**Cochlea: anatomy, non-linearity and computational modelling**

**Anatomy:**

The anatomy of the ear can be separated into three parts, the outer middle and inner ear. The inner ear contains the cochlea, a snail shaped, spiral structure. The cochlea’s base is narrower than its apex with higher frequency sounds being mostly represented at the base and low frequencies at the apex. Within a cross section of the entire cochlea, there are three sections. The scala vestibuli at the top, scala media in the middle and scala tympani at the top. The scala vestibuli and tympani are connected at the apex. The scala media and tympani are separated by the basilar membrane. Vibrations travel from the scala vestibuli around the apex to the scala tympani and reach the sound window. The basilar membrane supports hair cells which transduce vibrations into action potentials. This occurs in the organ of Corti which contains approximately three times more outer hair cells compared to inner hair cells. As vibrations travel around the top and bottom scala, the basilar membrane is excited in waves, which cause the hair cells to move back and forth. The hair cells also comprise smaller hairs called cilia which are linked to the tectorial membrane, above the organ of Corti. The cilia are connected to each other through tip links. The movement of the cilia results in opening of ion channels, causing the release of neurotransmitters (Heller, 2012; Newton & Vallely, 2006).

**Band-pass filters:**

Regarding the way in which tones and pitch are represented in the cochlea, a place to frequency, or tonotopic representation is often used. As mentioned above, the base of the cochlea selectively responds to higher frequencies and the apex to lower frequencies. Specifically, the basilar membrane is organised in band-pass filters. These filters represent different areas along the basilar membrane which selectively respond to a specified range of frequencies. These band-pass filters are organised in an overlapping manner and each filters response to a signal is the highest when the signal frequency corresponds to the filter’s central frequency. Signal frequencies which correspond to the outer parts of the filter respond less. The bandwidths of the filters increase as frequency increases near the base, meaning overlap also increases. During the presentation of a complex tone, this results in the excitation of only filters with centre frequencies equivalent to the lower harmonics, due to the limited overlap. The increasing bandwidth of the filters in higher frequencies, means that higher harmonics excite multiple filters with centre frequencies close to one another. The lower and higher harmonics are resolved and unresolved respectively due to the distinctness of the firing responses afforded by the bandwidths of the filters (Oxenham, 2008).

Another often used theory of representation is temporal. This means that pitch determines the timing of action potentials due to phase locking. Phase locking refers to the periodic firing of action potentials at certain phases of the sound wave. Therefore, two action potentials are separated by a multiple of the sound wave period. According to temporal explanations of complex tones, unresolved harmonics create temporal waveforms while interacting with the band-pass filters. Timing, interval and periodicity information can be used to identify the pitch (Oxenham, 2008).

Finally, a combination of the two theories can be used to explain pitch representation. In this case, temporal information about firing intervals determines pitch provided that the firing also corresponds to the correct place on the basilar membrane according to the tonotopic representation. Auditory nerve fibres along the basilar membrane respond at different phases, depending on frequency. The difference in firing phase determines the pitch of a sound. Crucially all three theories have strengths and downfalls, but there is no clear evidence supporting one prevailing theory. Although some have suggested that information from lower number harmonics is more important than temporal envelope information from unresolved harmonics, there is still some information to be extracted from that (Oxenham, 2008).

**Non-linearity:**

The traditional view of the cochlea proposed by Helmholtz, conceptualises it as a linear Fourier analyser. Were this to be true, band-pass filters would be organised linearly, separating the components of a sound as identified by the Fourier transform. However, as established before this is not the case and non-linearity is present in the transduction of mechanical energy (Chertoff et al., 2000; Nobili et al., 1998). Therefore, the cochlea is better modelled as non-linear. This is highlighted through a variety of observable phenomena. For example, the displacement of the basilar membrane in response to a certain frequency is not a one-to-one ratio throughout the membrane and does not increase in amplitude linearly. The phenomenon is called compressive non-linearity and is due to energy dependent responding. Another example is otoacoustic emissions, which are echoes of inputted sounds, produced by the cochlea, that are related but not the same as the stimulus. They are sounds with a band-pass central frequency related to the location where they were generated in the cochlea (Nobili et al., 1998; Withnell & Yates, 1998; Allen, 2008). Furthermore, voltage-dependent motility can be seen in the outer hair cells, meaning they can change length in response to electrical charge. These phenomena can all be traced back to the outer hair cells and the frequency selectivity that characterises them. The non-linear organisation of band-pass filters results in equalisation of response and two-tone suppression. The former equalises significantly different sound frequency components, irrespective of their intensity. While the latter, refers to suppression of one tone by the adjacent tone (conceptually similar to lateral inhibition of neurons), due to masking (Allen, 2008; Nobili et al., 1998). Masking occurs when one stimulus’ reception is obstructed by the presence of another stimulus (Durlach, 2006).

Computational modelling is necessary to relate the above outer hair cell dependent phenomena (Allen, 2008).

Non-linearity of the cochlea can be observed in temporal masking, and harmonic and inter-modulation distortion. These are due to basilar membrane vibration, otoacoustic emissions, and cochlear microphonics. Cochlear microphonics are due to basilar membrane and cilia displacement (Chertoff et al., 2000).

Additionally, travelling waves would follow the combinations of linear harmonic oscillations as proposed by Helmholtz (Nobili et al., 1998).

**Gammatone filter:**

Reverse correlation refers to correlating auditory fibre responses to white noise to the same white noise, to measure the shape of the auditory nerve. The gammatone filter approximates these measured revered correlations and extracts amplitude and phase information.

Reverse correlation is used to correlate the auditory nerve fibre responses to white noise to the white noise. This is done to measure the auditory filter shape. The gammatone filter approximates these measured reverse correlations and produces/represents amplitude and phase information. The distinction between reverse correlation and the gammatone filter is that the output of the reverse correlation is an estimate of the impulse response of the fibre, while the gammatone function is an equation which analytically expresses the reverse correlation. The gammatone filter equation consists of both the gamma function from statistics and the cosine term expressing a tone. Therefore, the impulse response of a filter is like a burst at the centre frequency of the filter which is enclosed in a gamma-shaped envelope. The amplitude characteristic of the gammatone function is similar to the roex filter which can predict masking (Darling, 1991; Patterson &Nimmo-Smith, 1987).

Each envelope of the impulse responses is symmetric, while those of the auditory filter are symmetric with differences in onset and offset. This means that differential masking of short signals adjacent to large impulses could present inconsistencies between psychophysical observations and the gammatone filters (Patterson &Nimmo-Smith, 1987).

**Computational modelling:**

Note to self: filter gain is the ratio of the output voltage over the input voltage. If you have a filter that’s made to handle frequencies below 100 Hz then when inputted with a frequency of 100 Hz or lower, you want the gain to be close to one. That means that the waveform leaves the filter same as it went in (unity or 100%). But then if a waveform with frequency above 100 Hz is inputted the gain should be close to zero, meaning it enters the filter but doesn’t make it through as output.

Computational models of the cochlea and its frequency selectivity can be used in many ways, for example to estimate excitation patterns, or create speech recognition algorithms. Despite the non-linear mechanics involved in the cochlea, traditionally, models have used linear gammatone filters (Patterson et al., 1992). The non-linear organisation of the basilar membrane has been psychophysically shown through changes in the filters’ central frequency and their bandwidths. The non-linear model suggested in this paper is similar to the gammatone filterbank model, however, each filter is replaced by a dual resonance nonlinear filter unit. The model is made up of two parts, one that transforms a sound wave delivered to the outer ear into a velocity waveform of the stapes. The second part is the dual resonance nonlinear filter that simulates the velocity of the vibrations of the basilar membrane. The dual resonance non-linear filter models the velocity of the vibration of the basial membrane at a specific area in response to the velocity waveform transferred from the stapes. This waveform input follows two paths, whose summation is the output of the dual resonance non-linear filter. A linear one is applied and the input is filtered through a series of first-order gammatone filters and a series of second-order low-pass filers. The second path is non-linear where the input is filtered through a series of first order gammatone filters, followed by a non-linear gain and another series of gammatone filters. The effect of each path differs depending on the frequency of the inputted signal since the non-linear function is compressive, therefore its dominance over the linear path changes. For very low and high signals the dual resonance non-linear filter functions linearly, since the non-linear function is linear for very low frequencies and at high frequencies the linear function dominates the output. This follows Plack and Oxenhams (2000) findings that show linearity, followed by compressive non-linearity and by linearity again, as signal frequencies rise (Poveda & Meddis, 2001).

In the past filterbank models fell under two categories, gammatone filters and rounded-exponential filters. However, considering the nonlinear properties of the cochlea, nonlinear filterbank models have been developed. Another type of modelling includes biophysical details to ensure both cochlear processing and its underlying biological processes are accounted for. Therefore, this type differs from others which only model the perceptual properties of input and output in the cochlea. There are parallel, cascaded, transmission-line filter models among others. Models such as the dual resonance non-linear filter differ from others like the gammatone in that they attempt to model the mechanical properties and vibrations of the cochlea and basilar membrane. While the gammatone model solely reproduces the neural and perceptual data without incorporating elements of mechanics. The two approaches to modelling yield different results, with perceptual models making assumptions about the dominance of specific aspects like frequency selectivity which may not be correct. Mechanical models can also employ psychoacoustic data to accomplish parameter tuning. Another opposing pair of model types are parallel and cascaded models. The former employs separate, independent filters, inputted with the same signal, which ensures simplification despite the inaccuracy. However, phase dependant wave propagation is successfully simulated even in parallel filter models such as the gammatone and dual resonance non-linear model despite lacking the cascading architecture theoretically necessary for such modelling. Although each computational model has advantages, the cascade asymmetric resonators with fast acting suppression model had high agreement with experimental data while at the same time retaining computational power (aka low computational loss) (Saremi et al., 2016).

The cascade of asymmetric resonators with fast acting suppression model is based on the pole-zero filter cascade models (Lyon et al., 2010). It consists of second order filters and form a mechanical point of view, represents second-order asymmetric resonators whose resonance is determined by the centra frequency and damping ratio, which are controlled by poles and zeros (Saremi et al., 2016).

Filters have traditionally been described as power-frequency response (roex) or impulse response (gammatone). Filter description and specification can be accomplished using Laplace domain poles (roots of the denominator) and zeros (roots for the numerator) from electrical engineering. Filters in cascade functions cannot be easily described using impulse and frequency responses like the previous filter models but can be better described using poles and zeros. Linear filters can be made quasi-linear or signal-dependent by using input or output level to define parameters. Each equivalent rectangular bandwidth corresponds to about 0.89 mm on the basilar membrane, meaning 39 channels in 35 mm without overlap or 78 channels with 50% overlap. An asymmetric resonator is created by having a complex-conjugate pair of poles and one of zeros, which are positioned in a way to enable a peak in gain near the pole frequency and a gain drop at higher frequencies. A dynamic level-dependent positioning of the poles and an instantaneous cubic distortion as the output, ensure nonlinearity is introduced in the model. The fast-acting automatic gain control and the instantaneous part from an odd-order nonlinearity make up the compression of the model. The cascade of asymmetric resonators with fast acting suppression is based on the pole-zero filter cascade model coupled with an automatic gain control loop. It approximates the instantaneous firing rate of the auditory nerve as a function of place, meaning both filtering and compression are simulated (Lyon, 2011).

References

Chertoff, M. E., Lerner, D., Amani-Taleshi, D., & Nagai Y. (2000). Characterizing non-linearity in the cochlear microphonic using the instantaneous frequency. *Hearing Research, 145*, 190-202.

Nobili, R., Mammano, F., & Ashmore, J. (1998). How well do we understand the cochlea? *Trends in Neurosciences, 21*, 159-167.

Allen, J. B. (2008). *Nonlinear Cochlear Signal Processing and Masking in Speech Perception*. In: Benesty, J., Sondhi, M.M., Huang, Y.A. (eds) Springer Handbook of Speech Processing. Springer Handbooks. Springer, Berlin, Heidelberg.

Withnell, R. H., & Yates, G. K. (1998). Onset of basilar membrane non-linearity reflected in cubic distortion tone input-output functions. *Hearing Research, 123*, 87-96.

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification, 12*(4), 316-331.

Heller, E. J. (2012). *Why you hear what you hear: An experiential approach to sound music, and psychoacoustics.* Princeton University Press.

Durlach, N. (2005). Auditory making: Need for improved conceptual structure. *The journal of the Acoustical Society of America, 120,* 1787.

Darling, A. M. (1991). Properties and implementation of the Gammatone filter: A tutorial. *Speech Hearing and Language, Work in Progress, University College London, Department of Phonetics and Linguistics*, 34-61.

Patterson, R., & Nimmo-Smith, I. (1987, December). An efficient auditory filterbank based on the gammatone function. In *a meeting of the IOC Speech Group on Auditory Modelling at RSRE, 2*(7).

Lopez-Poveda, E. A., & Meddis, R. (2001). A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America, 110*(3107).

Saremi A., Beutelmann, R., Dietz, M., et al. (2016). A comparative study of seven human cochlear filter models. *The Journal of the Acoustical Society of America,140*(1618).

Plack, C. J., & Oxenham, A. J. (2000). Basilar-membrane nonlinearity estimated by pulsation threshold. Journal of the Acoustical Society of America, 107, 501-507.

Patterson, R., Robinson, K., Holdsworth, J., McKeown, Zhang, C., & Allerhand, M. (1992). *Complex sounds and auditory imaging*. In Auditory Physiology and Perception. Y. Cazals, K. Horner, & L. Demany. Pergamon, Oxford.

Lyon, R. F. (2011). Cascades of two-pole-two-zero asymmetric resonators are good models of peripheral auditory function. *The Journal of the Acoustical Society of America, 130*(6).

Lyon, R. F., Katsiamis, A. G., & Drakakis, E. M. (2010). History and future of sudatory filter model. *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, Paris, France, pp. 3809-3812.

Newton, V. E., & Vallely, P. J. (2006). *Infection and hearing impairment*. John Wiley & Sons.