Objective Estimation of the Listening Effort: Towards a Neuropsychological and Neurophysical Model

Daniel J. Strauss, Farah I. Corona–Strauss, and Matthias Froehlich

Abstract—Modern hearing aid fitting could be revolutionized by the availability of objective methods for the listening effort estimation. However experimental and theoretical research dealing with this subject is still in its infancy.

In this paper we present first results towards a neuropsychological and neurophysical model for the objective estimation of the listening effort by electroencephalographic data. Our model is based on intended endogenously driven top–down projections represented by corticothalamic feedback dynamics for auditory stream selection and their large-scale correlates in auditory evoked late responses. The predictions of the presented model are compared to experimental data obtained during different auditory tasks which required a graduated effort for their solutions. The experimental data verified the model predictions.

It is concluded that the proposed neuropsychological and neurophysical modeling of stream selection provides an appropriate framework for listening effort estimation. The presented preliminary results of an ongoing study are encouraging, however, further focal research is necessary in order to estimate in how far the presented model and future extensions might support modern hearing aid fitting in practice.

I. INTRODUCTION

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 [1] for a survey. The utilization of this adaptivity to reduce the listening effort in the individual patient is a major concern in the fitting of hearing aids. 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 [1], [2].

Rather isolated past research mainly deals with double stimulation paradigms using finite resources/capacity cognitive models [3], [4]. 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 [3], [4] and [5] 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.

Parallel to the development of the digital hearing aid technology, strengthened research in neuroscience – supported by the rapid development of enhanced brain imaging methods – has brought light in neural and cognitive auditory processing. Surprisingly, these results have rarely been exploited for the problem of listening effort estimation up to now. However, recent research pointed out the importance of non–audiological variables for hearing aid fitting, in particular, for modern and flexible digital devices, see [6] and [2]. Edwards [1] 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.

In this paper, we present a neuropsychological and neurophysical framework for the objective estimation of the listening effort and its first experimental validation by a time–scale single sweep processing of auditory late responses (ALRs).

II. METHODS

A. Towards a Neuropsychological and Neurophysical Model

Listening: Kiessling et al. [7] 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 [8], [9], [10], [11] 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 [12], 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 recombinced into a perceptual stream.
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 [13] is shown in Fig. 1 (left). The $d$–dimensional vector $w = (w_1, w_2, \ldots, w_d)$ denotes 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 $w$ 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 [1]. In noisy situations, there are many competing auditory streams. It could be argued that a distorted auditory system is of no use in such a setting than someone with normal hearing [1]. In noisy situations, there are many competing auditory streams. It could be argued that a distorted auditory system is of no use for subconscious and consciousness–driven 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 [13] is shown in Fig. 1 (left). The $d$–dimensional vector $w = (w_1, w_2, \ldots, w_d)$ denotes 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 $w$ 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 [1]. In noisy situations, there are many competing auditory streams. It could be argued that a distorted auditory system is of no use in such a setting than someone with normal hearing [1]. In noisy situations, there are many competing auditory streams. It could be argued that a distorted auditory system is of no use.

Multiscale ALR Modeling: Focusing on the architecture of our model, the weights $w$ of segregated streams are modulated by means of three corticofugal gains $G_1$, $G_2$, and $G_3$ as shown in Fig. 1 (right). These gains correspond to the neurophysiological connectivity in the following way: The modulation of the gains $G_1, G_2$, and $G_3$ is accomplished by means of a corticofugal transfer function $\Phi$ that represents the activity between cortical excitatory $e$ and inhibitory $i$ neural populations $C_{e,i},$ and subcortical incoming auditory stimuli $A_{in}(.\cdot)$, as given by $C_{e,i}(k,w) = \Phi(\Pi, G_1, G_2, G_3, L, t_d, \theta(k)),$ where the function $\Phi$ incorporates the dendritic low–pass filtering effects $L$, corticofugal time delays $t_d$, amplitude modulation $\Pi$, and a dispersion relation $\theta(k)$ that describes the mean field cortical pulse–wave propagation of the average response of populations of neurons [15]. Following [15], the response to an auditory stimulus $A_{in}(.\cdot)$ by considering the excitatory neural population as the main source of large–scale brain activity, is represented by $G(r, w) = \int_D \frac{C_{e,i}(k,w)}{A_{in}(k,w)} \Psi_s(k)dk$. Here $\Psi_s(\cdot)$ is a function that comprises parameters related to the auditory stimulus, while the domain $D$ represents a finite two dimensional approximation of the human cortex. ALRs as large–scale response of the brain are obtained by the inverse Fourier transform of the function $G(\cdot,\cdot)$. We refer to [15] for a detailed description of the numerical scaffold for multiscale evoked response modeling and to [13] for more discussions regarding the mapping of this model to the hearing path and its implementation to simulate auditory selective attention correlates reflected ALRs.

Prediction: Here $G_1$ takes positive and negative values regarding the regulatory effect in the flux of information between cortico–subcortical structures. Gain $G_2$ takes positive values mainly emphasizing the excitatory effects on cortex and thalamus, while Gain $G_3$ is also taking part in the regulation of inhibitory effects in TRN. Based on our numerical results in [13], [16], this forward ALR model predicts a larger wavelet phase synchronization stability (WPSS) (see Sec. II-B) for an increased top–down endogenous modulation of the bottom–up data in the range of the N1 wave in ALRs, resulting in stable, i.e., a rather small variations of the gains, especially $G_1$. Consequently, the prediction of the model is that an increased listening effort will lead to a larger WPSS of ALR single sweep sequences.

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 12 subjects (mean age: 26, standard deviation: 3.9; 5 females, 7 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 50µV.

**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 inter-stimulus interval (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 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. This is a necessary constraint as the WPSS depends on the stimulation frequency (see [17]) such that the results to different simulation frequencies could not be compared.

**Synchronization Stability:** For analysis of the ALR single sweeps, we used the wavelet phase synchronization stability (WPSS) that we introduced in [18], [17], [19] for the quantification of auditory attention in ALR single sweeps (the larger the synchronization stability, the larger the auditory attention paid to stimuli). Here we just present the mathematical definition of the synchronization stability and refer to [19] for detailed discussions.

Let $\psi(\cdot) = [a]^{−1/2} \psi((\cdot−b)/a)$ where $\psi \in L^2(\mathbb{R})$ is the wavelet with $0 < \int_{\mathbb{R}} |\Psi(\omega)|^2 |\Psi(\omega)|^{-1} d\omega < \infty$ ($\Psi(\omega)$ is the Fourier transform of the wavelet), and $a, b \in \mathbb{R}$, $a \neq 0$.

The wavelet transform $\mathcal{W}_\psi : L^2(\mathbb{R}) \rightarrow L_\psi^2(\mathbb{R}^2; \, \Delta_{dA})$ of a signal $x \in L^2(\mathbb{R})$ with respect to the wavelet $\psi$ is given by the inner $L^2$-product ($\mathcal{W}_\psi x(a,b) = \langle x, \psi_{a,b} \rangle_{L^2}$). We define the WPSS $\Gamma_{a,b}(x)$ of a sequence $x = \{x_m \in L^2(\mathbb{R}) : m = 1, \ldots, M\}$ of and subset of $N$ ($N \leq M$) ALR single sweeps by

$$\Gamma_{a,b}^N(x) := \frac{1}{N} \sum_{m=1}^N e^{i \arg((\mathcal{W}_\psi x_m)(a,b))}.$$  \hspace{0.5cm} (1)

In this study, we used the 6th-derivitive of the complex Gaussian as in [19]. Following our results in [17], [19] (experimental) and [16] (model), we used the scale $a = 40$ in the numerical analysis in Sec. III.

### III. RESULTS AND DISCUSSION

Figure 2 (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 (see Sec. II-A and Sec. II-B). The mean WPSS in range of the N1 wave (we chose 75ms to 150ms) showed a significant difference (Wilcoxon test, significance level p<0.05) between the EP and DP data. Moreover, the mean WPSS in this interval was larger in 11 out of 12 subjects. This shows that the WPSS increased with the effort to solve the auditory task in almost all the subjects.

In Fig. 2 (right) we have shown the topological difference of the mean WPSS for 100 sweeps for data of the DP and EP condition for one subject with multi–electrode recording as example. It can be seen that the major differences in the time interval of interest appears in the temporal lobe which also exhibited the largest activity in our attention / no attention study in [17].

![Fig. 2](image-url)

**Fig. 2.** Left: Grand average of the WPSS (over all the subjects) for EP and DP. Right: Topological difference of the mean (interval [75ms,150ms]) WPSS for ($M = 100$) for data of the DP and EP condition for one subject with multi–electrode recording as example (ipsilateral view). The larger the WPSS difference, the brighter the points.

Let us introduce the map $\Lambda_{a,b,X} : I = \{1,2,\ldots,M\} \rightarrow [0,1]$ with $\Lambda_{a,b,X}(j) = \Gamma_{a,b}^j(X)$ ($j \in I$). Figure 2 shows the cross recurrence plot (CRP) of $\Lambda_{a,b_1,X}(j)$ and $\Lambda_{a,b_2,X}(j)$ with $a = 40$, and $b_1$ and $b_2$ are two arbitrary time points in the interval of interest for one subject as example (for the CRP: embedding dimension: 2, time lag: 2, neighborhood size: 0.15, $\ell^2$–distance measure). Recurrence plots reveal all the times when the phase space trajectory visits roughly the same area in the phase space, see [20] for an introduction to CRPs. In the EP, there is no clear formation diagonal lines whereas in the DP conditions a clear diagonal line appears for larger values of $j$ (i.e., the lower right corners of the CRPs), showing larger nonlinear interrelations between the WPSS for $b_1$ and $b_2$ in the DP condition. In this sense, the CRP provides another visualization of our findings in [17] (experimental) and [13] (computational model) where a large WPSS was reflected in clear traces in a matrix representation of ALR single sweeps and fuzzy traces for non–focal attention conditions.

The presented preliminary results of an ongoing study are encouraging as they verified the model predictions given in Sec. II-A. They showed that auditory evoked potentials...
might provide access to listening effort related neural correlates. Note that the N1 and P2 wave is commonly used in paradigms related to auditory attention and stream selection, e.g., see [21], [22].

However, it is worth to emphasize that the presented results serve as indicators that the problem of listening effort estimation might be approached by the sketched neuropsychological and neurophysical framework involving forward and inverse modeling of auditory stream selection. Further research needs focus points in both, theoretical and experimental work. For instance, the use of a stationary system for the corticothalamic gain modulation in Sec. II-A and the limited number of neural processing unit does not allow for the inclusion auditory processing on higher levels, e.g., the N400 or P600 wave frequently used in auditory examinations of higher processing, e.g., [23].

The design of paradigms that are appropriate for hearing aid fitting using this framework is a major objective of our further research in this project. Moreover, the inclusion of recent experimental evidence of early corticofugal modulation of subthalamic processing units [24] in the presented model is also an interesting direction of further research from a theoretical point of view.

IV. CONCLUSIONS

It is concluded that the presented neuropsychological and neurophysical model of auditory stream selection provides an appropriate framework for listening effort estimation.

The presented preliminary results of an ongoing study are encouraging, however, further focal research is necessary in order to estimate in how far the presented model and future extensions might support modern hearing aid fitting.

V. ACKNOWLEDGMENTS

The authors would like to thank Corinna Bernarding for supporting the data acquisition.

REFERENCES