CORTICAL SEGMENTATION OF SYLLABLES

Harald Höge

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

Abstract: In the early times of automatic speech recognition, bottom-up segmenting of speech into syllables has been investigated. But this approach was not competitive to current solutions. Recent cortical measurements lead to the conclusion, that evolution has found a neural implementation, which perform reliably a bottom-up segmentation of the auditory signal into syllables. This segmentation is based on θ – oscillations, where the duration and position of each syllable is related to the duration and position of each cycle of the θ-oscillations [6]. For transporting information and steering rhythmic tasks, θ oscillations were observed in many locations of the brain, especially in the thalamus. Yet, nor the cortical location of the θ-oscillator for segmentation of syllables, nor the implementation of the θ-oscillator itself is known. Neural models of θ-oscillators for syllable segmentation are scarce. We follow the approach [9], where the θ-oscillator is built up by PING microcircuits. In [9] these PINGs are driven by the sum of auditory signals given in each critical band (CB). In this paper this approach is extended by using onset edge features extracted in CBs. First experiments showed that the CB-PINGs deliver spikes related to the onset of syllables, which corresponds to a specific phase of the θ-cycles. The timing of the spikes from the CB-PINGs differs slightly. By interconnecting appropriate CB-PINGs, it seems possible to reduce the differences leading to a ‘unified’ θ-oscillation.

1 Introduction

The syllable is basic for human speech perception [24]. In the auditory and sensory motor cortex (vSMC), the auditory features are transformed to an articulatory code related to articulatory gestures building the onset, the middle and the coda of syllables. This code is transported to the short-term memory, thus acting as a syllabic interface for cognition. This approach needs a mechanism to segment the auditory signal into chunks of syllables. Due to recent progress in measuring the activity of population of neurons, it seems, that evolution has found a solution to perform reliably a bottom-up segmentation of speech into syllables [6]. This segmentation is based on θ – oscillations, where the duration and position of each syllable is related to the duration and position of each cycle of the θ-oscillations.

In the 1980th, when automatic speech recognition systems started to be developed, it was aimed to mimic this bottom-up approach in segmenting speech into phonetic units [4]. This approach was implemented by ‘knowledge-based rules’ by inspecting the short-term spectra of speech [2]. In [1], the rules for detecting ‘syllable-initial stops’ indicating the onset of a syllable are evaluated. Also, the human brain detects such onset events within a syllable called edge features (see chapter 2). As described in chapter 3, these features are input to an θ-oscillator generating θ-oscillations to segment the auditory signal into syllables[1]. Yet still nowadays, no competitive algorithm mimicking the human approach has been found. Instead, the borders of phonetic units are determined by an algorithm, which combines acoustic features processed bottom up with top-down knowledge given by a language model [7]. In this approach, there exist no interface for syllables.

[a]o The θ-oscillations are used also to transport codes [25]

In the brain of humans and animals, especially in the thalamus, various locations have been detected, where oscillations are involved in transporting information and steering quasi-rhythmic actions [21]. Yet for solving the task ‘segmenting the auditory signal into syllables’ nor the neuronal location nor the detailed functionality of an θ-oscillator generating the θ-oscillation for this task is known.

This lack in knowledge is caused mainly by the immature technology in measuring in vivo the sensitivity of neuronal populations stimulated by speech. For such measurements high spatial and temporal resolution is needed, which nowadays can be provided by invasive measurements only e. g. cortical electrocorticography (ECoG) performed in clinical settings. Yet the spatial resolution of the electrode-pads measuring the activity of neurons is strongly limited, as the electrodes are spaced in large distances compared to the size of neurons. 4mm distances of electrodes have been reported in recent papers [8], [10]. Those distances lead to very sparce sampling of the activity of populations of neurons[2]. Thus, the knowledge to investigate the functionality of neuronal populations, is based on measuring the activity of few neurons, connected to electrodes by chance. Thus, most neuronal models are based on hypotheses, whose evidence is concluded from the measurements available and from other knowledge sources.

In this paper, where we propose a model of an θ-oscillator segmenting the auditory signal into syllables, we use following hypotheses:

H1- The θ-oscillator is built up by *inter-neuronal network gamma* (PING) microcircuits driven by features related to the envelope of the auditory signal.

H2- The features are based on the gradient of the envelope leading to *envelope features* and are based on the maxima of the gradient of the envelope leading to *edge features*.

H3- The features are extracted within critical bands (CB) leading to an architecture of the θ-oscillator producing oscillations for each critical band (concept of CB-PINGs).

Models for simulating a θ-oscillator segmenting the auditory signal into syllables are rare. We follow the biophysically inspired model described in [9], which uses hypothesis H1 but not H2 and H3. The evidence of the hypotheses H2 and H3 is based on following findings:

H2: evidence of the existence edge features is given by psycho-acoustic findings. Huggins [3] noted: …*The results suggest that the perception of timing in natural speech is based on events at the syllabic level rather than at the segmental level, and that it is important to maintain the rhythm of the sentence, as defined by the onsets of vowels (especially stressed vowels), if the sentence is to sound temporally fluent.* These psychoacoustic findings are in line to recent neuronal measurements in the superior temporal gyros (STG). Populations of neurons have been detected, which are sensitive to edge-features related to the maximal energy rise of the envelope of the auditory signal during the onset of vowels [8]. But the measurements provide not the information needed to simulate exactly the functionality of the edge-features. For simulation, in chapter 2 a ‘plausible’ implementation for the edge features is described. Due to the conclusion given in [8], envelope features seem not to be involved in the θ-oscillator.

H3: evidence for the CB-approach is based on psycho-acoustic experiments, where the envelope within critical bands is manipulated. Decreasing the modulations in critical bands [5, 22] showed specific decrease in the intelligibility of speech. Thus, the ‘correct’ temporal structure of the syllabic envelopes within critical bands is essential for perception. Consequently, we conclude, that along the ‘perception path’ the processing of features is performed within critical bands.

Yet, as in [9], it is argued that the θ-oscillator is based on a broad-band analysis derived from the envelope of the complete auditory signal. The main argument for a broad-band approach is given by the view, that the oscillations are generated without spectral analysis of auditory

[2] Typically, in a 2mm^2 area about 100 000 neurons are active.

signal. Further the broadband approach has the advantage that only a single θ-oscillation is generated by the θ-oscillator. Still another view is given in [16, fig.1], where it is hypothesized that the 'final' θ-oscillator is located in the ventral sensory motor cortex. This θ-oscillator could be driven by features windowed by CB-specific oscillations. Thus, this θ-oscillator has a distributed architecture.

2 The Features Driving the θ-Oscillator.

Human auditory features are extracted along the auditory pathway starting at the hair cells located at the basilar membrane and ending at the auditory cortex. On this path the spectra-temporal properties of the auditory signal provided by the hair cells is performed leading to *primary* features. The processing of the primary features is done within critical bands (CBs) by populations of neurons ordered tonotopically in the inferior colliculus (IC) located in the midbrain [11]. The *primary* features are tuned to all kinds of sounds important for reacting acoustic events in the environment. Precise models for simulating the primary features have been developed [20] based on measurement of mammals, which process the auditory signal in the same way. The auditory path ends with the transport of the primary features to the belt of the auditory cortex located in the Superior Temporal Gyrus (STG). In this human specific area, a transformation of the primary features to features tuned to the sounds of speech is performed. Recent ECoG measurements showed that from the posterior to the anterior axis of the STG two types of spectral-temporal features are located [10]. The first type of features, whose neural populations are located anterior, are sensitive to slow temporal modulation and to detailed spectral resolution. The second type of features, whose neural populations are located posterior, are sensitive to fast temporal modulation and poor spectral resolution. We assume that the edge features driving the θ-oscillator belong to this kind features.

The edge features are derived from the gradient of the envelope curve of the auditory signal extracted in critical bands (see fig.1). The instance and strength of the maximal rise of the envelope curve (maxima of the gradient curve) given at the onset of a vowel as measured in the STG [8] define the edge features (in the current implementation the strength is not used).

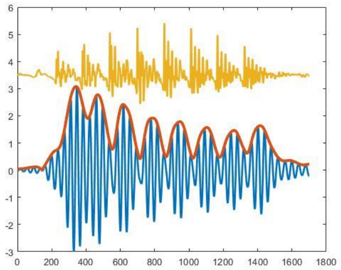

Figure 1 - phone */ae/* from the utterance 'Jack Webster'*;* top: speech signal; below auditory signal and its envelope analyzed for a critical band filter with center frequency 1409 Hz.

To implement the envelope features (not used in the current implementation of the θ-oscillator) and edge features we start to compute the primary features [13] based on the model [15]. Human primary features are generated in 24 critical bands, where in each band the AM-FM modulation of the auditory signal is extracted. Based on Fletchers importance function [12] perception of phones takes place in the frequency range of f_l = 100 Hz till f_u=9000 Hz covering the range of 21 critical bands. As the experiments (see chapter 4) are performed with speech sampled with 16kHz, we restrict the processing of the primary features to 19 critical bands covering a range of 3-21 Bark with center frequencies 248, 328, … 6330, 7423 Hz. To determine the edge features only the envelope of the auditory signal is used, and the spectral information is discarded. The neural implementation of the gradient of the envelope-curve is not known. We use as gradient the difference from neighbored samples. Samples of a smoothed gradient curve are processed within a 'frame' of 10ms. As concluded in [8] the neurons sensitive to edge features indicate the maximal rise of the syllabic envelope at the onset of vowels. To simulate this property, we determine the maxima of the samples of the smoothed gradient curve extracted in regions with positive gradient. Instead using the broad band envelop of the auditory signal as done in [8], the envelope curves are processed within critical bands. Thus, for each critical band edge features are extracted. As shown in chapter 4, at the instances of a maxima given by the edge feature, a spike is sent to a synapse driving the θ-oscillator.

3 The Θ – Oscillator

In section 3.1. the neural architecture of the θ-oscillator is described. The components of the architecture are interconnected PING microcircuits [17] driven by CB extracted edge features. This approach leads to the concept of CB-PINGs, where CB-specific PINGs are implemented. The simulation of the activity of the neurons constituting the CB-PINGs is performed by an approximated model of the biophysical Hodgkin-Huxley model as described in section 3.2.

3.1 The Architecture of the Θ – Oscillator

As shown in fig.2, the architecture of the Θ - oscillator is determined by the composition of microcircuits of CB-PINGs. The CB-PINGs are realized by PINGs as shown in fig. 3. Each CB-PING is related to one of the critical bands CB3 – CB21, and each CB-PING is driven by the envelope features and edge features extracted in this band (in the current implementation the envelope features are not used). The CB-PINGs generate as output θ-spikes $\theta_k, k = 3, \dots, 21$.

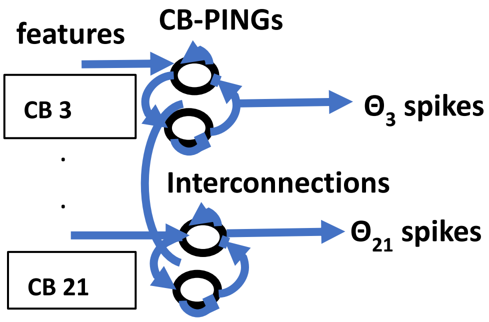

Figure 2 - Architecture of the θ-oscillator. The interconnections denote synaptic connections between Te/Ti neurons from different CB-PINGs.

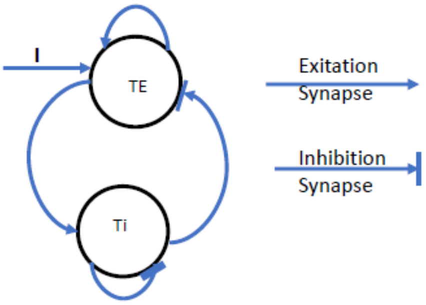

Figure 3 - A PING is realized by two coupled populations of neurons - an excitatory population (Te) and an inhibitory population (Ti), where each population is modeled by a single neuron (see section 3.2).

The output of each CB-PING should be a spike positioned at the onset of a syllable at the instance with largest increase of the CB-envelope at the onset of vowel. These instances can be related to the instances of phases $\omega t = 0, 2\pi, \dots$ of a sinus-wave $\theta = \sin(\omega t)$, because on these instances, the sinwave rise is maximal. The instant frequency $\omega(t)$ is given by the distances between neighbored spikes. The output of a CB-PING related to a CB_k are $\theta_k - spikes$ corresponding to sine wave oscillation $\theta_k, k = 3, \dots, 21$.

As described in section 3.2 the generation of the $\theta_k - spikes$ depends on the current flows within the Te_k-cells. The synaptic interconnections between the CB-PINGs influence these currents strongly. Changes in the strengths of the currents lead to changing instances of the $\theta_k - spikes$.

3.2 An Approximative Hodgkin-Huxley Model

The Hodgkin-Huxley model [17] determines the potential u of the cell-membrane of a neuron given by a first order differential equation: $\frac{du}{dt} = f(u, I^{syn}, I^{INP}, I^{DC})$ (1)

depending on u and the current flows I^{SYN}, I^{INP}, I^{DC}. The current flows are ion flows, which enter or leave the cell via gates. I^{SYN} denotes the flow induced by synapses and I^{INP} denotes the flow induced directly at the gates. I^{DC} is a constant current flow induced within a gate. In the original Hodgkin-Huxley model, each flow is dependent on complex gating variables leading to spikes with time dependent shape. In our approximated model, the spikes are modelled by a time independent triangle *tr* with different decreasing time for Te and Ti cells. The rising time is 5ms and the decreasing time is of 20ms/100ms for the Te/Ti cells. This approximation leads to a current flow I_i^{Syn} described by equation (2). I_i^{Syn} is composed by the output of all neurons *j* connected to the cell *i* via a synapse. t_j denotes the instance, where cells *j* ejects a spike. The spike-triangle *tr*(*t*), the gating weights g_{ij} and the difference $(u_{Syn} - u_i(t))$ *determine* the synaptic current flow from a cell *j* to a cell *i*. As seen in equation (2), u_{Syn} can take on two values dependent on the nature of the synapse.

$$I_i^{Syn}(t) = \sum_j g_{ij}\ \mathrm{tr}\big(t - t_j\big)\,(u_{Syn} - u_i(t));\ u_{Syn} = \begin{cases} 0mV\ \ exitatory\ synapses \\ -80mV\ \ inhibited\ synapses \end{cases} \qquad (2)$$

The currents I_i^{INP} is modeled by $I_i^{INP}(\mathrm{t}) = g_F\ F_i(\mathrm{t})$. (3)

g_F denotes a gaiting weight for a feature with value F_i connected to the gate of a cell *i*. Together with the approximations (2) and (3), the differential equation (1) is given by:

$$\frac{du}{dt} = \frac{g}{C}\ \big(u_{EQ} - u(t)\big) + I^{Syn}(t) + I^{INP}(\mathrm{t}) +\ I^{DC};\ \ u(t) > u_{TH} \rightarrow u(t + dt) = u_{RESET} \qquad (4)$$

At rest, when all flows are 0, *u*(*t*) approaches the value $u_{EQ} = -67mV$. Whenever *u*(*t*) reaches a threshold $u_{TH} = -40mV$, the cell emits a spike and returns to a value $u_{RESET} = -87mV$.

For each Te-neuron of a CB-PING, the synaptic current $I_i^{Syn}$is given by the spikes ejected by the neurons generating the edge features and by the spikes generated by the synapses connecting Te/Ti cells. The current $I_i^{INP}$is given by the samples of the gradient produces by the envelope features weighted by g_F. The weights g_{ij} and g_F determine the strength of the current flow into a Te-neuron of an CB-PING. The strength must be sufficient large to generate Θ-spikes. Yet biologically, a single ion channel cannot deliver such high currents. To achieve such high currents, many synapses must send spikes to the Te/Ti cells. To achieve high currents the single neurons shown in fig. 3 are realized biologically by populations of neurons performing the same operation simultaneously. Thus, the weights g_{ij} and g_F are a model for the sum of small weights from many synapses.

4 Experiments

4.1 Experimental Set Up

For running the experiments with speech data, an articulatory speech database was chosen in view of future experiments. From a professional British speaker, 1300 phonetically diverse utterances (read speech) were recorded together with a Carstens AG500 electromagnetic articulography [18]. The audio samples are down sampled to 16 kHz. The database is labelled automatically by forced alignment using the Combilex lexicon [19].

For simulating the θ-oscillator, the equations (2)-(4) are implemented in matlab. The parameters of the neurons are the same as described in [9]. Eq. (4) is solved by the approximation:

$\Delta\mathrm{u} = f(u(t), I^{syn}(t), I^{INP}(\mathrm{t}), I^{DC})\Delta\mathrm{t};\ u(t + \Delta\mathrm{t}) \approx u(t) + \Delta\mathrm{u};$ Δt=0.5 ms

Most computing time is needed for simulating the number of synapses implemented for interconnecting the Te and Ti neurons. Further the computing time is determined by the size of the timestep Δt used to solve the differential equation (1). Due to the smooth behavior of the curve of u(t), Δt=0.5 ms was chosen.

4.2 Generation of the Features

The envelopes of the auditory signal are extracted within 19 critical bands (CB3-CB21) with center frequencies 248, 328, … 6330, 7423 Hz. First experiments showed that most spikes generated by the edge features correspond to the maximal increase of the envelopes in the range of vowels. As shown in fig. 4, the spikes of the edge-features spike earlier than those of the broadband reference edge features. This may be caused that in different CBs the rise of the CB-envelopes due to spectral properties given by the movements of the articulators. Some of the generated spikes are error prune (mostly insertion errors). Errors can be derived by comparing the instance of the spikes delivered by the edge to the instances of the reference spikes from a broadband analysis. Further the instance of the spikes generated by the edge-features related to different bands show differences.

As shown in section 4.3 below, each CB-PING is able to reduce the insertion errors. Further, the network of connected CB-PINGs is able to diminish the temporal differences of the spikes generated by the not connected CB-PINGs leading to a 'unified' θ-oscillation.

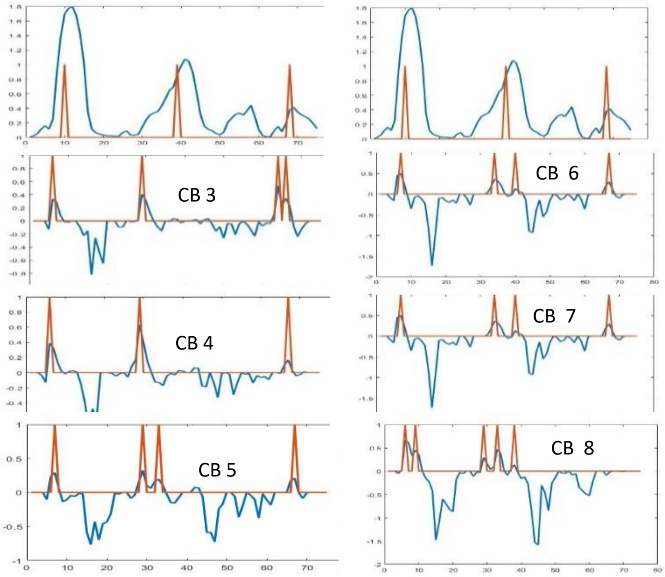

Figure 4 – features for the utterance *Jack Webster* realized by 3 syllables. The two curves at top (repeated right and left) show the envelope of the broad-band auditory signal together with the pulses generated at the maximum of the increase for vowel onset (edge feature of the broad band). The maximal increase is searched within a region given by the labels of the vowels (start/end)

Below: feature-curves extracted from critical bands CB3 - CB8. For each band the curves show the gradient of the envelope curve together with the spikes of the edge feature.

4.3 The Nature of the Θ - Oscillator

Two different architectures concerning the interconnections between the Te/Ti neurons are implemented. In one implementation, the CB-PINGs are not interconnected. Thus, each CB-PING produces its own θ_k-spikes. In the other implementation, the CB-PINGs are interconnected, where the output of each CB-Te/CB-Ti cell is connected to selected other CB-Te/CB-Ti cells via synapses. Fig. 5 shows the inhibition mechanism of the PINGs to inhibit 'follower spikes' from the starting spikes of an edge features within a syllable. This inhibition depends on the duration of the Ti-spikes. Whenever a follower spike is in the range of the duration of the Ti-spikes, the follower spike is deleted. Thus, the first spike of an edge feature arriving at the synapse of a Te-cells is the 'winning' spike. This strategy leads to a 'correct' Te-spike, whenever the first one is the correct one. In the current implementation the first spike often is not the correct one. This fact must be considered for future implementation of optimized edge features. For not interconnected CB-PINGs the θ_k of the different CB-PINGs show more or less large differences. Dependent on the gate variables of the synapses performing interconnections, the difference between the resulting θ_k can be manipulated. First experiments with CB-PINGs interconnecting the Te neurons only, show a tendency to 'unify' the different θ_k-oscillations, where instance of the reference Θ-spike is later than the instances of the unified oscillations.

A specific role has the ion current I^{DC} given in equation (4). In the simulation shown in fig. 5, I^{DC} has the value 0 for Ti-neurons and the value 20 for Te-neurons. Comparing the potentials u of the Te and Ti cells, the Te potential has the tendency to rise between the non-spiking regions caused by a positive I^{DC}. Due to this rising potential in time, the Te cells get more and more sensitive to the input of synapses connected. Thus, missing spikes of an edge feature can be restored by the spikes ejected from Te cells of other bands.

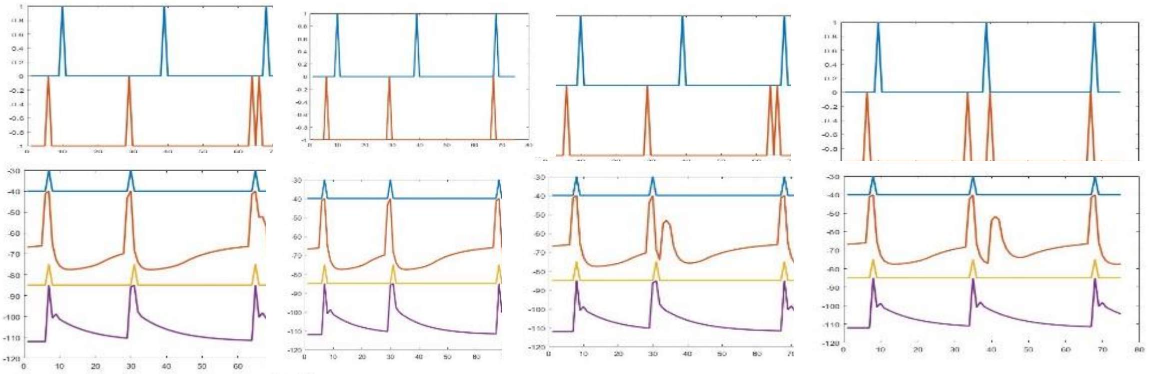

Figure 5 - Spike patterns of the Θ – oscillator for critical band CB3-CB6 for the utterance 'Jack Webster'. The CB - PINGs are not interconnected; CB-Te-neurons are driven by the spikes of the edge features. The triangle spike of the inhibiting TI-cells has a rising time of 20ms and a falling time of 150ms. The six curves for each CB$_k$; k=3-6 from left to right; first curve: reference spikes; second curve: spikes of the edge feature; 3. curve: θ_k −spikes; 4. curve: potential u of the Te-cell; 5. Curve: spikes of the Ti-cell; 6. curve: potential u of the Ti-cell

For future implementations, a learning algorithm must be implemented, which optimize the weights of the synapses to restore missing spikes und to minimize differences in the instances of the θ_k − spikes of different CBs.

5 Conclusion

The paper presents an implementation of a neural model of a θ-oscillator based on a CB-PING architecture. Using speech of a labeled database first experiments show that the θ-oscillator is able to produce spikes at the onset of syllables, which generate the θ-oscillations. The θ-oscillator is driven by edge features derived from the gradient of the envelope of the auditory signal. The temporal instances of the spikes of the edge features are error prune (insertion, deletion). Yet the experiments show that the θ-oscillator can reduce the errors of the edge features. Provisionally, the errors are given by comparing the instances of the edge features to reference spikes of edge features, extracted from by a broadband analysis of the envelopes using the labeled position of the vowels. A neuronal derived concept to evaluate the quality of the spikes is still missing due to the lack of reference spikes measured in the cortex.

The current implementation builds a platform for improving the edge features and the architecture of the CB-PINGs. To minimize the errors, a learning algorithm for optimizing the parameters of the θ-oscillator is needed. As the θ-oscillations lead to the concept to θ-syllables [23], the implemented θ-oscillator can help to study further the nature of the θ-syllable and the related articulatory code. This code is needed to implement Brain-Machine-Interfaces (BMI) to improve the communication abilities of handicapped people. Still the implemented algorithm for an θ-oscillator is far away to compete with the algorithm for segmenting, implemented in current ASR systems.

6 References

[1] LAMEL, L.: *Formalized knowledge used in spectrogram reading: acoustic and perceptual evidence of stops*. In *RLE technical report no. 537 (thesis)*, 1988.

[2] LOWERRE, P.T.: *The Harpy speech recognition system*. In *Ph.D. Thesis* Carnegie-Mellon Univ., Pittsburgh, PA. Dept. of Computer Science,1976.

[3] HUGGINS, A.W.F.: *On the perception of temporal phenomena in speech*. In *Journal of the Acoustical Society of America* 51, 1279-90, 1972.

[4] MERMELSTEIN, P.: *Automatic segmentation of speech into syllabic units*. In *The Journal of the Acoustical Society of America* 58, 1975.

[5] GHITZA, O.: *On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum*. In *Frontiers in Psychology*: 3:238, 2012.

[6] GIRAUD, A.L. AND D. POEPPEL: *Cortical oscillations and speech processing: emerging computational principles and operations*. In *Nat. Neuroscience* 15(4), pp. 511-517, 2015.

[7] NEY, H.; *The Use of a One Stage Dynamic Programming Algorithm for Connected Word Recognition*. In *IEEE Trans.* In *Acoustics, Speech and Signal Processing*, Vol. ASSP-32, No.2, pp.263-271, 1984.

[8] OGANIAN, Y. and E. F. CHANG: *A speech envelope landmark for syllable encoding in human superior temporal gyrus*. In *Science Advances*, 2019.

[9] HYAFIL, A., L. FONTOLAN, C. KABDEBON., B. GUTKIN, and A. GIRAUD: *Speech encoding by coupled cortical theta and gamma oscillations.* In *eLife*, DOI: 10.7554/eLife06213, 2015.

[10] HULLETT, P. W., L. S. HAMILTON, N. MESGARANI, C. E. SCHREINER and E. F. CHANG: Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. In *Journal of Neuroscience*, 36 (6) 2014 – 2026, 2016.

[11] WINER, J. A. and C. E. SCHREINER: *The Inferior Colliculus*. New York: Springer, 2005.

[12] FLETCHER, H. and R.H., GALT: *The perception of Speech and Its Relation to Telephony*. In *the Journal of the Acoustic Society of America*, Vol. 22, number 2, pp. 89-151, 1950.

[13] HÖGE, H.: *Modeling of Phone Features for Phoneme Perception.* In ITG, 2016

[14] HÖGE, H.: *On the Nature of the Features Generated in the Human Auditory Pathway for Phone Recognition.* In *Proc. Interspeech*, Dresden, pp. 1551-1555, 2015.

[15] CHI, T., P. RU, and S. A. SHAMMA: *Multiresolution spectrotemporal analysis of complex sounds*. In *J. Acoust. Soc. Am*. 118, August, pp. 887–906, 2005.

[16] HÖGE, H.: *Human Feature Extraction - The Role of the Articulatory Rhythm*. In *Proc. ESSV*, 2017.

[17] GERSTNER, W. and W. KISTLER: *Spiking Neuron Models*. In Cambridge University Bridge, UK 2002

[18] RICHMOND, K., P., HOOLE and S. KING: *Announcing the Electromagnetic Articulography (Day 1) Subset of the mngu0 Articulatory Corpus*. In *Interspeec*h, pp. 1505-1508, 2011.

[19] FITT, S., K. RICHMOND, and R, CLARK: *The Combilex lexicon*. www.cstr.ed.ac.uk/ research/projects/combilex

[20] CHI, T., RU, P., S. A. SHAMMA: *Multiresolution spectrotemporal analysis of complex sounds*. In *J. Acoust. Soc. Am*. 118, August, pp. 887–906, 2005.

[21] BUZSÁKI, G. and A. DRAGUHN: *Neuronal oscillations in cortical networks*. In *Science* 304, 1926–1929, 2004.

[22] K. B. DOELLING, L. H. ARNAL, O. GHITZA, and D. POEPPEL: *Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing*. In *NeuroImage* 85, 761–768, 2014.

[23] HÖGE, H.: *The nature of the Articulatory Code*. In *Proc. ESSV*, 2020.

[24] HÖGE, H.: *Using Elementary Articulatory Gestures as Phonetic Units for Speech Recognition*. In *Proc. ESSV*, 2018

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