31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

On the Cognitive Neurodynamics of Listening Effort: A Phase Clustering Analysis of Large–Scale Neural Correlates

Daniel J. Strauss, Farah I. Corona–Strauss, Corinna Bernarding,
Wolfgang Reith, Matthias Latzel, and Matthias Froehlich

Abstract— **An increased listening effort represents a major problem in humans with hearing impairment. Neurodiagnostic methods for an objective listening effort estimation could revolutionize auditory rehabilitation. However the cognitive neurodynamics of listening effort is not understood and research related its neural correlates is still in its infancy.**

In this paper we present a phase clustering analysis of large–scale listening effort correlates in auditory late responses (ALRs). For this we apply the complex wavelet transform as well as tight Gabor Frame (TGF) operators. We show (a) that phase clustering on the unit circle can separate ALR data from auditory paradigms which require a graduated effort for their solution; (b) the application of TGFs for an inverse artificial phase stabilization at the α/θ–border enlarges the endogenously driven listening effort correlates in the reconstructed time–domain waveforms.

It is concluded that listening effort correlates can be extracted from ALR sequences using an instantaneous phase clustering analysis, at least by means of the applied experimental pure tone paradigm.

I. INTRODUCTION

An increased listening effort represents a major problem in humans with hearing impairment [1], [2]. In the area of auditory habilitation and rehabilitation digital signal processing has revolutionized the hearing aid technology during the last decade. Modern digital hearing aids offer vast fitting capabilities, allowing a high degree of adaptation to the needs of the individual patient, see [3] for a survey. The utilization of this adaptivity to reduce the listening effort in the individual patient is a major concern in hearing aid fitting. However, fitting procedures which preferably require a minimum cooperation of the patient are still missing and the objective estimation of the listening effort based on auditory processing correlates represents an unsolved problem [2], [3].

Rather isolated past research mainly deals with double stimulation paradigms using finite resources/capacity cognitive models [4], [5]. Here the patients have to solve a primary task related to speech discrimination and a secondary task involving their (motor) reaction time to another secondary visual stimulus, see [4], [5] and [6] for a more recent review. However these paradigms require an increased patient cooperation, they are influenced by many non–listening effort related factors, and are *a priori* based on a crossmodal design, moving the focus away from the auditory modality.

Recent research pointed out the importance of non–audiological variables for hearing aid fitting, in particular, for modern and flexible digital devices, see [7] and [2]. Edwards [3] identified the potential of advances in hearing science and cognition for radical innovations in future hearing aid technologies. Pichora–Fuller [2] highlighted the importance of neuroscience research related to attention and cognitive effort estimation for hearing aid fitting and audiologic rehabilitation. Very recently, [8] reviewed the role of selective attention in normal and hearing impaired listeners. In particular, these authors discussed how hearing impairment may degrade selective attention.

Based on our results in [9] and [10], we present a phase clustering analysis of large–scale listening effort correlates in auditory late responses (ALRs) in this paper.

II. METHODS

A. Neurodynamics of Listening Effort

Listening: Kiessling et al. [11] described *hearing* as a passive function that provides access to the auditory world via the perception of sound, primarily useful to describe impairment by audiometry methods whereas *listening* was defined as the process of hearing with intention and attention which requires the expenditure of cognitive demands.

Auditory Scene Analysis: Let us now discuss *listening* from a bidirectional bottom–up (exogeneous)/top–down (endogenous) auditory processing point of view. Several computational theories have been developed to augment the bottom–up sensory processing with top–down feedback mechanisms, e.g., see [12], [13], [14], [15] and references therein. Employing schema based information stored in memory and driven by, e.g., expectation, experience, and emotions, these models generate predictions on higher processing areas which are projected to lower processing areas to influence the bottom–up information flow.

Let us now concentrate on top–down projections in the auditory modality. According to auditory scene analysis (ASA) and Bregman [16], the processing of exogeneous information reaching conscious states involve an analytical and a synthetical stage: the perceptual stimuli are decomposed into discrete sensory elements in the analytical stage; the sensory elements that are likely to have arisen from the same acoustical source are recombined into a perceptual stream

D. J. Strauss, F. I. Corona-Strauss, and C. Bernarding are with the Computational Diagnostics & Biocybernetics Unit, Saarland University Hospital and Saarland University of Applied Sciences, Homburg/Saar, Germany {strauss}@cdb-unit.de and with the Key Numerics GbR, Saarbruecken, Germany.

W. Reith is with the Clinic for Diagnostic and Interventional Neuroradiology.

M. Latzel and M. Froehlich is with the Siemens Audiologische Technik GmbH, Erlangen, Germany

in a process called auditory grouping (synthetical stage). In auditory grouping, we can differentiate between exogenous (primitive) and endogenous (schema–driven) grouping. Exogenous grouping is a purely data–driven process. Endogenous grouping utilizes the top-down projections mentioned before.

In the following, we use the term endogenous top–down projections or endogenous processing for subconscious and consciously driven top–down projections evoked by intention. A simplified representation of the auditory stream selection model by selective attention given in [9] is discussed in the following: Let the d–dimensional vector $\mathbf{w} = (w_1, w_2, \ldots, w_d)$ denote the weights of the segregated streams according to their assigned probability of getting selected (highest to lowest). The selection probability depends on exogeneous (e.g., physical stimulus attributes) and endogenous factors such as corticofugal top–down modulation by intentional processes. As listening is defined as an auditory process with intention and attention, we assume that it represents an effortful endogenous modulation of $\mathbf{w}$ in this scheme.

A common complaint of the hearing impaired is that listening in noisy situations is an exhausting experience, and a hearing impaired person is far more tired after some time in such a setting than someone with normal hearing [3]. In noisy situations, there are many competing auditory streams. It could be argued that a distorted auditory system is of course less effective in automatic ASA processes, especially, exogeneous processes and needs more endogenous modulation and effort, e.g., expectations driven by the semantic context of a stream which would be in accordance to the findings of [3], [2], [17].

Prediction: The probabilistic stream selection model which we sketched before can be mapped to the corticothalamic feedback loops, see [9] for the neurodynamics of attention and listening effort in [10]. This model predicts a larger phase synchronization of ALRs in the N1 range for a larger corticofugal modulation (over the thalamic reticular nuclei) of bottom–up information, see [9], [10] for details.

B. Subjects, Experimental Paradigm, and Signal Processing

Subjects were student volunteers from the Saarland University and Saarland University of Applied Sciences with normal hearing. A total of 20 subjects (mean age: 27, standard deviation: 4.1; 9 females, 11 males) entered the study. ALR single sweeps were recorded using a commercially available amplifier (g.tec USBamp, Guger Technologies, Austria) and electrodes were placed at the left and right mastoid, the vertex, and the upper forehead. Electrodes impedances were below 5kΩ in all measurements (filter: 1Hz–30Hz, sampling frequency: 512Hz). Two different paradigms were used with a distinct degree of difficulty to solve an auditory task. Artifacts where removed by an amplitude threshold of $50\mu V$.

Difficult Paradigm (DP): For DP we delivered 3 pure tones (1kHz, 1.3kHz and 1.6kHz) at 70dB (HL) of 40ms each in random order to the right ear at randomized inter-stimulus interval (ISIs) of 1–2s. Meanwhile, the left ear was presented with music which played the role as distractor. Subjects were required to pay attention to the stimulus and detect the target tone which was the 1.3kHz stimulation.

Easy Paradigm (EP): For EP we delivered just 2 pure tones (0.5kHz, 1.3kHz) at 70dB(HL) of 40ms each in random order to the right ear at randomized ISIs of 1–2s. Subjects were required to pay attention to the stimulus and detect the target tone which was the 1.3kHz stimulation.

The randomized stimulation paradigms were used to maximize the entropy of the experiment such that attention and effort is required to solve them. The rational for DP and EP is that solving DP requires more listening effort than solving EP. The subjects had to press a button after a target tone had been recognized. The number correctly identified target tones also served as control of the cooperation of the subject. For the numerical analysis in Sec. III, we considered just ALRs that were evoked by the target tone as it had the same frequency in both paradigms. This is a necessary constraint as our extracted time–scale features (see below and [18]) depend on the stimulation frequency such that the results to different simulation frequencies could not be compared.

The randomized stimulation paradigms were used to maximize the entropy of the experiment such that attention and effort is required to solve them. The rational for DP and EP is that solving DP requires more listening effort than solving EP. The subjects had to push a button after a target tone had been recognized. The number correctly identified target tones also served as control of the cooperation of the subject. For the numerical analysis in Sec. III, we considered just ALRs that were evoked by the target tone as it had the same frequency in both paradigms.

Phase Synchronization Stability: Using a phase clustering analysis on the unit circle (e.g., see [19], [20]), we have recently shown that the stability of the instantaneous phase can be used to efficiently monitor auditory attention in maximum entropy paradigms [18].

Given the sequence $\mathcal{X} = \{\mathbf{x}_k \in \ell^2 : k = 1, \ldots, M\}$ of M ALR single sweeps, we define the phase synchronization stability by the following equation

$$\Gamma^N_{\mu,\nu}(\mathcal{X}) = \frac{1}{N}\left|\sum_{k=1}^{N} e^{\imath\Phi_{\mu,\nu}(\mathbf{x}_k)}\right|, \quad N \leq M, \qquad (1)$$

where $\Phi_{\mu,\nu}(\mathbf{x}_k)$ represents the instantaneous phase, depending on the scale parameter μ and the time parameter ν. In the case of the continuous complex wavelet transform (CCWT), we obtain the wavelet phase synchronization stability (WPSS) by $\Phi_{\mu,\nu}(\mathbf{x}_k) = \arg\left((\mathcal{W}_\psi \mathbf{x}_k)[\mu, \nu]\right)$, where $\mathcal{W}_\psi$ is a discrete–time version of the wavelet operator to the wavelet ψ, see [21]. Another way to derive by the instantaneous phase stability by means of tight Gabor frames (TGFs), see [22] for an introduction to TGFs. For the TGFs we obtain the Gabor frame phase stability (GFPS) by $\Phi_{\mu,\nu}(\mathbf{x}_k) = \left((\mathcal{G}_\varphi \mathbf{x}_k)[\mu, \nu]\right)$, where $\mathcal{G}_\varphi$ is the tight Gabor frame analysis operator to the generating function φ.

Analysis of the Phase Stability: The WPSS is well localized in time and frequency using the Heisenberg adaptivity of the wavelets, i.e., a large support in time for low frequencies with a good frequency resolution and small support (good localization) in time for high frequency components.

As for the analysis of ALRs related to N1 wave modulations by attention [21], we use the WPSS for the forward analysis by Eq. (1) in this study with $\mu = 40$ which corresponds to the α/θ–border (6–9 Hz) for the chosen wavelet (the 6th derivative of the complex Gaussian function as in [21]).

It is easy to see that the synchronization stability as evaluated by the wavelet phase stability in Eq. (1) is a value in $[0, 1]$. We have a perfect synchronization stability for a particular μ' and ν' for $\Gamma_{\mu',\nu'} = 1$ (perfectly coherent phases) and a decreasing stability for smaller values due to phase jittering.

Phase Stability Modification: The GFPS provides a uniform decomposition of the ALR single sweep frequency band such that each modulation parameter μ can be directly associated with a frequency band in Hz. In contrast to the highly redundant CCWT, TGFs may provide a very efficient analysis as well as synthesis of the analyzed signal using the very same function, e.g., see [23] an references therein. For an efficient inverse transform of the CCWT, the δ–distribution reconstruction [24] is frequently used.

For the experiments in our study, we used a tight window generated from the Gaussian function with 128 modulations and $\nu = 1$. Note that these values result in an overcomplete Gabor frame decomposition with $m = 0, 1, \ldots, M - 1$ (m is the modulation index) frequency channels spanning the frequency interval $[mf_c, (m + 1)f_c]$ with $f_c = f_s/(2M)$ (f_s is the sampling frequency).

III. RESULTS AND DISCUSSION

In Fig. 1 we have shown the evolution of the WPSS for an increasing number of sweeps of one subject as example. More precisely, we used Eq. (1) to obtain the (discrete) function $\xi_{\mu,\nu}[n] = \Gamma^n_{\mu,\nu}(\mathcal{X})$ $(n = 1, 2, \ldots, N)$ with $N = 100$, $\mu = 40$, and ν corresponding to 110ms, i.e., a sample from the N1 wave range.

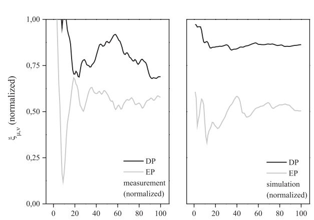

Fig. 1. The evolution of the WPSS over the sweeps as represented by the function $\xi_{\mu,\nu}[\cdot]$ (slightly smoothed for this graph and normalized).

It is noticeable that the WPSS provides a robust discrimination of DP and EP data sets, even for a small number of sweeps for the measured and simulated data. However, in the simulated condition, the oscillatory behavior is of course different due to the highly simplified noise model.

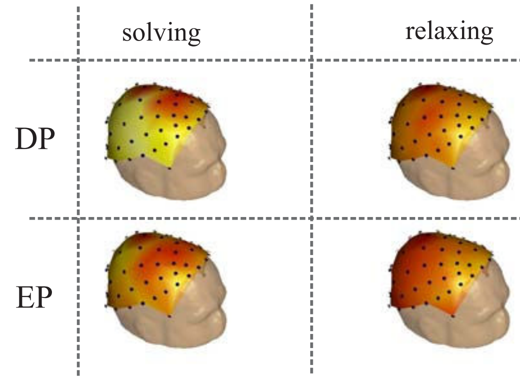

Fig. 2. Topological mapping of the WPSS for 100 sweeps in a paradigm / condition matrix using a 64 channel recording. The rows represent the WPSS while stimulating with DP and EP. The columns represent the WPSS while solving the paradigm and a relaxing condition. Normalized scale for WPSS: black/dark red(0) to white/light yellow (1).

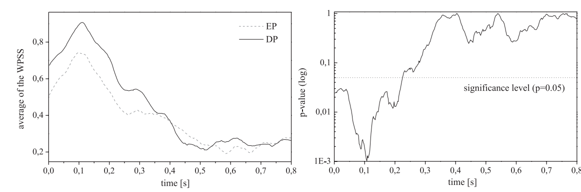

Fig. 3. Left: Grand average of the WPSS (over all the subjects) for EP and DP. Right: The results for the time resolved ANOVA.

Topological Mapping: In Fig. 2 we have shown a topological mapping of the WPSS for 100 sweeps in a paradigm / condition matrix using a 64 channel recording from one subject as example. The matrix shows the WPSS for solving DP and EP and for relaxing during the DP and EP stimulation, i.e., not solving the paradigms. It is noticeable that the WPSS is much larger in temporal and parietal areas while solving the DP as compared to the solution of EP and the relaxed condition, respectively.

DP/EP Discrimination: Figure 3 (left) shows the grand average (averaged over all the subjects) of the ipsilateral WPSS in Eq. (1) for a sequence of $N = 100$ sweeps per subject and paradigm, respectively. It is noticeable that the WPSS is much larger for the DP than for the EP, especially in the expected interval. In Fig. 3 (right) we have shown the results for the time resolved (one–way) ANOVA (over the subjects to evaluate the significance of the grand average in the plot above). It is noticeable that in the interval of interest of the N1 and P2 wave in ALRs, we have significant differences between the EP and DP condition. This shows that the increased effort to solve the more difficult paradigm DP is reflected in an increased WPSS of ALR sequences.

Artificial Phase Modification: In Fig. 4 (left) we have shown the results of an experiment using the phase modification and synthesis approach described in Sec. II-B. In particular, we have shown the averaged time domain waveform (using $N = 100$ sweeps) for a DP and EP data set. The EP data set was phase stabilized at the α/θ–border and synthesized. The α/θ–border corresponds here approximately to scale 40 in the WPSS experiments before. It is noticeable that the average of the ALR single sweep set $\mathcal{X}$ in the EP condition with this small band phase correction shows a large correlation to the DP data set in the N1 and P2 wave range.

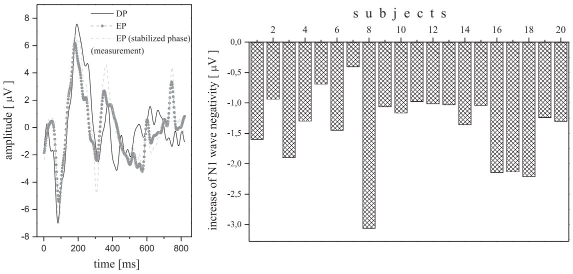

Fig. 4. Left: The averaged time domain ALRs for the DP and EP as well as EP with an artificial phase stabilization. Right: The increase of the N1 negativity after the artificial phase stabilization in the EP data for all the subjects.

In Fig. 4 (right) we show the increase of the N1 wave negativity after the phase stabilization for all the subjects (mean increase of the negativity: -1.4047 μV; standard deviation: 0.6208μV). It is noticeable that the phase stabilization results in a larger N1 negativity for all the subjects and thus has a crucial influence on the N1 amplitude for the experimental data. This shows that a small band phase reset (at the α/θ–border) in averaged ALRs – interestingly in way, such that a phase corrected (the the α/θ–border) EP data resembles DP data (without changing the energy of the individual sweeps in these bands). Although the discussion whether late evoked responses are due to a phase reset of the ongoing activity or just superimposed to the oscillatory electroencephalographic activity is a major research topic at the moment, e.g., see [25] and references therein, the clear discrimination of evoked from phase reset features is beyond the scope of this paper.

IV. CONCLUSIONS

We have presented a phase stability analysis of large–scale listening effort correlates in ALRs using forward and inverse transform techniques.

It is concluded that listening effort correlates can be extracted from ALR sequences using an instantaneous phase clustering analysis, at least by means of the applied stimulation paradigm.

REFERENCES

[1] C. B. Hicks and A. M. Tharpe, "Listening effort and fatigue in school–age children with and without hearing loss," *Jounral of Speech, Language, and Hearing Research*, vol. 45, pp. 573–584, 2002.

[2] M. K. Pichora-Fuller and G. Singh, "Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation," *Trends Amplif*, vol. 10, pp. 29–59, 2006.

[3] E. Edwards, "The future of hearing aid technology," *Trends in Amplification*, vol. 11, pp. 31–45, 2007.

[4] D. W. Downs, "Effects of hearing aid use on speech discrimination and listening effort," *Journal of Speech and Hearing Disorders*, vol. 47, pp. 189–193, 1982.

[5] B. Rakerd, P. Seitz, and M. Whearty, "Assessing the cognitive demands of speech listening for people with hearing losses," *Ear and Hear*, vol. 17, pp. 97–106, 1996.

[6] B. Larsby, M. Hällgren, B. Lyxell, and S. Arlinger, "Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal–hearing and hearing–impaired subjects," *International Journal of Audiology*, vol. 44, pp. 131–143, 2005.

[7] P. B. Kricos, "The the influence of nonaudiological variables on audiological rehabilitation outcomes," *Ear and Hearing*, vol. 21, pp. 7S–14S, 2000.

[8] B. G. Shinn-Cunningham and V. Best, "Selective attention in normal and impaired hearing," *Trends in Amplification*, vol. 12, pp. 283–299, 2008.

[9] C. Trenado, L. Haab, and D. J. Strauss, "Corticothalamic feedback dynamics for neural correlates of auditory selective attention," *IEEE Trans Neural Syst Rehabil Eng.*, vol. 17, pp. 46–52, 2009.

[10] D. J. Strauss, F. I. Corona-Strauss, and M. Froehlich, "Objective estimation of the listening effort: Towards a neuropsychological and neurophysical model," in *Conf Proc IEEE Eng Med Biol Soc*, vol. 2008:1, 2008, pp. 1777–1780.

[11] J. Kiessling, M. K. Pichora-Fuller, S. Gatehouse, D. Stephens, S. Arlinger, T. Chisolm, A. C. Davis, N. P. Erber, L. Hickson, A. Holmes, U. Rosenhall, and H. von Wedel, "Candidature for and delivery of audiological services: special needs of older people." *International Journal of Audiology*, vol. 42 Suppl 2, pp. 2S92–101, 2003.

[12] S. Grossberg, "Adaptive pattern classification and universal recoding: I. parallel development and coding in neural feature detectors," *Biological Cybernetics*, vol. 23, pp. 121–134, 1976.

[13] D. Mumford, "On the computational architecture of the neocortex II, The role of the cortico-cortical loops," *Biological Cybernetics*, vol. 65, pp. 241–251, 1992.

[14] S. Ullman, "Links sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex," *Cereb. Cortex.*, vol. 5, pp. 1–11, 1995.

[15] K. Friston, "A theory of cortical responses," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 360, pp. 815–836, 2005.

[16] A. S. Bregman, *Auditory Scene Analysis*. Cambridge, MA: MIT Press, 1990.

[17] S. B. D. M. Pichora-Fuller, K., "How young and old adults listen to and remember speech in noise," *J. Acoust. Soc. Amer,*, vol. 97, pp. 593–608, 1995.

[18] Y. F. Low, F. I. Corona-Strauss, P. Adam, and D. J. Strauss, "Extraction of auditory attention correlates in single sweeps of cortical potentials by maximum entropy paradigms and its application," in *Proceedings IEEE EMBS NER*, 2007, pp. 469–472.

[19] R. A. Dobie and M. J. Wilson, "Analysis of auditory evoked potentials by magnitude-squared coherence," *Ear & Hearing*, vol. 10, pp. 2–13, 1989.

[20] F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: phase synchronization and large-scale integration," *Nat Rev Neurosci*, vol. 2, pp. 229–239, 2001.

[21] D. J. Strauss, W. Delb, R. D'Amelio, Y. F. Low, and P. Falkai, "Objective quantification of the tinnitus decompensation by synchronization measures of auditory evoked single sweeps," *IEEE Trans Neural Syst Rehabil Eng.*, pp. 74–81, 2008.

[22] T. Strohmer, "Rates of convergence for the approximation of dual shift–invariant systems in $\ell_2(\mathbb{Z})$, journal=The Journal of Applied Fourier Analysis and Applications, volume=5, year = 1999."

[23] P. L. Søndergaard, "Gabor frames by sampling and periodization," *Advances in Computational Mathematics*, vol. 27, pp. 355–373, 2007.

[24] M. Farge, "Wavelet transforms and their applications to turbulence," *Annu. Rev. Fluid Mech.*, vol. 24, pp. 395–457, 1992.

[25] P. Sauseng, W. Klimesch, W. R. Gruber, S. Hanslmayr, R. Freunberger, and M. Doppelmayr, "Are event–related potential components generated by phase resetting of brain oscillations? A critical discussion," *Neuroscience*, vol. 146, pp. 1435–1444, 2007.