We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Issues in Creation of Bio-Compatible Cochlear Signal: Towards a New Generation of Cochlear Prosthetic Devices

Włodzimierz (“Vlad”) Wojcik

Abstract

A model of a fully functional cochlear prosthesis is presented here, simplified by taking into account only those evolutionary features of natural cochleae that contribute to their functionality. The proposed prosthetic device generates a bio-compatible digital signal which can be fed to the cochlear nerve. Subsequently, analysis of cochlear nerve signals is offered, both natural and artificial. To that end a number of mathematical theorems are formulated, proven, and then used to demonstrate that signals obtained from our prosthetic device are useful towards auditory pattern recognition, audio location, and even speech comprehension, as well as understanding and enjoyment of music.

Keywords: audio pattern recognition, auditory pathway, auditory cilia, cochlea, cochlear model, cochlear nerve signal, cochlear prostheses, mechanoreceptors

1. Introduction

Imagine a deaf patient who lost her outer, middle, and inner ears through disease or injury. The remaining parts of her auditory pathway (cochlear nerve, cochlear nucleus, superior olive, lateral lemniscus, inferior colliculus, medial geniculate body, and auditory cortex) are intact. Although profoundly deaf, the patient is not hopeless. She expects us to fit her with two electronic devices that would feed suitable signals to her cochlear nerves, thus allowing her to hear again.

The current cochlear implant devices are exceedingly basic: they normally assist or compensate for certain malfunctions of damaged parts of middle or inner ear (auditory cells, tectorial membrane, etc.) but offer only crude restorative signals spanning approx. 10 channels. To appreciate the inadequacy of the current state-of-the-art, the reader might want to peruse simulations at [1].

In contrast, healthy human cochlear nerve consists of approx. 30,000 neural axons, with almost every axon conveying one channel of a signal. Humans do not hold any records in this regard: feline cochlear nerves contain approx. 50,000 axon fibers. The information traveling through neural axons is digital in nature, due to all-or-nothing

neural axon hillock operation when generating signal. No wonder then that the problem of refining cochlear implants catches interest of some computer scientists.

Formation of the complete multi-channel, representative signal adhering to the biological cochlear nerve signal protocol is a daunting task, requiring deep understanding of the workings of outer, middle, and inner ears, which participate in the signal generation process, as well as the workings of organs in the auditory pathway, which receive and interpret these signals. While the functionality of ear organs is relatively well understood, the same cannot be said about the organs higher up the auditory pathway. We shed some light here on their functionality.

The biologists and physicians instill in us the awe for the complexity of biological organs. Ears are such organs. Outer ears are sound channeling and amplification devices, processing sound waves traveling through air. Middle ears are mechanical in nature: their tiny ossicles (malleus, incus, and stapes) controlled by equally tiny muscles amplify weak signals or dampen strong signals before conveying them to inner ear. In this way the inner ear is protected. Finally, inner ears are wonders of electromechanical and electrochemical precision, governed by the laws of fluid dynamics, etc.

Our goal is to build an electronic device creating a fully restorative cochlear signal without entering too deep into details of functionality of biological ears. We argue that our ears are unnecessarily complex for their function. They emerged through the process of evolution, not intelligent design. That evolutionary process consists of introducing random changes into the genotype (via crossover and mutation) and then ensuring of survival of the fittest individuals via natural selection. In this way only deleterious genetic changes are eliminated. Neutral changes are allowed to remain, thus unnecessarily increasing the complexity of emerging systems.

Biological textbooks offer many examples of such unnecessary complications introduced by the evolutionary process: consider the path of laryngeal nerve, allowing the brain to control the larynx. The nerve's inferior (recurrent) branch, innervating larynx muscles, is unnecessarily long: it descends from the brain into the thorax only then to return to the larynx. In medical textbooks it is frequently cited as an evidence of evolution: in giraffes this extra length can be measured in meters! This feature is a vestige of an evolutionary process—in fish throats and gills are in immediate vicinity so this detour did not exist when fish were the most complex animals.

Another example could be the structure of an eye. Human retina has a blind spot; also, tiny blood vessels supplying blood to the retina are positioned in front of it, rather than behind it, thus occluding some sensory cells (rods and cones). Such clumsy “design blunder” can only evolve through a process of mutations and natural selection. To compensate for that “blunder” the brain's visual cortex had to evolve unnecessary complexity. We have already decoded the fundamental workings of the visual cortex [2, 3]; here the same methodology is used to attack the problem of signal processing for auditory perception.

The complexity of the central nervous system (CNS) to a large degree is due to existence of structures compensating for its earlier structural inefficiencies. The difficulty in figuring out the inner workings of CNS stems from the fact that the biologists must study with equal diligence all parts of the CNS: those essential and those secondary to its functionality. They have no choice there: to distinguish between the roles of these parts is to understand the CNS functionality.

This multi-disciplinary chapter consists of two parts:

We start by presenting a simplified but functionally complete model of our auditory perception, together with an electronic model of a device capable of providing

restorative signal to the cochlear nerve, thus suggesting a new generation of cochlear prostheses. This functional description is targeted at electronic engineers who may have little biological background but will be tasked with creation of electronic circuits producing suitable e-cochlear signals.

Following that we demonstrate mathematically that signals so obtained carry information useful to the healthy part of the auditory pathway to perform the well-understood functions of audio location, as well as the more complex functions of audio pattern recognition and may even give rise to the subjective pleasure of music, so enjoyed by healthy humans. This section is written with some mathematical rigor.

2. The nature of auditory perception

Departing from the traditional phase-based or frequency-based models of cochlear and vestibular signals we offer simpler, unified, and biologically inspired insights that rely on massive parallelism, needed for real time auditory pattern recognition. Our signal analysis is also applicable to vestibular signals, used to balance body equilibrium and posture, as well as to stabilize retinal image.

To justify our approach to auditory perception modeling consider a hare hearing a noise: it uses its audio signal to detect a possible predator approaching, in order to take a prompt evasive action. Our hare uses its auditory pattern recognition skills live, as soon as it starts hearing the noise; this is necessary for hare's survival.

Luckily for all of us (hares included) the signals we perceive are causal. A signal $f(t)$ is causal, iff

$$f(t) = 0 \text{ for all } t < 0 \quad (1)$$

This work deals with causal signals, in the above sense.

To cope with causal signals, we build causal systems. An output $y(t_0)$ of a causal system at a given time t_0 depends only on its causal input $x(t)$, for all $t \leq t_0$. We create causal systems by building their causal models first.

The simplest (trivial!) causal system can be modeled as follows:

Given an input signal $x(t)$ for $0 \leq t \leq t_0$, we may create a causal model that copies the input signal to its output, that is, produces $y(t) = x(t)$. A memory-less causal system using such a trivial model is pretty much useless: it cannot understand the signal it processes.

We strive to create systems that use causal models in anticipatory fashion, that is, intending to forecast (within desired accuracy and meeting specific time deadlines) the input signal $x(t_1)$ at some future time $t_1 > t_0$ on the basis of a known input $x(t)$ for $0 \leq t \leq t_0$. For that we need an intelligent system.

Indeed, this is the essence of *intelligence* demanded of a system meeting challenges of its environment. A system is intelligent if it is capable of forecasting reliably enough its future causal input signals, while meeting forecasting deadlines. The longer forecast horizon and the better accuracy of prediction, the higher is the intelligence of such a system.

Example: Our hare hearing a noise may estimate the proximity, the direction, and speed of an approaching predator, and then timely identify and execute suitable evasive action. A hare capable of routinely saving its skin in this way is considered intelligent, that is, well adapted to its environment. To be so adapted the hare need not only to be versed in auditory perception, but also capable of real-time execution of its algorithms.

Example: Humans understand and like music. What is it exactly that we like?

When we say that we understand a causal signal we mean that we perceive patterns in it, and, using them, we are capable of building a predictive model of the signal. In this chapter we preoccupy ourselves with audio signals. Some of them we perceive as noise, some of them as music. How can we tell the difference? Or, is there a difference?

Following the logic just presented we say that we understand a particular audio signal, if we can forecast future signal values with satisfactory accuracy within some time horizon. To understand an audio signal means to have a useful model of it. We feel pleasure when we can accomplish this, and term such signal music (or speech, etc.) Other signals are termed noise. Additional pleasure is instinctively drawn from the realization that we are coping successfully with the challenges of our environment, and in particular, when listening to music that we are not in any danger regardless of the success or failure of our predictions; and so, we can hone our predictive skills in complete safety by listening to music, perhaps repeatedly.

Observe further that a previously taped “musical” signal can still be perceived as “noise” if reproduced at modified speed, or perhaps played backwards, thus making real-time forecasting impossible. This reinforces our view that the signal counts only as “music” if we can understand it, that is, identify patterns within it, use these patterns for forecasting its future passages while executing in real time the entire computation of forecasts, before the actual signals arrive. That successful verification of hearing our forecasts coinciding with reality as time advances (within “reasonable” tolerances) we call enjoyment of music or understanding of speech – at least in terms of adherence to relevant grammatical rules.

The larger goals of our research are twofold: (1) to further the understanding of the working of the CNS including human brain and (2) to learn how to build smarter devices, capable of interfacing with our CNS. Here we pursue both goals, using the methodology previously successful in decoding fundamentals of visual signal processing [2, 3]. Drawing on biological inspirations we limit our use of math tools to set theory and theory of metric spaces. After all, even simple animals seem to execute in timely fashion the forecasting models appropriate to their environmental niches. We doubt that simple animals have sufficient computing skills to execute complex math in real-time. In the spirit of Occam’s razor, we value simplicity. Simple models can be executed fast. When it comes to survival, speed counts!

Although the focus of this chapter is on cochlear signal processing, it is worth noting that all nerve signals are similar in structure and this similarity is of immediate practical interest in vestibular nerve signal processing. The vestibular system is responsible for sensing body equilibrium and posture, as well as for stabilizing eyeballs when we move our heads (a *Steadicam* function) [4–7].

3. Basic model of auditory perception

For the purposes of this chapter, we would like to offer a simplified engineering model of human auditory perception. At the outset we beg for indulgence of our biologically trained readers: detailed descriptions of this system are available elsewhere [8], including other chapters of this book.

Our description focuses on features of the system in need of further explanation, namely those of interpretation of signals by the auditory cortex. We limit our description of anatomy to essential features only.

The brain's auditory pathway consists of *cochlear nucleus*, *superior olive*, *lateral lemniscus*, *inferior colliculus*, *medial geniculate body*, *thalamus*, and terminating at *auditory cortex*.

Vigorous research activity is aimed at understanding the functionality of these brain nuclei by studying their neural structure. We propose a different approach, which we tested successfully when studying the functionality of human vision system [2, 3]. As pointed out already, we are convinced that these brain nuclei are exceedingly complex, as they contain structures essential to their function as well as those merely vestigial to the evolutionary process. To discern between those two kinds is not possible without understanding the functionality of the auditory pathway, which is the very goal of this research effort.

Facing this circular conundrum, we prefer another investigative approach: given the information traveling via cochlear nerve we attempt to deduce the structure and functionality of an abstract computer needed to account for auditory perception we (humans) experience. We depart from the assumption that intelligence must be wet or carbon-based; for us these are just details of a particular implementation. We prefer silicon.

Figure 1 depicts the schematic anatomy of the human ear. We mention in passing the well-known facts: the acoustic signal, that is, the air-pressure wave, traveling through the external auditory canal impedes on tympanic membrane, a.k.a. eardrum, causing it to vibrate.

The tiny bones of the middle ear (malleus, incus, and stapes) convey the mechanical vibrations of the eardrum to the cochlea where the sound perception begins.

The cochlea is a chamber filled with fluids (called perilymph and endolymph), which are in turn induced to vibrate. In engineering terms, the cochlea acts as an attenuating waveguide. It is most permeable to low frequencies, while strongly attenuating high frequencies.

Human ear can at best discern frequencies within the 20 Hz–20 kHz range. Other animals can hear within different frequency ranges, but the principle of operation of their cochleae remains the same: high frequencies are perceived only in the cochlear region close to the stapes, middle frequencies penetrate deeper, but only low

Figure 1.
Basic anatomy of the human ear.

frequencies can travel through entire cochlea. This distribution of frequency sensitivity along the length of the cochlea is referred to as *tonotopy*.

As the cochlear perilymph vibrates, it actuates the *auditory cilia*, that is, “hairs” of the “hair cells,” which convert these vibrations into electrical signals. These signals are then communicated (via *neurotransmitters*) to the nerve cells of the cochlear nerve, which passes the signals to the brain.

The hair cells are organized into the *organ of Corti*, shown in the cochlear cross-section of **Figure 2**.

We omit here again many details of primary interest to anatomists, while focusing mainly on the cochlear duct called *scala media* (shown in green in **Figure 2**). It is filled with *endolymph* and contains the organ of Corti, attached to the *basilar membrane*, being one of the walls of *scala media*.

Organ of Corti contains large number of mechanoreceptors, called *auditory cells*. Each auditory cell is equipped with 100–150 whiskers, called *auditory cilia*, all bathed by vibrating endolymph. Some of the longer cilia are attached to the *tectorial membrane*. Both basilar and tectorial membranes also vibrate with the endolymph.

Figure 3 shows the rudimentary schematics of an auditory cell.

At resting state (no sound wave) the auditory cell body maintains certain equilibrium, measured in terms of an electrochemical potential of cell body interior. It sends then no signal to its cochlear neuron.

The sound vibrations of the endolymph in *scala media* impact on the single cilium shown causing it to bend sideways.

Deflection of the cilium in one direction causes excitatory increase of potential inside cell body. When that potential reaches a certain threshold level, called *action potential*, the cell issues into its synapse a certain amount of a specific neurotransmitter, which in turn causes electric signal to travel through the cochlear neuron.

Deflection of the cilium in opposite direction has inhibitory effect on cell body potential, thus silencing the cell.

Figure 2.
Cochlear cross-section showing the organ of Corti.

Figure 3.
 Rudimentary schematics of an auditory cell.

In short, responses to vibration of the endolymph in scala media of the organ of Corti cause periodic changes in the potential of the auditory cell, thus causing a series of pulses travel through the *axon* of a cochlear neuron (i.e., *cochlear nerve fiber*) associated with that auditory cell.

The cochlear nerve is a bundle of axons of cochlear neurons. Consequently, signals traveling through cochlear nerve constitute simultaneous groups of series of pulses, each series being transmitted sequentially by a particular fiber of the nerve.

Given that the cochlea acts as non-linear waveguide, attenuating various frequencies differently, every place along its length experiences different sound pattern. For the purpose of pattern recognition, it would then be best if a number of mechanoreceptors (hair cells) were placed in exactly the same cochlear location.

However, every hair cell has finite dimensions, so it is impossible to place a number of them in the same location. Nature had no choice but to distribute them along the organ of Corti. No wonder then that different hair cells respond differently to various frequencies. All these cells are anatomically identical; their varying responses result from being exposed to different vibrations of the endolymph in their particular locations. This is the essence of tonotopy.

Nature has done everything possible to reduce the consequences of building the organ of Corti, populated by hair cells, as having non-zero dimensions, by filling the inner ear with liquid. The speed of sound waves in liquids is an order of magnitude higher than in gases. The relationship between the wavelength λ and the sound frequency f is $\lambda = v/f$, where v is the speed of sound in each of these media. From this it follows that for a given frequency f the sound wavelength in liquid is much longer than in gas, thus reducing the relative consequence of distribution of hair cells along the organ of Corti. The auditory cilia of these cells vibrate much more in sync when bathed by a liquid than they would if they were in gas. Furthermore, the inner ear is encased in the hardest bone found in a human body. This arrangement maximizes the agitation of cilia, by forcing the liquid to vibrate along the cochlea, and so perpendicularly to the cilia.

High audio acuity requires large numbers of auditory cilia, and so a larger organ of Corti in a longer cochlea. No wonder then that in simpler animals the cochleae are straight, while more complex animals have their cochleae coiled. It all has to do with the problem of packing of an elongated cochlea into a small cranial cavity within an exceptionally hard bone.

Loudness is a subjective quality of sound that is an attribute of auditory perception. We all classify sounds on some scale ranging from “quiet” to “loud.” However, sound experience is a creation of our brain, a correlate with a physical phenomenon (a sound wave) impacting on our ears. Sound waves can be measured in terms of physical values like power, amplitude, frequency, sound pressure, etc.

Given that our eardrums react to sound pressure, it seems logical to correlate loudness with sound pressure. This relationship, depicted in **Figure 4**, is particularly dependent on frequency. We can easily perceive relatively weak sound waves, as long as their frequencies lie within an interval of 1–4 kHz, this being the predominant frequency range of human speech. Sound waves outside of that interval (both of higher and lower frequencies) must be substantially stronger to be perceived as equally loud. This non-linearity arises due to the inertia of the mechanical parts of our ears (eardrums, malleus, incus, and stapes) as well as the inertia and drag of the auditory cilia against the surrounding liquid. Our novel cochlear prostheses need to account for all these nonlinearities.

To outline our modeling methodology, consider an arbitrary signal (shown in blue), traveling through endolymph and impacting on an auditory cilium of a hair cell, depicted in **Figure 5**.

That cell has an action potential $a_p = 0.25$ units and generates a packet of rectangular pulses (shown in red) inscribed within the endolymph pressure wave as observed at the cell location. Its neighboring cells may have action potentials set at different values, and together they describe the pressure wave with some accuracy. Note that this description is not exact, as one would need (mathematically speaking) an infinite number of cells at the very same location in scala media, all with different action potentials (forming a continuum) to describe the signal exactly. In that sense the pressure wave could be seen as curve enveloping all possible inscribed square pulse signals.

The illustration in **Figure 5** depicts a wave of sufficiently low frequency, so that the relevant auditory cell is able to inscribe a square pulse into each peak of the audio

Figure 4. Contours of equal loudness (red) as functions of frequency (Hz), ISO 226:2003 revision. Older ISO standard for 40-phon shown in blue.

Figure 5.
Hair cell with action potential $a_p = 0.25$ inscribes a square-wave signal (red) within a sample pressure wave (blue) of an endolymph.

wave. Not all audio cells are capable of such behavior. Due to their inertia, they need time to adjust their electrochemical potential levels. An inhibited call will need more time to bring its potential to the action level than a cell that was in a neutral state. After firing a pulse each cell requires certain amount of time to bring itself into the state of readiness to fire again.

The net result of this is that the cells do not always faithfully inscribe their pulses into the acoustic wave, although the firing of their pulses is stimulated by that wave. The inertia of all cells is similar and so they generate pulses of similar frequency characteristics. It is their location in the cochlea and their connections to relevant axons of the cochlear nerve that causes us to perceive varying pitch.

The emerging bundle of signals formed in this way, traveling along the cochlear nerve, carries to the auditory cortex the information about the ambient sound wave.

Specifically, there is only a finite number of cells with different action potentials in a particular region. That region being small, we can assume that all of them respond in sync to the same pressure wave. Our model deals therefore with bundles of square pulse approximations of the sound wave. According to the Shannon-Hartley theorem [9, 10] these approximations can be still exact even for a finite number of hair cells used, provided that their stimulating signal lies within certain harmonic limitations.

The frequency region of 20 Hz–20 kHz is not such limitation, as it only describes limits of the audible frequency response curve for human hair cells. Acoustic signals outside this range are merely inaudible but they do exist. Certainly, some harmonic limitations of an acoustic signal exist. Given that wave energy is proportional to the wave frequency, there must be some limit to the number of frequency components of a given audio signal for the signal to carry finite energy. We inhabit a universe filled with signals of finite energies—because of that we can exist.

Let us take a closer look at square pulse signals of finite length. There is a number of possible descriptions of a signal $s(t)$ consisting of n pulses:

$$s(t) = \langle t_0, t_1, t_2, t_3, \dots t_{2n-1} \rangle \tag{2}$$

where $\langle \bullet, \bullet, \dots, \bullet \rangle$ is a tuple of $2n$ values listed in strictly ascending order, that is, for every pair of values t_i, t_k such that $i < k$ we have $t_i < t_k$. The time values with an even index define the beginning of a pulse, the times with odd index represent the end of a pulse.

Alternatively, we can represent signal $s(t)$ as a set of mutually exclusive time-line intervals:

$$s(t) = \{[t_0, t_1), [t_2, t_3), \dots, [t_{n-2}, t_{n-1})\} \quad (3)$$

Using this description signal values for arbitrary t can be calculated as follows:

$$s(t) = \begin{cases} a_p & \text{for } t_i \leq t < t_{i+1} \text{ where } i \text{ is even;} \\ 0 & \text{for } t_i < t \leq t_{i+1} \text{ where } i \text{ is odd.} \end{cases} \quad (4)$$

In fact, we will deal mostly with normalized signals, that is, assuming values 0 and 1. Such signals can be obtained by dividing the above pulse signal by its action potential, viz. $s(t)/a_p$ yielding:

$$s(t) = \begin{cases} 1 & \text{for } t_i \leq t < t_{i+1} \text{ where } i \text{ is even;} \\ 0 & \text{for } t_i < t \leq t_{i+1} \text{ where } i \text{ is odd.} \end{cases} \quad (5)$$

Alternatively, having defined a function $\text{step}(t)$ as

$$\text{step}(t) = \begin{cases} 0 & \text{for } t < 0; \\ 1 & \text{otherwise,} \end{cases} \quad (6)$$

we may describe a normalized signal $s(t)$ as:

$$s(t) = \sum_{i=0}^{2n-1} (-1)^i * \text{step}(t - t_i) \quad (7)$$

Observe that whatever the signal descriptions—Boolean algebra applies to normalized signals.

Let us label as **true** the value 1 of a normalized signal, and as **false** its value 0. Then, given two arbitrary normalized signals $f(t)$ and $g(t)$, possibly consisting of a different number of pulses of arbitrary timing, we can define standard Boolean operations as follows:

$$\text{not } f(t) = \text{step}(t) - f(t) \quad (8)$$

$$f(t) \text{and } g(t) = f(t) * g(t), \text{ and finally} \quad (9)$$

$$f(t) \text{or } g(t) = f(t) + g(t) - f(t) * g(t) \quad (10)$$

These definitions allow us to formulate:

3.1 Self-test procedures for cochlear prostheses

Consider two arbitrary normalized signals $f(t)$ and $g(t)$, generated by two artificial hair cells of opposite orientation but of the same action potential and belonging to the same cochlear neighborhood. By *opposite orientation* we mean that both of them

cannot be excited simultaneously. For properly functioning such pair of artificial hair cells of a cochlear prosthesis the following conditions must hold for every t:

$$\begin{aligned} f(t) \text{ and } g(t) &= 0 \\ f(t) \text{ and not } g(t) &= f(t) \\ g(t) \text{ and not } f(t) &= g(t) \end{aligned} \tag{11}$$

This is so because $f(t)$ and $g(t)$ are mutually exclusive signals, while neither of them is the pure negation of another.

These test conditions can also be used to test cochlear prostheses as well as for hearing loss due to the loss of auditory cilia, caused by exposure to exceedingly high sound levels. Human hair cells have many cilia and initial hearing loss is usually not noticeable until large numbers of cilia are lost.

3.2 Audio location using pulse signals

Our simple methodology is sufficient for description of tasks like audio location.

Consider two hair cells with the same action potential, housed by two cochleae in two ears of an animal. A sound wave excites both cells, as per **Figure 6**.

An animal is capable of estimating the azimuth of the audio source by calculating the angle (shown) of the incoming sound. The essence of this well understood process is as follows: two hair cells, located in two cochleae of an animal are not excited simultaneously by an arriving sound wave. The delay in excitation of the hair cell more distant from the sound source can be used to calculate the extra distance the sound wave must travel to reach that cell (shown in **Figure 6**) Given that the base distance between the two cells (also shown in **Figure 6**) is an anatomical constant for

Figure 6.
Audio location process: One of two auditory cells in two different ears detects the same signal with a bit of delay.

a given animal, and the speed of sound in the medium the animal inhabits is also constant, then the angle of arrival of the sound wave can be calculated from the right triangle shown. In fact, an audio locating animal does not need to perform those complex trigonometric calculations. It just needs to turn its head to make both signals arrive simultaneously; it then faces in the direction of the incoming sound.

Observe further that for this method to work the length of the sound wave cannot be too long nor too short. Too long waves would not result in contrasting deflections of the corresponding cilia; for too short waves the uniqueness of the solution is lost (due to the periodic nature of the oscillations).

Optimal locating acuity occurs when the deflections of the cilia are of opposite phase. That is why the optimal sound wavelengths ensuring this remain in close correlation with the distance between two corresponding cells located within two ears of an animal. This explains why humans use frequencies much lower than bats, for example. It all has to do with different distances between ears of humans and bats. The square pulse nature of the signals generated by the hair cells enhances contrast between these signals, thus facilitating the audio location process.

The above examples are intended to demonstrate the capability of our approach to model processes already well understood. To gain understanding of more complex processes of audio pattern recognition, music, etc. we need to introduce several new concepts. We start with the design of the prosthetic device.

4. The cochlear model and prosthetic device

Our concept of a modular multi-channel device allows the number of supported channels to grow arbitrarily as technology advances. According to current practice, the device consists of two parts: the implantable part and the external part, worn by the patient. This is in order to maintain skin continuity thus avoiding infection. Both parts communicate via a trans-dermal electromagnetic link.

The external part is responsible for the formation of the multi-channel restorative signal and for transmission of the multiplexed signal via the electromagnetic link to the implantable part. The implantable part de-multiplexes the received signal and feeds proper channel signals to the relevant fibers of the cochlear nerve via a micro-array of tiny electrodes. The technological challenges of connecting microscopic multi-channel cables to proper fibers of a nerve are currently being addressed by work of several teams, led by Charles C. Della Santina at Johns Hopkins University (Vestibular Neuroengineering Lab), Daniel Merfeld, Wangsong Gong, and Richard Lewis at Massachusetts Eye and Ear Infirmary (MEEI) in Boston, James O. Phillips of the University of Washington, Andrei M. Shkel at the University of California, Irvine, Julius Georgiou at the University of Cyprus, and elsewhere [4]. Although these expert teams focus their work on the vestibular nerve, the technological challenges of connecting a micro-array of electrodes to any nerve remain the same.

In fact, we can use the cochlear nerve fiber layout to facilitate the connection. **Figure 7** shows the cochlea and the outgoing nerve fibers. Observe that due to the spiral nature of the cochlea the central fibers of the nerve are connected to the cochlear tip and so are responsible for conveying information about low frequency components of the audio signal, while the outer fibers deal with high frequency components. Connection of a micro-array of electrodes to the nerve can therefore be a two-step procedure. First, we attach a micro-connector configured so that its center wires mate with center fibers, and then we fine-tune the connection by creating a

Figure 7.
 Cochlear nerve layout: center fibers convey information regarding low frequency components of the audio signal; perimeter fibers deal with high frequencies.

detailed connection map, customized to each patient. We do that by passing stimulation signals to individual wires while asking the patient about frequencies perceived.

We focus now our attention on the formation of the restorative signal by an electronic device. That device consists of a microphone of a suitable directional characteristic, worn by the patient close to his non-functional ear, a transducer ensuring the transdermal connection, and a small electronic box, containing analog and digital circuitry.

To simulate our patient's cochlea, we will turn to the time-tested technology: that of a telegraph line. Early telegraph operators noticed that long telegraph lines tend to distort signals generated by telegraph keys. On the receiving end square pulse signals tended to have their edges rounded. When the line was too long, this distortion made the received signal unintelligible: dots of the Morse code would disappear, while dashes became dots with gentle slopes. To avoid this, telegraph lines have had certain maximum length, beyond which "repeaters" were used to refresh the signal. Early repeaters were just people transcribing the signals, later replaced by electrical devices.

While the goal of telegraph line builders was to maximize the quality of the signal and thus to maximize distance between repeaters by minimizing distortion, our goal is to minimize the size of the "bionic ear" worn by the patient. To that end we will construct a model of a bad telegraph line distorting the signal over short distances, and we will not use any repeaters.

A circuit shown in **Figure 8** is a model of a telegraph line segment or of neural axon segment. The wires of this segment present certain electrical resistance R [measured in ohms] and given that such line is never perfectly straight, a certain inductance L [henries]. The insulator between the two wires (air, plastic, etc.) is never perfect and so has a certain conductance G [siemens], and, together with the wires offers certain capacitance C [farads].

Figure 8.
Electrical model of a segment of a telegraph line or of a “bionic cochlea.”

In physics the concepts of conductance G and resistance R are bound by an inverse relationship: $G = 1/R$. In the circuit of **Figure 8** the values of G and R are unrelated; they are just intrinsic parameters of a line segment.

Such circuit distorts an input signal applied to its left terminals by selectively attenuating harmonic components of the signal: higher harmonics are attenuated more than the lower harmonics. A telegraph line or our “bionic cochlea” nerve fiber can be represented by a finite sequence of such segments, as per **Figure 9**.

In a real cochlea information contained in higher harmonics of a signal is gathered close to the entry of the cochlea (i.e., close to the *oval window*), while only lower frequencies can penetrate to the end of the cochlea. The cochlea has a special terminator (called *round window*). That terminator has a dual role.

First, it allows the non-compressive endolymph, encased in a very hard bone, to move along the cochlea, thus agitating the auditory cilia. Indeed, in certain congenital abnormalities the round window may be missing or rigidly fixed. Not surprisingly, people so affected suffer a huge hearing loss of about 60 dB.

In a healthy cochlea, low frequency endolymph movements do not cause slowly moving liquid to exert sufficient drag on the cilia (it is analogous to the ease of moving a paddle through water sufficiently slowly). This translates to lowering of auditory sensitivity as frequency decreases, shown in **Figure 4**. On the other side of the audible frequency spectrum, high frequency waves are of small amplitude, and, attenuated further by the cilial inertia similarly fail to sufficiently deflect cilia, again resulting in loss of sensitivity as per **Figure 4**.

Second: The round window prevents sound waves from bouncing off the end of the cochlea and traveling back towards the oval window. Should this happen interference or even standing waves could form in the cochlea, resulting in humans hearing subjective sounds in absence of external stimuli (a condition known as *tinnitus*). A healthy cochlea is an echoless chamber.

Following those anatomical hints, we will sample signal in various points along our e-cochlea. This e-cochlea should be echoless, too. If it were infinitely long, it would certainly be echoless, as the signal could never reach its end to bounce back. This is impractical, though.

Figure 9.
Electrical model of a telegraph line or of a “bionic cochlea.”

We can, however, calculate the input impedance of an infinitely long circuit depicted in **Figure 9**. Let that impedance be Z . If we terminate a finite circuit shown in **Figure 9** with a terminator Z (not shown), then viewed from the input side that circuit will be indistinguishable from an infinitely long circuit, and so will become echoless.

Incidentally, nature followed the same process when evolving cochleae. To increase audio acuity of more complex animals, it had to sample signals in many places along a cochlea, so the cochleae of higher animals had to become longer. In order then to fit them into a limited space of a cranial cavity in an extraordinarily hard bone the cochleae had to be coiled. Coiling a cochlea does not affect its acoustic properties much as coiling of a trumpet makes a trumpet smaller but does not affect its sound. With the exception of monotremes, all mammals have coiled cochleae.

We are now in a position to present the fundamental circuitry of our “bionic ear,” as per **Figure 10**. The electric equivalent of an audio signal (for brevity called “audio signal” henceforth) is fed into an e-cochlea made of a large but finite number of segments, and properly terminated by an echoless terminator Z (not shown).

The input audio signal fed to each segment is also input to an analog sampling circuit equipped with a basic A/D converter. The output of this converter is **true** when its input voltage is positive and is **false** otherwise. In this way a pulsating digital signal is obtained. The A/D converter can be constructed from a single transistor oscillating between the cut-off and saturation states or a similar device. It simulates the axon hillock of an auditory neuron.

Given that on short time scales the audio signal is practically symmetric about the timeline, only the signal values exceeding certain positive reference voltages need be sampled. Every reference voltage is determined by the ratio of resistances $R1/R2$. When the voltage of the audio signal exceeds that threshold value, then the diode depicted in the diagram charges the capacitor Cx . As long as this capacitor is sufficiently charged, the A/D converter outputs the value **true**. The resistor Rx slowly discharges the capacitor Cx . A careful choice of the decay time constant $RxCx$, customizable to each patient, controls the minimum duration of each pulse. That time constant describes the dynamics of an auditory cell changing its electrochemical potential.

Figure 10.
Electrical model of a “bionic ear”: each segment of the e-cochlea generates one digital signal.

Similarly, by adjusting values of the threshold ratio $R1/R2$ in each segment we can control the audio sensitivity of our e-cochlea to fit curves shown in **Figure 4**. Further tweaking is possible to customize this circuitry to the sensitivities and preferences of individual patients.

Note the simplicity of this design: Our “bionic ear” consists of a number of repeating segments, each segment yielding one channel of a digital signal. All segments are topologically identical, but may differ in element values R , L , G , and C to span the desired frequency spectrum in suitable number of steps, yielding the required number of channels. The last segment is capped with a suitable echoless terminator Z .

Due to current technological limitations this design probably does not allow us to create a sequence of 30,000 segments today, each segment generating a pulse signal to feed one fiber of the cochlear nerve, but it certainly overcomes current channel limitations. Furthermore, the modularity of the design allows us to build ever better, multi-channel bionic ears as technology improves.

This concludes our qualitative description of bionic ear electronics. To demonstrate that our device is capable of generating signals useful for sophisticated auditory perception, we need to introduce several mathematical concepts. We will also shed light on the way brain nuclei higher up the auditory pathway operate, with emphasis on audio pattern recognition and classification.

5. Variation on the theme of metric spaces

Mathematicians conceive a metric space as a set of objects, usually called “points,” between which a way of measuring the distance has been defined. In everyday life we use Euclidean distance. However, this is only one of the many possible ways of measuring distances.

Let S be our space of interest.

Definition 1. Measure of distance: Any real-valued function $d: S \times S \rightarrow \mathbf{R}$ can be used as a measure of distance, provided that for all $x, y, z \in S$ it has the following properties:

$$d(x, x) = 0 \quad (12)$$

$$\text{if } x \neq y \text{ then } d(x, y) > 0,$$

$$\text{i.e. the distance between two distinct points cannot be zero;} \quad (13)$$

$$d(x, y) = d(y, x), \text{ i.e. the distance does not depend on the direction of measurement;} \quad (14)$$

$$d(x, z) \leq d(x, y) + d(y, z), \text{ i.e. the distance cannot be diminished by measuring it via some cleverly chosen intermediate point } y. \quad (15)$$

The last property is frequently called the *triangle law*, because the length of each side of a triangle cannot be greater than the sum of lengths of two other sides.

The properties (12) and (13) imply that the distance between any two points cannot be negative and is actually positive if the points are different. In fact, we can state.

Theorem 1: Given any real-valued function $d: S \times S \rightarrow \mathbf{R}$ suitable for measuring-distances between points in S , the equality $d(x, y) = 0$ for any $x, y \in S$ implies $x = y$.

Proof: The equality $d(x, y) = 0$ means that a shortest travel from x to y does not make us cover any distance, therefore x and y must coincide, that is, $x = y$. ■

As examples consider a two-dimensional space $\mathbf{R}^2$ and two points $a = \langle x_1, y_1 \rangle$ and $b = \langle x_2, y_2 \rangle$. We may define distance $d(a, b)$ as:

Euclidean distance : $d_1(a, b) = \text{sqrt} \left((x_1 - x_2)^2 + (y_1 - y_2)^2 \right)$ or (16)

Manhattan distance : $d_2(a, b) = |x_1 - x_2| + |y_1 - y_2|$ or (17)

perhaps simply as : $d_3(a, b) = \max \{ |x_1 - x_2|, |y_1 - y_2| \}$ (18)

All three functions d_1 , d_2 , and d_3 meet the necessary requirements (12), (13), (14) and (15).

Definition 2. A **metric space** is the configuration $\langle S, d \rangle$, where S is a set of points, and d is some measure of distance between them.

In our study we are frequently interested in measuring distance between subsets of S , rather than only between points of S . One of the simplest subsets of S is a *ball*.

Definition 3. A **ball** of radius $r \geq 0$ around the point $c \in S$ is the set

$$\{x \in S | d(c, x) \leq r\}$$
 (19)

which we will denote as $\mathbf{B}(c, r)$ without mentioning either S or d when confusion can be avoided. The point c is called the center of the ball $\mathbf{B}$.

Note: In older math textbooks the term “sphere” is frequently used instead of a “ball.” This usage is currently being phased out, as we prefer now the “sphere” to mean the surface of a “ball.”

Observe that the “shape” of a ball depends on the way we measure distance, as per **Figure 11**.

It is worth noticing that the “volumes” of such balls may depend on the metrics used. In particular, the Manhattan ball d_2 is most specific about its center, that is, it has the smallest “volume” (i.e., area in $\mathbf{R}^2$), and also its metric function (17) computes faster than metrics (16) and (18). This is important for some of pattern recognition algorithms including visual and auditory perception, which, although fast and massively parallel, remain computationally intensive [2, 3].

A more detailed analysis of metric spaces can be found at [11].

Figure 11. The balls in $\mathbf{R}^2$ of same radius, with center in the origin of the system of coordinates, defined using metrics d_1 , d_2 , and d_3 in the previous example. Similar balls in $\mathbf{R}^3$ would be a sphere, an octahedron, and a cube, correspondingly. Generalizations of balls in higher-dimensional spaces are conceptually straightforward, although not easily drawn.

6. Computing distances between sets

Given any two non-empty sets $A, B \subset S$ we need to construct a function $D(A, B)$ to measure distance between A and B . That function should retain the properties (12), (13), (14) and (15). In particular, observe that the properties (12) and (13) imply that $D(A, B) > 0$ even if the sets A, B touch (i.e., have one common element), or intersect, or perhaps one of them contains another ($A \subset B$ or $B \subset A$). In fact, we must construct function D such that $D(A, B) = 0$ if and only if $A = B$.

Let d be our chosen function for measuring distance between points of space S . We will use this function to construct our function D .

Definition 4. Distance between a point and a set: Let $\langle S, d \rangle$ be a metric space and let $A \subset S$ be a non-empty set. A distance between a point $x \in S$ and A , denoted $\delta(x, A)$ is given by

$$\delta(x, A) = \inf \{d(x, a) | a \in A\} \quad (20)$$

It is the distance between x and a point $a \in A$ closest to x . Using δ , let us define.

Definition 5. Pseudo-distance between two sets: Let $\langle S, d \rangle$ be a metric space and let $A, B \subset S$ be two non-empty sets. A pseudo-distance from A to B , denoted $\Delta(A, B)$ is given by

$$\Delta(A, B) = \sup \{\delta(a, B) | a \in A\} \quad (21)$$

In other words, pseudo-distance from A and B is the distance from the most distant point $a \in A$ to its closest point $b \in B$. It is not distance, but merely pseudo-distance, because it is unidirectional, that is, the property (14) does not hold, given that $\Delta(A, B) \neq \Delta(B, A)$ in general. To make it bidirectional, we define.

Definition 6. Distance between two sets: Let $\langle S, d \rangle$ be a metric space and let $A, B \subset S$ be two non-empty sets. A distance between A and B , denoted $D(A, B)$ is given by

$$D(A, B) = \max \{\Delta(A, B), \Delta(B, A)\} \quad (22)$$

We are ready now to demonstrate that our construction of $D(A, B)$ has properties of a metric.

Theorem 2: Our newly constructed function D has the properties (12), (13), (14) and (15), and therefore can be used as a measure of distances between subsets of S .

Proof: Let $\langle S, d \rangle$ be a metric space and let $A, B, C \subset S$ be non-empty sets. We have

1. **Property (12) a.k.a. reflexivity:** Let us compute $D(A, A)$. Let us choose an arbitrary point $a \in A$. Using Definition 4 (see (20)) we obtain $\delta(a, A) = 0$, because the closest target in A from a is a itself, as it belongs to A . From that it must follow that $\Delta(A, A) = 0$ because the largest of all zeros is zero (peruse (21)). This implies further that $D(A, A) = 0$ because greater of two zeros is still a zero (see again (22)).
2. **Property (13):** Let $A \neq B$. Both sets being not empty, we can find an $a \in A$ such that $a \notin B$, or $b \in B$ such that $b \notin A$, or both (if this were not the case, then the sets A and B would be equal). Given that at least one of a and b does not belong to both sets, we must have either $\delta(a, B) > 0$ or $\delta(b, A) > 0$ or both. From this we have either $\Delta(A, B) > 0$ or $\Delta(B, A) > 0$ or both, and so $D(A, B) > 0$ as being greater of the previous two.

3. Property (14) a.k.a. commutativity follows directly from the Definition 6: $D(A, B) = \max \{\Delta(A, B), \Delta(B, A)\}$, so the order of arguments does not matter when evaluating $D(A, B)$.

4. Property (15) a.k.a. triangle rule: We need to show that $D(A, B) \leq D(A, C) + D(C, B)$.

A comment on travel from A to B via C : not any path will do. It is not sufficient to reach C from A at particular point $c_1 \in C$ and then continue the trip from another point $c_2 \in C$ to reach B to claim that we have traveled through C . Our itinerary must form a continuous path, that is, its first leg must end up at some point $c \in C$ exactly where the second leg begins.

Let then the points $a \in A$ and $b \in B$ be arbitrarily chosen departure and destination points such that $d(a, b) = D(A, B)$. Depending on the metric chosen the path from a to b of length $d(a, b)$ may not be unique. However, each of such paths may or may not pass through our intermediate point $c \in C$. If it does, then we do not modify that path, and so its length does not change. If it does not, then we need to modify it to pass through c . In that situation the length of that path can only increase. Given that $D(A, B)$ is computed using carefully chosen paths leading from $a \in A$ to $b \in B$ (or vice versa), such modifications of these paths can only result in leaving them intact or lengthening them, thus proving that $D(A, B) \leq D(A, C) + D(C, B)$. ■

Note that it also follows from this theorem that for our function D the Theorem 1 holds as well.

More importantly for the purposes of our chapter, observe that we may move the sets A and B within their space S so as to minimize $D(A, B)$. Even then, for A and B being different, the residual value $D(A, B) > 0$ will remain.

With the information regarding the spatial positioning of sets A and B deliberately destroyed through such preprocessing, that residual value $D(A, B)$ can be seen as measure of dissimilarity between sets. The function D can therefore serve also as a pattern classifier. With properly selected small value of $\varepsilon > 0$, the condition $D(A, B) \leq \varepsilon$ implies that A and B so preprocessed are sufficiently similar to be included in the same category.

7. Comparisons of audio patterns

We concern ourselves now with issues of pattern recognition, which in auditory domain call for comparison of audio signals. An audio signal is really a bundle of rectangular pulse signals traveling through fibers of a cochlear nerve. We will take full advantage of their properties arising from the fact that these elementary signals belong to time domain, that is, $\mathbf{R}^1$ space.

An audio pattern is an excerpt of an audio signal taken over a suitable time interval. Two audio patterns are similar, if all their elementary corresponding pulse signals are similar, that is, differ at most by an arbitrarily small time value of $\varepsilon > 0$.

7.1 Properties of elementary audio patterns

Our function D , although mathematically correct and theoretically useful in measuring distances and dissimilarities between sets, is only computable for sets

belonging to $\mathbf{R}^1$ space (time domain). Readers interested in circumventing this limitation are directed elsewhere [2, 3].

Consider now two corresponding elementary signals, shown in **Figure 12**, as red and blue, for clarity. They are excerpted from two different signal bundles. Clearly, they are not identical. Are they then completely different and unrelated, or is one a subtle plagiarism on the theme of another? Without any numerical measure of their dissimilarity, any two experts may have three different opinions on this matter. We will use our function D as a reconciliation tool.

For the purposes of comparison two signal bundles must be aligned, that is, both must start at the same time, which we will denote as time zero. This means that in each signal bundle there is at least one elementary signal starting at time zero.

However, not all elementary signals in that bundle need to start simultaneously. Indeed, their relative timing is one of the intrinsic characteristics of each signal bundle. We can now define:

Definition 7: The dissimilarity between two signal bundles is the maximum dissimilarity between two corresponding elementary pulse signals in each signal bundle.

We therefore turn our attention to comparisons of elementary pulse signals.

Let us compare first signals consisting of one pulse each. **Figure 13** shows all possible relationships between such two signals. We will be calculating the value of $\Delta(A, B)$, being the fundamental component of $D(A, B)$. In other words, we will be measuring pseudo-distance from A to B (the pseudo-distance in the inverse direction is measured in identical way). For clarity, in our graphs A will be shown in red, B in blue.

Furthermore, we adopt the convention that the value t_{A0} and t_{A1} are the start and the stop times of pulse A , and the values t_{B0} and t_{B1} are accordingly the start and stop times of pulse B .

Translating the formula (21) into plain multi-disciplinary English we write:

Figure 12.
Two pulse signals: are they really different? Unrelated? Or is there a similarity?

Figure 13.
Possible relationships between two pulse signals.

Pseudo-distance $\Delta(A, B)$ is the distance between the most distant point of A to its closest point in B.

This value describes the worst-case scenario when traveling from A to B in terms of the length of the shortest trip.

We are now in position to compute $\Delta(A, B)$ for all variants shown in **Figure 13**. We have:

Variant 1 : $\Delta(A, B) = |t_{A1} - t_{B1}|$, Variant 2 : $\Delta(A, B) = |t_{A0} - t_{B0}|$,
 Variant 3 : $\Delta(A, B) = |t_{A1} - t_{B1}|$, Variant 4 : $\Delta(A, B) = |t_{A0} - t_{B0}|$,
 Variant 5 : $\Delta(A, B) = \max\{|t_{A0} - t_{B0}|, |t_{A1} - t_{B1}|\}$, Variant 6 : $\Delta(A, B) = 0$ because $A \subset B$.

These variants can be easily identified and pseudo-distances computed, even by simple systems of several neurons. Observe symmetries in these expressions. On any parallel computer both values $|t_{A0} - t_{B0}|$ and $|t_{A1} - t_{B1}|$ can be calculated simultaneously, and then the proper value can be used as needed. Needless to say, we treat the brain as a parallel computer.

Do we have all the tools needed to compute distances between elementary pulse signals? Not yet. Not all elementary pulse signals consist of a single pulse each. In certain situations, two pulses in B may be needed to measure $\Delta(A, B)$. **Figure 14** depicts some of those cases. Signal A consists of one pulse as before, but signal B is now made out of two pulses. To cover these situations, we need to enhance our notation of pulse timing.

An elementary signal X consisting of n pulses we describe now as a tuple of $2n$ time values $\langle t_{X0,0}, t_{X0,1}, t_{X1,0}, t_{X1,1}, \dots, t_{Xn-1,0}, t_{Xn-1,1} \rangle$. We number pulses in signal X starting at 0 and ending at $n - 1$. In that sense the notations $t_{X4,0}$ and $t_{X4,1}$ mean start and stop times of the fifth pulse of signal X (the fifth pulse has a number 4 because computer programmers like it that way!).

Figure 14.
 Possible relationships between one-pulse signal A and two-pulse signal B.

Observe now what happens in variant 7 of **Figure 14**: the point in A most distant from B is that of timing in the middle of the gap between two pulses of B. We have here

$$\Delta(A, B) = |t_{B0,1} - t_{B1,0}|/2$$

More importantly, we have arrived at the following result:

Conclusion 1: When measuring pseudo-distances $\Delta(A, B)$ the points in A (if any) that are in the middle of gaps in B (if any) must also be considered, as well as the start and end points of A. Still, only the end points in B are of interest.

Therefore, in variant 8 the value $\Delta(A, B)$ must be calculated as

$$\Delta(A, B) = \max \{ |t_{A0,0} - t_{B0,0}|, |t_{B0,1} - t_{B1,0}|/2, |t_{A0,1} - t_{B1,1}| \}$$

Consider now the difference between variants 11 and 12. In variant 11 the pulse A extends beyond the middle of the gap between pulses in B, while in variant 12 it does not. Therefore, when calculating $\Delta(A, B)$ for variant 11, the value of the expression $|t_{B0,1} - t_{B1,0}|/2$ must be considered, while in variant 12 the existence of the second pulse can be ignored when calculating $\Delta(A, B)$ (although it is still relevant for calculation $\Delta(B, A)$). Indeed, should there be more pulses in B following the second pulse of B (variant 12), they all could be ignored when calculating $\Delta(A, B)$. Similar situation can arise when a multi-pulse signal B is timed so that the leading pulses of B can be ignored. This is good news, as this situation leads to vastly simplified neural calculations. Simplicity bestows speed; for survival speed is of essence!

Note further that the calculated values of $\Delta(A, B)$ depend on whether or not they are arrived at in real time. As an example, consider the situation depicted in variant 13 shown in **Figure 14**. According to our analysis, the pseudo-distance $\Delta(A, B)$ value is $\Delta(A, B) = |t_{B0,1} - t_{B1,0}|/2$. This is established knowing *a priori* that signal B consists of two pulses. If, however, we do the comparison between A and B in real time and current time is 12 units, then we are in the situation where the first pulse of B already ended but the second pulse did not begin. We cannot even know whether the second pulse exists. We have no choice but to keep calculating the distance between that part of A which we have experienced already and the only part of the pulse of B we heard so far. This will overestimate the value of $\Delta(A, B)$. That value will drop off immediately when we hear the onset of the second pulse of B. In fact, we have established now the following:

Conclusion 2: The calculated values of $\Delta(A, B)$ are a function of time. This function is continuous when we have an *a priori* knowledge of all pulses in A and B (i.e., we compare patterns A and B previously stored in memory), but only piecewise continuous if we calculate $\Delta(A, B)$ in real-time (i.e., without such *a priori* knowledge).

In short: When computing $\Delta(A, B)$ in real time, the number of pulses of B to consider is further limited. Only the pulses experienced so far can be considered. This simplification further enhances the speed of calculation thus increasing our chances of survival. The fact that our calculation is less exact is immaterial; our foe suffers from the same handicap.

If we survive the encounter with our foe, when safe we can replay both patterns from memory, refining our calculation and survival strategy. When at leisure we have more time to safely indulge in more complex calculations, refining our forecasting models.

This also explains why we enjoy the same recording of music differently depending on whether we hear it for the first time or not.

A question then arises: Suppose we enjoy a piece of music we hear for the first time. If that pleasure is tied to a certain comparison process, what are we comparing that piece to?

Conclusion 3: When listening to a piece of music (even for the first time) we keep predicting its future passages, on the basis of most recent passages in a certain active time interval. In that sense music must adhere to a certain “grammar” to be found pleasurable.

We are now ready to create a formalism of comparison of two elementary pulse signals.

7.2 Comparisons of elementary audio patterns

Given two multi-pulse elementary patterns A and B, we need to calculate their dissimilarity as a distance $D(A, B) = \max \{ \Delta(A, B), \Delta(B, A) \}$ as per formula (22). Again, we focus our attention at calculating the value $\Delta(A, B)$ (The value $\Delta(B, A)$ is computed in the same way).

Let A consist of n pulses, while B consists of m pulses. We can represent A and B as follows:

$$A = \bigcup_{i=0}^{n-1} A_i \quad \text{and} \quad B = \bigcup_{j=0}^{m-1} B_j \quad (23)$$

where A_i and B_j are single pulses.

The value of $\Delta(A, B)$ can now be calculated as:

$$\Delta(A, B) = \max \{ \Delta(A_i, B) | i = 0..n \} \quad (24)$$

In fact, we can accelerate this computation. We have already observed that when calculating $\Delta(A, B)$, not all pulses of B are of interest as possible closest destinations on our travel from A_i to B. We can calculate $\Delta(A, B)$ more efficiently as

$$D(A, B) = \max \{ \Delta(A_i, B^*) | i = 0..n \} \quad (25)$$

where B^* is a subset of consecutive pulses of B, namely

$$B^* = \bigcup_{j=k(i)}^{l(i)} B_j \quad (26)$$

where $k(i)$ and $l(i)$ are the first and last pulses of this consecutive sub-sequence of interest when calculating the pseudo-distance from A_i to B.

To give formal justification to this computational short-cut, we need to establish the following:

Theorem 3: Let A and B be two multi-pulse elementary patterns defined as per (23). When calculating the values of $\Delta(A_i, B)$, not all pulses of B are relevant. We can limit our attention to a subset of consecutive pulses of B, namely

$$B^* = \bigcup_{j=k(i)}^{l(i)} B_j$$

where k , which depends on i is an index of the first pulse of interest, and l , which also depends on i , is an index of the last pulse of interest.

Proof: Consider an arbitrarily chosen single pulse $A_i \subseteq A$. The starting moment of this pulse, timed at $t_{A_i,0}$ may or may not coincide with a pulse of B. If it does, then that pulse of B is the first pulse of interest. We call it B_k . The pulses preceding B_k (if any) cannot be the target destinations on our journey from A_i to B, because at the moment $t_{A_i,0}$ we are already in B.

If it does not, then the most recent pulse of B preceding the moment $t_{A_i,0}$ is our first pulse of interest. Again, we call it B_k . As before, the pulses preceding B_k (if any) cannot be the target destinations on our journey from A_i to B, because the trip from A_i to B_k is shorter.

In a special case that there is no pulse of B preceding the moment $t_{A_i,0}$ the first pulse of B is B_k .

To identify the last pulse of interest B_l we proceed similarly. The ending moment of pulse A_i , timed at $t_{A_i,1}$ may or may not coincide with a pulse of B.

If it does, then that pulse of B is the last pulse of interest. We call it B_l . The pulses following B_l (if any) cannot be the target destinations on our journey from A_i to B, because at the moment $t_{A_i,1}$ we are already in B.

If it does not, then the most imminent pulse of B following the moment $t_{A_i,1}$ is our last pulse of interest. Again, we call it B_l . As before, the pulses following B_l (if any) cannot be the target destinations on our journey from A_i to B, because the trip from A_i to B_l is shorter.

In a special case that there is no pulse of B following the moment $t_{A_i,1}$ the last pulse of B is B_l . ■

In plain multi-disciplinary English, all this means: it is possible to compare two elementary pulse signals by scanning them simultaneously and chronologically, even in real time. In this situation each signal we need to place in a moving time window, containing relevant pulses only, and keep advancing that window as time goes on.

As said before, every audio signal is a bundle of elementary signals, and therefore comparison of two audio signals also involves usage two advancing time windows. This is not new: we do that intuitively all the time when listening to music. We keep certain passages in a moving time window (i.e., some form of active memory). We consider these passages “current” and give them special attention. We feel pleasure if they continue to remain within a certain anticipated “grammar.” Older passages are kept in another, longer term memory.

We have been listening to and enjoying music since time immemorial. Until now it was just a beautiful emotion. Now we have a mathematical model of that emotion.

8. Final thoughts and recommendations

We are well on our way in gaining insight how our brains work. They are pattern comparison machines. When comparing complex patterns (visual, auditory, gustatory, etc.) complex patterns are decomposed into arrays of simpler patterns on which the comparisons can easily be made. These simple comparisons are made in real time and their results are then consolidated into a general conclusion about the complex patterns.

Our analysis of audio signal comparison processes makes us further deduce that:

- Two audio signal bundles must consist of the same number of elementary pulse signals to be comparable, and these elementary signals must be compared pairwise;

- When an elementary signal in one bundle ceases to exist (due to an accident or a disease), the corresponding pulse signal in another bundle is not fully useful to the brain;
- Implanting two incompatible cochlear prostheses in one patient is a grave error in the art, as they generate incompatible thus incomparable bundle signals;
- Implanting only one prosthesis is therefore an equally grave blunder, unless it is fully compatible with a healthy cochlea;
- Cochlear prostheses should therefore be implanted in pairs and tuned simultaneously to each patient. Such tuning should be an inherent part of the implantation procedure.

In the audio domain two patterns consisting of a different number of elementary patterns cannot be similar—they are intrinsically of a different kind. They cannot be readily compared. Children frequently ask a vexing question: “Daddy, when both of us listen to a piece of music we like, do we feel exactly the same sense of pleasure?” The correct answer is: these feelings are subjective, unknowable to others; they arise in our brains as responses to our individual audio signals. It has been said that each human has approx. 30,000 fibers in the cochlear nerve, but the actual numbers vary among the individuals. Also vary the numbers of cilia, geometries of cochleae, ossicles, etc. No two individuals are alike, and their audio signals are mutually incompatible. So are their emerging feelings of pleasure.

No wonder that the challenges in constructing cochlear prostheses are daunting! No single cochlear prosthesis will fit everybody. It must be customized to an individual.

In the current practice things are done the other way around. Every implant manufacturer swears that his device is “best” but dares to implant it into selected patients only, counting on the plasticity of their brains to adapt to the device. To makes things worse, we tend to implant only one device per person—to avoid “complications.” What complications? In fact, we introduce here unnecessary complications—humans (and other animals) have evolved to have two very similar ears. To ask of the human brain to adapt within a single lifespan to two different and mutually incompatible “ears,” thus generating signals not readily comparable is to ask too much! Practical problems emerge, viz.: how to perform audio location if you have two different hearing organs generating incompatible signals? No wonder that the current implant success rate leaves much to be desired.

Our approach offers new promises, but first and foremost we should customize the cochlear prostheses to our patients, not the other way around. We are well on our way.

Illustration credits

Figure 1. Chittka L. Brockmann: Anatomy of the human ear. Wikimedia Commons. Available from: http://en.wikipedia.org/wiki/File:Anatomy_of_the_Human_Ear.svg [Accessed: 13 June 2010], Creative Commons Attribution 2.5 GenericLicense.

Figure 2. Cochlear cross-section. Wikimedia Commons. Available from: <http://en.wikipedia.org/wiki/File:Cochlea-crosssection.svg> [Accessed: 13 June 2010], GNU Free Documentation license v. 1.2.

Figure 4. Contours of equal loudness. ISO 226:2003 revision.

Figure 7. Layout of the cochlear nerve. Wikipedia. Available from: <http://upload.wikimedia.org/wikipedia/commons/b/bc/Gray928.png> [Accessed: 17 July 2010], Public domain.

Acknowledgements

The author would like to thank his brother, Dr. W. Gregory Wojcik, MD, for his patient verbal advice and consultations on matters of biology and anatomy.

Author details

Włodzimierz (“Vlad”) Wojcik
Department of Computer Science, Brock University, St. Catharines, Ontario, Canada

*Address all correspondence to: vwojcik@brocku.ca

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- [1] Shannon RV, Zeng FG, Wygonski J, Kamath V, Ekelid M. Acoustic simulations of cochlear implants. Available from: <http://www.healthaffairs.uci.edu/hesp/Simulations/simulationsmain.htm> [Accessed: September 2010]
- [2] Wojcik V, Comte P. Algorithms for speedy visual recognition and classification of patterns formed on rectangular imaging sensors. *Neurocomputing*. 2010;**74**(1–3): 140-154. DOI: 10.1016/j.neucom.2009.10.029
- [3] Wojcik V, Comte P. Algorithms for speedy visual recognition and classification of patterns formed on rectangular imaging sensors. Technical Report: #CS-08-07. Department of Computer Science, Brock University. Available from: <http://www.cosc.brocku.ca/files/downloads/research/cs0807.pdf> [Accessed: July 2007]
- [4] Della Santina CC. Regaining balance with bionic ears. *Scientific American*. [Accessed: April 2010]
- [5] Crawford J. Living without a balancing mechanism. *British Journal of Ophthalmology*. 1964;**48**(7):357-360
- [6] Minor LB. Gentamicin-induced bilateral vestibular hypofunction. *Journal of the American Medical Association*. 1998;**279**(7):541-544
- [7] Della Santina CC et al. A multichannel semicircular canal neural prosthesis using electrical stimulation to restore 3-D vestibular sensation. *IEEE Transactions on Biomedical Engineering*. 2007;**54**(6):1016-1030
- [8] Standring S et al. *Gray's Anatomy: The Anatomical Basis of Clinical Practice*. 40th ed. New York, NY: Elsevier; 2009
- [9] Shannon-Hartley Theorem. Wikipedia. Available from: http://en.wikipedia.org/wiki/Shannon-Hartley_theorem [Accessed: September 2010]
- [10] Gokhale AA. *Introduction to Telecommunications*. 2nd ed. Florence, KY: Thomson Delmar Learning; 2004
- [11] Gleason AM. *Elements of Abstract Analysis*. Boston: Jones and Bartlett Publishers; 1991