SAGE reference

The SAGE Encyclopedia of Human Communication Sciences and Disorders

Neurophonetics

Neurophonetics studies neural bases of speech production and perception. Speech production starts with a proposition that is transformed into a verbal message by a series of pragmatic and language-specific rules and constraints, to be finally articulated with the help of motor control mechanisms and auditory feedback loop and monitoring processes. Speech perception involves auditory encoding of phonological features as a step in lexical access and comprehension of the message within the perception–memory–action loop. Just as the boundary between the more general fields of phonetics and linguistics is blurry, the area where neurophonetics and neurolinguistics meet is gray and some overlap is to be expected. There seems to be general agreement that neurophonetics covers acoustic, phonetic, and phonological aspects of speech processing to the level of lexical access. Neurophonetic research has found application in areas such as speech acquisition, clinical phonetics, and bilingualism. This entry provides an overview of the methodology used in neurophonetics, then explores the neural bases of speech processing, and finally addresses issues related to bilingualism and neurogenic speech disorders.

Methodology

Until the advancement of modern technology, *neurological cases* were the only source of information about possible neural bases of speech processing (e.g., aphasics, split-brain patients). In spite of its limitations (e.g., absence of premorbidity data, effects of comorbidity, lack of insight into relevance of locations other than the ones affected by the condition, inappropriateness of testing situations, varying diagnostic tools), this vast body of evidence is still valuable in studying the neurophysiological underpinnings of speech.

Groundbreaking data have been collected from *electrostimulation studies* that provide direct recordings of brain surface activity during various tasks in awake patients, but for obvious reasons, this method is reserved for medically justified uses. In transcranial magnetic stimulation, areas of interest on the brain surface can be stimulated through the skull by means of an electrical coil. Stimulation effects are very brief and timing is crucial. Although it has been used on healthy subjects, it may cause seizures and possibly affect memory.

Event-related potentials (ERP) are one of the most widely used noninvasive *electrophysiological methods*. It is a derivative of electroencephalography and is based on the fact that each event evokes electrical responses in the central nervous system that can be measured (after multiple repetitions of the same stimulus that are necessary in order to separate the signal from the noise). Phonetic and phonological processes typically have latencies up to 200 or 250 milliseconds (ms). Latency of approximately 50 ms (P50) is characteristic of processing place of articulation. Periodicity and vowel height have been linked with the latency of 100 ms (N100). The so-called mismatch negativity is a response to deviant stimuli. Deviant vowels in an oddball design typically elicit a frontocentral response with the latency of 100–250 ms. Such response is also found when an incoming stimulus conflicts with the top-down processing or with information from another modality and has therefore been considered as reaction to error. ERP has good temporal but poor spatial resolution, and it is not suitable for studying perception of continuous speech. Magnetoencephalography is based on electromagnetic fields that form during neural activity and displays deflections comparable to those of ERPs. It also requires multiple presentations of the same stimulus but is less affected by surrounding tissue than ERP.

Neuroimaging methods provide information about metabolic activity in reaction to a stimulus. Functional magnetic resonance imaging, based on blood-oxygen-level-dependent response, positron emission tomography, based on positron-emitting tracers after injection/inhalation of a radioactive isotope, and functional near-infrared spectroscopy, based on the brain's permeability to infrared rays, are the most common ones. Functional near-infrared spectroscopy is applicable to populations less suitable for functional magnetic resonance imaging or positron emission tomography (e.g., infants). These methods share good spatial but poor temporal resolution. Another problem common to all these techniques is that increased blood flow is present when both excitatory and inhibitory neurons are active. Diffusion-weighted magnetic resonance imaging and the related method of diffusion tensor imaging, based on the diffusion of water molecules in tissue, are used for tracking

white matter fiber bundles, which is relevant in connectivity research because it reveals not only connections among different regions of the brain but also their direction. Limitations of all of these methods include small sample size, artifacts, resolution issues, task restrictions, and neglecting nonstudied parts.

One of the most widely used *behavioral methods* is dichotic listening. In this method, two different auditory stimuli are presented to the left and right ears simultaneously and response time and/or accuracy are measured. It provides indirect indication of which hemisphere of the brain is more involved in processing various types of auditory information. Right ear advantage is typically found for speech stimuli, which is taken as evidence of greater involvement of the left hemisphere (LH). Contrary to this, nonspeech stimuli result in left ear advantage, suggesting greater activity of the right hemisphere (RH). Right ear advantage for syllables and left ear advantage for music stimuli have been found in 4-month-old babies. This method is closely related to divided visual field studies, which have revealed right visual field advantage for language stimuli and left visual field advantage for nonlanguage stimuli, indicating LH and RH dominance, respectively. Both methods are susceptible to attention manipulation/control and are suitable for very short stimuli.

Naturally, more *peripheral* methods—those that measure the outcome of actual perception/production rather than visualizing the underlying (central) neural processes (e.g., perceptual assessment, auditory and acoustic analysis, electromyography, glottography, ultrasound, electropalatography)—are widely used for studying speech production and perception.

The best and most reliable results are obtained by combining different methods and utilizing their individual strengths to yield converging evidence.

Neural Bases of Speech Processing

Speech production is predominantly an LH activity, especially in nonautomatic speech and more complex production (e.g., reciting poetry). Automatic speech (e.g., listing days of the week, swear words) is bilaterally or RH represented. In addition to areas essential for speech production (i.e., Broca's area, primary motor area and supplementary motor area, insula), basal ganglia and cerebellum as well as somatosensory and auditory feedback mechanisms constitute the intricate network involved in the process.

Speech perception seems to be less clearly left-lateralized and more widely distributed than production. This claim is partly based on the observation that in LH-injured patients, perception recovers faster than production. Bilateral activation of premotor areas has been recorded in tasks of syllable detection, subvocal repetition, and paying attention without overt response—which again suggests activation of motor programs consistent with the motor theory of speech perception and reveals neural bases of the phonological/articulatory loop in Alan Baddeley's working-memory model (involving numerous bilateral speech production and perception areas including Brodmann area [BA] 44, 33, 42, 40, supplementary motor area, insulas, cerebellum, and BA 1, 2, 3, 4).

LH is generally considered dominant for processing CV syllables (e.g., /pa/), but some studies report no significant asymmetry. With respect to individual *segments*, LH is predominantly activated during processing of stops and fricatives whose perception relies on transitions (e.g., /ð/, /Ө/) as opposed to those that are characterized by noise cues (e.g., /s/, /ʃ/, /z/, /ʒ/). However, there are indications that place of articulation and voicing may be represented in LH and RH, respectively. Semivowels (e.g., /w/) and liquids (e.g., /l/) are less clearly left-lateralized than consonants, but still more so than vowels, that exhibit bilateral activity or inconsistent lateralization. However, in difficult listening conditions (e.g., noise, distortion), LH clearly takes over. Segmentation task activates LH inferior frontal gyrus/inferior frontal cortex. Sensitivity to phonetic contrasts is represented in inferior frontal cortex and in temporal regions. Encoding new speech categories activates bilateral middle frontal gyrus, indicating its involvement in learning sounds of nonnative, foreign language (L2). Some electrophysiological data suggest that RH is involved in segmentation and prosodic timing at syllabic

level and in preprocessing of speech signals (at a prelinguistic level).

Positron emission tomography and functional magnetic resonance imaging studies have identified increased activity in superior temporal gyrus (STG) in response to frequency-modulated sounds in contrast to noise and in the area between STG and middle temporal gyrus in response to speech sounds in contrast to nonspeech stimuli. Both findings are taken as evidence that these areas are involved in spectrotemporal processing inherent in speech perception. Greater activity is usually recorded in the LH, but RH is active as well. This process is sublexical because it is found in words and pseudowords alike. The *phonological store* seems to be represented at the parietal–occipital–temporal junction in LH. It involves LH supramarginal and angular gyri and BA 40 bilaterally. Phonological working memory is represented in the inferior parietal cortex (roughly corresponding to BA 40). Phonological discrimination is represented in the LH planum temporale, and phonological word form representations seem to be stored in the LH anterior superior temporal lobe. Clinical and neuroimaging data clearly show that LH STG is crucial in speech processing because it houses general auditory systems and specialized networks for phoneme recognition. Areas involved in phonological access and storage partly overlap with general speech perception areas, but they are more strongly left-lateralized than speech perception.

Prosody is usually described as linguistic (i.e., intonation and accent position) or affective/emotional (i.e., emotional content, attitude, condition and mood of the speaker), and neurophonetic research is typically conducted along these lines. Brain-injured patients exhibit double dissociation in their ability to process prosody: LH-damaged patients have difficulty understanding words, with spared interpretation of emotional prosody; RH-damaged patients have intact comprehension of linguistically expressed meaning but are impaired in distinguishing among utterances on the basis of affective prosody (this extends to nonverbal expressions as well).

In tone languages, such as Thai or Mandarin Chinese, LH is dominant for lexical tones. However, speakers of one tone language do not process lexical tones of another in their LH by default. It is reserved only for language-specific tonal distinctions. In languages that use pitch accent such as Norwegian or Croatian, the little available data on laterality are inconclusive.

Research into neural bases of prosody has offered often contradictory answers. Although it is most frequently claimed that RH is dominant for affective/emotional and LH for linguistic prosody, it has been proposed that RH is dominant for all types of prosody or that the building blocks of prosody (i.e., duration, tone, and intensity) are lateralized differently or even that prosody may be a subcortical activity. However, it appears that production and perception of prosody require intact parts of RH that correspond to traditional speech/language areas in the LH.

Dual Processing Streams/Pathways

Acquisition and learning of speech sounds as well as their use in communication rely on connections between the auditory and motor areas of the brain. These connections enable associations between perceptual and motor representations as well as syntactic and semantic processing via two dorsal and two ventral pathways/streams. The (dorsal) pathway connecting dorsal premotor cortex with posterior STG/middle temporal gyrus via parietal cortex supports bottom-up auditory-to-motor mapping (as in speech repetition), and the pathway connecting BA 44 with posterior STG via arcuate fasciculus (AF) is in charge of complex syntactic processing (providing top-down prediction for the incoming input).

Immature model of interconnectivity has been found in children before the age of 6 years, and it is characterized by stronger connections between homotopic areas in the two hemispheres than between anterior and posterior regions. The mature pattern of connections between STG and inferior frontal gyrus is established between 5 and 18 years. However, this seems to be true of the *syntactic* connection (between STG and BA

44, probably due to not fully matured AF), whereas the *somatosensory* connection (between STG and premotor cortex) is present at birth and constitutes the basis for auditory-based phonological learning. This dorsal stream is crucial in speech acquisition.

The two ventral pathways are devoted to semantic and basic syntactic processing. The pathway connecting frontal cortex with temporal cortex, parietal cortex, and occipital cortex supports semantic processing and comprehension. The one that connects anterior inferior frontal cortex with anterior temporal cortex is in charge of local syntactic processing.

In the process of acquisition/learning, these pathways are important for recognition and perfecting movement patterns necessary for production, but it is also known that the primarily auditory areas are activated in speech production and that perceptual tasks activate the anterior language areas primarily associated with production. This (in addition to the discovery of mirror neurons in the 1990s) sheds new favorable light on the motor theory of speech perception since there is actual evidence that speech perception activates motor representations. However, higher level processes that involve comprehension (i.e., listening for meaning) seem to be independent of the speech motor system. It should be stressed that the initial enthusiasm in trying to relate the primate mirror neuron system to spoken language processing has been dampened by the findings that speech processing involves a much more complex network than predicted by hypothesis resulting from non-human primate studies.

Disorders

Some of the typical neurogenic speech disorders are dysarthrias, apraxia of speech (AOS), aphasias, and pure word deafness. Dysarthrias affect motor execution in some or all of the components of speech production: respiration, phonation, and articulation. Severity and type of dysarthria (which can be spastic, flaccid, ataxic, hypokinetic, hyperkinetic, or mixed) depend on the affected part of the nervous system: rolandic motor cortex, anterior cingulate cortex and supplementary motor area, upper motor neuron, brainstem, and basal ganglia or cerebellum, including its afferent/efferent pathways. Some of the characteristics include disorders of speech tempo, changes in voice quality, limited movement of articulators, and muscle weakness.

AOS does not involve motor execution problems, and the intended target (i.e., phonological form of the word) is preserved, but the problem lies in phonetic encoding (i.e., in transforming abstract representations of word forms into motor commands to articulators). It is characterized by segmental (e.g., phonetic distortions, phonemic paraphasias) as well as prosodic impairment (e.g., different kinds of dysfluencies) and inconsistent errors. Acquired AOS occurs after lesions to Broca's area and left anterior insula. Developmental AOS presents with similar speech production errors, but without consistent neuroanatomical basis, and is thought to have genetic origin.

Aphasias are generally described as language impairments, but phonemic errors in speech production (e.g., substitutions, omissions, additions, or combinations) are a common aphasic symptom. Absence of phonetic distortions distinguishes phonemic paraphasias in patients with aphasic phonological impairment from those suffering from AOS. Phonemic paraphasias are a result of faulty phonological encoding (i.e., incorrect selection or sequencing of phonemes). They extend to reading and writing, because reading/writing involve mapping from phonological to orthographic representations. Owing to this, failure to develop phonological awareness in children is an indicator of risk of dyslexia.

Severity and type of aphasia (which is commonly categorized on the basis of criteria of fluency, speech comprehension, and ability to repeat words/phrases) also depends on the affected brain areas. Damage to AF typically results in impaired speech repetition (characteristic of conduction aphasia), which supports the view that AF is crucial in auditory-to-motor mapping.

Pure word deafness (i.e., inability to understand speech in spite of normal peripheral hearing) occurs after lesions in the temporal lobe (mostly bilateral). The findings that it sometimes occurs after LH-temporal lesions but never after RH-temporal lesions support the notion that auditory processing of speech is left-lateralized. On the other hand, RH seems to be dominant for processing of music, environmental sounds, and speaker characteristics. A wider term, *auditory agnosia*, is used for such selective impairments. Like other types of agnosias, they are consequences of lesions in the so-called secondary cortical areas that are devoted to pattern/form recognition (*gestalts*), and as opposed to bilateral and modality-specific primary cortical areas, they are functional and lateralized. In other words, an auditory stimulus will reach primary auditory areas in STG in both hemispheres and then, depending on the type of the stimulus, it will be further processed in the hemisphere dominant for that type of stimuli (e.g., predominantly LH for speech, RH for nonspeech stimuli). Since the early 2000s, the more frequently used terms for primary, secondary, and tertiary/associative areas are *core*, *belt*, and *parabelt*.

Another way of looking at this is based on claims that the two hemispheres have different skills/priorities with respect to temporal and spectral analysis of auditory stimuli—LH having better temporal resolution and poorer spectral resolution abilities and RH having the opposite characteristics. This is in line with general LH-RH processing dichotomies, where LH is devoted to analytical, serial/sequential, and categorical processing; local representation; and high frequencies, whereas RH is devoted to holistic, parallel, and coordinate processing; global representation; and low frequencies.

Bilingualism

The key neurolinguistic question in studying bilinguals is nonnative, foreign language (L2) representation in comparison to native language (L1), and in that context the focus of neurophonetics has been on the representation of phonological store(s) and dealing with language-specific aspects. There is no evidence that L2 would involve a special module or structures not present in monolinguals, but the precise ratio of overlap and specificities is still under investigation. Due to neural plasticity, humans are able to master sound systems of languages different from their own even after L1 system is firmly in place, constituting a filter of sorts through which all novel sounds (and rules) are processed. Learning new criteria of categorical perception as well as learning to attach phonological relevance to tones in case of a tone L2, in addition to posterior speech areas (mostly STG and adjacent regions), activates frontal regions bilaterally. This bilateral activation does not necessarily mean that L2 prefers the RH; rather, it reflects the fact that unfamiliar tasks/materials increase processing demands, resulting in more neural tissue being engaged. It may also be an indication of employing other/additional strategies.

Final Thoughts

Neurophonetics also tries to address the question: Is speech special? In other words, is there a special module or network that is reserved exclusively for speech production and perception, or is speech just another function that relies on multipurpose structures and mechanisms? The *hardware* used for listening and speaking has other primary roles (e.g., determining the source and direction of sound or chewing, swallowing, and breathing, respectively). As for *software*, some clinical data, especially on agnosias, may suggest that speech processing engages a module/network of its own, but in light of other evidence, it seems more (or just as) likely that it is a matter of the signal's temporal and spectral properties. In terms of speech production, it is clear that there are motor control disorders that are characteristic exclusively of speech. However, neuroimaging data have not revealed any exclusively speech-specific modules/networks that would not be activated in a number of other functions as well. It seems that speech is not special from birth, but in the course of intense motor and perceptual learning during the period of speech acquisition, humans develop this innate ability to adapt general cognitive functioning to accommodate speech processing.

Neural basis of speech is best viewed as a comprehensive multimodal network distributed across and throughout the brain, with some focal and highly specific areas.

See also Anatomy of the Hearing Mechanism and Central Audiology Nervous System; Anatomy of the Human Neurological System; Neurogenic Communication Disorders; Neurolinguistics; Speech Perception, Theories of; Speech Production, Theories of

Vesna Mildner
http://dx.doi.org/10.4135/9781483380810.n412
10.4135/9781483380810.n412

Further Readings

Binder, J. R. (2008). The neural bases of communication: Evidence from neuroimaging. Retrieved July 20, 2017, from www.asha.org

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. Current Opinion in Neurobiology, 23, 250–254.

Hertrich, I., & Ackermann, H. (2013). Neurophonetics. WIREs: Cognitive Science, 4, 191–200. http://dx.doi.org/doi:10.1002/wcs.1211

Mildner, V. (2008). The cognitive neuroscience of human communication. Mahwah, NJ: Erlbaum.

Ziegler, W. (2011). Neurophonetics. In M. J. Ball, M. R. Perkins, N. Müller, & S. Howard (Eds.), The handbook of clinical linguistics (pp. 491–505). Malden, MA: Wiley-Blackwell.