Cognitive Models of the Hearing and Listening Mind

Barry Blesser and Linda Salter

In pursuing our sound-related research, our sound art-making, or our spatial designs, we tend to view hearing and listening through the lens of our own preoccupations. If we are researching a type of sonic experience, we focus on our research process as it explores a particular set of sonic stimuli. If we are composing a piece of sound art, we focus on our internal creative process as we manipulate the sounds. If we are designing spaces for human habitation, we focus on our sense of aesthetics and the nominal function the space will serve. In other words, our work tends to be self-contained and ballistic, as in, being concerned with the origin of the activity – the original sonic stimuli and our thoughts and feelings about them – and being much less concerned with the outcome or reception of our activity – how the sonic stimuli and our analyses, compositions, or designs are received by others and processed for meaning. If we don't focus on our own inner imperatives, nothing will be accomplished. If we focus only on our own ideas, our work is of limited usefulness and meaning for others.

All activity is part of a complex system; it is more than units of individual experience divorced from the interactions among the units, or between the units and their context. We need to fully comprehend and account for the total system in which the sonic stimuli and our work with them are embedded. This means considering the physical and psycho-socio-cultural contexts of the original set of sonic stimuli and our work with them, plus accounting for the way these sounds are being heard and listened to by those who receive our thoughts and actions with the sounds.

The matrix which produces the final product is intricate and interactive. When we expand beyond our own boundaries and concerns to understand hearing and listening not only from the points of view of creators and researchers of sound, but also to include the experience of the receivers of sound, we are achieving a wider and deeper impact for our activities. In addition to being aware of our own agendas, we become comfortable with those of the listener: what are they listening to? What are they hearing? How is their brain processing the sonic signal, and what meanings are they creating in their mind?

It is important to remember that while sound exists external to the perceiver, it is only within the perceiver's mind that sound achieves meaning. We construct our world, we don't receive it as a total and complete unit. Thus each of us hears a different sound, and each of us experiences a different meaning of that sound. Meaning is a construct that exists in the mind of the perceiver. We create the world in which we live; we hear with our mind, not our ears.[1]

This paper is an attempt to explore a multiplicity of ways in which the brain processes sonic neural signals to create meaning. We will present a collection of experiential fragments from which we can infer various types of brain processing of the sonic stimulus. In attempting to understand these fragments, we will build a representation or paradigm of the internal listener that we can then use to reflect on and refine our own research, composing, or spatial designing.

We are bombarded by sensory stimuli constantly, but we do not consciously process all these stimuli. It would be impossible with the brain size humans have developed. Current thinking[2] is that we have internal mental models of the outside world serving as templates that account for everyday sensory stimuli. As long as the sensory stimuli we are receiving match our internal brain models, we don't pay attention – these stimuli are processed on the unconscious level and become part of our automatically accessed, implicit memory. If we try to pay attention to how we function at this level, we become like the centipede who, in the children's story, can no longer walk when she is asked to account for the movement of a particular leg.

But we do consciously process external sensory stimuli under certain conditions. For example, we pay attention when there is a noticeable difference between an external stimulus and our internal model of the world. We also pay attention when we have a specific task to perform and we need specific sensory information to complete it. In addition, we pay attention when there is motivation to pay attention to specific stimuli because doing so will result in a desired pay-off.

By exploring some of the ways in which our brain processes external sonic stimuli, considering both conscious and subconscious activities, we can create insightful models of the complex procedures by which we achieve meaningful listening and hearing.

Neuro-plasticity: How the brain self-modifies and learns under the influence of motivation

The human brain is highly plastic: throughout our lifetime we continue to process sensory stimuli and, when useful, to learn and change in response to this processing. How do we learn, how does our brain change? One way is called Long-Term Potentiation (LTP). It is the process of strengthening the synaptic connections between neurons through increased brain activity. The increased neuronal activity provides chemical support for greater inter-neuron signal transmission, resulting in a long-lasting change in the brain. LTP is one of the significant cellular mechanisms underlying learning and memory. And of course, the opposite occurs: Long-Term Depression (LTD). This occurs when diminished stimulation of synaptic connections results in weakened chemical support for selective synapses. LTD is required because there appears to be a limit to the amount of synapses the brain can support – if the ceiling were reached, encoding of new information would become limited. We rely on both LTD and LTP (among other phenomena) as forms of synaptic plasticity – the ability to process and learn new things and to forget old things.[3]

Under the influence of motivation, an individual has a reason, or a pay-off, for changing their internal mental model of the world through active processing of sensory stimulation, resulting in LTP, and the forming of implicit and explicit memories. While this process of neural plasticity, or learning, at first takes place at the level of conscious awareness, it quickly becomes part of the unconscious level, outside of our conscious awareness: explicit learning becomes part of our implicit memory[4] – we just “know” things. What follows are several examples of these motivation-driven processes involving hearing and learning:

(1) In a research project to explore very early language learning, when babies were taught Mandarin Chinese by a human being in person, they learned well. Alternatively, if they were taught Mandarin virtually by an audio-visual teacher – i.e. not in-person – they learned very little. The same result held for just-audio teaching. It has been theorized that for the first group, the pay-off for learning Mandarin phonemes was not to learn the language in order to speak it; it was to actualize a connection to a real person. Humans like to connect to others. From the moment of birth, infants search for, create, enjoy and sustain connection with a caretaker, probably to maximize survival. In order to connect to the human teacher, the babies learned Mandarin phonemes – they pleased the human instructor. In the second group, there were no live humans involved so there was no payoff for learning Mandarin phonemes, there was no one to please and thus connect to.

Similarly, in a research project where babies learned from audio-visual screens with and without peers (other babies present in the learning environment), those who had peer partners, especially a variety of peer partners, learned better than those exposed to audio-visual screen learning by themselves. Again, babies are driven to connect to real persons; learning together with a peer provided a means to connect to another baby.

The payoff for putting in the work of focusing, concentrating, and remembering was the building of a real connection with a real human. In the first experiment, without a live teacher there was no payoff, so the infants didn't bother to put in the hard work. In the second experiment, with a peer partnering them, particularly a variety of different babies, they experienced a connection so they did work hard. Additionally, when the peers were varied, it is possible that they were also stimulated by newness in the experience, and so there was another payoff for putting in the effort to focus and learn.[5]

(2) Before shipping a piece of electronic audio equipment to a customer, it is tested to make sure it is performing at a high standard. At some point after the equipment has been installed and constantly used by audio engineers, these users might complain about the audio equipment's quality – these engineers could hear things not accounted for in the original product design. In this case, the users ship the equipment back to the manufacturer to be re-calibrated so as to respond to what the users are now hearing. The original designers of the equipment now have high motivation to increase the sensitivity of their own auditory processing, to add new neuronal connections which enable them to perceive and process more and more subtle auditory detail that is being heard by the constant users and therefore can be included in the design – there is a well-understood professional payoff for the original designers.

In this case, both groups experienced long-term learning (LTP) that encodes an ability to discern what they were hearing. The sonic stimuli are always present, but the users/designers needed exposure and training to hear more of these sounds. On the part of the user/engineers, their brains were being stimulated through continued exposure to audio signals, with the result of increasingly finer distinctions – increased connectivity – within their hearing synapses. On the part of the designer/engineers, they had a motivation to learn what the users had learned, so they could adjust their equipment to respond accordingly and provide support for the users.

(3) There is a lot of information in the sounds available for us to hear in any space. Sound reflects differently off the objects and the built surfaces (the four walls, floor and ceiling) in a space, and these different reflections can be heard, differentiated, and identified with training. It should be emphasized that these reflection-sounds are always present in the space, but unless there is motivation to hear them, to differentiate them from other background sounds, and to accurately interpret them, they are functionally not heard. If we need to use this information, if we have a motivation, for example if we are blind, we will pay attention to and

actively process this sonic information, enabling us to consciously hear the sonic reflections, to understand their meanings, and to engage in successful spatial reading.

We can expand our synapses for functional hearing of sounds in space. There is a piece of neural structure in the brain that identifies objects in a physical space. Usually this piece utilizes visual stimuli, but in the case of a person using echolocation sonic stimuli, we use it to map objects in space using only sound. If we need to, if we train ourselves to hear and interpret these sounds, we can "read" our surroundings – we can produce a functional "image" of our surroundings. Using sound, not only can we determine where the objects and built surfaces are, but also what shape and what size they are, what material they are made of, and in what direction and how quickly the object is/isn't moving relative to the self.

While the visual cortex integrates these reflective sounds to create a "picture" of the external world, it is actually not a visual picture, but rather is a workable construction of the physical external world. The brain constructs this reality in a complex way by combining the immediate auditory stimuli with a rich assembly of previous sonic experiences that form models for comparison with the current sonic cues. Sometimes additional sensory cues are used, such as smell and changes in skin temperature. In this way, we "see" by using our hearing/listening abilities combined with other sensory experiences.

This image-construction process is not aimed at achieving an accurate representation of the totality of the physical reality, but exists for the sake of utility. "Seeing" is carried out by the brain, not the eyes. We don't actually know how our brain "pictures" something. We only know that for those with high motivation, their brain is acquiring and using auditory information which allows them to exist in and move through a physical space successfully and safely. They are highly motivated to increase those neural connections (LTP) needed to perform this way-finding function by using sensory stimuli other than the usual visual stimuli.

There are many examples of individuals who have learned to navigate through space without using sight, but instead by using sound reflections. The singer Ray Charles is an excellent example. Another is Daniel Kish, who has been blind since he was one year-old. He uses echolocation to ride his bike through the woods. Daniel emits a clicking sound with his tongue, and "reads" the resulting reflections to inform him about his spatial environment; he calls his technique "FlashSonar."[6]

(4) Hearing is more than an addendum to seeing. Hearing is crucial in moving through space, even though we don't realize it. Research has shown that older people who have noticeably limited hearing fall more often than those who can hear well. This is true even when

accounting for the increased likelihood for failing in the elderly. It has been suggested that we need to be able to hear the sonic feedback from sound bouncing off objects and building-surfaces in order to fully experience and form a functional image of our spatial surroundings. Avoidance of falling certainly provides significant motivation for careful and attentive listening to our surroundings.

Perhaps, as we age and begin to lose some degree of hearing, those synapses that have formed through stimulation and allow us to form a functional "image" of a space begin to lose their chemical support as stimulation lessens. We experience Long-Term Depression of these synaptic connections, and thus we have less awareness of and knowledge of our surroundings in either implicit or explicit memory. We don't get the information we need to successfully and safely navigate our physical/sonic surroundings.[7]

(5) Cultures place differing degrees of emphasis on various sensory stimuli. For example, research shows that members of traditional cultures that particularly value their specialist musical heritage were better able to describe sounds, while individuals living in a traditional culture that produces patterned pottery were better able to describe shapes. As for tastes, members of Farsi- and Lao-speaking cultures showed almost perfect scores identifying tastes, while English speakers had difficulty speaking about even basic tastes. Describing smells appeared to be the most difficult sense for members of most cultures.[8]

We get a cultural payoff for investing in building neuronal connections for identifying, labeling and using particular sensory stimuli. If our culture or sub-culture (family, social group, community, profession, nationality, race, etc.) values the use of a particular sensory activity, we have motivation to invest the time and energy in building new neuronal connections to support using that highly valued sensory system. The opposite also holds true: minimal cultural and experiential payoff depresses the forming of such sensory neuronal connections.

Sub-groups such as particular professions can be expected to differ in the degree to which they develop and utilize different senses. Obviously, those in the food business can be expected to have heightened abilities to experience scent and taste. In Western culture, architecture was traditionally one of the five fine arts, with training focusing on aesthetic qualities. Even today, architects are trained and rewarded for visual excellence with famous awards and features in architectural magazines focusing on the looks, not the human functionality, of buildings. Thus, architects, and those in associated professions of design and planning, are motivated to see much more than to hear. They favor visual data over other sensory date in creating, evaluating and relating to their physical surroundings. As visually oriented experts, it is possible that they have less developed neuronal connections for hearing generally and hearing space in particular than those individuals whose professions are linked to hearing, such as musicians, teachers, actors and dancers, psychotherapists engaged in talk therapy, etc. However, if their professional payoff systems were to recognize the importance of listening and hearing by actually valuing the well-being of human users in their designed spaces, then design practitioners would be motivated to train their minds to be more active in listening and hearing. Design professionals would be motivated to build the neuronal connections that create and sustain the skills involved in using their auditory sense.

There are other professions where greater attention to sound and listening would produce recognizable payoffs, such as medical professionals, plumbers, teachers, and retailers. These individuals would most likely function noticeably more effectively if they paid more attention to what they heard. Professional success would create payoff – motivation – for building neuronal connections supporting the use of the auditory sense.

How the mind's ear functions

As indicated above, we are continuously bombarded by sonic stimuli. We pay active attention to only a small subset of these stimuli, and the content of that subset differs according to person, time, and circumstances. We hear with our mind, not our ears. Each of us hears something different, and at different times and under different circumstances, what we hear differs.

The sonic vibrations that reach our ears are different from the meanings we ascribe to these vibrations using cognitive processing. Sonic vibrations in the air do not have a one-to-one correlation with what we "hear" in our mind. The mind creates meaning for sonic stimuli by processing them through internal prisms such as personal experience and personality, cultural influences, social training, and immediate internal and external context. While we cannot apprehend and measure all the possible influences that our mind brings to bear on the cognitive meaning-creating process, some influences on perception can be measured. For example, the influence of vision, and the interference from ambient sounds such as conversation, loud music, or outside-world sounds like ambulance sirens and car horns are amenable to measurement because they are real-time events. However, there are many other influences that are harder to replicate and fully understand, including the perceiver's prior history, motivation and goals, training and experiences, age, emotions, memory, blood sugar/health/pain levels, etc. All of these influences combine within each perceiver to produce a hearing and listening experience unique to each perceiver, time and place.

The difference between the original sonic stimulus and the perceptual cognitive experience of that stimulus can vary considerably. For example, what we think we hear may or may not actually be present in the original sonic stimuli that reach our ears. All of us have experienced the difference between what we think we said and what a listener thinks we said, and vice versa, for example: “You said...” vs. “I never said ...” The meaning we extract from a sonic stimulus in this kind of situation depends on immediate perception, short-term memory, past experience and associated emotions, motivation, ambient interference, etc.

In addition, there are other experiences that have been studied in a laboratory setting where the external sonic signals and the internal sonic meanings don’t track closely.

The following discussions explore some of the ways in which what we hear is determined by our mental processing mechanisms.

(1) Subjects hear a sentence with a cough inserted in the middle. Subjects don’t report the cough as an interruption or as a blank spot in the flow of meaning. Rather, the subjects supply the missing word in the sentence as if it was there beneath the cough. One possible explanation for why we fill in the blank with meaning is that we have the capacity to call on prior knowledge and experience with the language to create a meaningful whole sentence. Our cognitive processing supplies the missing word without realizing we have done so. We create a functional gestalt based on what we expect the sentence to sound like. We are able to function in many situations because we have created, through experience, an internal model about how we expect the world to function. We don’t work off the most immediate sensory data, but rather, we refer to our internal model. Rather than building perception from bits of immediate sensory data, we match the sensory data to our internal expectations: we perceive what we expect to perceive, not what is actually there. The cough is assumed to be covering a word, and according to our knowledge of the language, we have a pretty good idea of what that missing word would be.

In another part of the experiment, subjects hear the sentence with a blank space instead of a word. This time they report hearing the blank space where the missing word would go. In this case, where there is an obvious mismatch between the expected sensory data and the actual sensory data – such as a gap in the flow of a sentence – that not-predicted sensory datum is noted, and if deemed necessary, the predicting model is adjusted. Because perception is a process of matching sensory inputs with internal predictions, we function on "automatic", not being aware of our sensory processing. It is only when there is a mismatch, a difference in the actual input (a gap) from what our internal mental model primes us to be receiving (a meaningful word), that we pay attention and note the input; in this case, we notice the gap.[9]

(2) Another example of hearing what is not in the stimulus is the McGurk effect. This occurs when there is a conflict between visual speech, meaning the movements of a speaker's mouth and lips, and auditory speech, which are the sounds the speaker is actually uttering. For example, using video manipulation, a listener is shown the mouth saying the syllable "ga" while the actual sound uttered is "ba". The listener's brain attempts to resolve the conflict between what it is hearing with what it is seeing. The resolution is that, in this case, vision overrides hearing: the speaker's lips trick the brain into hearing a third syllable – "da".[10]

The brain processes a sensory stimulus in a non-linear manner. Instead of the stimulus moving linearly from the sensory organ through various parts of the brain until it reaches the point of meaning assignment, we now believe there are multiplicities of feedback loops connecting different neurons or neuronal packages with the information moving through the brain forwards and backwards via feedback loops. This is termed "recurrence".[11] Because of this rich looping processing, the different senses influence the processing of each other. Vision affects hearing, affects smelling, affects physical feeling, and so on. This is another example demonstrating that we hear with our mind, not with our hearing sense. The visual and auditory sensory systems, densely tied in with each other, mutually influence each other. In this case, the result is what we term an illusion, but it is really the normal functioning of the brain as it tries to produce a unified story of events.[12]

(3) Learning to comprehend a spoken language occurs when the brain creates a model for speech where the parts are identified and combined in certain rule-based, meaningful ways, such that understanding ensues.

Spectrally rotated speech is a laboratory-created artifact in which the high frequencies and low frequencies of words in a sentence are artificially exchanged. This renders the speech unintelligible, perhaps because the linguistic rule is broken concerning which words and parts of a sentence are expected to be uttered in high or low frequencies. Much of language meaning is communicated through prosodic elements, not just through vocabulary and grammar.

Exploring how we successfully learn to understand this kind of speech can help us make sense of the processes by which the brain reconciles what is heard in the outside world with the internal mental model of speech that has been created in the brain. Under training, the listener is able, as a first step, to discern the larger structure – the grammar of the sentence. Over time, the listener is able to insert the differently pitched vocabulary into the large overall structure. A new internal model of comprehensible speech is formed in the brain. Sonic stimuli that, at first, appeared to have no relationship with previously-existing sonic meaning models gradually comes to track with the new model: nonsense becomes sense.

Meaningful language models, and all mental models, influence what we hear by matching pre-existing models with the incoming sensory data. The pre-existing models seek, at first, to match the new stimuli to what is already known and understood. The brain strives for a match in order to achieve meaningful sensory functioning. If the match cannot take place, the brain can do various things, including, as in this case, modifying the existing models to encompass the new sensory data. This is an example of the value of our neuroplasticity. This is how we learn new ways of doing things, such as hearing and understanding new languages. We also learn to match visual embodiments of a language with the sounds and meanings of the language, such as that which occurs when we learn to read, or learn to understand other cultural symbols such as traffic signs or dress codes.[13]

Overlapping and supplementary sensory contributions to cognition

We are complex beings who use overlapping and supplementary sensory input to form a rich picture of our environments. The looping forwards and backwards with multiple neuronal systems that occurs in the brain as we strive to create a meaningful and consonant set of models provides and supports this cross-sensory mutual influencing. The more we can know about our world using all our senses, the more successfully we can survive and thrive.

Film and video games are usually considered visual experiences, but in fact, sound is also central to the ability of the visuals to sell the story, characters and events to audiences. Sound has a strong link to emotions. Combining emotional stimulation via sound with emotional stimulation via vision creates a much stronger and more convincing experience than just one sense alone.

There are three kinds of sound which are used in film and game design: human voices – usually uttering the characters' dialog; music – usually forming the background for the foreground acting; and sound effects – these are either in the background to create an ambiance, or directly highlight the action. Sound produces contextual information which instantly identifies the characters and telegraphs the direction of the action; it also highlights the visceral content of the story and emotionally ties the audience to the product. Jesse Schell, a video game designer and CEO of Schell games has said: “sound is what truly convinces the mind it is in a place; in other words, ‘hearing is believing.’”[14]

To explore the power of sound, we can begin in pre-history with the sound effects created by simple musical instruments to accompany the storyteller’s tales around campfires, and by the minstrels performing in the courts of rulers in antique times. The Greeks and Romans believed that music could arouse powerful emotions in listeners; thus, music played an important role in the social communication and control achieved through performing plays and mass entertainments. In the Baroque era, musicians had to secure a patron – church, state, or privately wealthy individuals – in order to earn a living. Through the influence of patronage, music achieved a degree of significance by communicating the social views that benefitted the patrons. Emotive music accompanied by drama and dance, and supported by richly elaborate costumes and sets, powerfully impressed the public. Mutually enforcing, overlapping sensory stimulation had a strong effect on viewers, communicating and reinforcing messages of power, awe and control by the social institutions of the time.

In modern times, consider the way sound enhanced the communication capacity of early radio programs. In 1930, a program called *The Shadow* was introduced. The voiced introduction, uttered in a deep and portentous male voice, intoned: “Who knows what evil lurks in the hearts of men? The Shadow knows!” and was accompanied by an ominous laugh and a musical theme, an excerpt from a piece by Saint-Saëns. This instantly became an iconic phrase, and boosted the sales of written material about the crime fighting character known as “The Shadow”. A few years later, in 1941, the radio anthology of stories presenting mystery, terror and suspense called *Inner Sanctum* was introduced by the sound of an ominous, eerie, creaking door opening, accompanied by a spooky melodramatic organ score; this sound began and closed the broadcasts. Actually, the door-sound was made by a rusty desk chair that was slid around on the studio floor. When a helpful technician cleaned and oiled the chair, panic ensued and the Foley artist responsible for making the sound needed to vocalize the frightening squeak. I still recall, as a child, being terrified by that squeaky squealing door-sound; even today it causes me associated physical chills.

Foley sound effects create and perform everyday sounds that are used today in TV, movies and video-games; these highlight the action and make sure we believe it is happening. The sounds are made in a recording studio during post-production, and the goal is to make real and believable the activities we are seeing, such as fist fights, glass breaking, drinking, footsteps, wind and rain storms, keys jingling, etc. Foley artists are quite clever: a low-level fire burning sound can be produced by crinkling cellophane, a potato chip bag, or steel wool; the sound of running horses is created by pounding half-coconuts stuffed with cloth on dirt or sod; the sound of footsteps on snow is actually corn starch squeezed in a leather pouch; breaking bones are heard by snapping celery or carrots; a smashed head is heard by smashing pumpkins or watermelons, and endlessly on and on. We see a person knocking on a door, but we need to hear it too in order to believe it is really happening and we are there.

Music in opera and musical theater and ballet is central to creating character, building emotion, and advancing the story line. In film/TV/video games/commercials, music also plays a central part in creating and sustaining suspense, and identifying a character. Certain popular operatic arias, such as “Nessum Dorma” from Puccini’s *Turandot*, or the “Ride of the Valkyries” from Wagner’s *Die Walküre* are used over and over in movies to express bombastic and hyper-stylized emotionality. The American composer John Williams has given us memorable movie soundtracks that conjure mood and images. Think of the two-note theme of the shark in the movie *Jaws*. We rarely see the shark, but all we need to become totally frightened is to hear the musical theme. The Star Wars movies’ music provides us with a number of instantly recognizable musical themes. I would bet you can reproduce the Emperor’s March right now, just thinking about the movies.

Our focus may be on what we are seeing, but to be believed, we have to hear it happening. Our minds are eager to be manipulated into hearing and seeing and feeling the world that the art is creating.

Culture, expectations, prior experiences, and training contribute to forming our mental models

In the prior segments, we discussed the disconnect between the original sonic signal – the sound wave – and the sound we hear in our minds. We also discussed brain plasticity – the ability of the brain to change through experience. And, we discussed the idea that we use mental models of the auditory world to compare with the sonic signals we are receiving in order to achieve cognitive recognition and understanding of sound. We can put together these three ideas to understand the impact of culture, expectations, exposure, prior experience, and training in creating meaning for the sonic experience.

(1) Personal expectations and cultural training were central to the critical response of Boston’s musical and social elite in 1900 when the new Symphony Hall was first experienced. It was the first auditorium in the world to be designed in accordance with

scientifically derived acoustical principles. Today it is considered to be one of the top three halls in the world for hearing classical music, and the best in the United States. However, in the first few years of its existence, the Symphony Hall acoustic experience was disliked. Reviewers, musicians and listeners kept up a steady stream of criticism of the new hall's aural personality, quite different from that of the Boston Music Hall, the prior home of the Boston Symphony Orchestra.[15] If the acoustics in Symphony Hall is considered outstanding, making the sonic signature world-famous today, why was it denigrated over 100 years ago? What changed: the physical Hall, or the minds of the people listening?

Often, "different" is met with a sense of strangeness, and humans don't tend to like that which is new and strange. Mostly we tend not to like surprises that upset our sense of what life should be like. This was the case with Symphony Hall: the patrons' and critics' mental model of good sound did not match the actual sonic experience of the new space. Their response was that the experience must be wrong. Change is difficult to process. Change can bring feelings of danger; it can be a harbinger of something which we are unprepared to deal with. We feel disoriented and potentially might react with denial or annoyance.

However, change is part of human life, and our brains have a built-in way to address the new and process it effectively. The brain is sufficiently plastic that we can adjust our internal mental models of "the good" fairly quickly. After only a few exposures, the human brain begins to rewire itself. Neurons that are used frequently develop stronger connections among themselves – LTP, and neurons that are less used experience increasingly less connection – LTD. Eventually the unused neurons are pruned to make room for new neuronal creation. Ultimately, we accept the new, the uncomfortable becomes comfortable, and eventually it is judged to be good.

All of this cognitive processing occurs below the level of consciousness. Our hidden biases shape what we hear without our realizing it.[16] At first, the early listeners of the new Symphony Hall were biased against the sound that was so different from what they were used to. Within a few short years of exposure, they became used to what they were hearing, accepted it, and became biased in its favor. Exposure over time to the new Symphony Hall sound produced a reset of listeners' sensory and artistic expectations. And with new expectations came a new type of evaluation. Eventually listeners came to terms with the new normal. Audiences and musicians alike became familiar with the new hall's acoustics and could appreciate the bright, complex sound experienced in the new space.[17] Today that sound matches our mental model, and is now the model to be matched by other classical music venues.

Humans have a remarkable capacity to find a new normal. Mental resiliency is a significant tool in the human survival toolbox.

(2) In the early to mid-20th century, the acoustically "dead" sound of recorded and broadcast acoustics became the norm for quality sound. This acoustic signature sound was the result of several considerations. The recording and broadcasting technology of that time was still fairly primitive. Using close microphones allowed for the depression of noisy ambient sounds and the intensification of the crucial original musical sounds. "Precision" recorded sound was strongly marketed, which meant recordings that suppressed noise, distortion and spatial acoustics, and at the same time amplified the music. Any extra sound other than that made directly by a musician was considered noise.

This was the era of machines inside and outside of the home. Not only did anything machine-based such as recorded music gain favor, but the cacophony of progress itself became a noise problem. The optimal recorded soundscape came to be defined as emphasizing foreground sound against a background of silence. Performance spaces like the Hollywood Bowl, which was an open-air space with neither reflecting surfaces nor reverberation, and Radio City Music Hall, which was purposely designed so as to suppress sound to allow for pure electroacoustic augmented sound transmission now became the desired norm. The preferred sound was now similar to the sonic experience of listening to music on the radio in your own small, relatively non-reverberant living room.

Radio recording studios were built purposely with minimal reverberation because of the above considerations, and also because these studios accommodated many different kinds of sonic performances. For instrumental music, the recording process required the close-miking of each musician that captured only direct sound and not reverberation or spatial acoustics. This produced the signature recorded dry sound. In addition, it was important politically and economically for the privately owned broadcast networks to ensure that there was a noticeable difference between hearing live broadcasts and listening to studio/recorded broadcasts. The differences were not only in hearing the ambient crowd noise in live broadcasting, but also in hearing the local performance space's acoustics, including reverberation. Technology and economics combined drove the mental model.

The most popular classical music broadcasts of the time were by Arturo Toscanini (the best-known classical music conductor of his time) conducting the NBC Symphony Orchestra from a large, specially-built studio. These broadcasts were recorded in a reverberation-minimal space using close microphones, which was preferred by Toscanini who favored such a crisp non-ornamented sound. A single radio broadcast by Arturo Toscanini conducting the NBC

Symphony Orchestra in the 1930s reached more listeners than all of those individuals who attended live performances by the New York Philharmonic in its first 90 years.[18] Add to this the exposure gained by two decades of Toscanini's recordings produced in their radio performance studio, and it becomes clear how this new unornamented, drier acoustic sound became the desired public standard, the mental model.

As technology, economics and social ways changed, the pendulum swung back to appreciating the aesthetic value of reverberation and spatial acoustics. Recordings of live performances of modern popular music in real-life reverberant spaces became popular, and technology for adding reverberation to recordings developed. Today's mental model of quality recordings in music and in television, film and radio now requires spatial acoustics and reverberation. To us today, a dry recording sounds artificial and old-fashioned.[19]

Conclusion

If our intention is to better understand the human experience of hearing and listening to sound, we must view the totality of the experience: the original sonic stimulus or sound wave, the modification of this sonic wave by the physical environment within which it is generated, the internal-external physical/social/cultural and psychological environment within which the human sensory system perceives the auditory stimulus, and the ways in which the human mind processes the sonic stimulus to achieve meaning.

In this paper we presented models of the mind as listeners experience sound and process it to produce meaning. A model helps us understand the meaning of a particular event to a particular perceiver at a particular moment in time and space. Understanding the way humans process sound to achieve meaning is still very much a work in progress.

We suggest that heuristic, phenomenological, perception/neuropsychological based frameworks are the best approach, at this point in time, to providing insights that expand our understanding of the human hearing experience. We believe this approach produces better insights than the scientific approach where findings are the outcome of carefully constructed research projects, and results are couched in numbers, formulae, charts, and graphs – all expressing research findings with great clarity, thus producing a sense of confidence that something of significance has been learned. Unfortunately, sound and hearing are not easily amenable to this kind of research structure. (1) Reductionism gives us the illusion that we know something, when what we know is a small unit of laboratory-based experience with unknown significance. (2) The mental processing of sound is all internal and very difficult to access. (3) The experiencing of sound differs from person to person, and from event to event. (4) Audio signals are very difficult to measure in such a way as to produce data that reveals the meaning of the signal to the perceiver. (5) Cause/effect relationships are notoriously difficult to pin down when considering sonic sensory processing, because so much is unknown about the nature of the stimulus and about the individual doing the processing.

On the other hand, cognitive-science based models of the human experience more naturally approach the felt experience of human beings when dealing with the intangible, transient, endlessly variable activity of hearing sound. When it comes to hearing, there is often very little objective truth. The mind determines what we hear and what the meaning is. We need to focus on the mind and its multi-layered inter-acting processes if we want to understand hearing.

Notes

[1] Eagleman, D. *Incognito*: *The Secret Lives of the Brain.* Vintage Books. New York. 2011, pages 1-19
[2] Ibid, pages 20-54
[3] Lømo T. "Frequency potentiation of excitatory synaptic activity in the dentate area of the hippocampal formation*". Acta Physiologica Scandinavica. 68 (Suppl 277): 12.* 1966
Hebb DO (1949)*. Organization of Behavior: a Neuropsychological Theory.* New York: John Wiley*.*
Patihis L. *"*The historical significance of the discovery of long-term potentiation: an overview and evaluation for nonexperts*". American Journal of Psychology. 131 (3): 369–80.* DOI: 10.5406/amerjpsyc.131.3.0369*.* (October 2018)
[4] Eagleman, op cit, pages 57- 59
[5] Kuhl, P.K, F.M Tsao, H.M. Liu. "Foreign language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning." *Proceedings of the National Academy of Science.* USA 100:9096–9101. 2003
Lytle, S.R., A. Garcia-Sierra, P.K.Kuhl. 2018. "Two are better than one: Infant Language learning from video improves in the presence of peers". *PNAS.* October 2, 2018. 115 (40) 9859-9866; first published October 1, 2018; DOI: 10.1073/pnas.1611621115
[6] Kish, Daniel and Jo Hook, *Echolocation and FlashSonar: Guidebook to aid in mastery of echolocation skills*, American Printing House for the Blind, Louisville, KY, 2021
[7] Lin, F.R., and L.Ferrucci. 2012. "Hearing Loss and Falls Among Older Adults in the United States". *American Medical Association Archives of Internal Medicine* 172(4):369-371). 2012
[8] Asifa Majid, Seán G. Roberts, Ludy Cilissen, Karen Emmorey, Brenda Nicodemus, Lucinda O'Grady, Bencie Woll, Barbara LeLan, Hilário de Sousa, Brian L. Cansler, Shakila Shayan, Connie de Vos, Gunter Senft, N. J. Enfield, Rogayah A. Razak, Sebastian Fedden, Sylvia Tufvesson, Mark Dingemanse, Ozge Ozturk, Penelope Brown, Clair Hill, Olivier Le Guen, Vincent Hirtzel, Rik van Gijn, Mark A. Sicoli, Stephen C. Levinson. Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, 2018; 201720419 DOI: 10.1073/pnas.1720419115
[9] Eagleman, op cit, pages 48-49
[10] McGurk, H. and John MacDonald. Hearing Lips and Seeing Voices. *Nature,* volume 264, pages 746–748; December 23, 1976
[11] Bell, A. "Levels and Loops: the future of artificial intelligence and neuroscience". *Philosophical Transactions of the Royal Society of London. Series B*. 1999
[12] Eagleman, op cit, p 47
[13] Ibid, pages 38-44

[14] Schell, Jesse*. The Art of Game Design: A Book of Lenses.* 2nd Edition. CRC Press, Taylor and Francis Group. Boca Raton, FL. 2015
[15] http://ashtonallan.com; thesis; 201SymphonyHall PDF; *Acoustic Architecture Begins: Symphony Hall, Boston 1900; Part II Chapter 1*; pages 28-37; accessed online June 1 2021
[16] Eberhardt, J.L. *Biased.* Penguin Books, 2020
[17] Thompson, Emily. *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America 1900 to 1933*. MIT Press. Cambridge, MA. 2002
[18] Delong, Thomas. *The Mighty Music Box: The Golden Age of Musical Radio*. Amber Crest Books. 1980
[19] Blesser, B. and L-R. Salter. *Spaces Speak Are You Listening? Experiencing Aural Architecture.* MIT Press, Cambridge, MA. 2006, pages 103-123