

- Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2002). "Dynamic brain sources of visual evoked responses" *Science* **295**, 690-694.
- Mazaheri, A., and Jensen, O. (2008). "Asymmetric amplitude modulations of brain oscillations generate slow evoked responses" *J. Neurosci.* **28**, 7781-7787.
- McClelland, J. L., and Elman, J. L. (1986). "The TRACE model of speech perception" *Cognit. Psychol.* **18**, 1-86.
- Murry, T., and Singh, S. (1980). "Multidimensional analysis of male and female voices" *J. Acoust. Soc. Am.* **68**, 1294-1300.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (1997). "Language-specific phoneme representations revealed by electric and magnetic brain responses" *Nature* **385**, 432-434.
- Norris, D., and McQueen, J. M. (2008). "Shortlist B: a Bayesian model of continuous speech recognition" *Psychol. Rev.* **115**, 357-395.
- Obleser, J., Elbert, T., and Eulitz, C. (2004a). "Attentional influences on functional mapping of speech sounds in human auditory cortex" *BMC Neurosci.* **5**, 24.
- Obleser, J., Lahiri, A., and Eulitz, C. (2004b). "Magnetic brain response mirrors extraction of phonological features from spoken vowels" *J. Cogn. Neurosci.* **16**, 31-39.
- Parviainen, T., Helenius, P., and Salmelin, R. (2005). "Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia" *Cereb. Cortex* **15**, 1054-1063.
- Poeppel, D., Phillips, C., Yellin, E., Rowley, H. A., Roberts, T. P., and Marantz, A. (1997). "Processing of vowels in supratemporal auditory cortex" *Neurosci. Lett.* **221**, 145-148.
- Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P., Rowley, H. A., Wexler, K., and Marantz, A. (1996). "Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds" *Brain Res. Cogn. Brain Res.* **4**, 231-242.
- Rauschecker, J. P., and Scott, S. K. (2009). "Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing" *Nat. Neurosci.* **12**, 718-724.
- Salinas, E., and Sejnowski, T. J. (2001). "Correlated neuronal activity and the flow of neural information" *Nat. Rev. Neurosci.* **2**, 539-550.
- Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., and Huotilainen, M. (2004). "Orderly cortical representation of vowel categories presented by multiple exemplars" *Brain Res. Cogn. Brain Res.* **21**, 342-350.
- Titova, N., and Näätänen, R. (2001). "Preattentive voice discrimination by the human brain as indexed by the mismatch negativity" *Neurosci. Lett.* **308**, 63-65.
- von Stein, A., Chiang, C., and Konig, P. (2000). "Top-down processing mediated by interareal synchronization" *Proc. Natl. Acad. Sci. U S A* **97**, 14748-14753.

Low-frequency versus high-frequency synchronisation in chirp-evoked auditory brainstem responses

FILIP MUNCH RØNNE AND KRISTIAN GØTSCHÉ-RASMUSSEN

Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Lyngby, Denmark

This study investigates the frequency specific contribution to the auditory brainstem response (ABR) of chirp stimuli. Frequency rising chirps were designed to compensate for the cochlear traveling wave delay, and lead to larger wave-V amplitudes than for click stimuli as more auditory nerve fibres fire synchronously. Traditional click stimuli were believed to only excite high-frequency fibres synchronously. It is still currently unclear whether the broad-band chirp stimulus leads to increased synchronisation of both low- and high-frequency fibres. It is also unclear if both these groups of fibres contribute significantly to the overall wave-V amplitude. In the present study, ABRs were recorded from 10 normal-hearing listeners using low- and high-frequency band-limited chirps and clicks (0.1 – 1.5 kHz and 1.5 - 10 kHz) presented at a level of 40 dB HL. The results showed significantly larger wave-V amplitudes for both low and high-frequency band-limited chirps than for the filtered clicks. This demonstrates that the synchronisation of nerve fibres occurs across the entire frequency range at this presentation level, and this leads to significant increases in wave-V amplitudes. The increase for the low-frequency chirp was found to be clearly larger than that obtained at the higher frequencies.

INTRODUCTION

ABRs in response to transient sound stimuli represent the summed electric potential from many remotely located neurons, recorded via scalp electrodes. The click evoked ABR has 7 distinct waves, where wave-V is the most prominent. One key feature of the ABR wave-V is the peak latency which is dependent on both stimulus frequency (Neely *et al.*, 1988) and level (Dau, 2003). The frequency dependence is due to the tonotopic mapping on the basilar membrane (BM) with high-frequency at base and low-frequency at apex (Greenwood, 1990). Each frequency component of a stimulus is associated with a certain delay, and a click stimulus will thus elicit responses over a relatively large time span. This limits the synchronicity of the response, and thereby reduces the ABR amplitude evoked by such a stimulus (Elberling *et al.*, 2007). Frequency rising chirps have been designed to compensate for the cochlear travelling wave delay. The use of chirp stimulus leads to larger wave-V amplitudes than for click stimuli as more auditory nerve fibres fire synchronously (see Elberling *et al.*, 2007, for review). The increase in synchronicity has traditionally been argued to occur mainly at low frequencies, where the peaks of

the individual nerve responses are most delayed. E.g. Shore and Nuttall (1985) and Dau *et al.* (2000) argue that the low frequencies are the key to the improved wave-V amplitudes, as low frequencies are least synchronous with the more aligned high frequencies and the room for improvement thus is largest. However, the impulse responses of the nerve fibre responses at low frequencies have much longer duration (Kiang, 1965), and it is thus not possible to align all the excitation at low frequencies. However, a chirp is designed to align all frequencies (Elberling and Don, 2008), and the better alignment of high frequencies, with short impulse responses, could thus be an alternative hypothesis. It is currently unclear whether the broad-band chirp stimulus leads to increased synchronisation of both low- and high-frequency fibres. It is also unclear if both of these groups of fibres contribute significantly to the overall wave-V amplitude.

The research questions addressed in this paper are: 1) Is the increased wave-V amplitude (increased nervous synchronicity) observed for both high and low frequencies when stimulating with chirps instead of clicks? 2) Are high or low frequencies key to the increased wave-V amplitude observed when stimulating with broad-band chirps?

TEST DESIGN

Six stimuli were created. A broad-band click and a broad-band chirp, containing the frequencies from 100 Hz to 10 kHz, were used as reference. The click was a 100 μs standard click, and the chirp was identical to "chirp 3" in Elberling *et al.* (2010). Further low-frequency and high-frequency versions of the click and chirp were created. The method described by Elberling *et al.* (2007) was used. The phase delays for hence chirps and clicks were the same as used to create the broad-band stimuli. Both the high-frequency and low-frequency cut-off frequency was 1500 Hz. Fig. 1 shows the time series representation of the six stimuli. The power spectra of the two broad-band stimuli were identical. The summed versions of the low-frequency and high-frequency click, and the low-frequency and high-frequency chirp has also identical power spectra as the broad-band versions. The power of the low-frequency (-3.1 dB relative to broad-band condition) high-frequency (-0.6 dB relative to broad-band condition) stimulus are thus smaller than the power of the broad-band versions. Fig. 2 shows the power spectra of the stimuli, note that the two broad-band stimuli, the two low-frequency stimuli and the two high-frequency stimuli have identical spectra.

The six stimuli were linked to each other in terms of the power spectra as described above. Therefore only the broad-band click was calibrated, and the rest adjusted correspondingly. By inserting ER1-14 ear plug in a B&K Ear Simulator Type 4157 (IEC 60711) using adapter B&K DB 2012 the click was calibrated to a level of 75.2 dB peSPL. The reference equivalent threshold sound pressure level (RETSPL) for the click calibrated this way is 35.2 dB RETSPL (taken from the corresponding head and torso simulator measurement of Richter & Fedtke, 2005), and the measurements are thus carried out at 40 dB HL.

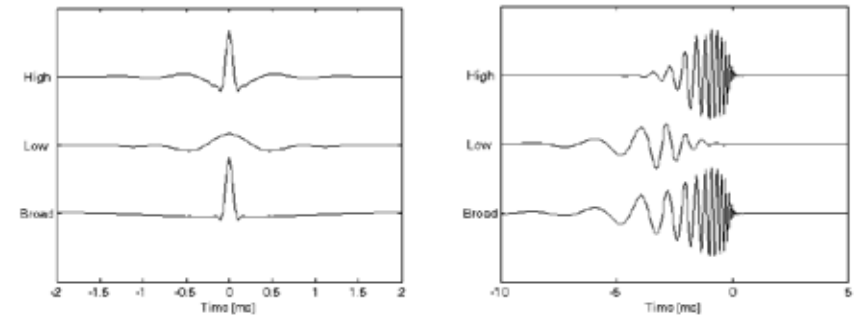


Fig. 1: The six stimuli used in the study. To the left are the clicks shown, to the right the chirps.

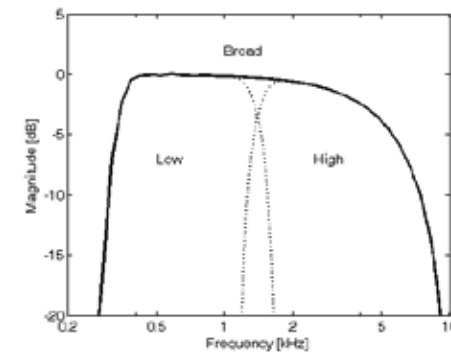


Fig. 2: Spectra of the different stimuli. The sum of the two hence low- and high-frequency clicks or chirps have the same power spectrum as the broad-band stimulus.

Test subjects

The ABR measurements were carried out at the Centre for Applied Hearing Research (CAHR), Technical University of Denmark. Ten normal-hearing test subjects (10 left ears) participated in the study. All subjects had normal hearing defined as pure tone thresholds equal to or better than 20 dB HL in the range from 125 Hz to 8 kHz. The subjects were all students between 20-30 years old (2 females and 8 males). The session lasted for maximally 1.5 hours including a short briefing and fitting of electrode cap. Only the left ear was tested.

Measurement procedure

The test subject was placed in an electrically and acoustically shielded booth. The signals were presented at 48 kHz sampling frequency through an Etymotic Research ER-2 insert earphone. The recording of the ABR was done using a Medical

Equipment ApS Synamps², which sampled the recorded signal at 10 kHz. The electrodes were placed at vertex (reference), ipsilateral mastoid, and forehead (ground). An electrode impedance below 1 kΩ was achieved for the majority of the test subjects.

The post-processing was done using MATLAB. The raw data was averaged, and filtered using a band-pass filter with cut-off frequencies at 100 and 3000 Hz. Wave-V was detected in a time interval from 0 - 7 ms after the offset of the stimulation. The wave-V amplitude was calculated as the difference in amplitude between the maximum amplitude and the minimum amplitude found in the subsequent 2 ms.

RESULTS

Fig. 3 shows the mean and the corresponding one-standard deviation of wave-V amplitudes of the 6 conditions measured. The broad-band click and chirp used in this study are identical to the ones presented by Elberling *et al.* (2010). They found an averaged click evoked wave-V amplitude of 0.368 μV and an averaged chirp evoked amplitude of 0.645 μV. This compares well with the amplitudes measured in this study.

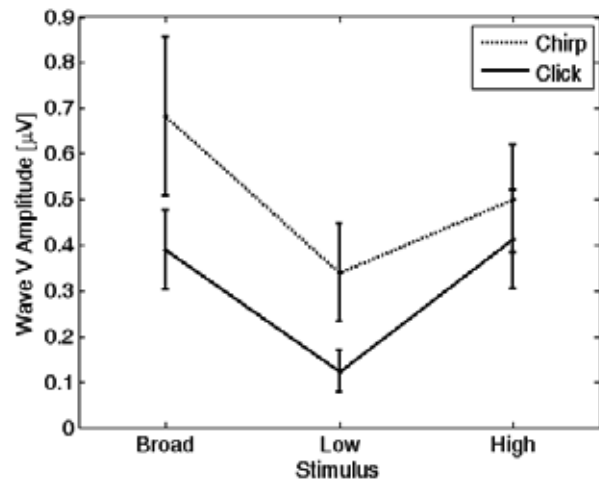


Fig. 3: Mean ABR Wave-V amplitude and one standard deviation plotted for each stimulus condition

The mean amplitudes indicate that the chirp stimuli generate larger ABR Wave-V amplitude compared to the click stimuli across all conditions. The high-frequency chirp condition is significantly different from both the broad-band chirp (High ≠ Broad: p value = 0.014) and the low-frequency chirp condition (High ≠ Low: p value = 0.005), indicating that both high and low frequencies are adding to the measured amplitude. It cannot be rejected that the high-frequency click gives rise to the same amplitude as the broad-band click (High ≠ Broad: p value = 0.614) indicating that

the broad-band click is entirely determined by the high-frequency contribution. The p-values were calculated using a one-sample t-test.

The difference between the click evoked and chirp evoked wave-V amplitude was calculated for each test subject to reduce the influence of the inter-subject variability. The mean and standard deviation of the improvements from click to chirp are shown in Fig. 4. A two-sample t-test was applied to analyse the data (see Table 1).

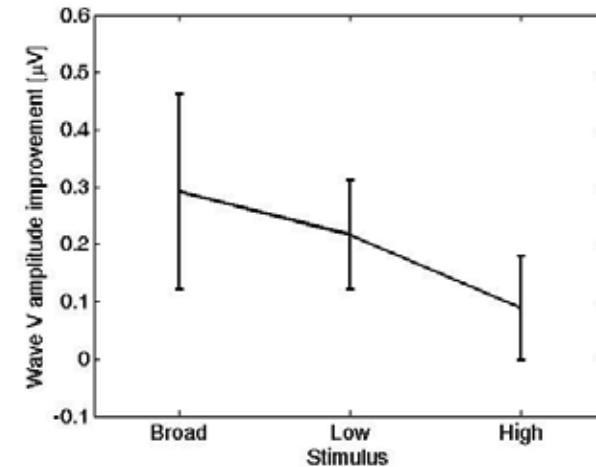


Fig. 4: Improvement in wave-V amplitude from click to chirp evoked responses. The mean and one standard deviation are plotted.

Hypothesis	P-value
Low > 0	<< 0.001
High > 0	0.006
Broad > 0	<< 0.001
Low ≠ Broad	0.237
High ≠ Broad	0.004

Table 1: Statistical analysis of data in Fig. 4, two-sample t-test

All three stimuli types show significantly larger amplitudes for chirps over clicks, supporting the hypothesis that the increased synchronicity happens over the entire frequency range. It is also shown that the high-frequency improvement was

significantly different from the broad-band improvement, and thus the high frequencies cannot be the entire explanation for the larger amplitude measured with a chirp instead of a click. It cannot be rejected that the improvement measured with the low-frequency stimuli are equal to the improvement of the broad-band conditions. These results will be further discussed in the discussion section.

DISCUSSION

This study investigated the frequency regions contributing to the chirp ABR Wave-V amplitude. It was found that an increase in ABR wave-V amplitude when stimulating with a chirp stimulus rather than a click, was observed both at lower and higher frequencies, indicating that the increased synchronicity of the nervous responses takes place across the entire frequency range. It was also shown that the high-frequency region cannot explain the improvement from click to chirp when stimulating with the broad-band stimuli. However, the improvements observed at the low-frequency conditions and the broad-band conditions were not significantly different, indicating that the lower frequencies can explain all the improvement from the click to chirp condition. This contradiction in the results, that the high-frequency improvement is significantly larger than zero, and that the low-frequency improvement is not significantly different from the broad-band improvement, would likely be clarified if more test subjects had been used.

Fig. 3 shows that high frequencies were the main contributor to the formation of ABR Wave-V amplitudes for both clicks and chirps. This was likely due to the fact that the high-frequency stimuli contains more power, and to the fact that the high-frequency basilar membrane responses have short impulse responses that were inherently better aligned than the longer impulse responses at low frequencies. However, the improvement from click to chirp at high frequencies was small.

In Fig. 5 the amplitudes of the low-frequency and high-frequency responses were added for each test subject and compared to the broad-band evoked amplitudes. It is clearly observed that the summed amplitude is larger than the broad-band evoked amplitude. This shows that the auditory pathway behaves nonlinearly. The explanation is that the outer-hair-cells (OHC) amplifies weak sounds more than louder sounds (compression) and the fact that the filtered responses gives rise to spread of excitation on the basilar membrane in the region surrounding the 1500 Hz cut-off frequency. The 1500 Hz region would in the broad-band conditions have been masked. The low level "off-frequency" excitation will be amplified by the OHC and the summed response of the two frequency limited conditions will thus be stronger than the one measured with the broad-band stimulus. The increased amplitudes observed with the summed low and high responses, are though equally large for both click and chirp stimulus. This leads to a very limited effect on the wave-V improvements shown in Fig. 4, and the possible uncertainty regarding the unmasked off-frequency effects were thus negligible.

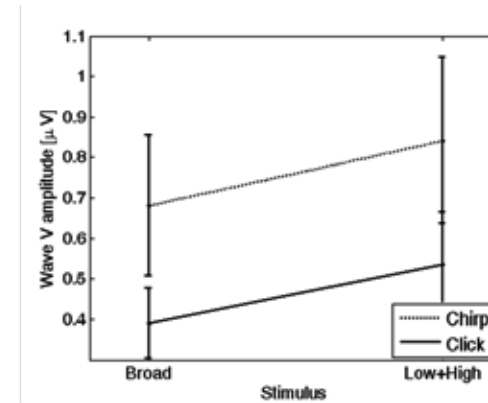


Fig. 5: ABR Wave-V for filtered stimuli are added for each subject and compared to data for broad-band. The mean and one standard deviation are shown.

CONCLUSION

This study examined the influence of frequency range on chirp evoked ABR at a presentation level of 40 dB HL. It was shown that both low and high frequencies contribute to the increase in wave-V when using a chirp stimulus instead of a click stimulus. This demonstrates that synchronisation of nerve fibres occur across the entire frequency range. However, the largest increase in wave-V is observed at lower frequencies.

REFERENCES

- Dau, T., Wegner, O., Mellert, V., and Kollmeier, B. (2000). "Auditory brainstem responses with optimized chirp signals compensating basilar membrane dispersion" *J. Acoust. Soc. Am.* **107**, 1530–1540.
- Dau, T. (2003) "The importance of cochlear processing for the formation of auditory brainstem and frequency following responses" *J. Acoust. Soc. Am.* **113**, 936–950
- Elberling, C., Don, M., Cebulla, M. and Stürzebecher, E. (2007). "Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay" *J. Acoust. Soc. Am.* **122**, 2772–2785.
- 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.
- Greenwood, D. (1990). "A cochlear frequency-position function for several species - 19 years later" *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Kiang, N. (1965). "Discharge patterns of single fibers in the cat's auditory nerve" Research monograph no. 35., The M.I.T. press, Cambridge, Massachusetts.
- 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,
- Richter, U. and Fedtke, T. (2005). "Reference zero for the calibration of audiometric equipment using 'clicks' as test signals" *Int. J. Audiol.* **44**, 478–487.
- Shore, S. E., and Nuttall, A. L. (1985). "High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts" *J. Acoust. Soc. Am.* **78**, 1286–1295.

Applying physiologically-motivated models of auditory processing to automatic speech recognition

RICHARD M. STERN

Department of Electrical and Computer Engineering and Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213 USA

For many years the human auditory system has been an inspiration for developers of automatic speech recognition systems because of its ability to interpret speech accurately in a wide variety of difficult acoustical environments. This paper discusses the application of physiologically-motivated approaches to signal processing that facilitate robust automatic speech recognition in environments with additive noise and reverberation. We review selected aspects of auditory processing that are believed to be especially relevant to speech perception, "classic" auditory models of the 1980s, the application of contemporary auditory-based signal processing approaches to practical automatic speech recognition systems, and the impact of these models on speech recognition accuracy in degraded acoustical environments.

INTRODUCTION

It is well known that human speech processing capabilities far surpass the capabilities of current automatic speech recognition and related technologies, despite very intensive research in automated speech technologies in recent decades. Indeed, these observations have motivated the development of feature extraction approaches for speech recognition systems that are motivated by auditory physiology and perception since the early 1980s, but it is only relatively recently that these approaches have become effective in their application to computer speech processing. We begin this paper with a brief review of some of the major physiological phenomena that have been the object of attention by developers of auditory-based feature extraction methods. We continue with a brief review of three seminal "classical" auditory models of the 1980s that have had a major impact on the approaches taken by more recent contributors to this field. We then discuss some of the topics that are foci of contemporary auditory models and provide some examples of current efforts. Finally, we describe the results of a limited number of representative experiments that demonstrate the effectiveness of auditory modelling for automatic speech recognition, concluding with a brief discussion of the attributes of these models that appear to be most effective in improving recognition accuracy.

RELEVANT AUDITORY PHENOMENA

Most classical and current auditory models are based on capturing a small number of rather basic physiological phenomena, all of which are quite familiar to researchers in