Chapter 16

Auditory accommodation

16.1 Introduction

Equipped with a rudimentary understanding of what distinguishes sharp and blurry auditory images, we are now able to explore the final prominent functional analogy between vision and hearing: accommodation. This adaptive feature is indispensable in vision and provides the basis for the sharp optical imaging when objects are positioned at arbitrary distances from the eye. In a broader context, the accommodation of the lens is part of a reflex that includes the adaptation of the pupil size and the convergence of the eyes that enables a proper combination of the two images in binocular vision. While there is much evidence of various adaptations that the auditory system makes over different time scales, none has been framed in analogy to ocular accommodation.

Because of the uncertainty about the exact architecture of the auditory imaging system, potential accommodating elements will be considered in terms of their physiological plausibility, supporting empirical evidence, and hypothetical advantage for hearing. Thus, should auditory accommodation exist, it may manifest as variable cochlear group-delay dispersion, time-lens focal time (curvature), neural group-delay dispersion, duration and/or shape of the aperture, auditory filter bandwidth, instantaneous sampling rate, synchronization regularity, coherent/incoherent product weighting in the inferior colliculus, or any combination thereof. Dynamic range compression, middle-ear acoustic reflexes, and other mechanisms that dynamically vary the gain applied in the signal chain—"**gain accommodation**", perhaps—are outside the scope of this work and will be mentioned only tangentially in relation to the medial olivocochlear reflex.

This chapter briefly reviews the main features of ocular accommodation. It then proceeds to hypothesize an imaging-centered analogous function of auditory accommodation along with some evidence that can be used to support this idea. The different parameters that can be hypothetically accommodated are then considered in terms of their usefulness and likelihood to exist. Finally, based on the analysis, we further hypothesize that so-called "listening effort" is an emergent response to difficulties in auditory accommodation.

16.2 Ocular accommodation

Accommodation is the automatic process that dynamically adjusts the optical power of the eye to maintain sharp focus on a visual target, with respect to its distance. The following brief summary— largely based on a review by Charman (2008)—attempts to communicate only aspects that may have relevance for the hypothetical analogous mechanism in hearing.

The biomechanical process in which the crystalline lens in the eye is stretched and flattened

(accommodated) for distant vision is a complex coordinated action of ciliary muscles, which are connected to the capsule that contains the lens (see Figure 4.7). A fine interplay of elastic forces is then responsible for the changes in shape and power of the lens. The accommodation of the lens is always accompanied with synchronized horizontal rotation of the eyes—**convergence** (also called **vergence**), which ensures that the two images lie in the central field of both eyes, so the visual cortex can produce a fused binocular image. Additionally, accommodation is generally accompanied with changes in the pupil constriction. These three processes are sometimes referred to as the **accommodation reflex** (also, the **near reflex** or the **near triad**), but only the change of the lens focal length is considered as accommodation proper. Normal accommodation is binocular and symmetrical, just like the pupil constriction, as the two eyes assume closely matched muscular movements (Campbell, 1960; Flitcroft et al., 1992). When asymmetrical targets are presented to the two eyes, the accommodation sets on a lens power that better matches the farthest object (Koh and Charman, 1998).

Accommodation is controlled automatically by the Edinger-Westphal nucleus in the midbrain that is located at the level of the superior colliculus. Control is done primarily via the (parasympathetic/cholinergic) oculomotor nerve, which also controls convergence and pupil constriction, although some minor inhibitory sympathetic innervation exists as well (Gilmartin, 1986). Despite its automaticity, it is possible to trigger accommodation by consciously attending to the object distance, or in the case of some individuals, to voluntarily control accommodation. The accommodative change of the lens power takes about 300 ms—a duration that most likely increases with age—with an additional 1000 ms that are needed for stabilization. However, even then, the focusing precision is imperfect and can fluctuate and may cause blur that is comparable to that cause by some of the higher-order aberrations of the eye.

The optical stimulus that cues the visual system to accommodate its focus has been debated extensively and several mechanisms may be at play (Toates, 1972). In recent studies, optical wavefront-vergence (not to be confused with the convergence reflex mentioned above) has been consistently shown to cue accommodation even in monocular monochromatic vision (Marín-Franch et al., 2017; Del Águila-Carrasco et al., 2017), in line with a classical hypothesis by Fincham (1951). This means that the eye is sensitive to the direction of retinal defocus, which corresponds to whether the blurred wavefront is convergent or divergent with respect to the focus. The response to these cues, however, is facilitated exclusively through foveal cones, before the control information is neurally fed back to the ciliary muscles (Toates, 1972).

16.3   What may auditory accommodation be like?

As was argued in the last chapters, auditory imaging is different from visual imaging, primarily because the acoustic phase information is slow enough that it can be processed directly by the neural system—at least at low carrier frequencies. This means that the intrinsic defocus of the system has the potential to differentiate sounds according to their degree of coherence. In particular, defocus differentiates the incoherent and coherent temporal modulation transfer functions (TMTF) according to their cutoff frequencies. In analogy, we would like to get a better understanding of the hypothetical auditory accommodation function, but also distinguish it from accommodation in vision[165].

Visual objects are critically characterized by their distance from the lens, which determines the

---

[165]The word accommodation has been occasionally used in auditory research in reference to a host of adaptive behaviors that involve listening over time (e.g., Holt and Kluender, 2000; Carlile et al., 2014). However, these word usages appear to have no direct relevance to auditory accommodation as is defined here.

focal length that the lens must assume for the eye to achieve sharp focus. Additionally, when the light intensity is high, the pupil (the spatial aperture) closes and lets less light into the system by eliminating high spatial frequencies, which causes an increase in depth of field and, as a result, sharper imaging. When this happens, the tolerance of precise focal accommodation can be somewhat relaxed, depending on the degree of sharpness afforded by the instantaneous depth of field. The situation is quite different in hearing, where depth of field manifests temporally as nonsimultaneous masking and appears to be exaggerated by the system for some types of acoustic objects. Therefore, unlike vision, the most relevant independent variable of the acoustic object is the degree of coherence of the source in its environment, rather than the distance from the object, which is factored into the propagated coherence function only indirectly.

16.3.1   Relevant empirical evidence

Several studies have demonstrated an apparent adaptation that the auditory system makes for reverberation, which results in improved reverberant speech intelligibility scores when the system is given sufficient time to adapt. The interpretation of this effect has been challenged in that it can be elicited without reverberation, only through the manipulation of the modulation depth and spectrum of the stimulus. After reviewing these results we analyze them with a coherence framework instead. Subsequently, it is maintained that auditory accommodation may be understood rather organically by considering the function of a dual coherent-noncoherent detection system.

The initial series of studies that may offer a window into accommodation-relevant hearing was pioneered by Watkins (Watkins, 2005a,b). In these studies, listeners were presented with a word token, "*sir*" or "*stir*", whose phonological category boundary was discretely varied between these two words using interpolation. The words were embedded in a carrier sentence and subjects were required to determine which of the two tokens they heard. But, the subjective categorical judgment between the two words can be impaired by reverberation, which reduces the temporal envelope cues necessary to hear the /t/ in "*stir*". Watkins (2005a) found that if the listeners are given a congruent acoustical context before the target—matching reverberation of both carrier and token—they are able to compensate for the blurring effect of reverberation, as the boundary between their categories is unaffected by the reverberation. This was unlike presentations with incongruent context, for which the token and context acoustics did not match, which caused a shift in the perceived categorical boundary. This effect was stronger for rapid speech than it was for slow speech presentations. Further, diotic presentations produced a larger effect than dichotic ones, which suggests a monaural mechanism. Also, single reflections enabled a similar but smaller compensation, maybe because they lack the long tail of reverberation decay that can inform the subjects about the appropriate acoustic context (Watkins, 2005b). Tests with narrow- and wide-band noises that were temporally comodulated as the speech context, showed that compensation becomes more effective the wider-band the context is and is almost completely ineffective with single auditory-channel-wide contexts (Watkins and Makin, 2007).

The interpretation of the auditory effect as a mechanism that specifically adapts to reverberation was challenged by Nielsen and Dau (2010). They replicated the experiment done by Watkins et al. using carriers with no reverberation: white noise, a silent interval, speech-shaped noise, and that same noise with random amplitude modulation at 4–8 Hz. In all cases but one, the category boundary was preserved as in the congruent case, despite the absence of any reverberation cues. The modulated speech-shaped noise showed a slight change in the category boundary, but was still close to congruent results. The authors suggested that the observed effect is the result of forward masking in the modulation domain, because the context in the near condition has a larger modulation depth, so it tends to cause larger modulation forward masking than the other signals that have smaller

modulation depth. Thus, if adaptation takes place here it may relate to the modulation and not the reverberation. This alternative interpretation was partially addressed by Beeston et al. (2014) using additional data that included the silent interval. They found that the adaptation was fine-tuned enough to have to rely at least in part on the reverberation information within the token word and not only on the preceding carrier context.

Perhaps a more compelling indication for adaptation to reverberation may be inferred from a related series of studies, which generalized the findings by Watkins et al. through testing of speech-in-noise thresholds in a room acoustical context. In several experiments it was found that when subjects were pre-exposed to the room acoustics they had an average of 2.7 dB improvement in performance, compared with the unexposed conditions (Brandewie and Zahorik, 2010). However, this improvement did not translate well to anechoic conditions (Brandewie and Zahorik, 2010; Zahorik and Brandewie, 2016). It was optimal (up to 3 dB improvement) for reverberation times between 0.4 and 1 s, but was reduced with a longer reverberation time of 2.5 s (Zahorik and Brandewie, 2016). Additionally, in order for the exposure to the room acoustics to be effective, listeners had to listen for a minimum of 850 ms (Brandewie and Zahorik, 2013). Listeners were insensitive to the exact location within the room, which strongly affects the spectral weighting of the response—a sort of a "room constancy" effect (Brandewie and Zahorik, 2018; cf. Weisser, 2004). Unlike Watkins (2005a), binaural listening was usually found to improve performance (Brandewie and Zahorik, 2010), although the test design in this case relied on contralateral presentation of speech and noise, which makes it difficult to generalize (Zahorik, 2019). In comparative behavioral tests of gerbils, which employed startle reflex responses as proxy for the threshold of sinusoidal amplitude modulation (AM) of broadband noise, reverberation compensation was not established (Lingner et al., 2013). This is maybe due to the gerbils' smaller head size and lack of corresponding binaural cues. It was suggested that gerbils may have never evolved to deal with reverberation in the first place, which makes the direct comparison to humans less telling.

Several physiological studies in animals have also demonstrated reverberation compensation in how AM signals are coded at the level of the inferior colliculus (IC), yet no compensation in the ventral cochlear nucleus (VCN). Single IC unit recordings of unanesthetized rabbits found less synchronized responses to sinusoidal AM narrowband noise in reverberant conditions compared to anechoic responses (Kuwada et al., 2012). Synchrony, spike rate, and neural gain were about constant for source distances of up to 40 cm, and degraded at larger distances, even after controlling for level, in what appears to be the result of growing interaural decorrelation (i.e., spatial decoherence). The same neurons usually responded to azimuth as well. However, when examined in individual azimuths, higher synchrony and gain were recorded than would be expected from the loss of modulation depth, suggesting again a compensatory mechanism for reverberation (Kuwada et al., 2014). Slama and Delgutte (2015) replicated and refined these results, by observing that the greatest reverberation compensation occurred for IC neurons that exhibited the most modulation depth compression, which itself may have had a cochlear origin. While envelope distortion, spectral coloration, interaural envelope disparities, and average interaural coherence could not explain the results, a subset of the neurons did respond to the interaural cues. These results are contrasted with recordings in the anesthetized guinea-pig VCN, where all unit types (primary, sustained/onset choppers) showed reduced temporal synchrony to pitch (speech intonation) in reverberant conditions (Sayles and Winter, 2008; Sayles et al., 2015).

16.3.2   Synthesis

The main effect of reverberation is to decohere the signal, so only the direct sound (the early portion of the signal) may have its phase information relatively intact, to the extent that it can

be detected coherently. The indirect, or reverberant, sound that arrives later is either partially coherent or incoherent, so its phase function cannot be recovered. For this portion of the signal, noncoherent detection or incoherent imaging are necessary, whereas a strictly coherent detection may be unnecessarily noisy. The speech signal itself poses a similar challenge in detection, since it contains a complex mixture of coherent and incoherent sound elements, which renders the speech partially coherent. Any loss of modulation depth can be directly associated with the impulse function of the room, and hence with its reverberation time (Schroeder, 1981), and effectively, with the degree of decoherence. Therefore, from the system engineering point of view, it makes sense that the auditory system adjusts its detection according to the specific combination of signal and acoustics, so that its detected product at the output can be used to extract the desirable information from the signal. The two interpretations reviewed in the previous section—that the system adapts directly to reverberation, or to the modulation depth changes—both refer to the same physics, but attribute the effect to different levels of explanation. We argue instead that attributing the changes to the degree of coherence is the most parsimonious level of explanation. This argument is very similar to the one that was made in the context of auditory depth of field in § 15.11.3. Either way, all interpretations require some form of time-variant signal processing that is not generally considered in classical models of hearing.

It can be concluded that insofar as the reverberation compensation effect exists, it is relatively small, based on the tests that have been reviewed above. Can this effect be thought of as auditory accommodation? The answer is a cautious yes, as it matches well the main requirement set forth above, as well as other features of ocular accommodation. First, it matches the main prediction of auditory accommodation set at the beginning of this subsection—it reacts to the amount of reverberation in the signal, which is related to the degree of coherence (§ 8.4.2). Second, it takes some time to be engaged—about three times longer than ocular accommodation—which suggests a dynamic mechanism as well. Third, its effect is stronger for faster speech stimuli, which suggests that higher modulation frequencies are implicated more strongly. This can be related to the cutoff of the TMTF—the main proxy variable that determines sharpness. Fourth, compensation may take place somewhere between the VCN and the IC (or in parallel in the dorsal cochlear nucleus, DCN), which matches with some areas that are hypothetically involved in accommodation (see §16.4).

All that said, if the reverberation compensation effect constitutes the hypothetical auditory accommodation, then it is relatively elusive and small, and may not be nearly as significant as in vision. As such, it may not make a very strong case for accommodation as a whole. However, other realistic signals but speech that are not purely coherent or incoherent have not been investigated frequently in the literature, and some real-world dynamic listening phenomena may be either unmapped, or misclassified as something else (e.g., a complex effect of attention or release from masking). Hypothetically, such effects may be strong enough to merit the evolutionary investment in their improvement. Therefore, the subsequent discussion does not necessarily apply to the reverberation compensation effect as the only possible candidate for accommodation effects, or even as the "it" effect, and may apply to other phenomena as well, which remain somewhat vague at present. Combining the insight obtained from the reverberation adaptation effect and the auditory depth of field, we shall be aware of the possibility that various masking release or enhancement effects can be reframed as effects of accommodation. As such, they could provide much more impressive advantage in detection than for reverberation compensation alone, which may be easier to justify in terms of the evolutionary investment that has led to its existence.

16.4   The hypothetical accommodating component(s)

The case for having an accommodating auditory imaging system will be laid out for all of the components that have been touched upon in this work, which are not necessarily independent. The most immediately attractive known system to explore is the olivocochlear efferent reflex, which shares many of the superficial properties of ocular accommodation. It will be analyzed in the context of the hypothetical time lens and the phase-locked loop that we associated with the organ of Corti. Although the analysis is speculative at this stage, it is valuable beyond the present analysis in further consolidating our understanding of the different system parts, how they interact, and why they matter. As such, this discussion will inspire, at least to some degree, the next chapter about hearing impairments. We shall primarily focus on upstream auditory areas that receive the stimulus before auditory retina, which is assumed to be the IC, and seem to be capable of adaptation.

Before delving into the individual component accommodation hypotheses, it is going to be instructive to put forth a strong overarching hypothesis of what the accommodation system does: **Auditory accommodation calibrates the signal detection to produce an image that is optimally sharp and as noise-free as possible. As most natural signals are partially coherent, the system is poised to produce a certain mix between noncoherent and coherent detection. The relative weighting of coherent and noncoherent products may be accommodated by manipulation of phase-locking, noise (decoherence), dispersion, and gain at different points in the auditory pathways. The system determines its optimal degree of coherence from previous stimuli and updates its operating points continuously, also using input from attention.**

In interpreting results from literature, several key points should be made. First, the auditory system may not be geared to work with completely coherent or completely incoherent stimuli, so the influence of both strategies is always present to some extent in normal hearing. Separating the contributions of the two may not be trivial, since they often produce very similar images. Second, much of the data in literature exclusively utilizes broadband (incoherent) and tonal (coherent) stimuli. A few studies also utilize narrowband (partially coherent), including speech, or other informative stimuli[166]. Therefore, some relevant evidence should be interpreted with care. Third, a related concern is that these non-informative stimuli are given meaning that is imposed by the experimenter, but may not correspond to what the system considers meaningful. For example, in masking experiments, it is common to treat the masker as noise and the probe as signal or target that the listener should detect. However, it has to be considered that the masker is an auditory object in its own right and may be taken to be more interesting for the system to detect. Strictly speaking, both types of stimuli contain almost no information, but might nevertheless resemble signals that have some significance to the listener in their everyday environment.

As was noted in the introduction, we do not deal directly with gain accommodation, although inasmuch as it exists, it should work in concert with all other accommodating components. Cochlear gain accommodation was implied recently by Carney (2018), who argued that efferent control of the hair cells may optimize the effective level they operate in, so that envelope level variations across channels may be adequately coded in the auditory nerve. This level is thought to match typical speech levels of 55–65 dB SPL, which has to be coded despite limited dynamic range in several elements of the system. This should impact the "fluctuation profiles" in the IC—not unlike our polychromatic image—elements of which can then receive attentional focus.

The section is organized according to the presumed order of signal processing of the auditory

---

[166]Based on the analysis in §A, we consider speech to generally be partially coherent, but also highly nonstationary. This means that different regions in time and frequency may be instantaneously incoherent or coherent, but they do not extend for very long.

stimulus that is relevant to dispersion and coherence. As such, we shall attempt any direct analysis of the role of the middle ear reflexes.

16.4.1   Cochlear group-delay dispersion

The cochlear group-delay dispersion (Figure 11.6) is just large enough to dominate over small variations in normal atmospheric conditions of up to about 1 km (Figure 3.3) and fluctuations in the group-delay dispersion of the ear canal (including its multiple sign changes, Figure 11.2). If it were any smaller, it could have resulted in unstable imaging, especially in variable environmental conditions, in all but a small range of the audio spectrum. On the other hand, cochlear group-delay dispersion that is larger than estimated can only have a relatively small effect on the total defocus, unless it is combined with additional neural group-delay dispersive changes. Either way, a large part of the cochlear group-delay dispersion may be a direct result of the spectral analysis property of the cochlea, which produces dispersion by having the auditory channels distributed along its length.

Therefore, hypothetical accommodation of the cochlear dispersion entails a rapid change to the global group-velocity dispersion in the cochlea, or a change in its spatial phase dependence—effectively a cochlear map change. The former may be accomplished by actively changing the velocity of the traveling wave, the viscosity of the cochlear perilymph and endolymph, or maybe the velocity of sound in the perilymph or endolymph (Donaldson and Ruth, 1996). However, no such dynamic change processes are documented in literature at present, to the best knowledge of the author. Additionally, it is not at all clear that the necessary anatomical mechanisms to efficiently realize such changes are feasible.

In conclusion, cochlear group-delay dispersion is an unlikely parameter for accommodation. A long-term (non-accommodating) change in cochlear mapping will be discussed in § 17.3.2 in the context of hearing impairments.

16.4.2   Time-lens curvature and the phase-locked loop

It has been commonly theorized that the primary role of the outer hair cells (OHCs) is to provide nonlinear amplification (see §1.2 and §2.2.3), but this function may not be universal among different animal clades with similar organs (Peng and Ricci, 2011). For example, it was recently found that short hair cells[167] in chickens do not produce any measurable amplification or tuning on the traveling wave (Xia et al., 2016). The temporal imaging theory hypothesized two new functions of the organ of Corti and OHCs—phase-locking through the phase locked loop (PLL) and phase modulation through a time lens. At present, the two new roles are speculative and it is not clear whether they interact with amplification, with each other, or with other functions associated with the OHCs.

The medial olivocochlear reflex

Attractive for hypothesizing any kind of OHC-related accommodation, there is a neural reflex mechanism in place to facilitate it—the medial olivocochlear (MOC) reflex, or MOCR. The MOC system shares some notable similarities with ocular accommodation. First, both systems are primarily cholinergic, where only in the eye it is part of the parasympathetic nervous system, while in the auditory system there is limited parasympathetic innervation (Linker et al., 2018). Both are bilateral reflexes, whose most dominant effect can be seen within hundreds of milliseconds (Backus and Guinan Jr, 2006; Charman, 2008; but see Salloom et al., 2023 for more recent evidence of shorter time constants

---

[167]There are two types of hair cells in the avian cochlea—**tall hair cells** (THCs) that are innervated with afferents and hence are functionally homologous to the mammalian inner hair cells, and **short hair cells** (SHCs) that have motile properties in common with OHCs.

in the MOCR). Both auditory and ocular reflexes may also be mediated by higher-level processing such as attention to particular targets. The MOC system is also found in a hierarchically similar place to ocular accommodation in the eye, since it is attached to the lens in both systems. However, in vision, the lens is activated from a midbrain nucleus whereas it is activated from the brainstem in hearing. Incidentally, it has been shown that visual working memory tasks modulate the MOC activity as well (Marcenaro et al., 2021; Vicencio-Jimenez et al., 2021).

Notable differences between the reflexes are that normal ocular accommodation is almost completely bilaterally symmetrical, whereas the MOC can have a marked asymmetry, which depends on the degree of symmetry of the stimulus that activates the monoaural ipsi- or contralateral reflex, or its binaural version (Guinan Jr, 2018). Additionally, the auditory MOCR has two or three associated time constants with its operation, where the hundreds of milliseconds one would be considered medium (Backus and Guinan Jr, 2006). Finally, ocular accommodation is tightly coupled to vergence and pupil constriction, which have no clear parallels in hearing, although the independent middle-ear reflex can be activated in similar conditions to the MOCR in some listeners (Mertes, 2020).

The role of the MOC system is not well understood. The different functions that have been attributed to it are controversial, especially since many aspects of normal hearing appear to be possible with severed MOC efferents (Scharf et al., 1997). The MOC efferents inhibit the OHC somatic electromotility and, therefore, generally reduce and linearize the amplification in the cochlea, which may entail dynamic range optimization in some conditions (Künzel and Wagner, 2017; Guinan Jr, 2018; Lopez-Poveda, 2018; Jennings, 2021). In humans, its effect is most prominent between 500 and 2000 Hz and medium sound pressure levels. Results from studies on humans tend to be inconsistent between methods and suffer from high noise where otoacoustic emission (OAE) techniques are employed as a proxy to its activity (Guinan Jr, 2018; Jennings, 2021). This sets several functions that are hypothesized for the reflex on a shaky ground, such as improvement of speech-in-noise perception, localization, or release from masking effects that improve tone thresholds in low-level noise (Lopez-Poveda, 2018). The interaction between the MOCR and masking has been particularly thoroughly investigated, although a conclusive understanding of this function still cannot be produced—especially when different experimental methods are contrasted (Jennings, 2021).

Aside from the broad range of results in literature, it is challenging to interpret the MOC data because any interpretation depends on secondary models that are themselves not necessarily free of controversy (e.g., various OAE measures, assumptions about the involved time constants of MOCR activation). Another serious difficulty in assigning a higher-level role for the reflex within the system as a whole is that the stimuli that are most frequently used in these tests do not carry any information. This category includes narrowband and broadband noise that elicits the reflex, as well as distortion products from pure tones that are used as signals. Thus, if the MOCR system has evolved to realize a certain process that is commonly encountered in naturally occurring circumstances, then pure tone(s), broadband noise, and dichotic stimuli are likely to be very poor representations of such circumstances.

Time-lens curvature

The auditory time lens directly affects the amount of defocus in the system, since it can counteract the chirping effect of the cochlear and neural dispersion (Eq. 12.15). We obtained a broad range of curvatures, which nevertheless excludes the curvature range that is necessary to achieve a sharp focus. We have mostly relied on the large-curvature time lens estimates that produced the smallest defocus from the obtained range, but found relatively little impact on curvature variations within the range in all the phenomena we examined, except for the stretched octave effect (§ 15.10.1). However, the small-curvature estimates we obtained based on two studies suggested a near zero effect of the time lens. Additionally, one of the two studies used to obtain the large-curvature

estimates was directly based on change in excitation of the MOCR in the gerbil, which seemed to have a dramatic effect on curvature (Guinan Jr and Cooper, 2008). Therefore, there seems to be evidence for variable time lens curvature.

We proposed that acoustic phase modulation can be theoretically achieved through modulation of the stiffness of the medium (§11.6.1), although we did not directly rely on this principle in estimating the time-lens curvature. Stiffness is also theoretically related to amplification by the OHCs. In vitro, the magnitude of the OHC stiffness decreases as a reaction to acetylcholine, which would imply facilitation of gain, rather than gain inhibition that the MOC is known to produce (Dallos et al., 1997). However, as Dallos et al. noted, in-vivo effects are generally more complex, which may account for these results that appear to contradict amplification. Cooper and Guinan Jr (2003) suggested that the stiffness changes are slow (10–100 s) relative to the rapid changes associated with amplificative negative damping. As for phase modulation, any change in stiffness mediated by the efferents—even a small one—is expected to affect the time-lens curvature, although the slow changes are nothing like ocular accommodation that has a dynamic effect in vision over shorter time scales.

It was mentioned in §11.6.2 that when the MOC efferent to the OHC is not stimulated, the traveling wave of the basilar membrane (BM) exhibits a slow phase modulation over the first few cycles of a click response (Guinan Jr and Cooper, 2008). The effect could be switched off or diminished by stimulating the efferents. It suggests that the OHCs automatically apply phase modulation as part of their nonlinear response, as was hypothesized in §11.6.1.

If the curvature estimates in Figure 11.13 are correct, then switching off the phase modulation can substantially decrease the time-lens term in the imaging equation (12.15), which is expected to bring it much closer to sharp focus, depending on the baseline phase modulation in the system (see §12.3). Eliminating or reducing the defocus by changing the lens curvature entails a smaller separation between the coherent and incoherent imaging responses. In terms of the modulation transfer function (MTF), the theoretical incoherent focused MTF has a broader bandwidth than the coherent MTF. In a focused system, sources of different degrees of coherence are no longer distinguishable through their modulation bandwidths. The advantage in assuming such processing for sound may be twofold. First, if the source of interest tends to be incoherent, then the object contour can be better defined by letting the high-frequency modulation content go through, which may improve the demodulation (but it is subject to constraints of the sampling rate). Second, this may be useful if the system does not attempt to emphasize any coherent sounds over incoherent one, but rather make them sound qualitatively similar. Another way of putting it is that decrease of focus increases the auditory depth of field, which may eliminate perceptual cues that can be used to differentiate between objects of different coherence types (remember that, somewhat unintuitively, a shallow depth of field provides an effective way to distinguish between objects; §15.11.3). Depending on the acoustic conditions, these functions may prove more or less useful in realistic listening situations.

The phase-locked loop

Motivation Similar considerations can be applied to the auditory PLL function as were applied in the time-lens analysis, except that its accommodation is considerably easier to justify. Accommodating the PLL can hypothetically enhance or degrade the phase locking performance that is achieved mechanically and is transformed to neural synchronization after transduction. Evidence of a relation between the MOC and phase locking is rather limited, but nevertheless consistent with the PLL theory of operation.

First, let us consider what might be achieved by accommodating the PLL. The main parameter that applies to all PLL orders is the loop gain, which determines its hold-in bandwidth (§9.4). It is the product of the different gains in the loop—the phase detector, the low-pass filter, and the

oscillator. In the auditory PLL, we specifically designated the somatic motility with the role of supplying additional power to the loop. The loop gain affects how efficiently the PLL locks onto a signal in noise, how much noise is rejected in the process, how quickly it takes to acquire the lock, how stable the PLL is, how broad the (pseudo) narrowband filtering appears, and how well the lock can be maintained with random instantaneous frequency modulation.

The motivation of modifying the PLL loop gain should depend on the kind of input signal, its signal-to-noise ratio, and the ideal strategy that the system can use to demodulate it. If the signal is coherent so it has a nonrandom phase function, then a coherent detection strategy that incorporates the PLL may be warranted and the loop gain may have to be set accordingly. But if the signal is incoherent, then phase locking to it, if at all possible, may provide little advantage. Worse, coherent detection of incoherent signals may require more computational resources and may result in excessive phase noise at the output. In this case, noncoherent detection may be advantageous and the PLL may be either bypassed or its contribution to the received signal reduced.

It should be noted that in higher-order PLLs, additional parameters may be tunable, which determine the filter properties. However, given how little is known at present about this system, we will not speculate about other parameters beside the loop gain.

Empirical evidence   The effect of the MOCR on phase locking on the level of the auditory nerve was physiologically estimated only a handful of times, using direct stimulation of efferent nerves that innervate the OHCs. In the cat, when the MOC was directly stimulated, the saturation point of the synchronization index to tones near and below the characteristic frequency was found to increase in level by 2–14 dB for on-frequencies (Gifford and Guinan Jr, 1983, Figure 4), and by 0–16 dB for off-frequencies (Stankovic and Guinan Jr, 2000, Figure 5). For the on-frequency tones, levels as high as 40 dB SPL achieve maximum synchronization, but these levels increase for off-frequencies, as they do not excite the fiber as much as on-frequencies. Using the same stimuli, the saturation levels were generally higher if measured using the spiking rates rather than synchronization. These increases were correlated with lowering of the operation point of the spiking rate in the auditory nerve and shift to considerably higher levels, which suggest a reduction of amplification and loss of compression (or linearization) and a release from neural adaptation. The only other relevant measurement found, somewhat trivially, that a contralateral stimulation by the same tone as the ipsilateral ear does not affect synchronization (Warren III and Liberman, 1989). As with the vast majority of auditory phase locking measurements, none of the above presented the time course of the lock acquisition or the tracking capabilities to nonstationary signals, which would have been essential to evaluate some of the PLL most important features (see §9.9).

A different effect of efferent stimulation on the OHC was found in in-vitro samples of the bullfrog's sacculus[168], where the hair bundle spontaneous frequency (in the range of 10–80 Hz) changed from its baseline, and its phase locking to external tones significantly deteriorated (Lin and Bozovic, 2020; Bozovic, 2021). According to our PLL model, the hair bundle phase locking is the main precursor for neural phase locking. Thus, if the results from the frog's hair bundle translate to the mammalian OHCs, we would expect to see a drop in phase locking in the auditory nerve. The data from Lin and Bozovic (2020), however, cannot be straightforwardly compared to the mammalian data, so generalizing these results for acoustic stimuli at different levels and higher frequencies in mammalian OHCs requires more research.

According to the PLL theory, the phase detector of the PLL is represented by the quadratic $f_2 - f_1$ distortion product that is emitted by the OHCs (§9.8.1). Therefore, the $f_2 - f_1$ level may

---

[168]The sacculus is part of the vestibular system that is found in all vertebrates. It contains a sensory epithelium with hair cells and supporting cells, as well as afferent and efferent innervation. The hair cells are sensitive to low-frequency vibrations and sound and are similar to those found in the auditory system (Fritzsch et al., 2013).

be an indication of the phase detector sensitivity K_m (§9.4) and its general function. The effect of contralateral MOCR is known to either suppress or enhance the quadratic product, depending on the stimulus properties, but it hardly affects the more dominant cubic distortion product $2f_2 - f_1$ (Brown, 1988; Kirk and Johnstone, 1993; Althen et al., 2012). Phase changes to $f_2 - f_1$ are also observed during suppression (Wittekindt et al., 2009). The effects are usually dependent on the frequencies of the primary tones and the contralateral stimulus. Large (± 11 dB) and spectrally non-specific effects of both suppression and enhancement of the DPOAE products can be triggered on a cortical level, which ultimately controls to the MOC through the corticofugal efferent network (Jäger and Kössl, 2016).

A possibly PLL-related effect that appears to be triggered by the MOC efferents is a slight broadening of the auditory filters. It was indirectly found in humans that by stimulating the contralateral ear with broadband noise, which caused the ipsilateral ear's bandwidth to decrease by a small amount, as the measured delay of evoked otoacoustic emissions (OAE) during reflex activation decreased by 5% at 500–2000 Hz, compared to baseline (Francis and Guinan Jr, 2010). Filter broadening in human was also demonstrated by measuring the filter shape using the notched-noise masking thresholds for tones, when the contralateral ear was stimulated with pink noise or narrowband noise (centered on the same frequency of 1000 or 2000 Hz; Wicher and Moore, 2014). It was found that the tip of the filter was the same as the quiet condition, but in the pink noise conditions the filter broadened by about 17% at the 2000 Hz and by a smaller amount at 1000 Hz, which was statistically insignificant. The results were cross-validated by testing the distortion-product otoacoustic emission (DPOAE) that changed by up to 2 dB only in the pink noise condition. These trends generally disclose small effects that are not always consistent with similar studies that employed somewhat different methods (Wicher and Moore, 2014).

Higher-level measures in humans usually attempted to estimate whether the MOCR has any effect on speech intelligibility. A recent study found that in a lexical task[169] the MOCR was strongly activated during a vocoded-speech presentation, much more than during speech-in-speech-shaped noise and speech-in-babble noise conditions, while during natural speech in quiet it was only moderately activated (Hernández-Pérez et al., 2021). The MOCR activity was recorded by monitoring the contralateral-ear click OAEs. Its effect was neurally modeled using the original speech tokens, which indicated that the received envelope function in the vocoded speech is closer to the natural speech than during the other conditions.

The significance of the MOCR for speech intelligibility in difficult listening conditions is roughly in agreement between two studies. The first one found that recognition of monosyllabic words in noise with unilateral vestibular neurotomized patients with sectioned efferents had better performance (up to 20%) in their healthy ear with intact reflex than in the operated ear, when presented with broadband noise to their the contralateral ear (Giraud et al., 1997). Similar results were reported by Zeng et al. (2000, Figure 7), although they were confounded by the hearing loss of the de-efferenated subjects. The speech reception threshold difference was 1–8 dB worse in the operated ears with large individual differences among four subjects.

Synthesis   There are at least two standard ways to synthesize the above findings and the role of the MOCR. One standard interpretation is that the auditory system manages the dynamic range and achieves release from neural adaptation that can happen after prolonged exposure to noise, which can lead to loss of fidelity (e.g., Clark et al., 2012; Künzel and Wagner, 2017). This model class does not explain why this reflex should depend on the contralateral stimulus, or why the uncompressed signal and noise should be perceived any more clearly than the compressed signal in noise, especially given that the system reacts to broadband noise that does not necessarily affect

---

[169]The lexical task paradigm entails the identification of whether a spoken token is a word or a non-word.

the signal passband. However, especially when combined with the function of the middle ear reflex, the model does point to that the auditory system may strive to process signals at a medium level, perhaps to maintain a convenient operating point (e.g., Carney, 2018). An alternative explanation is that the auditory system improves inputs at positive signal-to-noise ratio (SNR) by linearizing the input-output characteristics through gain reduction, which comes at the expense of negative SNR inputs—as may be gathered from masking experiments that are interpreted using the power-spectrum model (Jennings, 2021). Once again, the design logic from the system point of view is somewhat unclear, since the low SNR situations may become hopelessly inaudible, and the role of contralateral stimulus in activating this feature is still not well motivated.

An alternative, or perhaps a parallel, explanation to the MOCR advantage is that the auditory system sets its PLL (loop) gain according to the degree of coherence of the (preceding) stimuli. As phase locking may be less effective with the MOC inhibiting the OHCs ability to synchronize, it can imply that only the highest-level portions of the signals will be phase-locked, whereas the lowest portion will not, and appear more blurry. This coincides with the modeling in Hernández-Pérez et al. (2021), which found that envelope fidelity—information that can be extracted without phase locking using noncoherent detection—is enhanced when the MOCR is engaged. It also coincides with the dual-processing model by Shamma and Lorenzi (2013), who proposed somewhat different signal processing paths for the envelope detection and temporal fine structure (TFS) detection, which we have associated with noncoherent and coherent detections. In their TFS detection, the signal is saturated, so that no envelope cues remain in the spiking pattern—only temporal cues. Thus, the effect we saw that the MOCR stimulation caused in the saturation point of synchronization and spiking rate (Stankovic and Guinan Jr, 2000) may reflect exactly that—enhancement of the dynamic range and reduction of synchronization improves the noncoherent detection at the expense of coherent detection.

As speech signals are only partially coherent, either type of detection (coherent or noncoherent) can be used to detect it, as has been shown in the auditory literature in reference to envelope and TFS processing of speech (Lorenzi et al., 2006; Paliwal et al., 2011; see §6.4.3). Given that the quadratic DPOAE amplitude can be either enhanced or suppressed with the MOCR, such a control system may have some flexibility in setting the proportion of noncoherent to coherent detection, especially at medium levels. The large dynamic range of the DPOAE effects observed through corticofugal activation suggests that attentional mechanisms can indirectly control the detection strategy as well.

The PLL explanation can imply that the contralateral ear reflex has something to do with the signal interaural correlation, or rather, with an internal estimate of the spatial coherence in the system (§8.5). Broadband noise that is completely incoherent is known to trigger the contralateral reflex, whereas partially coherent (narrowband) noise does so partially, and coherent signals (tones) often do not (but see Althen et al., 2012). The same goes also for the ipsilateral signals, as they are usually tones that are incoherent, partially coherent, or coherent with the contralateral stimulus, respectively[170]. When an uncorrelated stimulus is presented to the contralateral ear, the system may register it automatically as an important signal, rather than "noise" per se. Then, enhancing the most general signal processing—noncoherent detection—may constitute a more robust heuristics to deal with such stimuli.

All in all, given the scant evidence for phase-locking effects in the MOC system, as well as the unavailable parameters of the hypothetical auditory PLL, the accommodation of the loop gain in the PLL is speculative, at present. Nevertheless, there appears to be merit in such a process and there are several strong findings that support such an effect in humans and animals. What remains

---

[170]See also Lilaonitkul and Guinan (2009) for a systematic comparison between ipsi-, contra-, and bilateral elicitor bandwidth effects.

relatively unclear at this point is whether the low-level shift that was observed in the synchronization index in the cat is relevant to humans and would it have any measurable effect in realistic listening conditions. Similarly, the bullfrog hair bundle synchronization data relevance remain to be seen in mammals.

PLL and time-lens accommodation

Perhaps unsurprisingly, the hypothetical accommodations of the time-lens curvature and the PLL that were discussed above can serve the same purpose in sound processing. The time lens accommodation was tied to an apparent reduction in defocus, as the incoherent response gets closer to that of the coherent one and the depth of field increases. In such a design, it makes perfect sense that no phase locking should take place, as incoherent sounds may be detected noncoherently, only using their envelope. The reduction in the loop gain of the PLL also biases the system for more noncoherent detection than coherent detection. Putting the two together, it seems that the MOCR is designed to reduce the level of coherent detection and increase the noncoherent envelope detection. This explanation is supported by the modeling and data from Hernández-Pérez et al. (2021). However, it may not match some of the hypotheses about the MOCR function that are found in the literature—none of which is in consensus at present (Lopez-Poveda, 2018). Data regarding unmasking of speech in noise has been particularly inconsistent between studies and methods (Smith and Cone, 2021). Given that speech is partially coherent and may be almost as equally well-recognized with coherent as with noncoherent detection, the comment made by Lauer et al. (2021) about the MOC function may be apt: "*In some cases, these effects may not be apparent because compensatory or redundant processes are likely in play.*"

16.4.3   The temporal aperture duration and shape

The aperture time was expressed earlier (Eq. 12.40) as a function of all three dispersive elements (cochlear, time lens, neural), because of physical and mathematical constraints. However, the aperture may be set independently of these constraints, if it turns out to depend on a purely neural mechanism. A simple neural correlate, for example, is the action potential in the auditory nerve, which factors into wave I of the auditory brainstem response. The typical width of the largest peak of the action potential is just under 0.5 ms long (e.g., Yoshie, 1968; Picton et al., 1974), which is approximately the same as in the high-frequency channels computed of 4–8 kHz (table 12.2). This correspondence fails at low frequencies, though, where the temporal aperture may not be purely neural and the aperture stop may depend on the cochlear channel (§13.4.1).

No less important than the duration of the aperture is its corresponding shape, as is expressed by the pupil function—effectively, a temporal window. For practical reasons, it was approximated here as a Gaussian function, which has a single parameter—width. However, although it was found to be a very faithful representation of the physiologically measured Chinchilla's pupil function, the measurement showed an asymmetrical fat tail on one side that deviates from the Gaussian (§12.5.4). We know from modulation transfer function theory in optics that the pupil function essentially defines the image properties—its contrast, sharpness, magnification, and aberrations. Therefore, an ability to accommodate the pupil shape should directly affect the image quality in some situations. Wave I itself is known to vary in shape between people, but its neural source may be mixed with contributions from nearly overlapping sources (e.g., the hair cell membrane potential) (Kamerer et al., 2020). The feasibility and plausibility of biophysically achieving such an accommodation are unknown.

The advantage of being able to vary the aperture duration is significant, as it affects the modulation transfer function characteristics, in concert with the neural group dispersion. Accommodation of the aperture time alone is probably analogous to the pupil function in vision, rather than to lens

accommodation. Inasmuch as we may be looking for an analog to the role of the pupil in vision that limits the amount of energy on the retina, then the middle-ear and MOC reflexes may be more suitable candidates. But if the pupil is required not only in restricting the amount of energy that is transduced, then direct adjustment of the auditory aperture may have some merit. This can be achieved, for example, through activation of the lateral olivocochlear (LOC) efferents that synapse to the auditory nerve dendrites to the inner hair cells. As the LOC efferents are unmyelinated, such an accommodation may be too slow to be useful in dynamic situations, though.

All in all, this potential candidate element for accommodation seems rather unlikely, at least as a parameter independent from the other dispersion parameters, or as something that can change quickly enough to dynamically track the signal.

16.4.4   Neural group-delay dispersion

Neural group-delay dispersion characterizes a piecewise conduit of information transmission before forming an image on the auditory retina—probably the IC. While this part of the system is hardwired, there is ample evidence that the brainstem has a range of dynamic capabilities that enable selective and rapid adaptation to different signals and conditions. A detailed review and analysis of the processes that mediate neural plasticity is out of the scope of this work (but see for example, Tzounopoulos and Kraus, 2009; Irvine, 2018). Instead, three very general properties of the auditory brain are mentioned below, which may function as short-term mechanisms that can constitute the plausible auditory neural machinery for group dispersion accommodation.

The first property is wide innervation of the descending auditory network in the brainstem, which allows for top-down control by higher auditory centers, through the formation of feedback loops. Thus, there are widespread descending projections from the IC to the cochlear nucleus (CN) and superior olivary complex (SOC), and from the thalamus to the IC, CN and SOC, among many others (Schofield, 2010; see Figure 2.4 for notable efferent projections). Some of the connections to the CN project directly from the acousticomotor area of the external nucleus of the IC (ICX), which is itself connected to the superior colliculus (SC), where it is coordinated with visual and tactile inputs (Huffman and Henson Jr, 1990). The SC itself may be indirectly controlling ocular accommodation as well, via the Edinger-Westphal nucleus (May et al., 2016). The exact function of these networks is not well understood, although the range of possible functions is constrained by the type of connections, e.g., excitatory or inhibitory (Milinkeviciute et al., 2017). Recent electrophysiological measurements demonstrated how the addition of context can quickly modulate the response at the subcortical auditory nuclei level. This was cleverly shown by having naive participants listen to stimuli made of three sine waves before and after revealing to them that these stimuli represent, in fact, sparse speech (Remez et al., 1981; Cheng et al., 2021). Hearing the stimuli as recognizable speech rather than tonal noise had a quick and dramatic effect on the frequency following response (FFR) amplitude, which led the authors to conclude that the subcortical facilitation of auditory processing could only be mediated by the descending efferent network from the cortex.

The second property of the auditory pathways is that neurons tend to have multiple receptor types, which allow for different neuromodulators that innervate the system to fine-tune the involved auditory functions and their associated signal processing in complex and nontrivial ways (Schofield and Hurley, 2018). The plasticity that arises as a result can take place at different time scales and may broadly reflect behavioral arousal, environmental stress, attentional inputs (e.g., regarding salience), past experience, and even social situations.

The third general property is a short-term synaptic plasticity that has been identified throughout the auditory pathways, which affects the amplitude, timing, and rate of neural discharges (Friauf et al., 2015). Synaptic plasticity can be observed in amplitude changes of postsynaptic responses

to presynaptic activation and hence, it modulates the synaptic transmission efficacy (i.e., signal processing speed and capacity).

As the dispersive properties of the auditory brainstem ultimately reflect its information transfer dynamics, it is reasonable to expect that the plasticity that is offered by processes as mentioned above can vary the neural dispersion in different pathways. When dispersion is changed in a frequency-dependent manner, the group-delay dispersion changes as well. Depending on the magnitude of this change, it has the potential to strongly affect temporal imaging. The effect of accommodating the neural dispersion can be substantial in setting the cutoff frequency of the low-pass response of the TMTF, which may be significant for tasks that require high-frequency modulation content. At this point, we are unable to tell how plausible it is that the group-delay dispersion is modulated at all, and if it is, then what its temporal dynamics is like and whether it can serve as auditory accommodation.

16.4.5   Filter bandwidth

The advantage of having variable-bandwidth filters is a degree of control of the received coherence in every channel: narrow channels produce more coherent outputs than wide channels (§ 8.2.8). Therefore, the sound of an incoherent object will appear more coherent in a narrower channel, which entails less blur, at the expense of spectral information loss from the full broadband sound, which should be processed temporally in a wide channel. This can be thought of in the extreme case of a high-Q resonant filter that oscillates with random noise as input—the filtered oscillation is quasi-tonal and partially-coherent, as it maintains the instantaneous phase of the broadband input, which varies slowly around the center frequency. Bandwidth accommodation is also likely to affect how polychromatic images are fused across channels. Also, in situations where off-frequency masking is dominant, the channel bandwidth may be able to decrease or increase the degree of masking—effectively, to adjust the contrast in the complex polychromatic images.

Evidence that the MOCR system affects the auditory filter bandwidth was brought up in §16.4.2 in the context of PLL loop gain accommodation. Another general mechanism that potentially affects the bandwidth involves the corticocortical and corticofugal descending auditory pathways, which have the capability to selectively modulate spectral, temporal, amplitudinal, and spatial responses of neurons, as was measured primarily in bats and mice (Suga, 2020). One of the common findings is of filter sharpening following electric stimulation, fear response, or a repeating tonal stimulation—mainly in the primary auditory cortex and the medial geniculate body (Suga, 2008, 2020), and in the IC (e.g., Yan et al., 2005). It was observed following the retuning of off-frequency neurons and the reacquisition of the best frequency that was most relevant to the task at hand. There is not much direct evidence for such bandwidth shifts more upstream, but excitatory descending projections found in the DCN of the mouse may suggest that similar changes in bandwidth may take place there (e.g., Milinkeviciute et al., 2017). Similarly, bandwidth tuning effects were found in chickens in response to inhibitory GABAergic inputs to the nucleus magnocellularis from the superior olivary nucleus (SON)—the avian nuclei that are analogous to the CN and superior olivary complex in mammals, respectively (Fukui et al., 2010). While the function of the SON is mainly associated with localization, other roles of this or other control pathways may induce similar tuning effects for processing other attributes.

Another mechanism to take into consideration is lateral inhibition, which sharpens the frequency response by suppressing off-frequency components in adjacent filters. It was originally found in insect vision (Hartline et al., 1956), and has been documented throughout the auditory pathways (Nomoto et al., 1964; Sachs and Kiang, 1968), but is probably better thought of as a universal sensory mechanism (Békésy, 1967). In the auditory nerve, it likely reflects the nonlinearity of the cochlea

(Ruggero, 1992). In the CN it may either preserve the cochlear response or sharpen it further (Rhode and Smith, 1986b; Rhode and Greenberg, 1994; Caspary et al., 1994; Kopp-Scheinpflug et al., 2002). This general mechanism is invoked to explain improved hearing in noise, as it is used to increase spectral contrasts (Kluender et al., 2003). However, the real-time effects and dynamic properties of lateral inhibition may be difficult to extrapolate from these low-level examples.

There are two caveats to bandwidth accommodation. First, it is not clear that the auditory filter bandwidth as is encompassed by steady-state auditory filter models, has a significant influence on temporal information processing that is relevant in imaging (see §15.5). The different neural mechanisms mentioned are bound to produce instantaneous bandwidths that are context- and situation-dependent. Either way, be it the auditory filter, or the neural sampler, or any other temporal constriction—the narrowest one in the processing chain functions as the aperture stop, which then has the limiting effect on the temporal modulation transfer function, which may depend on the filter bandwidth only at low frequencies §12.5. Second, even if broader channels can let in faster modulation frequencies, they still have to be sampled appropriately in order to make use of the extra temporal information that they can hold. This is discussed below in §16.4.6.

In summary, there are several general physiological mechanisms in place for sharpening or broadening the effective bandwidth of auditory channels. In the present state of knowledge, none of them stands out as a distinct mechanism that resembles accommodation, but this possibility will have to be considered in the future. Bandwidth manipulation midway in the auditory brainstem may be better thought of as analogous to spatial filtering in optics, which works to process the modulation band rather than the carrier band information. Therefore, more complex passband morphologies that are occasionally observed in the auditory brain may have to be reinterpreted accordingly. In contrast, vision performs spatial filtering only with the pupil, which is roughly a circular aperture that may have equivalent one-dimensional analogs in the auditory brainstem in the form of bandpass filters.

16.4.6   Sampling rate

It was argued in §14 that spikes in the auditory nerve correspond to discrete samples, and that each one may embody a narrowband image in its own right (§14.4.2). If this argument is accepted, then the entire ascending auditory pathway and the stochastic nature of neural firing are expected to encompass several points of imperfect resampling between synapses in the brainstem. Furthermore, it can contribute to the effective downsampling that is observable in the IC and further downstream— the gradual decrease of maximum spiking rates. It means that the highest modulation-frequency information, which is captured by the fastest sampling sequences, is lost in transmission before the IC, unless it is recoded (or is used, if it reaches its destination) earlier. However, the fact that the spiking is nonuniformly distributed suggests that an instantaneous cutoff frequency is a more correct way to approach sampling, instead of an exact Nyquist rate as in uniform sampling (§14.7). It brings about the possibility of generating more spikes in order to better sample the incoming stimulus, and thereby provide less opportunities for information loss in the resampling and downsampling process, on the ascending pathways to the IC.

There are several findings that support the idea that the spiking rate can be dynamically set by the system. Neural adaptation is one obvious example where the spike rate is known to be variable, as it decays after the signal onset (Kiang et al., 1965). As was mentioned in §16.4.2, the MOCR is known to reduce the operation point of the OHC process, which causes a corresponding drop in the saturation rate of the auditory nerve (also referred to as adaptation), which is interpreted as dynamic range management of the system (Lopez-Poveda, 2018; Guinan Jr, 2018). However, it also entails that higher-level signals can be coded with the same spiking rate as lower-level ones, which is

equivalent to a reduction of sampling rate. This change will have no discernible effect on tones and does not appear to not affect their loudness (Morand-Villeneuve et al., 2002), but more dynamic signals can be instantaneously undersampled as a result of MOC inhibition.

In more dynamic signals such as speech, the onsets may not be as well defined as with synthetic stimuli, but changes that can appear as onsets are important cues in phonetic segmentation in general (Delgutte, 1999). For example, changes between formants are marked with frequency glides that excite successive fibers in the auditory nerve. Each fiber reacts with an unadapted onset response at a higher spiking rate, enabling a more precise representation of the signal, which cannot be attributed only to intensity differences (that are more readily associated with rate changes) (Delgutte and Kiang, 1984). At the level of the CN, onset responses characterize most of the cells and fewer cells exhibit sustained responses. Using data from behavioral tests with speech-like stimuli, it was suggested that such adaptation effects enhance contrasts in complex sounds and may have a role in segmenting coarticulated speech (Kluender et al., 2003). This is supported by results from the aliasing detection study presented in §E (Experiments 4 and 5), in which adaptation was suggested to cause an almost one order of magnitude increase, on average, in threshold of temporal discrimination between successive pulses.

At higher levels of processing, sampling may be less directly relevant than in the auditory nerve and brainstem, as information downstream is coded more efficiently and sparsely, and at lower rates. Nevertheless, cortical firing rates have been shown to be modulated by focused attention, in hearing (Miller et al., 1972), vision (e.g., Moran and Desimone, 1985; Luck et al., 1997; Spitzer et al., 1988), and somatosensory processing (Hsiao et al., 1993).

In summary, there is no strong evidence that the sampling rate can be modulated en masse in the brainstem, although it is locally quite dynamic and may be affected by several subprocesses in different contexts. An open question is whether the increase in cortical firing rate "trickles down" to the brainstem, and if so, whether it requires focused or selective attention or whether other mechanisms can cause it. At the present level of knowledge, it does not appear as a very likely standalone mechanism to embody auditory accommodation.

16.4.7   Synchronization accommodation

Phase-lock coupling strength and noise

Synchronization accommodation would hypothetically involve controlling the precision of phase locking—the auditory response to the fine-structure of the stimulus. There are two general mechanisms that can control or modulate phase locking. They are the modification of the degree of coupling between the local oscillator and the signal path and the variation of the amount of sampling noise. The first process was considered in §16.4.2, where it was argued that the MOCR affects the level of phase locking that is provided by the OHCs. Activating the MOCR decreases the level of phase locking in the auditory nerve, which is equivalent to decoupling the PLL from the main signal path. We do not know if an opposite effect is possible with the MOCR (i.e., enhancing the coupling strength between the external signal and the local oscillator). Unless there are additional neural PLLs downstream (§9.10), then decoupling/coupling accommodation may limited to the cochlea.

Jitter

The second process that may accommodate synchronization is based on addition or removal of noise, which modulates the degree of coherence of the transduced signal. One way to add noise is to relax the sampling precision, as the spikes are synchronized to both carrier and envelope frequencies with finite precision. The small instantaneous deviations between the spikes and the incoming wave

affect the level of transmitted coherence of the input, as they correspond to proportional phase error—phase noise or **jitter** in the sampling. If the precision is perfect (zero jitter), then within- and across-channel cochlear-level coherence is conserved in the transduction process. Otherwise, it is eroded. While zero jitter is physically impossible, being able to control the finite level of jitter would theoretically enable the modulation of the degree of coherence of incoming signals and, which selectively subjects them to the defocus and to decoherence. So, the more incoherent signals are, the more defocus applies to them and the less effective coherent detection gets in demodulating them. Upstream, jitter may apply only to carrier phase locking and have little to no impact on slower timing patterns that belong to the envelope and to its processing. Thus, if the temporal precision of the envelope is also conserved downstream, then it may be used in the formation of auditory streams that can be attended to at a higher processing level (Singer, 1999; Niebur et al., 2002; Elhilali et al., 2009; Shamma et al., 2011).

Jitter exists during the transduction at the ribbon synapse between the IHC and the auditory nerve (Rutherford et al., 2021). This synapse reacts relatively slowly to high-frequency inputs, where jitter can limit the coding fidelity. It is considered a property of fiber type, i.e. low-, medium-, or high-spontaneous rate fiber. Whether this property can be accommodated to—say, using the slow LOC efferents—is unknown at present.

In the neural domain, there are several discontinuities between the auditory nerve and the CN in terms of synchronization, which varies between cell types and between the DCN and the VCN. The DCN appears to have poor phase locking capability, yet a spectrally fine-tuned response. In comparison, the VCN provides an excellent phase locking response, which is primarily projected to the SOC (associated mainly with localization processing) (Rhode and Smith, 1986b), but not exclusively, as it also projects ipsilaterally via the trapezoid body and the ventral nucleus of the lateral lemniscus (VNLL) to the IC (Oertel and Wickesberg, 2002). Particularly in the anteroventral cochlear nucleus (AVCN), phase locking is thought to be the result of multiple inputs that feed into "high-sync" bushy cells, which yield an improvement over the initial synchronization observed in the auditory nerve (Joris et al., 1994). The bushy cells are exceptionally well-designed for precise temporal coding on all cell levels (synapse, membrane, action potential, high frequency operation, quick repolarization, sustain operation; Kuenzel, 2019). Furthermore, these are multipolar cells with multiple inputs that have extensive neuromodulatory capabilities that can likely fine-tune these well-calibrated phase-locking features (see §16.4.4). While the exact function of such a neuromodulation may not be well-understood at present, it is possible that it has a role in accommodating the degree of synchronization in some or all of the subnuclei of the CN. The overall design here appears to be promoting decoherence in the DCN, which may be suitable for noncoherent detection, and coherence in the VCN, which may be more suitable for coherent detection. Their combined product at the IC is partially coherent to a degree that is determined by their individual contributions.

The analogous operation of jittering in continuous media is diffusion. Therefore, in optics and acoustics, it is sometimes achieved with diffusers.

Dither

The alternative to jitter is to add noise directly to the signal, which is sometimes referred to as dither. It has been discussed in the context of mechanical motion of the hair bundles, as it was found in hair bundles of the frog's sacculus in vitro that a small amount of noise that can be generated from Brownian motion, for example, can actually improve the SNR by up to 3 dB at low-SNR conditions, owing to a phenomenon called **stochastic resonance** (Jaramillo and Wiesenfeld, 1998; Indresano et al., 2003; Benzi et al., 1981). However, while it can be assumed with some confidence that these findings apply to audio frequencies in the cochlea as well, the effect was demonstrated using white noise and pure tones, so it may be difficult to generalize. Also, we do not know whether the source

of noise can be controlled internally within the cochlea, or if it strictly depends on external supply. The effect appears to contribute to coherent detection, but it may be inverted to do the opposite at different input SNRs. For example, since phase locking is known to diminish at high frequencies even though the OHC architecture seems to be the same along the entire spectrum, the possibility that the OHCs are designed to add dither at high frequencies through random movements—rather than phase lock—may be worth exploring.

Dither may be more commonly explored within the neural domain, where the amount of spontaneous spiking directly corresponds to noise. Spontaneous spikes were shown to have a desirable dithering effect on envelope detection in auditory nerve axons (Yamada and Lewis, 1999). Spontaneous activity in the central auditory system was shown to depend on the stimulus and may or may not relate to intended manipulation of the noise floor. For example, spontaneous activity in the DCN, VCN, and IC widely differs between different cell types and some are inhibited with the increase of contralateral input level or after onset (e.g., Syka and Popelář, 1984; Rhode and Smith, 1986a,b; Joris et al., 1994). Promisingly, early deep-neural-network simulations of the simplified DCN processing suggest that the addition of Gaussian noise at that stage can improve speech recognition accuracy after cochlear hearing loss (Schilling et al., 2022). Nonetheless, we do not know at present whether any of those internal noise floor variations can be modulated or have a desirable effect on detection—whether through decoherence or other signal processing tricks.

Relation to tinnitus?

Another perspective on synchronization may be gathered from recent findings and models of **tinnitus**—the perception of phantom sounds that do not correspond to external acoustic sources. In tinnitus, cortical synchrony is a robust correlate to the perceived phantom sound (Eggermont, 1984, 2012). An accommodation-relevant hypothesis is that (some forms of) tinnitus is an extreme manifestation of a naturally occurring process in the auditory system, whose normal function is to accommodate to the target stimulus. When such a system receives a complex signal that is composed of a mix of partially coherent stimuli, accommodation may be geared to selectively enhance synchronization in order to cohere and sharpen certain bands, or to decohere and to blur others. When a real stimulus is being processed, the synchronization accommodation is subtle and automatic (unconscious). In contrast, when nothing but noise is being detected (i.e., only weak cochlear activity, or spontaneous activity in the auditory nerve), then the system coheres to random patterns. If in addition there is sufficient central gain that is applied to the noisy channel, it may be perceived as audible noise with pitch strength that is inversely proportional to the bandwidth involved (i.e., narrow bandwidth—high pitch strength). In the normal system, this adaptation would seldom "glitch", but when it does it may be experienced as a spontaneous **transient tinnitus** (Flottorp, 1953)—an obscurity even within the tinnitus literature that is excluded from most surveys and is sometimes attributed to spurious activity of the OHCs (e.g., Eggermont, 2012, pp. 3 and 15). In hearing impaired listeners with mild tinnitus, an advantage is occasionally reported in speech-in-noise measures, compared to listeners with hearing loss without tinnitus (Husain and Khan, 2023). If accommodative synchronization indeed exists, then the synchronization process must include a control signal that modulates the relevant parameters of accommodation, as controlled by attention, for example. This hypothesis combines at least two of the prominent theories of tinnitus—elevated synchronization and central gain (see Henry et al., 2014a for a concise review)—but has a higher explanatory power, because it is less arbitrary as far as the complete system function is considered.

For this dual-hypothesis—that tinnitus is an abnormal form of synchronization accommodation—to be correct, two conditions have to be met. First, the hypothetical synchronization accommodation has to be applied before the imaging stage at the IC, i.e., while in the brainstem, or even in the cochlea. A related clue may be found in one prominent tinnitus theory, which traces its genera-

tion, but not its maintenance, to the DCN (e.g., Henton and Tzounopoulos, 2021). Fusiform cells in the DCN of guinea pigs were found to be hyperactive (with elevated spontaneous firing rate) with increased spontaneous synchrony both in noise- and in drug-induced tinnitus (Martel et al., 2019). However, this kind of plasticity is very slow in comparison with any useful reaction time for accommodation. Hence, the second condition is that the ability to induce a change in the amount of jitter or dither has to take place over short time durations. Tinnitus research has looked into long-term plasticity measured over days or longer, so there is no direct indication for this, to the best knowledge of the author. However, tinnitus studies generally exclude the normal, non-pathological, operation of the DCN, which may include some of the neuromodulatory mechanisms (of the kind mentioned in §16.4.4 and §16.4.5), which may be able to achieve synchronization accommodation over shorter time constants (Kuenzel, 2019).

It should be noted that it may be impossible to treat synchronization completely independently from sampling rate. In the VCN, it has been emphasized that the relatively low firing rates in comparison with the auditory nerve may pay off as a processing strategy that achieves an increased temporal coding precision (Keine et al., 2017; Dehmel et al., 2010; Kuenzel et al., 2011; for similar findings in trapezoid body, see Wei et al., 2017). The same cannot be said about the auditory nerve, where synchronization and spiking rate seem to be largely independent factors (Johnson, 1980),

Another clue that tinnitus may be related to an early-stage accommodation of synchronization is from recent auditory brainstem response (ABR) findings in tinnitus patients and normal-hearing controls. Synchronization can only be considered for signals that are coherent or partially coherent, but not for incoherent signals. Therefore, we would hypothesize to observe differential processing of different types of signals in the brainstem, depending on their degree of coherence. It was found by Tan et al. (2023) that in tinnitus patients the interpeak interval (latency) between waves I and V were shorter only with chirp stimuli (coherent), but not with (incoherent) clicks. The reduced interpeak intervals with the chirps at 45 dB (normal hearing level) were also correlated with lower speech intelligibility scores in those patients at 85 dB SPL. Tan et al. (2023) proposed that these results might be related to cochlear synaptopathy, but there was limited supportive evidence for this in the subject group of the study.

All considered, accommodation through synchronization may be an attractive feature, but does not have strong evidence to support it at present. The exact synchronization processes in normal hearing—especially fast ones that can respond dynamically (at the scale of hundreds of milliseconds)—are completely uncharted. Additionally, it is unlikely that such accommodation would work completely independently from the accommodation of the firing rate, gain, and channel bandwidth.

A lengthier discussion about tinnitus and accommodation will be deferred to the next chapter (§17.7.3), along with some interesting analogies of related disorders, which will make the case for synchronization accommodation more compelling.

16.4.8   Coherent and incoherent stream mixing

The imaging theory in this work has aggregated the neural pathways in the brainstem to a single parameter—the neural group-delay dispersion. In reality, each pathway may have somewhat different dispersive characteristics, as well as different internal noise characteristics, which are suitable for various kinds of processing. As an example that was mentioned earlier, depending on the particular cell type, in two out of the main three divisions of the CN, the DCN tends to have better spectral and worse temporal resolution, whereas the VCN has it the other way round (Rhode and Smith, 1986b,a; Joris and Smith, 1998). These nuclei project both ipsilaterally and contralaterally to the IC. This means that the IC receives multiple versions of the same stimulus (e.g., Ehret, 1997, p.

264 and Malmierca and Hackett, 2010, p. 26)—each one is potentially characterized by different defocusing and a different degree of coherence. Noise that is applied in either one of the pathways would reduce the degree of coherence selectively only to that processing path. Hypothetically, the IC can selectively weight the contributions of both inputs in order to optimize the received partial coherence, and therefore bring objects in and out of focus in the final mix that is then propagated to the auditory cortex.

A close variation of this idea is that the two main pathways of the DCN and VCN specialize in noncoherent and coherent detection (§5.3.1, §9.11), respectively. The output of these two detectors may be optimal for some signals, but not for others. Weighting their contributions to the complete image may endow the system with a broad range of hearing strategies that can be optimal for arbitrary inputs. A similar design to this was hypothesized to exist in the avian auditory system (Sullivan and Konishi, 1984; Warchol and Dallos, 1990) (see §2.4). It was also explored in some depth in §9.11 in relation to the PLL and with some evidence reviewed to support it—mainly from frequency-following response (FFR) studies in humans.

Compared to a few of the hypothetical accommodation mechanisms discussed above, these two hypotheses are attractive, as they seem relatively straightforward to implement. This is the case, especially given that there are already parallel signal processing pathways in the brainstem that converge in the IC with unclear role division. Nevertheless, the specialization of the VCN and DCN (or their avian analogs) serve only as circumstantial evidence, and at present there is no concrete proof that either variation of the mixing hypothesis holds. Furthermore, considering the cellular and functional diversity within the VCN and DCN, this role division may be too crude a classification. However, given the known capabilities of the auditory system to produce responses based either on coherent or on noncoherent detection, we speculate that some version of the second variation can describe the main split in the mammalian brainstem, which will be used as a hypothetical building blocks in the complete auditory model that is proposed in §18.2. Even so, the hypothetical mixing feature may not count as accommodation proper, if only because its assumed function does not resemble ocular accommodation or any of the other two ocular reflexes. Instead, the mixing of the two pathways may be a much more fundamental feature of the auditory system writ large. In this respect, the precise role of the third branch of the cochlear nucleus (the posteroventral cochlear nucleus, PVCN), which is not always easy to discern in the human CN, remains to be defined more closely.

16.5 What informs auditory accommodation?

Regardless of the specific accommodated variable, the hypothetical accommodation system has to infer from the signal itself how to accommodate. Of the different accommodation mechanisms considered, the various functions of the MOCR are probably the only ones that have clear triggering. Both ipsi- and contralateral reflexes appear to depend most strongly on the bandwidth of the elicitor signal, and on its absolute level (Lilaonitkul and Guinan, 2009). As the different studies tend to use broadband or narrowband white or pink noise to trigger the reflex, one may wonder whether periodic broadband, or other types of signals may elicit a similar efferent response. Triggering by the contralateral ear may indicate that the system strives to have some symmetry in processing between the two ears, perhaps to maintain some continuity of localization processing. Alternatively, a common mechanism of ipsilateral coherence detection between bands, or contralateral spatial coherence (interaural cross-correlation) has to be considered too, which can potentially inform the system about the optimal strategy to detect the image of interest for the listener. Such mechanisms may be realized using broadband coincidence detectors, which are thought to exist in the PVCN (Oertel et al., 2000; Lu et al., 2018) that, in turn, projects to the SOC, where the MOC bundle starts.

A problem with this hypothesis is that it appears that the MOC input comes from chopper units that provide sustained output (Brown et al., 2003), whereas the broadband coincidence detection properties are notable mainly in the octopus cells and are less quick and precise in the chopper cells (Lu et al., 2018). Either way, less speculative explanations can be established only after the accommodation function is elucidated.

16.6   Listening effort and accommodation

The term **listening effort** was introduced by Downs (1982) in order to operationalize the added cognitive difficulty that aided hearing-impaired listeners experience, which may not be captured by their speech intelligibility scores alone. Listening effort has been recently defined in a consensus paper as *"the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a* [listening] *task"* (Pichora-Fuller et al., 2016). The usefulness of having a concept of effort lies in the ability of listeners to relate to it (e.g., Luts et al., 2010). Also, it differentiates certain tasks and listener groups in a way that may be more sensitive than other measures such as speech intelligibility, especially in ceiling-performance conditions. However, listening effort requires a model to base objective measures on and is difficult to pin down precisely, because it is correlated and potentially confounded with attention, fatigue, motivation, and other high-level cognitive variables, which are themselves not necessarily clearly defined. Listening fatigue, for example, is thought to combine emotional, cognitive, and peripheral components (Hornsby et al., 2016). Or, in some models, the effects of cochlear hearing loss are thought to be compensated by cognitive and executive functions (Peelle and Wingfield, 2016). These complex relations within the concept of listening effort make it a particularly different to measure and pin down reliably (Shields et al., 2023).

While the analogous **visual effort** is not well-defined either, it is less abstract to speak of visual strain or fatigue, or accommodation and vergence efforts, which directly relate to ocular muscle actions, at least in the peripheral processing stage (e.g., Toates, 1972). For example, accommodation effort was found in situations where the visual system attempts to maintain a sharp focus despite low-light conditions. In extreme low light, it retains a fixed focus and exhibits decreased resolution (**night myopia**; Johnson, 1976; Charman, 2010, pp. 1.34–1.35). Vergence effort has been specifically shown to be associated with visual fatigue, or eye strain, for people spending much time in front of a computer screen (Tyrrell and Leibowitz, 1990). Although accommodation and vergence efforts are not the cause of the so-called **computer vision syndrome** (characterized by headaches, eye strain, image blur, neck pain, and more), they are typical in lengthy viewing of visually demanding targets, such as computer or smartphone screens (Rosenfield, 2011). Accommodation effort is not completely automatic since the "effort-to-see" can be mediated by attention (Francis et al., 2003). Accommodation may be voluntarily and effortfully controlled in some conditions, which in turn affects vergence too, as part of the accommodation reflex (see §4.3) (McLin and Schor, 1988). Another compensatory voluntary action is eyelid squinting, which sharpens vision by reducing the aperture and field stops that leads to mitigation of refractive errors, but which may cause eye strain as well (Sheedy et al., 2003). This voluntary action circumvents the inability to control the involuntary pupil constriction that normally determines the aperture stop of the eye.

The anatomical section of the eye that is behind the lens is strictly optical and unmistakably peripheral. In the auditory system, past the auditory nerve, processing is done in the central nervous system and may therefore be considered more firmly integrated with cognitive functions than the analogous parts of the eye. Most of the eye muscles have no analogs in (human) hearing, so an interoceptive monitoring of auditory accommodation—through sensation of strain, fatigue, and pain—is unavailable to auditory circuits. Physical fatigue of the eye may also tap into muscle fatigue more precisely, which is defined in some contexts as *"an exercise-induced reduction in the ability of*

muscle to produce force or power whether or not the task can be sustained", while noting that the reduction is task dependent (Enoka and Duchateau, 2008).

Inasmuch as forming a sharp image of attended objects is a goal of both the visual and the auditory systems, having analogous low-level accommodation processes that maintain sharpness may help with the disentanglement of low- and high-level components of listening effort and demystify some of its inner workings. Therefore, it is proposed here that listening effort is, in fact, **auditory accommodation effort**—it measures the activity of the process that has to be dynamically performed by the brainstem (perhaps along with other areas and with the mediation of attention, or even fluctuations in the metabolic demands within the organ of Corti) to maintain an optimal image quality, using the various mechanisms considered above.

16.7   Discussion

Although the existence and function of auditory accommodation remains hypothetical at this stage, this is nevertheless the first systematic exploration of the possibility of accommodation in hearing, to the best knowledge of the author. While the number of speculations in this chapter may appear prohibitive, it is maybe comforting to know that much uncertainty had characterized the understanding of ocular accommodation as well. For a long time after it was first discovered by Christoph Scheiner in 1619, it had been debated which one of several possible anatomical mechanisms is at the root of accommodation (Charman, 2008). Only in 1801 did Thomas Young prove that it is caused by changes in the focus of the crystalline lens. Given that a relatively larger portion of the auditory system is neural than the in visual system, hearing does not lend itself as conveniently to the analysis that vision received using optical principles alone. At the same time, the involved circuits in the brainstem are relatively compact, so substituting some of their actions with more intuitive closed-form operations may be possible, despite their complexity.

Of all the mechanisms that were reviewed, the MOCR stands out as the one that can be most readily analogized to accommodation in vision in terms of anatomy. Its existence itself is unmistakeable and its prevalence in different forms in all vertebrates is indicative that its function has to be biologically useful, perhaps more than has been gathered from results of various behavioral studies to date. Its main function, however, remains opaque according to the current literature, and we hypothesized that its role is to control the phase locking precision, which can preferentially enhance noncoherent (intensity envelope-based) detection at the expense of coherent detection. This function seems to be tied with the accommodation of the time-lens curvature, which controls the amount of defocus that determines the separation between coherent and incoherent parts of the stimulus. Consequently, such a change may impact the depth of field too, where it is dictated by the different degrees of coherence of the various elements in the acoustic scene.

The psychoacoustical and physiological evidence presented in §16.3 for what appears to be adaptation to reverberant fields does not clearly coincide with the effects that characterize the MOCR. One difference is their different reaction times, which is about 100 ms for the MOCR and almost 1 s for the reverberation adaption. Therefore, the latter may be a result of one of the other mechanisms that were discussed, or a combination thereof, which is placed somewhere in the brainstem. This can relate either to neural group-delay dispersion, synchronization, or coherent/incoherent mixing accommodation.

The accommodation analysis attempted as much as possible to tease apart the various mechanisms as parameters that can be manipulated independently of one another. However, in all likelihood, as was occasionally implied, sampling rate (rate coding), neural synchrony, neural group dispersion, and filter bandwidth are all tied together somehow. This was seen in the MOCR that affected bandwidth, sampling rate, phase locking, phase curvature, degree of coherence, and gain in

different amounts. Or, it can be inferred from the fact that the increase in the VCN synchronization rate seems to come at the expense of high firing rates. If a combination of such mechanisms turns out to be mostly working in tandem, then the analogy of auditory accommodation may be expanded to include an entire "auditory accommodation reflex" as in the near triad of vision.

By eliminating and refining some of the hypothetical accommodation mechanisms, more clarity may be obtained that can eventually reduce the overall apparent complexity of the auditory system, rather than exacerbate it. This may be obtained by creating more rigorous connections to higher-level phenomena such as listening effort and tinnitus, as well as by unifying parallel concepts from vision.