F. H. Guenther. **Neural control of speech**. Cambridge: MIT Press, 2016. 424 p. ISBN 978-0262034715.

Ekaterina V. Tomas

National Research University Higher School of Economics, Moscow, 101000, Russian Federation; ekaterina.tomas@gmail.com

Екатерина Викторовна Томас

Национальный исследовательский университет «Высшая школа экономики», Москва, 101000, Российская Федерация; ekaterina.tomas@gmail.com

Classical models of speech production have primarily focused on the psycholinguistic aspects of the process, identifying its components (i. e., meaning generation, lexical selection, functional assignment, phonological encoding and articulation) and how they interact [Bock, Levelt 1994; Dell, Change, Griffin 1999; Fromkin 1973; Garrett 1975 et al.]. This integral approach has provided us with a solid conceptual understanding of the various processes involved at every stage of speech production, suggesting, however, that each of these stages is likely to have its own unique operational mechanisms. Thus, a more detailed exploration of each of these individual components of the speech production system is now required to build more refined and realistic models of speech production, reflecting how this multicomponent task of communication is orchestrated by the brain. For example, the articulation of speech needs to be studied separately from the lexical selection process because each relies on its own specific set of neuroanatomical structures and has its individual mechanisms of functioning and pathogenesis.

"Neural Control of Speech" by F. H. Guenther is a brilliant attempt to tackle the problem using the "divide and conquer" approach. The book presents a detailed and comprehensive account of the neural processes involved in the articulation of speech based on the growing body of cross-disciplinary evidence from the fields of psycholinguistics, neuroscience, clinical linguistics, and computational modelling. It therefore provides a unified theory of how the brain performs the complex transformation of phonological units into precisely timed activation of motor neurons and the associated muscle fibres. The book also proposes a computational solution to the problem — the DIVA (Directions Into Velocities of Articulators) model implemented mathematically as an artificial neural network (ANN). Due to the nature of the articulation process, its key aspects (i. e. the acoustic features of the speech signal and the automated motor programs associated with its production) are best described by physics and human (neuro)anatomy, rather than psycholinguistic and cognitive mechanisms underlying other speech production processes such as meaning generation or lexical selection. Thus, the book has a strong neurophysiological focus, which seems to resonate with the Russian tradition of neurolinguistic and speech pathology research, dating back to the classical works by A. R. Luria and his colleagues [1964; 1976] (see also a recent review in [Akhutina, Pylaeva 2012]).

The book consists of ten chapters, and the first four cover the general theoretical background with an overview of the research methods, terms and concepts used throughout the book; here the author also introduces the DIVA model and its components. Chapters 5—7 discuss these components of speech control system in greater detail, including the "feedforward control" (FFC) of speech responsible for the initiation and delivery of the motor programs, and the auditory and somatosensory "feedback control" (FBC) systems supporting these processes. As is suggested by their names, the auditory FBC is associated with the auditory perception of acoustic formants of speech sounds, while the somatosensory FBC — with the position and motion of vocal musculature during articulation. Both the auditory and the somatosensory feedback systems are critically important at early stages of language development for mapping target phonemes to motor actions; during maturation and in adulthood, however, they carry out predominantly "system tuning" functions. Chapters 8 and 9 go beyond the level of individual phonemes and focus on the

suprasegmental patterns of speech production, specifically, on how speech sounds are organized into longer sequences and how the prosody supports this process. Here the author introduces the **G**radient **O**rder DIVA (GODIVA) model — an extension of the original model to account for neural computations underlying multisyllabic sequences. Finally, Chapter 10 focuses on speech motor disorders that disrupt fluent and smooth articulation. These are various forms of dysarthria, apraxia of speech, medial premotor syndromes and stuttering. Importantly, the severity of these impairments appears to depend on two factors: 1) on the patient's age, i.e. his/her linguistic skills prior to affliction; and 2) on whether it is the FFC or FBC that has been affected by the disorder. Thus, in adult speakers the damage to FBC typically results in minor deficits in their speech output compared to what is observed for the FFC system disruption. This is because the auditory and somatosensory feedback is believed to play a secondary role in the mature speech control system. In contrast, for young children the damage to either FFC or FBC has been shown to cause significant motor deficits, since the ongoing tuning of the FFC system is dependent on the intact auditory and somatosensory feedback mechanisms.

Neuroanatomy of speech articulation

Since the main focus of the book is on studying the neurophysiology of the speech control process, the cortical and subcortical structures involved in speech are discussed in great detail. The general overview starts with the description of their evolutionary predecessors, particularly the circuits involved in the production of learned voluntary vocalizations in non-human primates, for whom the reticular formation — part of the brainstem primarily involved in learned and innate reflexes — acts as the convergence zone of higher-level cortical projections. These projections come from the cerebral cortex via two pathways: limbic and motor. The former, also known as the cortico-basal ganglia loop, runs via putamen, pallidum and thalamus; it is associated with the motivation and readiness to vocalize, including such properties as global loudness and intensity of a vocalization. The motor pathway is the cortico-cerebellar loop, which connects pons, cerebellum, and thalamus and is responsible for higher-level coordination of the muscles involved in vocalization (see for details [Jürgens 2009]). Importantly, these pathways and structures carry out similar functions in humans, suggesting their fundamental evolutionary role in developing a verbal communication system.

The subsequent discussion of the human speech motor system is given from the functional periphery to the core. Thus, the lowest level of the neural system underlying speech is formed by the medulla, pons, and midbrain, which serve as a relay station for processing input / output sensory information. During bottom-up processes (i.e. involving receptive mechanisms), the cranial nerve nuclei in the medulla deliver tactile and proprioceptive information from speech organs and also auditory information from the cochlea. Similarly, when executing top-down commands, which come from higher-level neural structures through corticobulbar tract, the cranial nerves project back to the muscles to deliver the motor programs.

Above these primary level of structures, the motor cortical commands are processed and shaped by two reentrant subcortical loops, both running through thalamus. These loops and their basic functions are in essence the same as those proposed for non-human primates: a cortico-basal ganglia loop and a cortico-cerebellar loop. The former plays a key role in motor program selection and initiation. The latter is responsible for finely timed muscle activations required for rapid speech.

Finally, the most sophisticated computations for speech articulation are carried out in the cerebral cortex. Here, several major sites are distinguished. At the lowest level of cortical processing lies the *Rolandic cortex*, it is responsible for integrating somatosensory and motor representations and also for generating motor commands. It is further divided into the postcentral gyrus (particularly, primary somatosensory cortex) and the precentral gyrus (primary motor cortex and a portion of premotor cortex). Together with the cortico-cerebellar loop, these gyri form FFC — the system responsible for producing and coordinating highly articulated movements of multiple articulators during speech. However, precentral and postcentral gyri are also an important part

of the somatosensory FBC system as they are involved in processing tactile and proprioceptive information about the produced segments.

Another important component of the FBC control system is located in the *primary auditory cortex*, specifically in the superior temporal gyrus. It is involved in processing auditory attributes of speech segments, such as their fundamental and formant frequencies, etc. The supplementary motor area (SMA), located ventrally to the Rolandic cortex, plays a major role in movement initiation (together with the basal ganglia in the cortico-basal ganglia loop). It is thought to be responsible for stringing speech sounds in longer sequences and also for the motivation to speak. Importantly, all the cortical structures discussed above function bilaterally, and thus unilateral damage typically results in minor deficits or quick recovery of normal / quasi-normal functioning: the intact hemisphere takes over the functionality of its counterpart. This is not the case, however, for the structures responsible for the highest levels of speech motor planning. In the majority of the population these are left-lateralized and include the left inferior frontal gyrus (i.e. *Broca's area*), left anterior insula and posterior superior temporal gyrus (*Wernicke's speech area*). Damage to these regions results in apraxia of speech when the patient loses the ability to generate the motor program for speech sounds. In contrast, right-hemisphere counterparts of these regions appear to support sensory FBC mechanisms. However, there is growing evidence that functional lateralization can also be observed for other cortical areas (e.g., SMA and auditory cortex) and subcortical structures, including basal ganglia, cerebellum, and thalamus [Poeppel 2003; Gil Robles 2005]. Despite this, in cases of unilateral damage, the most severe impairments of speech articulation result from the disruptions in the insular cortex and the left inferior frontal gyrus. Importantly, since the high level speech motor planning is more closely related to language and cognition than to articulation, these left-lateralized language areas are beyond the scope of the book; thus, they are discussed very briefly and are not included in the DIVA model.

The DIVA model

In 1950's researchers started characterizing speech production as a control process. Since then many computational solutions of increasing sophistication have been proposed to model this process. The DIVA model is specified both neurally and mathematically, providing a unified account for a broad range of acoustic, kinematic, and neuroimaging data. Its name reflects this integral approach: the DIVA model describes how the brain transforms the desired **D**irections of movements **I**nto **V**elocities of the **A**rticulators. From the mathematical viewpoint, the model is a computer-implemented ANN, consisting of equations for neural activities and synaptic weights. The model has been tested on empirical data and in computer simulations with the ANN controlling an articulatory synthesizer that produces simulated articulator movements and acoustic signals. For example, the author compares the results of the computer simulations based on the DIVA model to the empirical data coming from multiple experimental fMRI studies, and highlights similarities in the distributions of the activated cortical areas observed during real and simulated articulation.

The book intentionally leaves out the mathematics underlying the model, sacrificing these technical details for the sake of conceptual understanding, i. e. the model's neural specification. From this perspective, the functional components of the DIVA model are linked to the brain regions involved in the generation of articulatory movements. Recall that the FFC system is responsible for fluent and automatic production of speech sounds, and the two interacting FBC systems — the auditory and the somatosensory controls — support self-monitoring and tuning used when the attempted targets do not match the productions. Interestingly, although the existence of the FBC has been recognized for years in speech science and thus has often been included in other computational models of speech production [Houde, Nagarajan 2011; Tremblay et al. 2003], the unique feature of the DIVA model is that it relies on somatosensory information in addition to auditory control. Despite the evidence that viewing a speaking face enhances the ability to perceive

speech in adults [McGurk, MacDonald 1976], based on the studies of language acquisition in blind children [Landau, Gleitman 1985], visual perception seems to have limited effect on articulation. Therefore, the DIVA model does not have a visual feedback controller among its components.

The DIVA model assumes that speech processes in the FFC and FBC are executed in stages. Every structural component of these systems is formed by several nodes carrying out smaller tasks and each has specific neural correlates (see Figure 1 for details). A full cycle of speech sound production starts with the activation of "speech sound map" node, which neurally corresponds to an ensemble of neurons in the left ventral premotor cortex (vPMC). Once the sound map is activated, motor commands are sent to the motor cortex through FFC and FBC systems. First, the FFC generates the previously learned motor program via the "initiation map" responsible for movement initiation at the appropriate time and the "articulatory map", which executes the readout of the motor program. The former runs through the cortico-basal loop (SMA — basal ganglia — thalamus — SMA), and the latter is activated by cortico-cortical projections from the left vPMC (i.e. following the speech sound map readout) to the ventral primary motor cortex of the precentral gyrus bilaterally, supplemented by a cerebellar loop (i.e. pons — cerebellar cortex — thalamus).

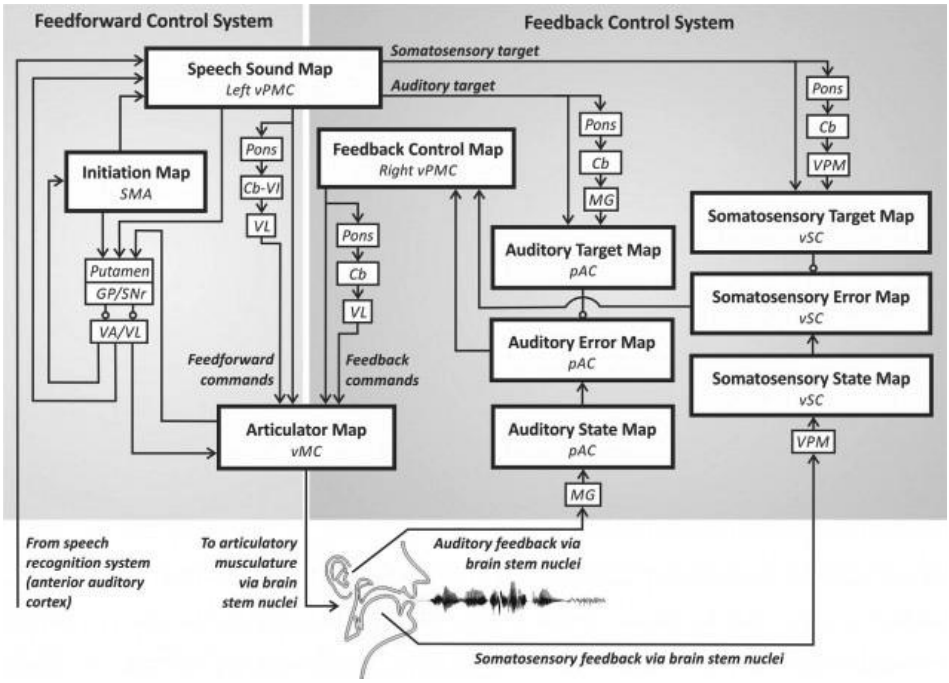

Figure 1. Neural correlates of the DIVA model from Guenther's "Neural control of speech".

After the initiation of the movement, the auditory FBC system checks whether the produced signal matches the target. In the DIVA model, a target is implemented as a time-varying region that encodes the allowable variability of the acoustic signal throughout the syllable. Thus, the sound is classified as the "correct phoneme" if its acoustic properties are within this phoneme's acoustic region. Using the regions rather than point targets allows accounting for a wide range of speech production phenomena, including motor equivalence, contextual variability, anticipatory coarticulation, carryover articulation, and speaking rate effects [Guenther 1995], and it is thus an important feature of the model. The information about the acceptable target regions, or the "auditory target maps", is assumed to be stored in the higher-order auditory cortical areas — the posterior auditory cortex (pAC) — and to be encoded in the axonal projections emanating from speech sound map nodes in left vPMC, both directly and via cortico-cerebellar loop.

The incoming auditory information about the produced signal, which is delivered to pAC via thalamus, is called the "auditory state map". It is compared to the target for the current sound (i.e. auditory target map). When the auditory feedback is outside the target region, this triggers activation of the auditory error node, or the "auditory error map" in the pAC. Here the error flag is transformed into corrective motor commands sent to "feedback control map" in the right vPMC, which in turn projects to the ventral motor cortex, launching "articulator map" to perform the adjusted movement.

In a similar manner, tactile and proprioceptive information about how well the production matches the target is processed by the somatosensory FBC. The DIVA model posits that this task is carried out in ventral somatosensory cortex (vSC), which stores information about the acceptable somatosensory target regions ("somatosensory target maps") and processes incoming sensory signal associated with the current sound ("somatosensory state map") and the error information in cases of mismatch between the two ("somatosensory error map"). If the error map is activated, the corrective motor command is sent to the FBC (the right vPMC) to rerun the sound articulation program. Importantly, this subdivision of the FBC system into auditory and somatosensory components is clear only during early stages of language learning, when children rely on the auditory information before they begin to articulate speech sounds, and thus build their tactile and proprioceptive representations. Later on, however, the error maps transform into a continuum of somato-auditory error representations stored in superior temporal gyrus, Sylvian fissure and supramarginal gyrus.

However, several important questions about the FBC remain unclear. Specifically, the reliability of the FBC system during spontaneous speech and particularly its role in self-monitoring for very young children. For example, it has been shown that listening to speech sounds with altered acoustic properties leads to compensatory adjustments during repetition in four-year-old children and adults, but not in toddlers [MacDonald et al. 2012]. In addition, while **in perception** the full repertoire of native phonemic contrasts is available to children before their first birthday (see for review [Best 1994; Werker 1989]), the differentiation of these contrasts **in production** typically occurs much later [Zharkova 2005]. This suggests that articulatory errors in children below the age of three years tend to "pass through" their auditory self-monitoring system undetected, which raises the question about the efficacy and the overall role of this mechanism during early development.

Another issue arises from the fact that the book focuses on the neural control of articulation, i.e. an element of the **production** process. However, speech cannot be easily divided into production and perception, particularly in a dialogue, when listening to an interlocutor leads to the activation of motor planning areas and facilitation of muscle even before the motor plan is developed (see for review [Wilson, Knoblich 2005]). In fact, even the self-monitoring FBC system cannot be separated from speech perception, because it involves processing one's own productions. Hence, while the "divide and conquer" approach allows building a comprehensible model useful for simulation and experiments, the future step is to combine it with the models of other speech production processes as well as to address the problem of divided attention — when the input information coming from the interlocutor and the environment is processed simultaneously with the information from the self-monitoring auditory FBC. This would allow us to have a more profound and complete understanding of the entire speech production and perception process and thus to make testable predictions on the full range of linguistic phenomena, including those observed in atypical populations such as patients with aphasia, children with language-learning difficulties, and others.

REFERENCES

Akhutina, Pylaeva 2012 — Akhutina T. V., Pylaeva N. M. *Overcoming learning disabilities: A Vygotskian-Lurian neuropsychological approach*. New York: Cambridge Univ. Press, 2012.

Best 1994 — Best C. T. Learning to perceive the sound pattern of English. *Advances in infancy research*. Rovee-Collier C., Lipsitt L. (eds.). Hillsdale: Ablex Publ, 1994. Vol. 8. Pp. 217—304.

Bock, Levelt 1994 — Bock K., Levelt W. J. M. Language production. Grammatical encoding. *Handbook of psycholinguistics*. Gernsbacher M. A. (ed.). New York: Academic Press, 1994. Pp. 741—779.

Dell, Change, Griffin 1999 — Dell G. S., Change F., Griffin Z. M. Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Review*. 1999. Vol. 23. Pp. 517—542.

Fromkin 1973 — Fromkin V. A. *Speech errors as linguistic evidence*. The Hague: Mouton, 1973.

Garrett 1975 — Garrett M. F. Syntactic process in sentence production. *Psychology of learning and motivation: Advances in research and theory*. Bower G. (ed.). Vol. 9. New York: Academic Press, 1975. Pp. 133—177.

Gil Robles 2005 — Gil Robles S. The role of dominant striatum in language: A study using intraoperative electrical stimulations. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2005. Vol. 76. No. 7. Pp. 940—946.

Guenther 1995 — Guenther F. H. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*. 1995. Vol. 102. Pp. 594—621.

Houde, Nagarajan 2011 — Houde J. F., Nagarajan S. S. Speech production as state feedback control. *Frontiers in Human Neuroscience*. 2011. Vol. 5. DOI: 10.3389/fnhum.2011.00082.

Jürgens 2009 — Jürgens U. The neural control of vocalization in mammals: A review. *Journal of Voice*. 2009. Vol. 23. No. 1. Pp. 1—10.

Landau, Gleitman 1985 — Landau B., Gleitman L. R. *Language and experience: Evidence from the blind child.* Cambridge: Harvard Univ. Press, 1985.

Luria 1964 — Luria A. R. Factors and forms of aphasia. *Ciba foundation symposium. Disorders of language*. De Reuck A. V. S., O'Connor M. (eds.). London: John Wiley & Sons, 1964. Pp. 143—167.

Luria 1976 — Luria A. R. *Basic problems of neurolinguistics*. The Hague: Mouton, 1976.

MacDonald et al. 2012 — MacDonald E. N., Johnson E. K., Forsythe J., Plante P., Munhall K. G. Children's development of self-regulation in speech production. *Current Biology*. 2012. Vol. 22. No. 2. Pp. 113—117.

McGurk, MacDonald 1976 — McGurk H., MacDonald J. Hearing lips and seeing voices. *Nature*. 1976. Vol. 264. No. 5588. Pp. 746—748.

Poeppel 2003 — Poeppel D. The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time". *Speech Communication*. 2003. Vol. 41. No. 1. Pp. 245—255.

Tremblay et al. 2003 — Tremblay S., Shiller D. M., Ostry D. J. Somatosensory basis of speech production. *Nature*. 2003. Vol. 423. No. 6942. Pp. 866—869.

Werker 1989 — Werker J. F. Becoming a native listener. *American Scientist*. 1989. Vol. 77. Pp. 54—59.

Wilson, Knoblich 2005 — Wilson M., Knoblich G. The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*. 2005. Vol. 131. No. 3. Pp. 460—473.

Zharkova 2005 — Zharkova N. Strategies in the acquisition of segments and syllables in Russian-speaking children. *Developmental paths in phonological acquisition*. Tzakosta M., Levelt C., van der Weijer J. (eds.). *Leiden Papers for Linguistics*. 2005. Vol. 1. Pp. 189—213.