Buenos Aires – 5 to 9 September, 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Plenary Lectures: Paper ICA2016-915

How the brain makes sense of complex auditory scenes

Barbara Shinn-Cunningham[(a)]

[(a)] Boston University, United States, shinn@bu.edu

Abstract

Everyday listening involves a complex interplay between the ear, which transduces sound energy into neural responses, and the brain, which makes sense of these inputs. Historically, research on the ear tended to ignore the fact that what we can perceive in sound depends on what task the brain is engaged by, while research on cortical processing of sound ignored the complexity and sophistication of how the ear works. This paper explores how everyday perceptual abilities depend jointly on how the ear encodes information (and individual differences in the fidelity with which it does so) and how attention and other state dependent variables change the information we perceive.

Keywords: psychoacoustics, cocktail party problem, hidden hearing loss, auditory attention,

How the brain makes sense of complex auditory scenes

1 Hearing research from an engineer's perspective

Historically, engineers were very involved in psychoacoustics, motivated by an interest in how to build speech communication systems (e.g., the telephone; [1]). The goal of much of this early work was to determine the smallest discriminable difference (the just-noticeable difference or JND) in various acoustic features of sound. By quantifying JNDs in different dimensions, one could determine what information had to be preserved in a transmitted speech signal in order to ensure there was no perceptible loss of quality. More generally, because communication bandwidth was restricted, researchers wanted to understand how different degradations in signal features affected speech intelligibility [2]. By quantifying how limits in spectro-temporal resolution affected speech understanding or musical quality, engineers could analyse how to trade off costs versus effectiveness, and make rational design choices. Such approaches are still used today to guide development of coding systems such as mpeg and other schemes [3, 4].

Critical concepts about how hearing operates came from this type of work—ideas that continue to be fundamental to how hearing researchers understand the auditory system [2]. For instance, the idea that any channel of information coming out of the cochlea has a best frequency as well as a "critical frequency band" [5], or range of acoustic frequencies to which it responds, grew out of research like this. Similarly, this approach to hearing research provided insight into how sound intensity maps to perceived loudness, how interaural time and level differences affect perception of sound direction, how one sound masks both simultaneous and non-simultaneous sounds, and how important different parts of the acoustic spectrum are for conveying speech meaning [6-10].

While this kind of approach provides great insight into auditory processing, it typically assumes that a listener makes optimal use of all perceptible information in the acoustic signal. For instance, in binaural hearing, one influential model (the equalization and cancellation model) predicts a great deal about how binaural cues enhance a listener's ability to detect a pure-tone signal embedded in noise [11]. The model assumes that the listener perfectly combines noisy representations of the left and right ear signals, which accounts well for listener abilities in this simple task. However, a change in sound that may be perceptible in a simple experiment (where there listener knows exactly what to listen for and no other sound features vary) is often undetectable in a more complex context, like when a listener is not sure what to listen for or when many different sound features vary simultaneously (e.g., [12-14]).

The idea that information that is perceptible might not always be perceived, even when it is relevant for a task, can seem almost mysterious, given how well optimal processing models work in simple tasks—to the point that when listeners fail to perceive such information, they are said to suffer from "informational masking." The typical definition of informational masking is perceptual interference caused by distracting sound on perception of a "target" sound that cannot be explained by the information (or lack thereof) available in the pattern of firing in the auditory nerve" (e.g., see [15-17]). The use of this kind of blanket term underscores that the processes that lead to these "failures" of perception are not explained by standard psychoacoustics. The analysis techniques that allow one to relate early sensory coding to simple discrimination (JND) tasks, where processing is close to optimal, does not give good insight into how central, cognitive mechanisms impact performance.

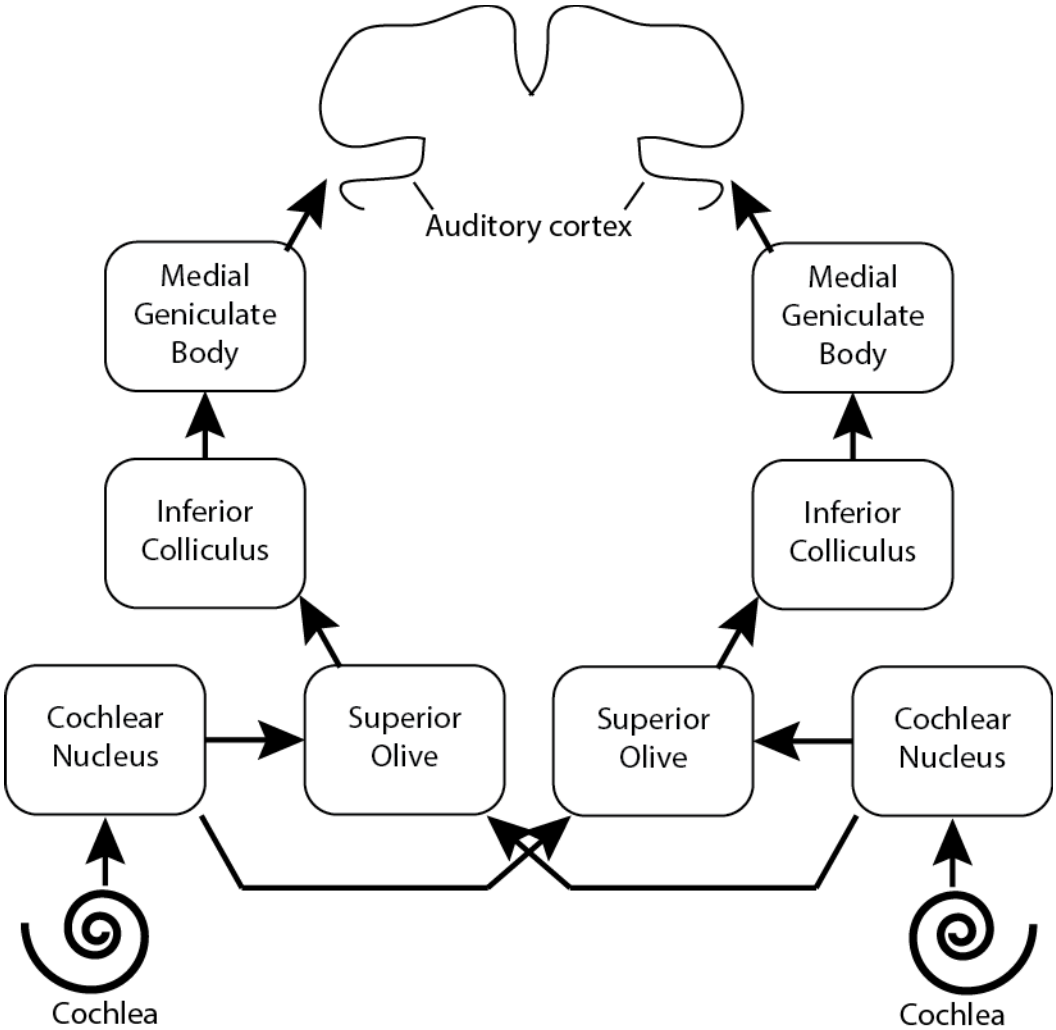

Figure 1: A typical block diagram of the ascending auditory system when drawn in the worldview of the traditional psychoacoustician. Note that the processing stream terminates in primary auditory cortex.

This kind of thinking can be understood by considering a typical diagram of the auditory system, drawn by a scientist from this lineage, shown in Figure 1. The different stages of processing from cochlea up to the inferior colliculus and even the medial geniculate body are shown; however, the processing pathway ends at the point where acoustic information arrives at the neocortex. If one were to take home a message from this kind of illustration, one would think that cortical processing has no effect on auditory perception.

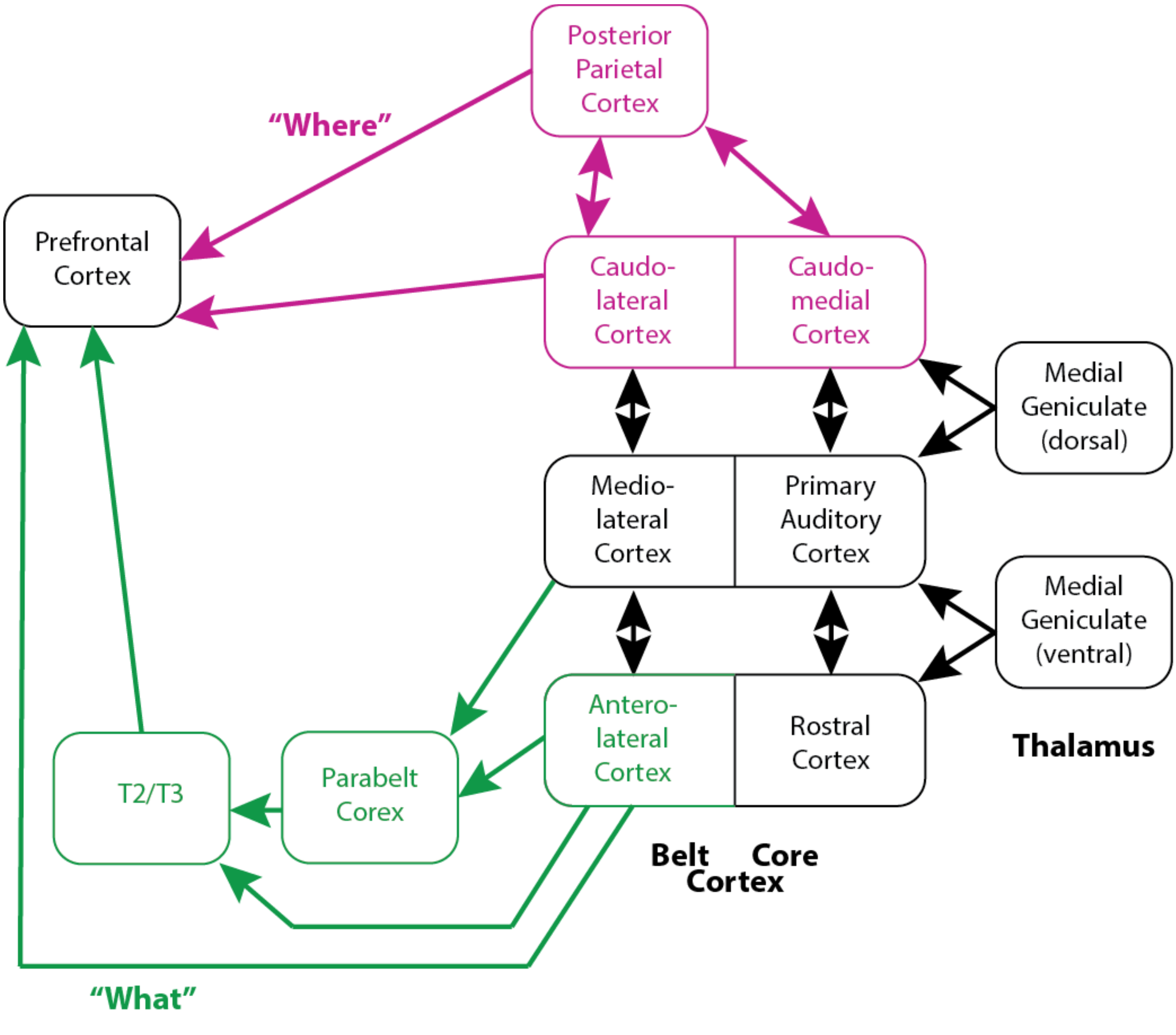

Figure 2: A typical block diagram of the auditory system when drawn in the worldview of the traditional cortical physiologist. Note that the inputs to the model come from the thalamus, without any consideration of processing in the brainstem. (Adapted from [18]).

2 Hearing research from a psychological perspective

In contrast to traditional psychophysics, traditional research in psychology and in cortical processing of sound often explicitly addressed how high-level processes like attention and memory impact (and interfere with) perceptual abilities. Hearing research growing out of this lineage includes auditory cognitive psychology, cognitive neuroscience, and cortical physiology. Such work considers, for instance, how different brain networks are engaged when focusing auditory spatial attention, storing and retrieving auditory information, processing language, and other high-level tasks [19-21].

This cortical-focused work was critical to developing insight into how cortical processing areas contribute to sound processing. However, this style of research often assumed that information was available in cortex in some relatively abstract form. Often, models of cortical processing did not consider how sound was encoded in subcortical portions of the auditory pathway, or how this encoding affected the more central processing that was being considered.

This view of the world can be captured by considering the block diagram shown in Figure 2, taken from a seminal paper on how acoustic information is processed to extract where sounds are ("where") versus "what" they are [18]. In this conception of the auditory processing pathway, information arrives at the sensory cortices—somehow. But the representation is abstract, and not tied very directly to the acoustic inputs. Once in cortex, however, abstract properties of the sound are extracted in different, parallel pathways, which link primary sensory auditory cortex in the temporal lobe with areas in the parietal and frontal lobes.

3 Viewing the auditory system holistically

The truth is that how sound is encoded in the subcortical portions of the auditory pathway have an enormous influence on what information can be processed in cortical networks. Yet, what we perceive and act upon depends on the processing in neocortex—what most people picture when they think of "the brain."

For instance, speech sounds that "normal" listeners perceive in full detail may not even be present in the neural responses in the brainstem of a listener suffering from elevated hearing thresholds. Understanding what sound energy gets represented in the cochlea and what does not determines the information that both "where" and "what" cortical pathways have available about a speech signal, and therefore how well this processing will work. This in turn will determine what the hearing-impaired listener will understand.

The problem is subtler than this, however. Recent evidence shows that animals and humans may suffer from hearing loss that is "hidden" to normal hearing screenings, even if they have normal hearing thresholds [22-24]. Specifically, auditory nerve fibers may die, degrading the representation of sound in the brainstem, even though listeners may be able to detect that the sound is present. Such deficits can cause real problems for listeners trying to communicate in everyday settings [25-27]. These communication deficits likely arise because the speech that they are trying to understand is not represented with good fidelity, which not only can interfere with understanding directly, but can make it difficult to focus selective attention on the desired speech signal and filter out competing sounds that interfere with speech understanding.

On the other hand, listeners *must* selectively focus attention on sounds of interest and filter out unwanted sounds in order to make sense of the source of interest [28, 29]. Cortical control networks literally change what information the listener processing, throwing out information in order to enable “selective attention” [30, 31]. If listeners filter out the information that matters for a task, they will not make use of the information, even though it is available at the auditory nerve. Similarly, if they cannot filter out unimportant information, they will be confused and not be able to parse the source that they are interested in. This is why understanding cortical processing is critical to understanding how listeners communicate in everyday settings.

Luckily, many researchers now are taking a more holistic view, considering how the ear and the brain interact to govern hearing. In my own laboratory, we attack the problem of understanding how listeners process sound in everyday settings by considering jointly both sensory coding in the subcortical portions of the auditory pathway, the brain networks that control perception and action in cortex, and how they interact to affect perception.

4 Conclusions

How we communicate in everyday settings depends on how sound is encoded subcortically, what information is gated through to the cortex, and how that information is then processed in the cortex. Interactions between the ear and the brain govern our perception. Understanding these interactions is the only way to truly understand everyday communication.

Acknowledgments

This work was funded in part by grants from the National Institute on Deafness and Other Communication Disorders, the National Science Foundation the Oticon Foundation, and the Coulter Foundation.

References

[1] Fletcher, H. and R.H. Galt, The perception of speech and its relation to telephony. Journal of the Acoustical Society of America, 1950. 22: p. 89-151.

[2] Allen, J.B., Harvey Fletcher's role in the creation of communication acoustics. J Acoust Soc Am, 1996. 99(4 Pt 1): p. 1825-39.

[3] Le Gall, D., MPEG: A video compression standard for multimedia applications. Communications of the ACM, 1991. 34(4): p. 46-58.

[4] Huber, R. and B. Kollmeier, PEMO-Q&# 8212; A new method for objective audio quality assessment using a model of auditory perception. IEEE Transactions on Audio, Speech, and Language Processing, 2006. 14(6): p. 1902-1911.

[5] Greenwood, D.D., Auditory masking and the critical band. The journal of the acoustical society of America, 1961. 33(4): p. 484-502.

[6] Zwicker, E. and B. Scharf, A Model of Loudness Summation. Psychol Rev, 1965. 72: p. 3-26.

[7] Searle, C.L., L.D. Braida, M.F. Davis, and H.S. Colburn, Model for auditory localization. Journal of the Acoustical Society of America, 1976. 60: p. 1164-1175.

[8] Ruotolo, B.R., J. Stern, Richard M., and H.S. Colburn, Discrimination of symmetric time-intensity traded binaural stimuli. Journal of the Acoustical Society of America, 1979. 66(6): p. 1733-1737.

[9] Kryter, K.D., Methods for the calculation and use of the articulation index. The Journal of the Acoustical Society of America, 1962. 34(11): p. 1689-1697.

[10] ANSI, American national standard methods for the calculation of the Articulation Index, S3.5. 1969, New York: American National Standards Institute, Inc.

[11] Durlach, N.I., Equalization and Cancellation Theory of Binaural Masking□Level Differences. The Journal of the Acoustical Society of America, 1963. 35(8): p. 1206-1218.

[12] Spiegel, M.F., M.C. Picardi, and D.M. Green, Signal and masker uncertainty in intensity discrimination. J Acoust Soc Am, 1981. 70(4): p. 1015-9.

[13] Watson, C.S., W.J. Kelly, and H.W. Wroton, Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. J Acoust Soc Am, 1976. 60(5): p. 1176-86.

[14] Melara, R.D. and L.E. Marks, Interaction among auditory dimensions: Timbre, pitch, and loudness. Perception & psychophysics, 1990. 48(2): p. 169-178.

[15] Brungart, D.S., Informational and energetic masking effects in the perception of two simultaneous talkers. J Acoust Soc Am, 2001. 109(3): p. 1101-9.

[16] Freyman, R.L., U. Balakrishnan, and K.S. Helfer, Spatial release from informational masking in speech recognition. Journal of the Acoustical Society of America, 2001. 109(5 Pt 1): p. 2112-22.

[17] Kidd, G., C.R. Mason, P.S. Deliwala, W.S. Woods, and H.S. Colburn, Reducing informational masking by sound segregation. Journal of the Acoustical Society of America, 1994. 95(6): p. 3475-3480.

[18] Rauschecker, J.P. and B. Tian, Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proceedings of the National Academy of Science, 2000. 97(22): p. 11800-6.

[19] Wu, C.-T., D. Weissman, K. Roberts, and M. Woldorff, The neural circuitry underlying the executive control of auditory spatial attention. Brain research, 2007. 1134: p. 187-198.

[20] Bodner, M., J. Kroger, and J.M. Fuster, Auditory memory cells in dorsolateral prefrontal cortex. Neuroreport, 1996. 7(12): p. 1905-1908.

[21] Catani, M. and D.K. Jones, Perisylvian language networks of the human brain. Annals of neurology, 2005. 57(1): p. 8-16.

[22] Schaette, R. and D. McAlpine, Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. J Neurosci, 2011. 31(38): p. 13452-7.

[23] Plack, C.J., D. Barker, and G. Prendergast, Perceptual consequences of "hidden" hearing loss. Trends Hear, 2014. 18.

[24] Bharadwaj, H.M., S. Masud, G. Mehraei, S. Verhulst, and B.G. Shinn-Cunningham, Individual differences reveal correlates of hidden hearing deficits. Journal of Neuroscience, 2015. 35: p. 2161-2172.

[25] Ruggles, D., H. Bharadwaj, and B.G. Shinn-Cunningham, Why middle-aged listeners have trouble hearing in everyday settings. Curr Biol, 2012. 22(15): p. 1417-22.

[26] Ruggles, D., H. Bharadwaj, and B.G. Shinn-Cunningham, Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. Proc Natl Acad Sci U S A, 2011. 108(37): p. 15516-21.

[27] Ruggles, D. and B. Shinn-Cunningham, Spatial selective auditory attention in the presence of reverberant energy: individual differences in normal-hearing listeners. J Assoc Res Otolaryngol, 2011. 12(3): p. 395-405.

[28] Shinn-Cunningham, B.G. and V. Best, Selective attention in normal and impaired hearing. Trends in Amplification, 2008. 12(4): p. 283-99.

[29] Shinn-Cunningham, B.G., Object-based auditory and visual attention. Trends Cogn Sci, 2008. 12(5): p. 182-6.

[30] Choi, I., L. Wang, H. Bharadwaj, and B. Shinn-Cunningham, Individual differences in attentional modulation of cortical responses correlate with selective attention performance. Hear Res, 2014. 314: p. 10-9.

[31] Eramudugolla, R., D.R. Irvine, K.I. McAnally, R.L. Martin, and J.B. Mattingley, Directed attention eliminates 'change deafness' in complex auditory scenes. Curr Biol, 2005. 15(12): p. 1108-13.