On the cost of introducing speech-like properties to a stimulus for auditory steady-state response measurements

Søren Laugesen[1,*], Julia E. Rieck[1,2], Claus Elberling[3], Torsten Dau[4], and James M. Harte[1]

[1] *Interacoustics Research Unit, Kgs. Lyngby, Denmark*

[2] *Faculty of Sciences, Free University, Amsterdam, The Netherlands*

[3] *Virum, Denmark*

[4] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

For the purpose of objectively validating hearing-aid fittings in pre-lingual infants, auditory steady-state response (ASSR) measurements are investigated. This paper examines the cost of introducing speech-like features into the ASSR stimulus, which is done to ensure that the hearing aid processes the stimulus as if it were real speech. The main findings were a reduction in ASSR amplitude of 4 dB and an increase in detection time by a factor of 1.6, while detection rates were unaffected given sufficient recording time.

INTRODUCTION

The success of new-born hearing-screening has led to very young infants being fitted with hearing aids. Standard tools for validation of these fittings (aided audiometry, questionnaires, etc.) are either impossible or highly unreliable in very young infants. Therefore, objective methods based on electrophysiology are investigated (e.g., Punch *et al.*, 2016). Here, an approach using the auditory steady-state response (ASSR) is considered (Picton *et al.*, 1998).

Aided-ASSR measurements are associated with several challenges. As the ASSR stimulus passes through a hearing aid, it must be ensured that correct gain is applied and that the correct signal processing features relevant for speech are selected. This can be achieved by manipulating the settings of the hearing aid, e.g., by turning off problematic helping systems, such as noise reduction or directionality (Billings *et al.*, 2007; Carter *et al.*, 2013; Easwar *et al.*, 2015). However, this weakens the ecological argument that the hearing aid is in a normal mode of operation. An alternative approach is to construct an ASSR stimulus with sufficiently speech-like properties that the hearing aid automatically classifies the stimulus as speech. The benefit in terms of strengthening the counselling of the infant's parents suggests the latter approach, because in that case the hearing aid can be fitted to the infant and tested in the exact same setting as it will be used in daily life. In addition, a speech-like stimulus will corroborate the relevance of the measurement for both parents and clinicians.

*Corresponding author: slau@iru.interacoustics.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
 ISBN: 978-87-990013-6-1.

For this purpose, narrow-band (NB) CE Chirps® (Elberling and Don, 2010) were modified to have speech-like properties, and it was verified that this stimulus indeed is classified as speech by the hearing aids tested, by observation in the fitting software. However, since the NB-CE chirps were designed for optimal efficiency, it is expected that any change, such as adding speech-like features to the chirps, will come at a cost of reduced ASSR amplitudes and increased detection times. This was investigated in the present experiment. In addition, the observed changes to the measured ASSR amplitudes and response times are compared to an objective characterisation of the speech-modified versus the standard chirps, based on modulation power.

METHOD AND MATERIAL

In order to isolate the effects of the stimulus modifications, the experiment was carried out with young adult normal-hearing test subjects ($N = 10$) and stimulation provided through insert phones. Individual real-ear measurements (REM) were performed in terms of the real-ear unaided gain (REUG) and these results were used to shape the stimuli to mimic the free-field stimulation eventually needed.

The NB-CE Chirps® consist of four one-octave-wide chirps centred at 500, 1000, 2000, and 4000 Hz, and the speech modifications (patent pending) were derived from the International Speech Test Signal (ISTS, Holube *et al.*, 2010). The root-mean-square levels of the individual octave-band chirps of both stimuli were scaled to match the one-octave band levels of the ISTS at its nominal broad-band level of 65 dB SPL, see Fig. 1. The chirp-band-specific repetition rates were 90.8, 94.7, 102.5, and 96.7 Hz, respectively, and each chirp-train was designed with alternating polarity.

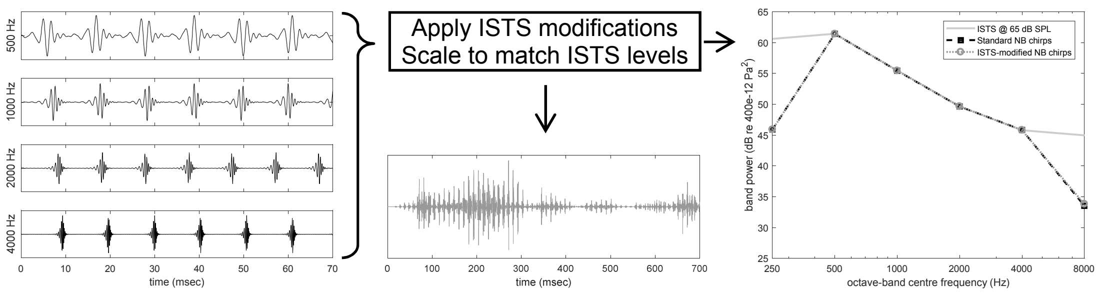

Fig. 1: Sketch of the construction of the stimuli. Left: individual one-octave-band chirps. Middle: resulting waveform of the ISTS-modified chirps. Right: resulting one-octave band spectra compared to that of the ISTS.

ASSR recordings were made using standard clinical 4-electrode montages (high forehead ground, ipsi- and contra-lateral mastoids active, and cheek reference), with 15 minutes of recording time per condition, in one ear at a time. The test subjects were lying comfortably on a bed in a darkened and sound-treated room. Test and re-test recordings were made for all conditions. The Interacoustics Eclipse unit was used as a front-end, with line-level signals routed to an RME Fireface UC soundcard that also delivered the stimuli to the Etymotic Research ER-1 insert phones. Recording and playback were handled by custom Matlab software. The sampling frequency was 32 kHz and the analysis block length was 65536 samples, corresponding to 2.048 sec.

The recordings were analysed with a 40 µV artefact rejection threshold, weighted averaging (John *et al.*, 2001), and simple F-test detection (Dobie and Wilson, 1996) for each of the first 6 response harmonics separately. (A multi-harmonic detector for the ISTS-modified stimulus has yet to be developed.) The outcomes considered were:

- Noise levels, estimated across 20 frequency bins distributed evenly around each response harmonic, excluding frequency bins close to harmonics of 50-Hz line noise, GSM interference, etc.
- Noise-corrected ASSR amplitudes (Dobie and Wilson, 1996). The estimated noise power was subtracted from the response-bin power to yield the noise-corrected power used to compute the ASSR amplitude, which was converted to dB to allow analysis of the relative changes in response with stimulus type.
- Detection times, evaluated as the first time a response was detected with a 5% criterion in successive weighted averages, ignoring Bonferroni correction.

The outcomes were finally analysed with a mixed-model ANOVA with *Test ear* as a random effect and *Stim freq*, *Stim type*, and *Harmonic* as fixed effects.

EXPERIMENTAL RESULTS

Figure 2 shows the number of successful detections in terms of percentages (detection rates). The upper panel shows results for the first 6 harmonics individually, whereas the lower panel shows the detection rates accumulated across harmonics, meaning that an ASSR is detected in either of the first n harmonics, $n = 1, \ldots, 6$.

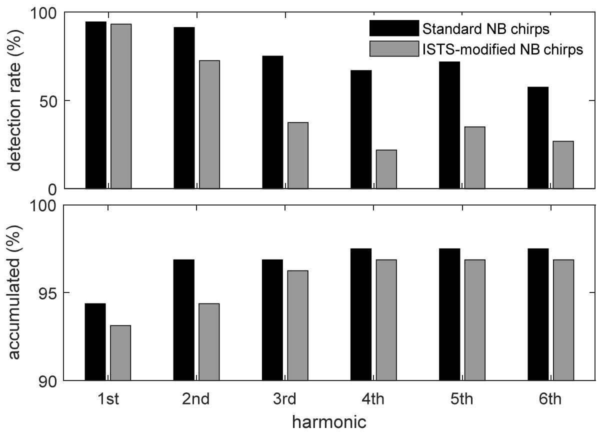

Fig. 2: Detection rates across *Test ear* and *Stim freq* for each *Harmonic* separately (top) and accumulated (bottom), for each *Stim type*.

Considering only the dominant 1st harmonic, the detection rates are very similar for the two types of stimuli. Individually, the higher harmonics provide fewer detections for the ISTS-modified than for the standard stimulus, but nevertheless the accumulated percentages are also very similar.

Figure 3 displays ASSR magnitudes (top panels) and detection times (lower panels) for each harmonic, stimulus band centre frequency, and stimulus type. Both three-way interactions in the statistical models (*Harmonic* by *Stim freq* by *Stim type*) were statistically significant. There are several interesting observations to make from Fig. 3. With increasing harmonic number, the ASSR magnitudes are generally reduced while detection times are increased, as expected. It should be noted that, particularly for the ISTS-modified stimulus, the number of detected responses decreases towards the higher harmonics, which implies that the estimated mean values

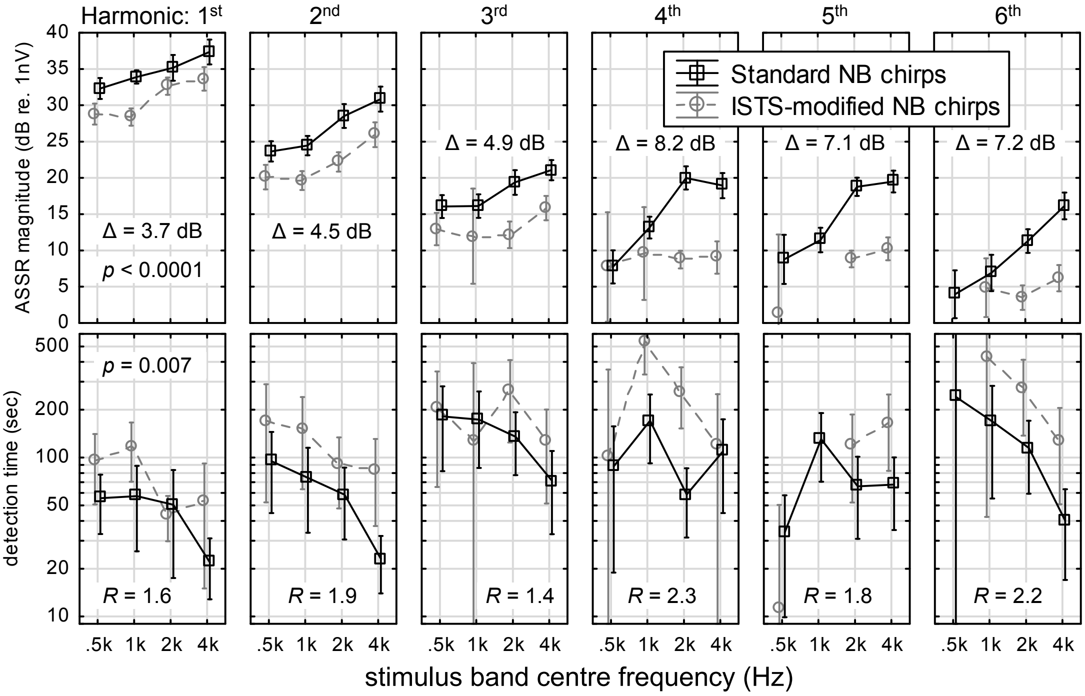

Fig. 3: Noise-corrected ASSR magnitudes (top) and detection times (bottom, logarithmic scale) for each *Harmonic*, *Stim freq*, and *Stim type*, averaged across all ears in which detections were obtained. Error bars indicate 95% confidence intervals. Mean *Stim type* magnitude differences, Δ, mean response-time ratios, *R*, and interaction *p*-values are indicated.

and their patterns are less reliable. Considering the difference between the two stimulus types, the ISTS-modified stimulus produces lower magnitudes and longer detection times than the standard stimulus, which again was expected. It is noteworthy that the patterns of magnitude versus stimulation frequency seem stable up to the 3rd harmonic in terms of a constant difference between the stimuli, while greater mean differences (the inserted Δ-values) and differences in slopes between stimuli can be observed at the higher harmonics. The detection-time data are more variable, as indicated by the wider error bars, but it is striking that the relative increase in detection time between stimuli (the inserted *R*-values) is almost constant across harmonics. This is in contrast to the observed increase in the Δ-values for the ASSR magnitude.

Figure 4 shows the estimated noise levels across conditions where detection was obtained. These results show a reduction in noise level with increasing harmonic number, which was expected since biological noise typically has a 1/*f*-shaped spectrum towards low frequencies (Pritchard, 1992). Note that the plotted *Harmonic* by *Stim type* interaction just fails to reach significance ($p = 0.06$).

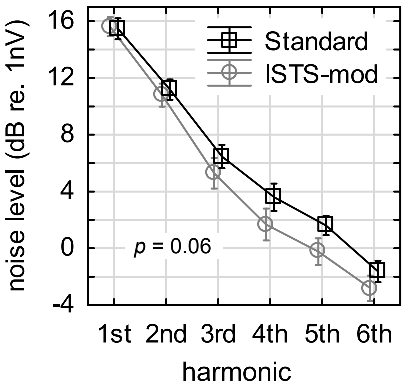

Fig. 4: Mean noise levels across *Test ear* and *Stim freq*.

STIMULUS ANALYSIS

To characterise the stimulus waveforms, an un-normalised modulation spectrum analysis was applied. This analysis was introduced under the assumption that the ASSR is driven by a non-linear representation of the stimulus (for example, after rectification), here modelled in terms of envelope power. This approach requires that the stimuli under comparison are scaled to the same level, in this case the nominal SPL of 65 dB. The two stimuli (both consisting of all four NB chirps) were first passed through gammatone filters (Johannesma, 1972) corresponding to the stimulus frequency band of interest, to mimic the frequency specificity of the auditory system. The results are displayed in Fig. 5 for the two stimuli and two representative stimulus-band centre frequencies: 500 and 2000 Hz. For the standard stimulus (left- hand panels), the modulation power is almost entirely represented at the response harmonics, i.e., the repetition rate of the respective stimulus band and its harmonics. There are smaller modulation power components present at frequencies not belonging to any stimulus repetition rate; these occur because of interactions among the four different repetition rates that are present at the same time in the stimulus. For the ISTS-

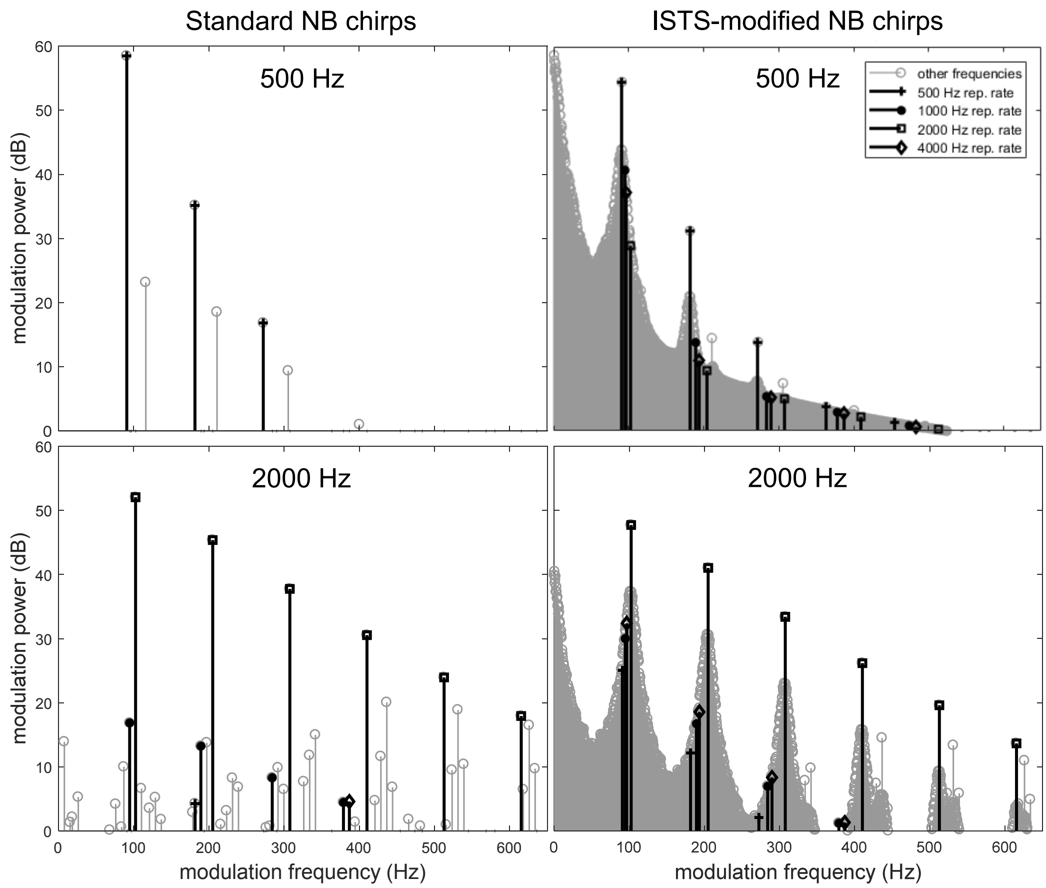

Fig. 5: Un-normalised modulation power spectra of the two stimuli, computed with gammatone filtering at 500 Hz (top) and 2000 Hz (bottom). The stimulus harmonics for each stimulus frequency band are highlighted.

modified stimulus, the modulation power is clearly smeared out around the respective repetition rate and its harmonics. This is a consequence of the additional temporal envelope fluctuations imposed from the ISTS. The modulation power at the response harmonics still stand out in the right-hand panels of Fig. 5, but their magnitudes are reduced compared to the standard stimulus. Table 1 lists the change in modulation power for all four stimulation bands and the first three harmonics. These results show that the estimated change in modulation power is similar across the first response harmonics and among the stimulus bands. The mean reduction is $\Delta_{mod.power}$ = 4.5 dB.

Response harmonic	Analysis band			
	500 Hz	**1 kHz**	**2 kHz**	**4 kHz**
1st	4.0	5.4	4.3	4.9
2nd	3.9	5.3	4.4	4.9
3rd	3.1	5.1	4.4	4.9

Table 1: Reduction in modulation power (in dB) due to the ISTS-modifications to the stimulus.

DISCUSSION

First, the relation between stimulus modulation power and observed ASSR magnitude is considered. For reference, two examples of similar data (‘physiological input/output curves’) are shown in Fig. 6. Both sets of results indicate a slope of about s = 0.8 between dB measures of modulation power and response magnitude. Applying this to the dominant 1st-harmonic data from the present experiment yields

$$\widehat{\Delta_{ASSR}} = \Delta_{mod.power} \times s = 4.5 \text{ dB} \times 0.8 \text{ dB/dB} \approx 3.6 \text{ dB},$$

which agrees very well with the observed Δ_{ASSR} = 3.7 dB from Fig. 3 (top-left panel). In addition, the modulation power analysis reproduces the trend that ASSR magnitude drops more rapidly across harmonics for the lower stimulation frequencies than for the higher (Fig. 3, top row), at least considering the standard NB chirps. This agrees with the successively fewer stimulus line-spectrum components present within an auditory (or gammatone) filter towards lower stimulus band centre frequencies. For the ISTS-modified chirps, the higher-order harmonic responses appear to be limited by the noise floor, which probably conceals the aforementioned effect.

Secondly, the cost of introducing the ISTS-modifications to the NB-CE chirps is considered. By comparing the changes to ASSR magnitude and detection time between the two stimuli, it is seen that the observed reduction in ASSR magnitudes is out of proportion with the increase in detection time, particularly at the higher harmonics. For example, at the 1st harmonic, Δ_{ASSR} = 3.7 dB suggests an increase in detection time by a factor of R = 2.4, where R = 1.6 was observed; at the 6th harmonic Δ_{ASSR} = 7.2 dB suggests R = 5.3, with R = 2.2 observed. The (albeit non-significant) difference in the noise-level patterns (Fig. 4) may hint at lower noise levels for the ISTS-modified relative to the standard NB chirps towards the higher harmonics, which would partly offset the effect of lower ASSR magnitudes on detection time. In addition, note that the detection times were determined from successively averaged un-weighted spectra, whereas the ASSR magnitudes were determined from weighted-

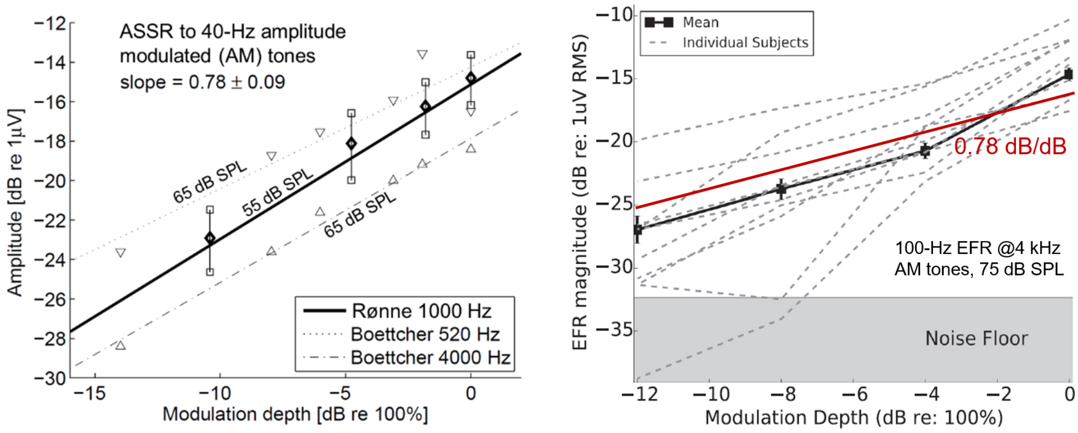

Fig. 6: Two examples of physiological input/output curves. Left: 40-Hz ASSR measurements (Rønne, 2012; Boettcher *et al.*, 2001). Right: 100-Hz envelope-following response (EFR) measurements (Bharadwaj *et al.*, 2015). The regression line from the left panel is superimposed on the right panel.

average spectra across all accepted blocks of the full 15-minute recordings.

Finally, it is encouraging to observe very similar detection rates for the two stimuli (Fig. 2). The reduced response contribution from successively higher harmonics seen for the ISTS-modified versus the standard NB chirps, observed in Fig. 2, top panel, indicates that a multi-harmonic detector, e.g. the q-sample detector (Cebulla *et al.*, 2006) may provide less benefit for the ISTS-modified stimulus compared to what has been found for standard stimuli. On the other hand, the accumulated detection rates in Fig. 2, bottom row, show bigger improvement from including more harmonics for the ISTS-modified than for the standard stimulus.

CONCLUSIONS

The consequences of adding speech-like properties to the NB-CE Chirps® for ASSR recordings – for the purpose of hearing-aid validation in infants – were investigated in young adult normal-hearing test subjects. The main findings were:

- Detection rates were very similar for the speech-modified and the standard stimuli.
- ASSR magnitude decreased by about 4 dB (for the dominant 1st response harmonic).
- Detection times increased relatively less, by a factor of 1.6.

The reduced response magnitude and increased detection time seem acceptable, given the potential for allowing aided ASSR recordings to be carried out with hearing aids in their daily-life mode of operation.

The un-normalised modulation power spectrum including pre-processing through gammatone filters appears to be a useful tool for characterising the efficacy of complex stimuli for ASSR measurements.

Future work will extend the investigations to infants fitted with hearing aids.

REFERENCES

Bharadwaj, H.M., Masud, S., Mehraei, G., Verhulst, S., and Shinn-Cunningham, B. G. (**2015**). "Individual differences reveal correlates of hidden hearing deficits," J. Neurosci., **35**, 2161-2172.

Billings, C.J., Tremblay, K.L., Souza, P.E., and Binns, M.A. (**2007**). "Effects of hearing aid amplification and stimulus intensity on cortical auditory evoked potentials," Audiol. Neurootol., **12**, 234-246.

Boettcher, F.A., Poth, E.A., Mills, J.H., and Dubno, J.R. (**2001**). "The amplitude-modulation following response in young and aged human subjects," Hear. Res., **153**, 32-42.

Carter, L., Dillon, H., Seymour, J., Seeto, M., and Van Dun, B. (**2013**). "Cortical auditory-evoked potentials (CAEPs) in adults in response to filtered speech stimuli," J. Am. Acad. Audiol., **24**, 807-822.

Cebulla, M., Stürzebecher, E., and Elberling, C. (**2006**). "Objective detection of auditory steady-state responses: comparison of one-sample and q-sample tests," J. Am. Acad. Audiol., **17**, 93-103.

Dobie, R.A. and Wilson, M.J. (**1996**). "A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," J. Acoust. Soc. Am., **100**, 2236-2246.

Easwar, V., Purcell, D.W., Aiken, S.J., Parsa, V., and Scollie, S.D. (**2015**). "Evaluation of speech-evoked envelope following responses as an objective aided outcome measure: effect of stimulus level, bandwidth, and amplification in adults with hearing loss," Ear Hearing, **36**, 635-652.

Elberling, C., and Don, M. (**2010**). "A direct approach for the design of chirp stimuli used for the recording of auditory brainstem responses," J. Acoust. Soc. Am., **128**, 2955-2964.

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (**2010**). "Development and analysis of an international speech test signal (ISTS)," Int. J. Audiol., **49**, 891-903.

Johannesma, P.I. (**1972**). "The pre-response stimulus ensemble of neurons in the cochlear nucleus," Symposium on Hearing Theory, Eindhoven, Holland, 58-69.

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

Picton, T.W., Durieux-Smith, A., Champagne, S.C., Whittingham, J., Moran, L.M., Giguère, C., and Beauregard, Y. (**1998**). "Objective evaluation of aided thresholds using auditory steady-state responses," J. Am. Acad. Audiol., **9**, 315-331.

Pritchard, W.S. (**1992**). "The brain in fractal time: 1/f-like power spectrum scaling of the human electroencephalogram," Int. J. Neurosci., **66**, 119-129.

Punch, S., Van Dun, B., King, A., Carter, L., and Pearce, W. (**2016**). "Clinical experience of using cortical auditory evoked potentials in the treatment of infant hearing loss in Australia," Semin. Hear., **37**, 36-52.

Rønne, F.M. (**2012**). "Modeling auditory evoked potentials to complex stimuli," PhD thesis, Technical University of Denmark.