Neurophonetics

Roll, Mikael

Published in:
Proceedings of Fonetik 2019

2019

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Neurophonetics

Mikael Roll
Department of Linguistics & Phonetics, Lund University, Sweden
mikael.roll@ling.lu.se

Abstract
This is a short review focusing on some research areas of neurophonetics: the neural underpinnings of speech processing and the time course and components of its neurophysiological correlates.

Introduction
The relation between the brain and speech processing started to be studied in the 19th century through cases of aphasia, where speech or comprehension was impeded due to brain lesions. During the 20th century, brain imaging and measurements of neuronal activity made it possible to record brain functions during speech processing and even relate them to healthy brain structures. This short review highlights some research areas starting with the anatomy of speech processing and continuing with neural correlates of online speech processing.

Anatomy of phonetic processing
Broca’s and Wernicke’s areas
Already by the mid 19th century, a speech control center was identified in the left frontal lobe (Bouillaud, 1825; Dax, 1865). Specifically, Broca’s area in the left inferior frontal gyrus (IFG), was found to coordinate “the movements of articulated language” (Broca, 1861). Ascribing complex coordination of movement to Broca’s area as a primary function is in line with its anatomical location, rostral to primary motor and premotor cortex areas controlling lip, jaw, tongue, and larynx movements involved in speech. Thus, a recent proposal suggests increasingly complex motor planning to be represented in a rostral-going direction along the frontal lobe starting from primary motor cortex in the pre-central gyrus (Badre & D’Esposito, 2009; Uddén & Bahlmann, 2012).

During the second half of the 19th century, a “center of acoustic images” was proposed in the superior temporal gyrus (STG) (Wernicke, 1874), which was later called Wernicke’s area. The originally proposed region of STG, lateral to primary auditory cortex in Heschl’s gyrus (DeWitt & Rauschecker, 2013) is thought to be homologous to an area responding to communication calls in macaque monkeys (Rauschecker & Tian, 2000). In humans, it is activated by segmental and prosodic phonological features (Mesgarani, Cheung, Johnson, & Chang, 2014; Tang, Hamilton, & Chang, 2017). Cortical thickness in Wernicke’s area has been found to correlate positively with speed of processing word accents in native Swedish speakers, supporting this region’s importance for storing native phonological patterns (Schremm et al., 2018).

Connections between language areas
Although he had no firm proposal on the neural pathways underlying it, Wernicke (1874) also identified a “conduction aphasia,” affected by lesions in the connection between Wernicke’s and Broca’s areas. Later research has shown that speech processing depends on two main streams of auditory processing, both originating in Heschl’s gyrus (Hickok & Poeppel, 2004).

Dorsal stream
The dorsal stream proceeds posteriorly from Heschl’s gyrus via the planum temporale through inferior parietal cortex and the superior longitudinal fasciculus
(SLF)-arcuate fasciculus in anterior direction to premotor cortex and Broca’s area (Makris et al., 2004). Whereas the arcuate fasciculus has been argued to have a special function in syntactic processing (Skeide, Brauer, & Friederici, 2016), the SLF is generally engaged in sound localization, spatial navigation and sensorimotor integration (Makris et al., 2004). The last aspect is important for language. Thus, the dorsal stream is involved in sensorimotor functions such as phonological working memory, inner speech, and word repetition, and is probably crucially involved during language learning and predictive processing (DeWitt & Rauschecker, 2013; Hickok & Poeppel, 2004; Roll, Söderström, Frid, Mannfolk, & Horne, 2017; Saur et al., 2008).

Ventral stream

From primary auditory cortex, the ventral stream continues through the anterolateral part of Heschl’s gyrus and planum temporale to STG (DeWitt & Rauschecker, 2013). It then extends in anterior direction over STG until reaching the anterior superior temporal sulcus (STS). From the anterior temporal lobe it connects to the anterior part of Broca’s area by joining the occipitofrontal fasciculus through the extreme capsule (Friederici, 2017; Saur et al., 2008). The general function of the ventral stream is auditory object recognition. It also recognizes known words and shorter phrases through a system of increasingly complex sound representations (DeWitt & Rauschecker, 2013). Thus, primary auditory cortex is “tonotopically” organized, meaning that different sound frequencies map to different locations. Mid STG hosts hierarchically organized representations of phonetic features (Mesgarani et al., 2014) and speaker-normalized F0 patterns (Tang et al., 2017). Following the ventral stream in anterior-going direction sensitivity to increasingly complex unities is found: syllables, words, and shorter phrases.

Motor involvement in speech perception

Although the ventral stream seems to host a system for phoneme-to-word recognition in its own right, the dorsal stream is also consistently found to be activated during speech processing. In this vein, the motor and premotor areas controlling the articulators involved in producing a speech sound are also activated when perceiving the same sound (Pulvermüller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). The motor areas indeed seem to facilitate speech perception, since disabling e.g. the part of motor cortex controlling lip movements using transcranial magnetic stimulation (TMS) decreases performance in discrimination of syllables along a /ba/-/da/ continuum (Smalle, Rogers, & Möttönen, 2015). Activation of the caudal part of planum temporale during speech perception might be part of a more general system of auditory representations of motor activity, like hammering, sawing etc. (Warren, Wise, & Warren, 2005). In sum, motor activity engaging the dorsal stream is likely to be involved in speech perception, but is probably most crucial when information needs to be completed, under noisy circumstances, during active prediction (e.g. listening to slow or disfluent speech), and during language learning.

Online phonetic processing

Mismatch negativity (MMN)

Event-related potentials (ERPs) and magnetoencephalography (MEG) can record online speech processing with high temporal precision. Rapid phonological processing has been detected in the ERP effect mismatch negativity (MMN) and its magnetic counterpart. MMN experiments use an “oddball” paradigm. This means that a standard stimulus is presented with high frequency of occurrence interspersed deviant stimulus occurring with low frequency. The standard stimulus is thought to maintain a constant activation of its memory
trace, which the deviant interrupts, resulting in an instantaneous increase of its trace activation. (Näätänen, 1992). The MMN responds more strongly to existing phonological contrasts in a given language (Dehaene-Lambertz, 1997). It has also shown to be more left-lateralized, with sources in STG, for speech stimuli than for complex noise, giving further support for phonological representations in Wernicke’s area (Shtyrov, Pihko, & Pulvermüller, 2005). MMN effects have further been found e.g. in response to stress (Honbolygó, Csépe, & Ragó, 2004; Zora, Riad, Schwarz, & Heldner, 2016; Zora, Schwarz, & Heldner, 2015) and phonotactic probability (Bonte, Mitterer, Zellagui, Poelmans, & Blomert, 2005). The latencies reported for the MMN have varied between 100 and 300 ms. The phonological mapping negativity (PMN) is a similar response that is slightly later timed: 250-350 ms (Connolly & Phillips, 1994). The difference between MMN and PMN is that the latter occurs without an oddball paradigm, with the context making phonemes unexpected.

Preactivation negativity (PrAN)
A recently proposed component overlapping the MMN in latency (136–200 ms after word onset) but which is found without using oddball paradigm and mismatch is the pre-activation negativity (PrAN) (Roll et al., 2017; Söderström, Horne, Frid, & Roll, 2016). PrAN is a speech perception component indexing the predictive value of phonemes at word onset. In this way, it is larger for word onsets with small cohorts of frequent lexical competitors. PrAN has been seen for predictively useful segmental phonemes (Roll et al., 2017), word accents (Roll, 2015; Roll et al., 2015; Söderström et al., 2016; Söderström, Horne, Mannfolk, Westen, & Roll, 2017; Söderström, Horne, & Roll, 2017), and boundary tones (Söderström, Horne, Mannfolk, Westen, & Roll, 2018). Increased degree of coarticulation between the first two phonemes of words (Lindblom & Sussman, 2012) would be likely to move the onset of PrAN to well before 136 ms.

Prosodic phrases and working memory
Intonation phrase boundaries give rise to a slow, positive-going waveform, a ‘closure positive shift’ (CPS) in ERP studies (Roll & Horne, 2011; Steinhauser, Alter, & Friederici, 1999). During silent reading, commas cue implicit phrase boundaries also reflected in a CPS (Steinhauer & Friederici, 2001). Even without explicit cues a CPS is elicited if a phrase boundary is strongly expected (Toepel, Pannekamp, & Alter, 2007). Roll, Lindgren, Alter, and Horne (2012) adjusted reading speed during silent reading so that one, two, or three syntactic phrases were read within a time span of 2.7 s, similar to the time limit of 2-3 s proposed for phonological short-term memory (Baddeley, Thomson, & Buchanan, 1975). A CPS was produced every 2.7 s independently of how many phrases were read within the time constant. The results indicate that readers try to construct implicit prosodic phrases with speech that fit into their phonological working memory. Time-driven implicit prosodic phrases have later been observed to guide syntactic parsing (Schremm, Horne, & Roll, 2016).

Conclusions
Speech perception involves a ventral stream for auditory object recognition with increasingly complex phonological representations and a dorsal stream important for auditory-motor integration. The first clear neural signs of phonological processing occur around 100 ms following stimulus onset and can reflect increased activation due to unexpected phonemes or increased pre-activation of anticipated word endings.

Acknowledgements
This work was supported by the Swedish Research Council (2018.00632), Marcus
and Amalia Wallenberg foundation (2018.0021), and Knut and Alice Wallenberg foundation (2018.0454).

References


Zora, H., Riad, T., Schwarz, I.-C., & Heldner, M. (2016). Lexical specification of prosodic information in