Chapter 18

General model and discussion

18.1   Introduction

Several ideas that are new to hearing science have been introduced in this work. While a suitable imaging theory for the auditory system has been the centerpiece of this work, a prominent role has been given to both communication and coherence theories, which have been considered to be indispensable stepping stones for the development of a comprehensive auditory theory. The inclusion of these perspectives within the acoustic signal path that is received by the auditory system had to be motivated by a de-idealized presentation of real-world acoustics compared to what has been traditionally considered in hearing science texts. However, this realistic and more complex account is rewarded by being able to readily produce hypotheses for explanations of a wide range of auditory effects that have so far been largely left out of standard theory. This repackaging of the auditory narrative also produced promising links to both vision and neuroscience—either in way of analogy (to the eye) or through the integration of analog and neural information processing paths (the auditory brainstem and inferior colliculus). Several speculative elements and notable uncertainties have been foundational for the development of this work, but they will eventually have to be rigorously tested, in order to obtain confidence in the more advanced aspects of the new theory.

In this concluding chapter, the full functional model of the auditory system is considered, as aggregated from the previous chapters. It will facilitate the discussion of some of the strengths and weaknesses of this work, which will lead us to make a few suggestions for improvements and further investigations. The chapter concludes with a short discussion of some of the overarching themes that go beyond hearing alone, but can be impacted by the theory.

18.2   A provisional functional model of the mammalian auditory system

18.2.1   Introduction

Various functions of the auditory system, which have not been considered previously in standard theory, have been presented in this work. Given the high number of known and hypothesized functions of the auditory system, it will be useful to try and put together the newfound elements along with the familiar ones in a single model. This should help us synthesize a more cogent understanding of what the auditory system actually does and identify the strengths and weaknesses of the proposed theory.

There are two major difficulties in coming up with a model that encompasses the entire au-

ditory system, at least up to the midbrain. First, the operational details of the cochlea are not uncontroversial. Especially, the organ of Corti and the outer hair cells (OHCs) are implicated in several functions—amplification is chief among them—whose functional inner workings are not entirely proven empirically and whose intricate mechanics is still being studied. It affects other important hearing characteristics, such as the auditory filter sharpness and distortion product generation. Second, the exact roles of the auditory brainstem remain fuzzy. The various nuclei are correlated with the detection of multiple acoustic features and specialized responses, which do not necessarily add up to clear, reducible, or nameable functions. Therefore, discussions about this critical part of the system tend to be somewhat vague so much so that entire nuclei and pathways have little more than conjectured roles associated with them (see §2.4).

There are three dominant classes of auditory models in literature. The first class treats the acoustic input as a time signal and uses standard signal-processing techniques—both linear and nonlinear—to account for the various transformations that the signal undergoes. Typically, the processing gives rise to some desirable auditory feature extraction or a specific response, which can also be cross-validated psychoacoustically or electrophysiologically. This model class does not always adhere to known physiological correlates, which should carry out the signal processing operations in reality. Nevertheless, it is useful in simulating and predicting different experimental responses and can be highly insightful in forming hypotheses about the physiological roles of the different auditory organs. Occasionally, the simplest models in this class have no particular processing that can be attributed to the brainstem (e.g., a "decision making device" that immediately follows the rectification of the transduced signal), or to any other auditory nuclei for that matter. In this and similar cases, these nuclei are occasionally (and implicitly) explained away as "relay neurons".

The second class of models usually traces the various neural pathways and their connections and tries to infer what their function is. These models emphasize the excitatory or inhibitory, and ascending or descending nature of of the projections. Occasionally, there would be additional emphases on the types of cells and neurotransmitters involved, and on their relative topological positions. In their most basic form, these models are nothing more than a connection diagram, which tends to be both vague and complex, due to multiple pathways and cells with no clear role in the emergent system function. A subset of these models hypothesize particular circuits that assume a desirable operation (e.g., acoustic feature extraction), which comprises specific neuron types with known responses to particular stimuli. In these models, the circuit interconnections and complete operations can be hypothetical, or based on partially known correlates that are available from physiological data. Yet more advanced physiological models relate directly to the biophysical properties of the cells, such as the different time constants of presynaptic and postsynaptic potentials, depolarization thresholds, and complex timing delays. This approach enables precise simulation of the neural signal between selected parts of the system, which can mimic the response of known acoustic signal types.

The third class of models combines elements from both signal processing and physiology and attempts to provide a phenomenological account of hearing. Such models can be highly effective, although they do not necessarily bring out any intuition as for why the system architecture and signal processing is the way it is.

18.2.2 The model

The explanation level that has been sought after in the present work lies somewhere in between these three model classes. Its focus is on the purpose of the auditory system as a whole and the functions that are required to realize it. The functional realization, though, is not necessarily unique, and more than one signal processing or physiological configuration can be conceived that can plausibly

perform it. Therefore, the model grossly localizes some functions in the system physiology, but does not commit to the details of operation. Furthermore, because of the relative poverty of parameters that were computed here with certainty, this kind of system description does not yet offer a clear path to simulation and remains high-level.

This section briefly presents how the various old and new components form a single auditory system. The provisional model is illustrated in Figure 18.1 and includes both well-known components (in black), as well as new ones that are hypothesized in this work (communication in green and imaging in blue). There are still some uncertainties, which are most prominent in the brainstem, where the exact function is less obvious. Below is a general description of the model.

The top section of Figure 18.1 includes the standard communication and information system building blocks (see Figure 5.1). The "acoustic transmitter" is made from an optional information source that modulates the acoustic object. The object radiates energy into a noisy acoustic channel and is then received by the ear—the "auditory receiver". The channel additionally introduces various distortions to the signal, as was described in §3.4. The auditory receiver input port is the outer ear, which is responsible for spectral shaping of the acoustic wave caused by the scattering from the pinna and the concha, and for emphasis of canal resonance frequencies. Additionally, the ear canal allows only plane-wave propagation (below 4 kHz, in humans), which is a key requirement for imaging. The middle ear receives the vibrations from the eardrum and acts as an acoustic impedance transformer, yet unlike standard transformers, it is frequency dependent and produces a reactive component in its output. This linear mechanical structure injects significant power to the signal. The middle ear also receives control input from the acoustic reflex efferents, whose circuit is omitted from the diagram.

The middle section of Figure 18.1 continues from the stapes / oval window, where it is coupled to the cochlear perilymph. The vibrations are transformed to a traveling wave in the basilar membrane (BM) from the base to the apex. A traveling wave excites the tectorial membrane as well, which is also omitted from the figure. Cochlear group-delay dispersion is indicated by the uncoiled BM that is shaped as a prism (in the diagram), which also exhibits the basic (broad) frequency selectivity—indicated with a bandpass filter bank. Only the most basal channel is displayed all the way through, to avoid clutter in the diagram. In the apex, there is some uncertainty about the bandpass nature of the filters, which is illustrated with a low-pass filter.

The hypothetical time lens is due to the phase modulation inside the organ of Corti. It is followed by the standard but not yet empirically settled cochlear amplifier, which is a function of the OHCs. The novel phase-locked loop (PLL) function is also associated with the OHCs and drives the inner hair cells (IHCs), which transduce and rectify the stimulus (for details about the hypothetical PLL circuit itself, see §9.8). If these three functions (time lens, amplifier, PLL) are indeed realized by the organ of Corti, then it is possible that they are either nested or occur in parallel, and not in series, as is displayed in the figure. So, a gain stage was assumed to be part of the auditory PLL, but it may also operate standalone.

Different efferent inputs to the OHCs are indicated on the diagram with no elaboration (medial olivocochlear, MOC) and IHCs (lateral olivocochlear, LOC). Slow Type-II afferents from the OHCs are indicated as well. The main output from the cochlea to the auditory nerve is through fast Type-I afferents that receive the rectified input from the IHCs. Each neural firing represents a sample of the continuous stimulus (or the deflection of the IHCs) and its temporal window constitutes the temporal aperture stop of the system, which is drawn as a vertical opening. Implicit in the figure is the high number of auditory nerve fibers for every hair cell (§2.5.2). In the brain, (nonuniform) sampling is always accompanied with encoding (see §14), which appears in the same block.

The bottom section of Figure 18.1 roughly illustrates the auditory brain function. First, neural group-delay dispersion is placed at the input, but should be distributed in the entire brainstem, in a manner that is unknown at present (with the exception of data by Morimoto et al., 2019 of
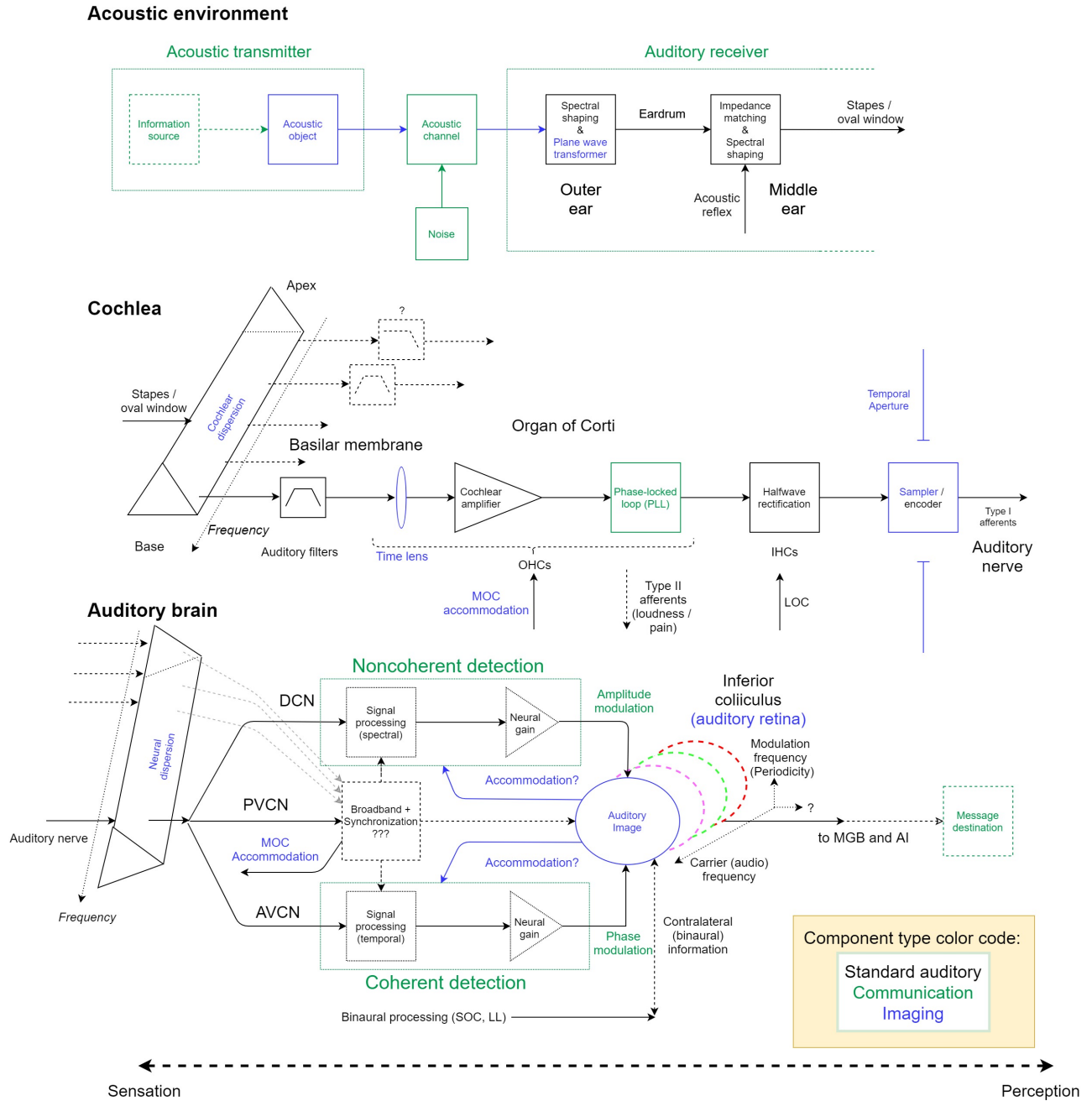
Figure 18.1: A functional diagram of the provisional monaural auditory system, which contains standard elements (in black), as well as communication elements (in green) and imaging elements (in blue). The dotted lines around the message source and destination indicate that they are optional. See text for details.

differences between waves I and V, see §11.7.2). From the auditory nerve, the neural pathway splits to three branches (acoustic stria): the dorsal cochlear nucleus (DCN), the posteroventral cochlear nucleus (PVCN), and the anteroventral cochlear nucleus (AVCN).

This work adopted the communication engineering differentiation between coherent and noncoherent detection. Combining it with various findings about the cochlear nucleus in mammals and their avian analogs, the following rough functions are hypothesized for the two main branches of the DCN and AVCN. The DCN is hypothesized to primarily realize noncoherent detection functions, which mainly imply spectral and (real) envelope processing. The AVCN is hypothesized to realize the coherent detection functions, which usually imply temporal processing and also include the binaural processing of the system (not shown in the figure). Hypothetical neural gain function is indicated in both branches, which may be varied using accommodation efferent input from the inferior colliculus. Neural gain—or really, attenuation—may apply the relative weighting of the coherent and the noncoherent detected products that arrive to the midbrain to form a partially coherent image. Auditory accommodation may modulate other circuits beside gain as well, so it is displayed with no clear target. The function of the PVCN is not entirely clear, although it is known to involve broadband processing and provide synchronization to periodic stimuli. This is realized using coincidence detectors with variable delays from different parts of the cochlea—effectively canceling out some of the dispersion from earlier processing. The PVCN is also the source of the MOC reflex, which we considered to be part of auditory accommodation as well. While not shown in the diagram, the neural system continues to nonuniformly sample the stimulus beyond the auditory nerve.

We have referred to the inferior colliculus (IC) as the auditory retina in this work, since it is where the image appears and the different processes converge, in analogy to the retina that contains extensive convergent neural circuitry within vision (§1.5.2). Its structure is laminar and each lamina processes a narrow band of carrier frequencies. In addition to tonotopy, periodicity—or rather modulation frequency that is more applicable universally—seems to be at least one additional orthogonal dimension of the image. In this model, only the coherent, noncoherent, binaural, and broadband inputs are shown, although almost all lower auditory nuclei project to the IC (Figure 2.4). As a communication target, the image is also where demodulation takes place (as detection). Like the retina, there is considerable signal processing and information compression that may be taking place in the IC. Given that in some mammals it was found that specific IC cells are sensitive to primitive (tuned) stimuli types, it suggests that the analogy between the IC and the retina must not be oversimplified to cast the IC as a simple screen and a detector.

The main output from the IC is to the medial geniculate body (MGB) and it continues from there to the primary auditory cortex (A1) and the rest of the cortex, where perception is thought to emerge. Potentially, it is the destination of where a message that was acoustically sent is being received, which complements the optional information transmission on the other side of the processing chain. As was discussed in §5.3.2, the communication system can work more or less independently of the potential intent to use it to send messages. This destination resides around the same areas where perception may be taking place. In contrast, auditory sensation may encroach into the auditory brainstem. Therefore, the borderline between the two is not well defined and we do not imply that perception can be truly localized to one area of the brain.

The auditory model in Figure 18.1 represents the main general functional blocks in the system, but remains relatively vague or agnostic regarding several key areas—mainly of the auditory brainstem. Some of the gaps in the model highlight the gaps in the present theory on the whole, whereas others reflect more general gaps in knowledge that are inherited from hearing and brain research, as well as the coarse grained perspective of this model. Either way, it contains certain speculations that will have to be tested in experiment. These gaps and weaknesses are reviewed in the next section.

18.3   Known gaps and weaknesses in the proposed theory

There are several aspects of the theory that have been left out of this work, either by design, or because of lack of sufficient knowledge and available methods at present. Below are several known gaps and weaknesses in the theory, as it stands at the time of writing. Some of these gaps can be appreciated directly from the model in Figure 18.1.

18.3.1   Analytical approximations of the dispersive paratonal equation

The dispersive paratonal equation that was originally derived by Akhmanov et al. (1968, 1969) has been adopted from optics without changes (except for renaming it). While both acoustical and electromagnetic fields are assumed to be scalar, the acoustical problem deals with the audio frequency range that is many orders of magnitude lower than light. This is correct as long as the slowly-varying envelope condition is satisfied, which ensures that the group-velocity dispersion can be expanded around the carrier frequency and is well-approximated up to second-order. These conditions are synonymous with how we defined the paratonal approximation.

Two additional strong assumptions were made to be able to work with a tractable equation that may have inadvertently skewed some of the predictions. First, absorption curvature (or gain dispersion) was neglected, as is customary in optical imaging. The somewhat less common linear absorption (§3.4.2) was also not considered. Linear absorption is known to have phase effects that are similar to dispersion (Siegman, 1986, pp. 358–359) and can be similarly incorporated into a complex group velocity (Eq. 10.14), which is a physically unintuitive concept and was not pursued here. However, even though the data about absorption are unavailable, we cannot rule out that the absorption curvature is of the same order of magnitude as the dispersion. This uncertainty is exacerbated by the abnormal low-frequency results we obtained in the cochlear curvature modeling (§12.4), where the channels may no longer be narrowband. Also, the strict psychoacoustical modeling (§F) has turned out complex parameter predictions that are only partially consistent with the mixed physiological modeling of §11. Consistency between the data sets may have to eventually incorporate absorption curvature. The uncertainty in the parameters is also compounded by the lack of distributions and confidence intervals for the obtained predictions (both physiological and psychoacoustical).

The second strong assumption we made was of medium linearity. A level-dependent group-delay dispersion would require a nonlinear paratonal equation that takes the form of the nonlinear Schrödinger equation. It is sometimes applied to model the amplifying medium that is placed inside the resonator of a laser device, and it is arguable, therefore, that the active parts of the cochlea should be similarly modeled. While this equation has a closed-form solution (Shabat and Zakharov, 1972), applying it here would have greatly complicated the analysis at this initial exploratory stage and would rest on shaky grounds with respect to the modeled cochlear medium. Most of the dispersive physiological data were obtained from low-level measurements at 40 dB SPL, which could warrant a linear treatment, but these results were then mixed with others that were obtained with observations done at higher levels. It should be noted that the level dependence of the phase may be negligible around the characteristic frequency (CF) in the cochlea and the auditory nerve (Geisler and Rhode, 1982). The extent of the nonlinear effect in more central auditory areas is unknown.

Adding level effects to the simple imaging equations should make it easier to incorporate compression into the theory, which has been purposefully avoided in order to simplify the analysis. The exclusion of level effects also entails that hearing threshold effects are not considered. As the spike rate in the auditory nerve is lower with low-level inputs, the sampling operation will be affected (degraded) by level as well, as reconstruction (should it exist) depends on less spikes—fewer samples.

18.3.2   Multiple roles assigned to the organ of Corti

Two new roles were hypothesized for the organ of Corti and the OHCs: PLL and time lens. That these two functions are not in contradiction to one another, or to the amplification function that is often associated with the OHCs, is uncertain.

The PLL model is framed in general terms and has a relatively strong empirical basis to it—both in supportive physiological and psychoacoustical evidence, and in the known properties of the PLL as a general model for a nonlinear oscillator (§9). However, the associations and details of the different components that are hypothesized in the complete loop (i.e., the phase detector, filter, oscillator, and feedback coupling) may require a future update when new observations are made available. Perhaps the most daring speculation in the auditory PLL model is assigning a central role to the quadratic intermodulation distortion product, and assuming a role for internal infrasound information, which is currently not backed by direct evidence (§9.8.1). It also assumes a critical role for the self-oscillations of the OHC hair bundles that can synchronize to external coherent stimuli. Despite the relative confidence in the PLL model, none of its parameters were estimated in the present work and none are available in literature about phase locking. Elementary specifications such as the order of the PLL, pull-in range, lock-in time are all missing. These parameters affect, for example, the PLL ability to retain lock of a linear chirp.

The time lens function has been derived on a more phenomenological basis (§11.6). While the physical mechanism that was presented for acoustic phase modulation should not be particularly controversial or complex, the existence and derivation of its magnitude has been based on four studies that yielded two clustered sets of curvature values. All four required us to invoke the cochlear frequency scaling assumption to convert the data to humans, on top of the uncertain time lensing mechanism, and a very large gap between the estimates from the two animals. The analysis also intersected with the controversy regarding the true bandwidth of the auditory filter in humans, which has not been settled yet. This led us to repeatedly consult a broad range of curvatures throughout this work. Most effects were rather insensitive to curvature changes, but most appeared more adequate with the large curvature estimate—either based on broad-broad or narrow-band auditory filters. The uncertainty is compounded by the high likelihood that the time-lens curvature is adaptive, by virtue of the auditory accommodation—the MOC reflex (MOCR). Hypothetically, this mechanism could account for the existence of the small-curvature estimates of the time lens. The connection between the time lensing and the MOCR was deemed attractive in that it may be able to accommodate the auditory depth of field, although no direct proof was available to corroborate this idea.

It should be remembered that temporal imaging can be achieved without a lens, using a pinhole camera design (Kolner, 1997). The very short auditory pupil function may be analogized to a pinhole, so such a design is not completely far-fetched. However, a stronger constraint that motivated the acceptance of the time lens existence is the common-sense requirement that the magnification of the auditory system should be positive and close to unity. A negative magnification would entail a time-reversal of each sample of the envelope. Unless the cochlear and neural dispersions were misestimated in both magnitude and sign, unity positive magnitude cannot be achieved without a lens. The only other test we identified for directly assessing the time lens has been to model the psychoacoustical octave stretching effect using the image magnification, which is dominated by the time lens magnitude (§ 15.10.1). This approach seemed promising but highly imprecise using the available data. All in all, the status of the time lens remains less than conclusive.

18.3.3   Neural group-delay dispersion

The neural group-delay dispersion was computed from the difference between simultaneous auditory brainstem responses and evoked otoacoustic responses of the same subjects (§11.7). This calculation cannot differentiate the different neural paths that exist between the auditory nerve and the IC and therefore it imposes a blanket dispersion to all the different parts of the auditory brainstem. Although it is not impossible that the different pathways have about similar dispersion dependence, they are unlikely to all be exactly the same. Initial data on the chirp slope that is required to maximize either wave I or wave V of the auditory brainstem responses, suggested that the auditory nerve and brainstem have different dispersions (Morimoto et al., 2019). In the general hypothetical model of the complete auditory brainstem of Figure 18.1, the different pathways are specifically assumed to have different functions—coherent and noncoherent detection. In the imaging theory, each one corresponds to another extreme solution using the same neural dispersion. However, in reality, the parametrization of the different path dispersions may produce different effects. Supporting evidence to this effect may be found in frequency-following responses (FFR) of amplitude modulated tones in humans, which were associated with different group delays and brain areas that generated the envelope synchronization and phase locking to temporal fine structure (King et al., 2016).

18.3.4   Modulation filtering

Throughout the auditory brain, there are units that are tuned either to carrier or to modulation frequencies, or to both. The existence of bandpass modulation-frequency filters is not accounted for by the present theory. Such filtering can be analyzed using the various modulation domain transfer functions. The analog in optics would be of spatial filtering, which is achieved in the inverse domain and affects the modulation band—the spatial frequencies. This optical technique provides relatively simple image processing capabilities, which can be implemented also in analog means (e.g., by constructing a special pupil function). The exact roles of these filters in the auditory brain may be specific to the species and its acoustic ecology. The degree to which these filters are eventually perceived may vary, especially with consideration to integration across channels.

18.3.5   The role of the PVCN

A particular challenge in the present model is the third pathway from the auditory nerve—the PVCN—whose function is not very clear. Neurophysiological research has mainly dealt with the DCN and AVCN and in the past has not always distinguished between the AVCN and the PVCN. Interestingly, the avian auditory system has analogs to all the main auditory organs in the brain, but has only two pathways split from the auditory nerve, which correspond to the AVCN and to a combination of the DCN and PVCN (see §2.4). The roles of the two pathways can fall in the purview of the assumed coherent and noncoherent blocks from the communication model. However, despite several anatomical and morphological differences between humans and other mammals, there is little difference in their hearing in terms of general mechanisms and responses.

The literature suggests three main differences between avian (mainly songbirds) and human hearing perception (Dooling et al., 2000): 1. Lack of avian high-frequency hearing, with a bandwidth that is usually cut off below 10 kHz. 2. More acute temporal processing (shorter time constants). 3. Higher sensitivity to small and fast details within sound sequences, which go unnoticed by human listeners. Although these differences (especially 1) may be attributed to anatomical differences in the cochlear and hair cell structure, they are most likely related to higher-level mechanisms. What

is particularly baffling about 2 and 3 is that they exhibit superior performance to humans[178], despite an absence of an entire auditory pathway. Therefore, a more refined model may be helped by comparative research between mammals and birds.

18.3.6 Accommodation

Accommodation was introduced in §16 as a plausible feature that the auditory system likely has, in analogy to vision, except that it is more deeply embedded in neural mechanisms. Several possible mechanisms were considered, but they were all speculative to different degrees. Given the amount of efferent connections in the system, it is not hard to imagine that some particular circuits can be recast as accommodation. However, the present theory remains uncertain as for which one it is. Perhaps the most likely hypothesis is related to the degree of coherence of the image, which is a function of the source coherence, and the products of the coherent and noncoherent detection pathways. This can then corroborate an interpretation of the MOCR system as one that adapts the time-lens curvature and the PLL gain to dynamically skew the processing to be more coherent or noncoherent, depending on the stimulus and situation, and thereby also affect the auditory depth of field. In addition, this dual processing logic may entail modulation of the relative neural gain of the pathways that are responsible for each kind of processing. Other mechanisms may exist that accommodate almost every other parameter in the system, either as part of an accommodation reflex, or independently. In all cases, the number of unknowns stands in the way of a comprehensive understanding of the system operation.

18.3.7 Coherence model

Throughout the analysis, we have appealed to coherence considerations to motivate a large number of effects, but it was done without specifying a particular model to compute auditory coherence. Although it should be possible to use the various formulas for continuous coherence in the traveling-wave domain, they may lose validity in central processing. Also, the nonstationarity that characterizes typical acoustic signals has to be matched with proper time constants that may be longer than the aperture time—perhaps depending on the processing stage. Furthermore, there are clearly two dimensions of coherence—both within channel that is applicable to classical imaging and communication, and across channel where object coherence (in its psychoacoustic sense) comes into play.

   Another important aspect of coherence that we relied on is that the weighted sum of the coherent and incoherent images gives rise to the partially coherent image that "appears" on the IC. The details of the specific images, their weightings, and how they are accommodated are missing too, although the logic of this operation is fairly straightforward. However, the combination of these two images has been a central point of this theory and will have to be better formalized.

18.3.8 No simulation

No simulation of the complete auditory imaging system has been provided in this work. Such a simulation will have to be attempted when the dispersion parameters and pupil function of the system are known with more certainty. This simulation requires numerical computation that has several hurdles to overcome, depending on the method used (Goodman, 2017, pp. 121–153). First, discretized quadratic transformations (as defocusing implies) require computational oversampling

---

[178]Superior performance here means that less information is lost. Alternatively, if information loss or integration of information over long time frames is key (Weisser, 2019), then the avian hearing superiority may be questioned.

to avoid aliasing of the output. Second, nonuniform sampling must be implemented as well, to adequately represent the neural transduction and spiking, yet this is not a standard element of the optical methods. The auralization of such transformations may have poor sound quality due to the apparent periodicity pitch of the low-frequency sampled pulses, or noisiness, depending on the sampling regularity and the type of detection assumed. This means that an anti-aliasing filter has to be implemented, unless aliasing can be converted to noise through nonuniform sampling, as theory predicts. A third issue has to do with the broadband configuration of the entire set of auditory filters, which have to be combined in a way that is able to preserve features of both amplitude and intensity imaging.

A related problem in simulation can take place in the hypothetical reconstruction stage of the sampled signals, which requires using unmodulated carriers that correspond to the CF of the auditory filter, especially in coherent detection and imaging. It is not unlikely that the combination of the inputs from multiple fibers and channels is essential for the production of a perceptually seamless and continuous sound. It should be remembered, though, that the auralized response may not sound good without appropriate decoding that is not entirely known. Perception deals with an auditory code that is recoded several times before the smooth tonal experience is perceived, downstream. The pitch percept, for example, may be likened to color, in the sense that it is not directly caused by the carrier frequency, but is partly determined by place-coding that is transformed to a higher-level representation in the cortex, which corresponds to pitch.

18.3.9 Across-channel response

Even though the significance of the polychromatic image has been emphasized, the present theory is based on single-channel analysis and it remains uncertain about how across-channel integration and broadband response is achieved in reality. Physiologically, it can be achieved in a few ways. Broadband units (e.g., the octopus cells) are the simplest solution that has been proposed so far, but there are few observations that demonstrate innervation by multiple channels, which can facilitate the cross-channel integration (for example, McGinley et al., 2012). Another option is that the channels are interconnected, say, by the interneurons in the brainstem, and have the ability to mode-lock, which gives rise to enhanced harmonicity sensitivity (§9.7.2). This may work in tandem with the helical geometrical structure of the dorsal lateral lemniscus, as was hypothesized by Langner (2015). In the same work, Langner (2015) argued for a powerful neural model that combines units from the DCN, PVCN, and VCN, which together detect the broadband signal periodicity by synchronizing to its envelope phase. The resultant periodotopy on the IC laminae is orthogonal to the carrier frequency tonotopy. In principle, Langner's model may apply to aperiodic or quasiperiodic modulation and also to unresolved periodicity, which is sometimes referred to as residue pitch. However compelling this model may be, it is unknown at present whether the exact circuitry that it hypothesizes exists in the brainstem.

18.3.10 Lateral inhibition

Lateral inhibition is crucial in both vision and hearing, but has not been treated in this work and, rather, it has been implied throughout, as is sometimes done in auditory models. It should be distinguished from lateral suppression in the cochlea that is more automatic and perhaps less goal-oriented. Because of the substantial overlap between the auditory filters, it is important to consider the effect of lateral inhibition that defines stable channels and affects the integration of information between channels. Linearly frequency-modulated (FM) stimuli, which have been repeatedly considered as the staple signals in temporal imaging, are particularly prone to excite multiple auditory channels over a

short time span (depending on their slope). Thus, these signals inherently challenge the simplistic single-channel models that have been presented in this work. On the one hand, we know that the auditory system "takes care" of the integration, so that these stimuli always sound seamless, even as they cross several channels. On the other hand, it requires an active weighting process across channels, which has to include the signal phase as well. In any case, inhibition should play a role in modeling these processes.

18.3.11   Binaural hearing

A major topic that has been deliberately left out of the analysis is binaural integration of monaural images. The binaural system in hearing has often been studied as an independent subfield, partly due to what appears to be dedicated pathways from the superior olivary complex to the auditory cortex and beyond. The fact that almost all auditory pathways converge in the IC should provide some clues for an integrated monaural-binaural model.

All that said, binaural perception is readily included within a general coherence theory of hearing, since coherence has been long used in binaural research. According to this reformulation, interaural cross-correlation is really a spatial coherence measurement (§8.5). Additionally, the familiar interaural time and level differences are also readily seen as special cases of the complex degree of coherence measurement that are sensitive to different cues of partial coherence. The entire structure of the binaural system is reminiscent of an interferometer, as was first noted by Cohen (1988), which is the kind of measurement device that produces the most precise data in physics. Therefore, interferometry science may have something to say about the ears as well, although this direction has not been consulted until very recently (Dietz and Ashida, 2021).

18.3.12   Neuromodulation

The topic of neuromodulation was briefly invoked in §16 as a possible general mechanism to realize auditory accommodation. Mounting evidence suggests that various neuromodulatory pathways in the auditory system, including the brainstem and midbrain, can have significant effects on sound processing that take place over different time scales. At present, these effects are excluded from all standard auditory models, which may create the false impression that the signal processing of the auditory system is static, or that plasticity occurs only on a high and slow processing level. Undoubtedly, inclusion of these effects greatly complicates the understanding of this system, but it also has the potential to bring its modeling much closer to reality, by introducing various points of calibration and plasticity that can be individually matched.

18.3.13   Multiple time lenses and PLLs

We modeled the auditory channel as though it contains a dedicated time lens and PLL, which were associated with the apparent uniformity of the hair cells and supporting cell tissues in the organ of Corti. This implicit assumption is unsubstantiated at present. The PLL, in particular, depends on a local oscillator for its function, which we took as the hair bundle self-oscillations, knowing that in vitro, hair bundles of different hair cell types tend to oscillate (Martin and Hudspeth, 1999). However, inasmuch as the spontaneous otoacoustic emissions (SOAEs) represent existing oscillators in the cochlea, they portray a rather limited picture—only a subset of subjects have measurable SOAEs and their retuning is not as flexible as we would like to see in a general-purpose PLL. A charitable interpretation would then be that the oscillations are usually too faint to be measurable, that they are suppressed by their neighboring oscillators, or that they remain active only during

supra-threshold stimulation that does not allow for external measurement in a favorable signal-to-noise ratio. Furthermore, we do not know what the effect is of chaining parallel PLLs that may all couple and synchronize to one another. Can they be isolated? How do they suppress one another? For a treatment of a related problem of coupled nonlinear oscillators in the cochlea, see Wit and Bell (2017).

Another difficulty in the theory is that mechanical phase locking does not work much beyond 4 kHz in mammals, whereas the OHCs extend to much higher frequencies. There, the PLL function cannot work, unless it is somehow able to synchronize directly to the envelope. Another more remote option is that the OHCs are desynchronized at high frequencies to the extent that they decohere coherent signals and gradually facilitate their power amplification and noncoherent detection.

To the extent that they are independent of PLLs, multiple parallel time lenses may also give rise to complex nonlinear effects, although less consequential, perhaps. The layout of small and channel-specific time lenses might be compared to compound eyes of insects and crustaceans, which do not have a single crystalline lens, but rather numerous small optical units (**ommatidia**)—complete with their own lenses that are directly connected to a few photoreceptor cells (e.g., Nilsson, 1989). Unlike the ear, the ommatidia are not coupled, or at least not in such a complex way. But maybe somewhat like the many channels of the ear, the images from every ommatidium have to be integrated into a whole in the animal's brain. This interesting association has not been explored any further.

18.4   Novel contributions to auditory theory

Despite its shortcomings, it is important to not lose sight of the insights that the temporal imaging theory has to offer to the understanding of auditory theory more generally. The primary ones are briefly discussed below.

18.4.1   Temporal auditory imaging

The temporal imaging theory, as was introduced into optics by Kolner and Nazarathy (1989) and Kolner (1994a), has been applied to the mammalian auditory system (§§ 10 to 13). The theory treats complex-envelope pulses as the objects to be imaged and operates in the realm of dispersive media. Such a foundation makes little sense in an acoustic world that comprises mostly pure and complex tones and where AM-FM modulation is a rarity. However, once we let realistic curvature-ridden signals and environments replace the idealized stimuli—sounds that had been inherited from an idealized musical world and may have occupied the collective auditory scientific spotlight as a consequence—a dispersive point of view becomes much less puzzling. This is the domain where paratonal sound processing resides.

The introduction of a rigorous imaging framework to hearing enables us to make intuitive predictions that are based on analogies from optics and vision, as long as two substitutions are made: the spatial envelope (the monochromatic object) should be replaced with the temporal envelope, and, in some cases, the distance from focus of the optical object should be substituted with the degree of coherence of the acoustic signal. These substitutions open the door for auditory focus, sharpness, blur, depth of field, aberrations, and accommodation as valid and useful concepts in hearing. A third substitution between color and pitch is more complex, though, because the relative bandwidth of hearing is considerably larger than vision that enables harmonic perception, and auditory phase perception allows for coherent (interference) phenomena that do not exist in vision that is strictly incoherent.

Temporal imaging requires several second-order phenomena in the auditory system to assume relatively prominent roles in the signal processing chain. Predominantly, these phenomena include

the group-delay dispersion of the cochlea and the auditory brainstem pathways, the exact temporal window shape that is achieved in transduction and works as a temporal aperture, and more speculatively, the phase modulation in the organ of Corti that works as a time lens. In current literature, these parameters have been either neglected or presented as physiological happenstance rather than essential to the signal processing of the ear. This neglect is not coincidental and is unsurprising in the case of unmodulated tones or signals with low-frequency modulation. Such stimuli are usually unaffected by the typical low-pass character of the modulation response and they do not disperse easily. This is also because the associated curvatures have very small magnitude, which has little influence on sustained narrowband signals. Additionally, the auditory filters work well within the narrowband approximation in normal-hearing conditions, so low-frequency modulation response generally dominates unresolved auditory stimuli. In contrast, high-frequency modulation that is more prone to dispersion gets resolved in adjacent filters and may become noticeable only in broadened filters and through a hearing impairment. Therefore, relatively imaginative experiments may be needed to uncover exceptional cases in this system.

Another interesting point, which has not been dwelt on in too much depth in the text is the significance of the pupil function in imaging. In optics, the imaging system is completely described by the generalized pupil function, which contains information about resolution, blur, and aberrations. In case we are interested in the demodulated product of the hearing system, then the auditory pupil function should be equally important. In the text we have used a Gaussian pupil that is analytically tractable and seemed to have yielded several useful results and insights, despite its oversimplified nature.

18.4.2 Sampling

An additional layer of theory that is required for the temporal imaging to work is the treatment of neural transduction as a sampling process that may or may not be amalgamated with the neural encoding operation. This is not a new idea in hearing theory, as various psychoacoustic and physiological models assumed the discretization of the stimulus (see §14.2 and §E.1). However, sampling has not been embraced as a standard part of the auditory signal processing chain and is rarely considered a mandatory function of the auditory neural code. Whenever a discrete representation of sound is employed, sampling is usually taken for granted and sampling theory is not consulted. This means that several interesting sampling-related effects have not been traditionally considered in hearing such as aliasing, the instantaneous sampling rate effect on temporal resolution, or the tradeoff between aliasing distortion and noise that can be achieved with nonuniform sampling (as was demonstrated spatially in the retina; see §14.7).

A related effect that is under-explored in hearing is the possible interaction, or even downright equivalence, of neural adaptation and the notion of having a variable sampling rate (§E). Nevertheless, it should be plausible and even intuitive that the auditory system densely samples the signal around the onset—the temporal analog of spatial edges—to provide higher-resolution information about the points where the signal exhibits the most variation due to transient effects.

18.4.3 Modulation domain

A central point that this work emphasized is the conditional independence of the carrier and the modulation domains—independence that applies only in strict narrowband and stationary conditions that are not particularly realistic. The idea of a separate modulation spectrum and periodicity maps is not new and can be traced back to Licklider (1951b). But even when it is discussed at length, the subtext in hearing theory has been that the significance of the modulation domain is secondary to

the carrier domain. This is mathematically, physically, and perceptually misleading and yet it is not a far-fetched impression that one may get by exclusively considering unmodulated (stationary) stimuli. The present theory attempts to dispel some of the confusion regarding modulation in hearing by maintaining consistent representation of the acoustic wave (the object), and the communicated signal, which is eventually perceived as an image. For the price of restricting the usage of Fourier frequencies that are unvarying and extend to infinity mainly to carrier frequencies, we obtain a unifying mathematical framework that encompasses several theories that are relevant to hearing. A corollary of this move is the recognition that the envelope and temporal fine structure are interdependent, so that AM-FM signals are inevitable, which reflects a signal representation that has a carrier and a complex envelope, rather than a frequency-modulated carrier and an amplitude-modulated envelope, separately. This representation also facilitates synthesis with coherence and PLL theories.

This point of view also highlights the informational content of the sound source object. In hearing, the acoustic signal contains useful information both in the carrier and in the modulation domains. Due to the finite bandwidth of the auditory filters, there is some fluidity between the carrier and modulation domains for signals with high modulation rates that are on the limit of being resolved by adjacent filters. This means that temporal scaling of the signal, or changes in the filter bandwidth can result in information moving from one domain to the other. Unresolved high-frequency modulation is theoretically more prone to phase distortion (over-modulation)—that is, to image aberration. Unresolved high-frequency modulation is also constrained by undersampling errors, which can potentially make the quality of received information lower than the resolved version of the same signal. Aberration, over-modulation, undersampling, and movement between carrier and modulation domains have rarely if ever been treated in hearing theory, so it is hoped that they may inspire more holistic ways of thinking about normal and impaired auditory processing.

18.4.4 Coherence and defocus

The key parameter that repeated in the context all all theories that were synthesized here is the degree of coherence of the signal. Even though coherence has been used occasionally in hearing theory, its application has been inconsistent and has lacked rigorous ties to scalar wave coherence theory. Along with the static frequency interpretation of acoustic time signals, this may have underpinned an idealized point of view in which coherence is an auxiliary signal processing concept, rather than a dynamic property of the acoustic source and transmitted signal. In reality, signals tend to be neither coherent nor incoherent, but rather, partially coherent.

The differential processing of incoherent and coherent signals under defocus can readily explain the defocus unmistakable existence within the auditory system. With a degree of coherence that can vary over the entire range between completely incoherent to completely coherent, it is reasonable that the auditory system takes advantage of its defocusing property to further differentiate these two signal types, as a key to segregate different sources. As signal informativeness is often implied by its coherence, an ability to manipulate signals accordingly may provide a considerable advantage for hearing in noisy, reverberant, and other distorting conditions. Moreover, coherence plays a dominant role in imaging theory and readily distinguishes between coherent and incoherent modulation transfer functions—something that has not been previously accounted for in auditory theory (re broadband and tonal temporal modulation transfer functions, TMTFs).

18.4.5 Coherent and noncoherent detection

Once we accept that information is carried by signals that are not mathematically idealized and that coherence is a field property that propagates in space, hearing becomes amenable to be treated as

a communication detection problem. Known signal detection techniques are distinguished as either coherent or noncoherent, which intuitively refers to the similarly worded imaging optics distinction, but is not exactly the same. Coherent detection preserves the carrier phase and generally relies on a local oscillator in the receiver, whereas noncoherent detection does not. Both detection types are advantageous in different conditions. Coherent detection is ideal for signals in noise and for carrier synchronization applications, whereas noncoherent detection is much easier to implement and less prone to failure as a result. Each method is capable to suppress noise from the other kind: noncoherent detection can provide immunity against coherent noise, and vice versa. Noncoherent detection strongly resembles the traditional power spectrum (and phase deaf) hearing model that has been dominant and effective in many applications, despite accumulating cracks that have been revealed in this approach (§ 6.4.2). Many of the countervailing observations against the power spectrum model have to do with various phase (temporal fine structure) effects that are thought to be related to the phase locking property of the auditory system.

The unifying perspective proposed here is that the auditory system has a dual detection capability, which can universally receive arbitrary signals. Such a system should dynamically optimize its detection according to the coherence of the signal, its environment, and various internal factors of the listener. On a high level, it functionally calls for auditory accommodation system that is somewhat analogous to vision. We were able to identify a range of processes that can fall into the purview of accommodation, but much still remains to be revealed. In particular, the idea that the organ of Corti functions as a PLL that can facilitate coherent detection and perhaps respond to accommodation—probably in unison with its effect on the time lens—has the potential to clarify several murky points in the present understanding of the system. These ideas may be a step toward understanding the interrelationship between phase locking, envelope synchronization, temporal fine structure, and the elusive duality in processing (i.e., spectral and temporal) that the system seems to manifest in almost any task.

Nonetheless, contrary to the traditional distinction between coherent and noncoherent detections, it appears that the system eventually combines the two detected products and produces a partially coherent output. This strategy is accounted for much more readily with principles from imaging optics, rather than communication engineering alone.

18.5   Missing experimental data

Part of what makes the theoretical claims in this work speculative is the relative poverty of direct data that could be harvested from literature to substantiate them. This is understandable given the small magnitude involved in the associated quantities, their obscure function, and a general attendance to other auditory properties that have been better motivated. In this section, we highlight some types of data that are most sorely missing for this theory.

18.5.1   Human dispersion parameters

At this stage, the auditory temporal imaging theory rests on insufficient objective/physiological data that were assorted from several sources—usually in different measurement conditions (i.e., species, stimulus level). We also generalized from data measured on very small populations, although we do not know what the spread of the dispersion parameters is in the normal population. Therefore, better estimates of the frequency-dependent cochlear and neural group-delay dispersions are needed—no matter how small in magnitude they appear to be.

Possibly, alternative methods should be conceived to obtain these estimates of the dispersion parameters that require less transformations or corrections (e.g., from animal models to humans,

from cadavers to living, or from high to low frequencies). In the present estimates, there was no way to altogether avoid employing some methods that have not been mired in some controversy. Therefore, some dataset choices (e.g., picking one dataset out of the four derived neural group-delay dispersion calculated that are physically plausible, §11.7.3) were preferred because they provided a somewhat superior prediction to independently measured data (e.g., the psychoacoustic curvature measurements §12.4). While the differences have usually been small and the results rather stable across several datasets, more certain data are clearly needed to establish sufficient confidence in the theory and predictions.

Perhaps the most uncertain parameter is the time-lens curvature when the lens is "relaxed". Then, if it indeed turns out that the curvature can be accommodated, then its bounds in humans should be clearly established, as well as the conditions that lead the system to actively accommodate to.

Another parameter that has been largely missing is the pupil function—its shape and effective temporal aperture and their frequency dependence. Importantly, any noninvasive methods that can be developed to directly measure this function in humans can go a long way to simplify the derivation of individualized transfer functions. Should dispersive pathologies exist in the listener's hearing, they should appear in the generalized pupil function.

Both auditory brainstem response (ABR) and otoacoustic emissions (OAE) are particularly attractive candidate methods for noninvasively determining the various parameters in humans. However, the underlying theory—especially of OAE—must be settled before they can be harnessed for these tasks. Furthermore, both ABR and OAE may require some corrections to precisely segment the desirable dispersive element without contributions from the adjacent dispersive segments.

Ultimately, superior estimates of the various dispersive parameters should decisively determine what the correct configuration of the auditory system is and whether it matches the one that has been championed in this work. This means that if either the cochlear or neural dispersive path curvature is measurably zero, then the possible image is going to be different and may entail that some of our conclusions are erroneous. If an active phase modulation capability producing the time lens effect is not found, then we would be dealing with an auditory pinhole camera design, which is similar but not identical to our model imaging system and could offer less flexibility in hearing. The combination of these three components determine the focusing capability of the system, which, for all we know, may be configured completely differently between different animal species. Therefore, generalizations between species must be done with care.

18.5.2   Behavioral data

A critical section in this work was based on high-quality cochlear phase curvature data by Oxenham and Dau (2001a), which allowed us to validate the physiological predictions and derive estimates for the aperture time. However, these results are based on relatively few subjects ($N = 4$) and are extremely demanding for listeners. A modified method that is based on Békésy's tracking technique was demonstrated, which could reduce the measurement time per frequency from 45 minutes to 8 minutes only (Rahmat and O'Beirne, 2015; see also Klyn, 2015). While an even shorter measurement duration will be preferable, this is undoubtedly a promising direction.

Ideally, the individual time-lens curvature can be measured in a more robust way than by using the stretched octave effect, which is also a tedious measurement. More importantly, if time-lens accommodation is confirmed, then the adequate "relaxed" curvature should be established (analogous to the emmetropic state of the eye). Then, the limits of the smallest and the largest curvatures should be established across the population along with the required conditions to elicit them.

Another perspective that can be rather readily tested is how normal-hearing and hearing-impaired

listeners rate or respond to stimuli that can be classified as sharp or blurry. A continuum can be generated between the two extremes, based on different principles of (de)coherence, which listeners can judge. Listeners' willingness to apply this measure and their consistency may be revealing. Ideally, such results can be correlated to the acoustical features of the signals. This may provide additional data about possible dispersive impairments that some subjects may experience.

Better controlled data of various psychoacoustic effects may be useful to establish whether a completely behavioral test battery can be used to fully characterize the dispersion parameters of a listener. This was attempted in §F with mixed results that were not straightforward to interpret, but which could have benefited from better controlled data. This refers to octave stretching, frequency-dependent temporal acuity, and beating data, in addition to the phase curvature using Schroeder phase complex.

It will also be important to find out what the individual spread is in the normal hearing population. We relied mostly on (small) population averages, but it is unknown how relevant these values are for individual listeners.

18.5.3   Acoustics

On the physical acoustics front, there is a total lack of literature on the group-delay dispersion in various media, as well as on methods of estimating it. The straightforward way is to directly obtain it from the second-derivative of the frequency-dependent phase function. However, this quantity may be generally very small, so measurements should be done carefully due to possible sensitivity to small changes in position, boundary conditions, equipment, etc. These data should be complemented with the group-delay absorption to find out how it compares to dispersion. Additionally, it can be useful to confirm that the Kramers-Kronig relations are indeed applicable in typical media using bandlimited measurements.

Group-delay dispersion measurements should include the phase response of the outer ear canal, which were estimated to be negligible, yet tend to erratically fluctuate between positive and negative values (§11.2). Most curious may be testing for the possibility of dispersion distortion, which is caused when the same modulation information is carried by different modes with different group velocities and spatial distributions. This is expected to take place in the vicinity of the eardrum at frequencies above 4 kHz, but the size of this effect is unknown—especially given the small acoustical path involved relative to the wavelength. This is a new distortion to acoustics, which can have important implications on information transfer fidelity.

Acoustic phase modulation constitutes the time-lens operation (in the quadratic approximation), but has not been investigated in acoustic research. However, there should be multiple ways to achieve it, depending on which medium parameters are being modulated (e.g., density, compressibility) and over what spatial and temporal extent. This is an uncharted territory that may produce interesting theoretical and technological insights in acoustics.

Finally, detailed coherence data that describe the range of natural and synthetic stimuli are missing. The data provided in §A are limited in scope and were mostly presented to get an idea of the range of values that are involved and the effects of room acoustics, nonstationarity, and analysis filter selection, among others. Such data would be required in order to establish a more rigorous coherence theory for the auditory system itself.

18.5.4   Cross-disciplinary digging

Communication theory and imaging Fourier optics have been used as the central sources of inspiration and analogy for the present theory. However, the analogy breaks down due to the idiosyncrasies of

the auditory system, such as the large number of overlapping channels, the dual nature of the system that leads to partially coherent images, the mixed acoustic-neural domains, the multi-purposedness of hearing, and the mechanical complexity of the cochlea. Still, both communication and optics have a wealth of theory, methods, and designs, along with extensive empirical records, which undoubtedly contain additional clues that are relevant for the hearing system. For example, the topic of ultrawide-bandwidth transmission and reception is relatively new in communication engineering and it matches the auditory system bandwidth, by definition. This topic has not been consulted in this work more than at the most superficial level, and is expected to harbor ideas that are relevant to hearing.

18.6　Overarching themes

Given the interdisciplinary nature of this work, some of its assertions may impact closely related disciplines outside of hearing. This chapter concludes with a short discussion of a few of them.

18.6.1　Imaging as a unifying sensory principle

This work has demonstrated how mammalian hearing can be reframed as a temporal imaging system, which can be formulated with the analytical tools developed for light and spatial imaging optics. While hearing and vision have been repeatedly juxtaposed and contrasted throughout history (§1.3), the present theory provides the most comprehensive account to date for the extent of this analogy. Specifically, an auditory image has been rigorously shown that shares the basic conceptual properties of the visual image such as focus and blur. However, the dominant dimensionality of the auditory image is "antonymous" to that of vision—temporal instead of spatial—and the number of carrier channels involved is orders of magnitude larger than in human vision. The physical signal that is applicable to hearing is also associated with completely different frequency and wavelength ranges, which results in a different operating principle than vision—coherence-dependent focus, instead of distance dependence. Therefore, the resemblance and analogy between the two image types is nontrivial. However, there are enough parallels between the two, which can lead to some form of theoretical union. It implies that some general theoretical principles can be studied only once and be applied to both vision and hearing as special cases.

　　This begs the question—how universal is the imaging operation? Hearing requires complex energy transformations, but seems to have evolved a similar mathematical solution as vision, at least in humans and many mammals. It is also known that in some species of lizards and tautara (see Figure 2.5), the pineal complex has an eye-like organ—the parietal eye that has a role in thermoregulation—complete with a lens, cornea, retina, and photoreceptors with sensitivity overlapping the visual portion of the electromagnetic spectrum (Tosini, 1997). Does that imply that other sensory channels may operate using similar principles? On the face of it, it seems highly unlikely, as the vast majority of senses are not based on radiated waves, as hearing and vision are, where the dimensional transformation is relatively neat. It may be that other sensory modalities require transformations that are not at all trivial and are intuitively challenging, even though they become comparable on the level of the cortex and assume similar mathematics to imaging along their sensory pathways.

18.6.2　Perception and sensation

Perception and sensation are usually treated together in psychology and philosophy texts, although the emphasis on perception is significantly greater. Visual perception is typically taken as the gold standard for perception in general (§1.4.4). This approach mirrored in the present work for practical

reasons, as the auditory system has turned out to resemble vision in some unexpected ways. But we have argued throughout that the auditory periphery encroaches well into the central nervous system, much like the eye, since the optic nerve is part of the central nervous system. It means that the borderline between auditory sensation and perception is fuzzier than in the visual system. It is generally understood that sensation operates on a low level of processing of the peripheral nervous system, which implies that it is automatic and unconscious. The sensory input is therefore conveniently fed into perception, which occurs somewhere in the brain and is typically associated with the cortex.

However, in the present theory we have a system for which this usual role division fails, inasmuch as the auditory physiology reveals either the sensory or the perceptual roles. The evidence for this is that the same operation that the visual system performs in optical periphery (diffraction), hearing performs in the brainstem (dispersion). It does not imply that there is consciousness that is associated with the high-level sensory perception. But it does raise the question of whether perception is produced incrementally and only manifests fully in the cortex. Our theory is therefore reminiscent of the ancient Greek philosophers' view, who did not have separate notions for sensation and perception (Hamlyn, 1961).

18.6.3   Analog and digital computation in service of sensation

Another facet of sensation is related to biological computation. We obtained a mathematical analogous function in the auditory neural domain and in the visual peripheral (ocular) domain. Traditional views (e.g., Marr, 2010) take for granted the analog information processing of the eye, prior to the retina, which is responsible for the sharp visual image. In contrast, the analogous auditory brainstem function is part of the signal processing of the system. We have a fully neural system that realizes a function (group-delay dispersion) that the eye does with the vitreous humor (the eyeball)—diffraction. In strict information processing terms, it means that the eye works as an analog computer, whereas the ear is neural (if we shy away from using the word "digital" in the context of the brain).

The inclusion of analog computation in the biological toolkit is not a subtle narrative change, because it demonopolizes the neural system from being the only biological element that can process information. At least in hearing and vision, biology has been more accommodating in utilizing diverse physical information processing principles that were available through evolution to achieve the desirable computational goal.

See Sarpeshkar (1998) for further ideas about combining analog and digital information processing in the brain and also MacLennan (2007) for a discussion about analog computation in biology.

18.6.4   Fast and slow processing

Our analysis has made explicit a dual processing strategy of the auditory system, which has been in plain sight for a long time, although it has not been called by name (§9.11). In the simplest terms, the ear combines coherently and incoherently detected images into a partially coherent image. One of the main differences between the two are the time scales. Coherent processing requires phase locking to carrier frequencies that may have to be matched with high sampling rates, whereas incoherent processing can track the slow envelope, thereby discarding much of the phase information with no apparent loss of auditory performance (i.e., speech intelligibility). This kind of processing may be simpler to accomplish, as it aggregates signals over several channels and is not particularly "fussy" about instantaneous changes of the signal. It likely requires slower sampling rates and can be performed with much more relaxed tolerance, which may free up brain resources to other tasks, if

necessary.

This fast and slow processing is remarkable in that it emerges at such a low level in auditory processing (in the cochlear nuclei). Such ideas have been popularized in various guises of brain processing, but never at such a basic processing level. For example, according to Stanovich and West (2000), reasoning may be attributed either to "**System 1**", which is inherently fast and requires heuristics that can lead to wrong conclusions, or to "**System 2**", which is slow and effortful, but is more logical and potentially precise. In another influential brain theory, McGilchrist (2009) associated processing of details with the left hemisphere and processing of the whole with the right hemisphere, so the two operate on two wildly different time scales.

If the dual auditory processing hypothesis will be confirmed, then it may have the potential to show that the distinction between fine-grained and coarse-grained processing is a fundamental processing feature of our brain. It may even work as a recursive principle that can be repeatedly applied at different levels of processing.

18.6.5　The auditory literature

A large portion of this work involved the harvest of existing auditory research literature for available data. Having reviewed what feels like innumerable papers for the sake of this research, I have no doubt that there is much relevant data that wait to be utilized, which can potentially save resources for truly novel research. Many of these publications have shown an immense level of accomplishment and creativity, which I can only be grateful for that I could access. By embracing more literature from the past, I believe that it should be possible to unify additional concepts that appear disparate today. This process may also be used to rule out long-standing hypotheses that no longer have merit. The alternative may entail endless divergence with growing challenge to close off loose ends and settle open questions that may otherwise be left unanswered.