Traces of Sound

Reflections of Sounds Unheard

Henrik Frisk and Sanne Krogh Groth (eds.)

Lund University

PUBLICATIONS FROM THE
SOUND ENVIRONMENT CENTRE
AT LUND UNIVERSITY REPORT NO. 20

This publication can be ordered through Lund University at:
https://www.ht.lu.se/serie/ljud. Email: skriftserier@ht.lu.se

Cover photo by Jacob Kirkegaard from a field recording
for his work Crossfire (2022).

Publications from the Sound Environment Centre
at Lund University No. 20
ISBN 978-91-89874-19-0 (tryck)
978-91-89874-20-6 (digital)
DOI https://doi.org/10.37852/oblu.255
ISSN 1653-9354

Typesetting by Gunilla Albertén, Media-Tryck, Lund University
Printed in Sweden by Media-Tryck, Lund University, Lund 2024

Nordic Swan Ecolabel 3041 0903

Contents

Affective Touch and the Auditory Envelopment Hypothesis

Thomas Lund

In this essay I develop one aspect of 'slow listening', a concept put to the Audio Engineering Society (AES) and Tonmeistertagung conferences (Lund & Mäkivirta 2018), and which resulted in the hypothesis that auditory envelopment has physiological commonalities with affective touch.

Thanks to non-invasive in vivo experimental techniques, we have a better understanding of human perception, including the verification or rejection of psychological theories. The big questions are being revisited, such as how much we actively seek stimuli rather than receiving them, the many ways in which time can affect sensation, the distinctions between dynamic and static conditions, pattern recognition, unexpected sensory connections, or identifying signalling mechanisms in the body that complement neural pathways, to name a few. The immediate sensory recognition of external stimuli was an obvious place to start when we began investigating ourselves, and that is still the premise of most listening tests, relying for example on ABX comparisons (ITU 2015). However, the sensory system is constantly bombarded with competing information, leading to a data deluge that cannot be processed in real time. The brain is no larger or more energy-consuming than it absolutely must be, so an efficient way of cutting down on both size and energy is to rely more on prediction and less on sensory input. This is now an accepted model of perception—and

https://doi.org/10.37852/oblu.255.c613

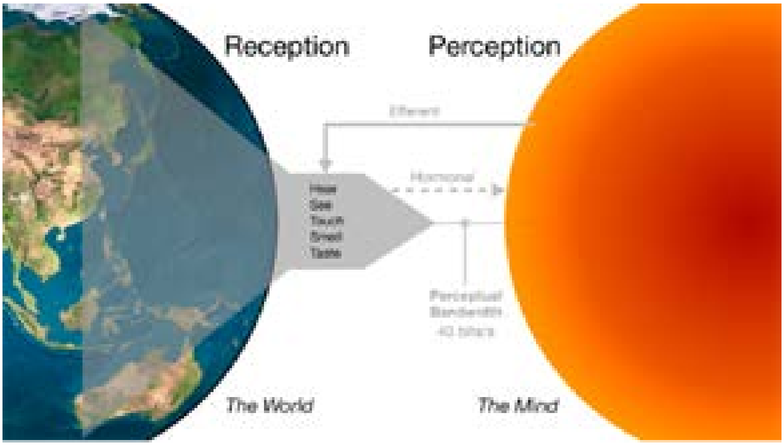

FIGURE 1 The five primary senses involved in exteroception, including the two funnels of human apprehension, afferent pathways and a modest perceptual bandwidth, efferent pathways, and hormonal longer-term modulation, the latter shown shown as a dotted line.

a recognized difference between children and adults (Gregory 1980; Siegler 1996; Friston 2012; Parr 2021). Therefore, we have to distinguish between the *reception* of stimuli, and how it is *perceived*, personally and dynamically.

Reception has a steep funnel associated with it (Fig. 1). Our sensory organs are attuned to conditions that pertain on Earth (the relevant mechanical waves, appropriate electromagnetic spectrum, suitable change-rates, etc.) and generally matter for the survival of a creature our size and composition, with our lifespan. Perception is distinguished from reception and based on a second funnel between the exterior world and our mental notion of it (Fig. 1).

Perception is entirely subjective. It is the outcome of sentient brain processing based on experience, expectation, mood, attention and—to some extent only—reception, and the perceptual bandwidth is known to be surprisingly low from a number of studies based on diverse methodologies (Lund & Mäkivirta 2018). New evidence for a human perceptual bandwidth of only around 40 bits/s is also given in a study of oral information rate across 17 languages (Coupé 2019) as 'biology

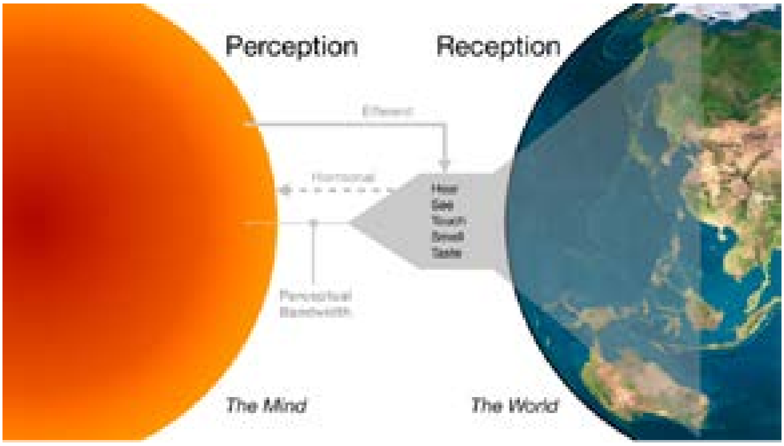

FIGURE 2 Human perception in adults in an updated model where active inference plays a major role. We are not passive receivers; we primarily reach out, based on prior experience and expectation.

under communicative pressures'. Due to the low perceptual bandwidth, perception is primarily a reach out, driven process (active inference) in humans and animals alike; sometimes, but not always, including overt behaviour (eye movements, sniffing, whisking). The brain is also a highly active participant in hearing, not only in the decoding of minute temporal information, but also as the main element of a sense that relies systematically on dynamic adjustments, comparable to saccadic movements of the eye (Friston 2012). In hearing, we overtly adjust head and body position when listening, but we also invisibly make use of a substantial number of efferent nerve fibres. Efferent fibres in the auditory nerve send information back to the middle and inner ears, adjusting each ear of the reception system itself, frequency selectively over a range of more than 60 dB. The active aspects of sensing (Fig. 2) say more than the traditional picture of reception (Fig. 1).

The systemic conditioning of the body may be driven by hormones rather than by neural pathways—for example. by the release of monoamines such as 5-hydroxytryptamine/serotonin, dopamine, melanin, or histamine into the tissues and bloodstream—by circadian rhythms, or by reward and pleasure responses by the brain. In a recent review (Lund & Mäkivirta 2018), we proposed the term 'slow listening' and with it the appropriate procedures for any attempt at a comprehensive, subjective assessment of sound, to complement untrained and trained listening, used in standard testing. Trained listening, for example, is required when detecting and understanding short-duration sounds, such as the phonemes or syllables of a particular language. Time thus plays a determining role in sensation, imposing formative changes on a lifetime scale, but it can also have an influence on intermediate timescales, where the word 'feeling' might be more appropriate than the immediacy of 'sensation'.

Sense of change

Our senses are tuned towards detecting change and movement, especially as identifiable patterns. We separate static from dynamic on timescales that are relevant to us; for example, to distinguish other agents from the scenery. Physiologically, the senses have change thresholds and/or filtering. The Eustachian tube is an example of such an ancient mechanical filter, focusing hearing on the kind of dynamic pressure changes that matter most in an evolutionary perspective. The same is true of eye physiology. We are highly sensitive to certain movements, and 99 per cent of our visual field (outside the fovea) is specialized in retaining important functionality in low light (scotopic vision). In a recent study, the responsible rod receptors in the retina are also found not to sacrifice detection of movement, while obtaining more light sensitivity than the cone receptors of the fovea (Field 2019).

From an evolutionary perspective, the next step is to determine if change and movement are threatening or benign. At close range we use touch for such assessments, and a class of low-threshold C-tactile mechanosensitive skin fibres are the biological substrate for the newly discovered affective and rewarding properties of touch, the Sahlgrenska Academy in Sweden being global pioneers in the field. C-tactile receptors respond to skin stroking of just the right force, speed and temperature, triggering a rewarding reaction from the striatal cluster in the brain. This causality is now widely accepted and regarded a potentially socializing mechanism of friendly touch (Vallbo 1993; Olausson 2010; McGlone 2014).

However, permitting other living beings to be intimately close to confirm their friendliness is risky, so we naturally use additional cross-modal indicators to evaluate movement, behaviour, and motives in others. Hearing is involved when interpreting sounds and words, but as described below, the auditory system might possibly also detect 'Goldilocks conditions' comparable to friendly touch, which might serve a soothing and socializing purpose too.

Envelopment

Listener envelopment (LEV) has an agreed meaning in acoustical engineering, but that is not how the term is used in this essay. Here, envelopment is a definable sensation or feeling, requiring a human listener to be part of the equation. There may also be a measurable, biochemical component to encountering it, for example a local or systemic release of monoamines.

Natural places where we tend to experience envelopment include sharing a large acoustical space with an orchestra or choir, where the concert hall or the church is part of the listening experience. The swirling sound patterns created in such pristine spaces were studied and described decades ago (for example, Griesinger 1998), and noted by the author when responsible for widely recognized reverb designs, used with music for playback in relatively small reproduction rooms. On those occasions when envelopment is felt outdoors, observable movement is generally involved: breaking waves when standing on a seashore, trees in a forest, or falling rain. Generally, though, we do not come across envelopment outdoors, where the norm is sound from uncorrelated, discrete sources located at different places in 3D space. Floyd Toole has recently suggested[1] using 'immersive' to describe an

1 Private Communication with Floyd Toole. Oak Park, CA. January 2020.

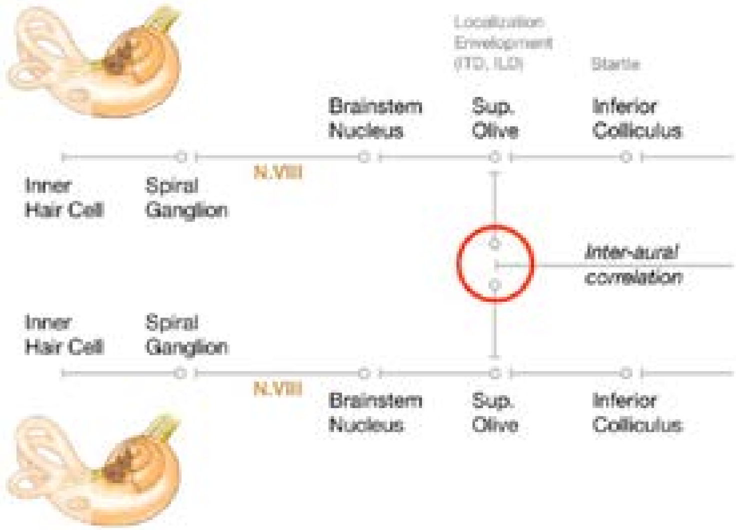

FIGURE 3 Elements of the early, primary auditory pathway, connecting to the thalamus and auditory cortices. Fast interaural synapses of the brainstem (e.g. calyx of Held) are illustrated by the red circle.

outdoor sound scenario, which is often how sound for cinema or drama is produced, but 'enveloping' when qualifying an in-door music recording.

The early auditory pathway of the brainstem has several ahead-of-consciousness qualities associated, for example, startle reflex, localization, proximity, and detection of change (Fig. 3). The inter-aural correlation nodes include the fastest firing synapses of the body (Hermann 2007), up to 700 Hz or 20–50 times the norm, with increased fidelity at low frequencies (Kaczmarek 2023), and they remain largely active also during sleep.

Such refined, energy-hungry anatomical structures must have good evolutionary reasons, which could be to detect aural patterns of movement precisely and quickly, possibly pre-categorizing them as friendly or threatening. If this were the case, it would be reasonable to theorize that it influences behaviour from positive (bonding) via neutral (relaxation) to negative (hostile or mobilization), with a suitable urgency.

Threat reactions would be triggered by fast internal signalling (neural pathways), while positive assessments of just the right type could be slower with systemic conditioning (hormonal), such as in response to friendly touch. Besides a general understanding of song and music as signs of friendliness or bonding, specialized physiology could contribute to such ubiquitous human reactions associated with music (Salimpoor 2011), or possibly even be their foundation.

Because of the way acoustics work, singing or playing music indoors is an effective way of promoting inter-aural fluctuation patterns in listeners. How such patterns move around, as a result of varying pitch, depends on the acoustical space, but classic concert halls all share this quality when music is played. Most listeners also find that kind of spaces remarkably stimulating, and there is experimental, qualitative evidence that our striatal system indeed engages when we listen to music or sing in certain indoor conditions—when envelopment is experienced. However, music in general has also been shown to trigger striatal responses in listeners (Zatorre 2013), so our experiments were done with neutral, non-music, low frequency (LF) test signals to investigate listener responses when subjected to varying degrees of inter-aural correlation. Further, tests were done with untrained as well as trained listeners.

We found everyone could hear a difference, and even children aged 6 associated correlated LF sound with a small space and uncorrelated LF sound with a larger space. In one experiment, delivering audio using open Sennheiser HD600 headphones, seven children aged 6–9 were asked if they could tell a difference, and then explain what they heard in their own words. They all could hear a difference and used words with similar meanings, such as 'small' or 'locked up' (Dan. *lille, spærret inde*) about correlated LF stimuli and 'large' or 'free' (Dan. *stor, fri*) about uncorrelated, abstract LF stimuli. Additional studies are required, but qualitative results from experienced listeners and the pilot study with children (Fig. 4) suggest humans from an early age naturally distinguish between three classes of enveloping auditory stimuli between 20 and 700 Hz, noted across all test subjects: (1) a high, static LF correlation, associated with a small listening space; 2)

FIGURE 4 Child aged 6, listening to a selection of loudness-normalized, correlated and uncorrelated abstract LF sounds.

a low, dynamic random LF correlation, associated with movement and/or a large, indoor listening space; and (3) a low, dynamic pattern LF correlation, again associated with movement and/or a large, indoor listening space.

The prerequisites for a hall or test signal to generate a pronounced feeling of envelopment appear to be a low interaural correlation at low and very low frequencies at the listening position, with a sensation of identifiable, moving patterns possibly amplifying the effect further. In the case of dynamic patterns, to maximize the feeling of envelopment, the perceived velocity of movement should likely remain within the 'Goldilocks' conditions at least in terms of velocity and strength. Further studies are required, but angular velocities around 0.5 rad/s have been used successfully in pilot studies.

The research agenda

Based on the present hypothesis, reactions to auditory envelopment are a potential confounder when the striatal effects of music listening are investigated, so the two should be studied separately. To investigate the outcome when subjects are only experiencing pleasantly dynamic interaural LF conditions, non-music, 20–700 Hz interaural test signals should be used as stimuli. In addition to self-reporting and behavioural assessment, dynamic (long-term) brainstem and/or brain responses can be used as objective measures of the physiological response, for example by means of functional near-infrared spectroscopy (fNIRS) or functional MRI (fMRI). The latter, however, has proven a challenge, due mainly to interfering noise generated by the scanner itself. If stimuli are administered through over-ear, on-ear, or in-ear headphones, care must be taken to restore an LF and VLF influence comparable to a natural in-room experience, for example similarly stimulating both aural and haptic pathways. If stimuli are administered via an in-room loudspeaker system, it is important not to distort test signals significantly by the reproduction room acoustics, which is not trivial with small rooms and LF test signals (Fig. 5).

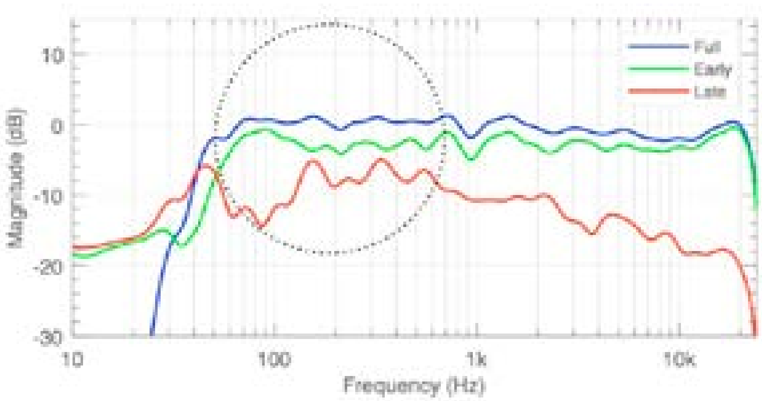

FIGURE 5 Test signals investigating audibility or reaction to potentially enveloping stimuli must ensure the reproduction room does not distort the stimulus, especially in the important 50-700 Hz range (the dotted circle). Measurement for a given loudspeaker and subject placement in a particular room. Genelec GRADE acoustical report.

More than just music

For thousands of years people have sung or played music indoors, whether in caves, cathedrals or concert halls, and it has a recognized potential for stimulating listeners. There are, however, theoretical reasons and qualitative evidence that the sensation or feeling of envelopment in itself, and not only the music, could be a factor in striatal engagement and therefore an important quality to preserve in the recording, distribution, and reproduction of audio.

The hypothesis may be further studied using quantitative subjective trials and a mix of music and test sounds, or by the objective measurement of activity in the striatal system. If proven, certain auditory patterns of 'friendly movement' are not just personally stimulating, they could also promote bonding and social behaviour.

References

Coupé, C., Y. M. Oh, D. Dediu & F. Pellegrino (2019), 'Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche', *Science Advances* 5. doi: 10.1126/sciadv.aaw2594

Field, G. D., V. Uzzell, E. J. Chichilnisky & F. Rieke (2019), 'Temporal resolution of single-photon responses in primate rod photoreceptors and limits imposed by cellular noise', *Journal of Neurophysiology* 121, 255–268. doi: 10.1152/jn.00683.2018

Friston, K., R. A. Adams, L. Perrinet & M. Breakspear (2012). 'Perceptions as hypotheses: Saccades as experiments', *Frontiers in Psychology* 3. doi: 10.3389/fpsyg.2012.00151

Gregory, R. L. (1980), 'Perceptions as hypotheses', *Philosophical Transactions of the Royal Society B: Biological Sciences* 290(1038).

Griesinger, D. (1998), 'General Overview of Spatial Impression, Envelopment, Localization, and Externalization', paper given at the 15th AES International Conference on Audio, Acoustics & Small Spaces (Oct.), paper 15-013. www.aes.org/e-lib/browse.cfm?elib=8095

Hermann, J., M. Pecka, H. von Gersdorff, B. Grothe & A. Klug (2007), 'Synaptic transmission at the calyx of Held under in vivo-like activity levels', *Journal of Neurophysiology* 98(2), 807–820. doi: 10.1152/jn.00355.2007

ITU (International Telecommunication Union) (2015), 'Recommendation BS.1116: Methods for the subjective assessment of small impairments in audio systems' (Feb.), updated 18 July 2023. www.itu.int/rec/R-REC-BS.1116

Kaczmarek, L. K. (2023), 'Modulation of potassium conductances optimizes fidelity of auditory information', *Proceedings of the National Academy of Sciences* 120(12). doi: 10.1073/pnas.2216440120

Lund, T. & A. Mäkivirta (2018), 'On Human Perceptual Bandwidth and Slow Listening', paper presented at the 2018 AES International Conference on Spatial Reproduction: Aesthetics and Science (July), paper P6-2.

McGlone, F., J. Wessberg & H. Olausson (2014), 'Discriminative and Affective Touch: Sensing and Feeling', *Neuron* 82(4), 737–755. doi: 10.1016/j.neuron.2014.05.001

Olausson, H., J. Wessberg, I. Morrison, F. McGlone & Å. Vallbo (2010), 'The neurophysiology of unmyelinated tactile afferents' *Neuroscience & Biobehavioral Reviews* 34, 185–191. doi: 10.1016/j.neubiorev.2008.09.011

Parr, T., N. Sajid, L. Da Costa, M. B. Mirza & K. J. Friston (2021), 'Generative Models for Active Vision', *Frontiers in Neurorobotics* 15. doi: 10.3389/fnbot.2021.651432

Salimpoor, V. N., M. Benovoy, K. Larcher, A. Dagher & R. J. Zatorre (2011), 'Anatomically distinct dopamine release during anticipation and experience of peak emotion to music', *Nature Neuroscience* 14, 257–262. doi: 10.1038/nn.2726

Siegler, R. S. (1996), *Emerging minds: The process of change in children's thinking* (Oxford: OUP).

Vallbo, A., H. Olausson, J. Wessberg & U. Norrsell (1993), 'A system of unmyelinated afferents for innocuous mechanoreception in the human skin', *Brain Research* 628. doi: 10.1016/0006-8993(93)90968-s

Zatorre, R. J. & V. N. Salimpoor (2013), 'From perception to pleasure: Music and its neural substrates', *Proceedings of the National Academy of Sciences* 100(supp. 2), 10430–10437. doi: 10.1073/pnas.1301228110