AUDIO ANALYSIS BY A MODEL OF THE PHYSIOLOGICAL AUDITORY SYSTEM

Turicchia L., De Poli G., Mian G.A.

Dipartimento di Elettronica e Informatica
Università di Padova
lorenzo.turicchia@pd.infn.it

Nobili R.

Dipartimento di Fisica "G. Galilei"
Università di Padova

ABSTRACT

In this paper, an analysis of flute attacks processed by a model of the physiological auditory system is presented. In flute performance, the musician uses consonants (/d/, /g/, /k/, /t/ and /p/) in order to create particular effects as hard and soft attacks. These effects are very important in music interpretation. We found that the model discriminates the sounds very well and better than spectral analysis as worked out by standard methods. The model responses appeared very detailed and allowed attack classification by the application of very simple pattern recognition techniques.

1. INTRODUCTION

The analysis of sound signals differs substantially from that of other signals. As in the auditory system of animals the acoustic signals are heavily transformed by the hearing organ, the knowledge of how the latter works is essential for the successful detection of audio contents. The simulation of the natural process is expected to detect sound components that are important to sound perception, whilst bypassing the unimportant ones.

Usually, the audio analysis algorithms follow a TOP-DOWN approach, the processing technique being guided by psycho-acoustic experiments (psycho-acoustic models).

In our study, a BOTTOM-UP approach was applied instead. Some processes occurring in the auditory system were taken into account in order to develop an algorithm that reproduces psycho-acoustical effects (physiological model).

This approach proved to be more reliable than other, especially in transient audio analysis, for three main reasons:

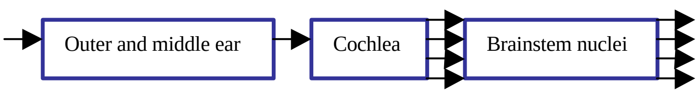
1. Natural selection provided mammals with the best auditory organ in the animal kingdom, and it did so because the basic task that the hearing function had to perform was detecting as fast as physically possible all those sound minute features that allow the prompt identification of a community member, as well as their operative meaning. Conceivably, this is the main reason why natural signal processing is so efficient. We expect therefore that its study will unveil many interesting things.
2. Signal processing techniques inspired by natural systems are expected to facilitate the extraction of signal features that are relevant at psycho-acoustic level.
3. In transient analysis, it appears very difficult to develop efficacious heuristic algorithms because of the very short transient duration (~10 ms). The process performed by the auditory system, which is highly non-linear, solves this problem.

Up to a few years ago, it was generally believed that the detection of acoustic information was almost entirely performed by the nervous system and that the auditory organ simply worked as a passive interface between the world of sounds and the brain. The discovery that the core structure of the peripheral auditory organ, the cochlea, plays a fundamental role in acoustic signal processing is a recent achievement of contemporary science. This is not so surprising, considering that the accessibility of the cochlea to experiments is severely limited by the anatomical structure. The internal architecture of the cochlea is very sophisticated and peculiar, and its cellular components and structures are characterised by extremely unusual properties [1,2,3]. This indicates, among other things, for how long and to what extent Nature had to be working, selecting an impressive collection of minute functions and improving their control mechanisms, in order to optimise the performance of the auditory function.

2. AUDITORY MODEL

We model the auditory system as composed by three parts:

1. Outer and middle ear
2. Cochlea (inner ear)
3. Brainstem acoustic nuclei



2.1. Outer and middle ear

Since at moderate sound regimes both the outer and middle ear are linear systems, they can be simply modelled as a band-pass filter [1].

2.2. Cochlea (inner ear)

Owing to the unique properties of its intrinsic dynamics, the cochlea works as a "paradoxical filter" that combines response speed with remarkable frequency selectivity. The latter property is not achieved through the fine-tuning of selective frequency filters but rather through a non-linear operation known as *tone-to-tone suppression*.

Another important feature of the cochlear response, which depends on hydrodynamically generated phase ordering of local oscillations, is to disentangle the attacks and transients of complex sounds in an optimal manner for the time domain analysis performed by the acoustic nuclei of the nervous system. This also permits the detection of the fundamental frequency (pitch) of voiced sounds, even in the absence of this component. These functions are crucial for the recognition of speech consonants and the discrimination of different speakers.

Briefly, the cochlea can be modelled [4,5] as an array of non-linear oscillators, each one of which is coupled instantly to all the others through hydrodynamic forces transmitted by the fluid that fills its interior. The input to the cochlea through the ossicle chain of the middle ear is also transmitted hydrodynamically to the oscillator array. Owing to the different physical parameters of the oscillators (mass, stiffness and damping constants), the response of the array to a frequency tone peaks at a frequency dependent location within the cochlea. The typical response is a travelling wave characterised by a phase delay that increases monotonically along the oscillator array, accumulating a few cycles up to the location of the response peak. Because of these features, the cochlea, as a frequency filter, differs substantially from a standard spectrum analyser. Physically, each oscillator is realised by a segment of the cochlear partition comprised between the elastic basilar membrane (BM) and the tectorial membrane (TM) (Figure 1).

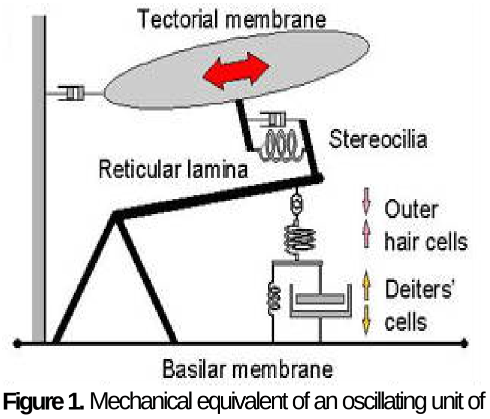


Figure 1. Mechanical equivalent of an oscillating unit of the cochlear partition

The motion equation of a BM segment is:

$$m(x)\frac{\partial^2}{\partial t^2}x(x,t)+h(x)\frac{\partial}{\partial t}x(x,t)+[\frac{\partial}{\partial x}s(x)\frac{\partial}{\partial x}]\frac{\partial}{\partial t}x(x,t)+$$
$$+k(x)x(x,t)=F_s(x,t)+F_{BM}(x,t)-F_U(h(x,t))$$

where $x\in[0,1]$ is the normalized BM coordinate in the longitudinal direction, t is time, $x(x,t)$ is the BM vertical displacement, $m(x)$ is the mass per unit length, $h(x)$ accounts all resistances affecting the absolute vertical motion of the organ of Corti, $k(x)$ is the elastic constant of the cochlear partition, $[\frac{\partial}{\partial x}s(x)\frac{\partial}{\partial x}]\frac{\partial}{\partial t}x(x,t)$ is the shearing resistance term, $F_S(x,t)=-G_S(x)\frac{\partial^2}{\partial t^2}s(t)$ is the stapes force term representing the force per unit BM length caused by the stapes motion and transmitted by the fluid to a BM segment at x (where $s(t)$ is the stapes displacement at time t and $G_S(x)$ is the stapes force propagator), $F_{BM}(x,t)=-\int_0^1 G(x,\overline{x})\frac{\partial^2}{\partial t^2}x(\overline{x},t)d\overline{x}$ is the BM force term representing the force per unit BM length acting on the BM at x due to the motion of all the BM segments and simultaneously transmitted by the fluid (where $G(x,\overline{x})$ is the Green's function of the fluid pressure field over the BM), and $F_U(h(x,t))$ the force term to oppose damping (i.e. undamping term) representing the force generated by the outer hair cell motors when the stereocilia undergo TM displacements $h(x,t)$. This force term vanishes at $h(x,t)=0$ and has a sigmoidal shape to account for the saturation properties of the cochlear amplifier.

The motion equation of TM's section is:

$$\overline{m}(x)\frac{\partial^2}{\partial t^2}h(x,t)+\overline{h}(x)\frac{\partial}{\partial t}h(x,t)+\overline{k}(x)h(x,t)=$$
$$=-C(x)\frac{\partial^2}{\partial t^2}x(x,t)$$

where $\overline{m},\overline{h},\overline{k},C(x)$, represent mass, damping, stiffness and coupling constant, respectively.

In this work we neglected the shear viscosity between adjacent section [i.e. $s(x)$=0] and reduced BM and TM motion equations to only one equation. This reduction is possible because at resonance, the first and third terms at the left hand side of the above equations cancel leaving after time integration:

$$h(x,t) = -\frac{C(x)}{\bar{h}(x)}\frac{\partial}{\partial t}x(x,t)$$

Therefore the undamping term is:

$$F_U(h(x,t)) = S\left(-\frac{C(x)}{\bar{h}(x)}\frac{\partial}{\partial t}x(x,t)\right) = -Q(x)\bar{S}\left(\frac{\partial}{\partial t}x(x,t)\right)$$

where $S(\ldots)$ and $\bar{S}(\ldots)$ are a suitable sigmoid-shaped functions. We modeled one segment of the organ of Corti by equation:

$$m(x)\frac{\partial^2}{\partial t^2}x(x,t) + h(x)\frac{\partial}{\partial t}x(x,t) - Q(x)\bar{S}\left(\frac{\partial}{\partial t}x(x,t)\right) +$$

$$+k(x)x(x,t) = -G_S(x)\frac{\partial^2}{\partial t^2}s(t) - \int_0^1 G(x,\bar{x})\frac{\partial^2}{\partial t^2}x(\bar{x},t)d\bar{x}$$

In this equation we expressed cochlear non-linearity through $\bar{S}(\ldots)$, a sigmoid function that in low-sound regime becomes linear and compensates for the damping factor as described by the following equation:

$$h(x)\frac{\partial}{\partial t}x(x,t) - Q(x)\bar{S}\left(\frac{\partial}{\partial t}x(x,t)\right) \cong$$

$$\cong [h(x) - Q(x)]\frac{\partial}{\partial t}x(x,t) = h_{LOW}(x)\frac{\partial}{\partial t}x(x,t)$$

At larger amplitudes, the transducer saturate, the undamping is overcome by viscous losses and the basilar membrane response approaches that of a passive cochlea:

$$h(x)\frac{\partial}{\partial t}x(x,t) - Q(x)\bar{S}\left(\frac{\partial}{\partial t}x(x,t)\right) \cong h(x)\frac{\partial}{\partial t}x(x,t)$$

The transition between the two different damping regimes makes the basilar membrane input/output function highly non-linear

In conclusion we considered the cochlear model as a nonlinear integro-differential equation:

$$\int_0^1 W(x,\bar{x},\partial t)x(\bar{x},t)d\bar{x} = -G_S(x)\frac{\partial^2}{\partial t^2}s(t)$$

where

$$W(x,\bar{x},\partial t) = \{m(x)\frac{\partial^2}{\partial t^2}x(x,t) + h(x)\frac{\partial}{\partial t}x(x,t) +$$

$$-Q(x)\bar{S}\left(\frac{\partial}{\partial t}x(x,t)\right) + k(x)x(x,t)\}d(x-\bar{x}) +$$

$$+G(x,\bar{x})\frac{\partial^2}{\partial t^2}x(\bar{x},t)$$

Note that, although the model is only a simplified representation of the real cochlea, it accounts for the instantaneous hydrodynamic coupling among the BM sections. This characteristic is not taken into account in many other models based on the transmission line paradigm.

2.3. Brainstem nuclei

Brainstem nuclei behaviour is very complex. We implemented one principal feature: the inter-peak time computation of the auditory nerve response. This is a very important feature because it permits the tracking of very fast frequency sweeps. Indeed sound processing performed by the cochlea is not sufficient to extract many sound information like the position of formant frequencies in vocal sound. For this reason we introduced a further stage, providing a time-domain simulation of the auditory nerve processing, which is used to compute peak-to-peak interval as detected at nerve response level.

3. TRANSIENT ANALYSIS SIMULATION

We tested the model on two types of signals: *voice signals* [6]and *flute attacks*. In both cases, we obtained that the model discriminates sound transients very well.

In this paper we present an analysis of flute attacks. In flute performance, the musician uses consonants (/d/, /k/, /g/, /p/ and /t/) in order to create particular effects like hard or soft attacks. These effects are very important in music interpretation but are hard to recognise with standard spectral analysis because the transient is very fast (~10 ms). In figure 2,3,4,5 and 6 we present the model response to /d/, /k/, /g/, /p/ and /t/ flute attacks. We formed a histogram of intervals occurring between the peaks of the auditory nerve response. Vertical axis measures the interval duration, horizontal axis the time at which the peaks occur, plotted in different grey colour which are weighted proportionally to their occurrence rate (white =0).

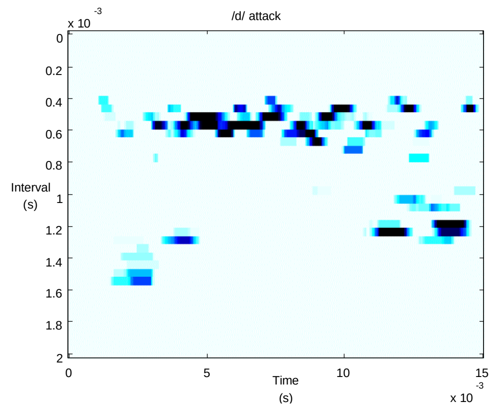


Figure 2 - Model response to /d/ flute attack.

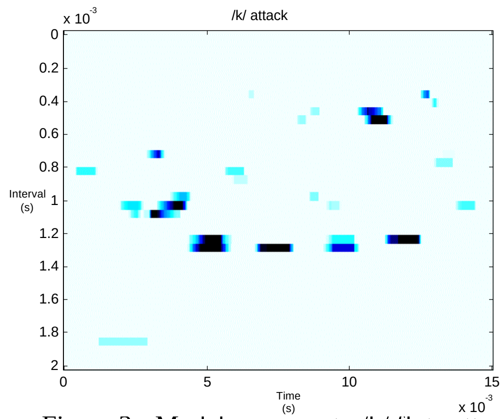


Figure 3 - Model response to /k/ flute attack.

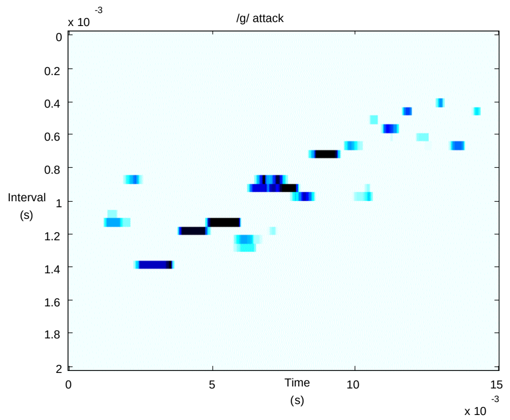


Figure 4 - Model response to /g/ flute attack.

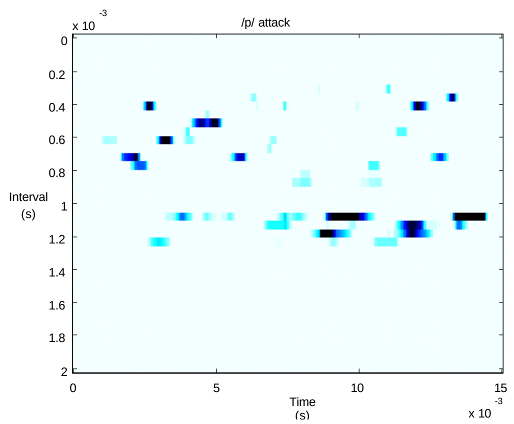


Figure 5 - Model response to /p/ flute attack.

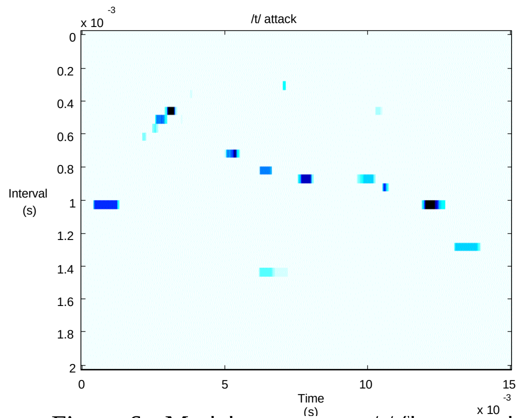


Figure 6 - Model response to /t/ flute attack.

We found that the model discriminates sounds very well and better than the usual spectral-analysis-based methods would allow. The model responses appeared very detailed and permitted the attack classification through very simple pattern recognition techniques.

4. CONCLUSION

Transient analysis is generally hard to be solved by standard spectral methods. We advanced some arguments to prove that models mimicking the physiological processes can provide valid strategies for signal analysis.

In this paper we used a simplified model of the mammalian cochlea and a putative brainstem nuclei function addressed to a detailed analysis of flute attacks. We obtained a good discrimination of transient sounds. The method here described might be successfully employed as the first stage of a micro-gesture extraction machine.

5. REFERENCES

[1] J.O. Pickles (1988). An introduction to the physiology of hearing. Academic Press, London 1988.

[2] A.W. Gummer (1996). Resonant tectorial membrane motion in the inner ear: Its crucial role in frequency tuning. Proc. Natl. Acad. Sci. USA, Vol. 93, pp. 8727-8732, August 1996.

[3] S.M. Khanna, L.F. Hao (1999). Reticular lamina vibrations in the apical turn of a living guinea pig cochlea. Hearing Research, 132 (1999) 15-33.

[4] Mammano, F. and Nobili, R. (1993). Biophysics of the Cochlea. Linear Approximation, J. Acoust. Soc. Am., 93: 3320-3332

[5] Nobili, R. and F. Mammano, F. (1996). Biophysics of the cochlea. II. Steady-state non linear phenomena, J. Acoust. Soc. Am., 99: 2244-2255.

[6] Nobili, R., Turicchia, L., Mammano, F. (2000) - Frequency analysis by the cochlea: spatial and temporal aspects-INFM Meeting (Genova, June 2000).