Opinion
Nina Kraus1,2,3,*, and Travis White-Schwoch1,†

The auditory system is stunning in its capacity for change: a single neuron can modulate its tuning in minutes. Here we articulate a conceptual framework to understand the biology of auditory learning where an animal must engage cognitive, sensorimotor, and reward systems to spark neural remodeling. Central to our framework is a consideration of the auditory system as an integrated whole that interacts with other circuits to guide and refine life in sound. Despite our emphasis on the auditory system, these principles may apply across the nervous system. Understanding neuroplastic changes in both normal and impaired sensory systems guides strategies to improve everyday communication.

Learning, Language, and Communication
Nervous system plasticity has been observed across the animal kingdom from single cells to sophisticated circuits. Sensory systems are prodigious in their ability to reshape response properties following learning, and in the auditory system plasticity has been observed from cochlea to cortex. This learning is fundamental to our ability to function in and adapt to our environments. Experience navigating this sensory world drives language development—perhaps the most remarkable auditory learning task humans accomplish—and it is necessary to understand the principles that govern this plasticity to devise strategies to improve language and communication in normal and disordered systems.

Here we argue that cognitive, sensorimotor, and reward ingredients engender biological changes in sound processing. The mechanisms behind these changes lie in two sets of dichotomous systems: (i) the afferent projections that transmit signals from ear to brain and the efferent projections that propagate signals from brain to ear; and (ii) the primary and nonprimary processing streams that suffuse the auditory neuraxis (Figure 1). We highlight experiments that advance our understanding of the neurophysiological foundations underlying auditory processing (see Glossary) and that offer objective markers of auditory processing in humans. Finally, we place learning in the context of a distributed, but integrated, auditory system.

Rethinking the Auditory System: A Distributed, but Integrated, Circuit
Traditional models characterized the auditory system as series of relay stations along an assembly line, each with distinct functions [1–3]. While these hierarchical models recognized the interconnectivity of the system, the emphasis was to characterize each nucleus’s specialization. The idea was that understanding each station would build each block necessary to construct the auditory circuit, and this ‘inside-out’ approach has contributed greatly to our understanding of auditory neurophysiology.
We propose a complementary ‘outside-in’ approach. Our view is that the auditory system should be thought of as a distributed, but integrated, circuit (Figure 1). Any acoustic event catalyzes activity throughout the auditory neuraxis and we argue that sound processing – and any assay thereof – is a reflection of this integrated network. Although each structure is specialized to perform a specific function, this specialization has evolved in the context of the entire circuit. To understand auditory learning, then, we are forced to move past a focus on an individual processing station as a single locus of activity, expertise, or disorder.

Our view is consistent with an emerging trend in neuroscience to consider the interplay of multiple processing stations and the ‘give and take’ between cortical and/or subcortical systems underlying human behavior [4–8].

**Plasticity in the Human Auditory System: A Double-Edged Sword**

We regard everyday auditory experience as a learning process that shapes the nervous system, not least because auditory experience is necessary for the maturation of basic auditory circuits [9–11,106]. These changes may be exacerbated – for better or worse – and cases of expertise and deprivation both contribute to understanding how experience shapes auditory circuitry [12]. Neuroplasticity must therefore be viewed as a double-edged sword. The cognitive, sensorimotor, and reward ingredients of auditory experience drive plasticity and a hypothesis based on this framework is that insults to any of these domains dictate the resulting phenotype.
Of particular interest in our research program is the neural coding of fast auditory events, such as the details that convey phonemic identity in speech (Box 1). Our laboratory has developed an approach to index the influence of life experience on the neural coding of these fast acoustic details called the frequency-following response (FFR). We have previously referred to this as the auditory brainstem response to complex sounds (cABR), but fear that this terminology
undermines the integrated and experience-dependent nature of the activity it indexes. The FFR is as complex as the eliciting stimulus and we use ‘FFR’ to refer to the product of aggregate neural activity in the auditory midbrain that reflects the coding of aggregate speech features, including activity that ‘follows’ both transient and static acoustic cues; because the auditory midbrain is a ‘hub’ of intersecting afferent and efferent auditory projections, in addition to projections to and from non-auditory cortices, its response properties are shaped by this constellation of cognitive, sensory, and reward input (Figure 2, Key Figure). Thus, despite its subcortical basis, the FFR reflects the distributed, but integrated, activity emblematic of auditory processing.

This research emphasizes the imprint of changes to the auditory system that affect the automatic sound processing that is always on and cannot volitionally be turned off, even after training has stopped [13,14]. Thus, biological infrastructure in the auditory system is influenced by an
individual’s life in sound. No two people hear the world exactly the same way because acoustic experiences impart enduring biological legacies (Figure 2).

**Cognitive Influences on Auditory Processing**

The cognitive component of our framework is grounded in these principles: (i) listening is an active process that engages cognitive networks; (ii) the precision of automatic sound processing in the brain is linked to cognitive skills such as attention and working memory; and (iii) the cognitive systems engaged during listening selectively modulate the aspects of sound that are behaviorally relevant. The legacy of this repeated, active engagement is engrained in the nervous system over time as listeners make sound-to-meaning connections.

Speech understanding relies on the ability to pull on cognitive functions such as working memory and attention [15–18]; engagement of these systems strengthens the neural circuits that facilitate listening [19]. One study showed that cognitive factors shape auditory learning in an experiment comparing two groups of rats [20]. The first group trained to attend to frequency contrasts in a set of tones, whereas a different group trained on intensity contrasts; crucially, identical stimulus sets were used in the two groups. Cortical maps changed along the acoustic dimension that was trained, demonstrating that what is relevant to an animal dictates map plasticity (see also [107]).

In humans, several studies show links between the integrity of the neural processing of sound and cognitive abilities [21], suggesting that the legacy of cognitive engagement is revealed through the precision of neural function. Additionally, this suggests that training to strengthen a cognitive skill propagates to sensory systems [22–24].

Many of these insights come from studies of music training, which provides a model to understand the biology of auditory learning [25–27]. Making music requires an individual to engage multiple cognitive systems and to direct attention to the sounds that are heard, produced, and manipulated. The physical act of producing sound – through instrument or voice – mandates intricate motor control and stimulates auditory–motor projections [28]. In addition, music is an inherently rewarding stimulus that elicits activity throughout the limbic system [29]. The musician's brain has been finely tuned to process sound and the musician is a case to explore what is possible in terms of experience-dependent plasticity.

Regarding cognitive–sensory coupling, individuals with music training exhibit stronger neural coding of speech in noise concomitant with heightened auditory working memory [30]. Contrast this with a bilingual, who exhibits stronger neural coding of pitch cues concomitant with heightened inhibitory control [31]. A musician needs to pull out another instrument’s ‘voice’ from an ensemble while mentally rehearsing a musical excerpt, facilitating the processing of signals in a complex soundscape and exercising working memory. However, a bilingual needs to actively suppress one mental lexicon while using voice pitch as a cue to activate the appropriate one. Whereas music training is associated with superior speech recognition in certain types of background noise [32,33] (cf. [34]), the cognitive systems engaged through bilingualism create a different situation. Bilinguals have superior recognition of non-speech sounds in noise but inferior recognition of speech in noise, due to cognitive interference from the mental lexicon they are attempting to suppress during active listening [35,36]. Thus, the impact of this cognitive–sensory coupling for everyday listening skills depends on what constellation of cognitive and sensory skills are rehearsed. This juxtaposition illustrates an important principle of auditory plasticity: cognitive systems tune into particular details of sound and selectively modulate the sensory systems that represent those features (Figure 2). By analogy, then, auditory learning may be thought of as a ‘mixing board’ more than a single ‘volume knob’, with distinct aspects of neural coding selectively modulated as a function of the precipitating experience [37,38]. This contrast
also reinforces the notion of a double-edged sword in experience-dependent plasticity and adds a layer of nuance: within an individual some listening skills may be strengthened whereas others may be suppressed.

There is similarly a tight interplay between cognitive and sensory losses; older adults with hearing loss exhibit faster declines in working memory, presumably because degraded auditory acuity limits opportunities for cognitively engaging and socially rewarding interactions [39]. Training these cognitive skills, however, cascades to boosts in sensory processing. For example, older adults have delayed neural timing in response to consonants but not vowels [40]; auditory–cognitive training that directs attention to consonants (including built-in reward cues) reverses this signature aging effect [23]. Similar phenomena are observed following cognitive interventions in the visual system [24,41].

These studies illustrate that identical neural pathways are imputed in disorder and its remediation and are consistent with the view that both should be conceptualized as auditory learning. They demonstrate how fine-grained aspects of sound processing are selectively modulated based on the cognitive demands and bottlenecks of the experience (Figure 2). Moreover, these cases exemplify the coupling between the integrity with which the nervous system transcribes sound and the cognitive skills important for everyday listening.

**Sensorimotor Influences on Auditory Processing**

The sensorimotor component of our framework is grounded in these principles: (i) the infrastructure responsible for encoding basic sound features is labile; (ii) extreme cases of deprivation and expertise illuminate mechanisms that apply to a typical system; and (iii) the entire auditory pathway—including the hair cells—can be thought of as sites of ‘memory storage’ because response properties reflect the legacy of auditory experience(s).

Basic sensory infrastructure has a potential for reorganization. The most extreme examples comprise cases of profound deprivation, such as deafness, blindness, or amputation, where sensory cortices are coopted by circuits dedicated to the remaining senses, but only after a period of adaptation (that is, learning) [42,43,108]. These extremes illustrate the brain’s potential for reorganization and the mechanisms underlying this remodeling.

In terms of expertise, music again offers a model for auditory learning. Musicians process sound more efficiently even when not playing music, suggesting that repeated active engagement with sound shapes the automatic state of the nervous system [19]. The imprint of music training extends all the way to the outer hair cells of the cochlea [44,45]. The musician model also demonstrates that sensory input alone is insufficient to drive neural remodeling: comparisons between children undergoing active music training (that engages cognitive, motor, and reward networks) and those in music appreciation classes have shown neurophysiological changes only in the former [46]. Thus, sensory input may be necessary, but not sufficient, for auditory learning [47].

With regards to language learning, evidence from songbirds demonstrates a causal role for the basal ganglia in song learning [48], suggesting a role for the motor system in language learning. We are just beginning to learn how the motor system is involved in auditory learning in humans, but it seems that motor acuity is tied to language abilities [49,50] and that training rhythmic skills can boost literacy skills [51]. The rhythm–language link may underlie the observation that music training confers gains in reading achievement.

Finally, we mention an example of sensory learning that on its surface appears to occur automatically. Infants quickly learn statistical regularities in the acoustic environment and this is thought to contribute to language acquisition [52]. However, not even these ostensibly passive
learning processes are exempt from cognitive influence: prior experience and active expectations guide statistical learning [53,54]. Thus, as young as infancy, listeners can connect incoming sounds to meaning and also exert meaning on incoming sounds, reinforcing the interplay between sensorimotor and cognitive systems in auditory learning.

**Reward (Limbic) Influences on Auditory Processing**

The reward component of our framework is grounded in these principles: (i) reward systems spark reorganization in fundamental auditory infrastructure; (ii) social and reward contexts gate auditory learning in humans; but (iii) limbic input can create the conditions to learn something that does not optimize auditory processing.

We learn what we care about. Consequently, the limbic system is likely to facilitate neural remodeling. Classic studies show that stimulation of the cholinergic nucleus basalis galvanizes cortical map reorganization [55,56] see also [109,110]. Aberrant sensory–limbic coupling, in turn, is involved in disorders such as tinnitus [57], but also in their treatment [47]. This again emphasizes that identical networks are implicated in conditions of both enhancement and diminution of sound processing.

Less is known about how the limbic system guides auditory learning in humans, in part due to practical limitations in controlling the expression of neuromodulators (although early evidence is promising [58]). Once again, music training provides a model: listening to and producing music activates multiple auditory–limbic projections [29,59]. Given that music training directs attention to minute details of sound in a rewarding context, it stands to reason that these neuromodulators play a role in the resulting neural remodeling.

The limbic system may also play a role in language development. It has been argued that infants must tune into the aforementioned statistical patterns in the auditory environment to jumpstart language learning, but that these computations are gated by social (i.e., reward) context [60]. For example, infants learn non-native phonemic contrasts when they are modeled by a tutor speaking ‘motherese’, but only if that tutor is present and interacting with the child – a video of the tutor is insufficient [61].

Deficits in reward input, then, are hypothesized to contribute to language impairment. Children with autism, for example, exhibit reduced functional connectivity between limbic structures and voice-selective auditory cortex, which suggests a decoupling of sensorimotor and reward networks during everyday listening [62]. Indeed, many children with autism show poor neural coding of prosodic cues that convey emotional and pragmatic information [63].

Children whose mothers have relatively low levels of education – a proxy for socioeconomic status – present a different case of deprivation. Children in these homes hear approximately 30 million fewer words than their peers; in addition, they hear two-fifths of the number of different words, meaning that both the quantity and the quality of their everyday linguistic input is impoverished [64]. Consider that a mother’s voice is perhaps the single most rewarding sensory cue available to a child. If the sensory input is impoverished, but the conditions are right for learning, what is learned may itself be impoverished. This linguistic impoverishment is reflected by poor neural coding and cognitive skills [65] (Figure 2). This is consistent with evidence from animal models that environmental deprivation constrains nervous system development; environmental enrichment, however, reverses this maladaptive plasticity [66], reinforcing the concept of auditory learning as a double-edged sword. This hypothesis also aligns with evidence that task reward structure shapes not only whether plasticity occurs, but how it manifests [67,68].

Taken together, these studies illustrate that, on the one hand, a lack of reward structure stymies the mechanisms of auditory learning. On the other hand, sufficient reward structure with an
impoverished content may cause learning of the wrong material. Presumably, this principle applies to the cognitive and sensorimotor aspects of auditory learning as well.

**Mechanisms of Learning**

Having laid the groundwork to understand that cognitive, sensorimotor, and reward systems are necessary to drive neural remodeling, the question arises: how do these systems influence automatic sound processing?

Two anatomical dichotomies help navigate this integrated circuit and its role in auditory learning: afferent versus efferent projections and the primary versus nonprimary pathways.

**Afferent and Efferent Projections: Bottom-Up Meets Top-Down**

The first dichotomy comprises the projections that feed signals forwards and backwards through the auditory system (Figure 1). The bottom-up afferent projections transmit information forwards to accomplish signal processing ('ear to brain') whereas the top-down efferent projections propagate signals backwards ('brain to ear'); both extend between the cochlea and cortex [69] and the latter mediates remodeling in subcortical structures [70,71].

Our proposal is that the efferent network shapes automatic response properties in cochlear and subcortical systems, which is why the basic response properties of the auditory system, such as otoacoustic emissions and electrophysiological responses, reflect life experiences in sound. It has been argued that similar mechanisms underlie both attention-driven online changes and long-term plasticity [67]. This leads to the hypothesis that if these experiences that engage cognitive, sensorimotor, and reward systems are repeated sufficiently they can, over time, facilitate functional remodeling by imparting a 'memory' to afferent processing [19,72] and future learning [73–75].

We hypothesize that the efferent system has become larger and more intricate evolutionarily with increasingly sophisticated auditory behaviors. Numerous complex auditory behaviors – many of which are important for listening in everyday situations – are similar across species. This includes the learning observed in animals with precocious auditory abilities such as bats [76], ferrets [77], and humans [78]. This may be due to convergent evolution, the independent evolution of a trait in distinct lineages based on the needs of the organism. These behaviors are perhaps most sophisticated in humans (Figure 3) and we speculate that the convergent evolution of efferent projections may underlie some of these behaviors and the key role that auditory learning plays in developing the skills necessary for effective everyday communication. If one accepts that language learning pulls on the circuitry necessary for auditory learning, one could imagine a role for the efferent system in language development and poor activation of these top-down networks as a chief factor in language impairment [79].

**Primary and Nonprimary Divisions**

The second dichotomy pertains to auditory structures such as the cochlear nucleus, inferior colliculus, thalamus, and cortex: the distinction between primary and nonprimary pathways (also known as lemniscal and paralemniscal, cochleotopic and diffuse, or highway and country road; Figure 4). Neurons in the primary pathway are biased to respond to auditory stimuli, whereas the nonprimary neurons are more multisensory. The primary pathway is tonotopically organized, shows sharp tuning, and strongly phase locks to the stimulus, whereas the nonprimary pathway is not especially tonotopic, has broader tuning, and does not time lock as strongly to stimuli [80]. Partly for these reasons, it is thought that the primary auditory cortex ('core') represents nearly all incoming signals whereas the nonprimary cortex ('belt' and 'parabelt') specializes for communication signals such as speech and music [81–83].
Adopting our systems-wide perspective, however, an additional distinction emerges: the primary processing stream preferentially codes fast temporal information whereas the non-primary stream codes relatively slow information [84–86]. This hypothesis is consistent with evidence from the rat trigeminal system [87] and primate visual system [88] that parallel pathways code fast and slow information. The functional consequences for language development are only beginning to be understood; however, preliminary evidence suggests that deficits in either fast or slow auditory–temporal processing may lead to language impairment, but may not necessarily co-occur [89–91].

Less is known with regard to learning and the primary versus nonprimary pathways. The nonprimary pathway has been implicated in rapid task-related plasticity, such as adapting to stimulus context [92] and classical conditioning [93]. During active listening, neurons in the prefrontal cortex first change their tuning, followed by neurons in the nonprimary auditory cortex, and then finally neurons in the primary auditory cortex [7]. This leads to the hypothesis that the nonprimary system is more labile than the primary circuitry and may facilitate rapid online learning and adaptation in connection to cognitive and reward circuits. We speculate that changes to the afferent pathway are biased towards stability: the system exhibits a more enduring physiology that resists transient changes and relatively few of the projections exhibit task-related evanescence (Figure 4). Conversely, the efferent pathway is biased towards in-the-moment changes; this evanescence facilitates phenomena such as selective attention to one speaker. This system is relatively less persistent in sound processing. The more an activity is done, repeated, and over-learned, the more likely remodeling will occur in the primary pathway and, eventually, influence afferent processing.
This primary-enduring, nonprimary-evanescent distinction may underlie the capacity to strike a balance between the stability of auditory processing and malleability in attention, adaptation, and learning (see Outstanding Questions).

**Concluding Remarks and Future Directions**

We have reviewed the auditory system’s ability to change. In particular, we have argued that cognitive, sensorimotor, and reward systems optimize auditory learning and that this learning underlies success in everyday language and communication. We have also argued that the auditory system should be thought of as a distributed, but integrated, circuit that is a moving target; for better or worse, its response properties change through the interplay of cognitive and reward circuits during everyday listening. Thus, both expertise and disorder should be considered from a common standpoint of neuroplasticity. While our emphasis has been on the auditory system, we argue that these principles extend to other sensory systems [4,6,8].

The recognition that states of decline, deprivation, and disorder should be viewed through a lens of plasticity suggests that they may, in part, be reversible. If the same pathways are responsible...
for expertise and disorder, the conditions that facilitate expertise may ameliorate communicative difficulties. Our framework therefore makes a clear case for auditory training as an intervention for listening and language difficulties, and – providing the training integrates cognitive, sensorimotor, and reward systems – early evidence is promising.

Finally, we have highlighted how measuring the integrity of sound processing at basic levels of the auditory system opens a window on human communication and the imprint of a life spent in sound. A healthy brain is labile and stable, able to adapt to new environments while pulling on knowledge and experience to make sense of the sensory world. Thus, in addition to motivating and informing interventions, our framework can help facilitate training by identifying an individual’s strengths and weaknesses in the neural processes important for everyday communication.

Acknowledgments

The authors are grateful to Trent Nicol for his input on the manuscript and to colleagues in the Auditory Neuroscience Laboratory, past and present, for the many experiments and conversations that motivated this framework. Supported by the NIH (ROI HD069414), the NSF (BCS 1430430), and the Knowles Hearing Center.

References

Trends in Cognitive Sciences