



Using Neural Circuit Interrogation in Rodents to Unravel Human Speech Decoding

Demetrios Neophytou¹ and Hysell V. Oviedo^{1,2*}

¹Biology Department, The City College of New York, New York, NY, United States, ²CUNY Graduate Center, New York, NY, United States

The neural circuits responsible for social communication are among the least understood in the brain. Human studies have made great progress in advancing our understanding of the global computations required for processing speech, and animal models offer the opportunity to discover evolutionarily conserved mechanisms for decoding these signals. In this review article, we describe some of the most well-established speech decoding computations from human studies and describe animal research designed to reveal potential circuit mechanisms underlying these processes. Human and animal brains must perform the challenging tasks of rapidly recognizing, categorizing, and assigning communicative importance to sounds in a noisy environment. The instructions to these functions are found in the precise connections neurons make with one another. Therefore, identifying circuit-motifs in the auditory cortices and linking them to communicative functions is pivotal. We review recent advances in human recordings that have revealed the most basic unit of speech decoded by neurons is a phoneme, and consider circuit-mapping studies in rodents that have shown potential connectivity schemes to achieve this. Finally, we discuss other potentially important processing features in humans like lateralization, sensitivity to fine temporal features, and hierarchical processing. The goal is for animal studies to investigate neurophysiological and anatomical pathways responsible for establishing behavioral phenotypes that are shared between humans and animals. This can be accomplished by establishing cell types, connectivity patterns, genetic pathways and critical periods that are relevant in the development and function of social communication.

OPEN ACCESS

Edited by:

Gordon M. Shepherd,
Yale University, United States

Reviewed by:

Daniel Llano,
University of Illinois at
Urbana-Champaign, United States
Santiago Jaramillo,
University of Oregon, United States

*Correspondence:

Hysell V. Oviedo
hoviedo@ccny.cuny.edu

Received: 09 October 2019

Accepted: 09 January 2020

Published: 30 January 2020

Citation:

Neophytou D and Oviedo HV
(2020) Using Neural Circuit
Interrogation in Rodents to Unravel
Human Speech Decoding.
Front. Neural Circuits 14:2.
doi: 10.3389/fncir.2020.00002

Keywords: speech-brain, auditory cortex (AC), animal models-rodent, cortical circuit, temporal processing and spectral processing

INTRODUCTION

During the Yugoslav wars of the 1990s ordering coffee could get one killed. In a Croatian coffee shop, the “Kava” pronunciation was associated with being a Croatian Catholic and led to the purchase of coffee without incident. But the “Kahva” pronunciation was associated with Bosnian Muslims and could potentially result in a bullet to the head (Dragojevic et al., 2015). This is not an isolated example: from Ebonics to the interpretation of the Constitution of the United States, language is of paramount importance to humans. Whether or not one believes that language

fundamentally constrains how an individual conceptualizes the world (Boroditsky, 2011), it is clear that it at least requires developmental programs that wire specific neural circuits. In this review article, we will briefly explore some widely-recognized computations the human brain performs to decode social communication and describe the animal studies designed to dissect potential circuit mechanisms that could serve as a blueprint to understand these processes.

Systems in the brain undergo time-limited periods of postnatal plasticity where synaptic connectivity adapts to the demands of the organism's environment. For instance, in the rat primary Auditory Cortex (A1), frequency representation (tonotopy) develops shortly after hearing onset in a time window between postnatal days 9–28. Rearing pups in white noise during this period can irreparably degrade spectral tuning (Zhang et al., 2002). Similarly, unnatural auditory experiences due to disease in early human infancy lead to degraded language abilities (Centers for Disease and Prevention, 2010). The implication of these findings is that exposure to the statistical structure of social calls provides the training signals that guide connectivity (Rauschecker, 1999; Levy et al., 2019). During critical periods, the brain relies on the statistical regularities of the auditory world (some sound sequences are more probable than others) to shape neural circuits in order to make detection and decoding of ethologically relevant signals faster in the future. In this review article, we focus mainly on studies from A1 (unless stated otherwise) because it is the first auditory area believed to represent perceptual features of sounds directly involved in decoding social communication (Wang et al., 1995; Nelken, 2008).

CIRCUIT FOUNDATION OF PHONEME DETECTION

There is general agreement regarding the basic computations a brain must perform to decode social communication. Mechanisms must exist to rapidly decode and bind phonetic elements and their temporal boundaries in a sequential and hierarchical manner. Studies in humans have made great advancements and fundamentally shape how we think about language processing (Yi et al., 2019). For decades it was unclear what was the most basic unit of speech that is decoded by neurons in A1. Recent studies using human cortical surface recordings (ECoG) from the superior temporal gyrus (STG) have discovered selectivity to phonemes in neural responses. Phonemes are the most basic units of speech sounds that have semantic meaning, and in the human STG responses are systematically organized by phonetic feature category (Mesgarani et al., 2014). Achieving this precise spectrotemporal tuning would require the wiring of phoneme detectors: neurons that preferentially respond to specific spectrotemporal features.

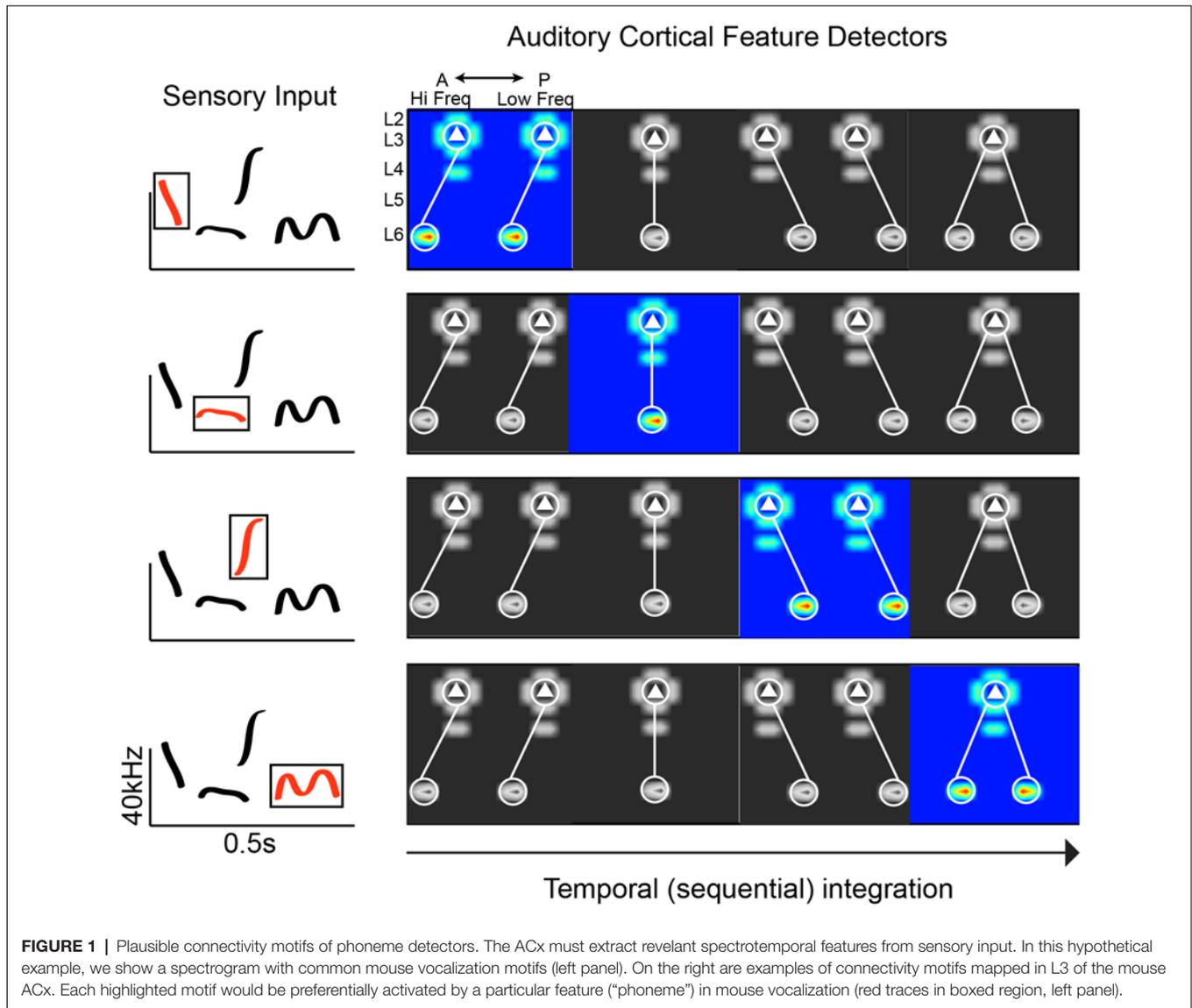
Given the limitations of experimental work in humans, we can more thoroughly interrogate potential circuit mechanisms in animal models that utilize social calls (e.g., birds and rodents). Circuit-mapping studies in the rodent A1 have begun to reveal connectivity schemes consistent with phoneme detection (Oviedo et al., 2010; Levy et al., 2019). Social calls

are spectrotemporally complex, requiring neurons to integrate across frequency channels. The superficial layers in A1 have been postulated to be a hub of spectral integration (Metherate et al., 2005; Winkowski and Kanold, 2013). Using circuit-mapping *in vitro*, we discovered along the tonotopic axis of the mouse A1 that Layer 3 (L3) excitatory neurons preferentially receive out-of-column inputs from neurons in L6 at higher frequency bands. The distance of this asymmetric L6 pathway ranges from 200 to 400 microns. Given the size of the mouse A1 (just over 1 mm along the horizontal plane covering 6 octaves; Guo et al., 2012), the L3 cells receiving spectrally shifted input from L6 neurons could be integrating across 1–3 octaves. This asymmetric pathway is absent along the isofrequency axis (orthogonal to tonotopy), where we found inputs that are columnar organized. Therefore, the functional anisotropy of A1 is directly reflected in the connectivity of neural circuits. We also found that differences in circuit-motifs correlate with differences in responsiveness to simple stimuli: L2 cells (which receive columnar input) show well-defined frequency tuning to pure tones, whereas L3 cells are largely unresponsive to pure tones (Oviedo et al., 2010). Though on average there are consistent circuit-motifs in A1, the correlation of input maps between pairs of neighboring neurons is very weak (0.3 within 100 microns, compared to 0.7 in the barrel cortex; Shepherd and Svoboda, 2005). Neighboring neurons can also respond very differently to the same stimulus (Hromádka et al., 2008; Oviedo et al., 2010). This could translate into a greater diversity of spectrotemporal decoders that maintain response flexibility (Figure 1). A relevant question is whether these circuit-motifs are unique to A1, or are also found elsewhere in the cortex. Compared to other cortical areas that have been mapped in detail, it does appear that A1 has a number of unique motifs (e.g., the aforementioned asymmetries). But the broader observation is that some circuit features are conserved across cortical areas (e.g., recurrent connections and a spectrum of columnar information flow), and some are unique to facilitate specific computational demands (Dantzker and Callaway, 2000; Bureau et al., 2006; Weiler et al., 2008).

REPRESENTATION OF THE MULTISCALE TEMPORAL FEATURES OF SPEECH

Social communication contains information at multiple time-scales, from fast fluctuations used for sound localization (~1 kHz), to slower modulations that correlate with word and syllabic structure (1–30 Hz). Different populations of cortical auditory neurons have shown distinct preferences to spectrotemporal modulation rates. There are neurons that represent amplitude modulations through the synchronized timing of their spikes (Wang, 2007; Wang et al., 2008). This synchrony is observed in response to slow temporal modulations (<5 Hz), which are associated with word and syllable boundaries (Arnal et al., 2015). Other populations represent amplitude modulations through a mean population firing rate. This rate code is observed in response to fast temporal modulations (<30 Hz) associated with phonemes (Arnal et al., 2015).

Moreover, the brain needs to rapidly decode multiplexed speech information before it is overwritten by new input



(Christiansen and Chater, 2016). This would ensure that we are able to comprehend the entirety of the speech signal and keep that information in a temporary buffer. One potential strategy is to recode the sensory input as it comes in, to capture each important component of the signal before a new input interferes (Brown et al., 2007). This results in a compressed representation of the information and minimizes the impact on echoic memory (sensory memory specific to audition; Pani, 2000). A candidate neural mechanism to accomplish this compression occurs in the form of synaptic depression at the thalamocortical synapse, where neural activity is transformed from explicitly representing the temporal structure of sound (i.e., subcortical input) to a rate code in A1 (Gao and Wehr, 2015).

To minimize the effects of echoic memory constraints aforementioned, the auditory system chunks the information into increasing levels of abstract representations of the sound during perception (Christiansen and Chater, 2016). It has

been proposed that this could be achieved with increasing temporal windows along a hierarchy to allow for an accurate representation of the chunks (see the section below for details; Hasson et al., 2008; Honey et al., 2012). A top-down predictive strategy has also been suggested to facilitate chunking: learned lexical information facilitates the chunking of the new incoming input. This anticipatory strategy tries to predict future input to allow for a more effective recoding of the information when it is eventually sensed (Christiansen and Chater, 2016).

Decoding social calls requires mechanisms to keep track of temporal landmarks: onsets and offsets. The majority of cells in the auditory system are responsive to sound onsets, and cells with excitatory responses to offsets have been widely described (Qin et al., 2007; Gao and Wehr, 2015). While the linguistic role of sound onsets is clear, the function of explicitly encoding temporal offsets has been more difficult to deduce. One accepted function is gap detection in continuous sounds. In particular,

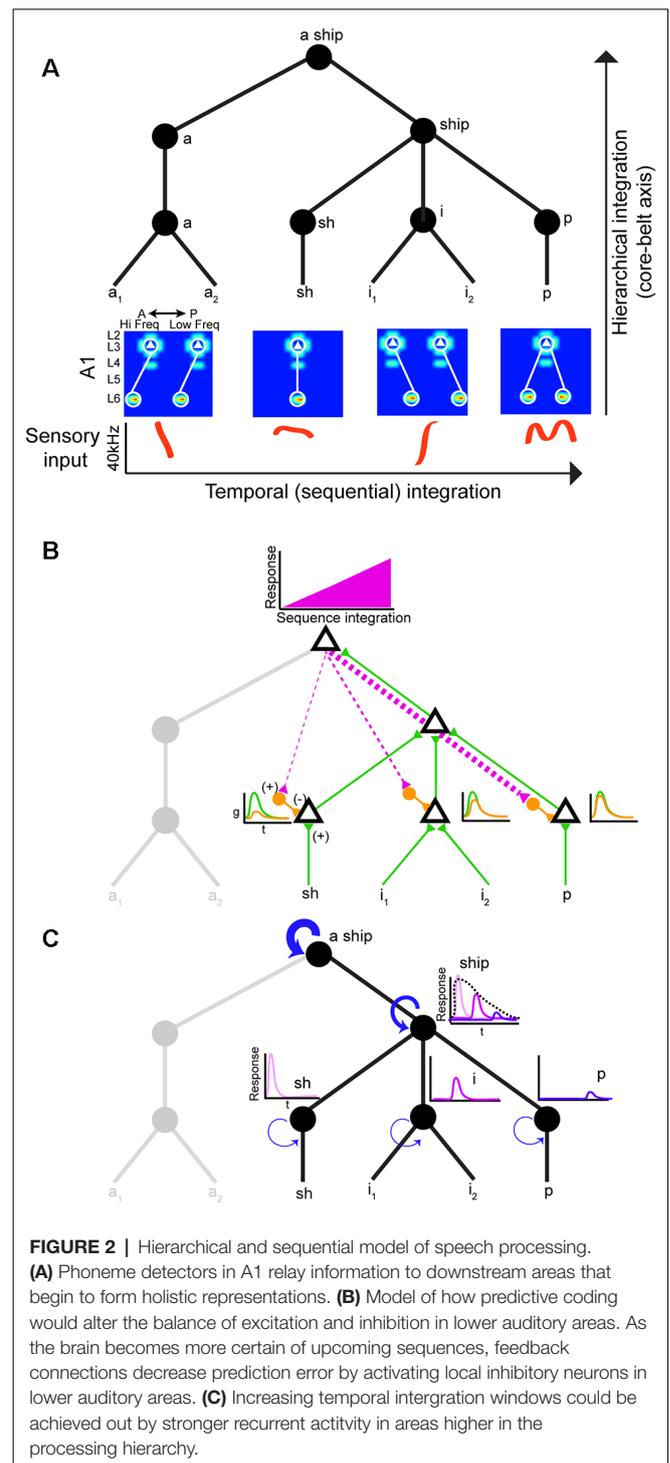
between-channel gaps (interruptions between distinct events) have been implicated in the discrimination of voice onsets in consonants (Kopp-Scheinflug et al., 2018). Distinct auditory pathways carry onset and offset information subcortically, but become integrated into A1 (Gao and Wehr, 2015).

Finally, social communication does not exist without context. Coarticulation is a microscopic example of context-dependency. In coarticulation, speech sounds influence the articulation of another in an anticipatory or carryover manner (Menzerath and de Lacerda, 1933; Ohala, 1993). At a macroscopic level listeners can predict upcoming words, based on the context, and incorporate them in ongoing linguistic processing (Van Berkum et al., 2005). Many context-dependent phenomena are due to syntactic regularities, which constrain the order of speech sounds. These regularities are exploited by the brain to reduce prediction errors and facilitate recognition (for potential mechanisms see section below; Leonard et al., 2015).

Animal studies have revealed potential network-level mechanisms underlying context dependency. An example is a sensitivity to rising or falling frequency sweeps, which is considered an important intonation cue in social communication. One candidate mechanism is the asymmetric organization of excitatory and inhibitory receptive fields in A1. There are populations of neurons with sideband inhibition that systematically changes its frequency bias along the tonotopic axis. This imparts directional selectivity (in frequency space) to neuronal responses (Zhang et al., 2003). Social communication circuits are also very sensitive to the fine temporal sequence of sounds. In a study of the rat A1, it was reported that neural networks could store estimates of tone order sequences for tens of seconds (Yaron et al., 2012). Nevertheless, neural mechanisms of this long-lasting temporal sensitivity remain unresolved (but see section below).

MEMORY DEMANDS FOR GROUPING AND BINDING SPEECH CONSTITUENTS

Neural circuits that decode social signals need a way to group sounds into meaningful categories in a sequential manner. Hierarchical and sequential processing are widely accepted models of linguistic representation (Uddén et al., 2019). There is psychophysical and empirical evidence for hierarchical processing in humans (Levelti, 1970; Pallier et al., 2011). For instance, there are areas for processing language that are sensitive to increases in syntactic complexity: they show a progressive increase in activation as the number of components with syntactic meaning increases. This parametric change in activation suggests the processing of smaller-sized constituents (Pallier et al., 2011). The neural architecture underlying hierarchical processing of social calls is not known, but the graph theory framework suggests nested tree structures where hierarchical and sequential distances can be clearly distinguished (Figure 2A; Uddén et al., 2019). The assumption is that areas higher in the hierarchy can represent longer sequences due to longer windows of temporal integration. Therefore, mechanisms that support echoic memory are required for accumulating and holding information online (Furl et al., 2011). The activity of



areas higher in the hierarchy would increase during sequence integration leading to increasingly stronger top-down input to areas lower in the hierarchy (Figure 2B). One potential mechanism to achieve predictive coding is inhibitory modulation from top-down input. These projections to lower areas can target local inhibitory neurons, which in turn reduce ongoing excitatory activity (Figure 2B; Pi et al., 2013). The prediction is that this increasing change in the balance of inhibition (top-down priors)

and excitation (bottom-up) would suppress prediction error signals (Friston and Kiebel, 2009). In rodents, there is evidence of hierarchical auditory processing: responses to vocalizations become more invariant between A1 and the suprarhinal auditory field (Carruthers et al., 2015). It is worth noting that information flow in the auditory stream is more nuanced: it is both parallel and hierarchical with extensive cross-talk between each auditory subfield (Hackett, 2011).

Echoic memory is necessary to concatenate phonemes, syllables, words, sentences, so we can make sense of language holistically in real-time. Social calls are dynamic and fleeting requiring mechanisms to maintain memory banks. There is evidence that A1 can support sequence memories on the order of tens of seconds (Yaron et al., 2012), but no circuit mechanism has been directly implicated. One way to implement longer temporal integration windows is recurrent connectivity (Figure 2C; Wang et al., 2018). However, it remains to be shown how increasing temporal integration windows are being implemented in a hierarchical fashion between different areas. Studies on the computational demands of working memory can offer some mechanistic insight into how echoic memory can be implemented by neural circuits. These studies have proposed neural networks with temporal dynamics that provide stable firing past the stimulus presentation (i.e., persistent activity). In addition to strong recurrent connectivity, several neural mechanisms have been identified that can lead to stable persistent activity. One is the activation of synapses with NMDA receptors. The long-time constant of these glutamatergic receptors can sustain stable persistent activity. Also, some form of negative feedback to maintain control of the firing rate in the presence of strong recurrent connections would be required (Wang, 1999). It would be fruitful to examine whether these circuit features change systematically along the auditory cortical hierarchy.

CIRCUIT FOUNDATION OF PARALLEL PROCESSING BETWEEN THE HEMISPHERES

Lateralization is a widespread strategy in the nervous system to assign distinct computational tasks to the left and right hemispheres. For processing social communication, it is believed that the left and right A1 specialize in processing information at different temporal scales. The left A1 is postulated to specialize in fast syntactic processes (identifying specific sequences in speech), and the right in slower temporal information (prosody and intonation; Arnal et al., 2015). Lateralized language processing in humans has been known for over a century (Broca, 1861; Long et al., 2016) and is crucial for normal function (Oertel et al., 2010; Cardinale et al., 2013). This division of labor has also been observed in other species including rodents (Ehret, 1987; Marlin et al., 2015), and in a recent study we began to elucidate the circuit mechanisms that could underlie lateralization by comparing the connectivity of the left and right A1 (Levy et al., 2019). We found significant hemispheric differences in the connectivity of L3 pyramidal cells in the mouse A1. In the left A1 projections from L6 to L3 arose out-of-column from

higher frequencies throughout most of the tonotopic axis. In contrast, the connectivity of the same pathway in the right A1 changed systematically with tonotopy, from lower to mixed to higher frequency bias. These distinct wiring schemes along the tonotopic axis suggest differences in spectrotemporal integration between the hemispheres.

To investigate the possible functional implications, we compared the responses to frequency sweeps of L3 neurons *in vivo*. In the left A1 there is a trend for L3 excitatory neurons to prefer downward sweeps regardless of their best frequency selectivity. This prevalent downward sweep selectivity could facilitate the left A1's responsiveness to ethologically relevant sequences such as downward pitch jumps, which are common components of mouse vocalizations (Holy and Guo, 2005). Pitch jumps spanning several octaves would activate a subset of these asymmetric integrators, and binding of their individual responses would occur at an auditory region downstream (Figure 1). On the other hand, in L3 of the right A1 we found that sweep direction selectivity changed along the tonotopic axis and was inversely related to best frequency tuning: cells with high-frequency selectivity preferred downward sweeps, whereas cells with low-frequency selectivity preferred upward sweeps and in between cells had mixed selectivity. Similar trends have been observed in the rat's right A1 (Zhang et al., 2003).

How does the brain dynamically control parallel, lateralized processes and information transfer without functional interference? Auditory sensory input stimulates both hemispheres and there are interhemispheric connections between the Auditory Cortices *via* the corpus callosum. One of the most debated questions is whether the impact of callosal projections is excitatory or inhibitory, which can determine the rules of information transfer and potential cooperativity. To a large extent the answer lies in the identity of the neurons targeted by callosal projections, but the targets and impact of interhemispheric projections on excitability remain unclear. A functional study in commissureless and normal primates suggested that the left A1 suppresses the right A1 (Poremba, 2006). Whereas hemispheric deactivation in cats suggested excitation is symmetric (Carrasco et al., 2013). In the rodent, almost every cortical layer (3, 4 and 5) is commissurally connected in a largely homotopic manner (Oviedo et al., 2010). Moreover, excitatory and inhibitory neuronal populations are commissural targets (Xiong et al., 2012). Hence, to answer the question of interhemispheric cooperation will require animal studies with very precise inactivation of specific neuronal classes in each hemisphere during vocal communication behavioral tasks.

CONCLUSIONS AND FUTURE DIRECTIONS

Human and animal studies should inform one another to gain a mechanistic understanding of how the brain decodes social communication. As recording techniques from humans improve, animal studies could serve as a guide to predict where (along the processing stream) and which signatures of specific auditory decoding operations should be examined. For instance, animal

studies could help to unravel proposed neural mechanisms of stream segregation (Lu et al., 2017). Moreover, experiments in animals may also be established as a complementary invasive platform to study mechanisms hypothesized from human observational work. Animal models can serve as a comparative template to determine the relative contribution of sensory experience and genetic programs in the development of social-communicative functions. Communication deficits are the most common disabilities in children, affecting 8–12% of preschoolers. The underlying pathological mechanism routinely involves the miswiring of the connections between neurons in the language centers of the brain. Animal models can help identify critical time points, neural elements, and molecular

targets to enable effective therapeutic interventions in human communication disorders.

AUTHOR CONTRIBUTIONS

HVO conceived and wrote most of the review article. DN helped write and revise the review article.

FUNDING

This project was funded by a Whitehall Foundation Research Grant (HVO), and an NSF Career Award to HVO (IOS-1652774).

REFERENCES

- Arnal, L. H., Poeppel, D., and Giraud, A. L. (2015). Temporal coding in the auditory cortex. *Handb. Clin. Neurol.* 129, 85–98. doi: 10.1016/B978-0-444-62630-1.00005-6
- Boroditsky, L. (2011). How language shapes thought. *Sci. Am.* 304, 62–65. doi: 10.1038/scientificamerican0211-62
- Broca, P. (1861). Remarques sur le sie'ge de la faculte' du langage articule', suivies d'une observation d'aphe'mie (perte de la parole). *Bull. Soc. Anat.* 6, 330–357.
- Brown, G. D., Neath, I., and Chater, N. (2007). A temporal ratio model of memory. *Psychol. Rev.* 114, 539–576. doi: 10.1037/0033-295X.114.3.539
- Bureau, I., von Saint Paul, F., and Svoboda, K. (2006). Interdigitated paralemniscal and lemniscal pathways in the mouse barrel cortex. *PLoS Biol.* 4:e382. doi: 10.1371/journal.pbio.0040382
- Cardinale, R. C., Shih, P., Fishman, I., Ford, L. M., and Muller, R. A. (2013). Pervasive rightward asymmetry shifts of functional networks in autism spectrum disorder. *JAMA Psychiatry* 70, 975–982. doi: 10.1001/jamapsychiatry.2013.382
- Carrasco, A., Brown, T. A., Kok, M. A., Chabot, N., Kral, A., and Lomber, S. G. (2013). Influence of core auditory cortical areas on acoustically evoked activity in contralateral primary auditory cortex. *J. Neurosci.* 33, 776–789. doi: 10.1523/jneurosci.1784-12.2013
- Carruthers, I. M., Laplagne, D. A., Jaegle, A., Briguglio, J. J., Mwilambwe-Tshilobo, L., R. Natan, G., et al. (2015). Emergence of invariant representation of vocalizations in the auditory cortex. *J. Neurophysiol.* 114, 2726–2740. doi: 10.1152/jn.00095.2015
- Centers for Disease and Prevention. (2010). Identifying infants with hearing loss - United States, 1999–2007. *MMWR Morb. Mortal. Wkly. Rep.* 59, 220–223. Available online at: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5908a2.html>. Accessed April 23, 2010.
- Christiansen, M. H., and Chater, N. (2016). The now-or-never bottleneck: a fundamental constraint on language. *Behav. Brain Sci.* 39:e62. doi: 10.1017/s0140525x1500031x
- Dantzker, J. L., and Callaway, E. M. (2000). Laminar sources of synaptic input to cortical inhibitory interneurons and pyramidal neurons. *Nat. Neurosci.* 3, 701–707. doi: 10.1038/76656
- Dragojevic, M., Gasiorek, J., and Giles, H. (2015). Communication Accommodation Theory. *International Encyclopedia of Interpersonal Communication*. doi: 10.1002/9781118540190.wbeic006
- Ehret, G. (1987). Left hemisphere advantage in the mouse brain for recognizing ultrasonic communication calls. *Nature* 325, 249–251. doi: 10.1038/325249a0
- Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300
- Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., and Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *NeuroImage* 54, 2267–2277. doi: 10.1016/j.neuroimage.2010.10.038
- Gao, X., and Wehr, M. (2015). A coding transformation for temporally structured sounds within auditory cortical neurons. *Neuron* 86, 292–303. doi: 10.1016/j.neuron.2015.03.004
- Guo, W., Chambers, A. R., Darrow, K. N., Hancock, K. E., Shinn-Cunningham, B. G., and Polley, D. B. (2012). Robustness of cortical topography across fields, laminae, anesthetic states and neurophysiological signal types. *J. Neurosci.* 32, 9159–9172. doi: 10.1523/jneurosci.0065-12.2012
- Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hear. Res.* 271, 133–146. doi: 10.1016/j.heares.2010.01.011
- Hasson, U., Yang, E., Vallines, I., D. Heeger, J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550. doi: 10.1523/jneurosci.5487-07.2008
- Holy, T. E., and Guo, Z. (2005). Ultrasonic songs of male mice. *PLoS Biol.* 3:e386. doi: 10.1371/journal.pbio.0030386
- Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., et al. (2012). Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* 76, 423–434. doi: 10.1016/j.neuron.2012.08.011
- Hromádka, T., Deweese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.* 6:e16. doi: 10.1371/journal.pbio.0060016
- Kopp-Scheinpflug, C., Sinclair, J. L., and Linden, J. F. (2018). When sound stops: offset responses in the auditory system. *Trends Neurosci.* 41, 712–728. doi: 10.1016/j.tins.2018.08.009
- Leonard, M. K., Bouchard, K. E., Tang, C., and Chang, E. F. (2015). Dynamic encoding of speech sequence probability in human temporal cortex. *J. Neurosci.* 35, 7203–7214. doi: 10.1523/jneurosci.4100-14.2015
- Levelt, W. J. M. (1970). Hierarchical chunking in sentence processing. *Percept. Psychophys.* 8, 99–103. doi: 10.3758/bf03210182
- Levy, R. B., Marquarding, T., Reid, A. P., Pun, C. M., Renier, N., and Oviedo, H. V. (2019). Circuit asymmetries underlie functional lateralization in the mouse auditory cortex. *Nat. Commun.* 10:2783. doi: 10.1038/s41467-019-10690-3
- Long, M. A., Katlowitz, K. A., Svirsky, M. A., Clary, R. C., Byun, T. M., Majaj, N., et al. (2016). Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* 89, 1187–1193. doi: 10.1016/j.neuron.2016.01.032
- Lu, K., Xu, Y., Yin, P., Oxenham, A. J., Fritz, J. B., and Shamma, S. A. (2017). Temporal coherence structure rapidly shapes neuronal interactions. *Nat. Commun.* 8:13900. doi: 10.1038/ncomms13900
- Marlin, B. J., Mitre, M., D'Amour, J. A., Chao, M. V., and Froemke, R. C. (2015). Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature* 520, 499–504. doi: 10.3389/fnbeh.2015.00311
- Menzerath, R., and de Lacerda A. (1933). *Koartikulation, Steuerung und Lautabgrenzung Eine Experimentelle Untersuchung*. Berlin: Dümmler.
- Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994
- Metherate, R., Kaur, S., Kawai, H., Lazar, R., Liang, K., and Rose, H. J. (2005). Spectral integration in auditory cortex: mechanisms and modulation. *Hear Res* 206, 146–158. doi: 10.1016/j.heares.2005.01.014
- Nelken, I. (2008). Processing of complex sounds in the auditory system. *Curr. Opin. Neurobiol.* 18, 413–417. doi: 10.1016/j.conb.2008.08.014
- Oertel, V., Knöchel, C., Rotarska-Jagiela, A., Schönmeier, R., Lindner, M., van de Ven, V., et al. (2010). Reduced laterality as a trait marker of schizophrenia-

- evidence from structural and functional neuroimaging. *J. Neurosci.* 30, 2289–2299. doi: 10.1523/jneurosci.4575-09.2010
- Ohala, J. J. (1993). Coarticulation and phonology. *Lang. Speech* 36, 155–170. doi: 10.1177/002383099303600303
- Oviedo, H. V., Bureau, I., Svoboda, K., and Zador, A. M. (2010). The functional asymmetry of auditory cortex is reflected in the organization of local cortical circuits. *Nat. Neurosci.* 13, 1413–1420. doi: 10.1038/nn.2659
- Pallier, C., Devauchelle, A. D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U S A* 108, 2522–2527. doi: 10.1073/pnas.1018711108
- Pani, J. R. (2000). Cognitive description and change blindness. *Vis. Cogn.* 7, 107–126. doi: 10.1080/135062800394711
- Pi, H. J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature* 503, 521–524. doi: 10.1038/nature12676
- Poremba, A. (2006). Auditory Processing and hemispheric specialization in non-human primates. *Cortex* 42, 87–89. doi: 10.1016/S0010-9452(08)70325-3
- Qin, L., Chimoto, S., Sakai, M., Wang, J., and Sato, Y. (2007). Comparison between offset and onset responses of primary auditory cortex ON-OFF neurons in awake cats. *J. Neurophysiol.* 97, 3421–3431. doi: 10.1152/jn.00184.2007
- Rauschecker, J. P. (1999). Auditory cortical plasticity: a comparison with other sensory systems. *Trends Neurosci.* 22, 74–80. doi: 10.1016/s0166-2236(98)01303-4
- Shepherd, G. M., and Svoboda, K. (2005). Laminar and columnar organization of ascending excitatory projections to layer 2/3 pyramidal neurons in rat barrel cortex. *J. Neurosci.* 25, 5670–5679. doi: 10.1523/JNEUROSCI.1173-05.2005
- Uddén, J., de Jesus Dias Martins, M., Zuidema, W., and Fitch, W. T. (2019). Hierarchical structure in sequence processing: how to measure it and determine its neural implementation. *Top. Cogn. Sci.* doi: 10.1111/tops.12442 [Epub ahead of print].
- Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 443–467. doi: 10.1037/0278-7393.31.3.443
- Wang, J., Narain, D., E. Hosseini, A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* 21, 102–110. doi: 10.1038/s41593-017-0028-6
- Wang, X. (2007). Neural coding strategies in auditory cortex. *Hear. Res.* 229, 81–93. doi: 10.1016/j.heares.2007.01.019
- Wang, X. J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J. Neurosci.* 19, 9587–9603. doi: 10.1523/jneurosci.19-21-09587.1999
- Wang, X., Merzenich, M. M., Beitel, R., and Schreiner, C. E. (1995). Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiol.* 74, 2685–2706. doi: 10.1152/jn.1995.74.6.2685
- Wang, X., Lu, T., Bendor, D., and Bartlett, E. (2008). Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience* 154, 294–303. doi: 10.1016/j.neuroscience.2008.03.065
- Weiler, N., Wood, L., Yu, J., S. Solla, A., and Shepherd, G. M. (2008). Top-down laminar organization of the excitatory network in motor cortex. *Nat. Neurosci.* 11, 360–366. doi: 10.1038/nn2049
- Winkowski, D. E., and Kanold, P. O. (2013). Laminar transformation of frequency organization in auditory cortex. *J. Neurosci.* 33, 1498–1508. doi: 10.1523/jneurosci.3101-12.2013
- Xiong, Q., Oviedo, H. V., Trotman, L. C., and Zador, A. M. (2012). PTEN regulation of local and long-range connections in mouse auditory cortex. *J. Neurosci.* 32, 1643–1652. doi: 10.1523/jneurosci.4480-11.2012
- Yaron, A., Hershenhoren, I., and Nelken, I. (2012). Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron* 76, 603–615. doi: 10.1016/j.neuron.2012.08.025
- Yi, H. G., Leonard, M. K., and Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron* 102, 1096–1110. doi: 10.1016/j.neuron.2019.04.023
- Zhang, L. I., Tan, A. Y., Schreiner, C. E., and Merzenich, M. M. (2003). Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature* 424, 201–205. doi: 10.1038/nature01796
- Zhang, L. I., Bao, S., and Merzenich, M. M. (2002). Disruption of primary auditory cortex by synchronous auditory inputs during a critical period. *Proc. Natl. Acad. Sci. U S A* 99, 2309–2314. doi: 10.1073/pnas.261707398

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Neophytou and Oviedo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.