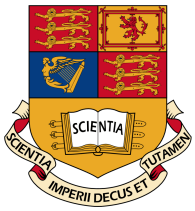


Imperial College London

Final Report for MEng Individual Project

Department of Biomedical Engineering

Computational Modelling of the Auditory Brainstem Response to Speech

Author:
Amna Samar Askari

Supervisor:
Dr. Tobias Reichenbach

Abstract

This report presents a unique computational approach which combines phenomenological and functional modeling to reproduce the electrical response to speech stimuli that is recorded from human subjects using scalp electrodes. This response is generated by subcortical neural regions in the auditory pathway and is called the auditory-brainstem response (ABR). Existing inner ear and Auditory Nerve (AN) models written in the Python coding language are combined with neuro-computational models for globular bushy cells in the Cochlear Nucleus (CN), to develop a common simulation framework that works with a speech sample of arbitrary length. Using spike train data analysis, an appropriate average measure for the collective electrical response along different points in the brainstem pathway i.e. at the AN level and at the CN level is devised. The resulting electrical signal from the CN is correlated with a novel representation of the speech signal that is constructed using empirical mode decomposition (EMD) methods. The cross correlation result is periodic, and has peaks at particular latencies, with the largest and most characteristic peak at 6.57ms. This supports the experimental efforts carried out which found a peak at the latency of 9.3ms that is meant to correspond with the output from the Inferior Colliculus (IC), which is further along the pathway from the CN and hence includes a larger delay.

Acknowledgements

I would like to thank a few key persons without whose help, this project would not be possible:

- My wonderful supervisor, Dr. Reichenbach, for inspiring and guiding me throughout the year with excellent advice, support and background knowledge.
- Antonio Forte, firstly for organizing the experimental ABR tests which are the basis for proving the success of this model. Secondly, for always providing detailed and clear answers to the endless amounts of questions I had (even till the very last moment).
- Shabnam Kadir, for assisting me in optimizing and debugging neuron models.
- Alexander Camuto and Jack Campbell, whom I built the firing rate estimation function with and who have helped me with various computational endeavors throughout the last two years.

Software Packages Used

The underlying code for this project is based on and has taken support from pre-existing code that has been developed over the years on inner ear and neuron models. Hence, I would like to credit the following open source packages that I have used which have been written by brilliant researchers in Auditory and Computational Neuroscience:

- **Cochlea 1.2.2** A package in Python that includes inner ear and cochlear processing models. https://github.com/mrkrd/cochlea
- **Brian 1.4.3** A package in Python that is used for constructing models for neurons and neuronal groups. https://github.com/brian-team/brian
- **Thorns 0.7.3** A package in Python that is used to analyze and display Spike train data. https://github.com/mrkrd/thorns
- **Cochlear Nucleus 1.4** A package in Python that includes detailed models of Globular Bushy Cells in the Cochlear Nucleus, which contains information on specific ion channels, synaptic dynamics and an accurate depiction of neurophysiological structures like endbulbs and the soma. https://github.com/mrkrd/cochlear_nucleus
- **Auditory Modeling Toolbox** A set of scripts in MATLAB that include most of the Auditory Neuroscience computational models and their supporting data [1] http://amtoolbox.sourceforge.net

Nomenclature

Clarifying repeated terminology in the report that has multiple known representations:

- Depending on the kind of stimulus, the recorded response from subcortical brain areas is referred to as different terms in relevant literature. The umbrella term used for recorded responses is the **AEP** or Auditory Evoked Potential. For transient inputs, this response is known as the **ABR** or Auditory Brainstem Response. For temporally varying inputs that have a fixed or changing frequency, the response is called the **FFR** or Frequency Following Response, and a subset of this includes the **EFR** or Envelope Following Response which is a term used for longer duration amplitude modulated tones. Lastly, for complex stimuli like speech, the response is called the complex ABR or **cABR** (Ronne, et. al 2013) [2].

- As this project is centered around evaluating responses to acoustic stimuli, it is important to clarify the units that these are measured in. Sound is created by changes in air pressure and hence measured in Pascals or **Pa**. In auditory research and in this paper, the term **dBSPL** is used to describe the 'Sound pressure level' (SPL) or the intensity of a stimulus in decibels (dB). This is the logarithmic ratio between the effective pressure value of the signal in Pa to the value of pressure for the lowest intensity sound that can be heard by most healthy individuals (20μPa).

Contents

1 Introduction

1.1 The recorded ABR

After being collected in both ears and before further analysis and perceptual registration in the Cortex (Steinschneider, et. al 1994) [3][4], an auditory stimulus i.e. sound is processed by a series of brain areas that are collectively termed as the 'Auditory Periphery' or the 'Auditory Brainstem', which is the region of interest for this project. Figure 1 below shows a brief outline of the auditory processing pathway including neural components of the brainstem:

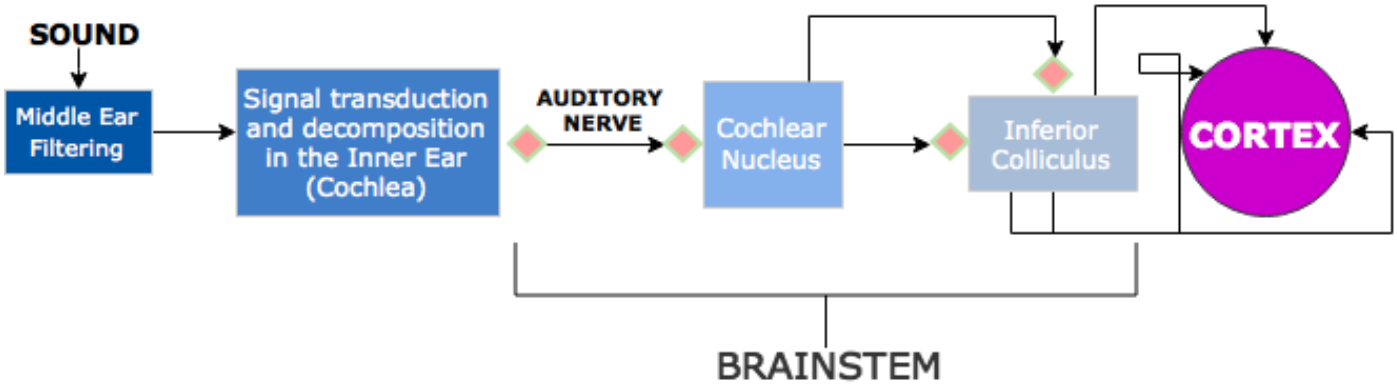


Figure 1: Display of sequential auditory processing stages after sound stimulus is received by the ear. Synapses with important dynamics are indicated by orange diamonds and parallel pathways that emerge after the Inferior colliculus (IC) are indicated using multiple arrows.

The collective electrical response from this region is the result of a concatenation of complex neural processes and can be recorded using electrodes placed on the scalp of a subject. The process is similar to Electroencephalography (**EEG**) and is hence noninvasive and measures overall electrical activity picked up by the electrodes. The recorded signal has different terminology depending on the kind of acoustic stimulus (*see Nomenclature above*), but will be referred to as the auditory brainstem response or 'ABR' throughout this report. Traditionally, the ABR has been used as a measure in detecting healthy hearing in infants [5]. However, over the years it has developed into a crucial tool for several research purposes including developing accurate clinical hearing assessments and investigating the mechanisms behind cochlear function to better understand the nature of conditions like cochlear neuropathy [6] [7] and other moderate forms of hearing loss. There have been significant experimental results from recording the ABR to complex signals like speech in several efforts to better understand the neural mechanisms behind attention [8] [9] and to improve and build perceptual models [10].

In the ABR hearing test, patients are made to hear a simple click which is a transient pulse of a particular amplitude or more appropriately, 'level' which is measured in dB SPL (*see Nomenclature above*). The experimentally determined ABR to this stimulus is a series of seven distributed peaks at particular time stamps, all within the span of 10-12ms [11][12]. The entire recorded response consists of more peaks that occur at later times but those aren't considered to be part of the ABR. Although the exact neural basis for each peak is not well understood, it has been suggested that the most robust peaks i.e. peak I, III and V are the result of synchronised neural activity of Spiral Ganglion cells in the Auditory Nerve, Spherical cells in the Anterior Ventral part of the Cochlear Nucleus and Principal cells in the Medial Superior Olive respectively[13] [14]. Therefore, each group generates a response at a particular time delay which is the time that it takes for the stimulus to reach it. Figure 2 shows a basic experimental set up, the recorded response and the associated neural basis for these seven peaks.

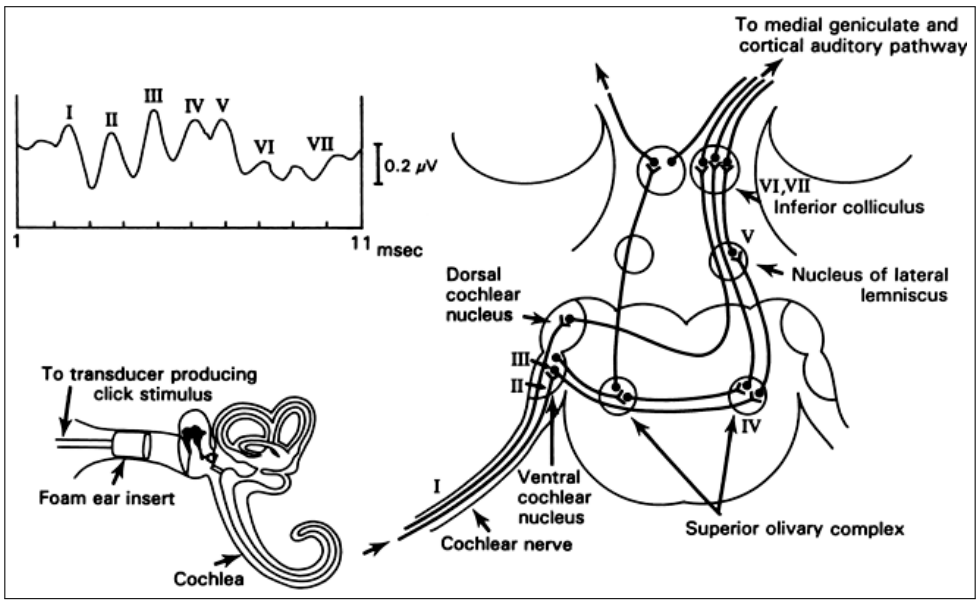


Figure 2: Brief sketch of experimental setup to record ABR for click stimuli and assignment of ABR peaks to particular neuronal groups along auditory pathway. Taken from Dr. Richard Vogel's Neuromonitoring blog[15]. *Note: the group responsible for the fifth peak here is in the Lateral Leminiscus which conflicts with other literature mentioned above.*

In response to stimuli with temporally varying characteristics such as frequency, the peaks form a continuous signal that resembles the temporal structure of the stimulus. For periodic stimuli, the response contains distinct peaks that correspond to its periodic peaks [16]. Hence, these responses are also known as 'Frequency Following' responses or FFR (*see Nomenclature above*). Stimuli that have been used to generate FFR's include chirps, tone bursts [11] and simple amplitude modulated tones [17]. For more complex signals such as speech that contain multiple frequency components that vary differently with time, it is more difficult to extract a characteristic response. As the amplitude of the ABR is in the order of microvolts and there is a lot of noise generated through recording mechanisms and the mains source, it is even more difficult to record a significant result in response to a non repetitive signal such as speech [8]. Therefore in efforts to analyse responses to speech, usually some alternative representation of the response is taken either using reconstruction based on statistical properties such as RMS, Fourier analysis [16] [10]or another signal processing routines. In Kraus et.al (2010), the most basic element of speech i.e. a vowel sound is essentially decomposed into a transient component and a periodic component, for which the response is analysed separately.

1.2 Modelling the ABR

After a thorough literature review and close analysis of computational models for ABR generation, I have concluded that the ABR modeling process can be divided into two stages:

1. Having a mathematical/computational representation for the non linearities in the Cochlea and IHC-AN synapse

2. Transformation between the neural representation of the output at the Auditory Nerve and the recorded potential at the scalp

Stage 1: The spread of excitation in the cochlea, where sound transduction occurs and the response of the auditory nerve (AN) , which carries this transduced signal into the brain are fundamental processes

behind ABR generation. Therefore, efforts to model these and the ABR are one in the same and have reinforced each other over the years. These models have been extremely important in developing cochlear implants [18] [19], researching hearing loss and the effects of cochlear lesions [20] and serve as a basis for more complex perceptual models [21] [22]. The earliest models assumed basilar membrane linearity but this was soon changed to develop non linear models that were more representative of what happens in the cochlea [23]. Some non-linear effects in the cochlea include compressive filtering, suppression, shift in tuning according to SPL, and envelope dependent hair cell behavior. AN response properties result from similar non linearities [24]. The link between the cochlea and AN fibres occurs at the synapse between the inner hair cell and the AN which is also known as the IHC-AN synapse. This synapse has extremely important properties that determine the behavior of the electrical signal in the AN, the most important one being that of adaptation [25] [23] which was explained in the 3 store diffusion model of Westerman and Smith 1998 [26] and included in a model introduced by the Carney lab in 1993 but fully implemented in Zhang et. al 2001 which models cochlear filtering using a set of parallel filterbanks. These filterbanks were specifically tuned to the characteristic frequencies (CF) of the AN fibres which are the values for which corresponding hair cells in the cochlea are most sensitive to and hence encode the signal for. [23]. An increase in stimulus level increases the breadth of component filters and hence leads to a broader excitation pattern in the cochlea [2] [11] [27]. The frequency dependency of the latency is also explained using the delay characteristics of these component filters [28]. Most models to date are phenomenological as they mimic the signal processing effects which occur in the auditory pathway and that have been uncovered by using different types and modes of stimuli and recording the result from them. There have been groundbreaking results from these models, including using chirp stimuli to uncover that the ABR is not the typical event based response whose peak indicates the on or off of a stimulus, but it conveys the amount of neural synchrony after cochlear processing [11] [29]. However, models can also be functional which means that they directly model the neural processes that are responsible for the response in question. The only time this has been successfully carried out is in Verhulst 2015 where the final ABR response is built from first principles using single unit AN fibre responses [12].

Stage 2: Most of the research discussed above has ended at producing an AN output. Dau et.al (2003), Ronne et. al 2012 and Verhulst et.al (2015) are efforts that have executed stage 2 i.e. onward from the AN. In Dau et.al (2003) and Ronne et. al 2012, the instantaneous rate of the AN fibres is summed across the different characteristic frequencies and then "deconvolved" with experimental data to produce an ABR result. Deconvolution is a complicated mathematical operation but is meant to be in spirit with the theory set out by Goldstein and Kiang in 1958 [30], which claims that the output of a neural unit can be expressed as a convolution of a rate function and its unique unitary response. Although this approach is successful in matching experimental results, its major flaw lies in the fact that the entire second stage is based on experimental data from one particular stimulus. Hence, it is not reasonable to use this sort of approach in modeling speech as it is not feasible to calculate a representative deconvolution waveform for speech as it is highly variable and complex. The sum of the instantaneous rate is again taken in the Verhulst approach, but instead of being used in a deconvolution operation, is taken as a dependent variable in a set of differential equations that model the behavior of the CN and the IC. Although this model is not exactly functional because it is based on the phenomenological representation of the CN and IC in Nelson and Carney 2004[17], it at least tries to represent the neurological phenomena that occur after the IHC-AN synapse. The issues with the second stage in these models in modeling speech is that the specific parameters derived in them depend on simulations or experiments with specific stimuli. The rate constants for the CN and the IC in the Carney model and the conversion constants from basilar membrane output to IHC output in the Verhulst model are highly specific and only suitable for clicks, chirps, tone bursts and amplitude modulated tones. It is difficult to model the CN and IC stages

from first principles as there are several different kinds of cells with different response properties and dynamics in each. Moreover, there is not just one CN but several parts of the CN (Posterior, Anterior and Dorsal) that have different distributions of cells and different levels of inhibitory and excitatory connections with AN fibres.

Moreover, there is an intermediary group of neurons where processing also occurs that lies between the two which is the Superior Olivary Complex. The role of this region in auditory processing is not well known or accurately depicted in literature. As shown in figure 1, the auditory pathway splits into several parallel channels on the way to and after the IC which is extremely difficult to directly model as some of them involve feedback from the Cortex.

In conclusion, to develop a model for speech, it is reasonable to use one of the existing models for Stage 1 i.e. cochlear processing however a unique approach must be devised for the second stage.

2 Methods

2.1 Experimental

The success of this computational model is determined by whether the result that it produces correlates with the experimental results. As discussed earlier, it is useful to use some sort of representative reconstruction of a speech signal for ABR tests as it is non repetitive and therefore doesn't lead to a significantly large response. Following the approach of Reichenbach et.al (2016) [8], a more sophisticated framework was built by Forte and Reichenbach et.al 2017 [9] to investigate the principles behind attention modulation. Empirical mode decomposition (EMD) was used to reconstruct a new signal, the fundamental waveform or F0, that has the same temporal variation as the fundamental frequency of the speech signal. A brief summary of the steps used to construct this waveform include downsampling the speech signal from 44.1 KHz to 8820 Hz, low pass filtering, windowing using autocorrelation and extraction of instantaneous frequency using Hilbert Transform methods. The following figure shows the speech sample used for the computational model along with its fundamental waveform:

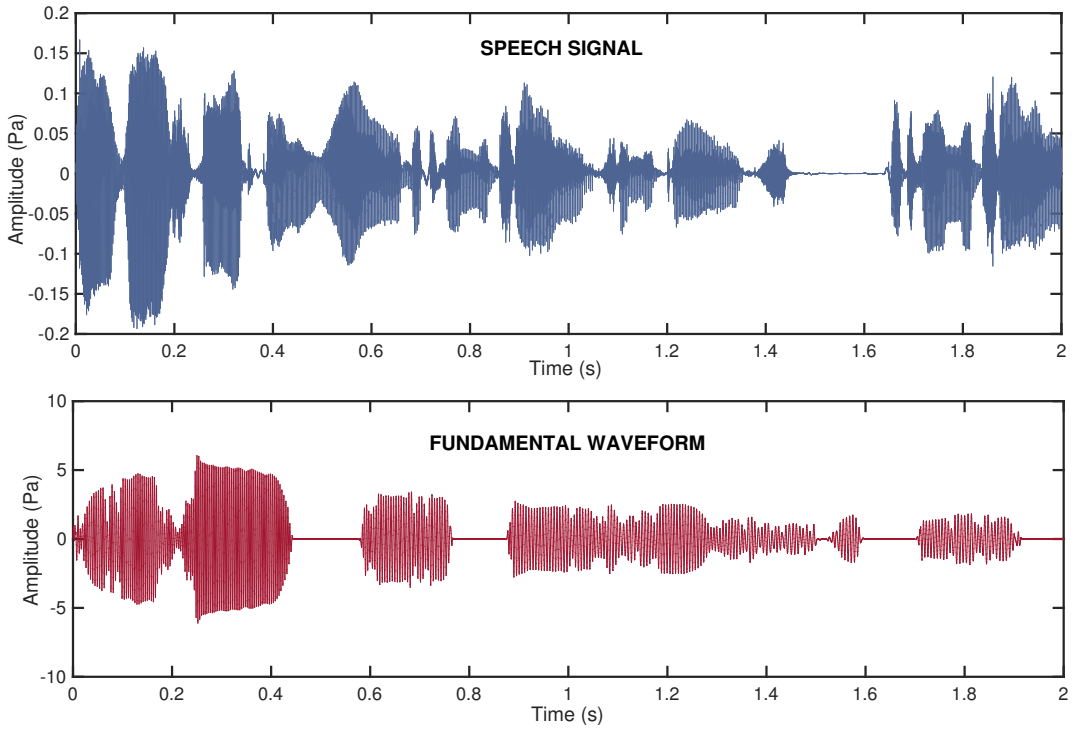


Figure 3: A two second sample of the speech signal, sampled at 44.1 kHz used for the computational model with its corresponding fundamental waveform or F0, sampled at 8820 Hz.

ABR recordings were carried out using scalp electrodes in 16 healthy adult volunteers aged 18 to 32 that agreed with written consent to participate in the study. Participants were made to sit in an acoustically and electrically insulated room and receive speech signals through ear phones. The speech signals were taken from audiobook recordings that were available online. These signals were sampled at a frequency of 44.1 Khz and delivered at 78 dBSPL. Two electrodes were positioned at the cranial vertex, two further electrodes were placed on the left and right mastoid processes, and the remaining electrode was positioned on the forehead as a reference for the ground voltage. The resulting ABR was obtained by averaging the signal recieved from both earphones, followed by noise removal, high pass filtering and bandpass filtering between 100Hz and 300 Hz. The high pass filter had a cutoff frequency

of approximately between 90-100Hz to ensure cortical signals, which usually have frequencies under 80 Hz, were removed. Details on the type of electrodes and associated electrical equipment can be found in the Appendices.

2.2 Computational

2.2.1 Cochlear Model Selection

A detailed comparison of various cochlear filter models that have been developed over the years can be found in Saremi and Steinfeld 2016 [31]. Initially, the DRNL filterbank was tested with a one second sample of the speech signal to analyse the output at different CF's. This is a sufficiently long sample length as it includes around 150-200 vowel sounds which have an approximate duration of 40ms [10]. Moreover, it is long enough to be termed as continuous speech, which is what we are testing the model to respond to. The algorithm was developed by the authors of the paper in the Carney lab and written in MATLAB code, which is found in the the Auditory Modelling Toolbox. The algorithm returns an array of CF's if they are not specified already. As the signal amplitude is technically still in Pa after being processed by the DRNL filterbank, a series of mathematical operations would be needed to carry out to convert the signal into spike trains. Moreoever, there have been significant additions to this model over the years and it has been optimized to have more flexibility and include a more accurate representation of the non linearities in the cochlea. Therefore, this model was rejected and the upgraded cochlear model of Zilany et.al (2014) is used for the final computational model, which is also from the Carney lab. 30 CF values were created in the range of 125 Hz to 8KHz according to Greenwood's cochlear function (Greenwood et. al 1990). Greenwood constructed an excellent mathematical representation of the relationship between the value of particular frequencies and where they are received with the highest sensitivity i.e. the location of the hair cell that is specifically tuned to that characteristic frequency (CF). This function is based on a thorough analysis of psychoacoustic and physiological data from various species including humans. The relationship between the distance from the apex of the Basilar membrane in metres and the frequency value in Hz is plotted in the figure below and can be expressed in the most basic representation as:

$$10^{Ax-k} \tag{1}$$

where a,k are constants that depend on the species. The value of A is 165.4 for humans and k is an integration constant that is usually taken to be 0.88 for humans.

The Zilany et.al (2014) model is dependent on source code written in C and has auxiliary scripts written in Python. The code was run through the Cochlea package, written in Python. This model includes a variety of parameters that can be tweaked including:

1. Subject species: the options include humans, cats and guinea pigs. For this project, human was chosen.

2. The type and number of AN fibres, which include high spike rate (HSR), medium spike rate (MSR) and low spike rate (LSR) fibres. The values chosen were based on the functional modeling in Verhulst 2015 [12] and were 13, 3 and 1 respectively which makes a total of 17 fibres.

3. The inclusion of Gaussian Noise and Power law adaptation. As noise was irrelevant to this study, the former was ignored. However, as adaptation in the IHC-AN synapse is an important principle we are trying to model, the latter was included.

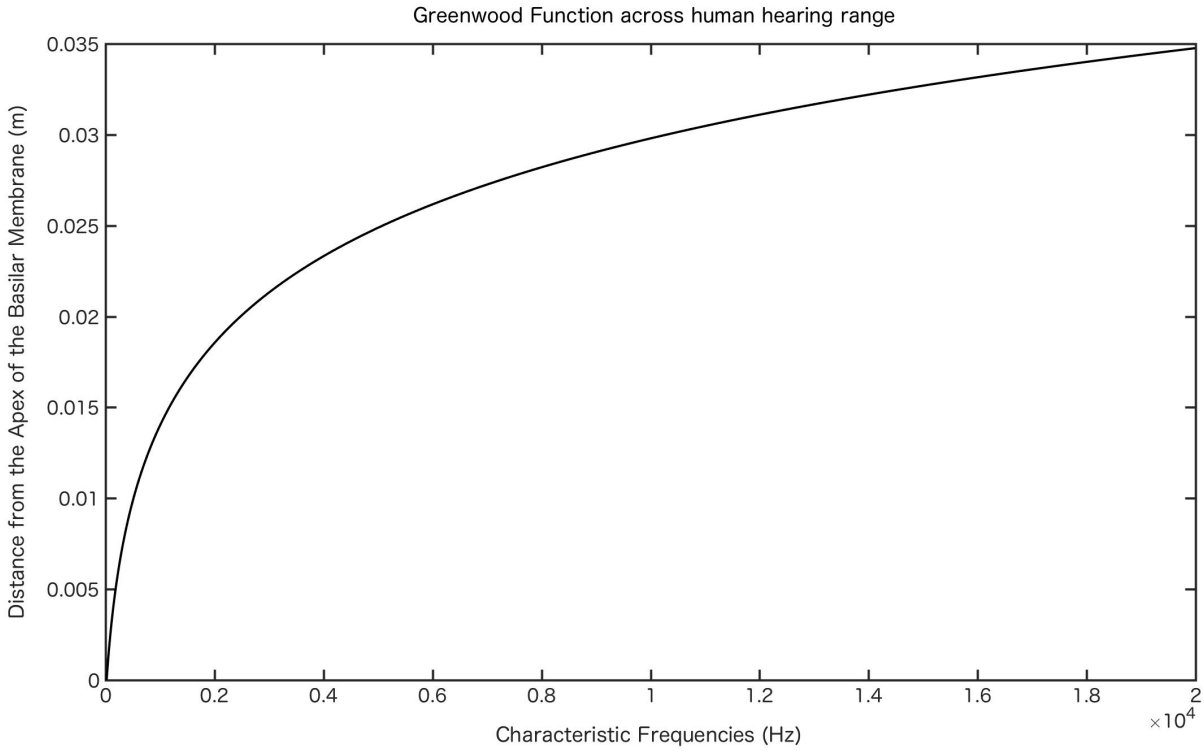


Figure 4: Greenwood Cochlear Function

The output of this model is recorded in a Pandas (Python package) dataframe which contains precise spike train information for each fibre i.e. specific times when spikes occured. This is then converted into a Neuron Group in the Brian simulator package in Python, using a local function in the Cochlea package, so that we are able to create simulations and connect this group with other groups along the brainstem. As this model runs on a sampling frequency of 100 KHz, the speech sample and the fundamental waveform F0 were upsampled to 100KHz to ensure a uniform sampling rate for the entire simulation and cross correlation calculations for the results.

2.2.2 Representing the brainstem after the IHC/AN synapse

In order to functionally model the neural processes after the IHC-AC synapse, the Cochlear Nucleus package in Python was used which has a detailed construction of Globular Bushy Cells (GBC) of the CN in the Brian package [32]. A group of 200 CN Neurons was linked to the 510 AN fibres (30 neurons per fibre). This includes various details of synaptic dynamics, Potassium and Sodium ion channels and end bulb characteristics. The only change that was made to the code was to ensure that the delay in CN processing was included by tweaking the 'Connections' class in Brian. After running the simulations with connected GBC's, the output was another Pandas dataframe with GBC spike train data.

2.2.3 Neural data processing and finding a measure for ABR

As the ABR represents collective electrical activity and not single unit responses, a measure was devised to average the spike rates based on the spike train data. Initially, Brian's local Population Rate function a few other inbuilt functions in spike analysis software were tested to calculate the average rate. Eventually, the Pandas dataframe was converted into an array which was exported into MATLAB and a MATLAB

function was written that converts the spike times into a binary sequence which contains 1's at the specific sample number when a spike occurs. This was carried out by identifying indices of spike times and converting them into a sample number. The resulting binary array has dimensions of [X,Y] whereby X is the number of neurons and Y is the length of the sample i.e. 100'000 time points (Duration of 1s * Simulation sampling Frequency of 100kHz). Then, using a linear kernel of a width of 2 samples, an average firing rate was calculated for the whole population of AN and CN neurons. This firing rate was a row vector of dimensions [1,Y] where Y is the length of the sample. This signal was cross correlated with the real and imaginary parts of the Hilbert transform of the fundamental waveform F0.

3 Results

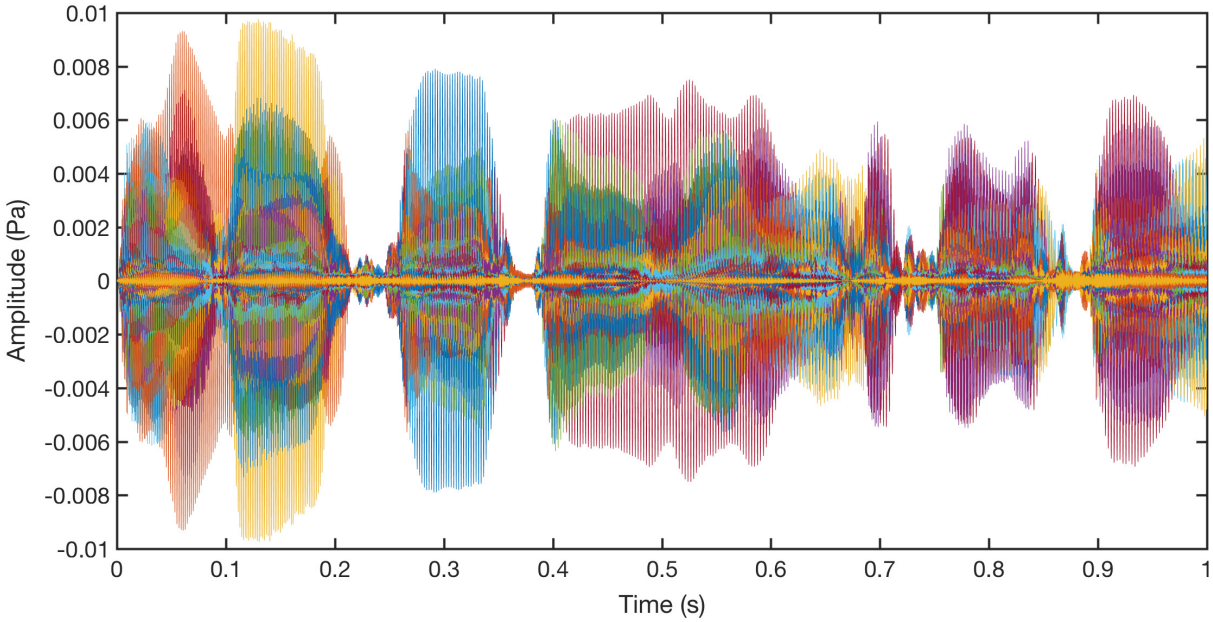


Figure 5: Output signals from 31 different CF channels in the DRNL filterbank. See Appendix for a more detailed plot.

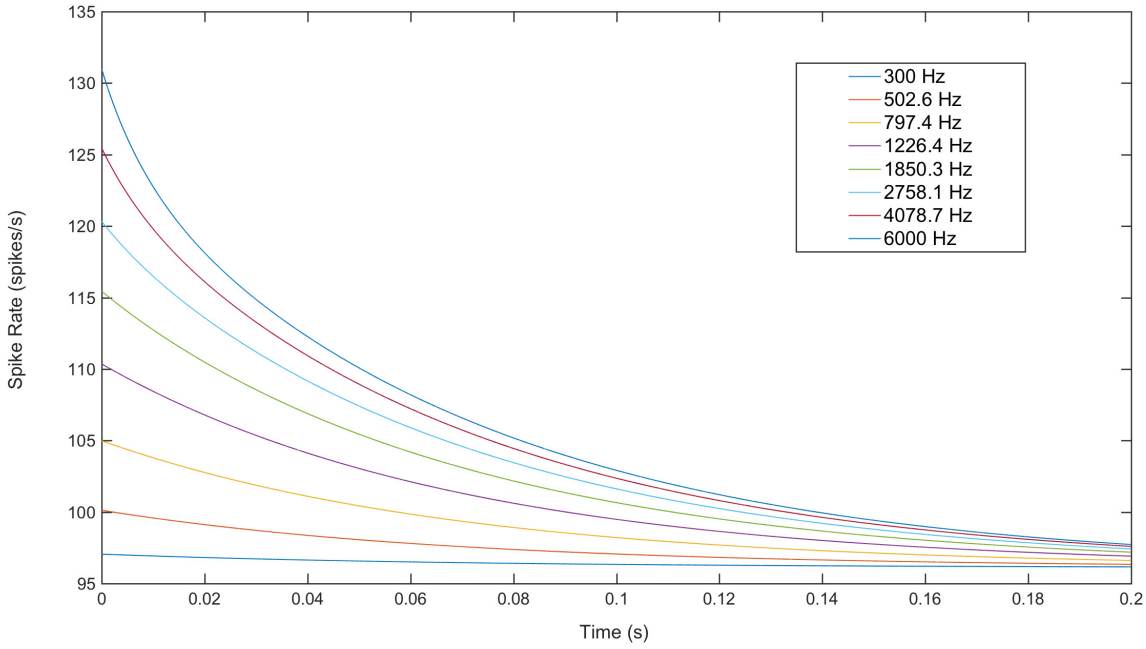


Figure 6: Spike rates generated from Zilany model from AN fibres across 8 characteristic frequencies under no stimulus. The rate shows exponential adaptation of the IHC-AC synapse which makes the value drop from an initial spontaneous value.

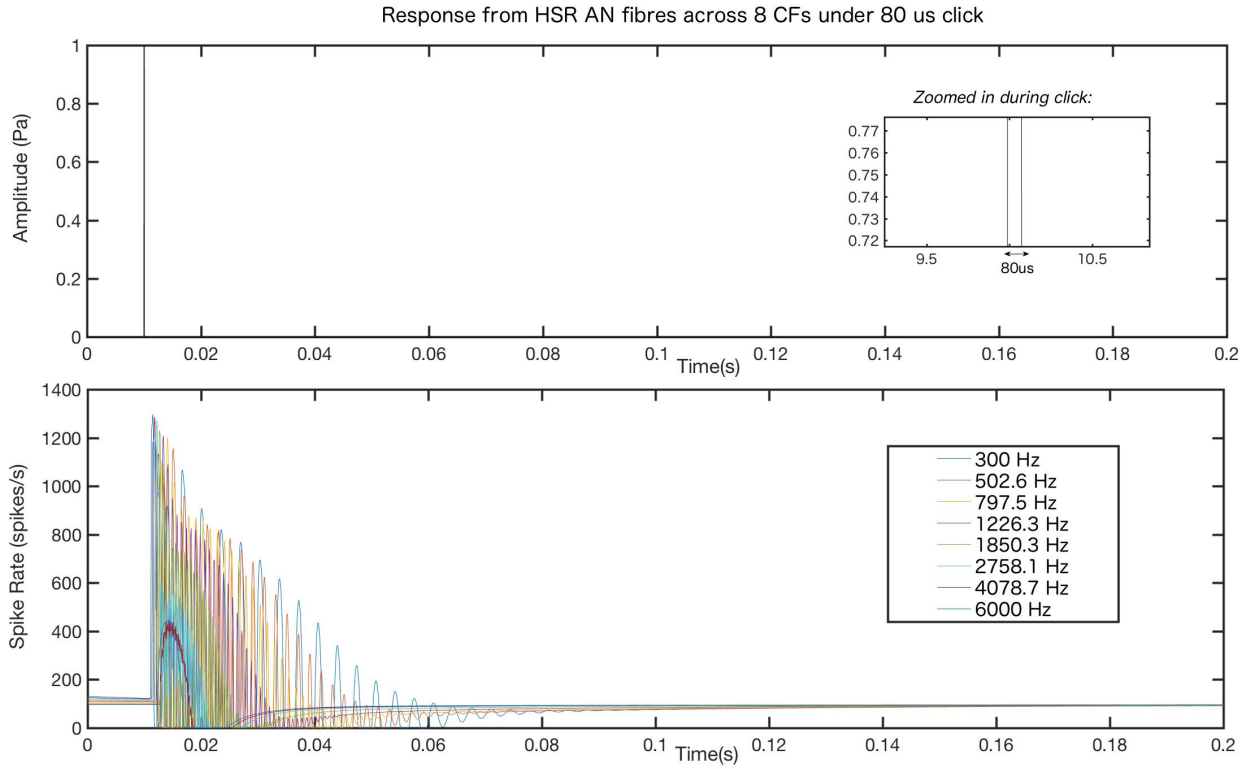


Figure 7: Response of AN fibres across 8 characteristic frequencies for Zilany model to an 80μs click, showing the exponential adaptation of the IHC-AN synapse. The click appears to be a vertical line in the first figure but is zoomed in in the same box to show that it has a finite duration.

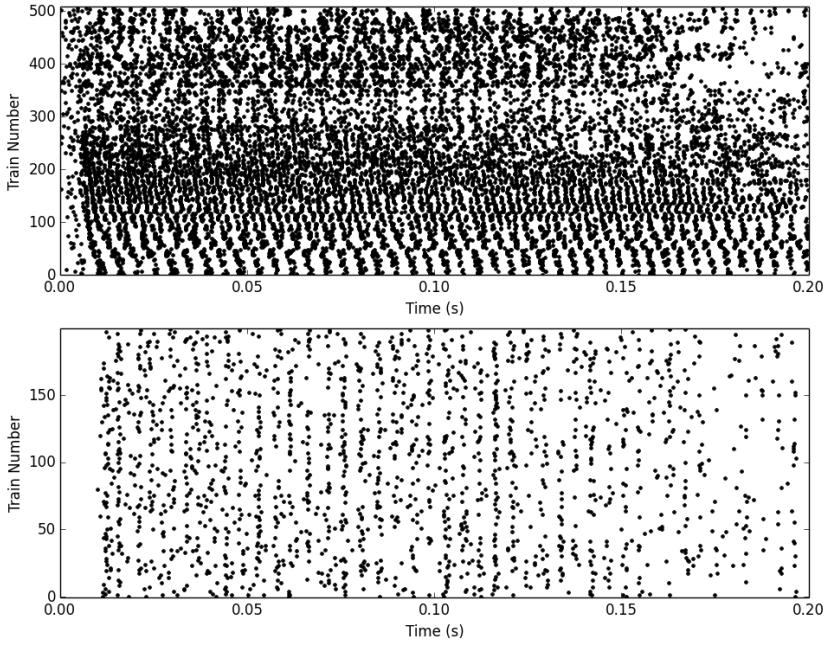


Figure 8: Raster plots indicating the presence of spikes at particular time values for AN and CN populations. The Y axis represents spike trains from each neuron.

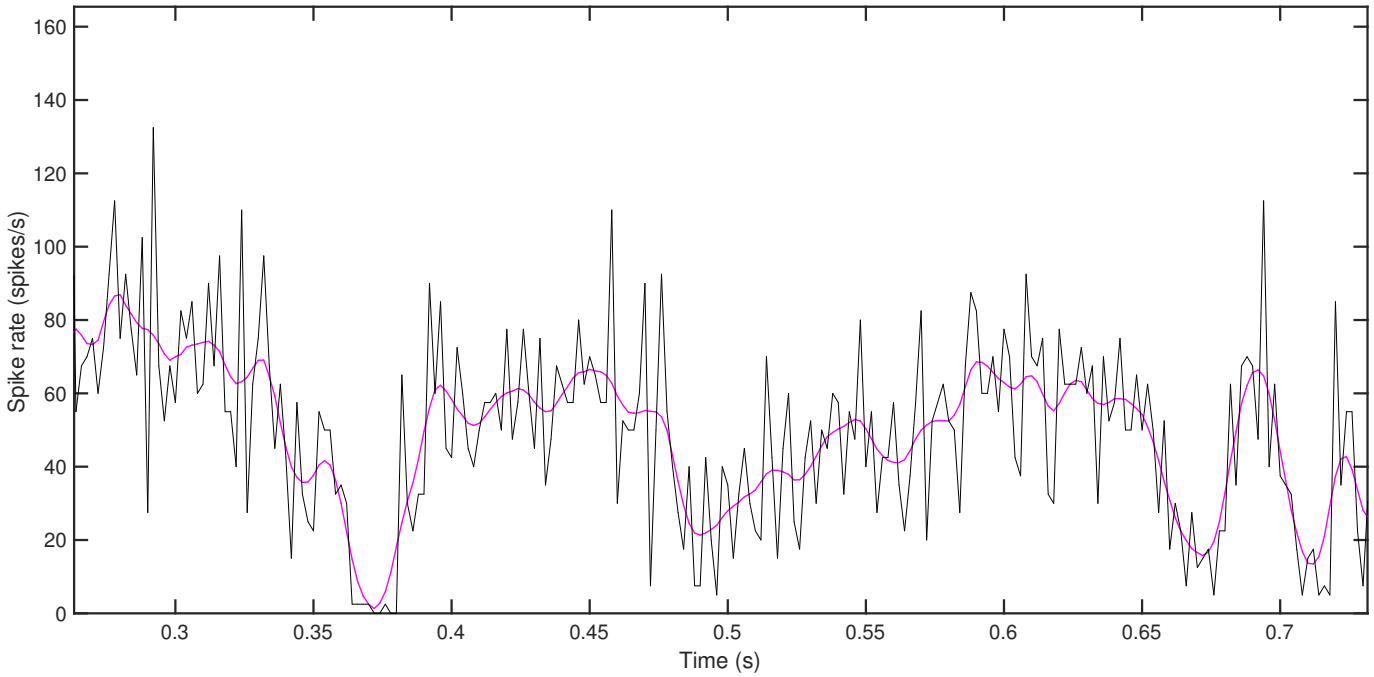


Figure 9: Spike rates calculated using Brian's local Population Rate function (black). Spike rates calculated in Brian using a smoothing Gaussian kernel (pink).

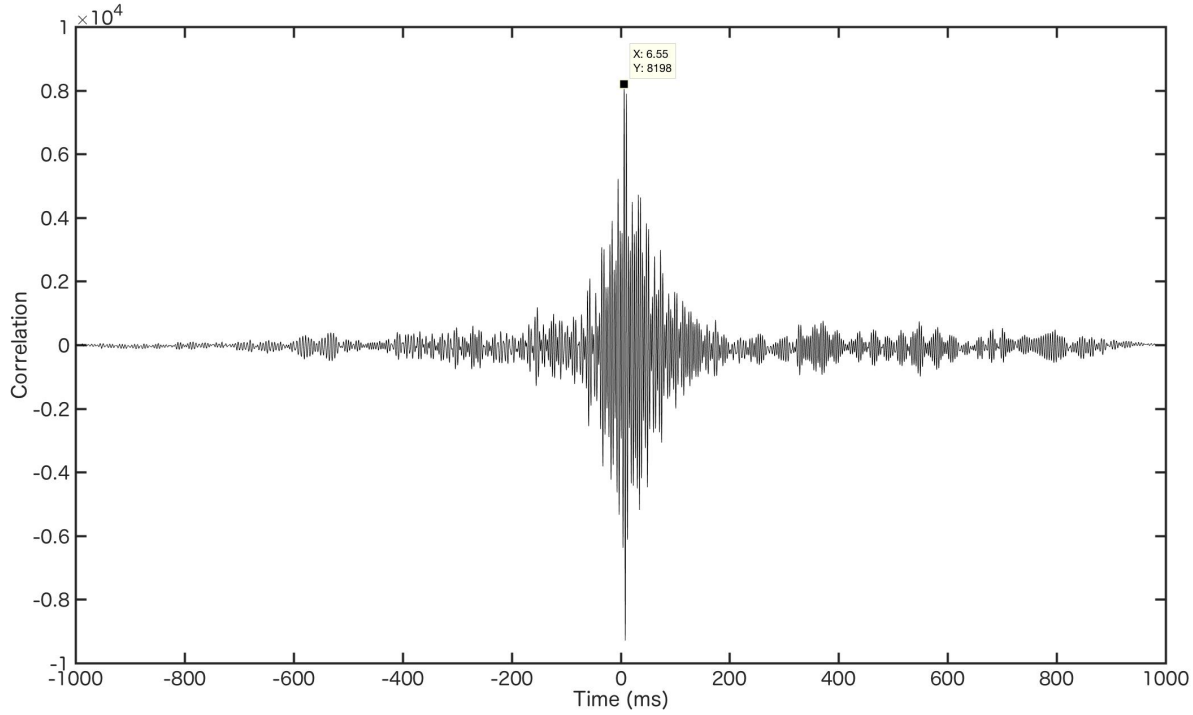


Figure 10: Cross correlation of CN Rates obtained from **MATLAB firing rate function** with the fundamental waveform F0 for the entire simulation length of 1s. The peak of this correlation at the latency of 6.55ms is highlighted.

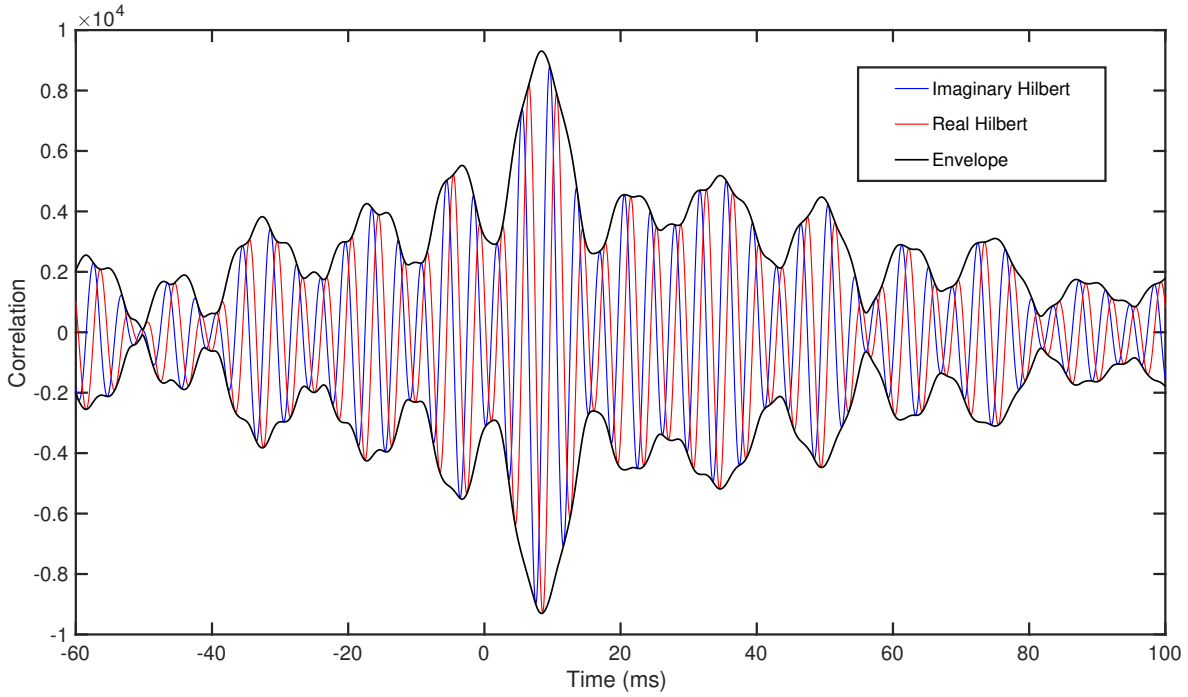


Figure 11: Cross correlation of CN Rates obtained from **MATLAB firing rate function** with the real and imaginary parts of the Hilbert transform for the fundamental waveform F0, along with its envelope for a 160ms range of time lags.

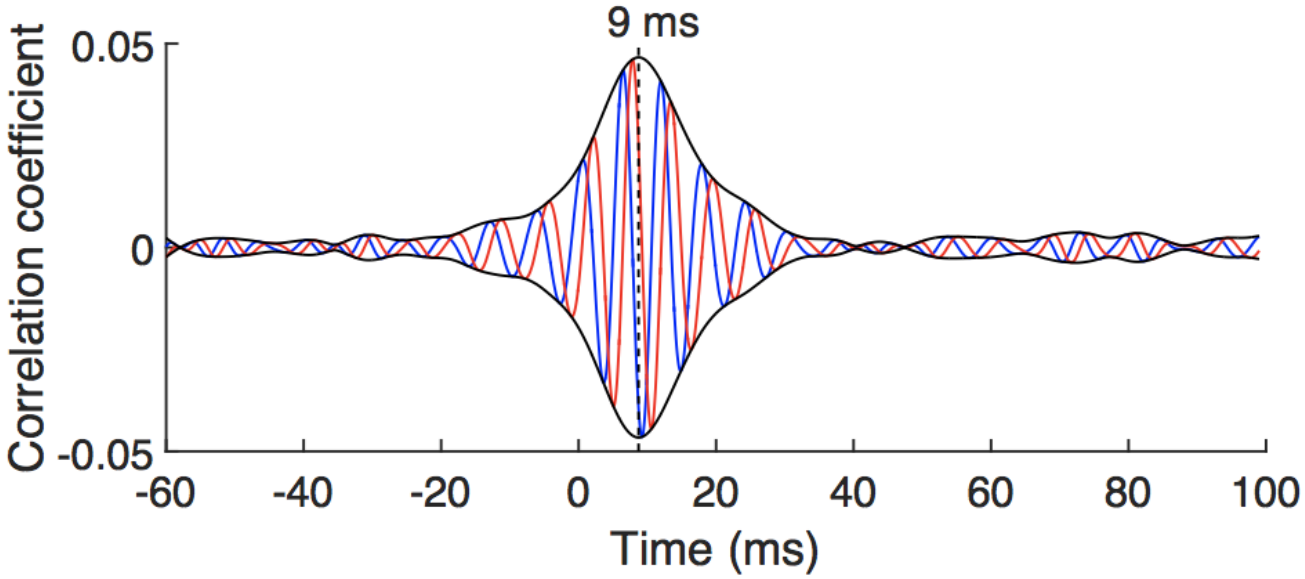


Figure 12: Cross correlation of CN Rates obtained from **experimental data** with the real and imaginary parts of the Hilbert transform for the fundamental waveform F0, along with its envelope for a 160ms range of time lags.

Figure 5 shows the result from the DRNL filterbank with the 1s long speech signal as an input. The components of 31 different CF's between 100kHz and 8kHz are represented in different colours on the same plot. An individual plot of these with comparative amplitudes can be found in the Appendices. Figures 6 and 7 are a result of testing that the Zilany et. al 2014 model worked locally after compiling

all the source code. For the purposes of these tests, 8 different characteristic frequencies between 300Hz and 6KHz were used. Figure 6 is the spike rate of the AN fibres plotted for 200ms of no stimulus. The initial values of the spike rates are a testament to the spontaneous response of the AN fibres, and the exponential decrease agrees with their behaviour in the litereature [33]. The model was then tested with a a basic 80 μs click. The peak of the response is seen to be at a small delay after the start of the click, which consolidates experimental and theoretical findings. Moreover, the shape of the onset peak followed by frequency specific exponential decays is consistent with the three store diffusion model for the IHC-AN synapse by Westerman and Smith, 1998 [26]. Figure 8 is a raster plot based on the spike train data for the AN fibres and GBC neurons. The raster plots for both AN fibres and GBC neurons are contained in the figure below. There is an observable delay in the start of spikes for GBC's. Moreover, there is a clear difference in spike train density between the two groups, with the AN exhibiting CF dependent delay characteristics which are consistent with the results in Dau et.al 2003 [11], Elberling et.al 2010 [29] and Ronne et.al 2013 [2]. Figure 9 shows the result of calculating rates using a local function in Brian, along with an attempt to smooth these rates using a Gaussian kernel. Although it closely represents the changes in spiking density, this rate function does not aptly capture the delay occurring in the start of GBC spike trains. Visibly, the smoothing process corrupts the periodicity of the waveform and hence does not give viable results for cross correlation. Figure 10 shows the result of cross correlating the CN rates obtained from the MATLAB function that was written, for the entire simulation duration of one second. There is visibly a peak at the latency of 6.55ms which corresponds to being generated by the GBC's in the Cochlear Nucleus.

As the brainstem response may occur at a phase that is different from that of the fundamental waveform, the experimental and computationally generated ABR is correlated with the Hilbert transform of the fundamental waveform that has a phase delay of 90°. The two correlations can be viewed as the real and imaginary part of a complex correlation function that traces out the ABR at any phase delay. The amplitude of the complex correlation informs then on the strength of the brainstem response. Experimentally, it was found that the amplitude of the complex correlation peaked at a mean latency of 9.3 ± 0.7 ms, and statistical analysis showed that this peak was significantly different from the noise in fourteen out of sixteen subjects. This latency value agrees with that found previously regarding the brainstem's response to short repeated speech stimuli (Skoe and Kraus 2010) [34]. The computational peak for the complex correlation is found at 9.5 ms, which agrees with the experimental data. Figures 11 and 12 show the computational and experimental result with these various waveforms respectively.

4 Discussion

We have presented a combined experimental and computational approach to quantify and model the ABR to speech signals. The success of the computational model has been determined by the similarity in latency values for the cross correlation of the ABR with the fundamental waveform of the speech signal. The advantages of the computational model are that it is highly flexible and can be used with a variety of physiological parameters, any speech signal of arbitrary length and is relatively computationally inexpensive. The total time taken for the simulations is around 58 seconds. However, the model could be made more accurate by increasing the number of CF's which could be done by shifting some of the processing operations to the GPU so that large matrices can be handled in parallel pools. The fundamental waveform is an essential component of our modelling approach. This waveform is constructed using EMD methods which unlike Fourier decomposition that takes the entire temporal content of the signal and converts it into only frequency content, instantaneously decomposes the speech signal into time varying frequency components in such a way that they oscillate at the fundamental frequency. This avoids the time-frequency resolution inversion problem that is common with Fourier analysis. The most basic unit of a speech signal can be taken as the vowel sound. Based on how the vowel sound is generated by the vocal folds, its corresponding waveform is a small increase or a decrease from the fundamental frequency value, coupled with higher resonances that are known as formants (Skoe and Kraus, 2010) [34]. Results from chirp simulations in Elberling et.al 2012 and Ronne et.al 2012 prove that a specific location on the BM is excited by a broader range of frequency components, which justifies running our analysis based on the fundamental frequency of the speech. Established studies have found that the brainstem responds at the fundamental frequency of a speech stimulus even when that frequency itself is removed from the acoustic signal (Galbraith and Doan 1995). The computational measure that was devised for the ABR keeps the rapid periodicity of GBC neuronal firing rates intact and results in an excellent periodic cross correlation with the fundamental waveform. A drawback of our computational model is that it doesn't include the response from the Inferior Colliculus (IC), which is believed to play an essential role in phase locking and producing the ABR [10] [34]. This is why the reported latency with the fundamental waveform is at 6.55ms which is slightly smaller than the experimental findings. However, the latency obtained with the complex correlation function is around the same value i.e. 9.5ms. This is a slight discrepancy and can be due to the fact that experimental data is based on a small sample or that the CN is primarily responsible for the ABR. Given current research, the latter is a slightly implausible conclusion as established efforts have been made to prove that the IC is an important player in the generation of ABR. The accuracy of the model is owed to the detailed GBC representations that are included in the Cochlear Nucleus python package. GBC's receive their main excitatory inputs from AN fibres through large synapses called modified endbulbs which have a unique set of dynamics that are exactly replicated in the package. Although the CN has other kinds of neurons that are believed to play a role in auditory processing i.e. Spherical bushy cells, Multipolar cells and Octopus cells, these have been ignored for the current model. A similar approach was taken in Verhulst et.al 2015 where only bushy cells were considered according to the argument that their time constant values and excitation and inhibition properties closely represent the collective electrical behavior of the Cochlear Nucleus [12] [17].

5 Future Work

The current model can be greatly optimized to include further cell populations in the CN, and also connect to an entirely new group of IC cells using the Brian package. IC modeling would require a

similar approach to that taken by the GBC modeling in the Cochlear Nucleus package, which means that synaptic dynamics and neuronal behavior needs to be modeled from first principles. However, the results presented here are highly encouraging and an incredible starting point for making more sophisticated computational representations of the auditory neural pathway. Further additions to this model could include an inclusion of noise, more speech signals and perhaps an integration with cortical feedback.

References

[1] Peter Søndergaard and Piotr Majdak. The auditory modeling toolbox. In Jens Blauert, editor, *The Technology of Binaural Listening*, pages 33–56. Springer, Berlin, Heidelberg, 2013.

[2] Filip Munch Rønne, Claus Elberling, James Harte, and Torsten Dau. *Modeling auditory evoked potentials to complex stimuli*. PhD thesis, 2013.

[3] Mitchell Steinschneider, Joseph C. Arezzo, and Herbert G. Vaughan. Tonotopic features of speech-evoked activity in primate auditory cortex. *Brain Research*, 519(1-2):158–168, jun 1990.

[4] Mitchell Steinschneider, Charles E. Schroeder, Joseph C. Arezzo, and Herbert G. Vaughan. Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(1):30–43, jan 1994.

[5] Guidance for Auditory Brainstem Response testing in babies. Version 2.1. Technical Report March, 2013.

[6] Yasheng Yuan, Fuxin Shi, Yanbo Yin, Mingjie Tong, Hainan Lang, Daniel B Polley, M Charles Liberman, and Albert SB Edge. Ouabain-induced cochlear nerve degeneration: synaptic loss and plasticity in a mouse model of auditory neuropathy. *Journal of the Association for Research in Otolaryngology*, 15(1):31–43, 2014.

[7] Robert V Harrison. An animal model of auditory neuropathy. *Ear and hearing*, 19(5):355–361, 1998.

[8] CS Reichenbach, C Braiman, ND Schiff, AJ Hudspeth, and T Reichenbach. The auditory-brainstem response to continuous, non-repetitive speech is modulated by the speech envelope and reflects speech processing. *FRONTIERS IN COMPUTATIONAL NEUROSCIENCE*, 10, 2016.

[9] Antonio Elia Forte, Octave Etard, and Tobias Reichenbach. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. 2017.

[10] Bharath Chandrasekaran and Nina Kraus. The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, 47(2):236–246, 2010.

[11] Torsten Dau. The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *J Acoust Soc Am*, 113(2):936–950, 2003.

[12] Sarah Verhulst, Hari M Bharadwaj, Golbarg Mehraei, Christopher A Shera, and Barbara G. Shinn-Cunningham. Functional modeling of the human auditory brainstem response to broadband stimulations. *The Journal of the Acoustical Society of America*, 138(3):1637–1659, 2015.

[13] J R Melcher and N Y Kiang. Generators of the brainstem auditory evoked potential in cat. III: Identified cell populations. *Hearing research*, 93(1-2):52–71, apr 1996.

[14] B C Fullerton, R A Levine, H L Hosford-Dunn, and N Y Kiang. Comparison of cat and human brainstem auditory evoked potentials. *Electroencephalography and clinical neurophysiology*, 66(6):547–70, jun 1987.

[15] Richard Vogel. ABR (Auditory Brainstem Response), Neuromonitoring.

[16] Nicole Russo, Trent Nicol, Gabriella Musacchia, and Nina Kraus. Brainstem responses to speech syllables. *Clinical Neurophysiology*, 115(9):2021 – 2030, 2004.

[17] Paul C Nelson and Laurel H Carney. A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America*, 116(4):2173–2186, 2004.

[18] Hung Thai-Van, Sebastian Cozma, Florent Boutitie, François Disant, Eric Truy, and Lionel Collet. The pattern of auditory brainstem response wave v maturation in cochlear-implanted children. *Clinical Neurophysiology*, 118(3):676 – 689, 2007.

[19] Charles A. Miller, Kathleen E. Woodruff, and Bryan E. Pfingst. Functional responses from guinea pigs with cochlear implants. i. electrophysiological and psychophysical measures. *Hearing Research*, 92(1):85 – 99, 1995.

[20] Amin Saremi and Stefan Steinfeld. Effect of metabolic presbyacusis on cochlear responses : A simulation approach using a physiologically-based model. *J Acoust Soc Am*, 134(4), 2013.

[21] Gregory Hickok. The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of communication disorders*, 45(6):393–402, 2012.

[22] Emily B J Coffey, Sibylle C Herholz, Alexander M P Chepesiuk, Sylvain Baillet, and Robert J Zatorre. frequency-following response revealed by MEG. 2016.

[23] Xuedong Zhang, Michael G Heinz, Ian C Bruce, and Laurel H Carney. A phenomenological model for the responses of auditory-nerve fibers: I. nonlinear tuning with compression and suppression. *The Journal of the Acoustical Society of America*, 109(2):648–670, 2001.

[24] Mario A Ruggero, Nola C Rich, Alberto Recio, S Shyamla Narayan, and Luis Robles. Basilar-membrane responses to tones at the base of the chinchilla cochlea. *The Journal of the Acoustical Society of America*, 101(4):2151–2163, 1997.

[25] C Ian and S A Muhammad. 19 th INTERNATIONAL CONGRESS ON ACOUSTICS MADRID , 2-7 SEPTEMBER 2007 COMPUTATIONAL MODELLING OF THE CAT AUDITORY PERIPHERY : RECENT DEVELOPMENTS AND FUTURE DIRECTIONS. (September):2–7, 2007.

[26] Larry A. Westerman and Robert L. Smith. A diffusion model of the transient response of the cochlear inner hair cell synapse. *The Journal of the Acoustical Society of America*, 83(6):2266–2276, 1988.

[27] Brian R Glasberg and Brian CJ Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1):103–138, 1990.

[28] James M Harte, Filip Rønne, and Torsten Dau. Modeling human auditory evoked brainstem responses based on nonlinear cochlear processing. (August), 2010.

[29] Claus Elberling and Manuel Don. A direct approach for the design of chirp stimuli used for the recording of auditory brainstem responses. *The Journal of the Acoustical Society of America*, 128(5):2955–2964, 2010.

[30] N. Y. S. Goldstein, M. H., & Kiang. Synchrony of Neural Activity in Electric Responses Evoked by Transient Acoustic Stimuli. *Jasa*, 30(2):107–114, 1958.

[31] Amin Saremi, Rainer Beutelmann, Mathias Dietz, Go Ashida, Jutta Kretzberg, and Sarah Verhulst. A comparative study of seven human cochlear filter models. *Jasa*, 140(3):1618–1634, 2016.

[32] Marek Rudnicki and Werner Hemmert. High entrainment constrains synaptic depression levels of an in vivo globular bushy cell model. *Frontiers in Computational Neuroscience*, 11:16, 2017.

[33] Muhammad S A Zilany, Ian C Bruce, Paul C Nelson, and Laurel H Carney. nerve : Long-term adaptation with power-law dynamics A phenomenological model of the synapse between the inner hair cell and auditory nerve : Long-term adaptation with power-law dynamics. 2390(2009), 2014.

[34] Erika Skoe and Nina Kraus. Auditory brainstem response to complex sounds: a tutorial. *Ear and hearing*, 31(3):302, 2010.

Appendices

.1 Details on experimental ABR recordings

- Speech signals were taken from audibooks on Greek mythology from: https://librivox.org
- Headphones used: ER-3C, Etymotic, U.S.A
- Conducting gel was sourced from Abralyt HiCl, Easycap, Germany and reduced the impedance between the skin and the electrodes to 5kiloohms.
- Electrodes were composed of Ag/AgCl and sourced from Multitrode, BrainProducts, Germany and were connected to a series of bipolar amplifiers
- The audio signals were measured by theintegrated amplifier as well through an acoustic adapter, from Acoustical Stimulator Adapter and StimTrak, BrainProducts, Germany

.2 Code

.2.1 Final simulation script written in Python

```
from __future__ import division, absolute_import, print_function
#   Importing packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

import brian
from brian import siemens,ms

import cochlea
import cochlear_nucleus.brn as cn
import thorns as th
import thorns.waves as wv
import scipy.io as sio


def main():


    Fs = float(100e3);
    #extracting speech signal and pre-processing
    dictsample = sio.loadmat('resampled.mat')
    sample = dictsample['newstory'];
    sample  = sample.flatten()
    duration = float(0.2);
    index = int(duration*Fs) ;
    mysample = sample[0:index];
    wv.set_dbspl (mysample,78) #setting the level to 50 dB SPL

    brian.defaultclock.dt = 0.01*ms;

    # Generate ANF trains
    anf_trains = cochlea.run_zilany2014(
        sound=mysample,
        fs=Fs,
        anf_num=(13,3,1),    # (HSR, MSR, LSR)
        cf=(125,8000,30),
        species='human',
        seed=0,
        powerlaw= 'approximate'
    )

    # Generate ANF and GBC groups
    anfs = cn.make_anf_group(anf_trains)
```

```
gbcs = cn.make_gbc_group(200)

# Connect ANFs and GBCs
synapses = brian.Connection(
    anfs,
    gbcs,
    'ge_syn',
    delay = 5*ms
)

convergence = 20

weight = cn.synaptic_weight(
    pre='anf',
    post='gbc',
    convergence=convergence
)

synapses.connect_random(
    anfs,
    gbcs,
    p=convergence/len(anfs),
    fixed=True,
    weight=weight,
)

# Run the simulation
cn.run(
    duration=duration,
    objects=[anfs, gbcs, synapses,gbc_spikes]
)
gbc_trains = th.make_trains(gbc_spikes)

CNspikes = gbc_trains['spikes'];
ANspikes = anf_trains['spikes'];

CN_spikes = np.asarray(CNspikes)
AN_spikes = np.asarray(ANspikes)
record = {'CN':CN_spikes,'AN':AN_spikes}
sio.savemat('SpeechANCN',record)

fig, ax = plt.subplots(2, 1)

th.plot_raster(anf_trains, ax=ax[0])
th.plot_raster(gbc_trains, ax=ax[1])
plt.show()
plt.tight_layout()
```

```
    if __name__ == "__main__":
        main()
```

.2.2 Function in MATLAB which converts Spike Train data into Binary or Raster data

```
function [Overallspikes,nspikes] = Raster(SpikeCell,T,StimFs)

for m = 1 : length (SpikeCell)
    a(m) = length(SpikeCell{m});
end

maxlength = max(a);

for mm = 1 : length (SpikeCell)
   SpikeData(mm,:) = [SpikeCell{mm},zeros(1,maxlength-length(SpikeCell{mm}))];
end

res = 1/StimFs;
npts = length(T);
indices = single(SpikeData./res);
indices = round(indices) ;
size_spikes = size(SpikeData);
i = size_spikes(1);
j = size_spikes(2);
BinarySpikes = zeros(i,npts);
nspikes = 0;

for s = 1 : i
    for p = 1 : j
        if indices(s,p) ~= 0
            BinarySpikes(s,indices(s,p)) = 1;
            nspikes = nspikes + 1 ;
        end
    end
end

Overallspikes = sum(BinarySpikes);

end
```

.2.3 Firing rate estimation function written in MATLAB, for various kernel functions

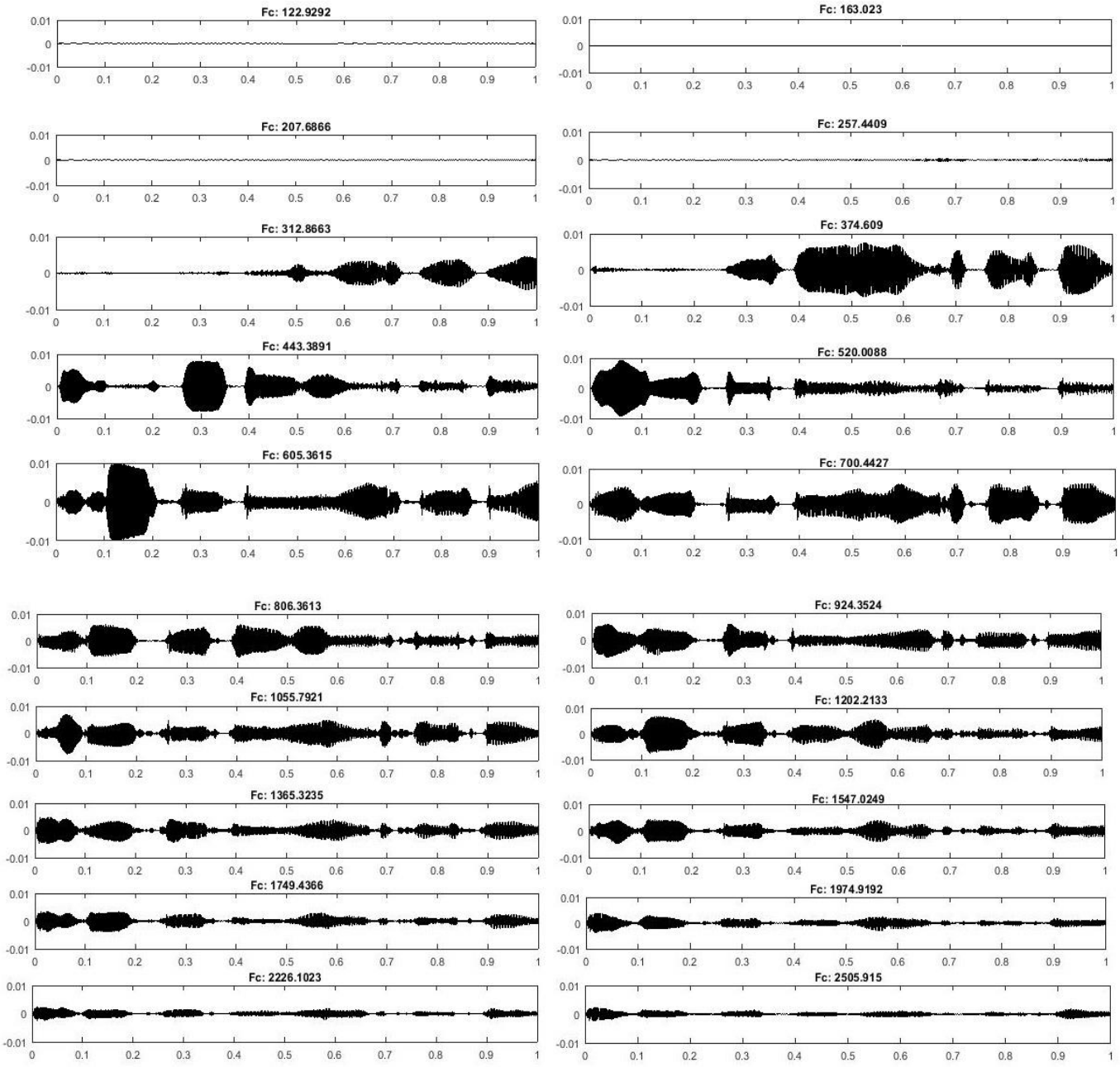
```
function l = fr_es(A, dt,kernel)

if(length(A)==1)
    l = A;
end

if (~kernel)
index = 1;
for j = (dt+1):(length(A))
l(index) = sum(A((j-dt):j))./(dt);
index = index +1;
end

else

l= zeros(1,length(A));

indices = find(A==1);

for i = 1:length(indices)
if (strcmp(kernel,'Boxcar'))
    if (ceil(indices(i) + sqrt(3)*dt) < length(A)+1 && ceil(indices(i) - sqrt(3)*dt) > 0)
        range  = indices(i)-sqrt(3)*dt:indices(i)+sqrt(3)*dt;
    elseif(ceil(indices(i) + sqrt(3)*dt) > length(A) && ceil(indices(i) - sqrt(3)*dt) < 1)
        range = 1:length(A);
    elseif(ceil(indices(i) + sqrt(3)*dt) > length(A) && ceil(indices(i) - sqrt(3)*dt) > 0)
        range  = indices(i)-sqrt(3)*dt:length(A);
    else
        range  = 1:indices(i)+sqrt(3)*dt;
    end

    range = ceil(range);

else
    if (indices(i) + 5*dt < length(A)+1 && indices(i) - 5*dt > 0)
        range  = indices(i)-5*dt:indices(i)+5*dt;
    elseif(indices(i) + 5*dt > length(A) && indices(i) - 5*dt < 1)
        range = 1:length(A);
    elseif(indices(i) + 5*dt > length(A) && indices(i) - 5*dt > 0)
        range  = indices(i)-5*dt:length(A);
    else
        range  = 1:indices(i)+5*dt;
    end
end
```

```
if (strcmp(kernel,'Gaussian'))
l(range) = l(range) + gauss(range,indices(i),dt);
elseif (strcmp(kernel,'Exp'))
l(range) = l(range) + exponent(range,indices(i),dt);
elseif (strcmp(kernel,'Boxcar'))
l(range) = l(range) + box(dt);
end
end

end
end

function f = gauss(x, mu, s)
mu2 = ones(1,length(x)).*mu;
p1 = -.5 * ((x - mu2)/s) .^ 2;
p2 = (s * sqrt(2*pi));
f = exp(p1) ./ p2;

end

function f = exponent(x, mu, s)
mu2 = ones(1,length(x)).*mu;
p1 = -sqrt(2) * (norm((x - mu2)/s)) ;
p2 = (s * sqrt(2));
f = exp(p1) ./ p2;

end

function f = box(s)

f = 1/(2*sqrt(3)*s);

end
```

.3 DRNL plots

The following page contains the result from the DRNL filterbank that was tested on 1s of the speech sample. The axes have been regularised across the plots therefore some amplitudes appear smaller than others however the relative amplitudes are an accurate depiction of how frequencies are decomposed in the cochlea. The axis labels have been removed for clarity but the X axis is Time in seconds and the Y axis is Amplitude in Pa. The specific CF values are written above each plot.

P.T.O.


Fc: 122.9292
Fc: 163.023
Fc: 207.6866
Fc: 257.4409
Fc: 312.8663
Fc: 374.609
Fc: 443.3891
Fc: 520.0088
Fc: 605.3615
Fc: 700.4427
Fc: 806.3613
Fc: 924.3524
Fc: 1055.7921
Fc: 1202.2133
Fc: 1365.3235
Fc: 1547.0249
Fc: 1749.4366
Fc: 1974.9192
Fc: 2226.1023
Fc: 2505.915
Fc: 3164.8545
Fc: 4462.5801
Fc: 3982.5659
Fc: 6256.5479
Fc: 5592.9795
Fc: 7819.2051
Fc: 6995.7495