Modeling auditory evoked brainstem responses to speech syllables. Can variations in cochlear tuning explain argued brainstem plasticity?

FILIP M. RØNNE[1,2,*], JAMES HARTE[1,3], AND TORSTEN DAU[1]

[1] *Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[2] *Eriksholm Research Centre, Rørtangvej 20, 3070 Snekkersten, Denmark*

[3] *Institute of Digital Healthcare, WMG, University of Warwick, Coventry, UCVA 7AL, UK*

Hornickel *et al.* (2009) and Skoe *et al.* (2011) measured and analyzed brainstem responses (ABRs) in response to the synthetic syllables /ba/, /da/ and /ga/, in normal and learning-impaired children. They reported a co-variation between the differences in average phase lag between the three syllable-evoked responses (called average phase-shifts), and speech-intelligibility performance (used as a predictor for learning-impairment). It was argued that, due to the reported normal peripheral hearing of both groups, the co-variation was evidence for neural differences in the brainstem, likely related to brainstem plasticity. They suggested brainstem functionality can be influenced by cortical structures to increase the difference between syllable responses. This study developed an ABR model capable of simulating ABRs to a variety of stimuli. The model was used to investigate whether the state of the peripheral hearing could be another possible explanation for the decreased average phase shifts observed for the learning-impaired children. Specifically, by changing the cochlear tuning of the model and evaluating the simulations based on models with broad versus sharp tuning (yet keeping all tuning estimates within normal audiometrical and wave-V latency range), it was observed that broader tuning systematically lead to smaller phase-shifts between the syllable-evoked ABRs.

INTRODUCTION

Auditory evoked potentials (AEP) have been used to assess the neural encoding of sound both for clinical and research purposes. Most studies have focused on the auditory brainstem response (ABR) as they are less affected by attention and sleep than potentials with origin at higher neural stages. The ABR has also been observed to be unaffected by training. However, a number studies have recently investigated and found evidence of plasticity[1] of the complex ABR (cABR), both

---

*Corresponding author: fmr@eriksholm.com

[1] physiological changes of the nervous system due to, e.g., learning

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

considering short term training effects and long-term experience effects. Hornickel *et al.* (2009) and Skoe *et al.* (2011) measured brainstem responses to the synthetically created syllable-stimuli /ba/, /da/, and /ga/, in normal and learning-impaired children. Both groups of children were reported to have normal audiometric thresholds and ABR wave-V latencies. Skoe *et al.* (2011) developed a 'cross-phaseogram' from the time-varying cross-power-spectral-density between two ABR recordings. When analyzed in time-frames, the outcome was a spectrogram-like representation of the phase-lag as a function of time and frequency. From the cross-phaseogram an averaged phase-shift between two syllable-evoked responses was obtained. The average phase-shift was shown to correlate with reading abilities and speech-in-noise perception, such that large phase-shifts correlated with good performance in the speech-in-noise test. Hornickel *et al.* (2009) and Skoe *et al.* (2011) argued that this result was evidence for plasticity in the brainstem, as the group with the good behavioral performance had undergone long-term learning. Thus, better performance was an indication of learning that had affected both the behavioral performance and the electrophysiological brainstem recordings. This paper challenges the reasoning behind this interpretation. By modeling it attempts to show that individual variations in cochlear tuning, all within normal-hearing boundaries, significantly affect the average phase-shifts, thus showing that the measures of the peripheral hearing chosen by Hornickel *et al.* (2009) are not sufficient to conclude that the individual spread in the peripheral hearing does not affect the average-phase shift group differences between normal and learning-impaired children.

METHOD

ABR model

The ABR model used in this study was similar to the model of Rønne *et al.* (2012). However, the auditory-nerve (AN) model used to compute the summed activity pattern was updated such that the Zilany *et al.* (2009) AN model was used instead of the Zilany and Bruce (2007) model. This update was made as the Zilany *et al.* (2009) has an improved IHC-AN stage producing more realistic adaptation properties. As the syllable-stimuli are of longer duration, a precise adaptation is beneficial. The change of the AN model required a recalculation of the unitary response (UR). The UR (based on standard cochlear filter tuning) was calculated, following Rønne *et al.* (2012), as the deconvolution of a 95.2 dB peSPL grand average click-evoked ABR recording (Elberling *et al.*, 2010; Rønne *et al.*, 2012) and the summed activity pattern obtained by simulating the response to an identical click-stimulus.

The simulated cABRs were at the output filtered with a 2nd order band-pass filter with cutoff frequencies at 70 Hz and 2 kHz. These filter settings were identical to the output filters of Hornickel *et al.* (2009) and Skoe *et al.* (2011).

Stimuli

Synthetic /ba/, /da/, and /ga/ syllables (Hornickel *et al.*, 2009; Skoe *et al.*, 2011) were used, that only differ in the frequency content of the second formant, f_2, of the first 60 ms, corresponding to the consonant part of the stimuli. The second formants decrease in the [ga] stimulus from 2480 Hz, in the [da] from 1700 Hz, and increased in the [ba] stimulus from 900 Hz, reaching a steady-state frequency (corresponding to the /a/ part of the syllable) of 1240 Hz in all 3 stimuli. The /a/ vowel-part of the syllables was the same for the three syllables, consisting of the formant frequencies f_0 = 100 Hz, f_1 = 720 Hz, f_2 = 1240 Hz, f_3 = 2500Hz, f_4 = 3300 Hz, f_5 = 3750 Hz and f_6 = 4900 Hz. All three stimuli were calibrated to have a root-mean-square (RMS) level of 1, and were presented to the model at a level corresponding to 80 dB SPL, which was also used in the study by Skoe *et al.* (2011).

Cross-phaseogram

Skoe *et al.* (2011) proposed a cross-phaseogram to illustrate the phase-differences and thus the time delays between two cABR recordings. Each recording was divided into 20-ms time frames with 19-ms overlap. A Hanning window was applied, resulting in a 3-dB main lobe width of 141 Hz. The cross power spectrum density, i.e., the power spectrum density of the cross correlation, was computed between each pair of frames from the two recordings. An artificial frequency resolution of 4 Hz was obtained by zero padding, effectively acting as a smoothing operation. Finally, the unwrapped phase (in radians) was extracted and plotted as a function of time (midpoint of the 20-ms frames) and frequency. Skoe *et al.* (2011) also proposed the average phase-shift to simplify the cross-phaseogram into a single number that could be compared to other measures, such as psychoacoustic speech-in-noise performance. The average phase-shift (in π radians) was calculated on the formant transition period (15 to 60 ms) of the syllable-evoked cABR in the frequency range of 70 to 1100 Hz.

Weighted cross-phaseogram

The cross-phaseogram weights time-frequency bins with little activity as high as bins with much activity. This limits the use of the cross-phaseogram, as it is impossible to distinguish between time-frequency bins of presumable little importance due to low activity from bins of major importance due to large activity. A weighted cross-phaseogram is therefore suggested here. It was created by deriving the energy from each of the two syllable-evoked cABRs in similar time-frequency bins as those chosen in the Skoe *et al.* (2011) cross-phaseogram. The two resulting matrices were summed and normalized with the average bin activity. This matrix was then multiplied bin-per-bin with the original cross-phaseogram.

Variability of cochlear filter tuning

Cochlear filter tuning and basilar-membrane (BM) delay are inherently related (Eggermont, 1979; Bentsen *et al.*, 2011; Verhulst *et al.*, 2013), such that broader filters lead to shorter delays. Elberling and Don (2008) measured derived-band latencies from a total of 81 normal-hearing subjects (hearing thresholds $<$ 15 dB HL), at four different band center frequencies (bCF; 710, 1400, 2800, and 5700). ABR wave-V latency and an inter-subject standard deviation (SD) were derived. The BM delay was achieved by subtracting the wave I-V delay (4.1 ms) and the synaptic delay (1 ms). A representation of the variation of cochlear filter tuning in normal-hearing subjects can be obtained from the mean latencies $\pm$ 1 standard deviation. The stimulus of Elberling and Don (2008) was a click presented at approximately 90 dB peSPL.

Eggermont (1979) derived a theoretical relation between the cochlear filter tuning, Q_{10}, and the average number of cycles in the impulse response up to the latency (minus 1 ms of synaptic delay) of the derived band CAP, N_{av};

$$N_{av} = \frac{0.5}{\pi^2}\left(\frac{5(1+\gamma)(2+\gamma)}{12\gamma}Q_{10}-1\right)\left(2+ln\frac{5(1+\gamma)(2+\gamma)}{12\gamma}+lnQ_{10}\right) \qquad \text{(Eq. 1)}$$

where N_{av} can be calculated as $(CF/1000)*\tau_{CF}$, where τ is the BM latency of at the CF. $\gamma = 2$ is representative of a normal cochlea (Eggermont, 1979), and Q_{10} values can thus be derived. To convert the Q_{10} values into Q_{ERB} values, the conversion from Ibrahim and Bruce (2010) was applied:

$$Q_{ERB} = \frac{Q_{10}-0.2085}{0.505} \qquad \text{(Eq. 2)}$$

Fig. 1 shows the Q_{ERB} values derived from Elberling and Don (2008)'s measured delays ±1 SDs and ±2 SDs. New tuning-curve estimates were obtained from the ±1 SD and ±2 SD based Q-estimates, by multiplying the Shera *et al.* (2002) estimates by a constant offset (broader tuning-estimates multiplied by 0.80 and 0.60, sharper tuning-estimates by 1.15 and 1.28). The four suggested tuning curves were implemented in the ABR model. For each simulated condition, a new UR was calculated.

RESULTS

Table 1 shows the average phase-shifts obtained in Skoe *et al.* (2011) and the corresponding values obtained from the simulations.[2] Both experimental results and simulations show the largest phase-shift between /ga/ and /ba/, which also differs most in their frequency spectrum. Also, both the data and the simulations show that the phase-shift between /ga/ and /da/ is smaller than the phase-shift between /da/ and /ba/. In Fig. 2, weighted average phase-shifts for all syllable comparisons and all five different tuning-curve implementations are shown. Although the growth of the phase-shift with increasing tuning amount is non-monotonic, a trend is clearly observed, where sharp tuning leads to larger phase-shifts. This confirms that the state of the

---

[2] Coloured cross-phaseograms describing the results in details are shown on the poster (available from `http://www.eriksholm.com/~asset/cache.ashx?id=26052&type=14&format=web`).

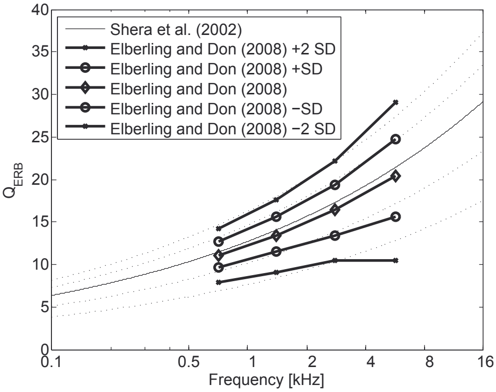

Fig. 1: Q_{ERB}'s calculated based on Elberling and Don (2008)'s measured derived band latencies (diamonds). In circles and triangles, Q_{ERB} estimates based on Elberling and Don (2008)'s measured latencies ±1 SD an ±2 SD. Also shown is the Shera *et al.* (2002) tuning (solid line) which is implemented in the standard ABR model. The alternative tuning curves (dotted lines) are fitted to the Elberling and Don (2008) based tuning (±1 SD and ±2 SD) and implemented in the model.

auditory periphery affects the cross-phaseogram and weighted average phase-shifts. The implications for the Hornickel *et al.* (2009) and Skoe *et al.* (2011) studies are discussed further below.

	Skoe et al. (2011)	Simulations	Simulations (weighted)
/ga/-/ba/	0.317 ± 0.040	0.353	3.040
/da/-/ba/	0.288 ± 0.031	0.243	2.163
/ga/-/da/	0.208 ± 0.028	0.141	1.660

Table 1: Average phase-shifts of Skoe *et al.* (2011) recordings (left column), simulated average phase-shifts (center column), and weighted average phase-shifts (right column). The average is taken across the region from 15 to 60 ms, and from 70 to 1100 Hz.

DISCUSSION

Unweighted versus weighted cross-phaseogram

The cross-phaseogram and the average phase-shifts were developed by Skoe *et al.* (2011) and have proven to be valuable tools for investigating phase-shifts between different frequency components of the recorded (or simulated) cABR. However, the

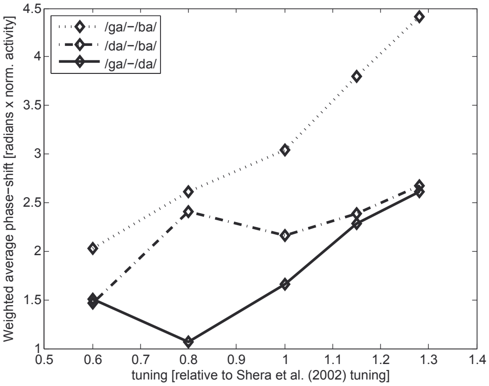

Fig. 2: Weighted average phase-shifts for each of the syllable combinations, for both broad (0.60 and 0.80), standard (1.00) and sharp (1.15 and 1.28) tuning.

equal weighting of all time-frequency bins limits the value of the average phase-shift (Skoe *et al.*, 2011), since a bin with little activity will hardly influence the cABR generation. In fact, a time-frequency bin with little energy is likely to be dominated by measurement noise, and the average measure might thus emphasize noise. In the simulations presented in this study, noise is not included. This makes a comparison between simulations and data in the terms of the average phase-shift difficult, as a systematic phase-shift at bins with little activity will be included in the simulated average phase-shift, whereas such a phase-shift is likely to be influenced or masked by measurement noise in the data-derived average phase-shift. This could be solved by adding noise to simulations. However, this would imply that the model would no longer be deterministic, which has not been considered in the present study.

Implications of changing cochlear tuning on Skoe et al. (2011) conclusions

Hornickel *et al.* (2009) and Skoe *et al.* (2011) found correlations between learning-impairments of children, and recorded cross-phaseogram phase-shifts (peak latencies in Hornickel *et al.*, 2009) between syllable-evoked cABRs, such that a small average phase-shift was an indication of learning-impairment. A basic assumption of Hornickel *et al.* (2009) was that the two groups of normal and learning-impaired children have equally good peripheral hearing (equal audiograms and ABR wave-V latencies). Hornickel *et al.* (2009) argued that this was the case as all subjects had audiometric thresholds below 20 dB HL and had normal ABR wave-V latencies. However, given the possible variation of ‘normal’ BM tuning, an alternative explanation for the Hornickel *et al.* (2009) results can be hypothesized. A broad cochlear tuning leads to shorter peak-latencies for all three stimuli. Further, the traveling-wave delay decreases logarithmically with increasing stimulus frequency (e.g., Neely *et al.*, 1988; Elberling *et al.*, 2010). A broad tuning would thus lead to a decreased difference between the cABR peaks, and thus a smaller phase-shift. Phase-shift differences similar to the one Skoe *et al.* (2011) finds between the groups of normal and learning-impaired children, could thus be hypothesized to also be found when measuring cABRs to two normal-hearing groups but with different cochlear tuning.

The results from this modeling study showed that there is indeed a relation between filter tuning and weighted averaged cross-phaseogram values, where sharper tuning leads to larger phase-shifts. Although this relation was not strictly monotonic, it does indicate that the phaseograms are sensitive to changes in the auditory periphery. Whether this finding offers an alternative explanation for the results of Hornickel *et al.* (2009) and Skoe *et al.* (2011) is, however, questionable. That would require the assumption that the group of learning-impaired children had significantly overall broader cochlear tuning than the normal children. Although this hypothesis is not unlikely, this study cannot verify such a claim. That would require a major study, where the cochlear tuning of learning-impaired and normal subjects were measured carefully and correlated with weighted average phase-shifts. Thus, the conclusion of this study is that the huge spread of normal-hearing cochlear-tuning likely leads to a huge spread in weighted average phase-shifts.

Skoe *et al.* (2011) concluded that the correlation between learning-impairment and average phase-shifts showed plasticity of brainstem. This conclusion was based on the assumption that the state of the auditory periphery was equal (i.e., normal hearing) in both groups. However, this study has indicated that the cochlear tuning of the normal-hearing subjects does have a significant effect on the average phase-shift, and does thus challenge the underlying assumption of the conclusions from Hornickel *et al.* (2009) and Skoe *et al.* (2011). Further, this study has shown that the use of audiograms and click-evoked ABR wave-V latencies are unlikely to be precise enough to claim that the cochlear tuning is similar between two groups.

SUMMARY AND CONCLUSION

This study evaluated the performance of an ABR model to simulate cABR responses to three synthetic syllables. The ABR model was shown to predict phase-shifts between the responses to the three syllable stimuli. It was shown that altering the cochlear tuning influenced the simulated phase-shifts, illustrating that the state of the auditory periphery is crucial when analyzing responses based on the cross-phaseogram. The results suggests that the assumption of Hornickel *et al.* (2009) and Skoe *et al.* (2011) that the peripheral hearing was similar between their two groups of test subjects might be flawed, and that the following conclusion that the larger phase-shifts for the non-learning-impaired children was the consequence of plasticity might thus be wrong.

Bentsen, T., Harte, J.M., and Dau, T. (**2011**), “Human cochlear tuning estimates from stimulus-frequency otoacoustic emissions,” J. Acoust. Soc. Am., **129**, 3797-3807.

Eggermont, J. (**1979**), “Narrow-band AP latencies in normal and recruiting human ears,” J. Acoust. Soc. Am., **65**, 463-470.

Elberling, C., and Don, M. (**2008**). “Auditory brainstem responses to a chirp stimulus designed from derived-band latencies in normal-hearing subjects,” J. Acoust. Soc. Am., **124**, 3022-3037.

Elberling, C., Callø, J., and Don, M. (**2010**). “Evaluating auditory brainstem responses to different chirp stimuli at three levels of stimulation,” J. Acoust. Soc. Am., **128**, 215-223.

Hornickel, J., Skoe, E., Nicol, T., Zecker, S., and Kraus, N. (**2009**). “Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception,” Proc. Natl. Acad. Sci. USA, **106**, 13022-13027.

Ibrahim, R.A., and Bruce, I.C. (**2010**), “Effects of peripheral tuning on the auditory nerve’s representation of speech envelope and temporal fine structure cues,” in *Neurophysiological bases of auditory perception*. Edited by E.A. Lopez-Poveda and A.R. Palmer, Med Elect, Hear Life, 15th International Symposium on Hearing, Salamanca, Spain, June 2009, pp. 429-438.

Neely, S., Norton, S., Gorga, M., and Jesteadt, W. (**1988**). “Latency of auditory brainstem responses and otoacoustic emissions using tone-burst stimuli,” J. Acoust. Soc. Am., **83**, 652-656.

Rønne, F., Harte, J., Elberling, C., and Dau, T. (**2012**). “Modelling auditory evoked brainstem responses to transient stimuli,” J. Acoust. Soc. Am., **131**, 3903-3913.

Rønne, F., Harte, J., and Dau, T. (**2013**). “Modelling human auditory evoked brainstem responses to speech syllables,” Proc. Meet. Acoust., **19**, International Congress on Acoustics, Montréal, Canada, pp. 050120.

Shera, C., Guinan, J., and Oxenham, A.J. (**2002**). “Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,” Proc. Natl. Acad. Sci. USA, **99**, 3318-3323.

Skoe, E., Nicol, T., and Kraus, N. (**2011**). “Cross-phaseogram: Objective neural index of speech sound differentiation,” J. Neurosci. Meth. **196**, 308-317.

Verhulst, V., Bharadwaj, H., Mehraei, G., and Shinn-Cunningham, B. (**2013**). “Understanding hearing impairment through model predictions of brainstem responses,” Proc. Meet. Acoust., **19**, International Congress on Acoustics, Montréal, Canada, pp. 050182.

Zilany, M.S.A., and Bruce, I.C. (**2007**). “Representation of the vowel (epsilon) in normal and impaired auditory nerve fibers: Model predictions of responses in cats,” J. Acoust. Soc. Am., **122**, 402-417.

Zilany, M.S.A., Bruce, I.C., Nelson, P.C., and Carney, L.H. (**2009**). “A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics,” J. Acoust. Soc. Am., **126**, 2390-2412.