AES Reviewed Paper at Tonmeistertagung 2018

Presented* by VDT in cooperation with the

Central European Region of the Audio Engineering Society (AES).

On Human Perceptual Bandwidth and Slow Listening

T. Lund[1], A. Mäkivirta[2]

[1] *Genelec OY, 74100 Iisalmi, Finland, Email: thomas.lund@genelec.com*

[2] *Genelec OY, 74100 Iisalmi, Finland, Email: aki.makivirta@genelec.com*

Abstract

Seventy-four recent physiological and psychological studies revolving around human perception and its bandwidth were reviewed. The brain has ever only learned about the world through our five primary senses. With them, we receive a fraction of the information actually available, while we perceive far less still. A fraction of a fraction: The perceptual bandwidth. Conscious perception is furthermore influenced by long-term experience and learning, to an extent that it might be more accurately understood and studied as primarily a reach-out phenomenon. Considering hearing, time is found to be a determining factor on several planes. It is discussed how such sentient observations could be taken into account in pro audio, for instance when conducting subjective tests; and the term "slow listening" is devised.

1. Introduction

The paper presents a broad review of selected articles and studies that have a bearing on auditory perceptual bandwidth in humans, particularly in the framework of professional audio applications, including listening tests.

In pro audio, double blind ABX testing is regarded as best practice. ITU-R BS.1534, ITU-R BS.1116 and similar standard procedures are generally used as the foundation for extensive statistical analysis, and often able to provide repeatable results. BS.1116 assumes that subjects have the capability to hear all phenomena tested for fully developed prior to participation. However, both guidelines also suggest the use of audio stimuli of 10-25 s duration [1, 2].

User interfaces where quick recognition could be a question of life and death address the limited conscious bandwidth of humans; a horn in a car, flight deck controls (visual design, phasic alerts, haptic alerts), medical equipment etc. A smartphone does every trick in the book to get noticed, a magician knows how not to.

Regardless, subjective audio testing currently puts high faith in that seemingly scarce resource, conscious bandwidth.

Thanks to new non-invasive in vivo experimental techniques, we are getting a better understanding of how *time* can affect sensing. Furthermore, explicit consciousness, the feature considered a hallmark of humans, has lost in importance over recent decades, if physiological and psychological findings have been interpreted correctly [3].

Basing subjective tests solely on explicit conscious and immediate responding may not provide a good enough understanding of long-term effects of a salient experience.

1.1. Interior and exterior

Even thousands of years before antiquity, our forefathers were aware of sensing, its importance and its limitations. The 32.000-year-old painting of an owl in Grotte Chauvet, for instance, recognize ears and eyes as main attributes, see Fig. 1 left. In contrast, we don't really know how music of the great composers sounded just 300 years ago.

Socrates, Plato and later philosophers, with idealism, reason and eventually scientific method, helped us get a firm stand on previous generations' shoulders, and collectively break one step away from the cave. However, every child born today

*: Please note that the AES Reviewed Papers at Tonmeistertagung can be published by both, AES and VDT, in print, online and as PDF download.

still spends many years to learn using her or his senses, the same way a long gone baby in Ardèche had to.

Fig. 1: More than 30.000 year old paintings from Grotte Chauvet.

Having noticed how unreliable perception was, René Descartes famously took refuge only in his mind, and declared consciousness to be anchored in our awareness of our own awareness, *je pense, donc je suis*. Enlightenment idealist George Berkeley took the opposite stance, that things only existed by being perceived, an idea that reverberated even into quantum physics, and from there to hypotheses on consciousness and brain topology today [4, 5].

19th century experimental psychologists Ernst Weber and Gustav Fechner discovered our senses to be logarithmic and developed the concept Just Noticeable Difference (JND), still widely relied on in a variety of sensory studies.

However, around 1850, Hermann von Helmholtz pioneered systematic studies on physics and perception. Based on research, he came to the radical conclusion that consciousness does not have access to all data and intermediate results that produce sensation [6]. He called the phenomena "unconscious inference" (in German "unbewusster Schluss"), but he also studied other aspects of involuntary entry into consciousness. Helmholtz's findings were refuted for decades, but they are now at the very heart of recent studies and models, detailed later.

1.2. Individual and collective

In our era, where extensive data about almost any highly specific topic is instantly available, the combination of discoveries from different fields is sometimes what matters more. A modern methodology and encouragement of inter-disciplinary problem-solving is shown in Fig. 2, the quadrant model of Integral or Integrative Perspectivism as proposed by philosopher Ken Wilber [7]. The concept is to consider a question or problem from all quadrants.

Besides from being a methodology under further development at Danish and German universities [8, 9], Integrative Perspectivism has proven a valuable tool for stimulating discussion in international standardisation involving stakeholders from different fields; for instance, engineers, physicians, psychologists, artists, economists and politicians.

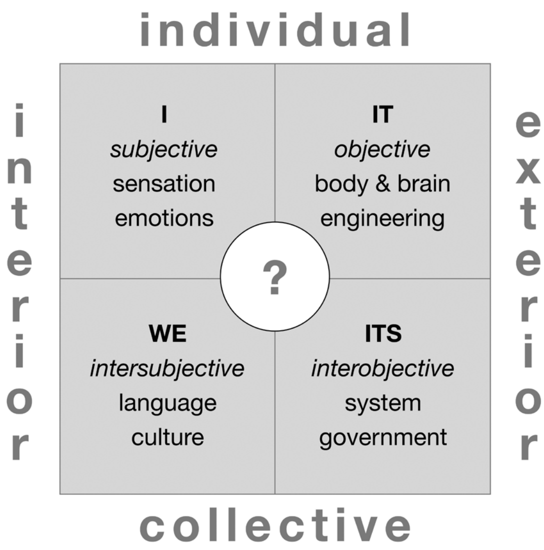

Fig. 2: The quadrant model. Left side subjective, right objective.

Economic utilitarian societies tend to appreciate the exterior (objective) right-hand quadrants more, so the model is also a reminder of more profound human interests, spanning millennia rather than stock cycles.

2. Nomenclature and method

Our sensory system is bombarded with information, leading to a data deluge that cannot be processed in real time. The role of attention is to select information to meet current behavioural goals. Consciousness does not select information but summarizes it, potentially on a longer timescale.

Conscious is used in a narrow, sentient definition: The ability to discriminate, categorize and (to some extent) recall external stimuli that fall within our reception range.

Attention means selective attention. It is the way we can focus on certain aspects of sensory stimuli, mono or multi-modal. The two are therefore entirely different.

The hard problem of consciousness, sentience, the feeling of perception, is important when recognising human art from the past, but outside the scope of this review.

2.1. Reception and perception

A plethora of receptors generate *afferent* nerve impulses about the status of the body, as well as our close and more distant surroundings. Via nerve fibres and a number of synapses, impulses can be relayed to the central nervous system (CNS), allowing the brain to participate in the reception, and occasionally to produce a sensory experience [10].

Reception obviously has a steep funnel associated: We are at a given physical location in space and time, and reception is tuned for conditions that generally matter on planet Earth, for a creature our size and composition, having a certain life-span etc. Our reception apparatus registers only a certain mechanical wave frequency range (hearing, touch), a certain electromagnetic frequency range (seeing, touch), certain smell/taste dimensions etc.

Considering hearing, the brain indeed is an active participant, not only in the decoding of minute temporal information, but also as the main element of a sense relying heavily on internal tuning. Hearing therefore also features a substantial number of *efferent* nerve fibres. Such fibres constantly send information back to the middle and inner ears, adjusting the reception system itself.

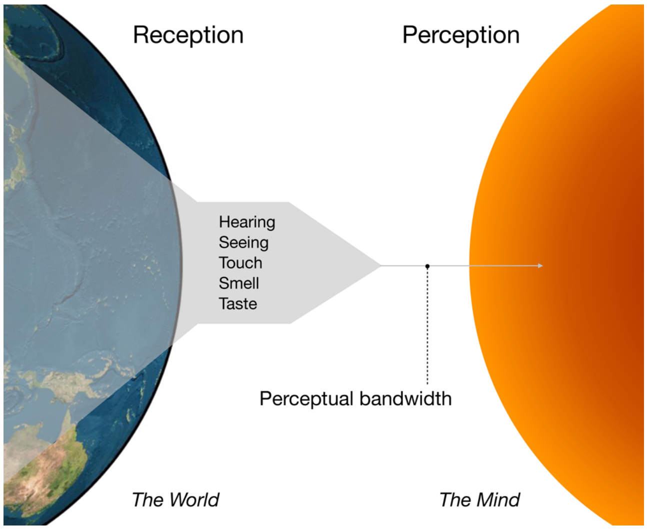

Fig. 3: The two funnels of human apprehension, this review focusing on the second, i.e. perceptual bandwidth.

Perception is distinguished from reception and introduces a second funnel between the exterior world and consciousness, see Fig. 3. Perception is entirely subjective. It is the outcome of sentient brain processing based on experience, expectations, mood, attention and - to some extent - reception.

In this review, *exteroception* is used about stimuli (appearing to originate) from outside the body, and *proprioception* about our sense of movement and relative position of body parts. Because *interoception* in psychology is currently taking on a broader meaning than sensing of internal body status (hunger, pain, body ownership etc.), we shall refrain from further using that term [11].

Perceptual bandwidth is the rate by which we can (consciously at the moment) register sensory stimuli, external or internal.

2.2. Inclusion criteria

A more than 50-year-old finding was inspiration for this review: An estimation of how quickly humans can acquire new information from the exterior. We have searched for results confirming or rejecting Karl Küpfmüller's observation of a modest upper limit for data "input-velocity" (Nachrichtenfluss) [12], and for new findings revolving around the topic. The precise number of bits/s was considered less relevant than verification of the pronounced sensory speed bottleneck itself.

New in-vivo human research and diagnostics tools have been introduced or refined in recent decades, potentially tipping discussions about the perceptual bandwidth to a certain side, or discovering entirely new connections and insights: Non-invasive measurements of various brain event-related potentials (ERP), auditory brainstem responses (AEP) and nuclear magnetic resonance imaging (MRI), especially its temporal version, functional MRI (fMRI); a technique based on hemodynamic response. Neurons require glucose when firing, and the brain has no local energy storage.

Understanding of human exteroception has improved dramatically from early studies based on PET scans 25 years ago, to fMRI today, with better temporal resolution and typically around 1x1x1 mm spatially (voxel 256x256). Animal trials have also been employed, for instance when human experiments were considered unethical.

This review is based on articles and papers listed in PubMed, Embase or PsycINFO, searched first for an AES Convention paper [13] using MeSH terms and keywords between April and June 2017. January to June 2018, a second search was performed, because several trials had unexpectedly reported a pronounced reach-out element to exteroception, leading to updated inclusion criteria. More than 350 articles were found to meet the (new) inclusion criteria. A secondary sorting process was carried out based on title and abstract, limiting the number of studies to 103, thereby adding risk of false negatives and bias.

A tertiary sorting was performed for a condensed presentation of results compatible with this TMT conference paper format, and the number of reported studies limited to 74. Newer studies were prioritized over older. Independent searches between the authors were not systematically performed. Only publications available in English, German, French or a Nordic language were considered for inclusion.

3. Results

Reviews were divided into six categories about perceptual bandwidth, latency of sensation, short duration sound, auditory stress, active sensing and unconscious inference, including long-term learning effects.

3.1. Perceptual bandwidth

Based on physiological evidence, our brain, mostly via cranial nerves, receives sensory data from millions of neurons when we are awake. However, according to Karl Küpfmüller, perceptual bandwidth is limited to 40-50 bits/s in total, and vision and hearing are the only senses that are even close to saturating that bottleneck; while touch, smell and taste can contribute 5 bits/s or less [12].

According to [14], "numerous experimental results indicate that the bandwidth of human perception is severely limited." Consciousness is fairly sparse, and more limited than our introspection would suggest [15].

Several studies also find that during audio-visual stimulation, when participants attend to *just one* modality, the sensory cortical areas serving the attended modality show increased activity and the ignored sensory cortices show a decrease in activity, compared with when attention is directed to the other modality [16, 17, 18, 19]. However, in *divided* attention experiments, audio and visual cortices are not depressed, though performance does decrease. Competition for bandwidth appears to happen elsewhere in the brain, possibly where resources are shared, but without the mechanisms being fully understood [19].

Perceptual decoupling (PD), i.e. to disengage attention from perception, is another example where fMRI has provided new insight by charting the meandering between sensory experience and self-generated thoughts uncoupled from perception. PD is shown to happen frequently in daily life, and even called "pervasive" among specialists such as pilots, thus modulating perceptual bandwidth in unforeseeable ways [20]. A high cognitive load can reduce perceptual bandwidth further. Inattentional deafness in aviation was observed in 1/3 of high load cases where subjects overheard alarms they were instructed to report [21].

Finally, studies suggest there could be no fundamental difference between normal perception and hallucinations or pathological conditions, besides from a grading of how much exteroception is used to anchor an experience at all [22].

3.2. Latency of sensation

When physical events happen within our reception range, we cannot experience it instantly. Besides from a delay in the reception system, nerve potentials travel at a finite velocity, and each synapse with its chemical conduction takes time too. The 12 pairs of cranial nerves are categorized as part of the central nervous system (CNS). They can be regarded afferent and efferent highways, for example carrying sensory inputs to the brainstem without introducing extra, time-consuming synapses.

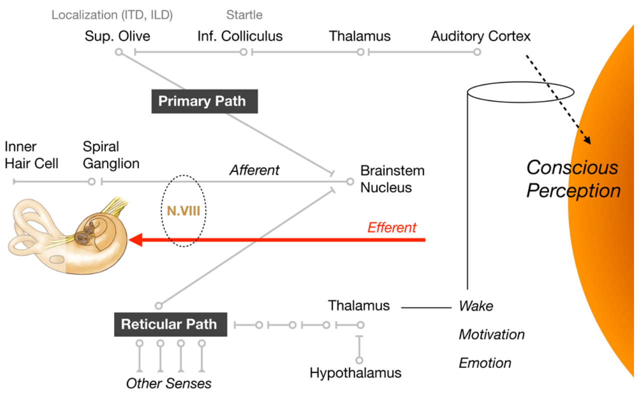

Fig. 4: Auditory nerve (N. VIII), primary and reticular pathways.

Once sensory information arrives at the brainstem, it follows a primary and a reticular path, the latter being used for prioritization between the senses when allocating bandwidth via our perception switchboard, the midbrain's thalamus. Considering the primary auditory pathway, this is where localization and several reflexes origin ahead of conscious awareness, see Fig. 4.

Our conscious mind without attention, however, is associated with a latency of 400 ms where a backwards referral system is hypothesized to make us believe events happen now, in order for a conscious mental image to synchronize with our motor actions [23, 24, 25, 26], see Fig. 5.

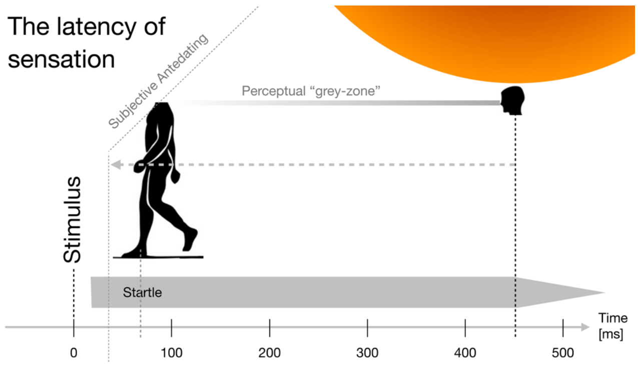

Fig. 5: Latency of sensation, reflexes and perceptual grey-zone.

More recently, in an ERP study based on vision, it has been found that latency depends on whether access is driven by stimulus saliency alone or by a combination of expectations and sensory evidence. Considering the latter, latency of specific neural signatures decreased from 300 ms to 200 ms [27]. Similar effects have been found for hearing [28].

It has long been debated whether perception is time- discrete or continuous, and theories have been proposed that rhythmic brain activity support the periodic sampling of the environment [29, 30]. Cortical theta rhythms (around 1/4 Hz) observed in human scalp EEGs indeed have just been found to be “strong evidence of rhythmic perceptual sampling as an inherent property of the frontoparietal attention network in humans”, even independent of task structure [31]. An underlying perceptual sampling mechanism could potentially also further modulate the timing of sensation.

3.3. Perception of short duration sounds

Basic temporal decoding skills are innate: We are born with the ability to detect differences among all the 500 phonetic distinctions used in the world's languages. Within a few years, however, only the roughly 50 contrasts that are used in our native language are selectively promoted, while we lose the ability to distinguish among the rest [32, 33, 34].

Using Mismatch Negativity (MMN) [35], several studies confirm a perceptual relationship between phonemes in language and rapidly changing acoustic information in music: "By showing similar effects of linguistic and musical expertise, these results argue in favor of common processing of duration in music and speech." [36].

It has also been found that for syllables requiring resolution of temporal cues, musicians are better at identifying unusual phonemes than non-musicians [37]. Subjects that speak multiple languages with pronounced difference between phonemes (Finnish and French) have increased short-term decoding skills in common with musicians [36], confirmed also in [38].

The left auditory cortex is specialized in the processing of acoustic transients as well as short sounds in language [39], musicians listen differently than non-musicians by engaging motor cortex [40], and the style of music played shapes auditory skills [41].

Music learning in childhood helps language learning [42]; and phonemic templates are formed early in life, so some non-native consonant contrasts are difficult to learn later [43]. Lasting impact of early language learning is also reported in [40], even if exposure does not continue into adulthood.

Finally, a study found that from the early phase of memory encoding, the auditory cortex plays a prominent role in impulse control and emotional learning [45], possibly pointing to it having a more general influence on mental timing in humans.

3.4. Auditory stress and impairment

Some sound stimuli can temporarily or permanently cause reception to degrade, but studies find that not only hair cells in the cochlea are subject to impairment. Recent animal studies have shown that sound exposure can lead to neuronal degeneration, even when hair cells recover and thresholds return to normal. The neurotransmitter, glutamate, becomes toxic with excessive sound exposure (excitotoxic), causing swelling and damage to nerve terminals [46].

Degeneration of cochlear peripheral axons may also be an important component of presbycusis (age-related hearing loss) despite a near-normal hair cell population [47], and remain hidden in (younger) subjects when tested using standard clinical methods [48].

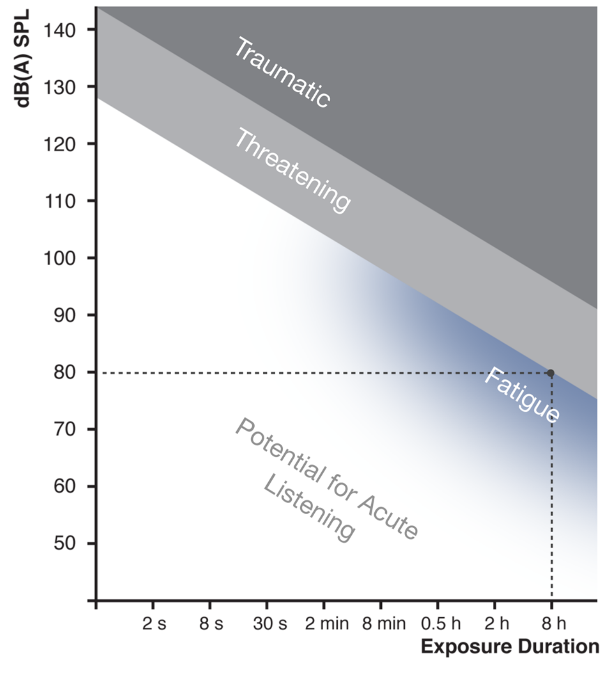

Fig. 6: Safe and potentially harmful sound exposure in adults per day, i.e. sound pressure integrated with time. Adopted from [49].

Besides from overall sound exposure, the quality of sound can also be a factor [49], these variables having the potential to shorten the time it takes before fatigue manifests: Excessive high frequency energy, lack of "quiet transients" and interaural strangeness or unnaturalness. The brain's fastest firing synapses (calyces of Held) are employed in ear comparison, and consequently also prone to fatigue [49, 50], albeit possibly not atrophy.

Recent studies estimating adverse effects of sound exposure per day (24 hours) are summarized in Fig. 6.

3.5. Active sensing

Are we passive receivers of sensory information from our environment, or actively collecting it? Contemporary trials agree on the important roles learning and experience play in perception. However, theories differ in terms of neuronal implementation, largely falling into two categories: A bottom-up accumulation model, and a top-down prediction-delta model [51, 52], with the latter gaining more support recent years [53, 54]; or at least the two complementing each other [55].

Since Alfred Yarbus' seminal work on eye movements was translated into English [56], it has been evident that saccadic eye movements may be modulated top-down by various task demands. A new study finds that implicit learning also guides saccadic movements, but in a qualitative different way than when goal-driven [57].

Different levels of expertise in detecting (pathological) visual features strongly affects saccadic eye movements, to an extent that "expert behaviour is manifested in distinct eye-movement patterns of proactivity, reactivity and suppression." Furthermore, "the fewer fixations needed to detect a visually salient abnormality, the more advanced the level of development." [58].

Saccadic eye movements are also hypothesised to be a more general sign of sensory reach-out whereby we sample the environment in order to test predictions [59, 60]. In such case, reach-out behaviour would likely not relate only to space (vision), but also to time (hearing).

EEG based studies indeed find that the rhythm of amplitude modulated noise and frequency modulated complex tones modulate perception, i.e. immediate explicit detectability is modulated by rhythm [61, 62]. Another auditory-based study concludes that regularity in non-attended items in itself does not capture attention, and that findings can be reconciled as resulting from mechanisms that minimise surprise [63].

It is also reported that real-time sensing is prone to sacrifice optimality in order to ensure self-consistency [64, 65], at times showing preference for a percept that is partially inferred, yet paradoxically considered more reliable than a percept based on external input [66].

3.6. Unconscious inference

This paragraph also includes subsequent learning, cognitive activities carried out unconsciously, and exteroception that produce (more than usual) delayed consciousness.

Unconscious inference in perception is not questioned in recent work. A commonly used mathematical model today, the Bayesian framework, even formalizes Helmholtz's original ideas [6, 55, 67, 68].

Objects can be attended to without consciously being perceived, i.e. attention is able to select invisible objects [69].

A new study finds that sound can unconsciously, not through attention, disambiguate the contents of visual awareness by facilitating perception of audiovisually congruent stimuli [70]. Other trials have found alpha wave EEG (electro-encephalography) response to hypersonic sound that could not be reported based on explicit consciousness [71, 72].

Positive and negative (ironic) priming affects what involuntarily enters consciousness [73], and problem solving can be performed without consciousness or attention [74, 75].

Various ERPs are under scrutiny as objective prediction error markers, for instance MMN, P300 and contingent negative variation (CNV), where each one's relation to expectation and attention are being examined [76].

Recognition memory, and other types of memory relevant in this context, are being studied specifically [77]. Applying a test protocol that is sensitive to detecting mnemonic mechanisms operating outside of conscious perception, by manipulating subjects' awareness at study and/or at retrieval, it has been found that non-conscious processes can drive accurate recognition memory in either case, or even in both [78].

Responses in auditory cortex are preserved during sleep in rats, and human subjects have been able to learn new sounds during REM or light non-REM sleep [79, 80]. General anesthesia and intensive care units are abundant with reports of patients hearing and learning when they were clearly not conscious [81, 82], and contact has been established to a number of patients in a persistent vegetative state, i.e. having the lowest score, 3, on the Glasgow Coma Scale [83, 84, 85].

4. Discussion

As a natural consequence of our limited perceptual bandwidth, we have to reconsider the notion of sensory *reception*. Could we be systematically underestimating the efferent component of a sensory experience? The studies reported suggest that we generally are, and how Fig. 7 would be a more relevant representation of human probing of the environment than Fig. 3.

Thanks to higher temporal resolution brain imaging techniques, evidence is also mounting that perception is predominantly a reach-out process, and that it could be driven mainly by exteroceptive prediction errors.

Most trials included were based on vision, where brain imaging can be combined with an objective marker, saccadic eye movements, but no studies suggest the fundamental principles to be much different for hearing. On the contrary, our two highest perceptual bandwidth senses are likely to make use of the same basic information gathering and filtering strategies, though specialized in two different dimensions, spatial and temporal.

Aside from temporal neural decoding, reach-out features of hearing would thus include body and head movements.

As organisms, we must be adaptive, but also conserve resources, so reach-out in exteroception would be just enough that we were able to resolve most uncertainty about the environment quickly. Unexpected findings are taken into account by the brain as potential updates of future sentient predictions.

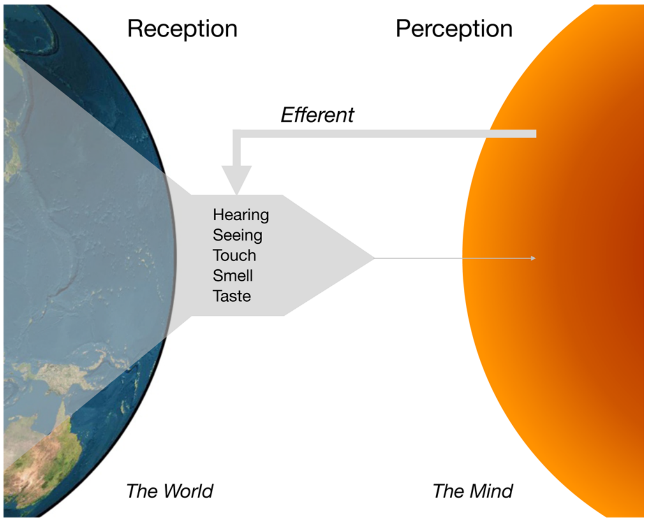

Fig. 7: Exteroception is highly influenced by experience and efferent neural activation.

The active reach-out consensus in general was a surprise finding of this review, though the notion has been held by perceptual psychologists for decades, taking off from the work of J. J. Gibson [86]. Both hypotheses, however, would be readily compatible with a limited perceptual bandwidth in humans.

If sensing primarily serves to correct experience + expectation, the time it takes to assess auditory stimuli close to what one would with unlimited time, depends almost entirely on familiarity with the features and artefacts evaluated.

From the opposite perspective, we gradually become what we sense, without having much notion of what there might be "outside the cave". From generation to generation, our species is getting better at conveying explicit information (symbolic, language), which though can be biased [87], but some of our knowledge is still tacit, and some of that relies on personal sensing. We consequently owe children reasonably unlimited sensory examples so they can fully develop sentient faculties including relevant reach-out behaviour.

Children do not just learn; they reach curiously out to the world and also learn how to learn [88]; but short-duration phonemic contrasts outside our mother-tongue are difficult to learn once we grow up. Another auditory feature, absolute pitch, the ability to identify the pitch of a sound without absolute reference, has also repeatedly been linked to musical training early in life [89, 90], and to learning a native language where pitch is a strong convenor of meaning, possibly supported by genetic disposal [91].

Considering hearing, part of the learning would be familiarization with the language of all humanity - music - performed using discrete acoustical instruments in a fine hall. Under such settings, we are able to combine musical features and the full spectrum of auditory acuity with head movement and other reach-out actions: Frequency range, dynamic range, localization, imaging and envelopment.

Since the arrival of composers exploring the potential of abstract instrumental form, for example L. v. Beethoven, music became an even more prominent conveyor of emotions and engagement. Any such musical work, old or new, clearly deserves excellent preservation for future generations to enjoy. Old music scores are subject to today's interpretation, so our experience of such compositions unfortunately are severed, compared to, for instance, visual art of the past. Future generations need not need to have the same kind of regrets about our time.

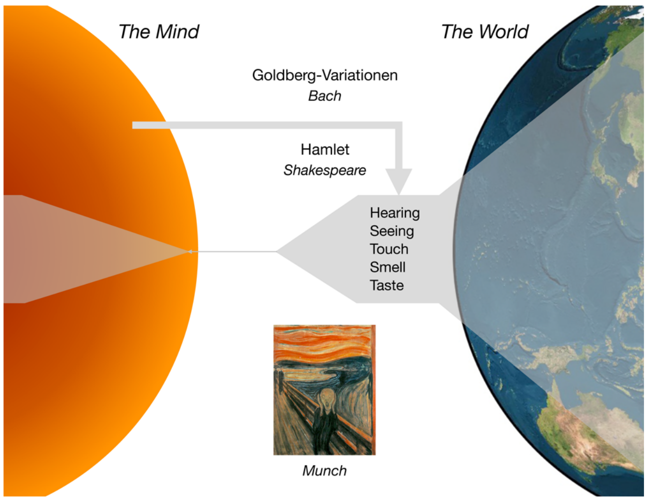

Fig. 8: Perception illustrated as a reach-out, modest return phenomenon; art being the exception.

One way to interpret "art", based on the issues discussed here, would be its unusual ability to make perceptual bandwidth momentarily widen in the receiver, for instance when listening to J. S. Bach, seeing an E. Munch painting or reading W. Shakespeare, Fig. 8. If we look at Fig. 1 right, such a quality may arguably be preserved for 30.000 years; again pointing to aspects of being human that have remained fundamentally unchanged; in essence the upper left-hand quadrant of Fig 2.

4.1. Implications for AES

New technology has the potential to influence all individual and collective aspects of our lives. Tech-robust phonemic contrasts, for instance, could help promote some words and depreciate others, possibly entire languages, like pilot's alphabet has found general use because of poor sound quality in phones, to maintain verbal exchange of meaning at all [92].

Because hearing is not only employed in explicit communication, but also offers a rare, relatively high bandwidth channel capable of carrying tacit information across generations, we must be careful not to over-simplify auditory content and distribution due to current cognitive or technical limitations.

Using traditional subjective testing, it has proven difficult to argue clearly in favor of higher data-rates than 48 kHz/24 bit linear PCM per channel [93]. The same kind of tests, however, have also been used to promote lossy data reduction, where most audio information is discarded, though many interested in sound today notice warbling "space monkey" artefacts and collapsed imaging across platforms, be it broadcast, YouTube, music streaming or phone. That kinds of artefacts might be felt more gravely now as more listeners have become "experts", than when the codecs were originally tested.

Consequently, all three temporal time-scales - the 400 ms grey-zone, fatigue and long-term learning effects - should be taken more into account in future audio standardisation, so our society is not used to rubber-stamp vulgarisation of audio in recording, storage or distribution. From a practical perspective, an automatic date of withdrawal (DOW) e.g. five years later, could be associated with any AES approval of a new audio format or watermarking based on, for instance, less than 48 kHz/24 bit linear per channel. If the new technology is still considered transparent at that time, a continued approval could be issued.

4.2. Subjective testing

Pre-testing should ensure subjects are entirely familiar with artefacts to be detected and reported, when testing is based on conscious responding (BS.1116); but taking into account the mechanisms reviewed whereby perception of auditory stimuli may be paused, delayed or altered (not yet recognized), it is indicated in subjective pro audio testing to generally be highly aware of:
- listener experience,
- listener attention,
- listening duration.

Experience includes intimate familiarity with the features tested as well as the listening environment. Identical or at least standardized listening rooms should be used.

Attention means allocating perceptual bandwidth fully to hearing and focused listening, or less.

Time means enough to satisfy every experience / learning / fatigue-probing criteria, or less. The time required could depend on whether or not the subject is multilingual or a musician, on his or her age, sound pressure level etc., but three practical categories are suggested: Easy listening, trained listening and slow listening.

4.3. Easy listening

Investigation of topics persons with normal hearing should generally be able to evaluate. For instance, if sound is too loud, voice is intelligible, reproduction is flat or immersive. Besides from understanding a language, there is no need to invoke temporal grey-zone skills (Fig. 5) in easy listening.

4.4. Trained listening

Investigation of topics that require conscious listening with attention, for instance relating to the temporal grey-zone (transients), dynamic balance, spectral balance, assessment of imaging etc.

Because experience plays such a fundamental role for our starting point when subjected to sensory stimuli, listeners should either use a room and equipment they know intimately, or have plenty of time to get to know an acoustic environment before any tests are performed. Based on a limited perceptual bandwidth and 8 hours of dedicated listening per day, getting to know a room and equipment in any detail would take at least a week, but assuming years would be safer.

This would also be compatible with what we have discussed with many audio professionals, and a story multi award-winning mastering engineer Bob Ludwig told at the 2015 AES convention in New York: He avoids listening at different levels because of the time it takes to mentally re-calibrate. Instead, clients are invited to use headphones in case they wish to listen louder.

Elaborate aspects of trained listening are given in [94], and ITU-R BS.1116 provides more extensive practical recommendations with regard to subjects, screening, training etc. than the casual ITU-R BS.1534. However, we might still underestimate the time required for pre and post listening learning and fatigue assessment, or at least do not emphasize the importance of various temporal elements strongly enough.

4.5. Slow listening

For investigation of audio questions of possible long-term influence, we should consider all four quadrants of Fig. 2, including both interior and exterior aspects over time, but especially when examining any quality which may not be completely familiar.

This review found several mechanisms behind long-term effects of sound exposure from myological or neurological *fatigue*, to various factors causing delayed or modulated entry into consciousness, some clearly via long-term *learning*. Slow listening should therefore at least employ the time an experienced listener needs to potentially quantify fatigue, i.e. typically hours under completely known and controlled conditions, including listening level to keep outside the grey sound exposure areas of Fig. 6.

In case what is tested for is unfamiliar, slow listening could take as long as it would for the subject to learn a new language, maybe more. When reaching out, we first need to know of what to reach for; so subjective tests, even producing repeatable results, may have little long-term relevance if confined in time.

4.6. Research agenda

A variety of recent sensory studies have described exteroception as a mainly experience-based, reach-out driven process. Future research should focus on whether Fig. 7, Fig. 8 or a third principle best reflects a general model of human hearing.

Regardless, both temporal aspects of training (grey-zone, 400 ms) and learning (long-term, years) should be taken into account when testing subjectively for certain audio qualities in content and reproduction. Furthermore, what is broadly categorized as listener fatigue, could possibly be an exhaustion of afferent and efferent pathways, or of one of them, with the outcome not necessarily being the same.

However, objective criteria, not immune to falsification, and repeatable procedures need to be established so “more testing time” does not become a way of defending just any claim.

fMRI sensory studies have typically investigated vision where temporal challenges are few, compared to hearing. New windows into human perception will surely still be opened this way, and “big questions” also verified by nonparametric analysis, where less is assumed a priori [95].

Still, there is a limit to the temporal and spatial resolution of fMRI, and stimuli may need repetition for responses to manifest above the test noisefloor. Furthermore, current scanning methods are typically too noisy for examining topics of interest to pro audio.

A physiological, auditory equivalent of saccadic eye movements, but working in the time domain, could be a stepping-stone from the high spatial resolution of vision to the high temporal resolution of hearing. ERPs such as MMN, P300 or CVN could possibly be tuned for dedicated pro audio applications, for example to investigate early auditory processing and neural exhaustion taking place in the brainstem, while satisfying requirements for repeatability.

Using new ERPs and protocols that mask stimuli from auditory awareness, it might be possible instead to show non-conscious recognition. Some specific trials have already been conducted this way [71, 72], but long-term effects should ideally be tested also to investigate learning with potential efferent consequences, possibly showing less or altered efferent neural activity in expert subjects [58].

5. Conclusion

A broad selection of studies on reception, perception and consciousness, focused on hearing and perceptual bandwidth, were reviewed, divided into categories and summarised.

Previous experience, learning and expectations play the major roles in what is recognized when a human adult probes the environment; and several studies propose perception to be the outcome of primarily an active reach-out process, only sometimes including overt behavior.

Experimental results confirm the bandwidth of human perception to be severely limited, to an extent of it being a scarce resource. Hallucinations or mental illness may be understood as a free-wheeling mind with exteroceptive anchoring further depreciated or lost. Still, our primary senses covering space and time, seeing and hearing, under normal conditions offer higher perceptual bandwidth than touch, smell or taste.

Findings have been related to pro audio listening tests, depending generally on explicit consciousness with attention and short duration stimuli; and complementary methods, compatible with studies reported, have been suggested. The strength of evidence is not only down to p value, especially if we are not asking all the right questions; or if we are not factoring-in time everywhere it is potentially of influence.

The title of ITU-R recommendation BS.1116 includes the words “small impairment”, so we know something is not perfect. What may change over time is the personal and collective evaluation of the significance of the impairment – maybe the “small” is not small enough after all. Because of possible learning and fatiguing effects, “slow listening” is especially important when testing systems or technology where qualities could be unfamiliar.

Science relies on empirical data-gathering, repeatability and verification; but it also relies on theory, and a willingness to strike out for new ones, subject to additional measurements and verification [96]. With all the relevant in vivo research on human perception emerging these years, time is right to consider if a more prominent role should be granted that elusive phenomena – time – also in professional audio.

7. References

[1] ITU-R BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems" *ITU-R Recommendation* (2015).

[2] ITU-R BS.1116-3, "Methods for the subjective assessment of small impairments in audio systems" *ITU-R Recommendation* (2015).

[3] Oakley, D. & Halligan, P., “Chasing the Rainbow: The Non-conscious Nature of Being” *Frontiers in Psychology* vol. 14, iss. 8 (2017).

[4] Hameroff, S. & Penrose, R., “Consciousness in the universe. A review of the ‘Orch OR’ theory” *Physics of Life Reviews*, vol. 11, iss. 1 (2014).

[5] Jedlicka, P. “Revisiting the Quantum Brain Hypothesis: Toward Quantum (Neuro)biology?” *Frontiers in Molecular Neuroscience* vol. 10, iss. 366 (2017).

[6] Helmholtz, H. v., "Treatise of physiological optics: Concerning the perceptions in general" *in T. Shipley, Classics in psychology.* (1925, original published 1856).

[7] Wilber, K., “Sex, ecology and spirituality. The spirit of evolution” *Shambhala Books* (1995).

[8] Tønnesvang, J. et al., "The Four Quadrant Model" *Klim Publishing Books* (2015).

[9] Kleineberg, M., “The Blind Men and the Elephant. Towards an Organization of Epistemic Contexts” *Knowledge Organization 40* vol 5, pp. 340-62 (2013).

[10] Boron, W. F. & Boulpaep, E. L., "Medical Physiology" (2nd ed.), *Elsevier* (2011).

[11] Ceunen, E. et al., “On the Origin of Interoception” *Frontiers in Psychology*, no. 7 (2016).

[12] Küpfmüller, K., "Nachrichtenverarbeitung im Menschen" *Springer Verlag, Taschenbuch der Informatik No. 3*, pp. 429-454 (1974).

[13] Lund, T. & Mäkivirta, A., “On the Bandwidth of Human Perception and its Implications for Pro Audio*” in proceedings of the 143rd AES Convention*, NYC (2017).

[14] Cohen, M. A. et al., "What is the Bandwidth of Perceptual Experience?" *Trends in Cognitive Sciences* vol. 20, no. 5, pp. 324-335 (2016).

[15] Lamme, V. A. F., “How neuroscience will change our view on consciousness” *Cognitive Neuroscience* vol. 1, iss. 3, pp. 204-20 (2010).

[16] Shomstein, S. & Yantis, S., ”Control of attention shifts between vision and audition in human cortex” *The Journal of Neuroscience* vol. 24, no. 47, pp 10702-6 (2004).

[17] Johnson, J. A. & Zatorre, R. J., “Neural substrates for dividing and focusing attention between simultaneous auditory and visual events” *NeuroImage* vol. 31, no. 4, pp 1673-81 (2006).

[18] Salo, E. et al., ”Brain activity during auditory and visual phonological, spatial and simple discrimination tasks” *Brain Research* vol. 1496, pp 55-69 (2013).

[19] Moisala, M. et al., “Brain activity during divided and selective attention to auditory and visual sentence comprehension tasks” *Frontiers in Human Neuroscience* (2015).

[20] Smallwood, J. & Schooler, J. W., "The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness" *Annual Review of Psychology* vol. 66, pp. 487-518 (2015).

[21] Giraudet, L. et al., "P300 event-related potential as an indicator of inattentional deafness?" *Public Library of Science,* vol. 10, issue 2 (2015).

[22] Seth, A. K., “A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia” *Cognitive Neuroscience* vol. 5, no. 2, pp 97-118 (2014).

[23] Libet, B. et al., "Subjective referral of the timing for a conscious sensory experience: a functional role for the somatosensory specific projection system in man" *Brain* vol. 102, no. 1, pp 193-224 (1979).

[24] Ray, P. G. at al., "Physiology of perception: cortical stimulation and recording in humans" *Neurology* vol. 52, no. 5, pp. 1044-1049 (1999).

[25] Libet, B., "Timing of conscious experience: reply to the 2002 commentaries on Libet's findings" *Consciousness and Cognition* vol. 12, no. 3, pp. 321-331 (2003).

[26] Pockett, S., "The great subjective back-referral debate: do neural responses increase during a train of stimuli?" *Consciousness and Cognition* vol. 15, no. 3, pp. 551-559 (2006).

[27] Melloni L. et al., "Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness" *The Journal of Neuroscience* vol. 31, no. 4, pp 1386-1396 (2011).

[28] Seibold, V. C. et al., "Temporal attention shortens perceptual latency: a temporal prior entry effect" *Psychophysiology* vol. 48, no. 5, pp. 708-717 (2011).

[29] VanRullen, R. et al., "The blinking spotlight of attention" *PNAS* vol. 104, no. 49, pp 19204-209 (2007).

[30] Bush, N. E. et al., "Whisking mechanics and active sensing" *Current Opinion in Neurobiology* vol. 40, pp. 178-88 (2016).

[31] Helfrich, R. E. et al., "Neural Mechanisms of Sustained Attention are Rhythmic" *Neuron* vol. 99, pp. 854-65 (2018).

[32] Eimes, P. D, et al., "Speech Perception in Infants" *Science*, vol. 171, no. 3968, pp. 301-306 (1971).

[33] Kuhl, P. K. et al, "Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age" *Science*, vol. 255, no. 5044, pp. 606-608 (1992).

[34] Martin, A. et al., "Learning phonemes with a proto-lexicon" *Cognitive Science* no. 37, pp. 103-124 (2013).

[35] Näätänen, R. et al. "Early selective-attention effect on evoked potential reinterpreted" *Acta Psychologica* vol. 42, no. 4, pp. 313-329 (1978).

[36] Marie, C. et al. "Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds" *Cortex* vol. 48, no. 4, pg 447-457 (2012).

[37] Zuk, J. et al., "Enhanced Syllable Discrimination Thresholds in Musicians" *Public Library of Science* vol. 8, no. 12 (2013).

[38] Cooper, A. et al., "Thai Rate-Varied Vowel Length Perception and the Impact of Musical Experience" *Language and speech* vol. 60, no. 1, pp. 65-84 (2017).

[39] Johnsrude, I. S. et al., "Left-hemisphere specialization for the processing of acoustic transients" *Neuroreport* vol. 8, no. 7, pp. 1761-1765 (1997).

[40] Alluri, V. et al., "Connectivity patterns during music listening: Evidence for action-based processing in musicians." *Human Brain Mapping*, vol. 38, no. 6, (2017).

[41] Vuust, P. et al., "Practiced musical style shapes auditory skills" *Annals of the New York Academy of Sciences*, vol. 1252, pp. 139-146 (2012).

[42] Zhao, T. C. & Kuhl, P. K., "Musical intervention enhances infants' neural processing of temporal structure in music and speech" *Proceedings of the National Academy of Sciences of the United States* vol 113, no. 19, pp 5212-5217 (2016).

[43] Bomba, M. D. et al, "Phoneme discrimination and mismatch negativity in English and Japanese speakers" *Neuroreport* vol. 22, no. 10, pp. 479-483 (2011).

[44] Bowers, J. S. et al., "Preserved implicit knowledge of a forgotten childhood language" *Psychological Science* vol. 20, iss. 9 (2009).

[45] Grosso, A. et al., "Auditory cortex involvement in emotional learning and memory" *Neuroscience* vol. 299, pp. 45-55 (2015).

[46] Kujawa, S. G & Liberman, M. C., "Adding Insult to Injury: Cochlear Nerve Degeneration after 'Temporary' Noise-Induced Hearing Loss", *The Journal of Neuroscience* vol. 29, no. 45 pp. 14077-85 (2009).

[47] Viana, L. M. et al., "Cochlear neuropathy in human presbycusis: Confocal analysis of hidden hearing loss in post-mortem tissue" *Hearing Research* vol. 327, pp. 78-88 (2015).

[48] Liberman, M. C. et al., "Towards a differential diagnosis of hidden hearing loss in humans" *PLoS One*, vol. 11 (2016).

[49] Gourévitch, B. et al. "Is the din really harmless? Long-term effects of non-traumatic noise on the adult auditory system", *Nature Neuroscience* vol 15, pp. 483-491 (2014).

[50] Hermann, J. et al., "Synaptic Transmission at the Calyx of Held Under In Vivo–Like Activity Levels", *Journal of Neurophysiology* vol. 98, pp. 807-820 (2007).

[51] Luce, R. D., "Response Times: Their Role in Inferring Mental Organization" *Oxford University Press* (1986).

[52] Gregory, R. L., "Perceptions as hypotheses" *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 290, iss. 1038 (1980).

[53] Hesselmann, G., et al., "Predictive Coding or Evidence Accumulation? False Inference and Neuronal Fluctuations" *PLoS One*, vol. 5 (2010).

[54] Hickok, G., "Predictive coding? Yes, but from what source?" *Behavioral and Brain Sciences*, vol. 36, no. 4 (2013).

[55] Kersten, D. et al., "Object perception as Bayesian inference" *Annual Review of Psychology* vol. 55, pp. 271-304 (2004).

[56] Yarbus, A. L., "Eye Movements and Vision" *Springer* (1967).

[57] Jiang, Y. V. et al., "First saccadic eye movement reveals persistent attentional guidance by implicit learning" *Journal of Experimental Psychology Human Perception & Performance* vol. 4, no. 3 (2014).

[58] Bertram, R. et al., “The effect of expertise on eye movement behaviour in medical image perception” *PLoS One*, vol. 8 (2013).

[59] Friston, K. et al., “Perceptions as hypotheses: saccades as experiments” *Frontiers in Psychology* no. 3 (2012).

[60] Parr, T. & Friston, K., “The active construction of the visual world” *Neuropsychologia* vol. 104, (2017).

[61] Henry, M. J. & Obleser, J., “Frequency modulation entrains slow neural oscillations and optimizes human listening behavior” *Proceedings of the National Academy of Sciences*, vol. 109, no. 49, (2012).

[62] Hickok, G. et al., “The Rhythm of Perception: Acoustic Rhythmic Entrainment Induces Subsequent Perceptual Oscillation” *Psychological Science*, vol. 26, no. 7 (2015).

[63] Southwell, R. et al., “Is predictability salient? A study of attentional capture by auditory patterns” *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 372, iss. 1714 (2017).

[64] Stocker, A. A. & Simoncelli, E. P., “A Bayesian model of Conditioned Perception” *Advances in neural information processing systems* no. 20 (2007).

[65] Peters, M. A. K. et al., “Perceptual confidence neglects decision-incongruent evidence in the brain” *Nature Human Behaviour*, vol. 1, iss. 7 (2017).

[66] Ehinger, B. V. et al., “Humans treat unreliable filled-in percepts as more real than veridical ones” *eLife*, vol. 6 (2017).

[67] Luce, R. D., “Detection and recognition” *Handbook of Mathematical Psychology*, pp. 103-189 (1963).

[68] Lorenz, R. et al., “Neuroadaptive Bayesian Optimization and Hypothesis Testing” *Trends in Cognitive Sciences* vol. 21, no. 3 (2017).

[69] Norman, L. J. et al., "Object-based attention without awareness" *Psychological Science* vol. 24, no. 6, pp. 836-843 (2013).

[70] Plass, J. et al., "Automatic auditory disambiguation of visual awareness" *Attention, Perception and Psychophysics* (2017).

[71] Oohashi, T. et al., “Inaudible high-frequency sounds affect brain activity: Hypersonic effect” *Journal of Neurophysiology*, vol. 83, no. 6 (2000).

[72] Fukushima, A. et al., “Frequencies of Inaudible High-Frequency Sounds Differentially Affect Brain Activity: Positive and Negative Hypersonic Effects” *PLoS One*, vol. 9 (2014).

[73] Allen, A. K. et al., "Entry of involuntary conscious contents from ambiguous images" *Psychology of Consciousness* December, pp. 326-337 (2016).

[74] Reverberi, C. el al., "Deduction without awareness" *Acta Psychologica* vol. 139, no. 1, pp. 244-253 (2015).

[75] Hassin, R., “Yes It Can: On the Functional Abilities of the Human Unconscious” *Perspectives on Psychological Science* vol. 8, iss. 2 (2013).

[76] Chennu, S. et al., “Expectation and Attention in Hierarchical Auditory Prediction” *The Journal of Neuroscience*, vol. 3, pp. 11194-205 (2013).

[77] Barak, O. & Tsodyks, M., “Working models of working memory” *Current Opinion in Neurobiology* vol. 24, pp. 20-24 (2014).

[78] Chong, T. et al., “Recognizing the unconscious” *Current Biology* vol. 24, iss. 21 (2014).

[79] Sela, Y. et al., “Responses in Rat Core Auditory Cortex are Preserved during Sleep Spindle Oscillations” *Sleep* vol. 39, no. 5 (2016).

[80] Andrillon, T. et al., “Formation and suppression of acoustic memories during human sleep” *Nature Communications* vol. 8, no. 179 (2017).

[81] Lawrence, M., "The unconscious experience" *American Journal of Critical Care*, vol. 4, no. 3, pp. 227-232 (1995).

[82] Tapson, K., "'It's the hearing that is last to go': a case of traumatic head injury" *Brain Journal of Nursing* vol. 24, no. 5, pp. 277-281 (2015).

[83] Owen, A. M. et al., “Detecting awareness in the vegetative state” *Science* vol. 313, no. 5792 (2006).

[84] Hampshire, A. et al., “Assessing residual reasoning ability in overtly non-communicative patients using fMRI” *Neuroimage Clinical* vol. 2, pp. 174-183 (2012).

[85] Owen, A. M. & Coleman, M., “Functional neuroimaging of the vegetative state” *Nature Reviews Neuroscience* vol. 9, pp. 235-43 (2008).

[86] Gibson, J. J., “The senses considered as perceptual systems” *Boston: Houghton Mifflin* (1966).

[87] Caliskan, A. et al., “Semantics derived automatically from language corpora contain human-like biases” *Science* vol. 356 (2017).

[88] Siegler, R. S., “Emerging minds: the process of change in children’s thinking” *Oxford University Press* (1996).

[89] Moulton, C. “Perfect pitch reconsidered” *Clinical medicine* vol. 14, no. 5, pp. 517-9 (2014).

[90] Leite, R. B. et al., “Music Proficiency and Quantification of Absolute Pitch: A Large-scale Study among Brazilian Musicians” *Frontiers in neuroscience* vol. 10 (2016).

[91] Deutsch, D. et al., “Absolute pitch among American and Chinese conservatory students: prevalence differences and evidence for a speech-related critical period” *Journal of the Acoustical Society of America* vol 119, no. 2, pp 719-22 (2006).

[92] http://www.spelltool.com/en/index.php

[93] Reiss, J. D., "A Meta-Analysis of High Resolution Audio Perceptual Evaluation" *Journal of Audio Engineering Society* vol. 64, no. 6 (2016).

[94] Bech, S. & Zacharov, N., "Perceptual Audio Evaluation – Theory, Method and Application" *Wiley publications* (2006).

[95] Nichols, T. E. & Holmes, A. P., "Nonparametric permutation tests for functional neuroimaging: A primer with examples" *Human Brain Mapping* vol. 15, no. 1 (2002).

[96] Koopmans, T. C., "Measurement without theory" *The Review of Economic Statistics*, vol. 29, no. 3, pp. 161-72 (1947).

Illustrations

Figures 3, 4, 5, 7, 8 original artwork by the authors. Citation granted with declaration of source.