

THE NATURE OF THE ARTICULATORY CODE

Harald Höge

*Universität der Bundeswehr München
harald.hoege@t-online.de*

Abstract: The paper investigates the nature of the articulatory code (AC), which contains the messages needed for cortical processing to perceive and produce speech. The paper claims that the AC is a multi-item neural code identical for both modalities - speech perception and speech production - and that the transport of the AC between the short-term memory and the speech processing areas is performed by coupled Θ/γ -oscillations. Due to the limitation given by the state of the art in measuring cortical activities, the structure and content of the AC is not deciphered yet. Using findings from research in phonetics and evolution and using invasive as well as non-invasive measurements of cortical activities of non-human and human primates, the paper is focused to verify the hypothesis, that each message of the AC describes the structure and content of a Θ -syllable. The items of each message represent elementary articulatory gestures composing syllables and are coded based on bundles of manner & place like articulatory features. Further the items are generated by spike patterns of neurons connected to complexes of neurons sensitive to articulatory features and are transmitted by modulated γ -oscillations embedded in the Θ -oscillations. A final proof derived from cortical measurement for the proposed AC is still missing. Thus, there is a need for further cortical measurements.

1 Introduction

Goldstein et. al. [8] state: *Language can be viewed as a structuring of cognitive units that can be transmitted among individuals for the purpose of communicating information.* Given this definition, two questions must be answered:

What is the nature of the cognitive units implemented in the human cortex?

What is the nature of the transmission process within the brain and between communicating brains?

The paper has as goal to review evidence for the hypothesis, that the cognitive units are specific articulatory gestures embedded in a syllable and that these articulatory gestures are coded as a multi-item neural code [19] transmitted by coupled oscillations. In the paper this code is called the *Articulatory Code (AC)* and the coupled oscillations are called the *Articulatory Rhythm (AR)*.

Due to the limited abilities of current technology to measure cortical activity, the questions stated above are not answered yet. There exist at least two areas of applications, where the answer to these questions is from high relevance. The first area concerns the development of Brain-Computer-Interfaces (BCI) also called Brain-Machine-Interfaces (BMI) for restoring the communication abilities of handicapped persons [1,2]. The second area concerns the use of the knowledge about the articulatory code to develop cortical models for speech perception, speech production and speech learning leading to cortical inspired automatic speech recognition, speech synthesis and learning systems.

To achieve evidence for the nature of the articulatory code and the articulatory rhythm, in the following findings from psycho-acoustic experiments and findings applying principles from evolution, as well as knowledge gained from neuronal measurements from the cortex are combined.

Regarding the principles of evolution, in the following the statement of Darwin [3]: ‘*man bears the indelible stamp of his lowly origins*’ is applied. Translating this statement to the functionality of the human brain leads to following findings [4]: *comparison the structure and functionality of neuronal complexes of mammals with equivalent complexes of humans the ‘primary’ functionalities are the same but the more advanced of humans are performed by add-on complexes located in their neighborhood.* Further the principles of neuronal processing are the same for all mammal cortices including the human, especially the use of oscillations in sampling and transmission of information and the processing in a six-layer cortical structure [6]. Regarding the evolutionary path from the monkeys to humans, the main developments in vocal communication are the development of the vocal tract [5] and the perception of gestures [6, 8]. In the evolutionary path, facial, manual and body gestures and finally articulatory gestures have been developed. The importance and the role of articulatory gestures is stated by Arbib [6]: *The theory of gestures is grounded in evidence from brain imaging [e.g. 7] that there is a human mirror system for grasping—i.e., a brain region activated for both grasping and observation of grasping—in or near Broca’s area.* Interpreting ‘grasping’ as speech production and ‘observing grasping’ as perception, leads to the conclusion that gestures act as cognitive units and that the structure and transmission of the articulatory code is the same as from that of the ‘grasping code’. Due to the mirror principle the articulatory code must be the same for speech perception and speech production.

The idea, that articulatory gestures are the basis for communication, was promoted by Brownman and Goldstein [9] defining the term ‘gestures’ by: *A gesture is identified by the formation and release of a characteristic constriction within one of the relatively independent articulatory subsystems of the vocal tract (i.e. oral, laryngeal, velic).* The geometry of the gestures (see fig. 1) are based on the concept of manner & place features describing articulatory constrictions. These constrictions define phonemes as standardized by phoneticians [10].

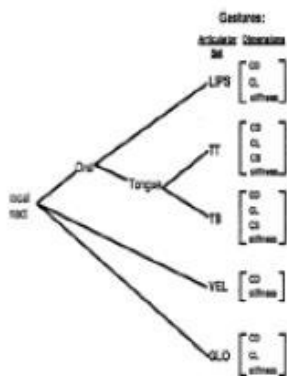


Figure 1 - articulatory geometry (Brownman and Goldstein, 1989) described by the parameters: constriction degree (CD), constriction location (CL), constriction shapes (CS), and the stiffness of the articulatory subsystems lips, tongue tip (TT), tongue body (TB), velum (VEL) and glottis (GLO).

Neurons sensitive to place & manner like features have been found in the cortex. During speech perception and speech production those neurons have been detected in the auditory and sensorimotor cortex (e.g. Bouchard, Mesgarani et. al. [11,12]). In the following those features are called *articulatory features*. Brownman and Goldstein [9] regarded the *gestures* as defined above as independent atoms. This view has been confirmed by Bouchard, Mugler et.al. [11,13] performing high gamma ECoG (see chapter 2) measurements during speech production. The areas regarded are the ventral primary motor cortex (M1v), the ventral pre-motor cortex (PMv), and the inferior frontal gyros (IFG). Yet, the study of the movement of articulators showed, that their timing pattern depend on the position (onset, coda) within a syllable [14]. These finding lead to the concept of syllabic ‘molecules’, which are defined as the combined actions of *gestures* building parts of a syllable. Based on this concept Höge [15] proposed a set of specific molecules called OVC - gestures, which are articulatory gestures describing the onset, the vowel-kernel and the coda of a syllable.

Derived from the timing patterns of gestural movements, psycho-acoustic measurements demonstrated, that the rhythms of larger phonetic units as syllables or foots are coupled to the timing of gesture movements. From these findings the concept of coupled oscillation [16] has been derived in phonetics. As described in chapter 3 coupled oscillations play an important role in cortical speech processing. In speech perception Giraud and Poeppel [17] claim, that coupled theta/gamma (Θ/γ)-oscillations segment the auditory signal into syllables and their phonetic units. In the following the coupled Θ/γ -oscillations are called the articulatory rhythm (AR). Further in this context, the term Θ -syllable has been defined by Ghitza [18] stating: *We shall conclude by proposing the Θ -syllable as an information unit defined by cortical function—an alternative to the conventional, ambiguously defined syllable¹ ... The theta oscillator (the master) provides segmental parsing; assuming perfect tracking, a theta cycle is aligned with a segment that is often a $V\Sigma V$ (Σ stands for consonant cluster).* Due to this definition in the following the term ‘syllable’ is defined to be a Θ -syllable.

A further hint concerning the nature of the AC is the mechanism developed during evolution for transporting information between distant cortical areas. This mechanism has been developed by the need to transport more and more information between increasing amount of specialized distinct cortical areas². Regarding the transmission between cortical areas and the short-term memory, Lisman and Jensen [19] derived following theory: *we review evidence for the following hypothesis: that the dual oscillations form a code for representing multiple items in an ordered way. This form of coding has been most clearly demonstrated in the hippocampus, where different spatial information is represented in different gamma subcycles of a theta cycle.*

Till now, the articulatory code (AC) combined with the articulatory rhythm (AR) has not been verified by measurements of neuronal activities. The paper is focused on the goal to review evidence for the following ‘**AC-Hypotheses**’:

H1: The AC is identical for speech perception and production. The AC is based on articulatory features.

H2: The AC is a multi-item neural code constructed by a message with embedded items. Each message relates to a Θ -syllable with embedded sub-codes of elementary articulatory gestures composing the Θ -syllables.

H3: The AC is transmitted using coupled Θ/γ -oscillations.

The paper is organized as follows. Chapter 2 gives a short overview in measuring neural activities. Chapter 3 focuses on the goal to review evidence of the AC-hypotheses.

2 Measurements of Neuronal Activities

To measure the activity of neurons, non-invasive and invasive methods have been developed. Many investigations have been done using non-invasive methods as electro-encephalogram (EEG) or magnetoencephalography (MEG). Due to their limited spatial resolution only the averaged activities of large neuronal areas can be measured. But for deciphering the AC, the activity of many single neurons must be measured synchronously. Recently a new non-invasive method – a brain stimulation method - has been developed, where oscillations as used for transmitting neural codes (see section 3.3) influence the activity of neurons [20]. Yet the field of oscillatory brain stimulation is still in its infancy.

Invasive methods as described by Buzsáki [21] have the needed spatial resolution but the number of neurons to be measured synchronously is still quite restricted: *The invasive method*

¹ The term *ambiguously defined syllable* refers to the concept of syllables as defined by phoneticians.

² This kind of distribution of information needs less energy as the transmission based on dendrites

'Electrocorticography' (ECoG) is becoming an increasingly popular tool for studying various cortical phenomena in clinical settings. It uses subdural platinum–iridium or stainless-steel electrodes to record electric activity - the local field potential (LFP) - directly from the surface of the cerebral cortex, thereby bypassing the signal-distorting skull and intermediate tissue. The spatial resolution of the recorded electric field can be substantially improved ($<5\text{ mm}^2$) by using flexible, closely spaced subdural grid or strip electrodes (see fig. 2 and 3).

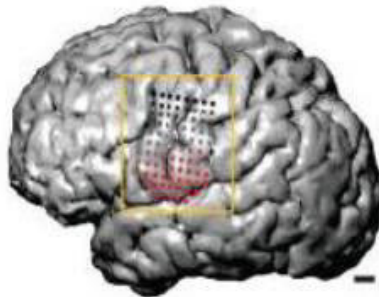


Figure 2: a typical measurement scenario in a clinical setting. The measurements are done on persons during their clinical treatment of epilepsy (Bouchard and et.al.,2013): ECoG measurements in the ventral sensorimotor cortex (vSMC), MRI reconstruction of single subject brain with electrodes (dots); about 30 electrodes were connected to neurons delivering useful information

Electrical events at deeper locations can be explored by inserting metal or glass electrodes, or silicon probes into the brain to record the LFP. Recording the wide-band signal (direct current to 40 kHz) — which contains both action potentials and other membrane potential-derived fluctuations in a small neuronal volume — using a microelectrode yields the most informative signal for studying cortical electrogenesis (see fig.3).

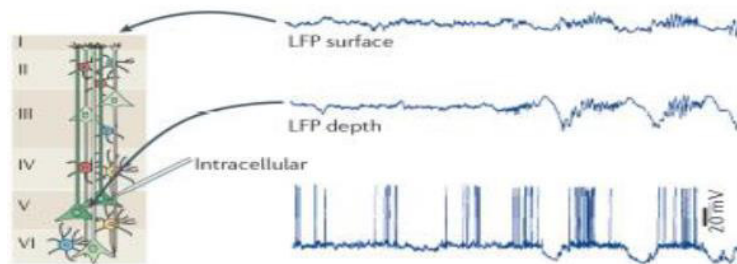


Figure 3 - Simultaneously recorded LFP traces from the superficial ('surface') and deep ('depth') layers of the motor cortex in an anaesthetized cat and an intracellular trace from a layer 5 pyramidal neuron [21].

3 Findings Derived from Cortical Measurements

As pointed out in chapter 1, speech production and speech perception can be regarded as a process of the production and perception of target-oriented gestures (grasping). As discussed in section 3.1 this view has been confirmed by ECoG measurements leading to the components of the AC. The role of coupled Θ/γ -oscillations – the articulatory rhythm (AR) - is investigated in sections 3.2 and 3.3. Section 3.2 focuses on the role of the AR in speech perception. In section 3.3. the processes of transporting the AC between cortical areas is regarded.

3.1 Articulatory Features and Elementary Articulatory Gestures

Mugler et.al. [13] argue, that the steering of the articulators is performed by target-oriented *gestures* and that the steering principles for articulators are the same as for manual target-oriented gestures (grasping). In speech production the targets are specific tuned (*manner*) constriction at specific *places* in the vocal tract described by *articulatory features* as defined in chapter 1. Thus, in speech production the articulatory features are closely related to target-oriented gestures. Due to the mirror principle stating that there exist neurons sensitive in producing and perceiving the same gestures, it is concluded that in speech perception and in speech production the articulator features must be 'mirror features', i.e. they must be common to perception and production. This conclusion is underpinned by ECoG measurements observing the activity of neurons for utterances of syllables as depicted in fig. 4. During speech perception, Mesgarani et.al. [12] measured articulatory features in the STG stating: *Furthermore,*

selectivity of phonemes is organized primarily by manner of articulation distinctions and secondarily by place of articulation, corresponding to the degree and the location of constriction in the vocal tract, respectively. Given these findings, complexes of neurons located in the STG must transform the speaker dependent auditory signal to speaker independent articulatory features. To the authors knowledge this transformation is not deciphered yet. During speech production, Bouchard and Mugler [11,13] measured articulatory features in the ventral sensorimotor cortex (vSMC). During speech perception in the vSMC neurons sensitive to articulatory features have been found [28]. Thus, the vSMC seems to be the location of the ‘mirror-features’, and it is concluded that speech perception and speech production work both with the same set of *articulatory features*. These findings are the basis for the AC-hypothesis H1. Yet, to the knowledge of the author it is not verified by neural measurements, whether the articulatory features are completely identical in speech perception and speech production.

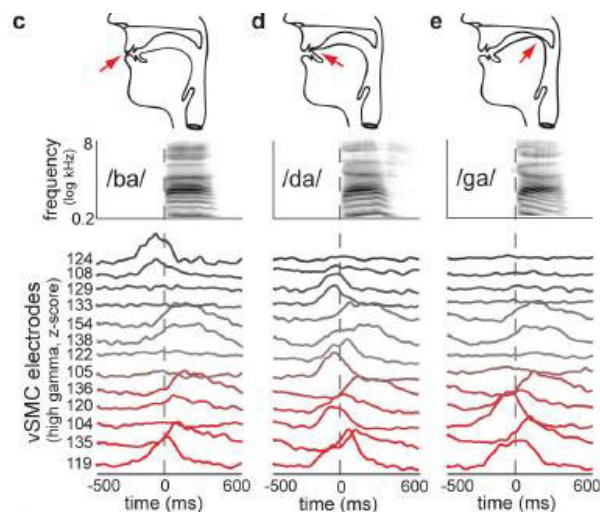


Figure 4 - ECoG measurements in the vSMC during speech production [11]. Rows c, d, e: Above: the articulatory position and the spectra of the utterances /ba/, /da/ and /ga/, with different articulatory features for ‘place’. Below: The z-score of electrodes active for different articulatory places

For steering the movement of articulatory *gestures*, in the vSMC somatotopic ordered populations of neurons³, controlling the muscles of single articulators⁴, have been found. In [11,13] it is discussed, how the motor commands for generating EAGs representing phonetic units are planned. In different cortical areas neurons were found, whose activity hint to complex mechanisms for steering the articulators. Bouchard et al. [11] note: *It is not any single articulator representation, but rather the coordination of multiple articulator representations across the vSMC network that generates speech. Analysis of spatial patterns of activity revealed an emergent hierarchy of network states, which organized phonemes by articulatory features.* Mugler et. al. [13] found in the inferior frontal gyros (IFG) and in the ventral premotor cortex (PMv) neurons related to phonemes⁵. From these findings it can be concluded that the mechanism of the steering the articulators in speech production is based on a transformation of a code defining complex elementary articulatory gestures (EAGs) to the coordinated steering process of *gestures* of single articulators. As the articulatory rhythm is related to the Θ -syllable (see next section), the EAGs must be related to syllables. Höge [15] argued that the Θ -syllable is composed by specific EAGs, which are specific articulatory gestures describing the onset, the vowel-kernel and the coda of a syllable (OVC-gestures). In addition, this ap-

³ Using electrical stimulation, [26, 27] already described the somatotopic organization of face and mouth representations in human vSMC.

⁴ In [2] this somatotopic ordered neural area is used for a BCI.

⁵ This relation between gestures and phonemes are in line with the phonetic framework of scores [22].

proach is in line with structure of multi-item codes (see section 3.2), where the code of the EAGs are sub-codes of the AC.

3.2 The Articulatory Rhythm of the Θ -Syllable

Regarding the functionality of the cortex during speech perception, Giraud et.al. [17] conclude, that cortical processing ‘packages’ the information of the auditory signal into syllables using Θ/γ -oscillations stating: *Recent data show that delta, theta and gamma oscillations are specifically engaged by the multi-timescale, quasi-rhythmic properties of speech and can track its dynamics. We argue that they are foundational in speech and language processing, ‘packaging’ incoming information into units of the appropriate temporal granularity... The faster ‘phonemic’ gamma oscillations are ‘nested’ in the slower ‘syllabic’ oscillations. Through theta-gamma nesting, concurrent syllabic and phonemic analyses can remain hierarchically bound.* Thus, the Θ -oscillations are adapted to the rhythm of the syllables. To entrain the oscillations, edge-features, which determine the start and end of Θ -cycle, must be extracted. For this purpose, in the STG neurons sensitive to the edge feature have been found (e.g. ‘Speech/non-speech-detectors’)⁶. To verify these findings, Hyafil et. al. [23] performed neural simulations to entrain Θ -oscillations using the envelope of the auditory signal segmented successfully the auditory signal into syllables. Further they assumed that the γ -cycles segment the auditory signal into phonemes. Yet the simulation results were poor. Höge [15] assumed that the γ -cycles segment the auditory signal into OVC-gestures. But the results were not promising too. Thus, it seems, that the γ -cycles are not used for segmentation. As discussed in section 3.3, the transmission of the AC is performed by coupled Θ/γ -oscillations. Thus, in perception the use of γ -cycles could be limited to carry the sub-code of the perceived EAGs.

3.3 The Neural Code

As discussed in the introduction specific neural codes are transmitted via coupled Θ/γ -oscillations. As described in [19] this kind of transmission has been observed for steering target-oriented manual gestures. The position of the hand with respect to the goal is stored by place cells located in the hippocampus. The information of the place cells is transmitted by coupled oscillations, where the longer cycles of the low frequency oscillations build a ‘frame’, where the short cycles of the higher frequency oscillations are embedded (see Fig. 5).

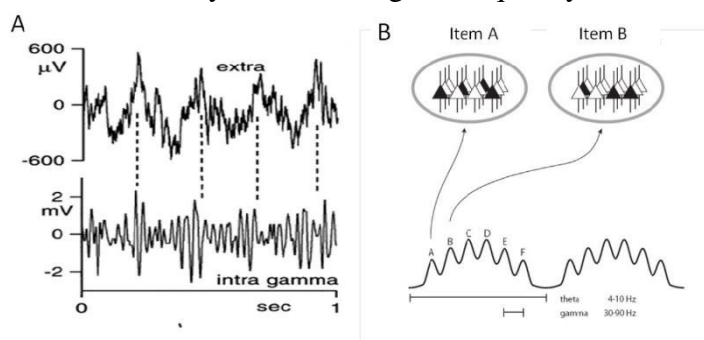


Figure 5 - Neural code organized by theta and gamma oscillations [19]. (A) Simultaneous extracellular (top) and intracellular (bottom) recordings from the hippocampus. Intracellular gamma is due to IPSPs, the amplitude of which is modulated by the phase of theta. (B) Schematic of the theta-gamma code. The ovals at top represents states of the same network during two gamma cycles. Different ensembles are active in different gamma cycles.

⁶ A special topic in the performance of perception is the integration of visual input as lip-reading to improve the edge features and the perception of phonetic units. Due to [24], the integration point is the vSMC. In order to achieve integration, the visual features must be transformed into articulatory features.

The position of a short cycle within the frame determines the position of the place cells. Thus, a frame ports a multi-item message, where the position of an item within the frame is relevant for decoding the message.

In rare cases a multi-item neural code has been deciphered [25]. Due to the fact, that articulatory gestures are goal-oriented gestures and that the steering principles of the articulators are the same along the sensorimotor area, the assumption of hypothesis H2, that the articulatory code (AC) is a special multi-item neural code, is quite probable. Thus, the messages of the AC are Θ -syllables and the items of the AC are sub-codes of EAGs.

4 Conclusion

To verify the AC-hypotheses H1-H3 following conclusions are drawn.

H1: Based on ECoG measurements, during speech perception and speech production neurons sensitive to articulatory features have been found. Additionally, applying the mirror principle of manual gestures to articulatory gestures it can be concluded that the articulatory features are identical in speech perception and in speech production. Still neuronal measurement on this equality are missing.

H2&3: Assuming, that articulatory gestures are controlled and perceived in the same way as manual gestures, there is large evidence that the AC is a multi-item neural code transmitted by Θ/γ -oscillations. Due to the segmentation of the auditory signal into syllables during perception and due to the syllable dependent control of articulatory gestures in speech production, it can be concluded that the messages of the AC are syllables. Further there is large evidence, that the items of the code are EAGs building syllables. The set of EAGs is still unknown and must be deciphered. Due to the plasticity of the cortex this set is learnt in the babble phase of babies and could be speaker specific. Thus, the AC may be not speaker independent but speaker dependent.

The final proof of the correctness of the AC-hypotheses can be given only by measuring simultaneously the activity of many single neurons and the related LFPs active in their areas - mainly in the sensorimotor cortex – a long way to go.

5 References

- [1] Chaudhary, U. et. al.: *Brain-computer interfaces for communication and rehabilitation*. In *Nat. Rev. ANeurol.*, 12, pp. 513-525, 2016.
- [2] Ramsey, N. F., Salari, E., Aarnoutse, E.J., Vansteensel, M. J., Bleichner, M. G. and Freudenburg, Z. V.: *Decoding spoken phonemes from sensorimotor cortex with high density ECoG grid*. In *NeuroImage*, 10.011, pp. 1-11, 2017.
- [3] Darwin, C. (1871). The descent of man. In *Great Books*, Encyclopedia Britannica.
- [4] Kaas, J.H.: *The Evolution of Brains from Early Mammals to Humans*. In *Wiley Interdiscip Rev Cogn Sci*. 4(1), pp. 33–45, 2013.
- [5] Fitch, W.T., de Beur, B., Mathur, N., and Ghazanfar, A.: *Monkey vocal tracts are speech ready; science advances*. In *Science Advances*, 2016.
- [6] Arbib, M. A., Liebal, K., and Pika, S.: *Primate Vocalization, Gesture, and the Evolution of Human Language*. In *Current Anthropology* Vol. 49, 6, 2008.
- [7] Grafton, S. T., Arbib, M. A., Fadiga, L., and Rizzolatti, G.: *Localization of grasp representations in humans by PET. 2. Observation compared with imagination*. In *Experimental Brain Research*. pp. 112:103–111, 1996.
- [8] Goldstein, L., Byrd, D., and Elliot Saltzman, E.: *The role of vocal tract gestural action units in understanding the evolution of phonology*. In Arbib, M., editor. *Action to Language via the Mirror Neuron System*. Cambridge: Cambridge University Press: pp. 215-249, 2006.

- [9] Browman, C.P. and Goldstein L. *Articulatory Gestures as Phonological Units*. In *Haskins Laboratories Status Report on Speech Research*, 99, pp. 69–10, 1989.
- [10] Ladefoged, P., and Johnson, K.: *A Course in Phonetics*. Wadsworth Cengage Learning, 7th Edition, Boston, 2015.
- [11] Bouchard, K. E., and Chang, E. F.: *Control of Spoken Vowel Acoustics and the Influence of Phonetic Context in Human Speech Sensorimotor Cortex*. In *The Journal of Neuroscience*, 34(38): pp. 12662–12677, September 17, 2014.
- [12] Mesgarani, N., Cheung, C., Johnson, K., and Chang, E.F.: *Phonetic Feature Encoding in Human Superior Temporal Gyrus*. In *Science*, 343(6174), pp.1006–1010, 2014.
- [13] Mugler, E.M., Tate, M.C, Livescu, K., Templer, J., Goldrick, M., and Slutzky, M.W.: *Differential Representation of Articulatory Gestures and Phonemes in Precentral and Inferior Frontal Gyri*. In *the Journal of Neuroscience*,38(46), pp. 9803–981, 2018.
- [14] Byrd, D.: *Articulatory Timing in English Consonant Sequences*. In *Thesis UCLA California*, 1994.
- [15] Höge, H.: *Using Elementary Articulatory Gestures as Phonetic Units for Speech Recognition*. In *Proc. Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*, 2018.
- [16] Tilsen, S.: *Multiscale Dynamical Interactions Between Speech Rhythm and Gesture*. In *Cognitive Science* 33, pp:839-879, 2009.
- [17] Giraud, A.L. and, Poeppel, D.: *Cortical oscillations and speech processing: emerging computational principles and operations*. In *Nat. Neuroscience* 15(4), pp. 511-517, 2015.
- [18] Ghitza, O.: *The theta-syllable: a unit of speech information defined by cortical function*. In *Frontiers in Speech psychology*, Article 138, pp.1-5, 2013.
- [19] Lisman, J. E., and Jensen, O.: *The Theta-Gamma Neural Code*. In *Neuron*, 77(6), pp. 1002–1016, 2013.
- [20] Vosskuhl, J., Strüber, D., and Herrmann, C. S.: *Non Invasive Brain Stimulation: a Paradigm Shift in Understanding Brain Oscillations*. *Brain*. In *Frontiers in Neuroscience*, 2018.
- [21] Buzsáki, G., Anastassiou, C. A., and Koch, C.: *The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes*. In *Nat Rev Neuroscience* 13(6), pp.407–420, 2016.
- [22] Nam, H.: *Gestural coupling model of syllable structure*. In *Thesis, Yale University*, 2007.
- [23] Hyafil, A., Fontolan, L. Kabdebon, C., Gutkin, B., and Giraud, A.: *Speech encoding by coupled cortical theta and gamma oscillations*. In *eLife*, 2015.
- [24] Park, H., Kayser, C., Thut, G., and Gross, J.: *Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility*. In *eLife*, pp.1-17, 2016.
- [25] O'Keefe, J., and Dostrovsky, J.: *The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat*. In *Brain Res.* 34, pp.171–175, 1971.
- [26] Penfield, W., Boldrey, E.: *Somatic Motor and Sensory Representation in The Cerebral Cortex of Man Studied by Electrical Stimulation*. In *Brain*, 60, pp. 389–443, 1937.
- [27] Foerster, O.: *The cerebral cortex in man*. In *Lancet*, 221, pp. 309–312, 1931.
- [28] Connie Cheung, C., Hamilton, L., Johnson, K., Chang, E.F.: *The auditory representation of speech sounds in human motor cortex*. In *eLife*, 2016.