Mapping of functions to brain regions : A neuro-phonetic model of speech production, perception, and acquisition

Bernd J. Kröger[*] **& Stefan Heim**[**]

1. Introduction : The functional Model

A model has been computer-implemented which is capable of producing and/or perceiving speech items (sounds, syllables, words, or short utterances). The organization of the model is given in Fig. 1 (see also Kröger et al. 2009a). Its *cognitive linguistic module* is not modelled in detail in this neurophonetic approach but it can be assumed that this module is subdivided into a procedural and a declarative part (Ullman 2001). Here it is assumed that phonemic word forms are selected from a mental lexicon (Levelt 1992, Levelt et al. 1999, Indefrey & Levelt 2004) forming the main part of the declarative memory. These forms pass linguistic processing modules including the syllabification module (procedural memory and appropriate processing modules) and subsequently they build up a chain of phonemically specified speech items on the level of the phonemic map (Fig. 1). The subsequent part of the model is the *phonetic or sensorimotor module*. Within all parts of the model, a *neural map* is defined as an ensemble of neurons located in a specific brain region which can be associated with a distinct cognitive or sensorimotor representation or *state* of a speech item. Different *neural activation patterns* occurring within a neural map represent different neural states and different speech items. Coming back to the phonemic map, thus the phonemic description of each speech item generated by the cognitive linguistic module is coded by a distinct neural activation pattern or neural state within the phonemic map.

Subsequently within the phonetic or sensorimotor module, *speech production* can be separated in sensorimotor feedforward and feedback control (cf. Guenther 2006 and Guenther et al. 2006). Feedforward control is the direct generation of articulatory movements from a specific phonemic state. Sensorimotor feedback control is activated during production in order to control, whether the phonetic realization of a speech item is correct and, if it is not, to correct its production.

[*] Department of Phoniatrics, Pedaudiology, and Communication Disorders, University Hospital Aachen and RWTH Aachen University, Germany bkroeger@ukaachen.de

[**] Department of Psychiatry and Psychotherapy, University Hospital Aachen and RWTH Aachen University, Germany sheim@ukaachen.de and: Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Germany s.heim@fz-juelich.de

Sensorimotor feedforward control starts from the phonemic state. If the syllable under production is a *frequent syllable* within the speaker's language – i.e. an already well practiced or "overlearned" syllable (see the speech acquisition model of the model, described below) – the phonemic activation on the level of the phonemic map leads to a co-activation of the appropriate auditory, somatosensory, and motor plan state for that syllable via the *phonetic map*. Thus motor and sensory states for frequent syllables are assumed to be learned during speech acquisition and stored within the phonemic-phonetic, phonetic-sensory, and phonetic-motor mappings (arrows between the appropriate maps in Fig. 1). The phonetic map as well is built up during speech acquisition and speech items are ordered within this map with respect to phonetic features (phonetotopy, see Kröger et al. 2009b). In terms of neurocomputing, the phonetic map is a self-organizing map (SOM, see Kohonen 2001), representing the associations between the phonemic, motor, and sensory representations for all types of frequent speech items within the target language. Thus the phonetic map links each neural state within the phonemic map with an appropriate neural state of the *motor plan map* and one of the *sensory maps (auditory and somatosensory map)*. From the viewpoint of self-organization, the phonetic map is a part of the mapping between phonemic, motor, and sensory maps. The phonetic map can be interpreted as *hyper- or supramodal neural map*, connecting the phonemic, motor and sensory states of a speech item. We hypothesize that this level is an explicit level of speech relevant mirror neurons (cf. Fadiga et al. 2002, Fadiga and Craighero 2004, Rizzolatti and Craighero 2004). All maps and mappings described thus far form the *mental syllabary* as is postulated by Levelt and Wheeldon (1994). *Infrequent syllables* are not processed by the mental syllabary but by a separate motor planning module (Fig. 1), generating the motor plan of a syllable on the basis of subsyllabic units (cf. Levelt and Wheeldon 1994, Levelt et al 1999). This motor planning module is linked with the phonetic map since it profits from the phonetic knowledge on production of frequent syllables stored within the phonetic map. A hypothetical organization of the motor planning module is described in Kröger et al. (accepted).

While feedforward control as described above is the main control mechanism within normal (adult) speech production and implemented in our model for frequent syllables, online *sensorimotor feedback control* for supervising the ongoing flow of speech production (cf. Guenther 2006 and Guenther et al. 2006) is dominantly activated during production of infrequent syllables and during speech acquisition. Feedback control starts with the auditory and somatosensory processing of the articulatory and acoustic signals produced by the speaker itself (Fig. 1). Lower level somatosensory signals are directly projected to and processed by the motor programming and execution module. Higher level somatosensory and auditory information (e.g. how a speech item "feels like" during production and how it "sounds like") is projected to the somatosensory-phonetic and auditory-phonetic processing module via the somatosensory and auditory map. These *current external sensory feedback states* (external state ES in Fig. 1) are compared with *acquired or trained sensory states* (trained and stored during speech acquisition and activated via the phonetic-to-sensory mappings; TS, trained states in Fig. 1) also activated on the level of the sensory processing modules for the current speech item under production. In the case of differences between acquired and current sensory states, an error signal can be generated in order to correct the current forward production (cp. Guenther et al. 2006).

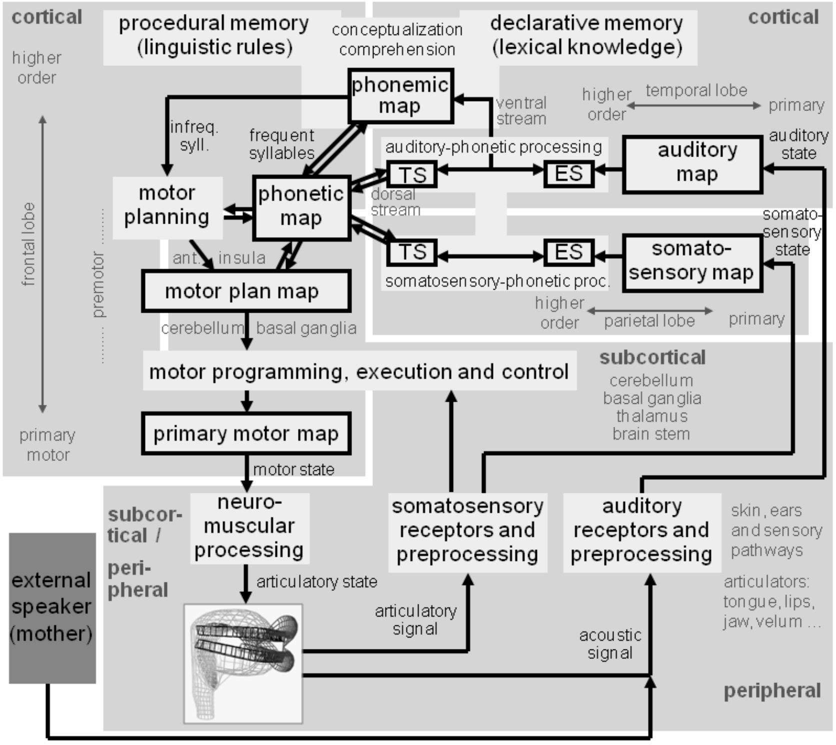

Figure 1. Structure of the neuro-computational model of speech production, perception and acquisition. Boxes with black outline indicate neural maps, arrows indicate neural mappings, and boxes with no outline indicate processing modules, comprising maps and mappings, which are not specified in detail in this figure. The linguistic module (i.e. procedural and declarative memory) as well as the motor programming and execution module is not modelled in detail in our approach. “Trained states” and “external states” are abbreviated as TS and ES (see sensory processing modules). Ventral and dorsal perceptual processing pathways are introduced with respect to Hickok and Poeppel (2007).

A specific feature of our model is the separation of a higher and a lower level of motor representations, i.e. the separation of a *motor plan level* and a *primary* *motor level* (cf. the organization of movement control in action theory, Fadiga and Craighero 2004 and Kröger et al. in press). The *motor plan* of a syllable comprises a score of executable articulatory actions (Kröger and Birkholz 2007). Actions are goal-directed speech gestures describing higher level motor features such as "produce a labial vocal tract closure for [b]", "produce a glottal opening for [p]", or "produce a velopharyngeal opening for [m]" as well as their temporal coordination within a syllable or word. Subsequently, *motor programming and execution* leads to a concrete specification of each gesture and subsequently to a concrete specification of articulator positioning and movement on the level of the *primary motor map*. For example, a labial closing gesture involves coordinated movement of at least three articulators, i.e. the lower jaw and the lower and the upper lips and each of these articulators is controlled by an ensemble of different motor units. Thus the concrete realization of a gesture is not specified on the higher motor plan level but spelled out during motor programming and execution.

Three *function modes* can be differentiated in our model, (i) speech acquisition mode, (ii) speech production mode and (iii) speech perception mode. Each mode uses more or less the whole model as introduced above. *Speech acquisition* is separated in babbling and imitation in our approach (cf. Oller et al. 1999). During *babbling training* the model (now representing a toddler) starts with the production of random motor plans comprising prelinguistic coarse actions (Kröger et al. 2009a). These motor plans are executed and lead to (pre-linguistic) model articulator movements. The articulatory-acoustic model then produces the appropriate sensory states. Thus an ensemble of prelinguistic motor states and associated sensory (auditory and somatosensory) states (i.e. a training set) is generated for *training the central self-organizing map* (i.e. the phonetic map) of the model in its prelinguistic phase. Thus preliminary sensorimotor knowledge is gained during this training. After babbling training the model (the toddler) is capable of associating motor plan states with auditory states, which is a mandatory prerequisite for imitation training. During *imitation training*, the model is stimulated by language-specific external auditory states (i.e. speech items produced by an external speaker, i.e. mother or caregiver). In parallel to the auditory state the phonemic state of each training item (syllables or words) is given now as well, since the toddler associates phonetic forms with meaning in the phase of imitation (the detailed process of developing lexical concepts is beyond the scope of our model). For each external speech item a motor plan state can now be generated on the basis of the sensorimotor knowledge gained during babbling training and this motor state can be executed. If the external feedback auditory state of this production trial deviates from the original auditory state produced by the external speaker, corrections can be introduced via the auditory phonetic processing module until the motor plan of the speech item is satisfactorily for being stored. Thus an ensemble of motor states, appropriate sensory and phonemic states is built up for training the language-specific phonemic-phonetic mapping for all frequent speech items of a language during imitation. Thus far a *model language* comprising five V(owel)-items {V = /i/, /e/, /a/, /o/, /u/} and all one- or two-syllabic words which can be composed of CV-

and CCV-syllables, with C(onsonant) items out of { C = /b/, /d/, /g/, /p/, /t/, /k/, /m/, /n/, /l/} and with CC-clusters out of {C1 = /b/, /g/, /p/, /k/; C2 = /l/} has been trained in our model (Kröger et al., in press_2). Babbling and imitation training is realized here with some temporal overlap. If this temporal overlap of the babbling and imitation phase is strong, i.e. if imitation starts early with respect to babbling, many trails are needed in order to get acceptable imitation items, due to incomplete sensorimotor babbling knowledge. If temporal overlap is less, i.e. babbling mainly precedes imitation training, this problem does not occur during imitation but in this case babbling can be non-effective, since babbling may occur for those items which are not of primary importance for the toddler's mother tongue. Thus the temporal overlap of babbling and imitation training helps *to shape the babbling training set* in order to train sensorimotor relations in those regions of the motor space which are important for a specific target language. Thus this temporal overlap helps to prevent babbling of items which never occur in a specific language (i.e. non-effective babbling items).

After babbling and imitation training the model is capable of *producing* well trained speech items in the feed-forward mode. Feed forward production was already introduced above. It is important to mention again that the activation of the phonemic state of a well trained speech item leads to a co-activation of the appropriate (already learned) motor plan state and (already learned) sensory state via phonetic map (Fig. 1). Actual sensory feedback state activation patterns are generated via the sensorimotor feedback-loop (see above : feedback control). These actual sensory feedback state activation patterns are compared with the learned sensory states. If learned and feedback sensory states deviate markedly, corrected motor plans can be generated. Moreover the phonetic map and its mappings towards sensory maps and motor plan map are modified if the deviation between learned and actual feedback sensory states persists. This leads to *adaptation* (Guenther 2006).

After babbling and imitation training the model is also capable of *perceiving* speech items, i.e. the model is capable of doing identification and discrimination tasks for speech items. Auditory-only perception or audio-visual perception (see Kröger and Kannampuzha 2008, but not indicated in Fig. 1) can be performed by our model. During perception (in comparison to production) the *bi-directionality* of the phonemic-phonetic mapping becomes apparent : Identification of an acoustic speech items means activation of the appropriate auditory state, then leading to a co-activation of an appropriate phonetic and phonemic state via the auditory-to-phonetic and phonetic –to-phonemic mappings.

2. MAPPING FUNCTIONS TO BRAIN REGIONS

The model introduced thus far is a neurocomputational model. The postulated maps and mappings result from neurophysiological knowledge and in addition are based on functional needs occurring during the development of a neurocomputational model capable of producing and perceiving speech items. Thus from a neurophysiological viewpoint there are two main questions which remain to be answered : (i) Are the maps and mappings postulated in this model occurring in the central nervous system (i.e.in cortical, subcortical or peripheral regions)? (ii) If yes, is it possible to specify the location of these maps and mappings in detail?

Since the structure of the model was grounded on neurophysiological knowledge (Kröger et al. 2008), the first question can be answered positively for most of the maps and mappings introduced in our model (Fig. 1). The cognitive linguistic part of speech production ends with a phonological specification of a current speech item under production following lexical retrieval and syllabification (Indefrey and Levelt 2004). The cognitive linguistic network is located mainly in the left frontal and temporal lobe (ibid.). Here, phonological processing occurs in particular in Brodmann's area (BA) 44 as part of Broca's region and in the posterior portion of the left posterior superior temporal gyrus (pSTG; e.g. Burton et al. 2000, Démonet et al. 1992, Zatorre et al. 1996). These two regions interact during phonological processing in language production and comprehension (e.g. Heim et al. 2003), exhibiting differential temporal dynamics during comprehension (Thierry et al. 1999) and production (Heim and Friederici 2003). In language comprehension, activation in the pSTG precedes that in BA 44, whereas the reverse pattern is observed for language production. This finding is in line with the functional interpretation of the pSTG as a phonological word form store (mental lexicon in terms of Levelt et al. 1999) and left BA 44 as a region involved in the feature manipulation of phonemes and in the process of syllabification) of phonemes (e.g. Indefrey and Levelt 2004).

The neurophysiological basis for the distinction between a motor plan level and a primary motor level for speech actions is discussed in Kröger et al. (in press). Functional neuroimaging demonstrated that the left insula, premotor and motor cortex, as well as subcortical regions and the cerebellum are of relevance for speech motor planning (Dronkers 1996; see also Ackermann and Riecker 2004, Heim et al. 2002; for a recent meta-analysis cf. Eickhoff et al. 2009). New insight into the dynamics of the brain networks involved in the processing between lexical-phonological selection and motor output comes from two studies using dynamic causal modelling (DCM). DCM is a method that elucidates the effects regions exert on each other as well as the influence of context variables on the connectivity in a network of brain regions. Whereas the study by Heim et al. (2009) further corroborated the notion that it is in particular left BA 44 that initiates the cascade of post-phonological information processing which ends in the primary motor cortex, the study by Eickhoff et al. (2009) dissociated a sub-network relevant for motor planning (phonetic map, planning module, and motor plan map in Fig. 1) and from that involved in the programming and execution of articulation (programming and execution module and primary motor map in Fig. 1), two phases which are also dissociated in our present neurophonetic model. The DCM analysis revealed that the motor plan is processed in a network comprising the insula, basal ganglia, and cerebellum, with information flowing from the insula to the latter two regions. Then, activation further propagates from cerebellum and basal ganglia to the premotor cortex (BA 6); it is this second network in concert with primary motor cortex (BA 4) which is relevant for the programming and execution of motor plans (see also Hillis et al. 2004; Riecker et al. 2005; 2006). Motor planning in addition may comprise parts of the premotor cortex (SMA or BA6, see Riecker et al. 2005).

Somatosensory and auditory feedback processing as well as auditory processing of signals produced by other speakers are processed in the parietal and temporal lobe. The auditory pathway from ear to brain passes through peripheral and subcortical regions before reaching the primary auditory cortex (BA 41, 42, i.e. location of the auditory map in our model). A temporal short-term memory capable of memorizing the sound of syllable sized units occurs within the unimodal auditory and temporal multimodal sensory cortex (STG or BA 22, i.e. the location of the auditory-phonetic processing module, Fig. 1). Within this region (posterior part of STG) the comparison of trained and external auditory states takes place in order to generate an auditory error signal for correcting the production of a speech item (Guenther 2006, p. 354). The somatosensory pathway from muscles or dermis of speech articulators or vocal tract walls allows feedback control in an inner non-conscious subcortical loop for controlling ongoing motor programming and execution (Fig. 1). Somatosensory feedback signals in a second somatosensory pathway reach the speech organ regions of the primary somatosensory cortex (BA 3, i.e. location of the somatosensory map in our model). A temporal short-term memory capable of memorizing, how the production of a syllable “feels like”, occurs within the unimodal somatosenory cortex (BA 1, 2, 5, and anterior BA7) and within the parietal multimodal sensory cortex (Gyrus angularis and gyrus supramarginalis, i.e. posterior BA7, BA 39, and BA 40, i.e. the location of the somatosensory-phonetic processing module, Fig. 1). Within this region (especially gyrus supramarginalis) the comparison of trained and external somatosensory states takes place in order to generate an somatosensory error signal for correcting the production of a speech item (Guenther 2006, p358).

3. Concluding Remarks

Computer modeling, i.e. exact quantitative modeling of neural activation and neural processing is an important complement to functional brain imaging studies, since these studies are not currently capable of drawing a detailed picture especially of neural processing. But models need to be based on knowledge gained by functional brain imaging studies in order to be realistic. Our current model is a detailed model for speech production and speech perception including speech knowledge, gained during early phases of speech acquisition. The structure of this model is in line with current neurophysiological knowledge as well as with other models of speech production and speech acquisition (e.g. Guenther2006, Guenther et al. 2006) and speech perception (e.g. Hickok and Poeppel 2007). But one open questions concerning our model is that concerning the existence of the phonetic map. Neural mappings between phonemic and motor as well as between phonemic and sensory maps are assumed also by Guenther et al. (2006). If these mappings are assumed to be self-organizing (as are all cortical mappings) an internal neural layer, called “phonetic map” is needed. It will be a goal of our future work to answer this question and if this answer is positive it will be a further goal to estimate the cortical location of such a phonetic map, i.e. the location of a hypermodal map between phonemic, motor, and sensory representations of speech items.

ACKNOWLEDGMENTS

This work has been supported in part by the German Research Council DFG, research grant Nr. KR 1439/13-1 and grant Nr. KR 1439/15-1.

REFERENCES

Ackermann H. Riecker A., 2004, The contribution of the insula to motor aspects of speech production : a review and a hypothesis, *Brain and Language* 89, p. 320-328.

Burton M. W., Small S. L. & Blumstein S. E., 2000, The role of segmentation in phonological processing : An fMRI investigation, *Journal of Cognitive Neuroscience* 12, p. 679-690.

Démonet J. F., Chollet F., Ramsay S., Cardebat D., Nespoulous J. L., Wise R., Rascol A. & Frackowiak R., 1992, The anatomy of phonological and semantic processing in normal subjects, *Brain* 115, p. 1753-1768.

Dronkers N. F., 1996, A new brain region for coordinating speech articulation, *Nature* 384, p. 159-161.

Eickhoff S. B., Heim S., Zilles K. & Amunts K., 2009, A systems perspective on the effective connectivity of overt speech production, *Phiosophical Transactions of the Royal Society* A 367, p. 2399-2421.

Fadiga L., Craighero L., Buccino G. & Rizzolatti G., 2002, Speech listening specifically modulates the excitability of tongue muscles : a TMS study, *European Journal of Neuroscience* 15, p. 399-402.

Fadiga L. & Craighero L., 2004, Electrophysiology of action representation, *Journal of Clinical Neurophysiology* 21, p. 157-168.

Guenther F. H., 2006, Cortical interaction underlying the production of speech sounds, *Journal of Communication Disorders* 39, p. 350-365.

Guenther F. H., Ghosh S. S. & Tourville J. A., 2006, Neural modeling and imaging of the cortical interactions underlying syllable production, *Brain and Language* 96, p. 280-301.

Heim S., Eickhoff S. B. & Amunts K., 2009, Different roles of cytoarchitectonic BA 44 and BA 45 in phonological and semantic verbal fluency as revealed by dynamic causal modeling, *NeuroImage* 48, p. 616-624.

Heim S., Friederici A. D., 2003, Phonological processing in language production : time course of brain activity, *Neuroreport* 14, p. 2031-2033.

Heim S., Opitz B., Müller K. & Friederici A. D., 2003, Phonological processing during language production : fMRI evidence for a shared production-comprehension network, *Cognitive Brain Research* 16, p. 285-296.

Heim S., Opitz B. & Friederici A. D., 2002, Broca's area in the human brain is involved in the selection of grammatical gender for language production : evidence from event-related functional magnetic resonance imaging, *Neuroscience Letters* 328, p. 101-104.

Hickok G. & Poeppel D., 2007, Towards a functional neuroanatomy of speech perception, *Trends in Cognitive Sciences* 4, p. 131-138.

Hillis A. E., Work M., Barker P. B., Jacobs M. A., Breese E. L. & Maurer K., 2004, Re-examing the brain regions crucial for orchestrating speech articulation, *Brain* 127, p. 1479-1487.

Indefrey P. & Levelt W. J. M., 2004, The spatial and temporal signatures of word production components, *Cognition* 92, p. 101-144.

Kohonen T., 2001, *Self-organizing maps*, Berlin, Springer.

Kröger B. J. & Birkholz P., 2007, A gesture-based concept for speech movement control in articulatory speech synthesis, in A. Esposito, M. Faundez-Zanuy, E. Keller & M. Marinaro (eds.), *Verbal and Nonverbal Communication Behaviours*, Berlin, Springer, p. 174-189.

Kröger B. J. & Kannampuzha J., 2008, A neurofunctional model of speech production including aspects of auditory and audio-visual speech perception, Proceedings of the International Conference on Audio-Visual Speech Processing, Moreton Island, Queensland, Australia, p. 83-88 (www.speechtrainer.eu).

Kröger B. J., Lowit A. & Schnitker R., 2008, The Organization of a Neurocomputational Control Model for Articulatory Speech Synthesis, in A. Esposito, N. Bourbakis, N. Avouris & I. Hatzilygeroudis (eds.), *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction*, LNAI 5042, Berlin, Springer, p. 121-135.

Kröger B. J., Kannampuzha J. & Neuschaefer-Rube C., 2009a, Towards a neurocomputational model of speech production and perception, *Speech Communication* 51, p. 793-809.

Kröger B. J., Kannampuzha J., Lowit A. & Neuschaefer-Rube C., 2009b, Phonetotopy within a neurocomputational model of speech production and speech acquisition, in S. Fuchs, H. Loevenbruck, D. Pape & P. Perrier (eds.), *Some Aspects of Speech and the Brain*, Frankfurt, Peter Lang, p. 59-90.

Kröger B. J., Kopp S. & Lowit A., 2010, A model for production, perception, and acquisition of actions in face-to-face communication, *Cognitive Processing* 11, p. 187-205.

Kröger B. J., Miller N. & Lowit A. (in press), Defective neural motor speech mappings as a source for apraxia of speech : Evidence form a quantitative neural model of speech processing, in R. Kent & A. Lowit (eds.), *Motor Speech Disorders*.

Kröger B. J., Birkholz P., Lowit A. & Neuschaefer-Rube C. (in press_2), Phonemic, sensory, and motor representations in an action-based neurocomputational model of speech production (ACT), in B. Maassen & P. van Lieshout (eds.), *Speech Motor Control : New developments in basic and applied research*.

Levelt W. J. M., 1992, Accessing words in speech production : stages, processes and representations, *Cognition* 42, p. 1-22.

Levelt W. J. M. & Wheeldon L., 1994, Do speakers have access to a mental syllabary?, *Cognition* 50, p. 239-269.

Levelt W. J. M., Roelofs A. & Meyer A., 1999, A theory of lexical access in speech production, *Behavioral and Brain Sciences* 22, 1-75.

Oller D. K., Eilers R. E., Neal A. R. & Schwartz H. K., 1999, Precursors to speech in infancy : the prediction of speech and language disorders, *Journal of Communication Disorders* 32, p. 223-245.

Riecker A., Kassubek J., Gröschel K., Grodd W. & Ackermann H., 2006, The cerebral control of speech tempo : Opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures, *Neuroimage* 29, p. 46-53.

Riecker A., Mathiak K., Wildgruber D., Erb M., Hertrich I., Grodd W. & Ackermann H., 2005, fMRI reveals two distinct cerebral networks subserving speech motor control, *Neurology* 64, p. 700-706.

Rizzolatti G. & Craighero L., 2004, The mirror neuron system, *Annual Review of Neuroscience* 27, p. 169-192.

Thierry G., Boulanouar K., Kherif F., Ranjeva J. P. & Démonet J. F., 1999, Temporal sorting of neural components underlying phonological processing, *Neuroreport* 10, p. 2599-2603.

Ullman M. T., 2001, A neurocognitive perspective on language : the declarative / procedural model, *Nature Reviews Neuroscience* 2, p. 717-726.

Zatorre R. J., Meyer E., Gjedde A. & Evans A. C., 1996, PET studies of phonetic processing in speech : Review, replication, and reanalysis, *Cerebral Cortex* 6, p. 21-30.