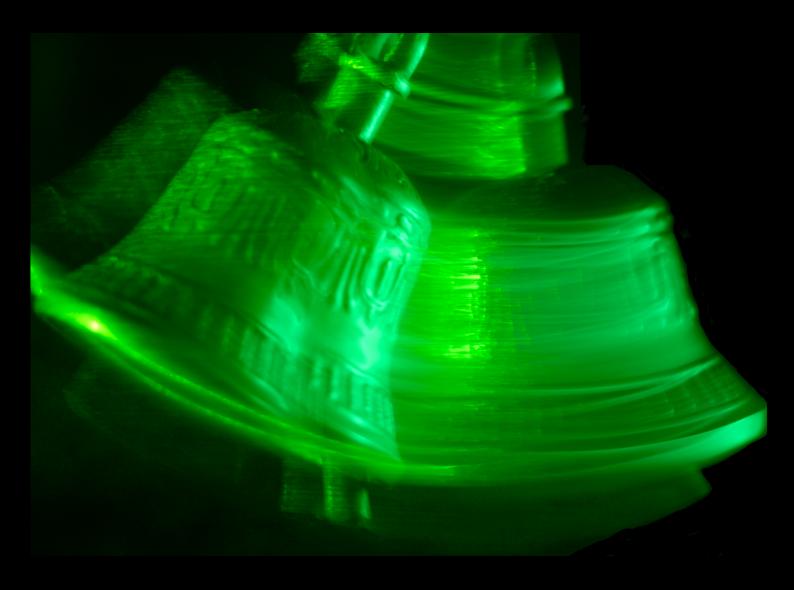# Treatise on Hearing
The Temporal Auditory Imaging Theory
Inspired by Optics and Communication

Adam Weisser

Treatise on Hearing:
The Temporal Auditory Imaging Theory
Inspired by Optics and Communication

Adam Weisser, PhD

June 2, 2024

All photographs and illustrations are by Adam Weisser,
unless credited otherwise.

v0 published September 2021
v1, v2, v3 published November 2021
v4 April 2022
v5 May 2022
v6 May 2023
v7 this version May 2024

Contact the author at weisser@f-m.fm

Abstract

Contrary to traditional thinking about hearing in which the broadband audio spectrum is taken as a whole, modern hearing science has gradually uncovered how the channel-based temporal envelope and its own spectrum are often prioritized by the auditory system. This is achieved through various processing mechanisms at different stages between the auditory brainstem and cortex, which operate on the temporal envelopes both within single auditory channels and between channels of different frequencies. Without loss of generality, it is possible to formulate the temporal envelope as a complex function that varies slowly around a fast center carrier frequency. The complex envelope includes all frequency and amplitude modulations, and hence includes the signal onset and offset cues, by definition. Tracking the transformations that the complex envelope undergoes between the acoustic source and the listener's brain should therefore be one of the key points of hearing theory. However, no systematic treatment of the complex envelope transformations relevant to hearing exists. Rather, only fragmentary treatments are available that primarily rely on empirical findings that pertain to particular stages of hearing.

The new theory of mammalian hearing that is presented here attempts to bridge this gap in the science by consulting the two disciplines that offer the most extensive analytical tools that deal with complex envelope transformations. The first one is imaging optics, which deals with the spatial envelope that propagates between an object and an image and undergoes diffraction and refraction—as is the basis for vision. The second is communication theory, which devises various types of temporal modulations to transfer information between a receiver and a transmitter, over a noisy channel.

Drawing from optical physics, it is argued that an auditory image is formed in the midbrain (inferior colliculus) of an object that is located in the acoustical environment of the listener. Using the space-time duality, it is shown that the ear is a temporal imaging system that comprises three transformations of the envelope functions: cochlear group-delay dispersion, cochlear time lensing, and neural group-delay dispersion. These elements are analogous to the familiar transformations from the visual system of diffraction between the object and the eye, spatial lensing by the crystalline lens, and second diffraction between the lens and the retina. However, unlike the eye, it is established that the human auditory system is naturally defocused, so that coherent stimuli do not react to the defocus, whereas completely incoherent stimuli are impacted by the defocus and may be blurred by design. It is argued that the auditory system can use this differential focusing to enhance or degrade the images of real-world acoustical objects that are partially coherent, predominantly. In addition to the imaging transformations, the corresponding inverse-domain modulation transfer functions are derived and interpreted with consideration to the nonuniform neural sampling operation of the auditory nerve. These ideas are used to rigorously initiate the concepts of sharpness and blur in auditory imaging, auditory aberrations, and auditory depth of field.

In parallel, ideas from communication theory are invoked to show that the organ of Corti functions as a multichannel phase-locked loop (PLL) that constitutes the point of entry for auditory phase locking. It provides an anchor for a dual coherent and noncoherent auditory detection further downstream in the auditory brain. Phase locking enables conservation of coherence between the

mechanical and neural domains.

Combining the logic of both imaging and phase locking, it is speculated that the auditory system should be able to dynamically adjust the proportion of coherent and noncoherent processing that comprises the final image or detected product. This can be the basis for auditory accommodation, in analogy to the accommodation of the eye. Such a function may be achieved primarily through the olivocochlear efferent bundle, although additional accommodative brainstem circuits are considered as well.

The hypothetical effect of dispersion and synchronization anomalies in hearing impairments is considered. While much evidence is still lacking to make it less speculative, it is concluded that impairments as a result of accommodation dysfunction and excessive higher-order aberrations may have a role in known hearing-impairment effects.

Summary

Below is an informal summary that provides a concise and broad overview of the main ideas found in this work.

Vision and hearing

Out of the five traditional senses—vision, hearing, touch, taste, and smell—hearing and vision superficially share the most in common—something that has led to recurrent juxtapositions and comparisons over millennia. For start, the peripheral organs themselves, the eyes and the ears, are placed in proximity and at similar height on the human face, they both come in pairs, and both provide near-nonstop information from the distance about the immediate and remote environments. On top of that, both are central for communication and both are used expressively in many art forms.

As the understanding of the senses has matured over the last two centuries, additional characteristics have stood out, the main one being that both hearing and vision are based on physical wave stimuli that radiate toward the body—sound or light waves— albeit at very different characteristic speeds, wavelengths, and frequencies. With the advent of psychophysics, analogies between visual and auditory perception were made clear too, which occasionally turned out to have parallels in the respective structure of the relevant brain area or its presumed method of processing the stimuli.

The human eye works like a camera, complete with an object, a lens, a pupil to limit the amount of light, and a screen that is the retina, on which an optical image appears upside down. Seeing the image simplifies the understanding of the eye, as it is intuitively clear that the ideal image is a demagnified replica of the object, which has to be as sharp and free of aberrations (various distortions in the two-dimensional image) as possible. The eye achieves focus using a variable focal length of its lens, which is controlled by accommodation. Once the image hits the retina, it is transduced by photoreceptors that also filter the light into three broad frequency ranges, which form the basis for color perception. After initial signal processing in the retina, a neural image is sent to the visual cortex through the optic nerve and through the thalamus.

A quick inspection of the ear does not reveal an equally obvious mechanism of operation and certainly nothing that looks (or sounds) like a lens or an image, let alone a two-dimensional one. Optics does not apply here, so an intricate combination of principles from physics, engineering, and biology must be invoked to explain its operation. The ear also does not work as a tape recorder, which might be thought of as the associative analog to the camera for sound. It does not "record" the incoming sound as a broadband signal, let alone "play" it back as is. Rather, after a series of complex mechanical transformations, the organ of Corti in the cochlea filters the broadband sound into numerous narrowband auditory channels that are sometimes independent of each other, and yet interact in other cases, according to complex rules that must be uncovered through experiment. The sound is finally perceived as broadband, as though the input has been perceptually resynthesized, following signal processing in different brain areas related to the auditory system.

Hearing points

The present theory attempts to show how the principles of optical imaging apply to hearing notwithstanding. The recipe for doing so is straightforward as long as several tropes of hearing science are shed. For the sake of this summary, let us treat the following statements as correct. They will all be properly demonstrated and motivated in the main text:

1. Time is for hearing what space is for vision.
2. Wave physics does not stop at the auditory periphery.
3. Constant frequency is the exception, not the rule.
4. The ear is not a lowpass receiver, but rather a (multichannel) bandpass system.
5. Acoustic source coherence propagates in space according to the wave equation.
6. Information arriving to the auditory brain is discretized and the rules of sampling theory must apply.

Each one of these statements on its own may not be particularly novel or controversial, at least in some contexts of hearing theory, but once their totality is internalized, new ways to understand hearing inevitably arise.

Inferring the auditory image

The visual image is spatial, since it is distributed over the area of the retina, where a two-dimensional projection of three dimensional objects appears as a pattern of light. In imaging analysis, it is customary to "freeze" the progress of time at an arbitrary moment and look at a single still image, which contains much of the information from the object and its environment, all simultaneously available within same image. The passage of time entails movement of the object(s) and observer, which can be understood as incremental changes to the reference still image. In contradistinction, it is exceedingly difficult to make sense of a "frozen" image of sound—for example, a particular geometrical configuration of the traveling wave in the cochlea—which carries limited information about the auditory scene on the whole. Here, it is necessary to let time pass and hear how the different sounds develop and interact in order to be able to say something meaningful about the acoustic situation they represent and how they are distributed in space. Hence, we intuitively arrive at the space-time analogy between vision and hearing, which was epitomized in Point 1.

Further pushing the logic of Point 1 entails that if the visual image occurs in space, then a hypothetical auditory image must occur in time. Then, we should expect that just as the optical object-image pair in vision can be represented as a spatial envelope that propagates between points in space (e.g., between the object origin and the image origin), so should the acoustical object-image pair of hearing be representable by a temporal envelope, between reference points in time. Both object types should have a center frequency that carries the envelope, as there is no physical difference in the way that the information about the objects is borne by waves.

Another important substitution is to find the temporal equivalent of diffraction, which in optics determines how different parts of the light waves interfere and change their shape, as a result of scattering by various boundaries and objects on their path to the screen. Diffraction is really a general term for wave propagation from the object, which may or may not encounter scattering obstacles on its way to the screen. Switching between the spatial and temporal dimensions (Point 1), diffraction is replaced by group-velocity dispersion, in which the shape of the temporal envelope is impacted by differential changes to its constituent frequencies. We recall that the cochlea itself is inherently a dispersive path, so that the acoustical signal that arrives into the cochlea is automatically dispersed.

Note that while we often talk about dispersion, we are actually concerned with its derivative—the group-velocity dispersion—that goes by different names, such as group-delay dispersion and phase curvature.

Next, if there is any chance for us to construct an imaging system within the auditory system, we should be also looking for a temporal aperture—something that limits the duration of signal that can be processed at one point in time (really, sampled) for a given chunk of acoustical input. Here, the neurons that transduce the inner hair cell motion produce spikes that are limited in time by definition, so there is a time window in effect that continuously truncates the signal into manageable chunks.

While the above steps are relatively simple endeavors, completing the identification of the temporal imaging system in the ear requires bolder steps—borderline speculative. First, we require an additional dispersive section in the auditory brain, regardless of the acoustical signal representation that is now fully neural (Point 2). Current science has it that there is no neural dispersion in the auditory brain, whereas the present work claims otherwise, as can be demonstrated by several measurements. While the precise magnitude of dispersion is difficult to ascertain, it is readily evident how no cochlear measurement of the group delay based on otoacoustic emissions has ever matched the auditory brainstem response measurements that include the brainstem as well. This discrepancy translates to a non-zero group-velocity dispersion of the path difference, which is mostly neural.

The second speculative step is to identify a lens. A temporal imaging system requires a "time lens", rather than a spatial lens, although it is not strictly necessary (this is because a pinhole camera—the simplest imaging device possible—does not have a lens, but still produces a sharp image, as long as the aperture is very small). A time lens performs the same mathematical operation as the spatial lens in the eye, only over one dimension of time instead of over two dimensions of space, and where frequency is a variable rather than a constant (Point 3). Indeed, an inspection of the dynamic properties of the organ of Corti that were recorded in several state-of-the-art physiological measurements reveals a phase dependence in time, frequency, and space that is symmetrical in shape. Such phase response can be readily attributed to a time lens, which is modeled using a quadratic phase function—a form of phase modulation.

Putting the system together

We have now identified the four necessary elements of a basic temporal imaging system in the ear: cochlear dispersion, cochlear time lens, temporal aperture, and neural dispersion. The values of the different elements can be roughly estimated for humans following an analysis of available physiological data from literature. These estimates can also be cross-validated using human psychoacoustic data from other sources.

Although we are dealing with sound, now we are in the conceptual realm of optics, which has devised a number of powerful analytical tools to characterize the image and objectively assess its quality. For example, it is possible to compute whether the above combination of elements produces a focused image (putatively, inside the brain)—a temporal envelope carried by a center frequency that is propagated to the midbrain or thereabout, where the "auditory retina resides". Surprisingly, the answer is a definite "no". The auditory image is defocused, unlike the optical image that appears on the retina in normal conditions. However, in vision, we know how information about the object is superior when the image is focused (for instance, try to read a blurred text from the distance).

Why should hearing be any different than vision and be defocused and not sharply focused? Why would we want to hear sounds that are blurry rather than sharp? To be able to answer these questions requires us to revisit the idea of group-velocity dispersion and establish its relevance to realistic acoustic signals. This will help us establish the meaning of sharpness and blur in hearing.

Spatial blur

In spatial imaging, there are two general domains of blur. In the geometrical one, light "rays" are traced following refraction. Every object can be thought of as a collection of point sources in a continuum, from which light rays diverge in all directions, each carries the information about the point it emanated from. The goal of imaging is to collect the rays so they form the same pattern of light in another region in space as they do at the object position. This generally includes a linear scaling factor—magnification—which does not have direct bearing on the fidelity of the image that is otherwise a one-to-one mapping of the object in space. Deviations from the one-to-one mapping in two dimensions—when the rays do not converge exactly where they should in order to reconstruct the object—are called aberrations. An out-of-focus imaging system has a "defocus" aberration, which entails a lack of convergence of the rays coming from different directions on the screen, so that information from different points of the object is "mixed" at the position of the image, in a way that is visible. The geometrical form of blur is the most dominant one when the wavelength of the light is much smaller than the object and also smaller than the different obstacles in the optical path to the image.

On the other extreme, when the wavelength of the light is comparable to the details of the object or the aperture, or other things that scatter the light on the way to the screen, then the effect of diffraction may be visible in the image as different interference patterns that can distort the details of the object. These may be thought of as diffraction blur, although the underlying mechanism is very different from geometrical blur. Effects here, if they are visible at all under normal conditions, tend to appear along edges and around very fine details of the image.

With these rough definitions in mind, we note that sharpness is simply the absence of visible blur, or blur that is quantified to be below a certain threshold.

Auditory blur and coherence

Back to hearing, how does the auditory image—really, the temporal envelope carried by high frequency—ever becomes blurred? First, we have to transform the two types of blur to temporally-relevant phenomena (invoking Point 1 again). An example of geometrical blur is relatively easy to see, since it manifests in reverberation. Here, the information from the source contained within a point in time (or rather, an infinitesimally short interval) arrives to the receiver mixed with information originating in other points in time. The mixing is asynchronous, so it adds up randomly and does not interfere. A corollary is that direct or free-field (anechoic) sound suffers from no geometrical blur at the input to the ear.

The spatial diffraction blur can be analogized to temporal dispersion blur. Group velocity dispersion entails that every component of the temporal envelope propagates at somewhat different velocity. Temporal obstacles in time (i.e., those that can be expressed as filters or time windows) that are approximately proportional to the period of the carrier wave may impose a differential amount of delay to the different components of the envelope spectrum. The result is similar to interference in time and is, therefore, an analogous form of blur of the temporal envelope to that of diffraction blur of the spatial envelope.

It appears that in order to know whether the different types of blur ever apply in reality, it is necessary to know how much bandwidth the acoustic source occupies. If the bandwidth is very narrow (with the extreme being that of a pure tone), then group-velocity dispersion is unlikely to have any effect, because only a single velocity is relevant per given frequency. However, as the bandwidth becomes wider, more and more frequencies are subjected to differential group-velocity dispersion, so the envelope may become blurry if the dispersion is high or if it is accumulated over

large distances. But, there is an effective limit imposed on the bandwidth here, because the auditory system analyzes the acoustic signal in parallel bandpass filters, each with a finite bandwidth, that together cover the entire audio spectrum. Therefore, the maximum relevant bandwidth of a signal has to be related to the auditory filter in which it is being analyzed. Either way, it is realistic signals that are naturally modulated in frequency and do not have constant frequency (Point 3) that may experience the effects of group-velocity dispersion most strongly.

Another effect of the bandwidth that is related to geometrical blur has to do with the degree of randomness of the signal. Unlike deterministic signals, a truly random signal does not interfere with a delayed copy of itself. Therefore, the notion of geometric blur as in reverberation applies best to random signals that do not interfere, and only mix in energy without consideration of the signal phase. In general, the more random a signal is, the broader its bandwidth is going to be, since its amplitude and phase cannot point to one frequency at all times. This reasoning is best captured by the concept of (degree of) coherence (also called correlation in the context of hearing and acoustics). It describes the ability of a signal to interfere with itself. A completely random signal, such as white noise (broadband spectrum), does not interfere with itself and is considered incoherent. A deterministic signal, such as a pure tone, can interfere with itself and is considered coherent. Roughly translating these terms into more intuitive understanding, a coherent source sounds more tonal—like a melodic musical instrument—whereas an incoherent sound source is more like noise. However, the vast majority of real-world sounds are neither completely tonal nor are they completely noise-like, so they can be classified as partially coherent.

All in all, it appears that the auditory defocus is applied differentially to different types of signals. Coherent signals that are largely unaffected by dispersion, are also unaffected by defocus. In contrast, partially coherent signals are made more incoherent by defocus, whereas incoherent signals remain incoherent also after defocus.

The modulation transfer function

Returning to standard imaging theory, it should not come as a great surprise that the imaging process and quality differ depending on the kind of light that illuminates the object: coherence matters. This becomes immediately apparent when analyzing the spatial frequencies of the object, which make the spectrum of the spatial envelope that is being imaged. A major result in imaging optics is that whatever diffractive and geometrical blurring (or other aberration) effect beyond magnification (i.e., linear scaling) exist in the system, they can all be expressed through its "pupil function", which then leads to the derivation of the so-called modulation transfer functions (it is different for coherent and incoherent illumination).

In exact analogy to spatial optics, we obtain similar, degree-of-coherence-dependent modulation transfer functions in the temporal domain that incorporate the effects of dispersion in the aperture and blur due to defocus. Such functions are nothing new in hearing science, but so far they have been obtained only empirically without explanation for why the coherent and incoherent functions are different, whereas the present theory contains the first derivation of these functions from the basic principles of auditory imaging. Once these functions become available, all sorts of predictions may be offered to explain different auditory effects that have been also measured only empirically until now.

The theoretical modulation transfer functions that have been obtained here are only good as first approximation and there are clear discrepancies from experiment in several cases. It is argued that a major reason for the discrepancy is the discrete nature of the transduction. Irregular spiking in the auditory nerve further downstream in the auditory system is tantamount to repeatedly sampling the original signal at nonuniform intervals, which tends to degrade the possible image at the output of

the system. In very subtle contexts it may also create perceivable artifacts, which are not captured by the analytically derived (continuous) modulation transfer function (Point 6).

Coherence conservation and the phase locked loop

It is necessary to take a small detour in the auditory imaging account and introduce another element to the discussion that is borrowed from communication and control theories. In the brief mention of coherent signals above we sidestepped an important question: while we know that coherence propagates in space according to the wave equation (Point 5), do we also know for certain that it is conserved in the ear? Notably, does the transduction between mechanical to neural information conserve the degree of coherence of the original signal? A big clue seems to suggest that the answer is yes: signals are known to phase lock in the auditory nerve, following transduction by the inner hair cells. This applies to coherent (tonal) signals, and to a lesser degree to other signals, where incoherent signals only lock to the slow envelope phase and not to the random carrier. Phase locking to coherent signals is special in hearing compared to vision, where the phase of the light wave changes too rapidly to be tracked by a biological system.

Phase locking is a hallmark of coherent reception—a form of information transfer in communication theory, which tracks the minute variations in the phase of the (complex) temporal envelope of a signal. Realizing this form of reception requires an oscillator, which is an active and nonlinear component within the receiver. (In the complementary noncoherent reception, which only tracks the slowly-varying envelope magnitude, an oscillator is not strictly necessary.) A closer inspection of the cochlear mechanics and transduction suggests that phase locking first emerges at the cochlea, before it becomes manifest in the auditory nerve. Therefore, it seems reasonable to look for the components of a classical coherent detector (Point 4) that provides phase locking, i.e, a phase-locked loop (PLL). For a PLL to be constructed, we require a phase detector, a filter, an oscillator, and a feedback loop that returns the output to the phase detector. All these components can be identified within the organ of Corti and the outer hair cells.

Conventional theory holds that the outer hair cells perform cycle-by-cycle amplification for the incoming signal, but theory and experiment are still in disagreement as for how this process exactly works. The PLL model does not necessarily clash with this standard amplification model, and might even interact with it, by incorporating amplification into its own feedback loop. A similar argument may be made about the time lens and the PLL—their function may not necessarily be in conflict.

As it currently appears in the main text, the auditory PLL model is strictly qualitative and, accordingly, speculative. However, it does provide the missing link for coherence conservation between the outside world and the brain—a link that has been glossed over until now and is critical for the understanding of how the auditory system handles different kinds of stimuli according to their degree of coherence.

Auditory imaging concepts

With the imaging system specified, we can now turn to explore some of the hallmark concepts of imaging theory and apply them to hearing, beyond those of auditory sharpness and blur.

First, we calculate the temporal resolving power between two pulses—analogous to the resolving power between two object points, as is mandatory information in telescopy, for example—using the estimated temporal modulation transfer function. The predictions are comparable to empirical findings from relevant studies, especially at the 1000–8000 Hz range.

We also elaborate on the auditory analogs to monochromatic and polychromatic images. It is argued that the various pitch types can be thought of as the quintessential monochromatic image (pure tone, unresolved complex tone) or polychromatic image (resolved complex tone, interrupted pitch)—all of which highlight different periodicities in the acoustic object.

This understanding can then be used to hypothesize deviations from perfect imaging—various monochromatic and polychromatic aberrations. Monochromatic aberrations relate to the variation of the group delay within the auditory channel, whereas polychromatic aberrations to variations between channels. Several examples are given to the different types, based on known phenomena from the psychoacoustic literature, as well as a new effect.

Finally, we can hypothesize about the auditory depth of field that is temporal rather than spatial. It should be most clearly observable between objects of different degree of coherence. For example, it is argued that forward masking can be readily recast as small depth of field, in the case of incoherent (broadband noise) masker and coherent (pure tone) probe, since the boundary between them is effectively blurred as a result of forward masking. However, when the masker and probe are of the same type (e.g., incoherent and incoherent), then the depth of field is large, as the forward masking becomes longer.

Auditory accommodation and impairments

An even bigger leap in applying the analogy between optical / visual imaging and acoustical / auditory imaging is the search for auditory accommodation. In vision, accommodation is an unconscious mechanical process that varies the focal length of the lens to match the distance of the object, so to bring its image to sharp focus on the retina for arbitrary distance of the object. Accommodation involves several ocular muscles that are fed by an efferent nerve from the midbrain, which together with the output from the retina form a feedback loop.

We have already stated that the ear is defocused, so what could possibly be the use of auditory accommodation in this context? One attractive answer is to control the degree of coherence that enters the neural system, which as was argued above, is captured by the degree of phase locking. This is interesting because of much converging empirical evidence that shows how the hearing system can process sound either according to its slowly-varying temporal envelope (its magnitude), or using phase-locking to track the fast variation of the carrier phase (the so-called temporal fine structure of the stimulus). The difference between the two processing schemes runs throughout the very physiology of the auditory brainstem, which appears to have dedicated parts for each type of processing. Coming in full circle, this differentiation is akin to the types of imaging that exist in optics: coherent, incoherent, or a mixture of the two—partially coherent. It is also akin to the detection schemes that are used in standard communication engineering: either coherent or noncoherent. Applying a variable stage in coherence processing may be achieved, for example, by the medial olivocochlear reflex—an efferent nerve that innervates the outer hair cells and whose function is not well understood. But additional mechanisms to achieve the same function may exist. Once again, this involves considerable speculation at present, but evidence to support this and related possibilities does exist and is discussed in depth in the main text.

Although there are many uncertainties about the specifics of this system, the penultimate chapter is dedicated for hypothesizing what happens when things go wrong: what is the effect of faulty imaging that may translate into hearing impairments? There are several possible answers here, with dysfunction in hypothetical auditory accommodation being the most attractive candidate. Nevertheless, evidence here is difficult to gather and much work has to be done to uncover the basics before turning to these more challenging, yet important, questions.

"With the four-dimensional space curved, any section that we make in it also has to be curved, because in general we cannot give a meaning to a flat section in a curved space."

PAUL DIRAC (1963)

"Information is physical."

ROLF LANDAUER (1996)

Contents

List of acronyms

A1	Primary auditory cortex
A2	Secondary auditory cortex
A3	Tertiary auditory cortex
ABR	Auditory brainstem response
AC	Alternating current
AM	Amplitude modulation
ATF	Amplitude transfer function
AVCN	Anteroventral cochlear nucleus
BM	Basilar membrane
CCD	Charge-coupled device
CF	Characteristic frequency
CN	Cochlear nucleus
CMR	Comodulation masking release
cPSF	coherent Point spread function
CSF	Contrast sensitivity function
DC	Direct current
DCN	Dorsal cochlear nucleus
DNLL	Dorsal nucleus of the lateral lemniscus
DPOAE	Distortion-product otoacoustic emission
EEOAE	Electrically evoked otoacoustic emission
EOAE	Evoked otoacoustic emission
ERB	Equivalent rectangular bandwidth
FEM	Finite-element method
FFR	Frequency-following response
FM	Frequency modulation
fMRI	functional Magnetic resonance imaging
FWHM	Full-width half maximum
GDD	Group-delay dispersion
GVD	Group-velocity dispersion
IACC	Interaural acoustic cross-correlation
IC	Inferior colliculus
ICC	Central nucleus of the inferior colliculus
ICD	Dorsal cortex of the inferior colliculus
ICX	External nucleus of the inferior colliculus
IHC	Inner hair cell
INLL	Intermediate nucleus of the lateral lemniscus
iPSF	incoherent Point spread function
LCT	Linear canonical transform
LED	Light-emitting diode

LGB	Lateral geniculate body
LL	Lateral lemniscus
LOC	Lateral olivocochlear
LSO	Lateral superior olive
MDI	Modulation discrimination interference
MEG	Magnetoencephalography
MET	Mechanoelectrical transduction
MGB	Medial geniculate body
MLR	Middle latency response
MNTB	Medial nucleus of the trapezoid body
MOC	Medial olivocochlear
MOCR	Medial olivocochlear reflex
MSO	Medial superior olive
MTF	Modulation transfer function
NLL	Nuclei of the lateral lemniscus
NPLL	Neural phase-locked loop
OAE	Otoacoustic emission
OHC	Outer hair cell
OTF	Optical transfer function
PLL	Phase-locked loop
PM-HLL	Period-modulated harmonic locked loop
PSF	Point spread function
PTF	Phase transfer function
PVCN	Posteroventral cochlear nucleus
QM	Quadrature modulation
RC	Resistance capacitance
RMS	Root mean square
SC	Superior colliculus
SFM	Sinusoidal frequency modulation
SFOAE	Stimulus-frequency otoacoustic emissions
SHC	Short hair cell
SNR	Signal-to-noise ratio
SOAE	Spontaneous otoacoustic emission
SOC	Superior olivary complex
SON	Superior olivary nucleus
SPL	Sound pressure level
SPN	Superior paraolivary nucleus
TOAE	Transient otoacoustic emission
TFS	Temporal fine structure
THC	Tall hair cell
TM	Tectorial membrane
TMTF	Temporal modulation transfer function
UWB	Ultra-wideband
V1	Primary visual cortex
VCN	Ventral cochlear nucleus
VCO	Voltage controlled oscillator
VNLL	Ventral nucleus of the lateral lemniscus
VNTB	Ventral nucleus of the trapezoid body

List of symbols

a, b, c	Constants in polynomial (§F)
a, b, c, d, i	Coefficients of power law of ABR or OAE (§11.7.3, §17.3.1)
a, b, d	Linear canonical transform (LCT) coefficients (§C)
a	Complex temporal envelope
a	Arbitrary signal (§14)
A	Complex envelope spectrum
a_g	Ideal geometrical image of envelope (§13.2.1)
a_n	Complex temporal envelope in neural domain (§12.2)
a_s	Sampled signal (§14)
A_s	Sampled spectrum (§14)
b_n	Random eigenfunction coefficient (§8.2.6)
B	Bandwidth
B_n	Noise-equivalent bandwidth (§9.4)
c	Wave speed; Speed of light; Speed of sound
C	Circle of confusion (§4)
C	Channel capacity (§5.2.1)
C	Schroeder phase curvature (§12.4.1, §17.3.1)
C	Real part of Fresnel integral (§D)
C_T	Linear canonical transform (LCT) kernel for operator T (§C)
d	Temporal acuity; Detectable gap (§13.4.4, §15.6, §F)
d	Group-delay dispersion impulse response (§12, §13.2)
D	Aperture size (§4)
D	Group-delay dispersion transfer function (§12, §13.2)
D	Pulse sequence duration (§E)
D'	Pulse sequence duration minus the last pulse duration (§E)
D_a	Acoustic dipole strength (Table 3.1)
D_e	Electromagnetic dipole strength (Table 3.1)
$\boldsymbol{E}$	Electric field (vector) (Table 3.1)
E	Electric field (scalar) (§4)
ERB	Equivalent rectangular bandwidth of auditory filter
f	Frequency
f	Focal length (§4)
f	Arbitrary function (§3.2.1)
F	Fourier transform of arbitrary function f (§3.2.1)
$\boldsymbol{F}$	Arbitrary wave field (Table 3.1)
f'	Doppler-shifted frequency (§3.3.3)
f'	Remapped imaged frequency under binaural diplacusis (§17.4.2)
$f^{\#}$	F-number (§4.2.1)

f_0	Fundamental frequency
f_c	Carrier frequency
f_m	Modulation frequency
f_n	Cutoff frequency (Figure 9.3)
f_T	Focal time
$f_{T,gg}$	Focal time estimate of the gerbil and guinea pig time lens
$f_{T,h}$	Focal time estimate of the human time lens
f_s	Sampling rate
$f_T^{\#}$	Temporal imaging f-number (§ 15.11.1)
$\mathcal{F}$	Fourier transform
$\mathcal{F}^{-1}$	Inverse Fourier transform
g	Arbitrary function (§ 3.2.1)
g	Complex envelope / Mapping function (§ 5.3.1)
g_1, g_2	Fresnel integral variables (§ D)
g_T	Linear canonical transformed (LCT) function (§ C)
h	Impulse response function
$\tilde{h}$	Impulse response function with reduced coordinate; Point spread function (§ 13.2.1)
$\tilde{h}_d$	Defocused impulse response function for Gaussian pupil (§ 13.2.2)
h_{dr}	Defocused impulse response function for rectangular pupil (§ D)
h_L	Time-lens impulse response function (§ 10.3, § 13.2)
$\boldsymbol{H}$	Magnetic field (vector) (Table 3.1)
H	(Shannon's) entropy (§ 5.2.1)
H	Transfer function
H_d	Amplitude transfer function (ATF) of generalized Gaussian pupil (§ 13)
H_{dr}	Amplitude transfer function (ATF) of generalized rectangular pupil (§ 13)
H_L	Time lens transfer function (§ 10.3)
$\mathcal{H}$	Hilbert transform (§ 6)
$\mathcal{H}^{-1}$	Inverse Hilbert transform (§ 6)
$\mathcal{H}$	Optical transfer function (§ 13)
$\mathcal{H}_d$	Defocused optical transfer function with Gaussian pupil (§ 13)
$\mathcal{H}_{dr}$	Defocused optical transfer function with rectangular pupil (§ 13)
i	$\sqrt{-1}$
I	Intensity
I	Frequency interval (§ 15.10.1)
I_{coh}	Coherent part of intensity (§ 8.2.1)
I_{inc}	Incoherent part of intensity (§ 8.2.1)
I_{max}	Maximum intensity (interference)
I_{min}	Maximum intensity (interference)
I_{tot}	Total intensity (§ 8.2.1)
I_0	Object intensity (§ 4)
I_1	Image intensity (§ 4)
I_r	Intensity (radial) (Table 3.1)
I_θ	Intensity (azimuth) (Table 3.1)
I_ϕ	Intensity (elevation) (Table 3.1)
J_n	Bessel function of the first kind of order n
k	Wavenumber; Spatial frequency
k_x	Horizontal component of spatial frequency (Table 1.2)
k_y	Vertical component of spatial frequency (Table 1.2)

$\boldsymbol{k}$	Wavenumber vector
k_c	Wavenumber at carrier frequency
k_i	Imaginary-part of wavenumber function (absorption) (§3.4.2)
k_r	Real-part of wavenumber function (dispersion) (§3.4.2)
K	Gain (§5.3.1)
K	Stiffness (§11.6.1)
K	Loop gain (§9.3)
K_{BM}	Passive basilar membrane stiffness (§11.6.1)
K_f	Filter DC gain (§9.3)
K_m	Phase detector sensitivity (§9.3)
K_v	Voltage controller oscillator gain (§9.3)
l	Integer
$\mathscr{L}(T)$	Linear canonical transform (LCT) for kernel operator T (§C)
m	Integer
m	Message (baseband) function (§5.3.1)
m	Modulation depth (audibility, contrast)
m	Frequency slope (chirpiness) (§B.3)
M	Magnification
M'	Complex magnification (defocused system) (§12.2.2)
$M^*, M_f, M_{f'}, M_{f'}^*$	Distorted magnification under binaural diplacusis (§17.4.2)
m_0	Chirp slope of object (§12.4.1)
M_0	Magnification $-v/u$
m_1	Chirp slope of image (§12.4.1)
m_r	Rectangular pulse chirp slope (§B.3)
n	Index of refraction
n	Noise time-signal (§5.3.1)
n	Harmonic number
n	Integer
n_k	Neural group velocity dispersion coefficient (§11.7.3)
N	Noise power (§5.2.1)
N	Number of samples
N	Schroeder phase number of harmonics (§12.4.1)
N	Number of pulses in sequence (§E)
NS	Nonstationarity index (§A.4)
p	Pressure
p	Probability (§5.2.1)
P	Pressure (frequency domain)
P	Pupil function
$\mathcal{P}$	Cauchy principal value (§6)
$\mathcal{P}$	Generalized pupil function (§13.2.2)
P_e	Entrance pupil (§14.4.2)
P_g	Gaussian pupil function (§13.2.2)
P_r	Rectangular pupil function (eq. (13.32), §D)
q, q'	Linear canonical transform (LCT) variables (§C)
Q	Q-factor of bandpass filter
Q_{10}	Q-factor of bandpass filter at -10 dB points (§11.6.4)
Q_{ERB}	Q-factor of bandpass filter at the ERB bandwidth (§11.6.4)
r	Displacement

r	Received signal time-signal (§5.3.1)
r	Correlation coefficient (Pearson's r)
$\boldsymbol{r}$	Position vector
r_c	Critical distance (§8.4.2)
R	Fixed distance (§8.5)
R	Synchronization strength/index (footnote 85)
R_{pp}	Autocorrelation function of p (§8.2.2)
s	Time lens curvature
s	Time signal
s	Laplace-transform complex frequency (§9.4)
s	Sampler delta function array (comb function) (§14)
$\hat{s}$	Estimate of time signal (§3.3.3)
$s_{T,gg}$	Curvature estimate of the gerbil and guinea pig time lens
$\boldsymbol{S}$	Poynting vector (Table 3.1)
S, S_I	Intensity/power spectrum
S	Signal power (§5.2.1)
S	Power spectral density / Spectrum (§8)
S	Surface area of room (§8.4.2)
S	Imaginary part of Fresnel integral (§D)
S_a	Amplitude spectrum (§13.3.3)
s_h	Time-lens curvature in humans (§11.6.4)
S_ω	Acoustic point source strength (Table 3.1)
t	Time
t_0	Gaussian pulse width parameter of the object; Entrance pupil
t_1	Gaussian pulse width parameter of the image; Exit pupil
t'	Complex Gaussian pulse width parameter
T	Period
T	Integration time constant
T	Complex tone standard (§15.11.1)
T	Linear canonical transform (LCT) matrix operator (§C)
$\tilde{T}$	Integration variable (§13.2.1)
T_{60}	Reverberation time
T_a	Aperture time
T_c	Carrier period
T_s	Sampling rate period
v	Velocity (§3.3.3)
v	Distance between lens and screen (§4)
v	Neural group-delay dispersion
V	Visibility (§7.2.1, §8)
V	Volume (§8.4.2)
V	Voltage (§11.6.1)
v_g	Group velocity
V_c	Output voltage from loop filter (§9.4)
V_d	Output voltage from phase detector (§9.4)
V_i	Input voltage to phase detector (§9.4)
V_o	Output voltage from voltage controller oscillator (§9.4)
v_p	Phase velocity
v_{V-I}	Neural group-delay dispersion for the segment between wave I and wave V (§11.7.3)

$\Delta f'$	Iterated octave stretch in cents (§ 15.10.1)
Δf_h	Bandwidth of the time-lens phase modulation in humans (§ 11.6.4)
Δk	Wavenumber difference between beating components (§ 3.2.1)
Δl	Coherence length (§ 8.2.4)
Δn	Path change in index of refraction (§ 11.6.1)
Δt	Deviation from complex tone standard (§ 15.11.1)
Δt_{opt}	Optimal ratio between geometrical and dispersive blurs (Table 12.2, § 13.2.2)
$\Delta \tau$	Coherence time
$\Delta \omega$	Frequency difference between beating components (§ 3.2.1)
$\Delta \omega$	Bandwidth of narrowband signal
$\Delta \omega$	Phase ramp (§6.3)
$\Delta \omega$	Phase ramp (§6.3)
$\dot{\omega}$	Frequency velocity (§6.3)
$\ddot{\omega}$	Frequency acceleration (§6.3)
ϵ	Dielectric constant (permittivity) (Table 3.1)
ϵ_0	Dielectric constant in vacuum (permittivity) (Table 3.1)
ζ	z-axis coordinate in the traveling pulse coordinate system (§10,§11, §12.2)
ζ	Damping factor (Figure 9.3)
η	Displacement function (§ 3.3.1)
η	Mask coordinate (§ 4.2.2)
θ	Azimuth (Table 3.1)
θ	Angle from optical axis (§4)
θ_{min}	Angular resolution (§ 4.2.2)
κ	Adiabatic compressibility (Table 3.1)
λ	Wavelength
Λ	Triangle function
λ_n	Eigenvalue (§ 8.2.6)
μ	Magnetic permeability (Table 3.1)
μ	Spectral degree of coherence (§8)
μ_0	Magnetic permeability in vacuum (Table 3.1)
ξ	Mask coordinate (§ 4.2.2)
ρ	Fluid density (Table 3.1)
ρ_{BM}	Basilar membrane density per unit length (§ 11.6.1)
σ	Multiplicative factor (§ 12.4.2, §F)
τ	Time lag
τ	Time coordinate in the traveling pulse coordinate system, group delay
$\tilde{\tau}_0$	Reduced coordinate of the object time (§ 13.2.1)
τ_0	Initial time of pulse in the traveling pulse coordinate system
τ_1, τ_2	Time constants (Figure 9.3)
τ_e	Effective duration (§ 8.3.2, §A)
τ_g	Group delay
τ_p	Phase delay
φ	Phase function (time domain)
ϕ	Phase function (frequency domain)
ϕ	Elevation (Table 3.1)
ϕ_e	Phase error (§9.3)
ϕ_n	Schroeder phase for component n (§ 12.4.1)
$\phi_{\Delta f/2}$	Time-lens phase at the cutoff frequencies of Q_{10} (§ 11.6.4)

Mathematical conventions

In this work, a forward-propagating wave is implicitly taken to be the real part of a signal of the form:

$$p(x,t) = e^{i(\omega t - kx)} \tag{1}$$

This matches the convention that is typically used in signal processing, where the argument that contains the angular frequency ω is positive. It is also the convention in most of the optics and communication texts cited here, as well as several others that include, notably, Haus (1984); Siegman (1986); Kolner (1994a); Cohen (1995); Fletcher and Rossing (1998); Kinsler et al. (1999); Blinchikoff and Zverev (2001); New (2011); Couch II (2013) and Kuttruff (2017). Therefore, formulas that were taken from Morse and Ingard (1968); Jackson (1999); Whitham (1999); Born et al. (2003) and Goodman (2017) sometimes had to be adopted by changing the sign as they originally appeared in the form $p(x,t) = e^{i(kx - \omega t)}$.

The Fourier transform that is used in the text conforms to

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \tag{2}$$

whereas the inverse Fourier transform is

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega \tag{3}$$

Thus, throughout this work, lowercase signal functions are time domain and their spectral domain counterparts are capitalized.

Preface

This was never intended to be a book-long piece. My original wish upon starting to write it was to introduce into hearing the "paraxial" dispersion equation—perhaps in an article format. Naively, I was convinced that the beauty of the equation—or rather, the beauty in the analogy to vision—would be self-evident and immediately applicable. But as I was trying to come up with a way to introduce these ideas, it had quickly become clear that there is no straightforward way to motivate them at the present state of hearing science. At least, not without taking many steps back and venture into several scientific corners that have somehow escaped rigorous treatment. Wherever I looked, it had seemed as if some aspects of the science froze in time and have left me with an intuition about acoustic waves that would have been excellent at the turn of the previous century. It has taken considerably more research to be able to assort all the pieces that would be required to motivate the need for a dispersion equation in hearing. Some of these pieces were in plain sight, while some are in vogue in contemporary research, and yet others have been long forgotten. The price for picking up these pieces may be considered steep, though: let go of static frequencies and stationary signals, embrace coherence as a fundamental descriptor of sound, and learn to accept that the auditory system must know what it is doing much better than what we would like to think it should want to do.

Therefore, this is an invitation for the daring reader. I expect that a fair share of the topics that are touched upon will elicit resistance in some of the readers. Nevertheless, some of the topics—mainly in the introductory chapters—may appear timely and not nearly as controversial as the more advanced chapters. In more than one occasion I have resorted to indirect and unconfirmed methods that can associate the new ideas with an unwelcome, yet unavoidable, air of speculation. Also, the very broad scope of the present theory may simply be overwhelming, as there should be numerous points of contention that may be deemed worthy of rebuttal. This is all well, as far as I am concerned, as long as the ideas will be seriously considered and stir a long-needed discussion in the community.

My quest for a hearing theory has been initially triggered by a simmering dissatisfaction with the traditional presentation of auditory science. It had struck me as a bewildering collection of facts, which had to be learned by rote instead of through internalization of deeper concepts. Many results had to be empirically measured rather than derived based on higher-level principles. There did not seem to be a theory that is general enough to offer the necessary predictive power that may spare a direct experiment. Quite the contrary—predictions often seem to be proven wrong by experiment, which had made it difficult to develop an intuition for the inner workings of the ear and its inherent logic.

This is why stumbling across the temporal imaging theory in optics seemed like a possible key to my unrest. This theory is based on the space-time duality principle as formulated by Sergeĭ Aleksandrovich Akhmanov et al. (1968) and Akhmanov et al. (1969) and was later developed into a temporal imaging theory by Brian Kolner and Moshe Nazarathy (1989), and greatly elaborated by

Kolner (1994a) and his subsequent work. Here was an elegant theory that is completely analogous to the classical single-lens imaging theory from optics, only with the dimensions reversed, as the temporal envelope is used instead of spatial envelope (expressing the distribution of light of the optical object and image). Knowing that vision rests on spatial imaging that is neatly formulated using the paraxial equation and a double Fourier transform, there was an immediate allure of having a paraxial equation and a double Fourier transform expressed in time and frequency coordinates that can, rather organically, represent hearing.

Alas, the basic elements in the temporal imaging theory are group-velocity dispersion, time-lens curvature, and aperture time. What do these concepts have to do with hearing? Everything, it seems. But the reasoning behind it, which is the main subject of this treatise, has taken me the better part of the last four years to arrive at. Apart from my own time-consuming ignorance of many of the associated disciplines—some of them are routinely alluded to in auditory research— there appeared to be several gaps in the fundamentals of acoustics and hearing science, which had to be retraced and patched up in order to be able to tackle the idea of imaging with a degree of rigor that I thought the topic deserves.

There were two principal "culprits" that underpin the gaps in auditory science. The first one is the over-reliance on pure tones—a mathematically degenerate signal with no curvature, which carries little-to-no information and is not encountered in nature. This deficiency is addressed in several chapters that adopt the complex envelope and constant carrier formulation as the most general representation of waves, signals, objects, images, and communication functions. In turn, it opens the door for unifying the auditory concepts of temporal envelope and temporal fine structure with mathematically related concepts in acoustics, optics, and communication.

The second "culprit" is acoustic coherence theory, or the lack thereof. As a scalar wave theory, linear acoustics of plane waves is completely compatible with scalar wave theory in optics, which is also where classical coherence theory was developed. The main developments in coherence theory gathered momentum in the 1950s, at a point in which acoustics and optics may have been practiced by different scholars. Acoustics, and hearing too, imported a mélange of coherence-related concepts from several disciplines—each with its own jargon—that hardly coalesce into a consistent under-standing. The two chapters about hearing-relevant coherence theory, while not adding anything new to the science, are a first attempt to unify and revive these ideas in a manner that is consistent with wave physics, room acoustics, hearing (phase locking), communication engineering, and neu-roscience. Optical coherence theory provides a bridge that can be applied in sound, using Fourier optics, alongside some of the most insightful tools from imaging. The proposed amalgamation of the different coherence theories attempt to connect concepts of coherence with synchronization that manifest both in the mechanical and in the neural parts of the auditory system, and is thought to generally characterize perception throughout the brain.

With the availability of these introductory chapters, the motivation for the temporal imaging theory should be in place. I have put substantial effort in exploring some of the potential implications of temporal imaging—temporal modulation transfer functions, aberrations, accommodation, and dispersive hearing impairments. Thus, it is my hope that interested readers will be able to follow the wildly different approach to hearing that is presented in the advanced chapters of this work, despite the effort that it may require. While I cannot foresee the correctness of some of the hypotheses put forth, I will feel greatly rewarded to know that these ideas will have influenced future researchers in solving some of the more persistent challenges in our understanding of hearing and hearing impairments.

Preface to v6

While not exactly a second edition, v6 contains more significant updates to the theory compared to the one that first appeared in September 2021 and its subsequent public versions on arXiv (v1–v5; https://arxiv.org/abs/2111.04338).

The most substantial update is the addition of two more supporting studies to the section about the time lens (§ 11.6.2). Although human-relevant curvature values are difficult to extract from these animal data (§ 11.6.4), the current total of five different studies helps to push the idea that the cochlea contains a time lens away from the realm of speculation. Still, the data appear to be clustered in two curvature ranges of the putative time lens, which may be difficult to accept without the interaction with auditory accommodation—itself a speculative idea, albeit completely in line with the system physiology and the precedent of accommodation in vision. The addition of more data points to the time lens curvature had a slight cascade effect on many of the quantitative predictions in this work, which were corrected accordingly. It also led to a correction of an error in the related formula of the octave stretch effect in § 15.10.1. Unfortunately, this update produces an octave stretch effect prediction that is more limited and somewhat messier than in the previous editions, although still relevant. Throughout, the large-curvature time-lens estimates have been used, whereas the small-curvature estimates were no longer viable in most contexts. This is unlike the previous versions of the text, where the difference between the two curvatures informed us of the extreme values of the system curvature range.

Another major addition to this work is the summary, which aims to provide a briefer and lighter exposition to the ideas of the temporal auditory imaging theory and is more accessible than the technical abstract and long introductory chapters. The summary was constructed in a rather deductive manner, which rests on six points that were themselves gathered from the text (although without explicit reference to them later). Other changes are the additions of a couple dozen recent references (more than 1600 in total), a figure of the peripheral ear, and minor corrections to some figures and text.

Preface to v7

With the exception of a clearer presentation of the results of the experiments in §E, the changes in this version (v7) are all relatively minor: the usual correction of typos and references, as well as the inclusion of several new references from the past couple of years and a few ones that were overlooked before (notably, a mention of an imaging system in lizards with a parietal eye that is part of their pineal complex, the hypothesis that color vision in cephalopods is based on defocus, and the term "contrast sensitivity function" that is the correct visual equivalent to temporal modulation transfer function). However, to ease the future search and increase the visibility of this manuscript, this may be the final version that will appear on arXiv, for which I am grateful. Future updates may appear exclusively on the new web version to be found at http://www.hearingtheory.org.

About the author

Adam Weisser (born 1978) is an independent researcher with a broad academic and industrial background in hearing science and acoustical engineering. He holds a PhD in Hearing Science from Macquarie University in Sydney (topic: Complex Acoustic Environments, supervised by Jörg Buchholz and co-supervised by Gitte Keidser), MSc in engineering acoustics from the Technical University of Denmark (topic: Small Room Acoustics, supervised by Jens Holdger Rindel and co-supervised by Jan Voetmann), and BA Cum Laude in Physics from the Technion in Haifa. In between he spent eight years in the hearing aid industry in Denmark, primarily in the research and development of remote-microphone wireless technologies for the hearing impaired population. Before that, he also worked in high-frequency microwave chip design (MMIC) in Israel. This industrial experience involved basic acquaintance with communication engineering. Along with a previous training in the fundamentals of optical physics and a long standing interest in photography and cinematography, as well as extensive experience in music performance and production, these influences coalesced into a way of thinking about hearing that informed the ideas presented in this manuscript.

Acknowledgments

The kernel of this work emerged from a long-standing interest I have had in exploring analogies that exist between psychoacoustics and imaging and Fourier optics to improve my intuitive understanding of hearing. Early during my PhD studies I had realized that such analogies tend to be fragmentary, where they appear in print, or altogether nonexistent. Fortunately, I had the chance to delve into this question in the context of a journal club presentation I titled "Optical Hearing", on 5 October, 2017 at Macquarie University. Thus, I am indebted to the large group of participants in my talk and for this group for sharing their own presentations that had undoubtedly provided me with background knowledge and motivation that had helped me to crystallize my own ideas.

I would like to thank Jörg Buchholz, for giving me the necessary freedom to explore these uncharted territories while working on my PhD during that period in 2017–2018, and for providing the opportunity to be immersed in this research in the first place.

A big thanks goes to Nicholas Haywood for his insightful suggestion to explore aliasing as a proxy for discrete processing. While he was unavailable for direct cooperation, the results garnered from this original suggestion have gone a long way.

I am also thankful to David McAlpine for hearing out the idea in its initial and very raw form (along with Jörg Buchholz), and for getting me to pay attention early on to the inferior colliculus rather than to the auditory cortex as the main auditory hub.

Many thanks for Gojko Obradovic for reviewing the text and providing invaluable comments throughout.

Special thanks to Fabrice Bardy and Macarena Paz Bowen for audiological help in two measurements and for their general encouragement in these early stages.

Thanks also to Jody Ghani for her remarkable patience with transforming some abstract ideas into original technical drawings.

An early source of inspiration was a presentation given by James B. Lee, who provocatively tied together several ideas in optics, nuclear physics, and concert hall acoustics. He held quite unlikely talks in the 2016 meeting of the Acoustical Society of America in Honolulu (Lee, 2016a; Lee, 2016b), which attempted to conceptually link optics and acoustics in a way that was both fresh and insightful.

Sometimes people say things that resonate and can be later recognized in a completely different context in life. Such was something that my friend Ofer Meir said, who has also taken it upon himself to ensure that this work would see the light of day much sooner than it would have been without him. Thus, I am deeply grateful for his friendship and support.

The presence of close and loving friends throughout the process has been indispensable. Kelly Miles, who provided early inspiration and ongoing enthusiasm, Ophir Ilzetzki for invigorating discussions and sincere curiosity along the discovery process, Dotan Perlstein for his unyielding encouragement and for making sure that my feet remain on firm scientific ground, Jan Tomáš Matys for his unflinching confidence and moral support of this work, Augusto Bravo for his endless openness

for radical thinking and many fascinating conversations about science, Michael Yang for his firm friendship and taking this project seriously all along, and for Diogo Flores, who drove me with his excitement during the early parts of the work.

For their valuable advice at key moments along the way I am indebted to Timothy Beechey, Andrew Bell, Yaniv Ganor, Ami Goren, Barak Mann, Yehuda Spira, and Eduardo Vistisen.

For keeping things close at different points throughout the writing process and for their sincere support and friendship, I am grateful to Fadwa AlNafjan, Camilla Althoehn, Emily Arday, Javier Badajoz-Davila, Jerome Barkhan, John Beerends, Isabelle Boisvert, Ammalia Duvall, Jack Garzonio, Gady Goldsobel, Julia Gutz and Paul Springthorpe, Sharon Israel, Brent Kirkwood, Emilija Klovaitė, Rèmi Marchand, Jason Mikiel-Hunter, Juan Carlos Negrete, Allie O'Connor, Anders Pedersen, Heivet Hernandez Perez, Claudiu Pop, Matthieu Recugnat, Mariana Reis, Mariana Roslyng-Jensen, Ana Ruediger, Jeremy Rutman, Greg Stewart, Kramer Thompson, Lindsey Van Yper, Sarah Verhulst, Jay and Jane Woo, Jaime Undurraga, Gil Zilberstein and Stèphanie Èthier, and the late Liviu Sigler.

I am grateful to my sister, Oriyan Miller, and my brother-in-law, Roee Miller, for their continuous engagement, constructive questions, and much encouragement at numerous points throughout the discovery process.

And, finally, a huge thanks to my mother, Vivian Savitri, whose involvement and trust in this work have been essential from the early development of the ideas, which could have not been nearly as peaceful and resolute without her ongoing dedication to this project.

This work is standing on the shoulders of numerous scholars in hearing, acoustics, optics, communication engineering, information theory, signal processing, physics, neuroscience, and biology, without whom this treatise could have not been written. The comprehensive bibliography of this treatise is a testimony of my deep appreciation for their work that is frequently studded with unmistakable ingenuity. Of special help were the textbooks by Brian C. J. Moore (2013) about psychological acoustics and by James O. Pickles (2012) about the physiology of hearing, as well as the entire Springer series of handbooks on auditory sciences, edited by Richard R. Fay and Arthur N. Popper, which has been an indispensable source of knowledge of all things auditory. Two additional texts that provided an exceptionally useful overview of the historical and current state of hearing science from original perspectives were by Andrew Bell (2005) and Richard F. Lyon (2018). A book by Howard L. Resnikoff (1989) provided a unique point of view that related perception (mainly vision) to information theory and served as early inspiration. Finally, non-hearing texts that were used extensively were books about communication theory by Leon W. Couch II (2013) and Fourier optics by Joseph W. Goodman (2017). I was fortunate enough to attend a semester of Stephen G. Lipson's optical physics class in the Technion back in 1998, which evidently stuck for longer than I had realized at the time. The textbook of that course has thus remained an important reference throughout my writing (Lipson et al., 2010).

I could not have carried out this research without access to the Macquarie University Library, and to a lesser extent the Technion and Haifa University libraries. My gratitude goes to all the librarians that have worked behind the scenes.

About the text

This treatise is not intended to be an introduction to hearing science. It assumes that the reader is well-versed in basic hearing phenomena and has at least some acquaintance with wave physics, Fourier analysis, and linear signal processing. Readers that are already familiar with Fourier and geometrical imaging optics, as well as photography, astronomy, or microscopy enthusiasts, are likely to find certain optics-inspired passages relatively easy to follow—bordering on trivial. The same goes for readers with background in communication and radar engineering, who are going to find some sections relatively straightforward. While several chapters contain mathematical derivations that can be outside the comfort zone of the less mathematically-inclined readers, they are encouraged to gloss over them and focus on the qualitative descriptions that may be sufficient to develop the necessary insight. Nevertheless, a few topics should undoubtedly benefit from mathematical understanding— mostly those that introduce the basic space-time duality equations, the analytic signal, modulation and demodulation, coherence, and the various modulation transfer functions.

Some specific conclusions and derivations may raise interest among non-specialists as well. The derivation of the modulation transfer functions from the temporal imaging equations has not appeared in the optics literature previously, which has focused primarily on time-domain solutions.

There are several implications of the ideas expressed in this work that may also be of interest to perception, vision, and neuroscience specialists. If proven correct, then auditory imaging as presented here suggests that certain imaging principles are biologically common to both hearing and vision. This begs the question of whether additional sensory inputs are processed in a similar fashion, only with less obvious dimensional substitutions.

For the neuroscientist, the idea that the brainstem performs neural processing in hearing in part to achieve a function that is performed analogically in the eye may be curious as well. This suggests that biological computation is both analog and digital and that the segregation between the mechanical and neural domains may be at least somewhat contrived. It also underlines the significance of sampling considerations, which are usually taken for granted in the discussion of neural coding.

The scope of this work has been limited on purpose and largely excludes an in-depth treatment of some topics that have already received much attention in hearing science. Major topics that are only mentioned in passing are binaural processing, intensity and dynamic range compression, and lateral inhibition, as well as across-channel frequency weighting that is achieved by different segments of the auditory processing chain, which possibly contains the spectrotemporal modulated class of signals. The theory also formally deals with the auditory system up to the inferior colliculus, so higher-level effects (like attention or speech perception) are mostly avoided. Finally, in order to limit the scope of the literature reviewed, I tended to ignore most of the mathematical models of the associated ear parts, such as the cochlea, the auditory nerve, or of the complete system. These omissions notwithstanding, there was still much left to be explored in this work.

In the text, I have striven to remain agnostic about the particular cochlear mechanics that transduces the signal, as numerous publications and models have been exclusively dedicated to this problem and, for all I can tell, the jury is still out as for which one is (the most) correct.

Experimental data are still being reported regarding the cochlear mechanics and it is not unusual that they contradict predictions made based on different cochlear models or on classical observations done with outdated methods. However, I have posited two new functions of the organ of Corti (of a phase-locked loop and a time lens), which has made it almost impossible to retain this agnostic approach all throughout.

Chapter overview

The heart of this work—the temporal imaging theory—is contained in chapters §§ 10 to 14. Confident readers are encouraged to skim them before committing to the various introductory chapters. Results and implications are presented in the final chapters §§ 15 to 18, which are more qualitative in nature and likely have relevance to a wider audience within the auditory research community.

The introductory chapters survey a range of topics that are not necessarily new, but they attempt to tackle several acoustical issues in a fresh manner that is especially pertinent to hearing as a communication system that is embedded in a realistic world of arbitrary stimuli. Notable among them are chapters §6 about physical signals, and §7 and §8 about synchronization and coherence. Chapter §9 may be considered a standalone text that is introductory in spirit, but presents a novel hypothesis regarding the auditory phase-locked loop (PLL). This chapter was required for the assumption of coherence conservation between the external world and the auditory brain, but I believe that it may have far-reaching consequences beyond it, which are only superficially explored in the present work.

Appendices §A, §E, and §F feature results of small-scale measurements, which were necessary to corroborate some of the claims in the text and may be interesting in their own right, although only the first two may be understood without reference to the main text.

Some of the sections refer to audio demos, which can be found in the supplementary directories and are printed in SMALL CAPS. They are found in https://zenodo.org/record/5656125.

Below is an overview of the individual chapters.

Chapter §1 motivates the treatise and provides a brief review of current and historical hearing theories with emphasis on visual analogies. It dwells on existing attempts to define the acoustic object, the auditory image and object, and the inconsistencies and shortcomings they bring about. Using various physiological, functional, and physical considerations, it makes the case that a correct analogy between the ear and the eye has it that the cochlea of the inner ear is at an analogous level to the lens, whereas the inferior colliculus of the auditory midbrain is at an analogous level to the retina. A temporal imaging theory is then motivated using four additional perspectives: the prominence of direct versus reflected sound in hearing (unlike light in vision), imaging mathematics analogies between spatial and temporal equations, insights from communication about the physical transfer of information, and signal coherence propagation from the acoustic environment into the listener's brain.

Chapter §2 reviews the anatomical structure and physiology of the mammalian ear, with emphasis on humans, from the external ear to the auditory cortex. The review is deliberately high level in that it tends to neglect low-level details (e.g., cellular, biochemical) in order to crystallize a systemic perspective, where attainable. It is intended mainly for reference and for highlighting possible roles that have been attributed to the different components of the auditory system. The chapter concludes with a comparative section about some of the major differences between the auditory systems of humans and other mammals.

Chapter §3 presents a novel point of view on known aspects of real acoustic sources and environments. The idea behind this chapter is to highlight how the acoustics of realistic sounds and

environments diverge from classical linear descriptions. For this purpose, a general formalism is adopted for the representation of waves, which allows for straightforward incorporation of the concepts of dispersion, instantaneous envelope and phase, and group delay. The overarching difference can be boiled down to that between constant Fourier frequency representation and time-dependent complex envelope representation, which facilitates amplitude and frequency modulation. The effects of dispersion and other realistic acoustic signal degradations in realistic environments—both outdoors and indoors—are emphasized.

A very short introduction to physical optics is provided in Chapter §4, which revolves around spatial imaging. Several basic concepts in geometrical, wave, and Fourier optics are presented, as they provide the basis for the analogy with hearing in later chapters. The optics of the eye and the main elements in its peripheral physiology are presented. Finally, notable links and differences between imaging and Fourier analysis in acoustics and optics are mentioned.

Chapter §5 introduces a few basic information- and communication-theoretic concepts in a qualitative manner. Information theory is not applied directly in the work, but the physical propagation of information is taken to be the unifying element across the different stages of auditory processing. Several historical connections between information and hearing are briefly mentioned and it is argued that conservation of information—over the various signal transformations—has been taken as an implicit assumption of hearing theory. Actual communication systems can be described using generalized receivers and transmitters that deal with modulated signals. It is argued that hearing can be viewed as a communication system by assigning the appropriate roles of transmitter, channel, and receiver to the acoustic source, environment, and ear, respectively, and by recognizing that the intentional transfer of information is optional. The communication approach to information transfer is very similar to that used in simple spatial imaging, but there are some important differences between them that are highlighted as well.

Chapter §6 deals with the mathematical basis of physical and communication signals, which are used in all the theories that are relevant to this work. It begins from the analytic signal and the narrowband approximation, which gives rise to the important concept of instantaneous frequency. It then explores the roles of the temporal envelope and amplitude modulation in hearing and briefly reviews the role of phase in hearing, with emphasis on linear frequency modulation. Auditory phase perception has been a contentious topic, which gave rise to the concept of temporal fine structure as a proxy of auditory phase locking. However, several authors have indicated that the common way of applying these concepts in hearing has been inconsistent with the mathematics of broadband signals and with certain psychoacoustic observations. It is shown that the emphasis that has been put on the (mathematically) real envelope has led to auditory theory that treats hearing as a baseband (i.e., with low-pass characteristics, as though the system is capable of detecting sound down to 0 Hz) rather than a bandpass system. It is argued that a correct treatment of the system as bandpass is critical for embracing modulation and demodulation phenomena in hearing as a reality, rather than a metaphor. The existence of auditory demodulation along with a two-dimensional (carrier and modulation) spectrum are considered.

Chapter §7 is an exposition of the concepts of coherence and synchronization that are found in six different scientific fields that have some bearing on hearing: acoustics, optics, communication, neuroscience, auditory neuroscience and physiology, and psychoacoustics. While the essence of coherence as a concept may be shared between all six fields, it is obfuscated by the use of different jargons, sometimes for narrowly defined purposes. A standardized jargon is then proposed, which is used throughout the work. It largely adheres to the jargon used in optics of coherent and incoherent illumination and imaging, which overlaps with coherent and noncoherent detection in communication.

Chapter §8 draws heavily on optical coherence theory and summarizes its most important con-

cepts that include interference, the mutual coherence function, partial coherence, temporal and spatial coherence, coherence time, coherence propagation according to the wave function, spectral coherence, the effect of narrowband filtering, and nonstationary coherence. These ideas are then linked to data gathered about known sound sources and to the theory of room acoustics, which has used coherence rather sporadically. Other topics in hearing that relate to coherence are briefly mentioned as well, such as binaural hearing and coincidence detection. It is argued that partial coherence and coherence time have central roles in auditory perception. Appendix §A provides several quantitative figures from typical acoustical sources to substantiate some of the main claims of the chapter.

Chapter §9 introduces the concept of synchronization from nonlinear dynamical system point of view. It focuses on the phase-locked loop (PLL)—one of the most important circuits in communication engineering, control theory, and general electronics. It is shown how a PLL can conserve the degree of coherence of an input signal at the output. It is then hypothesized and demonstrated how the phase locking that characterizes the mammalian low-frequency hearing can be the result of an auditory PLL, which may be assembled from known functions of the organ of Corti and the outer hair cells: a phase detector from the distorting mechanoelectrical transduction channels, a loop filter from the outer-hair cell membrane, and the self-oscillating hair bundle as the voltage controlled oscillator. This hypothetical feedback process may be additionally amplified by the somatic motility of the outer hair cells and feed into the inner hair cell transduction path. Available evidence that supports this idea, as well as known gaps in the model, are discussed at length. The usefulness and likelihood of having dual coherent and noncoherent detection within hearing are discussed as well.

Chapter §10 derives the paratonal (originally, the "paraxial") dispersion equation that was first introduced in nonlinear optics and has been applied to scalar electromagnetic plane waves, but can just as well apply to pressure waves. This equation employs the space-time duality principle, which analogizes the spatial envelope to the temporal envelope of the wave field. The general solution of the wave equation requires the narrowband approximation—the decomposition of the wave into a fast moving carrier and a slow-moving complex envelope. Only the complex envelope is considered in the solution that has a fixed carrier. The solution requires the group-delay dispersion (or group-velocity dispersion), which is a fundamental property of the medium. It is expressed using the derivative of the standard (phase-velocity) dispersion and has not appeared in this name in acoustics before. The reciprocal operation to the group-velocity dispersion of the medium—that of a time lens, or a phase modulator—is also presented. Both group-delay dispersion and time lensing rely on quadratic phase transformations that can produce linear frequency modulation.

Chapter §11 goes through the signal transmission chain of the human ear and attempts to estimate its frequency-dependent dispersion parameters. The passive cochlea is known to be group-delay dispersive, and the magnitude of this dispersion is estimated to be much larger than that of the outer and middle ears. The group-delay dispersion associated with the neural pathways between the inner hair cells and the inferior colliculus—a quantity that has been considered to be negligible before—is estimated as well. It is speculated that the organ of Corti functions also as a time lens, and a physical principle of phase modulation is hypothesized, which has to do with the active change of stiffness that is caused by the outer hair cell electromotility. The time-lens curvature values for humans are roughly estimated based on animal data. The uncertainty in these estimates is very large and we derive approximate lower and upper bounds that may pertain to humans. Appendix §F offers an alternative derivation of the dispersion parameters using strictly psychoacoustic data instead of physiological data. The results are only partially consistent with the physiological estimates, because they introduce group-delay absorption to the parameters, which is largely ignored in this work and in optics, although it may be physically justifiable. Alternative explanations for the discrepancy are discussed.

Chapter §12 introduces the imaging equations, which require the three dispersive components to be in cascade—cochlear group-delay dispersion, time-lens curvature, and neural group-delay dispersion. The image of a pulse is computed and it is shown that it is inherently defocused in humans, based on the estimated parameters from Chapter §11. The same parameters are used to model available psychoacoustic data of the cochlear curvature in humans and high correspondence is found above 1 kHz. At lower frequencies, additional constraints of modulation bandwidth had to be introduced in order to obtain better estimates, although some uncertainty remains. The results also reveal the existence and the durations of the temporal aperture—a short sampling window that is associated with the different auditory channels. The estimates show close correspondence to additional human and animal data. It suggests that at high frequencies the aperture stop is determined by the auditory nerve, but at low frequencies it is determined by the cochlear filters.

Chapter §13 begins from the derivation of the impulse response of the defocused temporal imaging system for a single channel. The equations are then used to further derive the modulation domain transfer functions, which have not been previously introduced in optical temporal imaging, but are completely analogous to the spatial modulation and optical transfer functions from Fourier optics. Predictions are compared to available observations of the auditory temporal modulation transfer function. Interesting predictions and discrepancies are highlighted, where the dependence of the results on the degree of coherence of the stimulus is argued to be key.

The effect of sampling of continuous signals by the neural system is explored in Chapter §14. While sampling has been invoked several times in hearing models, the consequences of discretization have not been fully considered before. The significance of nonuniform sampling is discussed with respect to the modulation transfer functions, which are thought to degrade (lose modulation bandwidth) upon repeated resampling that occurs downstream, within the auditory pathways. The tradeoff between nonuniform sampling noise and aliasing from undersampling, as is known to take place on the retina, is hypothesized in the context of hearing. A psychoacoustic experiment that is interpreted as demonstrating the existence of auditory sampling is presented in Appendix §E, with emphasis on the effects of aliasing.

Chapter §15 explores in greater depth the idea of an auditory image based on all the previous findings and the principle of space-time duality. The concepts of sharpness, blur, focus, defocus, and depth of field are discussed, and a simple computation of the system temporal acuity is presented, based on its impulse response or the modulation frequency discrimination. A formal presentation of polychromatic images is made and pitch is discussed as a special case in auditory imaging that manifests in different ways. A subset of image aberrations from optics that can be relevant to hearing are discussed with speculations about the most significant auditory aberrations in humans. Ideas from masking theory are extrapolated to examine how supra-threshold stimuli sound in the presence of other sounds. Furthermore, nonsimultaneous masking is analogized to the auditory depth of field that applies temporally and is exaggerated by the signal processing of the auditory system. Most of the imaging effects considered are well-known auditory phenomena that are reinterpreted in light of the concepts of temporal imaging. Seven rules of thumb for auditory imaging are proposed that epitomize some of the analyses in the chapter.

Chapter §16 takes these ideas a step further and hypothesizes what an auditory accommodation function that is roughly analogous to accommodation in vision could be like. Different mechanisms are proposed for parameters that can be shifted within the system. The operation of the olivocochlear efferent bundle appears to be key in accommodating the PLL gain and/or the time-lens curvature. The plausibility of other mechanisms of accommodation is discussed. The coherence of the stimulus is a recurrent key parameter in the analysis that the system is hypothesized to react to. The idea that the system may be combining coherent and incoherent imaging products in different amounts is considered and complements the earlier discussion made in the context of coherent and noncoherent

communication detection in Chapter §9.

Chapter §17 brings together the ideas of auditory imaging and coherence and sets to find out if they can be used to shed light on known hearing impairments. Evidence for dispersive shifts in hearing-impaired individuals is examined and their effects are considered. Additionally, the possibility of aberration and accommodation impairments is explored, also with analogy to eye disorders. Out of the different impairments considered, accommodation disorders and excessive higher-order aberrations appear to have the highest likelihood to be detrimental, but more conclusive relations to known hearing disorders cannot be made before the entire temporal imaging theory is elucidated and more relevant data become available.

The treatise closes in Chapter §18, where a functional model of the hearing system is presented that encompasses all of its standard parts, as well as the imaging components and the PLL, which are roughly mapped to the different auditory organs. The chapter concludes with a general discussion that highlights open questions, limitations, weaknesses, and merits of the ideas presented in this treatise. It also highlights several topics that can benefit from future experimentation. Finally, certain novel ideas and issues that appear in this work are mentioned, which may find interest outside of hearing research alone.