference to 0 dB to the mean psychoacoustic threshold of 10 young subjects with normal hearing function. The pass/fail intensity criteria were fixed for both screening tests at 40 dB nHL (equivalent to 75 dB pSPL for the click and 62 dB SPL (overall) for the mixed amplitude modulated tones). Stimuli were calibrated using a Brüel & Kjaer sound level meter model Investigator 2260 and a microphone type 4165.

Recording procedure

All newborns were tested under natural sleep. The MSSR and cABR were obtained using the AUDIX 5 evoked response test system (Neuronic S.A., Havana). Electrode discs of Ag/AgCl were fixed with electrolytic paste at the forehead (positive), ipsilateral mastoid (negative), and contralateral mastoid (ground). Impedance values were kept below 5 kOhms.

For the cABR recordings the bioelectric activity was amplified with a gain of 100 000 and bandpass filtered between 0.02 and 2 kHz. A fixed number of stimulus presentations—3000—were used to obtain the response average, acquired in two subsets of repetitions or “replicates”. Trials containing electroencephalographic signals with amplitudes greater than $\pm 50 \mu\text{V}$ from baseline were automatically rejected from the averaging process. The statistics SDR, CCR, and the residual noise levels calculated during the averaging process were used as quantitative indicators for response identification and quality control of the recordings.

The MSSR tests were made with a semi-automatic software application of AUDIX 5, designed specifically for neonatal screening (Mijares et al, 2011). For the MSSR the bioelectric activity was amplified with a gain of 100 000 and band pass filtered between 10 and 0.3 kHz. Epochs of 8192 samples (digitized with a sampling period of 1.36 ms) were collected, transformed to the frequency domain, and averaged. Artifact rejection was carried out with shorter epoch sections of 512 points whenever the activity exceeded $\pm 50 \mu\text{V}$. Fast Fourier transforms were computed with a long sweep of 8192 samples, as well as HT2 and, with each of the 16 subsegments of 512 samples, CT2. The MSSR software uses specific stopping criteria for the screening test: (1) When, for both frequencies, a significant response is present, stabilized in three consecutive sweeps, and the number of averages is greater than 8; (2) When one or both frequencies are not statistically significant and the maximum number of 32 averages was reached.

Calculation in the frequency domain of MSSR statistics

HOTELLING T2 STATISTICAL INDICATOR

The HT2 evaluates the difference between two estimated means value (signal vs. noise) using a coefficient of maximum verisimilitude.

For these indicators, it is assumed that the noise mean is different to zero and at the same time, is different to the signal mean. A response was judged as significantly different from the residual noise at a criterion of $p < 0.05$. The formula for HT2 was:

$$HT2 = (N-1) \times \frac{[\text{Re}(Y^s) - \text{Re}(\bar{Y}^r)]^2 + [\text{Im}(Y^s) - \text{Im}(\bar{Y}^r)]^2}{\sum_{i=1}^N [\text{Re}(Y_i^r) - \text{Re}(\bar{Y}^r)]^2 + [\text{Im}(Y_i^r) - \text{Im}(\bar{Y}^r)]^2}$$

where Re and Im are real and imaginary parts of the complex number, Y^s is the spectral component where the signal is expected, and Y_i^r are the adjacent frequency components used to estimate the noise.

CIRCULAR T2 STATISTICAL INDICATOR

To calculate the CT2 the amplitude and the phase of the signal spectral components were used following the exact methodology described by Victor & Mast (1991). This test evaluates if an observed set of spectral components (corresponding to the response) comes from the noise population (with zero mean). The statistic of this test was calculated as follows:

$$CT2 = (N-1) \times \frac{\text{Re}(\bar{Y})^2 + \text{Im}(\bar{Y})^2}{\sum_{i=1}^N [\text{Re}(Y_i) - \text{Re}(\bar{Y})]^2 + [\text{Im}(Y_i) - \text{Im}(\bar{Y})]^2}$$

where Re and Im are real and imaginary parts of the complex number, \bar{Y} is the sample mean, and Y_i refers to one observation.

Calculation in the time domain of cABR statistics

STANDARD DEVIATION RATIO STATISTICAL INDICATOR

The SDR is a signal to noise ratio, as follows:

$$SDR = \frac{|X(t)|}{|R(t)|}$$

where $|X(t)|$ represents the norm of the total average vector (replicates averaged together) and $|R(t)|$ represents the norm of the plus minus reference average.

CORRELATION COEFFICIENT RATIO STATISTICAL INDICATOR

The CCR is a statistic that reflects the replicability between two sub-averages, computed as follows:

$$CCR = \frac{\sum_{t=1}^{np} Y_1(t) \times Y_2(t)}{|V1(t)| \times |V2(t)|}$$

where $Y_1(t)$ and $Y_2(t)$ represents the replicates or subaverage waveforms, $|V1(t)|$ and $|V2(t)|$ represents the norm of the vector, and np means the number of data points digitized in each waveform.

ROC analysis

The area under the ROC (AUC) was calculated using the nonparametric Wilcoxon two rank sum statistics. This method was preferred because it avoids distributional assumptions about the probability model of the data, without introducing an estimation bias (Hajian-Tilaki et al, 1997). The method, briefly consisted in the comparison between all possible pairs of subjects in the true positives p_k subset and the true negatives subset q_j . The method to compute the area is based on the average of the assignment function in data.

The assignment function $\Psi(q_j, p_k)$ is equal to 1 if $q_j < p_k$, is equal to 0.5 if $q_j = p_k$, and is 0 when $q_j > p_k$. Then the overall area is computed by the following equation, which is an unbiased estimator of the area:

$$\hat{\theta} = \frac{\sum_j \sum_k \Psi(p_k, q_j)}{m n}$$

To compare the areas between ROC curves, the hypothesis is tested that both curves have equal area. The test of DeLong et al (1988) was applied, which is based in a ratio statistic that measures the differences between both areas. This non-parametric comparison is used for preventing departures of normality in the studied data.

Results also were scrutinized in terms of correct detections, for two fixed values of false alarm rates – 10% and 20%, extracted from the ROC curve. The estimation of confidence intervals for these results was obtained using a bootstrap (Hastie et al, 2009). This technique constructs new artificial datasets that are used for compute new ROC curves. The variance of the new ROC curves is used to derive a confidence interval for the observed statistics. The construction of the new data sets is achieved by resampling, with replacement of the observations in the actual data, via an iterative process.

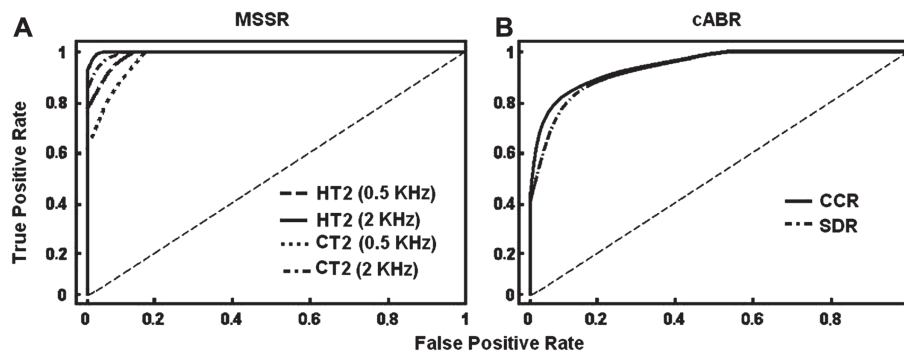


Figure 1. ROC curves summarizing the performance of the different statistical indicators. The y-axis corresponds to the true positive rate and the x-axis to the false positive rate, both of which are represented as proportions. Superimposed ROC curves for HT2 and CT2 statistics used to MSSR objective detection at 0.5 and 2 kHz are shown in the Figure 1, A. Figure 1, B shows ROC curves for SDR and CCR quantitative indicators used to cABR objective response detection.

Table 1. Comparison between AUC for each one statistical indicator used for the objective response detection of cABR and MSSR obtained at 0.5 & 2 kHz.

Test	Stimuli	Statistical tests			
		HT2	CT2	SDR	CCR
cABR	Click	–	–	0.91	0.92
MSSR	0.5 kHz tone	0.98	0.96	–	–
	2 kHz tone	1	0.99	–	–

Results

The ROC curves derived for cABR and MSSR statistics are shown in Figure 1. Each ROC curve represents the performance in automatic response detection of a single MSSR or cABR statistic. It is noteworthy that the ROC curves for both MSSR statistics (Figure 1, A) have larger AUCs than those for cABR statistics (Figure 1, B). The ROC curves more distant from the diagonal corresponded to the MSSR statistics obtained at 2 kHz, followed by the MSSR statistics obtained at 0.5 kHz, then those of both cABR statistics, which in turn overlapped.

The AUCs calculated are summarized in Table 1. It is noteworthy that the larger areas correspond to the MSSR statistics (HT2 and CT2) followed by those of the cABR (CCR and SDR). Furthermore, the AUCs for both MSSR statistics were slightly larger for the 2 kHz than 0.5 kHz responses. However, these differences did not reach statistical significance ($p < 0.19$ for HT2 and $p < 0.07$ for CT2).

Table 2 provides a summary of the results of the statistical comparisons between AUC using the DeLong et al (1988) method. For the MSSR elicited at 0.5 kHz the performance of the HT2 statistic was significantly better than the CT2 statistic. However, for the MSSR elicited at 2 kHz there were no significant differences in the performance of HT2 and CT2, though both MSSR statistics (HT2 obtained at 0.5 and 2 kHz; and CT2 obtained at 2 kHz) were significantly better than those used for cABR automatic detection.

The performance of each statistical indicator was also evaluated at false alarms rates of 10 and 20%. Either could be acceptable for screening applications. Table 3 provides the hits rates and confidence intervals calculated for each statistical indicator. For both false alarm rates there were clear advantages in the automatic detection of MSSR over cABR. The correct detection rates for MSSR statistics was between 90–100%, whereas cABR statistics could only identify correctly between 75% and 84% of the recordings at 10 and 20% false alarms rates, respectively.

Table 2. Statistical significance between AUC among all statistic tests studied for cABR and MSSR (0.5 and 2 kHz) objective response detection.

	CT2		SDR		CCR	
	0.5 kHz	2 kHz	0.5 kHz	2 kHz	0.5 kHz	2 kHz
HT2	0.009	0.138	0.008	0.000	0.010	0.000
CT2	–	–	0.077	0.000	0.090	0.000
SDR	–	–	–	–	0.211	0.211

Values statistically significant ($p < 0.05$) are shown in bold type.

Discussion

The performance of automatic detection methods for auditory evoked responses is important for hearing screening. This study uses ROC methodology to compare the performance of MSSR and cABR statistics in neonatal hearing screening. Results show significantly larger AUCs for MSSR statistics (HT2 and CT2) calculated in the AUDIX 5, than for the cABR indices SDR and CCR, suggesting an advantage of the MSSR technique for automation purposes.

ROC analyses were used successfully to compare the performance of various statistics in the automatic detection of either cABR (Sturzebecher, 1999; Norton et al, 2000; Cebulla, 2000) and MSSR (Valdés et al, 1997; Cebulla, 2001; Picton, 2001). However, there is only one previous report using such methodology to compare MSSR and cABR statistics (Savio et al, 2007), and there are some methodological differences between studies that should be discussed. In Savio et al (2007) the number of recordings classified with normal and impaired hearing was unevenly distributed, and the true positive represents 29% of the sample. Secondly, the gold standard used to classify the recordings was the audiological status of the child made 3–4 years after the screening test. Thus the influence of inter-current events between the initial and final testing cannot be ruled out.

Present findings confirm, via a more strictly defined test sample and a more rigorous design, the superiority of MSSR frequency domain statistics over those time domain statistics calculated for cABR detection, at least for the statistical analyses and method of recording as applied to the cABR here. Namely, the ROC curves were calculated in an equally distributed test sample of recordings obtained in healthy neonates evaluated in a single testing session. The importance of the balanced sample for the experimental design has been assessed by several authors (Montgomery, 2008; Bishop, 2006). The “balance” data set in which equal numbers of examples were selected from each of the classes permitted a more accurate

Table 3. Hits rate (correct detection of newborns without response) for a fixed false alarms rate of 10 and 20%, and its confidence intervals (in parentheses below) were computed with a bootstrap technique for cABR and for 0.5 & 2 kHz amplitude modulated tone steady-state responses.

False alarms	Hearing screening tests (% of successes)					
	cABR		0.5 kHz MSSR		2 kHz MSSR	
	CCR	SDR	CT2	HT2	CT2	HT2
10%	80 (78.9–80.4)	75 (74.7–76.0)	90 (89.0–90.0)	97 (95.8–98.3)	98 (97.5–99.5)	100 (98.7–100)
20%	84 (82.8–85.2)	84 (83.0–85.0)	96 (94.7–96.5)	97 (95.8–98.3)	100 (98.7–100)	100 (98.6–100)

model (Bishop, 2006). Thus, the limitations of previous approaches could be avoided, further substantiating the conclusion of advantages of MSSR technique for automated screening. Also, these results, obtained within directly in the context of neonatal screening, provides a first step for further development of MSSR screening technology.

The MSSR “superiority of performance” found here, and by Savio et al (2007), can be attributed to the periodic quasi-sinusoidal nature of the steady-state responses which, in turn, can be easily represented in the frequency domain and detected as distinct spectral components at the modulation frequency of each eliciting tone. Again, these properties facilitate not only signal identification but also its differentiation from the background noise. Contrastingly, transient signals such as cABR have complex waveforms which are more difficult to describe quantitatively and to differentiate from the overlapping noise.

The selection of the specific cABR indices used here and the relative advantages of one over the other also bear discussion. Among the many statistical indices used for automatic detection of transient evoked responses, the CCR and the SDR have interesting theoretical properties (Picton et al, 1983). These time domain signal-to-noise indices (or modifications derived from them) have been used successfully for objective detection of cABR (Picton et al, 1983; Valdés et al, 1987; Jimenez et al, 1988; Savio et al 2007). For this reason in the AUDIX system, these indices are calculated “on line” during the averaging process. They thus can be used as quantitative aids for response identification and quality control of the recordings. Published results comparing their performance have shown either a slight superiority of SDR over CCR (Valdés et al, 1987; Savio et al, 2007) or vice versa (Picton et al, 1983). Such differences were not evident from results of the present study, where both time domain indices showed similar AUCs and no statistical differences were found between them. Since differences between studies were rather small, these earlier differences in CCR and SDR performance likely reflect differences in the nature and size of the test sample and in the methods used for comparisons between indices.

A caveat apropos the findings of this study is that the statistics used for cABR detection were limited to those already included in the AUDIX system employed. Other time domain power ratio methods such as Fsp developed by Elberling & Don (1984), the frequency domain T2R developed and tested by Valdés et al (1987), or template matching methods (Jiménez et al, 1988) might be more effective in the detection of cABR. These indices should be also evaluated in further studies. However, considering that the AUC values reported for the automatic detection of cABR using either T2R or the Fsp are smaller than those found here for the T2 like statistics, it would seem that the MSSR advantage for automation likely would still hold.

The differences among frequency domain statistics for objective MSSR detection also should be evaluated further. Most published studies have evidenced a superiority in the performance of statistics based on both amplitude and phase information (Picton et al, 1987; Victor & Mast, 1991; Champlin, 1992; Dobie & Wilson, 1993, 1994 a,b; Sturzebecher & Cebulla, 1997; Valdés et al, 1997; Cebulla et al, 2001; Picton et al, 2001; Savio et al, 2007) over methods such as the coherence synchrony measure (CSM) or phase coherence (PC) relying on only the phase information (Victor & Mast, 1991; Dobie & Wilson, 1993; Sturzebecher & Cebulla, 1997; Champlin, 1992; Cebulla et al, 2001; Picton et al, 2001). The relative performance of CT2 and HT2 in the detection of 0.08 kHz steady state responses

though has been less studied (Valdés et al, 1997; Savio et al, 2007). Results here showing a slight superiority of HT2 over CT2 are in agreement with these reports. This could be explained by the more robust statistical properties of the HT2 (Valdés et al, 1997). First the assumptions for the noise are less restricted in the HT2. To calculate the CT2 an assumption is made about the noise (mean equal to zero) whereas the HT2 estimates the noise from actual spectral components at both sides of the signal. The noise however may be different from zero when there is a moderate contamination by recording artifacts, or in the presence of small jitters in the latency and amplitude of the averaged response (Valdés et al, 1997). Another statistical advantage of the HT2 resides in the use of more data points to estimate the variance (120 spectral components versus 16 samples of the signal frequency in the CT2). This will increase the degrees of freedom of the test and consequently enhances its accuracy (Valdés et al, 1997). Despite the statistical advantages of the HT2 pointed out by Valdés et al (1997), results of their study failed to reveal superiority in performance over CT2. This might be explained by the smaller sample size and the different statistical methods used to estimate and to compare the ROC curves.

In summary, although efficient technology and methods are available today for automated neonatal hearing screening, improvements still can be beneficial. In this endeavor choosing a statistical test for objective response detection and selecting an appropriate technique must be thoroughly investigated. An important consideration in this decision is naturally to ensure a performance yielding the smallest possible number of false positives and false negatives (Cebulla et al, 2001). In this regard, the present findings suggest the use of the MSSR HT2 as a strong candidate and certainly warrant further research and development for neonatal hearing screening.

Conclusions

These results confirm certain previous findings but further insights from the use of a more sizable test sample of recordings and a more rigorous design, apropos potential advantages of the MSSR for automated newborn screening. Based on ROC analysis both statistics examined—HT2 and CT2—proved to be more effective than the time domain statistics (SDR and CCR) employed in objective detection of the cABR. Also, there were some advantages found in the performance of the HT2 over the CT2 for MSSR objective detection. Overall, these results are promising and support the use of MSSR as the basis for further developed newborn screening technology.

Acknowledgements

The authors would like to thank Dr. Ross J. Roeser, Dr. John Durrant, and two anonymous reviewers for helpful comments.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- Bishop C.M. (2006). Introduction. In: M. Jordan, J. Kleinberg, S. Bernhard (eds.), *Pattern Recognition and Machine Learning*: 1st ed. New York: Springer, pp. 45.
- Cebulla M., Sturzebecher E. & Wernecke K.D. 2000. Objective detection of auditory brainstem potentials: Comparison of statistical test in the time and frequency domains. *Scand Audiol*, 29, 44–51.

- Cebulla M., Sturzebecher E. & Wernecke K.D. 2001. Objective detection of the amplitude modulation following response (AMFR). *Audiology*, 40(5), 245–252.
- Champlin C.A. 1992. Methods for detecting auditory steady-state potentials recorded from humans. *Hear Res*, 58, 63–69.
- Connolly J.L., Carron J.D. & Roark S.D. 2005. Universal newborn hearing screening: Are we achieving the Joint Committee on Infant Hearing (JCIH) objectives? *Laryngoscope*, 115, 232–236.
- DeLong E.R., DeLong D.M. & Clarke-Pearson D.L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837–845.
- Dobie R.A. & Wilson M.J. 1993. Objective response detection in the frequency domain. *Electroenceph Clin Neurophysiol*, 88, 516–524.
- Dobie R.A. & Wilson M.J. 1994a. Phase weighting: A method to improve objective detection of steady-state evoked potentials. *Hear Res*, 79, 94–98.
- Dobie R.A. & Wilson M.J. 1994b. Objective detection of 40-Hz auditory evoked potentials: Phase coherence vs. magnitude-squared coherence. *Electroenceph Clin Neurophysiol*, 92, 405–413.
- Elberling C. & Don M. 1984. Quality estimation of auditory brainstem responses. *Scand Audiol*, 13, 187–197.
- Hajian-Tilaki K.O., Hanley J.A., Joseph L. & Collet J.P. 1997. A comparison of parametric and nonparametric approaches to ROC analysis of quantitative diagnostic test. *Med Decis Making*, 17(1), 94–102.
- Hastie T., Tibshirani R. & Friedman J. 2009. *Model Inference and Averaging. The Elements of Statistical Learning*. Springer, pp. 261–293.
- Jimenez J., Carballo J., Riera J. & Pérez-Abalo M.C. 1988. Comparación de la eficiencia diagnóstica de diferentes medidas de normalidad para el PEATC. *Rev. Cub. Ciencias Biológicas*, 19, 139–141.
- Mehl A.L. & Thomson V. 1998. Newborn hearing screening. The great omission. *Pediatrics*, 101, 4.
- Mijares E., Herrera D., Santos E., Gaya J.A., Pérez-Abalo M.C. et al. 2011. Cribado auditivo neonatal con potenciales evocados auditivos de estado estable a múltiples frecuencias. *Acta Otorrinolaringológica Española*, 62 (2), 87–94.
- Montgomery D.C. 2008. Simple comparative experiments. In: J. Welter & J. Foxman (eds.) *Design and Analysis of Experiments*: 7th ed. John Wiley & Sons, pp. 25–27.
- Norton S.J., Gorga M.P. & Widen J.E. 2000. Identification of neonatal hearing impairment: Evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem response test performance. *Ear Hear*, 21, 508–528.
- Picton T.W., Dimitrijevic A., Sasha M. & Van Roon. 2001. The use of phase in the detection of auditory steady-state response. *Clin Neurophysiol*, 112, 1698–1711.
- Picton T.W., Linden R.D., Hamel G. & Maru J.T. 1983. Aspects of averaging. *Seminars in Hearing*, 4, 324–340.
- Picton T.W., Vajsar J., Rodriguez R. & Campbell K.B. 1987. Reliability estimates for steady-state evoked potentials. *Electroenceph Clin Neurophysiol*, 68, 119–131.
- Savio G., Mijares E., Pérez-Abalo M.C., Vega M., Lage C. et al. 2007. Eficiencia de la detección automática de potenciales evocados auditivos de estado estable a múltiples frecuencias evaluada mediante la metodología ROC. *Rev Logop Fon Audiol*, 27(1), 12–23.
- Savio G., Pérez-Abalo M.C., Gaya J., Hernández O. & Mijares E. 2006. Test accuracy and prognostic validity of multiple auditory steady-state responses for target hearing screening. *Int J Audiol*, 45, 109–120.
- Savio G. & Pérez-Abalo M.C. 2007. The auditory steady-state response: Generation, recording, and clinical application. In: G. Rance (ed.), *ASSRs and Hearing Screening*. San Diego: Plural Publishing. pp. 185–199.
- Sturzebecher E. & Cebulla M. 1997. Objective detection of auditory evoked potentials: Comparison of several statistical tests in the frequency domain on the basis of near threshold ABR data. *Scand Audiol*, 26, 7–14.
- Sturzebecher E., Cebulla M. & Wernecke K.D. 1999. Objective response detection in the frequency domain: comparison of several q-sample tests. *Audiology Neurotology*, 4, 2–11.
- Valdés J.L., Pérez-Abalo M.C., Martín V., Savio G., Sierra C. et al. 1997. Comparison of statistical indicators for the automatic detection of 80 Hz auditory steady state response. *Ear Hear*, 18, 420–429.
- Valdés M., Bobes M., Pérez-Abalo M.C., Perera M., Carballo J.A. & Valdés P. 1987. Comparison of auditory-evoked potential detection methods using signal detection theory. *Audiol*, 722, 166–178.
- Victor J.D. & Mast J. 1991. A new statistic for steady-state evoked potentials. *Electroenceph Clin Neurophysiol*, 78, 378–388.
- Vohr B.R., Carty L.M., Moore P.E. & Letourneau K. 1998. The Rhode Island hearing assessment program: Experience with state-wide hearing screening (1993–1996). *Pediatrics*, 133, 353–357.