TOWARDS A BRAIN-COMPUTER INTERFACE MODELLING THE PHONOLOGICAL SHORT-TERM MEMORY

Harald Höge

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

Abstract: The phonological short-term memory (pSTM) [2] depicts an interface for speech perception and speech production. In speech perception the information provided by auditory and visual (lip reading) sensors must be transformed into information needed for language processing. In speech production the information of language processing must be transformed into information needed for motor control of articulatory movements. Research on Brain Computer Interfaces (BCIs) is driven to develop therapies to correct disordered information exchange in the brain using neuromodulation [1]. For correcting disordered information, the input-output codes of the pSTM, must be known. To understand these codes, the functionality of the pSTM must be known. The paper focuses on how the pSTM encodes (learns) and recalls (perceives) syllables. The pSTM is unique to human restricting invasive measurements as electrocorticography. Thus, the functionality of the pSTM is poorly understood. To increase knowledge, the paper works with following ***pSTM-hypothesis:*** neuronal functionality of the pSTM can be derived from the functionality of the episodic memory (EM), common to all mammals including humans. To provide evidence of this hypothesis, the paper shows the similarity of the tasks to be performed in both memories and the common use of a multi-item code transmitted by Θ-oscillations interfering with the synapses of the items stored in both memories.

1 Introduction

Research on Brain Computer Interfaces (BCIs) is driven to develop therapies to correct disordered information exchange in the brain using neuromodulation [1,3]. The paper supports the development of BCIs paving the way towards a neural model of the phonological short-term memory (pSTM) [2]. The pSTM is needed for speech perception as well as for speech production. During evolution, the pSTM developed together with speech production and speech perception to improve speech communication. Speech communication is a unique ability of humans, thus the pSTM is unique. Because the use of invasive measurements such as electro-cartography (ECoG), intracranial electrodes (iEEG) and deep brain stimulation (DBS) is very restricted to being applied to humans, deciphering the neural functionality of the pSTM has not achieved the maturity needed. For developing BCIs, the exact neural implementation as provided by microcircuits [9] must be known. For this demand most progress is achieved by applying invasive measurements, which have been applied to nonhuman mammals such as rats, rabbits [4]. Due to state of the art, progress in studying complex areas of the human brain is fastest, when the human areas are common for nonhuman mammals. In this case invasive measurements can be performed on nonhumans and the results are transferred to the human brain. The paper follows this approach to gain knowledge about the pSTM with following **pSTM-*hypothesis:*** the functionality of the pSTM can be derived from the neural functionality of the episodic memory (EM) common to all mammals including humans [10]. Using the pSTM-hypothesis the knowledge gained from the EM can be transferred to the pSTM. The paper is focused on providing evidence of this hypothesis. For both memories, in sections 2 and 3 the functionality, state-of-the art-models of architecture and the role of the Θ-oscillations are described. Section 4 is focused on providing evidence of the pSTM hypothesis.

Both memories, EM and pSTM exchange information of stored items via a multi-modal code [5] describing different modalities of an item (e.g. a multi-modal code of an item with two modalities: the auditory loudness and the visual brightness). The information of each modality has its own code encoded as modulation of a cycle of a γ-oscillation. For a given item, γ-cycles of all modalities are embedded in a single Θ-cycle (see fig.1 taken from [22]; all codes of each modality are stored within a specific modality-memory).

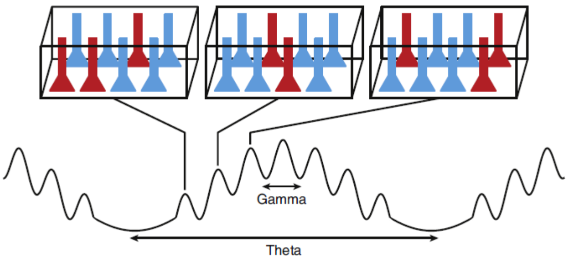

Figure 1 - Θ – γ oscillations in working memory. Θ – γ gamma coding scheme by which multiple memories are actively stored in the same network. Different memory items (3 are shown) are represented by different groups of active cells (black) that fire in different gamma cycles of a theta cycle. The different memories thus have different theta phase. If the multiple items are being held in short-term memory, the entire pattern repeats on subsequent theta cycles.

Within memory, N-methyl-d-aspartate receptors (NMDARs) perform long term potentiation (LTP) of synapses changing the strength of synaptic connections between the items. NMDARs are sensitive to the information encoded in Θ-oscillations. This mechanism is the basis for learning, where synaptic connections between items are strengthened or even created, whenever a sequence of perceived items encoded by the Θ-cycles fits to items stored in the memory. Within the hippocampus the role of NMDARs has been studied for humans and nonhumans leading to the conclusion, that. NMDARs are crucial for regulation of excitatory synaptic transmission and synaptic plasticity [13].

2 The Episodic Memory, and the Role of Θ-Oscillations

The human EM is located in the hippocampus of the medial temporal lobe (MTL) [10]. Research for providing the functionality of the human EM started with lesion models derived from people with specific lesions. Those models show that the EM is critical for the recollection of past personal experiences and the retrieval of previously learned facts about general knowledge [10]. The process of decoding and recalling information from the EM is not just memory access via an address of an item stored as done for the memories used by a computer. Instead, the EM must handle as input sequences of snapshots of the sensors describing the environment and must provide as output a sequence of codes describing the sequence of snapshots. Each snapshot is given either by a single item or by different connected items. There is high evidence that the code of each item is a multimodal code [5] as described in the introduction. There exist several theories concerning the nature of the items and how the items are connected. In the community of neuroscience regarding liaison models, the structural nature of the items is termed the 'neural code', where specific neurons spike whenever a specific item of a episodic snapshot is present [11]. Further it is hypothesized that the items could be abstract constructions (e.g. a 'friend'-neuron spikes, whenever this friend is part of the snapshot independent from the specific environment recorded by the visual sensor). As described in [4] the temporal/spatial order of the sequence of the snapshots of an episode stored in the EM is realized by specific synapses connecting neighboring snapshots of an episode. Whenever a neural ensemble realizing a specific snapshot is activated, the ensemble of the following snapshot is activated.

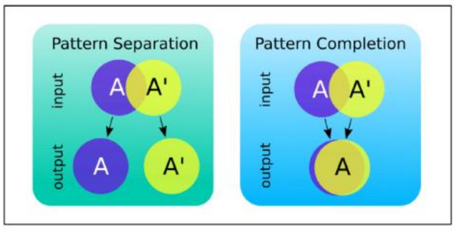

Figure 2- Concept of pattern separation and completion taken from [12].

Pattern separation makes similar input patterns (i.e., A and A') more distinct and reduces the overlap between representations. In contrast, pattern completion increases the overlap between two similar input patterns.

The timing of the activation process is steered by phases of the cycles of Θ-oscillations generated in the hippocampus [5,10]. The sequence of the phases is called precession [27]. Another aspect of the capabilities of the EM is given by the process of learning of new items and the detection of similar items. As shown in fig.2 two capabilities termed pattern completion and pattern separation are active in the EM.

2.1 The Architecture of the EM

The first neural architecture of the EM was presented in 1994 [4]. Fig.3. depicts the architecture of the EM given by its components, and their connections. In [4] the functionality of the EM is characterized as follows: *The neural ensemble CA3 is an auto associative memory and works together with the DG, and CA1. Via the entorhinal cortex the EM receives and transmits the information representing the episodes from respectively to the neocortex. In terms of cell numbers, information appears to be funneled from DG through the CA3 bottleneck and then spread out again into CA1. The output of CA1 returns via the subiculum to the entorhinal cortex, from which it is redistributed to neocortical areas.*

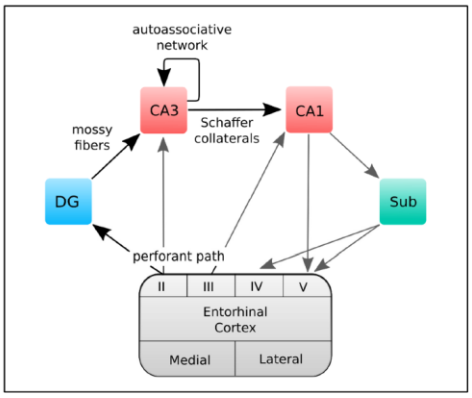

Figure 3 - Architecture of EM as presented in [12]

The basic components of the EM are the DG (dentate gyrus), the memories CA1 together with CA3, and the SUB (subicular complex). Output and input of the EM is the entorhinal cortex (EC), which is the interface to many regions of the neocortex especially to all regions delivering the information provided by the sensors (visual, auditory, emotion, odor). In figure 3 the input and output of the EM with the EC is shown in more detail given by connections to or from the layers II-V of the EC (For all mammals the structure of the EC and the neocortex is described by 6 denoted as layers I-VI. Layer I is near the scalp, whereas layer VI is deep in the cortex).

Nowadays more detailed functionality of the components of the EM has been deciphered (see the following subsections 2.1.1-2.1.2)

2.1.1 The DG

In the following the term ‘pattern’ is identical to the term ‘item’. The patterns are stored in *Dentate Granule Cells* (GC) [14]. As described in [14] the role of the DG can be stated as follows: *In the feed-forward pathway from the EC → DG → CA3 synaptic transmission and plasticity have been implicated in pattern separation based on the spatial and temporal segregation of the initially overlapping EC memory engrams* (engrams are items describing parts or the complete snapshot). In [14] two kinds of GCs are regarded building the set of the GCs: young and old GCs. The young GCs represent GCs, which are created for new patterns not yet observed. Thus, the set of GCs increases fast especially for young mammals. The young GCs are very specific, as they are derived from a specific snapshot. An old GC gathers all the patterns, which represent similar patterns. Thus, old GCs represent a more abstract presentation for cluster of patterns. To handle the properties of the GC, the DG must handle several tasks:

- The DG must segregate patterns (segmentation task)
- The DG must be able to generate new GCs (creation task)
- The DG must be able to decide if a pattern is a candidate for a new GC or if it belongs to an old GC (recognition task)

How the DG solves these tasks is not yet fully understood. NMDARs play an important role in performing these tasks e.g. in pattern segregation [14].

2.1.2 Memories CA3 and CA1

Haselmo [15] and Buzsáki [8], were pioneers in developing neural models of the EM. In [8] the hypothesis is set up that in evolution navigation in the real world of nonhuman mammal developed to navigation in the mental space: *we propose that mechanisms of memory and planning have evolved from mechanisms of navigation in the physical world and hypothesize that the neuronal algorithms underlying navigation in real and mental space are fundamentally the same. We review experimental data in support of this hypothesis and discuss how specific firing patterns and oscillatory dynamics in the entorhinal cortex and hippocampus can support both navigation and memory.* Thus, the mechanisms of the nonhuman EM are like the mechanism of the human EM. Haselmo [15] argues in the same line. He gives a more detailed mechanism of mental navigation including the information of specific neurons located outside the EM: *Entorhinal grid cells integrate the head direction input to update an internal representation of location and drive hippocampal place cells.*

[16] describes the memories CA3 and CA1 in more detail saying: *We showed that activation of CA3 ensemble recalls the memory and that synchronous activation of distinct cell populations in CA3, which corresponds to distinct events links these initially independent events. These findings suggest that the CA3 recurrent circuit expressing NMDARs mediates artificial association of memory events stored in CA3 ensembles.*

NMDARs in CA1 are involved in memory acquisition, indicating that NMDARs in the hippocampus are essential for multiple brain functions [26].

2.2 The Role of the Θ-Rhythm in the EM

In 1954 regular waves of 5-7Hz have been detected in the hippocampus of a cat [17]. Nowadays these oscillations are called Θ-*rhythm.* In 1978 O'Keefe and Nadel [18] measured the activity of the hippocampus of rats performing a navigation task for food-finding. They detected a cognitive map of place cells, where each cell is sensitive to a specific position of the rat. The activation of the place cells is triggered by phases of the Θ-*rhythm.* This property of the Θ-*rhythm* has been named *Phase Precession* [Buzsáki, 2002]. In the learning process the rat stores different ways (episodes) to the food in the EM. In the finding process the rat uses the visual cues provided by the current environment and compares these with the snapshots derived from the episodes stored. These two discoveries, the Θ-rhythm and the map, are the basic ingredients for understanding the functionality of the EM. As mentioned in [8], the two discoveries [17,18] are the basis of fundamental properties of memories. Thus, I conclude that EM can be seen as a blueprint for other memories and thus, this statement is in line with the pSTM-hypothesis.

3 The pSTM

The items stored in the pSTM are syllables. In [3] it is hypothesized that the cognitive units are syllables, which are encoded in the same way for speech perception as well for speech production. The neural code of a syllable is termed the AC-code. As described in [3,6] each syllable is represented by a multi-item code (see fig.1) called AC-code representing articulatory gestures given by the onset, kernel, coda of a syllable and its emotional state.

3.1 The Architecture of the pSTM

In various publications the term *phonological short-term memory* (*pSTM*) is called by different names as *phonological working memory* (pWM) [20]. According to [20], the understanding of the neural functionality of pWM is poor, because speech processing belongs to human cognitive abilities with high complexity, where multiple systems work together. The task of the pSTM is to handle items concerning speech perception, speech production, and language processing. The most widely used functional model for the pSTM is Baddeley's theory of working memory called the phonological loop [21]: *The model comprises a phonological store, which can hold memory traces for a few seconds before they fade, and an articulatory rehearsal process that is analogous to subvocal speech* (fig.4). *Memory traces can be refreshed by being retrieved and re-articulated. Immediate memory has a limited span because articulation takes place in real time - as the number of items rehearsed increases, it reaches a point at which the first item will have faded before it can be rehearsed. Evidence for the role of articulation comes from the word-length effect: immediate memory span declines as word length increases from one to five syllables. The suggestion that this reflects the slower rehearsal of interpretations assume that each item forms a cue or stimulus for the following item, with the result that once the initial item is activated, the sequence runs off relatively automatically*.

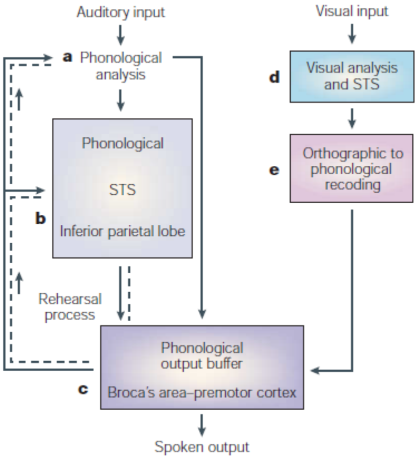

Figure 4 - A functional model of the phonological loop [21].

a Phonological analysis.

b Short-term storage (STS).

c Programming of speech output.

d Visual encoding.

e Grapheme-to-phoneme conversion. Auditory information gains direct access to a phonological store, after which it passes to an output buffer for recall or recycling through rehearsal. Visual material can be recorded verbally and then gain access to the phonological store through rehearsal.

3.2 The role of the Θ-Oscillations in the pSTM

In speech processing the duration of a Θ-cycle describing a multi-modal code of a syllable has the duration of the related syllable. Thus, the frequency of the Θ-rhythm is dependent on the speaking rate. It is speculated that the information carried by one Θ-cycle is encoded in four γ-cycles; three γ-cycles represent three articulatory gestures given by the onset, kernel, and coda of a syllable and the fourth γ-cycle represents the emotional state [3]. In [23], it is speculated that phases of the starting points of the γ-cycles are the starting point of the three related articulatory gestures. Further it is speculated that these starting points are acoustic cues specifically produced by the articulators to determine together with the phonetic content the temporal structure of a syllable.

A still unsolved question is the neural position and functionality of the Θ-oscillator. The Θ-oscillator must generate a Θ-cycle fitting to the start and duration of the related syllable. A plausible position is the hippocampus, where many oscillations are generated. As measured in [18] the frequency of the Θ-oscillator depends on the speed of rats searching food. In speech production the equivalent speed is the speed of the articulators. The emotional memory is also located in the hippocampus [3]. It seems that speed is also an emotional feature. In [24] in the hippocampus as well in the vSMC synchronous Θ-cycles were observed during speech production. Surprisingly the information carried in a Θ-cycle increased substantially during the trip from the hippocampus to the vSMC. This increase can be caused by following mechanism: In the hippocampus the Θ-oscillator is steered by emotional status including the speed feature leading to a Θ-cycle including the single γ-cycle encoding emotion. The Θ-cycles are transmitted to the vSMC via the pSTM, where the three γ-cycles needed to complete the AC-Code are added.

4 Providing evidence of the pSTM-Hypothesis

EM and pSTM are found in different areas of the brain: EM the hippocampus, pSTM in the inferior parietal lobule. The similarity of pSTM with EM can be explained by following strategy observed in evolution:

Due to the similarity of their functionality, the functionality of the pSTM can be derived from the 'copy and paste' strategy of evolution. During evolution the development of new skills of the brain (e.g. gathering more food) is not done by developing completely new neural ensembles but by a 'copy and paste' strategy: existing neuronal ensembles are copied and placed near old neuronal ensembles achieving improved capabilities. In evolution this strategy has been proven to be the fastest method to develop new skills. As the pSTM is younger concerning evolution, the functionality of the pSTM may be an adapted copy of the functionality of the EM.

4.1 Similarities in Tasks

In the pSTM the sequence of syllable plays the role of the sequence of the snapshots of the episodes stored in the EM. Further the items stored are encoded by a multi-item code and transmitted via Θ-cycles. The duration of a Θ-cycle active in the EM is given by the local speed of the items. In the pSTM the duration is given by the speaking rate i.e. the speed of the articulators. Thus, the speed determines the frequency of the Θ-cycles.

These findings lead to the conclusion that both memories have the same task: storing sequences of items related to the timing of the items. The memory must enable the recall of sequences and the handling the creation of new sequences.

Both memories, pSTM and EM, must provide a mechanism to extract information given the input of sensors and transform ‘snapshots’ of this information as sequential stored items. In the EM the transformation of the snapshots to items is performed by the DG (see 2.1.1), which detects fitting patterns taken from a set of patterns stored in the DG. Each item is encoded by fitting patterns. To detect fitting patterns, the DG performs three tasks: the segmentation task, the creation task and the recognition task (see 2.1.1). These 3 tasks must be performed also within the pSTM. Due to [3] the patterns are articulatory features, extracted from segmented auditory as well extracted from segmented visual input (segmentation task). In a learning process those patterns must be found, which belong to the language spoken. The found pattern (creation task) relates to the articulatory gestures building syllables of that language. According to fig. 4 these tasks must be performed by phonological analysis. In speech perception and speech production the neural mechanisms determining the timing of the movements of the articulators is not well understood (see conclusion).

4.2 Similarities in Architecture

Comparing phonological loop (fig.4) with architecture EM-architecture (fig. 3), it seems that the STS plays the role of CA3, where the recursive links are provided via the rehearsal process and the phonological output buffer takes the role of CA1. In fig. 4 the box ‘**c** Programming of speech output’ has at the left the output the information of the rehearsal process. This information can be interpreted as the sequence of AC-codes of the perceived auditory (speech)-input. These AC-codes enter the rehearsal system for planning the motor programs to steer the articulators (in rehearsal the programs are not executed). In parallel the sequences of the AC-codes perceived are fed back as sequences stored in the STS as done by the recursive links of the CA3.

In the feat-back the AC- sequences are channeled through the STS, where the path is realized by the synaptic connections between the neurons sensitive to the sequence of AC-codes.

Further for speech perception, the articulatory features extracted from the auditory input and the visual input (lip-reading) are combined in the ventral sensorimotor cortex (vSMC), where the combined articulatory features are transformed into the AC-code [3]. Thus, in fig. 4 the action ‘a’ performing Phonological analysis, including auditory and visual information delivers as output the AC- code.

5 Conclusion

According to state of art, from both memories EM and pSTM, the architecture and the role of the Θ-oscillation influencing NMDARs are described. To provide evidence of the pSTM-Hypothesis the strategy of evolution, the similarity of the tasks and the role of the Θ-oscillation are given.

To increase the knowledge of the pSTM concerning speech perception and speech production, some functionalities of the EM are of interest:

- The functionality of the DG can be transferred to speech perception, because the ability to perform pattern segmentation and classification without using a language model is not yet solved.
- The mechanism of precession used in the EM can be transferred to speech production, because in speech production the control of timing of the articulators is understood only partly.

Still the neural functionality of the EM has to be more deciphered in detail and the neuronal ensembles of the pSTM corresponding to those of the EM have to be identified.

6 References

[1] CHAUDHARY, U. et. al.: *Brain-computer interfaces for communication and rehabilitation*. In *Nat. Rev. Neurol.*, 12, pp. 513-525, 2016.

[2] ROMERO, L., V. WALSH, AND C. PAPAGNO: *The Neural Correlates of Phonological Short-term Memory: A Repetitive Transcranial Magnetic Stimulation Study*. In *Journal of Cognitive Neuroscience* 18:7, pp. 1147-1155, 2006.

[3] HÖGE, H.: *Towards a Brain Computer Interface for Speech Perception*. In *Proc. ITG*, 2023.

[4] TREVES, A., and E.T. ROLLS: *Computational Analysis of the Role of the Hippocampus in Memory*. In *Hippocampus*, Vol. 4, No. 3, pp. 374-391, 1994.

[5] LISMAN, J. E., and O. JENSON: *The Theta-Gamma Neural Code*. In *Neuron,* 77(6), pp.1002-1016, 2013.

[6] HÖGE, H.: *The nature of the articulatory code*. In *Proc. ESSV2020*, 2020.

[7] HÖGE, H.: *Cortical Segmentation of Syllables based on Phases of Θ-Cycles*. In *Proc. ESSV2025*, 2025

[8] BUZSÁKI, G. AND E.I. MOSER: *Memory, navigation and theta rhythm in the hippocampal-entorhinal system*. In *Nature Neuroscience* 16, pp. 130-138, 2013.

[9] CUTSURIDIS, V., STUART C., AND B. P. GRAHAM: *Encoding and Retrieval in a CA1 Microcircuit Model of the Hippocampus*. In *ICANN* 2008, pp. 238-247, 2008.

[10] SQUIRE L.R, C.E.L. STARK, AND R.E. CLARK: *The Medial Lobe.* In *Annu. Rev. Neurosci.* 27, pp.279-306, 2004

[11] KOLIBIUS, L.D., S.A. JOSSELY, AND S. HANSLMAYR: *On the origin of memory neurons in the human hippocampus*. In *Trends in Cognitive Sciences*, Vol. 29, No. 5, 2025.

[12] HANERT, A.K.: *Pattern Separation in the Human Hippocampus: Natural Lesion Models and Sleep-dependent Memory Consolidation*. In *Dissertation Philosophischen Fakultät der Christian-Albrechts-Universität zu Kiel,* 2019.

[13] NAKAZAWA K., T.J. MCHUGH, M.A. WILSON, AND S. TONEGAW: *NMDA receptors, place cells and hippocampal spatial memory.* In *Nat Rev Neurosci.*5(5) pp.361–72, 2004.

[14] NAKASHIBA, T., J. D. CUSHMAN, K.A. PELKEY, S. RENAUDINEAU, D.L. BUHL, MCHUGH, T.J. BARRERA, V.R. CHITTAJALLU, R. IWAMOTO, C.J. MCBAIN, M.S. FANSELOW, AND S.TONEGAWA: *Young Dentate Granule Cells Mediate Pattern Separation, whereas Old Granule Cells Facilitate Pattern Completion.* In *Cell,* 149 pp.188-201, 2012.

[15] HASSELMO, M.E.: *A model of episodic memory: Mental time travel along encoded trajectories using grid cells.* In *Elsevier: Neurobiology of Learning and Memory*, 92, pp. 559–573, 2009.

[16] NOMOTO, M., N. OHKAWA, K. INOKUCHI AND N. OISHI: *Requirement for hippocampal CA3 NMDA receptors in artificial association of memory events stored in CA3 cell ensembles*. In *Molecular Brain*, 16:12 ,2023.

[17] GREEN, J.D., AND A.A. ARDUINI: *Hippocampal Electrical Activity in Arousal*. In *Neurophysiology,*17: pp. 533-557, 1954.

[18] O'KEEFE, J. AND L. NADEL: *The Hippocampus as a Cognitive Map*. In *Oxford Univ. Press*, NewYork,1978.

[19] Herweg N.A., E.A. Solomon, M.J. Kahana: *Theta Oscillations in Human Memory*. In *Trends in Cognitive Science*. 24, S. 208-227, 2020

[20] SCOTT, T.L., T. K. PERRACHIONE: *Common cortical architectures for phonological working memory identified in individual brains*. In *NeuroImage* 202, 2019.

[21] BADDELEY, A.: *Working memory: looking back and looking forward*. In *Nat. Rev. Neusci.4*, pp. 829–839, 2003.

[22] LISMAN, J.: *Working Memory: The Importance of Theta and Gamma Oscillations.* In Current Biology 20 (11), 2010.

[23] Höge, H.: *Cortical Segmentation of Syllables based on Phases of Θ-Cycles*. In *Proc. ESSV2025,* 2025.

[24] CHRABASZCZ, A., W. J. NEUMANN, O. STRETCU, W. J. LIPSKI, A. BUSH, C.A. DASTOLFO-HROMACK, D. WANG, D. J. CRAMMOND, S. SHAIMAN, M. W. DICKEY, L. L. HOLT, R.S. TURNER, J. A. FIEZ, AND R.M RICHARDSON: *Subthalamic Nucleus and Sensorimotor Cortex Activity During Speech Production*. In *The Journal of Neuroscience,* 39 (14), pp. 2698 –2708, 2019.

[25] RAGHAVACHARI, S., J.E. LISMAN, M. TULLY, J.R. MADSEN, E.B. BROMFIELD., and M.J. KAHANA: *Theta Oscillations in Human Cortex During a Working-Memory Task: Evidence for Local Generator*s. In *J. Neurophysiology* 95, pp.1630–1638, 2006.

[26] TSIEN, J.Z., P. T HUERTA, AND S. TONEGAWA: *The Essential Role of Hippocampal CA1 NMDA Receptor–Dependent Synaptic Plasticity in Spatial Memory.* In *Cell,* Vol. 87, pp.1327–1338, 1996.

[27] ZHENG, J., YEBRA M., SCHJETNAN A.G.P, PATEL K., KATZ C.N., KYZAR M., MOSHER C.P., SUNEIL K., CHUNG J. M., REED C. M., VALIANTE T. A., MAMELAK A.N., KREIMAN G. AND RUTISHAUSER U. *Theta phase precession supports memory formation and retrieval of naturalistic experience in humans*. In *Nature Human Behaviou*r V.8, pp.2423–2436, 2024.