Time for Slow Listening

THOMAS LUND, *AES Member*, **AKI MÄKIVIRTA,** *AES Fellow*, **AND**
(thomas.lund@genelec.com) (aki.makivirta@genelec.com)

SIAMÄK NAGHIAN, *AES Member*
(siamak.naghian@genelec.com)

Genelec OY, 74100 Iisalmi, Finland

Conscious perception is 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, at least in adults. Considering human hearing, time is a deciding factor on several scales, and the sensory information flow rate, otherwise termed the perceptual bandwidth, is modest. We introduce the term “slow listening” and discuss how new findings from other fields of science should be taken into account in pro audio, for instance when conducting subjective tests, and when preserving content for future generations to enjoy.

INTRODUCTION

Are we passive receivers of sensory information from the environment or actively collecting it? A recent review on ways human perception is affected by time, unexpectedly led us to findings that require discussion from a specific pro audio perspective. For a full list of references, see [1].

Contemporary trials from several scientific fields conclude that our perceptual bandwidth is lower than we generally tend to believe. Instead, learning, experience, and temporal reach-out phenomena, for instance movement and rhythm, play major roles in perception.

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. However, subjective audio testing puts high faith in that seemingly scarce resource, perceptual bandwidth.

Thanks to new non-invasive in vivo experimental techniques, we are getting a better understanding of how *time* can affect sensing in various ways. 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 [2].

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

INTERIOR AND EXTERIOR

Even thousands of years before antiquity, our forefathers were aware of sensing, its importance, and some of 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. 1a.

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

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 has reverberated even into quantum physics and from there to wild hypotheses on consciousness and brain topology today [3, 4]. Nineteenth 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

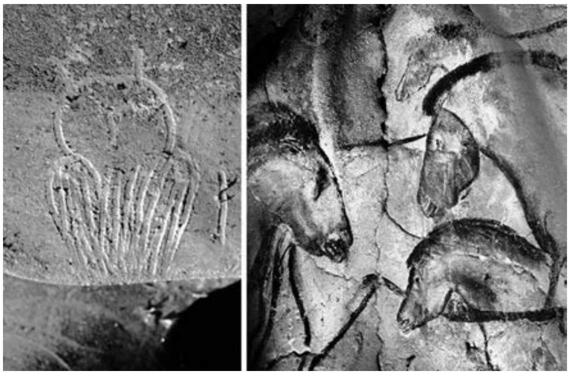

Fig. 1a, 1b. Illustrations from Grotte Chauvet, France.

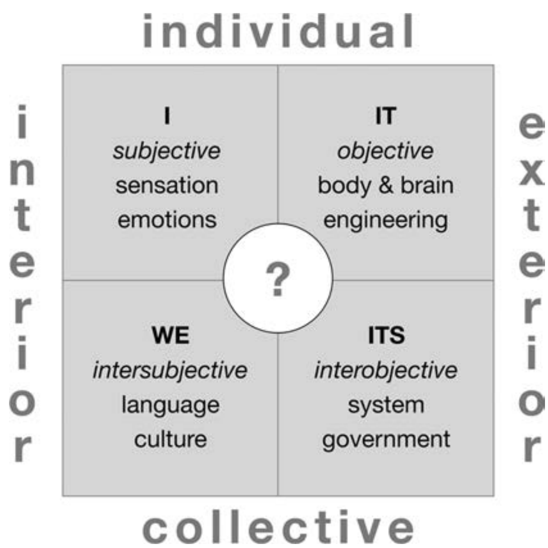

Fig. 2. The quadrant model. Left side is interior/subjective. Right side is exterior/objective.

intermediate results that produce sensation [5]. He called the phenomenon "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.

INDIVIDUAL AND COLLECTIVE

In our era, where extensive data about almost any highly specific topic is instantly available, the combination of discoveries from different disciplines remains essential, for example to propose and to test new hypotheses. 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 [6]. 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 [7, 8], Integrative Perspectivism has proven a valuable tool for stimulating discussion in international standardization involving stakeholders from different fields; for instance, engineers, physicians, psychologists, artists, economists, and politicians. 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.

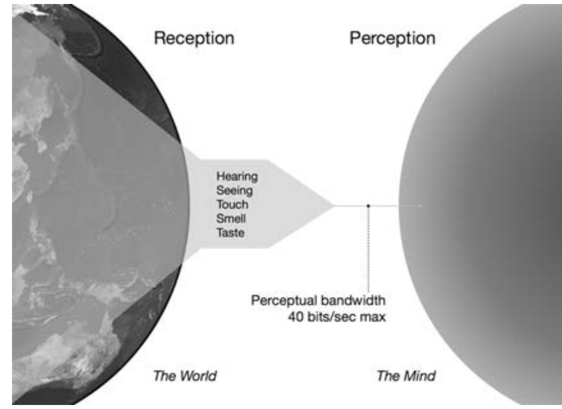

Fig. 3. The two funnels of human apprehension.

TERMINOLOGY

We use *conscious* 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.

A plethora of bodily receptors generate *afferent* nerve impulses about its status, as well as our close and more distant surroundings. *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, haptic), a certain electromagnetic frequency range (seeing, touch), certain smell/taste dimensions, etc.

Considering hearing, the brain 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 also makes use of a substantial number of *efferent* nerve fibres [9]. Such fibres send information back to the middle and inner ears, significantly adjusting the reception system itself.

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. *Perceptual bandwidth* is the rate by which we can (consciously, at the moment) register sensory stimuli. It measures the second funnel between the exterior world and consciousness. The review investigated Karl Kupfmüller's observation of a modest upper limit for human sensory information flow, "Nachrichtenfluss" [10].

SUMMARY OF REVIEW

Despite improved non-invasive, in vivo measurement techniques, we still have an incomplete understanding of human perception. However, consciousness is known to not have access to all data and intermediate results that produce sensation; where delayed entry into consciousness can have a wide variety of causes.

Attention is used to examine different aspects of sensing, mono- or multimodal, but the second funnel of Fig. 3 was confirmed to be steep. It is also not continuous, but time-modulated in ways not fully understood.

Fatigue, age, and other factors may limit the perceptual bandwidth further, and overwhelming or threatening situations can block most or all exteroception from entry into consciousness. However, phenomenal consciousness, what it is like to have an experience, has higher bandwidth than explicit consciousness, what can be reported.

Expectations, previous learning, and experience makes up most of adult perception. There is some evidence that this may be realized through a hierarchy of neural processes in which forecasts sent backward from higher levels result in prediction errors that are fed forward from lower levels, thereby updating the current model of the environment. Under such a regiment, even partly, constantly reaching out to our surroundings for verification and updates would be essential for maintaining a useful mental map and, therefore, for survival. Reach-out behavior in vision has the potential to change and improve gradually over time with learning and expertise, and overt behavior in animals has confirmed this to be the case for other sensory modalities as well.

Without training, humans generally do not understand or even recognize short stimuli inside a 400 ms temporal grey-zone, phoneme discrimination being one example. Decoding of short duration sounds in music and speech makes use of some of the same temporal grey-zone capabilities and cortical areas of the brain. Like language learning, training to recognize short duration sounds at a young age is most efficient, but brainstem and auditory cortex plasticity can be preserved even in old people.

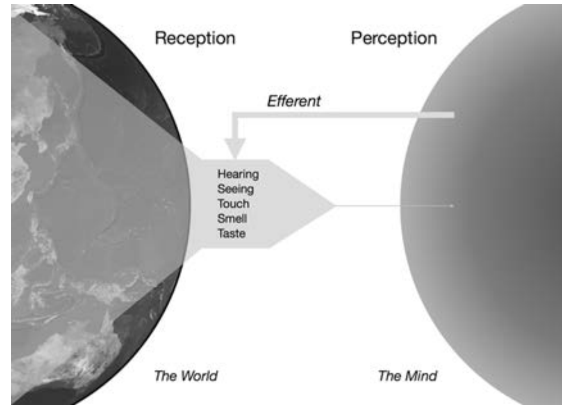

Fig. 4. Exteroception is highly influenced by experience and efferent neural activation.

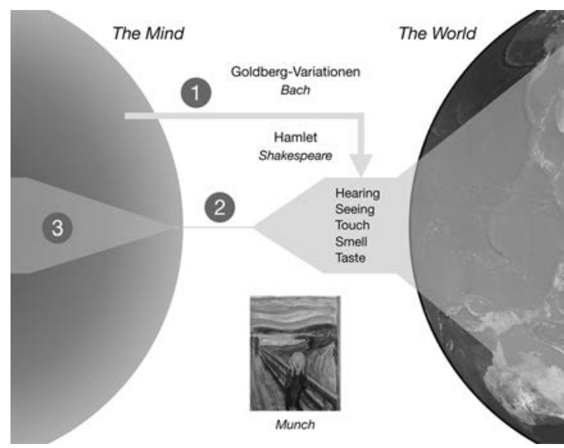

Fig. 5. Perception illustrated as a reach-out (1), modest return (2) phenomenon; art being exceptional (3).

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 components of a sensory experience?Recent studies suggest that we are and how Fig. 4 would be a more relevant representation of human perception than Fig. 3. Thanks to higher temporal resolution brain imaging techniques, evidence is also mounting that perception is predominantly an active process and that it could be driven mainly by exteroceptive prediction errors. Thus, Fig. 5 would appear even more appropriate than Fig. 4.

As organisms, we need to 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. Active sensing features of hearing includes body and head movements.

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 artifacts 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 can be biased, but some of our knowledge is still tacit and much of that relies on personal sensing.

We consequently owe children reasonably unlimited sensory examples so they can fully develop sentient faculties including relevant movements. Children do not just learn; they reach curiously out to the world and also learn how to learn [11].

Considering hearing, short-duration phonemic contrasts outside our mother-tongue are difficult to learn once we grow up. Another aspect of children's learning should 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.

One way to define "art," based on the topics discussed here, would be its exceptional ability to make perceptual bandwidth seem to widen explosively in the receiver, for instance when listening to Bach, seeing a Munch painting or reading Shakespeare, Fig. 5. Looking at Fig. 1b, 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.

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; potentially including crude machine learning ("AI") algorithms.

IMPLICATIONS FOR AES

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 [12]. The same kind of tests, however, have also been used to promote lossy data reduction, where most audio information is discarded, though anyone interested in sound today notice warbling "space monkey" artifacts and collapsed imaging across platforms, be it broadcast, YouTube, music streaming or phone. That kind of artifact might be experienced more gravely now than when the codecs were originally tested.

At least *three* temporal time-scales—the 400 ms grey-zone, auditory fatigue, and long-term learning effects—should be taken into account in audio standardization, so our society is not used to rubber-stamp vulgarization 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 DOW, a continued approval may be issued.

With regard to subjective testing, pre-qualification must ensure subjects are entirely familiar with artifacts to be detected and reported when testing is based on immediate conscious responding; but it is indicated to generally be highly aware of:

- Listener experience,
- Listener attention,
- Listening duration.

Experience includes intimate familiarity with the features tested and the listening environment.

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.

Easy Listening for investigation of topics people should generally be able to evaluate. For instance, if sound is too loud, voice is intelligible, or reproduction is flat or immersive. Besides from understanding a language, there is no need to invoke temporal grey-zone skills.

Trained Listening when investigating topics that require conscious listening with attention, for instance relating to the temporal grey-zone (short duration sounds), 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 eight 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.

Trained listening has been emphasized in literature, for instance [13], but we might still underestimate the time required for pre and post listening learning and fatigue assessment, or at least do not observe the importance of the various temporal elements strongly enough.

Slow Listening is used for investigating audio questions of possible long-term influence, where all four quadrants of Fig. 2 have to be considered.

Slow listening should 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. 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 too confined in time.

CONCLUSION

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 [14]. The strength of evidence in subjective pro audio testing is clearly not only down to p value, especially if all the right questions are not asked, or if we are not factoring-in time everywhere it potentially is of influence.

Despite continued research on human perception, our own inner workings may never be fully understood, let alone the "hard problem" of consciousness. However, new insight from all fields of science can and should be used to allow future generations the best possible experience of music and other sound art created today.

We must therefore now consider if a more prominent role should be systematically granted to that elusive quality—time–also in pro audio evaluation and testing. To that end, our society should also prevent the proliferation of time-frozen algorithms with a bearing on human perception and sentience from taking hold in production or distribution.

REFERENCES

[1] T. Lund and A. Mäkivirta, "On Human Perceptual Bandwidth and Slow Listening," *Proceedings of Tonmeistertagung*, AES reviewed paper, Cologne (2018). ISBN 978-3-9812830-9-9.

[2] D. Oakley and P. Halligan, "Chasing the Rainbow: The Non-Conscious Nature of Being," *Frontiers in Psychology,* vol. 14, no. 8 (2017 Nov.). https://doi.org/10.3389/fpsyg.2017.01924

[3] S. Hameroff and R. Penrose, "Consciousness in the Universe. A Review of the 'Orch OR' Theory," *Physics of Life Reviews*, vol. 11, no. 1 (2014). https://doi.org/10.1016/j.plrev.2013.08.002

[4] P. Jedlicka "Revisiting the Quantum Brain Hypothesis: Toward Quantum (Neuro)biology?" *Frontiers in Molecular Neuroscience* vol. 10, no. 366 (2017). https://doi.org/10.3389/fnmol.2017.00366

[5] H. v. Helmholtz, "Treatise of Physiological Optics: Concerning the Perceptions in General," in T. Shipley, *Classics in Psychology* (1925, original book published 1856).

[6] K. Wilber, *Sex, Ecology and Spirituality. The Spirit of Evolution* (Shambhala Books, 1995).

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

[8] M. Kleineberg, "The Blind Men and the Elephant. Towards an Organization of Epistemic Contexts," *Knowledge Organization,* vol. 40, no. 5, pp. 340–62 (2013). https://doi.org/10.5771/0943-7444-2013-5-340

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

[10] K. Küpfmüller, "Nachrichtenverarbeitung im Menschen," Springer Verlag, Taschenbuch der Informatik No. 3, pp. 429–454 (1974). https://doi.org/10.1007/978-3-642-65588-3

[11] R. S. Siegler, "*Emerging Minds: The Process of Change in Children's Thinking* (Oxford University Press, 1996).

[12] J. D. Reiss, "A Meta-Analysis of High Resolution Audio Perceptual Evaluation," *J. Audio Eng. Soc.,* vol. 64, pp. 364–379 (2016 Jun.). https://doi.org/10.17743/jaes.2016.0015

[13] S. Bech and N. Zacharov, *Perceptual Audio Evaluation–Theory, Method and Application* (John Wiley & Sons, 2006). https://doi.org/10.1002/9780470869253

[14] T. C. Koopmans, "Measurement without Theory," *Review Econ Stat.*, vol. 29, no. 3, pp. 161–72 (1947). https://doi.org/10.2307/1928627