A Brief History of Auditory Models

Leonardo C. Araújo[1] **, Tairone N. Magalhaes**[1] **, Damares P. M. Souza**[1] **,
Hani C. Yehia**[1] **, Maurício A. Loureiro**[1]

[1]CEFALA - Center for Research on Speech, Acoustics, Language and Music
Universidade Federal de Minas Gerais (UFMG)
CPDEE-Centro de Pesquisa e Desenvolvimento em Engenharia Elétrica, room 214
Av. Antônio Carlos 6627 – 31270-010 Belo Horizonte, MG

`leoca, tairone, damares, hani, mauricio @cefala.org`

Abstract. *This work presents a brief description of the human auditory system together with the history of human comprehension of the auditory function, its main features, and classic models used to represent it. First, a historical view of the hearing apparatus is presented. After that, the physiology of the peripheral auditory system is described. The process of acoustic propagation through the outer, middle and inner ear, as well as the mechanism of transformation of cochlea inner hair cell motion into neuron spikes are explained. Next, Flanagan's mathematical representation (based on physiological data acquired by von Békésy) of the passive relation between the sound that reaches the outer ear and the motion of the cochlea basilar membrane. Flanagan's model is followed by Lyon's model of the cochlea, Meddis' model of the inner hair cell, and Patterson's Auditory Image Model. Finally, the IPEM Toolbox is introduced as an example of music analysis system that incorporates an auditory model to perform acoustic analysis of sound based on human perception.*

1. Introduction

Over the past half century, the auditory system has been the focus of intensive research. Knowledge on physiology, psychology and engineering has provided the possibility for creation of models which until a certain extension tries to describe and mimic the hearing mechanisms. Mathematical and computational models are used to build quantitative simulations that describes a system and bring insights about the system itself or its behavior under certain conditions. Auditory models might be used to several purposes, such as help on tracking problems on hearing or speech fields. They can be also very useful on music research since these systems can simulate the mechanisms of sound perception, so it is possible to use it in musical timbre research, as has been done by a few [Cosi et al., 1994], [Toiviainen et al., 1995].

2. Hearing Physiology

Basically, the peripheral auditory system has three main parts: the outer, middle and inner ear. The outer ear is the visible portion of the ear. It includes the pinna (also called auricle), the ear canal and the eardrum. The pinna collects the sounds and directs them down into the ear canal. Its shape also help us to extract directional information from sounds. The ear canal is a tube about 2.5 cm long, and it acts as a quarter-wavelength resonator. So it enhances frequencies around 3400 Hz, the maximum sensitivity regions of human hearing. This region can be easily observed at the famous equal loudness curves.

Acoustic sound waves impinging on the outer ear is propagated down the external meatus leading to the eardrum which is set into vibration. These vibrations are then transmitted to the tiny ossicles of the middle ear. The ossicle chain is compounded of three units: the incus, the malleus and the stapes. The malleus, or hammer, is fixed on the eardrum. It makes contact with the incus, or anvil, which in turn connects with stapes, or stirrup, through a small joint.

The main function of the ossicles is to make an impedance transformation from the air medium of the outer ear to the liquid medium inside the cochlea (inner ear). As there is a great change in the propagation medium density, and so in its impedance, it is necessary an impedance transformation, trying to match those two far distant impedances, comprising their respective disparate realms. The lever action of the ossicles alone provide a force amplification of about 1.3 [von Békésy, 1960]. Another amplification arises from the change in effective area, the eardrum area is much greater than that of the stirrup on the order of 15:1, according to Békésy's measurements [von Békésy, 1960].
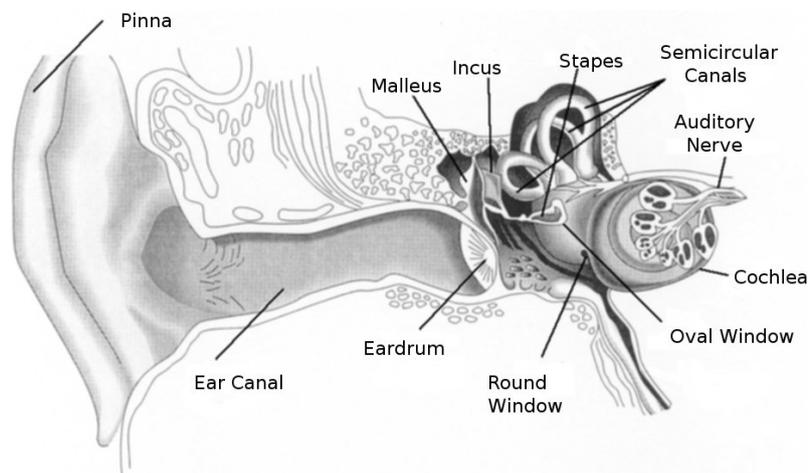

Figure 1: Outer, middle and inner ear

The inner ear is the part where the cochlea is located, just behind the oval window, which is connected to the stapes footplate. The inner ear also comprises the vestibular apparatus and the auditory nerve terminations. The cochlea is a snail-shaped organ filled with almost incompressible fluids. It is divided by two membranes: Reissner's membrane and the basilar membrane.

When the oval window is set into movement, a pressure difference on the fluid propagates down the cochlea. Thus inward motion of the ossicles results in an outward movement in the round window. The basilar membrane (Figure 2) is set in motion by this pressure difference, but the pattern of its movement varies, depending of its mechanical properties at a certain position. At its basal end, the basilar membrane is narrow and stiff, while at its apical end it is wider and more compliant. Because of these differences, the position of the peak in the pattern of vibration differs according to the frequency of stimulation. The basis responds best to high frequencies, while the apex responds best to low frequencies.

Mechanical motion of the basilar membrane leads to displacements of the inner hair cells stereocilia, which are located between the basilar membrane and the tectorial membrane, in a structure called organ of Corti.

3. Auditory Models

The crucial problem in modelling is deciding what is the optimal level of details and what are the possible simplifications. The amount of details and the computational (and why not, mathematical)
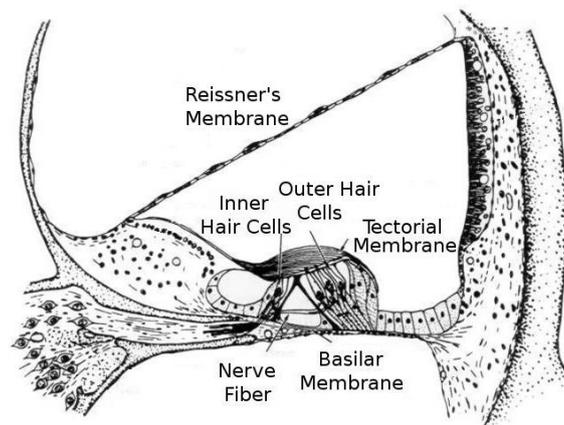
Figure 2: Section of the cochlea. (Adapted from Moore,1995).

burden go against each other. Increasing the level of details, leads us to complex and computational expensive models. Furthermore, all the effort put on those details modelling might be wasted if the final user of the models does not care about them.

This kind of representation subdue the dynamic characteristics and the nonlinearity of the signal. It also incorporates other phenomena from the perception mechanism such as critical bands, frequency masking and temporal masking.

Auditory models are being continually upgraded, and there are many different models available on the internet. The principal models will be described here.

3.1. Flanagan's Model

Flanagan, based on the physiological data measured by Békésy, has proposed a mathematical and also a computational model for the auditory mechanism [Flanagan, 1960] [Flanagan, 1962] [Flanagan, 1972]. His model is divided into two parts, one comprising the middle ear and the other the basilar membrane.

The fist block implements the physiological function embraced by the middle ear and the second the basilar membrane. The input in the first block, $p(t)$, is the pressure at the eardrum, its output, $x(t)$, is stapes displacement which will be the input to the next stage. The final output, $y_l(t)$, is the basilar membrane displacement at a distance l from the stapes.

The approximating functions are indicated in its frequency-domain (Laplace) representation by the functions $G(s)$ and $F_l(s)$. Those functions must be fitted to available physiological data. Flanagan assumes the ear to be mechanically passive and linear over the frequency and amplitude ranges of interest, than, rational functions of frequency with left half-plane poles (stable) can be used to approximate the physiological data. The fact of being modeled by rational functions make it feasible of lumped-constant electrical circuit implementation.

Flanagan uses as an approximation to $G(s)$ a third degree function, composed of a real pole and a pair of complex conjugated poles. $F_l(s)$ is approximated by a second order pair of complex conjugated poles, a real pole, a real zero and a delay factor.

This transfer-function is created this way to resemble the resonant properties of the membrane, approximating it as a constant-Q (constant percentage bandwidth) filter bank in character.

3.2. Lyon's Model

Richard F. Lyon [Lyon and Mead, 1988] developed a model of analog electronic cochlea based on the knowledge of how the cochlea works. His approach was to model the fluid-dynamic wave
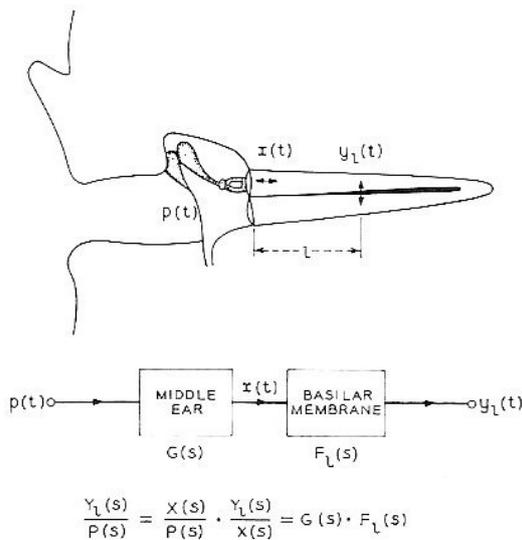
Figure 3

$$G(s) = \frac{1}{(s+a)[(s+a)^2 + b^2]}$$
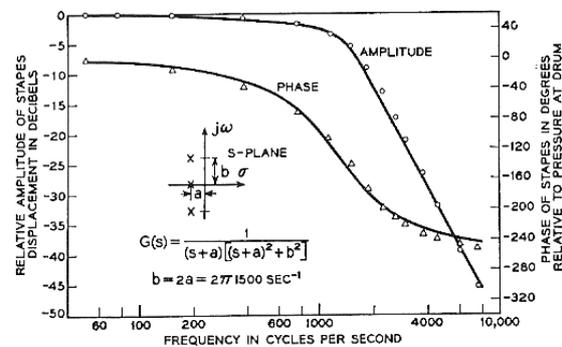
$$b = 2a = 2\pi 1500 \text{ SEC}^{-1}$$

Figure 4

medium of the cochlea by a cascade of filters based on the observed properties of the medium. The action of the active outer hair cells was modelled by a set of automatic gain controls (AGC) which simulated the dynamic compression of the intensity range on the basilar membrane.

A neural spike at the base of auditory nerve is only generated when the stereocilia of the inner hair cell are bent one way. When the cilia are bent the other way, no spikes are generated. So the inner hair cells acts like half-wave-rectifiers, and this unit was used to model them in the Lyon's cochlear model. The output of the model is the probability of firing along time for the neurones of the auditory nerve. This data received the name cochleagram.

Malcolm Slaney implemented an auditory model based on Lyon's model. He developed a toolbox for MatLab called auditory toolbox which includes several functions for auditory modelling including different models for cochlear processing, hair cell transduction and additional functions for spectral analysis and correlogram generation. The correlogram a is function that summarizes periodical information of the cochleagram to give us a better representation of an auditory image. It can be visualized as a movie, and there is also a function for their visualization in the auditory toolbox.

3.3. Meddis' Inner Hair Cell

Ray Meddis [Meddis, 1986] developed a model of inner hair cell based on its physiology. In this model, the permeability function controls the neurotransmitter release into the synaptic cleft. The spike probability on specific neurones of the auditory nerve is a function of the amount of neurotransmitter in the cleft. Interspike intervals are also taken into account. Spikes cannot occur in intervals less than 1 ms.

His inner hair cell model became popular between the auditory researchers. The auditory image model (AIM) (see section 3.4), a famous model of auditory processing developed by Roy Patterson et al. [Patterson et al., 1995] includes a stage of neural transduction using the Meddis hair cell.

This model was upgraded by Sumner et al [Sumner et al., 2002], creating a revised model of the inner hair cell and auditory-nerve complex. This model improves the previous in terms of the range of phenomenon simulated and is more consistent with recent developments in hair-cell physiology. It represents better the response of medium and low spontaneous rate fibers. The
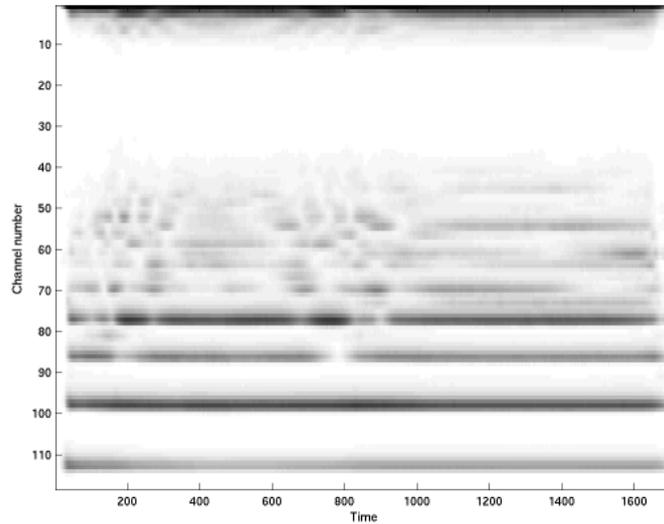
Figure 5: Example of cochleagram using a recorded clarinet sound (C4) as input to the model.

former only simulate the response of high spontaneous rate fibers at the fiber's best frequency.

3.4. Auditory Image Model (AIM)

The auditory image model includes several alternative modules comprising the stages of processing by the peripheral auditory system. These stages are (1) middle ear filtering, (2) spectral analysis, (3) neural encoding and (4) time-interval stabilization. The middle ear filtering is a simple linear filter that enhances middle frequencies. The spectral analysis can be done with a functional (gammatone [1]) or a physiological auditory (nonlinear transmission) filter. The output of this stage is the estimated basilar membrane motion (BMM) obtained in the presence of the input signal.

The neural encoding stage converts the BMM into a neural activity pattern (NAP). Two modules are provided for generating the NAP: a bank of meddis inner hair cells and a bank of two-dimensional adaptive threshold units, which rectify and compress the BMM, then apply adaptation in time and suppression across frequency.

The last stage summarizes the temporal activity at the output of the NAP stage based on the idea that periodic sounds give rise to static perceptions by the human listener. There are two modules included: an strobed temporal integration (STI) and a correlogram.

3.5. IPEM Toolbox

The IPEM Toolbox is an example of music analisys system which incorporate an auditory model to peform the perception-based acoustic analysis of the sound. The auditory model (Auditory Peripheral Module) was developed by Van Immerseel and Martens (1992) and implement some features that are important for music perception.

There is an outer ear filtering, an array of band-pass filters for simulate the cochlea and a hair cell model (HCM) that incorporates half wave rectification and dynamic range compression. The HCM also introduces distortion products that corresponds to the beating frequencies. There is also a low pass filter that extracts the envelope of each channel. This filter agree with the loss of syncronization at the primary auditory nerve.

---

[1]Gammatone filter bank is a constant bandwidth filter bank.The gammatone auditory filter can be described by its impulse response: $y_{tone}(t) = at^{n-1}e^{2\pi bt}cos(2\pi f_c t + \phi)$, for $t > 0$, where the parameters a and b determines the duration of the impulse response and n the filter's order.

This toolbox contains modules wich deal with different aspects of perception and can provide good estimations of sensory, perceptual and cognitive parameters of the sound.

4. Conclusions

The research carried out up to now has shed light on the mechanisms that govern the human auditory sense. However many issues involving the function of the auditory periphery are still unresolved. New publications suggest that factors such as displacements of the stereocilia due to the influence of Brownian motion and stochastic resonance enhances the detection of weak signals in the middle frequency range. The ultimate mechanism of perception and the transmission of neural information to the brain is still a cloudy subject. These processes are dealt as *black box* systems (we are unaware of what is going inside) but, still, we might be able of observing and measuring subjects behavior, in response to prescribed auditory stimuli. This approach together with new measurement techniques might give us the information necessary to draw conclusions and understand to which extent auditory physiology and psychology behavior are brought into harmony.

References

Cosi, P., Poli, G. D., and Lauzzana, G. (1994). Auditory modelling and self-organizing neural networks for timbre classification. *Journal of New Music*, 23:71–98.

Flanagan, J. L. (1960). Models for approximating basilar membrane displacement. *Bell System Technology Journal*, 39:1163–1191.

Flanagan, J. L. (1962). Models for approximating basilar membrane displacement ii. *Bell System Technology Journal*, 41:959–1009.

Flanagan, J. L. (1972). *Speech Analysis, Synthesis and Perception*. Springer-Verlang, end edition.

Lyon, R. F. and Mead, C. (1988). An analog electronic cochlea. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(7):1119–1134.

Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory recepter. *Journal of the Acoustical Society of America*, 79(3):702–711.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (1995). Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, 98:1890–1894.

Sumner, C. J., Lopes-Poveda, E. A., O'Mard, L. P., and Meddis, R. (2002). A revised model of the inner-hair cell and auditory-nerve complex. *Journal of the Acoustical Society of America*, 111(5):2178–2188.

Toiviainen, P., Kaipainen, M., and Louhivuori, J. (1995). Musical timbre: Similarity ratings correlate with computational feature space distances. *Journal of New Music Research*, 24:282–298.

von Békésy, G. (1960). *Experiments in Hearing*. McGraw-Hill, 1st edition.