A Paradigm Shift for Modelling Sound Sensation

John N. Mourjopoulos
Audio and Acoustic Technology Group
Electrical & Computer Engineering Dept.
University of Patras, 26500 Patras, Greece
mourjop@upatras.gr

ABSTRACT

How do we describe the exemplary acuity of humans to analyze and respond to sounds, particularly to music? Is our current knowledge sufficient to produce working computational models for such sensations? Has our perspective for the conceptual structure of such models changed and why is this important?

This work attempts to provide brief answers to these questions, focusing on a recent comprehensive model of binaural listening which is directed towards engineering applications in audio and acoustics [1]. It is also discussed how such model can formally approach the concepts of quality and fidelity in sounds and how it may be employed to demystify experienced listener and audiophile perception. A brief discussion of the conceptual and philosophical implications of such a model is also given.

1. INTRODUCTION

The auditory system has amazing capability in performing a wide range of functions. For a long time this performance challenges established knowledge in physics, psychophysics and cognitive science. In his unfinished 1866 essay "The Mechanism of the Ear", the great mathematician Bernhard Riemann wrote that one of the more intriguing problems to be understood is the ear's ability to perceive sound waves, the energy levels and physical displacements of which are so small that they cannot even be measured directly [2]. It is now known that the human ear's sensitivity is such that it can detect motions of atomic dimensions, that is, at the threshold level of 0 dB SPL, hearing can perceive tones generated by air particle displacement of approx. 1/10 the diameter of hydrogen molecule in which the energy transmission to the eardrum is on the order of 10^{-18} Joules. Nevertheless, to start explaining this hearing paradox, Riemann introduced the concept of analogies derived from an analytic abstraction of audition − a concept that, as it will be shown, is currently adopted by state-of-the-art computational auditory models, nearly 150 years after Riemann's death whilst he was working for this essay.

AABBA (Aural Assessment by Means of Binaural Algorithms), is a benchmark grouping of researchers, active in auditory modelling, with results recorded in a recent collective volume [1]. AABBA is revisited here in order to evaluate the principles behind current understanding of the sensations of sounds, at least as far as this can be implemented via computational models. For this, we shall review the methods and their underlying conceptual and philosophical implications for such auditory models, particularly, considering the principles behind a recent paradigm shift in their structure.

Over many decades, such hearing models have evolved facing the challenge of how to accommodate via analogies the extreme processing capabilities of the ear mechanism and the still largely unknown processes of brain cognition. Up to date, auditory modelling has been mostly based on the synthetic principle of interconnecting modules forming the auditory architecture, first described by Helmholtz [3], a contemporary of Riemann.

More recently, digital signal processing has allowed efficient modelling of signals, stimuli and transformations in a bottom-up abstraction of this synthetic auditory architecture. Significantly, the *Auditory Scene Analysis (ASA)* approach introduced by Bregman in the '90s [22] and its extensions into machine perception (*Computational Auditory Scene Analysis, CASA)* has proposed models of perception for complex auditory tasks via bottom-up and top-down concepts. Nevertheless, as will be shown there is a need to revise and enhance such traditional modelling approach in order to accommodate mechanisms and operations which cannot be accounted for by the synthetic approach.

The current article will examine the emergence, potential applications and consequences of recent proposals for a model based on the analytic abstraction, proposed by Blauert et al. [4], which can also address issues of auditory cognition. Any such computational analogy of cognitive analysis for the aural scenery must at first establish the primal reasons of humans to employ such a sensory channel. According to Blauert et al. [1], these are:

- Listening for awareness about the environment
- Listening for communication purposes
- Listening for pleasure

Here, we shall mostly consider auditory modelling for the pleasure-listening scenario. The paper is organised as following: Section 2 considers some of the current challenges facing binaural auditory modelling. Section 3 analyses the proposed structure for such models, which introduces the computational model of higher level perception. Section 4 discusses the conceptual implications from these recent developments and draws some final conclusions.

2. A PARADOX AND A MISSING LINK

2.1 Uncertainty and hearing

A recent publication [5] presents results indicating that the human hearing performance is superior to the absolute theoretical time-frequency accuracy limit for signals, as given by the Fourier uncertainty principle:

$$\Delta f \cdot \Delta t \geq 1/4\pi \quad (1)$$

In practice, this limit states that for signals such as sounds, the accurate observation for short signal durations results in low accuracy for the definition of frequencies representing their spectrum. Conversely, sound signals with well-defined frequency spectrum require observation over longer durations, hence being inaccurately defined in time. Therefore, the uncertainty principle imposes the absolute theoretical limit for the precision of the simultaneous physical measurement or observation of the duration and frequency of any signal such as the acoustic waves. The comparable Heisenberg uncertainty principle for the trade-off between momentum and position in particles applies to quantum mechanics and relates to phenomena having many orders of magnitude lower energy and higher frequency than audible sounds.
During the tests reported in [5], 12 human assessors were evaluated on how well they could simultaneously identify the duration and frequency of a sound via a series of 2−down−1−up listening procedures. They were asked to discriminate simultaneously whether a test note, having either Gaussian or transient-like amplitude envelope, was higher or lower in frequency than a leading note that was played before it, and whether the test note appeared before or after a third note, which was discernible due to its much higher frequency. The top score, achieved by a professional musician, violated the uncertainty principle of eq.(1) by a factor of about 13, displaying equally high precision in frequency and timing acuity. The score with the top timing acuity of 3 ms was achieved by an electronic musician working in precision sound editing. Such performance of the auditory cognition is superior to most known systems and can only result from processing via a non-linear (or under certain conditions, a chaotic) system. This result can be partially attributed to the highly non-linear properties of the auditory periphery. There is little evidence for other known non-linear systems that can even reach the uncertainty limit without introducing distortions that hinder observed responses. In contrast, audition acuity seems to be enhanced by such non-linear signal processing. However, such impressive performance in time-frequency identification may be also related to higher level cognitive auditory functions or to a yet unknown principle of combining lower auditory periphery processing with higher level cognitive adaptation. Note that the enhanced time-frequency acuity and, specifically, the performance accuracy in temporal detection tasks can be associated with recent findings relating to the selective temporal processing functionality of the neural transduction mechanisms associated with the onset neurons located at the cochlear nucleus [6].

However, the study of models that convert the signals into binary nerve action potentials (spike trains from the auditory nerve fibers, ANFs), presents significant difficulties since these must be treated as stochastic processes and require laborious statistical analysis. The principle of coding sound signals into spike trains has as yet an only partial systematic mathematical description framework [4,7] and, hence, traditional signal processing methods cannot easily be applied. Amongst other aspects, research in signal processing analysis of neural coding attempts to describe:

(a) Neuron identification, i.e. to identify neurons that encode certain signal features of interest, and
(b) Neural encoding, i.e. to establish functional relationship between feature and spike trains of identified neurons [7].

Here it must be noted that binaural audition apart from enabling spatial detection and localisation, provides significant advantages for most listening tasks compared to the monaural case, especially under adverse acoustic conditions. At the auditory-periphery level, there are established signal-processing models and mapping operations which can represent the binaural activity resulting from the combination of the separate monaural signals travelling through the Auditory Nerves (ANs). The binaural co-processing and encoding is realised at the Superior Olivary complex (SO) and the Inferior Colliculus (IC) of the binaural auditory architecture, as is shown in **Fig. 1**. The auditory signals activate the primary Auditory Cortex (CX) areas and ultimately other areas in the Brain Lobes (BL). Note that, as was previously explained, most such models rely on deterministic signal processing, omitting the binary coding into spike trains. This applies to the majority of models described in Section 3.

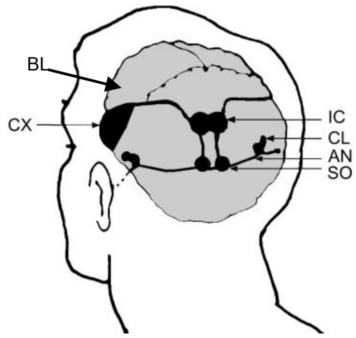

Figure 1. Schematic of the binaural system showing the relevant components (from [1], reprinted by permission)

2.2 A missing link in sound activation maps

Today, it is indeed widely accepted that most neurophysiological mechanisms beyond the auditory periphery are not properly understood. The binaural auditory system consists of pathways of connected modules, nerves and interfaces, where transductions, transformations and local processing are applied to the stimulus signal before main processing is undertaken higher up this path, namely, at the cortical level. Although the first synthetic view of this complex system unifying medical science, physiology, anatomy, physics and music theory was developed by Helmholtz more than 150 years ago [3], until recently, there is a lack of integrating framework for the analysis and computational modelling of audition from stimulus up to perception stage. It is accepted that the current integrative paradigms that simplify the auditory system's complexity (as shown in **Fig.1**) into manageable computational modules[1], follow the general bottom-up architecture shown in **Fig. 2** [8].

This paradigm illustrates the established view of the auditory perception including some anatomical abstractions for the various subsystems involved, that is, decomposition of various phases in the transduction of auditory information from the periphery to the brain, some simple description of the organization of neural information along the auditory pathway including a final stage which attempts to include perceptual inference in order to combine the psychophysical and the physiological knowledge [8].

Based on such bottom-up, signal-driven paradigm, for some time now, binaural activity maps have been produced via signal processing models, displaying the combined effect of the two ear cues (ILD, ITD, IACC, etc), which were successfully employed for many applications and for modelling localization, detection and interpretation of acoustics spaces [1].

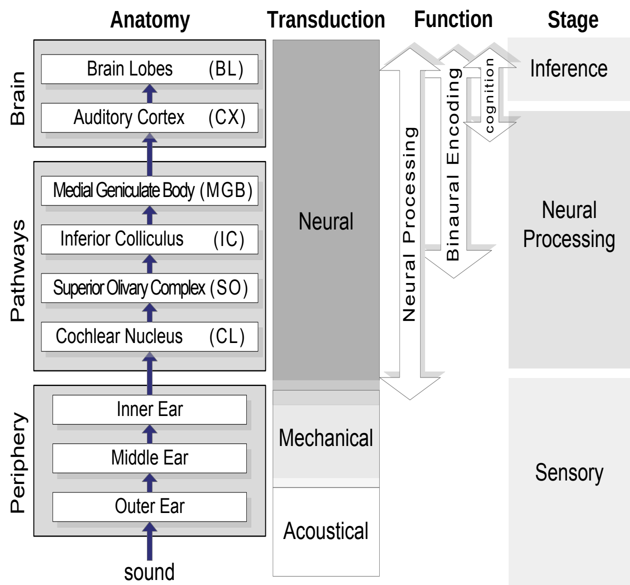

Figure 2. Decomposition paradigm of the various mechanisms of auditory perception (adapted from [8]).

In another parallel development, current brain-imaging technology via functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET) has provided displays for activation patterns of the human brain during auditory tasks [8]. The possibility of combining brain-imaging with the above model and with auditory-nerve level binaural maps, appears very promising: However the relationship between macroscopic observations of the regional cerebral activities displayed via fMRI and the signal analogues of the physiological excitatory or inhibitory responses at the cellular level is still far from being understood.

The missing link here results from the difficulty in assigning specific regional brain patches to cognitive auditory tasks especially for complex stimuli (such as speech and music). Such tasks appear to activate multiple regions in the brain beyond the known neuron-anatomical and topographically specialized regions for the auditory function. This is so possibly because auditory perception involves active cognitive functions such as association with previous experience (memory and learning) and depends on the adaptation to the significance of the observed auditory event in relationship to expected and known percepts. Hence, memory, learning and anticipation appear to introduce a dynamic and time-varying environment that restructures the interconnection of the brain's processing units, leading to currently unknown organisation and mapping beyond the obvious static, hierarchical, and sensory-specific topology. It is also evident that besides the specific problem of auditory perception, a still unknown brain organisation model based on ontology may hold the key to a better understanding of all perceptual processes [8].

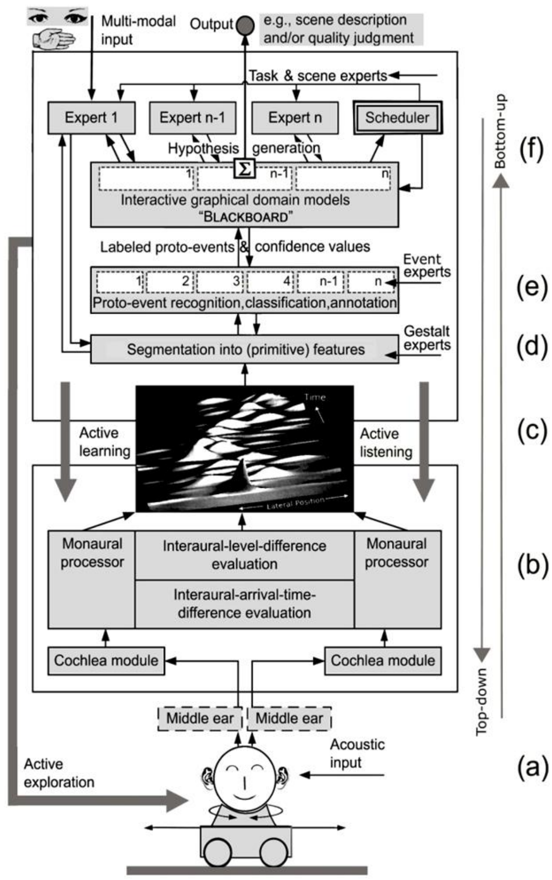

Figure 3. Schematic of the architecture of the comprehensive model of binaural listening proposed by Blauert et al. (from [4], reprinted by permission).

[1] It is referred to as "constructionist's" model in [8].

3. THE PARADIGM SHIFT

3.1 A model for binaural listening and perception

To bridge the missing link between the perceived sound and the general synthetic representation requires a modification of the established methodology that considers it as a set of interacting components.

This can be realised via an additional level of abstraction: one that considers the system as a functioning unit [2]. This approach was also proposed by Riemann and introduces a level of abstraction which describes the auditory system as a black box and studies its overall behaviour − i.e. what it does or accomplishes.

The conceptual implications of such a paradigm shift will be further discussed in Section 4 and has been recently addressed by Blauert et. al via computational models of binaural hearing, suitable for engineering applications [4, 1].[2] For such models, the adoption of top-down, hypothesis-driven functionality can be viewed as enabling the interaction between "black box" cognitive functions of experience, memory and adaptation with the sensory binaural data. These "black box" cognitive functionalities are trained via supervised or unsupervised procedures usually based on transformations from signal to symbolic features and utilizing machine learning and AI classification methods.

The general architecture of such a binaural model proposed within the AABBA consortium [4] and shown in **Fig. 3**, attempts to simulate bottom-up signal processing in the subcortical monaural and binaural pathways as well as hypothesis-driven processing as attributed to the cognitive parts of the central nervous system – the latter at least as far as needed in the specific application areas. In this figure, **(a)** illustrates the head-and-torso mobility, **(b)** the signal processing in the lower auditory system, **(c)** the internal representation of the binaural activity, **(d)** the rule-and/or data-driven identification and annotation of the perceptually-salient (primitive) features, **(e)** the rule-and/or data-driven recognition, classification and labeling of proto-events, **(f)** the scene-and-task representation, knowledge-based hypothesis generation, assessment, decision taking and assignment of meaning.[5]. Note that the output of the model can be either a scene description and/or a quality judgment of the acoustic input. Quality judgments with respect to audio events will be discussed in the following section. The model accommodates multi-modal inputs, for example, from visual or haptic signals that may moderate assessment and decisions at the expert stage **(f)**. Such multi-modality interaction of the model will be discussed further in Section 4. Note also that operations in the stages **(d) – (f)** may lead to active exploration of the aural scene via the mobility function of stage **(a)**.

Hence it is proposed that, by applying these ideas, the current architecture of binaural audition (e.g. see **Fig. 2**) may be able to accommodate in a comprehensive manner the important stages of perceptual inference and knowledge and to set up hypotheses based on this knowledge. The enhanced binaural model of **Fig. 3** contains a "brain", that is, expert components which "interpret" the output of the lower, signal-driven sections of the model. At some cases, this signal-driven (bottom-up) and hypothesis driven (top-down) processing can proceed in an interleaved manner focusing on states which make sense in a given specific situation. [4]

This approach may allow the concept that the listener model ("artificial listener") actively explores its aural world and develops it further in an autonomous way. Such autonomous sensory perception, adaptation and self-awareness form the basis of audition for robotic applications [10].

3.2 A general model of the perceived quality of sound

As was the case with the test discussed in Section 2.1, experienced listeners can discriminate features and detect qualitative aspects of audio signals of systems, of music performances, and of acoustic spaces, that are not often measurable from these signals via objective means. Listeners judging sound quality utilize auditory abilities worth mimicking by engineering applications and for some time now, controlled methods have evolved allowing some degree of prediction of some limited perceptually-inspired quantities, mostly audible distortions appearing in speech (e.g., PESQ) and audio signals (e.g., PEAQ) [11].

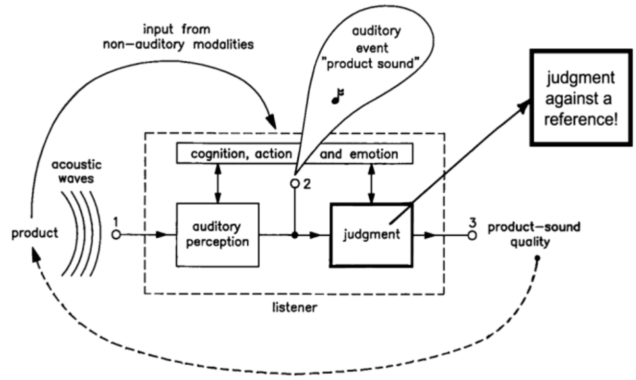

Figure 4. The "product" sound quality-assignment process (from Blauert and Jekosch, [12]).

However, in order to accommodate fully the impressive hearing system performance via a model following the paradigm shift discussed above, we have to consider the general architecture for quality-assignment introduced by Blauert & Jekosch [12, 13]. This architecture, shown in **Fig. 4**, accommodates the combined bottom-up and top-down abstraction discussed in the previous section and is mainly geared towards audio / acoustic engineering applications.

Quality is here regarded as the degree to which any auditory percept fulfills expectations and such comparison needs to be performed against a set of features and symbols provided by an internal reference or possibly via an external audio reference assigned as such by the listener. Such reference can be the manifestation of top-down sets of prior knowledge (i.e. subjective) features. Ideally, mainly due to the short-term duration of auditory memory

[2] An earlier version of such a model by Blauert has appeared in [9].

employed for comparison, any valid judgment of auditory quality may improve via a direct comparison to a reference template – as is the case with controlled listening experiments (e.g. ABX or MUSHRA tests [11]). However, for practical reasons, very seldom this is the case during a less formal listening scenario. In such cases, the listener will resort to abstract and possibly unreliable reference formed from past experience feature precepts. Additional biasing in the reference set may be due to multimodal stimuli, emotion and other reference-moderating factors [13, 14]. Also, depending on the application, the listener may judge an individual sound object or multiple sound objects composing an aural scene. The reference sets used for quality judgments fall under different levels of complexity according to the amount of intellectual abstraction involved [13]. In accenting order of abstraction, these levels may reflect

- *basic psycho-acoustic* features, (*L1*)
- *physical acoustic* features, (*L2*)
- *aural-gestalt* features, (*L3*) and
- *symbolic* or *semiotic* features (*L4*)

We may apply this analysis to the example of sound quality judgments made by experienced listeners, such as musicians or audio-mixing and mastering engineers [13].
Such listeners are able to identify primitive features in sound objects and scenes and can isolate them with the remarkable accuracy, often defying known theoretical laws as was described in Section 2.1. An active performing musician, a composer, or even an audiophile with genuine "golden ears" may utilize audition at the highest abstraction level, such as to consider relationships between sound features (see *L4*, in the list above) and semiotic or aesthetic features of the content (music), hence judging the *fidelity* of sound object or scenes with reference to such pure cognitive percepts (internal references). This clearly calls for a different level of perception than for a more casual listener of music who, at the audio scene analysis level, judges the plausibility, immersion or illusionary functionality of audio reproduction (e.g., see *L3* in the list above).
The above proposal for a layered model of audio quality may also accommodate the ambiguous case of audiophile listeners, who often claim to perceive qualitative aspects of audio systems and system components that may or may not correspond to instrumentally ("objective") measurable features, hence representing a practical manifestation of cognitive indeterminacy in audio technology. In a general sense, *the audiophile dilemma* concerning any qualitative judgment upon any audio system/component under consideration may correspond to any of the combinations in Table 1.
Assuming that an audio system/component has been properly calibrated and is objectively transparent (i.e. meeting measured specification standards), then it should conform to listener judgments falling at the two lower levels of abstraction (*L1* and *L2*, in the list above) and must result to unambiguous classification under class *C1*. However, this is not often the case.

	C1	*C2*	*C3*	*C4*
instrumentally measurable	*yes*	*yes*	*no*	*no*
perceptually measurable	*yes*	*no*	*yes*	*no*

Table 1 Possible combinations of perceptual ("objective") and instrumental ("subjective") judgments of audio quality.

Dismissing case *C4* as self-evident of an "audio fraud", it can be argued that both cases *C2* and *C3* may be attributed either to: (a) response moderating factors such as emotional status (placebo, confirmation bias, buyer's remorse, expectation bias, etc., [14]), or, (b) hypothesis-driven and reference-biased listening acuity utilised by an experienced audiophile listener. Such cases of experienced listeners and audiophiles illustrate that the cognitive functions activate references at the highest abstraction level of sound quality assessment which appear to overrule judgments based on lower levels. Therefore, such listeners can disregard evidence from psychoacoustic attributes, physical acoustic properties and aural gestalt features. Significantly, as it appears from the discussion in Section 2.1, such a procedure can provide judgments with far more accuracy than it is allowed by pure bottom-up, signal-driven approach. Furthermore, it shows that the proposed model can accommodate listening selectively at any level of abstraction or cases where the attributes and features from different abstraction levels are combined to form a comprehensive quality judgment. Significantly, the proposed model of sound quality is not bound to signal processing principles and, thus, can accommodate ambiguous and paradoxical cases in sound sensation as mentioned above.

4. DISCUSSION AND CONCLUSIONS

4.1 Challenges for models of auditory periphery

The human hearing mechanism has a performance superior to any single man-made instrument for similar applications. It is accurate in detecting, analysing and recognising acoustic events covering a $10^3/1$ range of frequencies, a $(32x10^{12})/1$ range of power ratios, and this is achieved over all spatial directions with a few degrees azimuth angle discrimination (down to 1° for frontal sound incidence). Considering everyday speech communication, which approximately requires a range of frequencies with a ratio of 10:1 and a power ratio of $10^4/1$, such specifications appear to be highly over-engineered. However, for speech recognition tasks, audition achieves exemplary performance especially under adverse acoustic conditions: It can isolate and recognise a preferred speaker under competing speech and maintains intelligibility in noisy and reverberant environments. It is also remarkable that such functionality is implemented via highly non-linear processes, starting at the inner ear stage which compresses the stimuli dynamic range into a $10^4/1$ ratio. However, it is not evident how such impressive acuity is reflected on the spike train features which code the sound stimuli.

In contrast to the well identifiable time and frequency domain properties of sound signals, the coded spike train responses have indeterminate and often chaotic features that cannot be fully described via known principles or analogies hence leading to a possible conclusion that information and acuity is lost via coding at the auditory periphery. For example, with respect to timing accuracy, such spike trains can be only considered as stochastic processes with temporal uncertainty due to Gaussian distribution timing jitter [6]. Furthermore, auditory nerve fiber coding exhibits selective temporal firing by progressive saturation of the transmitted information for steady state sounds (e.g., speech vowels) once locking to their pitch frequency has been established [6]. However such selective coding of onset signal features results to an impressive temporal resolution (20 μsec approx.). Such accuracy enables the previously mentioned horizontal plane sound localisation accuracy and the joint time-frequency discrimination which defies the mathematical theory (see Section 2.1). This is only a short list of paradoxes and missing links in our current knowledge of the auditory mechanisms, even at subcortical level, where the performance of the auditory system already challenges current knowledge in many scientific fields: physical acoustics with respect to the covered power range, mathematics and signal processing with respect to neural coding, and engineering with respect to ear transduction principles.

Nevertheless, over the last decades, besides partial knowledge of such principles, a number of perceptual mechanisms have been translated into computational models and into applications via digital signal processing analogies. For example, auditory masking models have been successfully utilised in convincing and game-changing digital audio technology applications such as the *mp3* coding format [15]. Similarly, successful models were developed utilising binaural activity features and mappings and have initiated significant developments in many applications in spatial audio reproduction [16]. Such established signal-processing models have paved the path for successful engineering application motivated by human hearing. Applications of this kind are now reaching a stage of maturity that requires addressing more important issues relating to sound sensation.

4.2 Challenges for models of auditory perception

It is accepted that beyond the auditory periphery, which is usually addressed by current models, higher level cognitive processes are even less understood and, hence, present an open challenge not only to engineers but to many interdisciplinary scientific fields. Although current brain-imaging technologies via fMRI and PET produce images of activation patterns in the human brain during auditory tasks, it is not yet known how to relate such maps to cognition or to perceptual features. Our knowledge is obscure regarding cognitive functions which associate and adapt stimuli to previous experience, that is, to the significance of the observed auditory event in relationship to expected and known percepts by way of utilising memory, learning and anticipation via the generation of hypotheses. Especially when listening to music, (listed as "listening for pleasure" function in the classification in [1], see also Introduction), there is evidence that apart from activation of the primal auditory cortex, which identifies the simple features of pitch and loudness, other areas in the brain lobes are also activated responding to more structured aspects such as harmony, melody or rhythm, together with areas responsible for the emotional state. Significantly, listener enjoyment appears to activate those regions of the brain usually associated with sexual and food pleasure, that is, the reward of such biological actions. There is also evidence that trained musicians activate correspondingly larger areas in the brain lobes [17].

Additionally, at such higher stages of cognition, different sensory modalities appear to strongly interact and humans identify objects, events, scenes and qualitative aspects often combining sensory modalities. Hence, auditory or visual events / objects are not only activating the corresponding areas of auditory or visual cortices but generate cross-modal activations and, according to recent results [18], the bottom-up processing of sensory stimuli is often accompanied by the top-down reconstruction of associated patterns in different modalities. Such cross-modal and multi-modal aspects appear to play a significant role in cognition and require yet unknown organisation of stimuli and responses with dynamic and adaptable interconnection structures in the brain and peripheral processing units. Therefore, any comprehensive model for sound sensation needs to address sound cognition in conjunction to other sensory modalities, especially with vision. Furthermore, as has been established in neuroscience [18], neuron ensembles in higher-order association cortices register associations among perceptual representations from multiple sensory modalities. It has been also suggested that information from different sensory channels converges somewhere in the brain to form modality-invariant representations, i.e. representations that reflect an object / event / scene independently of the modality through which it has been received. Such modality-invariant representations may be a first stage for identification of the neural process that allows recognition and response to supramodal sensory stimuli, i.e. functioning at a pure conceptual level [18].

4.3 Implications of the model paradigm shift

The recent paradigm shift in engineering model for binaural listening can accommodate many of the challenges, paradoxes and missing links identified previously and addresses limitations of the conventional models based on the bottom-up signal processing analogy. The model architecture described in Section 3.1 contains a complementary top-down, hypothesis-driven functionality that can enable interaction between the yet somehow obscure "black box" of cognitive functions of experience, memory, adaptation and multimodality to the bottom-up binaural cues and features extracted from the acoustic signals. Currently, machine learning, artificial intelligence and pattern recognition methods are employed to annotate and label signal features selected from well-defined rules and auditory “scenes”. Hence, for a given scenario or application, expert components trained by typical features, may “interpret” the output of the lower, signal-driven sections of the model. Significantly, as was also shown in Section 3.2, such architecture can also accommodate multi-layer judgments of sound quality, addressing functionalities often observed in experienced listeners that otherwise defy current signal processing or psychophysical conventions.

The implications of such paradigm shift extend to the conceptual and philosophical principles related to the relationship between perception, knowledge and reality.

At a first level, this new approach indicates a divergence from the dominant “*objectivistic or realistic*” approach favoured by most engineering methods. Such approach is associated with the *synthetic* principle of the modelling the observer as carrier of subjective sensations derived via a bottom-up procedure from stimuli from an objective reality.

The model described in Section 3.1 via its top-down structure can accommodate the approach of “*perceptionism*”, according to which the consciously perceived percepts represent essentially the real world [19]. Such concepts can be traced to the principles defined by Kant [20], according to which, all human experience can be related to quantity, quality, relation, and modality, modified by perceptual thresholds, various forms of comparative judgment, magnitude estimation, emotional response and scaling, as described by cognitive psychophysics which relates the physical world with its mental interpretation. Kant also divided all scientific propositions according to their logical form into *analytic* and *synthetic*. Facing the problem of the hearing mechanism, Riemann re-examined the interplay between such classical concepts of *synthesis* (the anatomist`s approach which builds up the knowledge by investigating the individual components) as was at the time established by Helmholtz, and of *analysis* (the hypothesis-driven approach which examines the tasks accomplished by the organ). He concluded that for the case of hearing, the analogy employed via the analytic abstraction has to do with some sort of associative thinking which is free and disciplined [2]. Analogy also allows a formal association between seemingly different types of phenomena, often related to creativity in arts and science, and was thus called "*the poetry of hypothesis*" by Riemann. Analogy is at best used to formulate hypotheses about such fundamental principles and the acceptable conditions that a system must satisfy in order to achieve its functionality without specifying the manner in which its components function and interrelate with each other. These crucial elements seem at last to be adopted into models of binaural hearing.

The analytic / synthetic logical distinction has been also divided by Kant according into their *a priori* or *a posteriori* validity, i.e. according to whether their claim to truth or falsehood was in no need of empirical backing (*a priori − analytic*) or was so in need after the synthetic process (*a posteriori*) [21, 20].

Nevertheless, from an engineering modelling perspective, any such new analogy of the largely unknown cognitive operations of the brain must be open for scientific scrutiny, and its specific implementation must be tested and verified. According to Popper [21], “...*Scientific knowledge proceeds from old problems to new problems by means of conjectures and refutations*”.

Consequently, testable cases of sound sensation need now to be investigated with respect to the proposed paradigm shift, hopefully resulting in solutions to the burning problems and envisaged applications to binaural listening as discussed in audio technology.

Acknowledgments

The author thanks Jens Blauert, D—Bochum, for constructive comments and corrections to this manuscript. This research has been co-funded by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Project: THALIS-UOA- ERASITECHNIS MIS 375435.

5. REFERENCES

[1] J. Blauert, ed.: “The Technology of Binaural Listening”, *Springer and ASA Press*, 2013.

[2] T. Ritchey, “Analysis and Synthesis. On Scientific Method-Based on a Study by Bernhard Riemann”, *Systems Research*, Vol. 8, No. 4, pp 21-41, 1991.

[3] H. Helmholtz, “On the Sensation of Tone as a Physiological Basis for the Theory of Music”, *Dover Publications*, 1954.

[4] J. Blauert , D. Kolossa, K. Obermayer, K. Adiloglu, “Further Challenges and the Road Ahead”, in J. Blauert, ed.: “The Technology of Binaural Listening”, *Springer and ASA Press*, 2013.

[5] J. N. Oppenheim, M. O. Magnasco: “Human Time-Frequency Acuity Beats the Fourier Uncertainty Principle”, *Phys. Rev. Lett*. 110, 044301, 23 January 2013.

[6] H. Wang · M. Isik, A. Borst , W. Hemmert “Auditory information coding by modelled cochlear nucleus neurons”, *J Comput Neurosciense*, 2010.

[7] II M. Park, S. Seth, A. Paiva, J. Principe “Kernel Methods on Spike Train Space for Neuroscience”, *IEEE Signal Proc. Magazine*, 149, July 2013.

[8] R. Munkong, B-H. Juang, “Auditory Perception and Cognition”, *IEEE Signal Proc. Magazine*, 98, May 2008.

[9] J. Blauert, “Binaural Auditory Models: Architectural Considerations”, in A.N. Rasmussen, P.A. Ostenhammel, T. Anderson, T. Poulden, edts: “Auditory Models and Non-linear Hearing Instruments”, *Proc. 18th Danavox Symposium*, 1999.

[10] S. Argentieri, A. Portello, M. Bernard, P. Danes, B. Gas, “Binaural Systems in Robotics”, in Blauert, ed.:

"The Technology of Binaural Listening", *Springer and ASA Press*, 2013.

[11] S. Bech, N. Zacharov, "Perceptual Audio Evaluation – Theory, Method and Applications", *John Wiley & Sons, Ltd*, 2006.

[12] J. Blauert and U. Jekosch, "Auditory Quality of Concert Halls−The Problem of References," *Proc. 19th Int.Congr. Acoust., ICA 2007* vol. 38, 2007.

[13] J. Blauert, U. Jekosch: "A Layer Model for Sound Quality" *J. Audio Eng. Soc.*,vol. 60 (1/2), 2012.

[14] "Audio Myths Workshop", *127th AES Convention*, New York 2009. http://www.youtube.com/watch?v=BYTlN6wjcvQ

[15] M. Bosi, R. Goldberg, "Introduction to Digital Audio Coding and Standards", *Kluwer Academic Publishers*, 2003.

[16] J. Breebaart, C. Faller, "Spatial Audio Processing: MPEG Surround and Other Applications", *John Wiley & Sons, Ltd*, 2007.

[17] A. Patel, "Music and The Mind", http://www.youtube.com/watch?v=ZgKFeuzGEns.

[18] K. Man, J. T. Kaplan, A. Damasio, K. Meyer "Sight and Sound Converge to Form Modality-Invariant", *The Journal of Neuroscience,* 32(47)*,* 2012 .

[19] J. Blauert, "A Perceptionist's View of Psychoacoustics", Archives of Acoustics, Vol. 37, No. 3, 2012.

[20] I. Kant, "Critique of pure reason", *Johann-Friedrich Hart-knoch,* 1781.

[21] K. R. Popper, "Objective Knowledge. An Evolutionary Approach", *Oxford University Press*, 1972.

[22] A.S. Bregman, "Auditory Scene Analysis", *The MIT Press*, 1994.