Chirp Analyzer for Estimating Amplitude and Latency of Steady-State Auditory Envelope Following Responses

Eduardo Martínez-Montes, Yalina García-Puente, Matías Zañartu, *Senior Member IEEE*, and Pavel Prado-Gutierrez

***Abstract*— OBJECTIVE: The envelope following response (EFR) is a clinically relevant evoked potential, reflecting the synchronization of the auditory pathway to the temporal envelope of sounds. Since there is no standard analysis of this potential, we here aim at contrasting the relative accuracy of known time-frequency methods and new strategies for the reliable estimation of the EFR amplitude and latency. METHODS: The EFR was estimated using explicit time-frequency methods: the Short-Term Fourier Transform (STFT) and the Morlet Continuous Wavelet Transform (CWT). Furthermore, the Chirp Analyzer (CA) was introduced as a new tool for the reliable estimation of the EFR. The applicability of the methods was tested in animal and human recordings. RESULTS: Using simulated data for comparing the estimation performance by each method, we found that the CA is able to correctly estimate EFR amplitudes, without the typical bias observed in the STFT estimates. The CA is more robust to noise than the CWT method, although with higher sensitivity to the latency of the response. Thus, the estimation of the EFR amplitude with any of the methods, but especially with CA, should be corrected by using the estimated delay. Analysis of real data confirmed these results and showed that all methods offer estimated EFRs similar to those found in previous studies using the classical Fourier Analyzer. CONCLUSION AND SIGNIFICANCE: The CA is a potential valuable tool for the analysis of the EFR, which could be extended for the estimation of oscillatory evoked potentials of other sensory modalities.**

Manuscript received April 14, 2020; revised September 16, 2020; accepted October 10, 2020. Date of publication October 21, 2020; date of current version January 29, 2021. This work was supported in part by ANID programs under Grant BASAL FB0008 and Grant MEC 80170124, and in part by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Grant P50DC015446. *(Corresponding authors: Eduardo Martínez-Montes; Pavel Prado-Gutiérrez.)*

Eduardo Martínez-Montes is with the Neuroinformatics Department, Cuban Center for Neuroscience, Havana 11300, Cuba (e-mail: eduardo@cneuro.cu).

Yalina García-Puente is with the Fabulas Laboratory, Department of Physics Engineering, École Polytechnique Montréal, Station Centre-ville, Montreal, QC H3T1J4, Canada (e-mail: yalina19@gmail.com).

Matías Zañartu is with the Advanced Center for Electrical and Electronic Engineering, and with the Department of Electronics Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile (e-mail: matias.zanartu@usm.cl).

Pavel Prado-Gutiérrez is with the Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile (e-mail: pavel.prado@usm.cl).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2020.3032835

***Index Terms*— Chirp analyzer, envelope following response, Fourier analyzer, latency estimation, time-frequency analysis.**

I. Introduction

THE envelope following response (EFR) is a scalp-recorded evoked potential that reflects the phase-locked neural activity of sensory pathways to amplitude modulations (envelope) of environmental stimuli [1], [2]. In the auditory domain, the EFR can be elicited by tones or broadband noise, modulated in amplitude by a chirp, i.e. a continuous sweep of modulation frequencies, each of which will be called hereinafter instantaneous modulation frequency (IMF). In some cases, the IMF increases linearly during the first half of the stimulus and decreases in the second half, with the same rate [1]–[3] (see Supplemental Material, Section A). If the sweep of IMFs is slow enough, the amplitude of the EFR can be interpreted as a measure of the capability of the neuronal system to respond to each IMF [2]. The EFR can also be evoked by natural vowels, thus reflecting the auditory processing of acoustic features of speech sounds [4]–[6].

The EFR could be clinically valuable for objectively evaluating the outcome of hearing aids, optimizing the fitting process of the hearing devices [5], [7]. As an objective measurement of temporal auditory processing, the EFR might be used in the diagnosis of speech disorders, including word deafness, deficits in speech discrimination, and dyslexia [8], [9]. Some of the advantages of evaluation tools based on estimated EFRs are the possibility of eliciting auditory responses using natural stimuli (running speech or stimuli with similar spectral components), the implementation of relatively short hearing tests, and the potential of detecting neural responses using statistical tests [4], [5], [10]–[12]. However, the extended clinical use of the EFR requires standard procedures for the acquisition and analysis of the response. Studies comparing the accuracy of the methods used to estimate this evoked potential are not available and their limits of applicability need to be established.

Methodologies for the analysis of the EFR that are based on fixed-frequency transforms (e.g., the discrete Fourier transform) are effective if steady-frequency stimuli are used [13]. However, the analysis of EFRs elicited by natural sounds with non-stationary envelope, such as speech, requires the use of time-frequency methods [14]. Consequently, the EFR has

been analyzed using Complex Demodulation methods [15], Short Time Fourier Transform (STFT) [12], [16], and the Continuous Wavelet Transform (CWT) using Morlet complex functions [17], [18]. Nevertheless, the Fourier Analyzer (FA) is the most widely used method in the analysis of the EFR [2], [5]–[8], [10], [11], [19].

There are also several methodologies for the analysis (and estimation) of the EFR latency, which is a critical parameter that allows for establishing the neural generators of the response. The latency of the EFR has been calculated as the time shift that maximizes the linear statistical correlation between the envelopes of the stimulus signal and the electrophysiological response [13]. However, in most cases, the apparent latency of the EFR is estimated based on an assumed linear relationship between the amplitude modulation frequency of the stimulus and the phase difference between the stimulus and the response, as determined by using the Fast Fourier Transform (FFT) or the FA [2], [3]. These methods assume that the electrophysiological signal is stationary and to our knowledge, no attempts have been done to estimate the latency for EFRs elicited by non-stationary stimuli.

In this work, we compare the accuracy of time-frequency methods for the analysis of the EFR, stressing their advantages and drawbacks in different scenarios. Furthermore, we introduce an "Envelope analyzer" as a novel methodology for the reliable estimation of this kind of response. Its rationale is similar to that of the Fourier Analyzer (FA) [2], [20] but instead of the classical Fourier basis, it uses the same non-stationary modulation signals of the stimuli (e.g. chirps) as the reference function. This follows the hypothesis that the response will match the amplitude envelope of the stimulus, and that the use of this reference function will lead to a more precise estimation of the response. In this work, we restrict the "Envelope Analyzer" to a particular chirp signal, and therefore we will call it Chirp Analyzer (CA). We compare the relative performance of the CA with that of other time-frequency methodologies. In this sense, the traditional FA is implicitly included in the comparison, as it can be derived from the STFT. Finally, we evaluate a new amplitude-based methodology to estimate the latency of the EFR, by measuring the time interval between the onset of every IMF in the stimulus and the maximum amplitude of the estimated response at the corresponding frequency. For these purposes, we use simulated data to validate the methods and illustrate their performance on real auditory responses of human and animal models.

II. Methods

A. Estimating the Amplitude of the EFR

In this section, we briefly describe the two time-frequency methods used in this study for the estimation of the EFR: the Short Time Fourier Transform (STFT) and the Continuous Wavelet Transform (CWT) using Morlet complex functions. Furthermore, we introduce the Chirp Analyzer (CA).

1) Short-Time Fourier Transform and Fourier Analyzer: We used the Goertzel algorithm to estimate the STFT (discrete version) at predetermined frequencies. Then, the estimated EFR was extracted as the absolute value of the complex Fourier coefficients in the time-frequency points that correspond to each instantaneous modulation frequency (IMF) of the stimulus. By using a rectangular window function, the STFT reduces to the Fourier Analyzer (FA) described in the literature [2], [20], which consists in correlating -in time domain- the signal in each temporal window with the sine/cosine reference functions for the IMF corresponding to the center of the window. Nevertheless, the performance of the STFT did not largely differ when different window functions (rectangular, Hamming, Hann) were tested in a preliminary experiment. Therefore, we used the STFT with a Hamming window as a representative of this family of time-frequency methods, including the FA. The window size was of 1 s for all IMFs, which implies a spectral resolution of 1 Hz in the whole time-frequency plane [14]. However, using the Goertzel algorithm and an overlapping of 0.923 s, we imposed a frequency step of 0.5 Hz and a time step of 77 ms.

2) Morlet Continuous Wavelet Transform: The continuous wavelet transform (CWT) of signal x(t) is defined by:

$$\mathrm{CWT}\,(\tau,\, f) = \int \mathrm{x(t)W(f,\,t-\tau)dt}$$

In our case, the function W(f, t) is the complex Morlet "mother wavelet":

$$\mathrm{W(f,\,t)} = (\sigma_{\mathrm{t}}\sqrt{\pi})^{-1/2}\mathrm{e}^{-\mathrm{t}^2/2\sigma_{\mathrm{t}}^2}\mathrm{e}^{\mathrm{i}2\pi\,\mathrm{ft}}$$

where the temporal support σ_{t} is inversely proportional to the spectral support σ_{f} [21]. We chose this wavelet due to its extensive use in the analysis of cognitively/clinically relevant oscillatory EEG signals, as it can be considered a Gaussian-windowed version of a Fourier Transform [14], [21], [22]. The magnitude $\mathrm{z} = \mathrm{f}/\sigma_{\mathrm{f}}$ is kept constant and determines the number of cycles of the wavelet covered in the temporal support $(\tau \pm \sigma_{\mathrm{t}})$ for every frequency. In this study, we chose z = 8, which means that the wavelet will have about 2.5 cycles, leading to a temporal/spectral support of 63.7 ms / 2.5 Hz around IMFs of 20 Hz and of 10.6 ms / 15 Hz around IMFs of 120 Hz. To allow direct comparison with the other methods, we also used a frequency step of 0.5 Hz and the wavelet coefficients were down-sampled to a time step of 77 ms. The EFR was estimated as the absolute value of the wavelet complex coefficients in the time-frequency points corresponding to the IMF of the stimulus.

3) Chirp Analyzer: Instead of using a Fourier basis, the Chirp Analyzer (CA) consists in correlating the signal x(t) with a non-stationary reference function φ (t) that represents the theoretical response, modeled as the analytic (complex-valued) function of the normalized amplitude-modulation chirp used in the stimulus:

$$\mathrm{CA}\,(\tau) = \int \mathrm{x(t)g(t,\,\tau)}\varphi\mathrm{(t)dt}$$

This procedure is carried out in overlapping rectangular windows $\mathrm{g(t,\,\tau)}$, for achieving a higher temporal precision with the same spectral resolution. Like in FA, the correlation is performed in the time domain, thus making FA and CA faster than both the STFT and CWT, which need to estimate

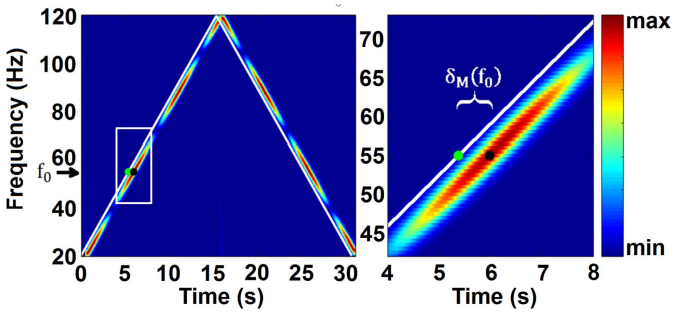

Fig. 1. Estimating the delay of the EFR response for each instantaneous modulation frequency f_0, based on the time-frequency map (left panel) obtained by the STFT or CWT methods. Right panel zooms in the region delimited by a white rectangle in the left panel. White lines represent the (τ_0, f_0) points corresponding to every IMF of the stimulus. The neural delay $\delta_M(f_0)$ for each frequency is estimated as the distance between τ_0 (green dot) and the time at which the highest amplitude of the signal is found (black dot).

coefficients for all frequencies at each time point. Correlations with both real and imaginary parts of the reference signal (complex chirps) leads to complex CA coefficients. The EFR is then estimated as the absolute values of the CA complex coefficients in the time points corresponding to each IMF. It is also noticeable that, like in FA, the phase of the complex CA coefficients would reflect the phase of the signal with respect to the reference function.

B. Estimating the Latency of the EFR

Typical approaches for estimating the latency of oscillatory electrophysiological responses are highly dependent on the measurement of phases with the FFT, which has been shown to be unreliable when the level of noise is high, or when the amplitude of the response is small [23]. As an alternative approach, we implemented an amplitude-based computation of the EFR latency that takes advantage of the fact that time-frequency methods, such as the STFT and the CWT, give amplitude estimates for all time points in each frequency. This allows us to look for the maximum amplitude of the response for every specific IMF, in a time interval after the time corresponding to that IMF in the stimulus. We set this time interval to 200 ms, to cover the entire latency range of auditory neurons [12], [19]. An estimate of the neural delay is then obtained as the difference between the time of the IMF in the stimulus and the time of the maximum EFR amplitude for that frequency (see Fig. 1).

In the case of the Chirp Analyzer, we firstly correlate the signal in each window with time-shifted versions of the real part of the reference signal (i.e. the chirp used in the stimulus). Then, the delay is estimated as the difference between the time point at which the highest correlation is achieved and the time point of the corresponding IMF in the stimulus.

We consider that, in a certain range of IMF, the latency of the neural generators of the EFR does not vary with the IMF. Therefore, the different estimated delays for all IMFs -due to the presence of noise, as well as numerical and model errors of each method- will be considered as random observations of the neural delay. As the statistical properties of these observations are unknown, we evaluated their mean, weighted mean (weights proportional to estimated amplitudes), median, and mode; as empirical estimates of the true neural delay.

An important practical aspect is the definition of the time-frequency precision to be used, i.e. the time step and frequency step of results (we avoid using the term resolution, which refers to the capacity of distinguishing different components in the signal due to the basis functions and time windows used in the analysis). Higher precision will help to estimate more accurate delays (although with a higher computational cost) but it will not lead to better estimates of the EFR, as long as the temporal resolution of the method remains the same. Therefore, we here used the highest time precision (lowest time step) for estimating delays. However, when the delay was known or was not of interest, we used the time step corresponding to the frequency step for all methods, since estimating the response in intermediate frequencies will not lead to better spectral resolution and the EFR will be practically the same. In all simulated/real datasets analyzed here (with a sampling frequency of 2 kHz), both STFT and CA were applied to 1-s-long segments, with 99.95% overlapping (0.5 ms time step) for estimating delays and 92.3% overlapping for estimating the EFR (77 ms time step) to match the user-defined frequency step of 0.5 Hz.

C. Simulated and Real Datasets

EFR was simulated with different shapes, resembling properties of the auditory EFR, e.g. the presence of amplitude peaks at specific (optimum) IMFs, and/or constant amplitudes in ranges of IMFs (red traces in Fig. 2). Nevertheless, simulations were not designed to exactly replicate a measured auditory EFR, but rather to test the capabilities of the methods. EFR simulations that differ from the typical auditory EFR have also been used in previous studies to analyze the effect of stimulus parameters (the IMF rate of change), and of the frequency resolution of the method [4]. The first simulated EFR is a full-wave rectified sine (*Sine-deep*), representing a response with more than one optimum IMF and zero amplitude at particular IMFs. The second is a sine wave vertically stretched by a factor of 0.5 and translated in 0.5 units (*Sine-low*), which represents a response with smoothly-changing amplitude. The third is a rectangular function (*Rect-deep*), which represents constant response amplitudes in specific IMF bands. These EFRs are multiplied by the reference chirp to simulate the electrophysiological responses, which are then delayed and perturbed with noise (the simulation model is detailed in the Supplemental Material, Section B1).

To analyze the effect of the noise on the estimation of the EFR, white noise was added with different peak signal-to-noise ratios (pSNR = 2, 1 and 0.1), defined as the ratio between the squared maximum of the signal (fixed to 1 in all simulations) and the variance of the noise. To explore the influence of the latency in the estimated EFR we simulated responses with latencies of 10, 50, and 100 ms. Following the procedure described by Aiken and Picton [4], delays were introduced by shifting the simulated EFR in time and completing with zeros the beginning of the recording.

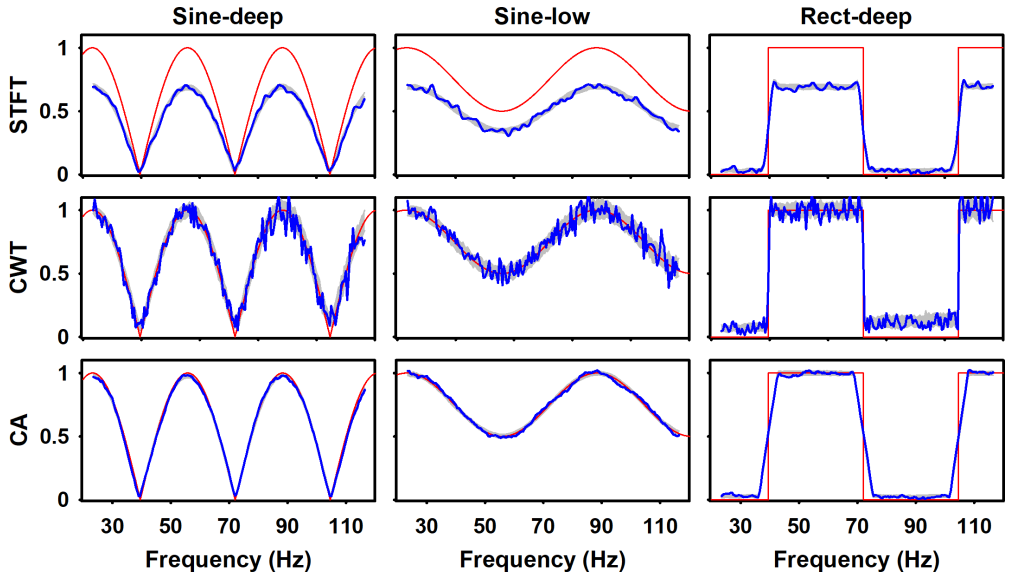

Fig. 2. EFR estimations using STFT, CWT, and CA from data created with simulated EFR (red lines) of different shapes (Sine-deep, Sine-low, and Rect-deep). Blue lines represent the estimated EFRs for a single run, while the grey shadows cover one standard deviation around the mean EFR, across 50 repetitions of each simulation. All signals were simulated with pSNR = 2. The amplitudes are reported in arbitrary units. Note that estimated EFRs do not have values close to the edges, as the first and last values correspond to the center of the first and last analysis windows. Also, recall that each IMF on the x-axis correspond to a progression in time according to the modulation linear chirp.

We simulated EFR with latencies that are higher than those reported in humans (30-40 ms maximum, [2]), with the objective of evaluating the methods in a wide range of possible physiological situations. In every simulated scenario, we created 50 realizations with different random noise and computed the mean and standard deviation of the results (estimated EFR and/or delays).

To test the methods in the analysis of real data, we used electrophysiological recordings obtained from an adult rat (male, 70 postnatal days) and a human baby (male, 2 years old) [3], [24]. The studies were performed under approval of the Animal Research and Ethics Committee of the Cuban Center for Neuroscience. The data acquisition and processing are described in the Supplemental Material, Section B2. In brief, the stimulus consisted of 30 sweeps of 100%-depth amplitude-modulated broadband noise, repeated without any pause between them. The intensity was 50 and 70 dB SPL for rats and 50 dB HL for human babies. Scalp recorded responses were synchronously averaged in the time domain and the EFR was extracted using a Fourier Analyzer with orthogonal reference sinusoids that matched the instantaneous frequency of the stimulus. Due to the effect of anesthesia on the EFR of rats [3] and the reproducibility of the EFR in human babies, [24], the analyses of the EFR in the real datasets was restricted to the 90- to 190-Hz IMF range.

III. Results

A. Analysis of Simulated Data

1) Estimation of EFR Amplitudes: All the methods were able to reproduce the forms of the simulated EFR, although the accuracy of the estimations strongly depended on the methods of analysis (Fig. 2). While the estimated EFRs with both the CWT and the CA were very similar in amplitude to the simulated response, the EFR estimated with the STFT showed a consistent bias toward lower amplitudes (of about 25% lower than the simulated amplitude). Quantitative comparisons between the simulated and the estimated EFRs are shown in the Supplemental Material, Section C. In summary, correlations were above 0.94 for all methods in all scenarios. Remarkably, the highest correlation between rectangular signals was obtained when the EFR was computed using the CWT. This can be explained by the higher temporal resolution of CWT as compared with STFT and CA, which allow to follow the abrupt changes in the rectangular simulation. The relative error (relative Euclidean distance between the estimated and simulated EFRs) were always higher for STFT (above 30%). The lowest errors for sinusoidal EFR were obtained when responses were estimated using the CA (below 5%), while for rectangular EFR resulted from the estimation by the CWT (15%).

2) Noise Robustness of the Estimated EFR: The influence of the level of noise on the estimation of the EFR was analyzed in the case of the Sine-deep simulation (Fig. 3). As expected, the reliability of the EFR estimation was affected when the pSNR decreased, for all methods. However, when outputs were contrasted in signals with equal pSNR, the CWT was the most sensitive to the presence of noise in the recordings (Fig. 3, middle row). The noise impaired the estimation of the EFR performed by using the STFT in a lesser extent, while the CA was the most robust method to increases of the noise level (Fig. 3, top and bottom rows, respectively).

Quantitative comparisons between the simulated and the estimated EFRs are shown in the Supplemental Material, Section D. Summarizing, the correlation between the estimated and simulated EFRs were higher than 0.95 in most of the cases. The exception was in very noisy scenarios (pSNR = 0.1),

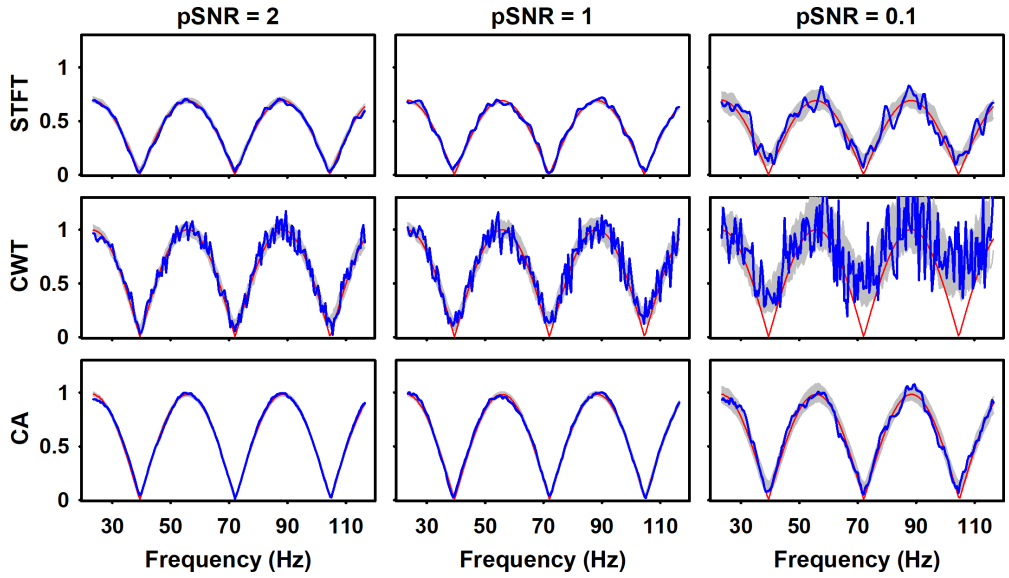

Fig. 3. EFRs estimated from data created with simulated Sine-deep EFR, using different levels of noise (pSNR). Blue lines represent the estimated EFRs for a single run, while the grey shadows cover one standard deviation around the mean EFR across 50 repetitions of each simulation. The red lines represent the EFR estimated when there is no noise affecting the simulated data. The amplitudes are reported in arbitrary units.

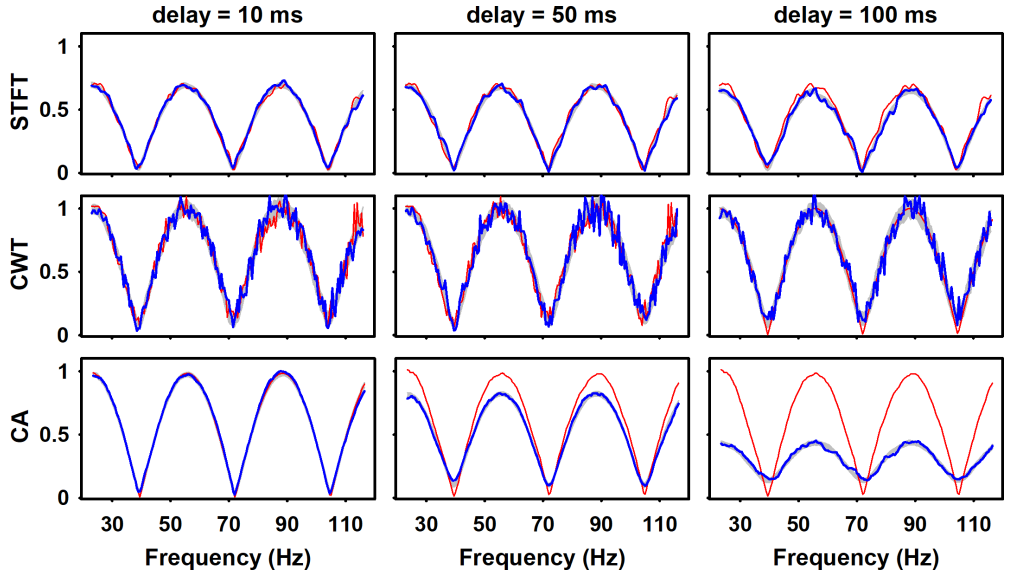

Fig. 4. EFRs estimated from data created with simulated Sine-deep EFR, using different delays with respect to the stimulus. Blue lines represent the estimated EFRs for a single run, while the grey shadows cover one standard deviation around the mean EFR across 50 repetitions of each simulation. The red lines represent the EFR estimated considering the correct delay in the response, i.e. delay-corrected responses. The amplitudes are reported in arbitrary units.

where mean correlation coefficients of 0.92 and 0.63 were obtained when the EFR was estimated using the STFT and CWT, respectively. The relative error was less affected by the increased noise level in the case of STFT (only 1% higher from pSNR = 2 to pSNR = 0.1), while the CWT was the most affected (37% higher). The CA showed an intermediate increase of 7%.

3) Influence of the Response Latency in the Estimated EFR: Since the estimated EFR amplitude might be biased when the latency of the response is not considered, we simulated signals using three different delays and compared the EFR estimated without considering the delay with that obtained when the mismatch introduced by the delay was corrected in the reference signal, hereinafter delay-correction (Fig. 4). Overall, the estimated EFR without considering the delay was similar to the delay-corrected EFR when using the STFT and CWT (Fig. 4, upper and middle panels). On the contrary, the amplitude of the EFR computed with the CA was underestimated when the latency of the response was 50 and 100 ms. This means that the EFR amplitudes estimated with the CA are more sensitive to the presence of a delay than those obtained with the other two methods. These results suggest that analyzing the estimated EFR amplitude for each frequency in a time window after the presentation of the corresponding IMF in the stimulus might be helpful to estimate the latency of the response, which is the topic addressed in the next section.

TABLE I
ABSOLUTE ERRORS OF ESTIMATED DELAYS FOR DIFFERENT NOISE LEVELS

	STFT			CWT			CA		
pSNR =	2	1	0.1	2	1	0.1	2	1	0.1
Wmean	**0.1 ± 0.5**	**0.4 ± 0.5**	3.2 ± 2.0	12.2 ± 2.4	14.1 ± 2.6	29.5 ± 2.0	7.3 ± 0.7	7.7 ± 0.9	7.5 ± 1.6
Mean	**1.5 ± 1.2**	**1.3 ± 1.1**	4.8 ± 2.4	20.5 ± 2.8	25.1 ± 3.0	33.4 ± 4.1	5.2 ± 2.6	5.3 ± 3.3	**1.5 ± 3.9**
Median	**0.2 ± 0.9**	**0.2 ± 1.0**	**1.3 ± 3.8**	11.7 ± 5.3	16.9 ± 5.3	30.3 ± 7.2	8.1 ± 0.8	8.5 ± 0.9	7.7 ± 2.3
Mode	46.1 ± 13.3	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.6 ± 2.8**
Regress	54.6 ± 12.7	59.1 ± 19.2	57.0 ± 24.5	53.4 ± 28.3	61.7 ± 24.1	48.7 ± 32.5	16.2 ± 0.1	16.1 ± 1.3	13.8 ± 4.9

Mean and standard deviation across 50 repetitions of the absolute errors (in ms) between the estimated delay and the delay of simulated signals, for different values of the peak signal-to-noise ratio. In all cases the EFR was simulated with the Sine-deep shape and a delay of 50 ms. In bold we highlighted mean values below 3 ms.

TABLE II
ABSOLUTE ERRORS OF ESTIMATED DELAYS FOR DIFFERENT SIMULATED DELAYS

	STFT				CA			
Simulated delay =	0	10	50	100	0	10	50	100
Wmean	8.2 ± 0.4	4.2 ± 0.4	**0.1 ± 0.4**	**2.6 ± 0.5**	**1.3 ± 0.2**	**2.2 ± 0.3**	7.4 ± 0.8	13.4 ± 1.1
Mean	11.8 ± 1.0	7.7 ± 1.3	**1.2 ± 1.1**	**3.0 ± 1.3**	9.1 ± 0.8	5.6 ± 0.8	5.6 ± 2.7	17.2 ± 2.1
Median	3.4 ± 0.5	**0.5 ± 0.8**	**0.1 ± 0.7**	**0.4 ± 0.8**	**0.0 ± 0.0**	**0.1 ± 0.2**	8.3 ± 0.8	10.2 ± 0.5
Mode	**0.0 ± 0.0**	10.0 ± 0.0	46.1 ± 13.3	45.8 ± 50.8	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.2 ± 1.2**
Regress	5.7 ± 15.7	10.3 ± 16.6	52.7 ± 12.2	113.2 ± 17.8	0.9 ± 9.5	10.9 ± 7.5	16.2 ± 0.1	16.1 ± 0.1

Mean and standard deviation across 50 repetitions of the absolute errors (in ms) between the estimated delay and the delay of simulated signals, for different values of the simulated delay. In all cases the EFR was simulated with the Sine-deep shape and pSNR=2. In bold we highlighted those values with both mean and standard deviation (of absolute errors) not higher than 3 ms.

4) Estimation of the EFR Latency: In addition to the measures of central tendency of instantaneous delays (mean, weighted mean, mode and median), we also estimated the delay as the slope of the linear regression between the estimated EFR phases and the IMFs. To this end, a Sine-deep EFR was simulated with a delay of 50 ms, and different pSNR. Table I shows, for each measure, the mean and standard deviation of absolute errors (difference between the simulated and estimated delays) across 50 repetitions. Errors typically increased as the pSNR decreased. Computation of EFR latencies using CWT consistently showed higher absolute errors as compared with the other methods. Therefore, the CWT was not considered for further analysis of latency. Using the STFT, the median, weighted mean and mean of all delays, offered good estimates of the simulated delay, with errors below 2 ms. However, the best results were obtained when using the mode of the delays estimated with the CA.

To provide more elements about the statistical measure most suitable for the estimation of the EFR latency, we simulated a Sine-deep EFR with different delays but fixed pSNR = 2 (Table II). Again, the median, weighted mean, and mean of latency estimated using STFT were very accurate in the case or large delays (50, 100 ms), while better estimates were obtained with CA for smaller delays. The mode of the values estimated by the CA was more accurate in all cases.

When EFRs with different shapes were simulated (Supplemental Material, Section E), the instantaneous delays were not reliable at IMFs for which the EFR amplitude was very small, which occurred in the case of Sine-deep and Rect-deep simulations. For high-amplitude responses, the mode of instantaneous delays estimated with CA offered the best latency estimates.

Given the sensitivity of the CA method to the presence of delays, we explored the practical usefulness of using the estimated delays for delay-correcting the estimated the EFR. We compared EFRs estimated with the CA assuming zero delay and the delay-corrected EFRs computed using the mode of the instantaneous delays obtained with the CA method (Fig. 5). The correction considerably improved the estimation of the EFR amplitude, even for those frequencies where the actual response's amplitude was very small or zero. The relative error of the reconstructed EFR was lower than 0.5% for small delays, and less than 2% for delays of 50 and 100 ms.

B. Analysis of Real Data

In the analysis of real electrophysiological recordings, the EFR computed with the three methods were compared with those described in previous studies using the FA method [3], [24]. Since these authors used a commercial acquisition system, the exact implementation of the FA, including any scaling or calibrating factor, was not available to us. This impeded us from directly comparing the amplitudes of estimated EFRs, as they could be differently standardized or given in different units. Therefore, we firstly normalized all responses to have the same maximum values as the corresponding EFR obtained with the MASTER system.

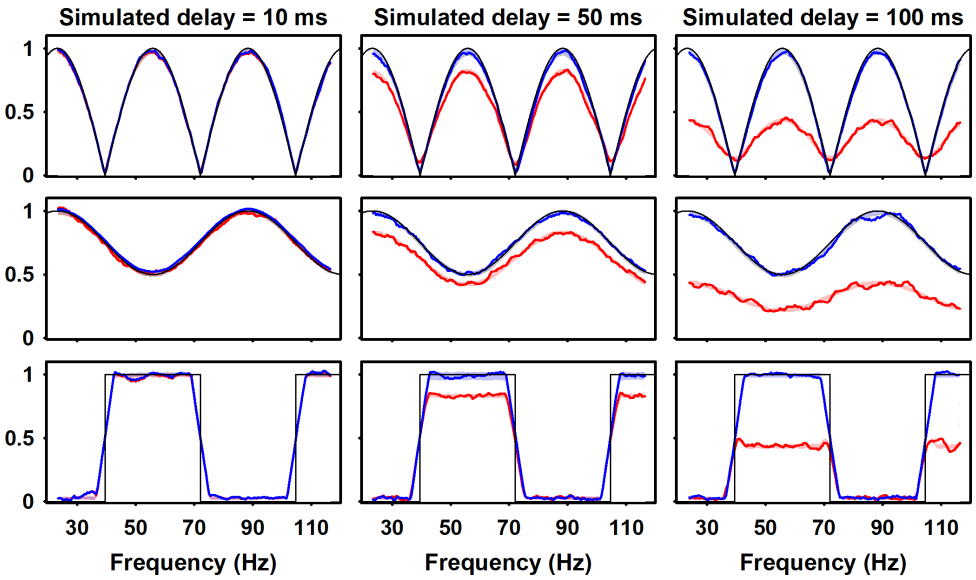

Fig. 5. Correcting the EFR obtained by the CA method with the estimated delay. Simulations were carried out using a pSNR = 2 for the three different EFR shapes. In each panel, the black line represents the simulated EFR; the red line represents a single repetition of the EFR estimated without correcting for the delay and the blue line represents the same single repetition of the EFR re-estimated using a correction with the delay obtained from the mode of instantaneous delays measured with the CA method. Red and blue shadows represent the area covered by one standard deviation around the mean EFR across 50 repetitions for the uncorrected and corrected estimation, respectively.

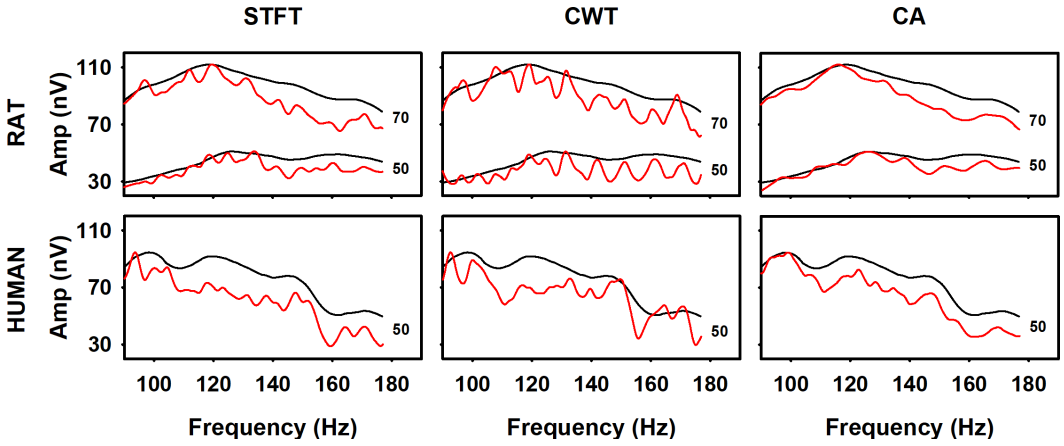

Fig. 6. EFRs estimated (red lines) using the three methods (without delay-correction) from real electrophysiological recordings of an adult rat stimulated with broadband noise at 50 and 70 db-SPL (upper panels) and from a human baby stimulated at 50 db-HL (lower panels). The responses were smoothed using a 7-point moving average and normalized such that all maxima values coincide with those obtained with the corresponding FA (black lines).

In general, the shape of the EFR estimated in an adult rat and a human baby with the three methodologies (without delay correction) did not remarkably vary from the corresponding ones estimated with the FA (Fig. 6). Nevertheless, when the responses were elicited by 70-dB-SPL stimuli, the EFR amplitudes of rats obtained with the STFT and CA in the 140–160 Hz range seemed to be slightly smaller than those computed with the FA. This effect is also present in the EFRs of the human baby for frequencies higher than 110 Hz. The EFRs obtained with the CWT presented the highest variability with more local extremes. This effect can be explained by the higher sensitivity to noise of the CWT and could be ameliorated by smoothing the response of CWT with a larger time window in a moving average or with a low-pass filter. Even in these conditions, the CA offered the highest correlation (up to 97%) and the smallest relative error (down to 10%) as compared with the EFR estimated with the FA (see Section F of the Supplemental Material).

Left panels in Fig. 7 show the non-normalized EFRs estimated by the three methods described in this study. The main differences observed are consistent with the results in simulations, namely: the smallest amplitudes corresponded to the evoked potentials obtained with the STFT, and the noisiest response was obtained with the CWT. Remarkably, the amplitudes of the EFR estimated with the CWT were higher than those obtained with the CA in the whole frequency range. This difference suggests a nonzero delay of the response, which leads to a decrease in the amplitude of the EFR estimated with CA. Consequently, we estimated the delay as the mode of instantaneous delays obtained with the CA, and then, corrected

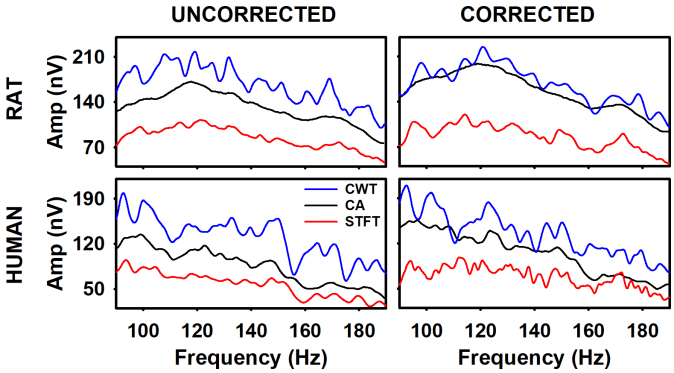

Fig. 7. Left column: Non-normalized EFRs estimated with the three methods described in this study, from a real electrophysiological recording in an adult rat stimulated with broadband noise at 70 db-SPL and in a human baby at 50 db-HL. Right column: the EFRs obtained for the same data as in the left column but correcting the amplitudes by using the delay estimated with the mode of instantaneous delays measured with the CA method.

all EFR estimations (Fig. 7, panels in the right column). For the EFR of the rat, the estimated delay was about 24 ms, while in the case of the human baby, the estimated delay was around 34 ms. Both delays are large enough to lead to underestimated EFR amplitudes with the CA, when the latency of the response is not considered. Consequently, as a result of the correction, the EFR computed with the CA showed a clear increase in amplitude, without introducing large changes in the general shape of the response.

IV. Discussion

In this work, we have introduced the Chirp Analyzer (CA) as a new method for estimating auditory envelope following responses (EFR). This method is based on using the envelope of the stimulus as the reference function for a sliding-window correlation with the recorded EEG. It provides advantages over time-frequency methods with fixed basis functions, such as the STFT and CWT. Although described for the analysis of auditory EFR, the methodology can be extended to other types of auditory stimulation and even other sensory modalities (e.g. visual, or somatosensory) in which continuous stimulation with non-stationary characteristics is extremely common.

A. Evaluating Time-Frequency Methods With Simulations

Simulations showed that the CA is more robust and reliable than the STFT and the CWT to estimate EFRs with different shapes, even in presence of noise (Fig. 2 and 3). Importantly, the lower EFR amplitude estimated with the STFT illustrated in those figures is due to the use of stationary reference signals that will never match the response in the entire windows of analysis, leading to a spectral underrepresentation of the signal. This bias might make the STFT method (and the equivalent FA) less reliable to estimate actual absolute EFR amplitudes.

The lower robustness of the CWT to estimate small EFR amplitudes, as compared to CA, is explained by the higher sensitivity of the CWT to the background noise. This is a consequence of the fixed inverse proportionality between the IMF and the temporal resolution implemented in the CWT, which strongly depend on the range of frequencies covered by the stimuli. In our simulations, the wide frequency range (20 to 120 Hz) led to a very high temporal resolution for higher frequencies, making the EFR estimation noisier than those offered by the STFT and the CA. As positive points, the high temporal resolution of the CWT allows to estimate abrupt changes in amplitude better than the other two methods and offers unbiased estimations of the EFR amplitudes (Fig. 2).

The CA method showed the lowest robustness to the presence of delays in the response, being of physiological and/or instrumental origin (Figs. 4, 5). This was not surprising as the CA is based on correlating the recorded signals with the reference chirp and any displacement between these signals will make several cycles to be in counter-phase, leading to a quick drop in the correlation. However, for delays lower than 50 ms, the underestimation of the EFR amplitude given by the CA is smaller than 15%, and even smaller for STFT and CWT.

Although all methods can be used without knowing the actual latency, strategies to minimize the influence of the unknown delays on the EFR estimation can consider using linear chirps with a slow sweep of frequencies in the stimuli, and an analysis window of sufficiently length to minimize the influence of transient activities. It has been shown that this response can accurately follow the temporal envelope of the stimulus when the IMF change is slower than 10 Hz/s [4]. In this scenario, the response can be considered locally stationary, which is relevant for the computation of amplitudes using the FA. Nevertheless, a more direct approach would be to introduce the knowledge about the latency of the response in the procedure for estimating the EFR. Although a fixed 10-ms shift has been used for correcting the mismatch between the reference and physiological signals [4], [16], [18], it is important to note that the latency of the response varies across experiments and subjects.

B. Evaluating the Estimation of Response Delays

Although the delay of the EFR has been computed using the phase information of the response [1], [3], [4], the circular nature of phases makes this estimation problematic, since the manipulations needed to convert them in a linear magnitude do not always ensure the correct unwrapping for a reliable regression. It has been shown that the estimation of phases using time-frequency methods based on Fourier basis are not reliable when the amplitude of the oscillatory response is low [23]. In our simulations, the STFT and CWT did not offer correct estimations of the delay when using the phase information of the response, since phases were consistently miss-estimated in most of the simulated scenarios (Tables I and II). Although the CA showed better estimates of delays, errors were always higher than 10 ms. We also confirmed that phase estimations worsen when the EFRs had very low amplitudes (see Supplemental Material, Section E).

The amplitude-based method proposed here for estimating delays allows computing latencies for each IMF or frequency range, which is in accordance with studies showing

that the physiological delay vary within restricted ranges of IMFs [2]–[4]. However, in our simulations we introduced the same delay for all IMFs, such that the latencies estimated for different IMFs can be considered as statistical variations of the neural delay. Therefore, the response delay can be computed by using the mean, the weighted mean, the median and the mode of the whole set of instantaneous delays. Results presented in Tables I and II showed that the STFT and the CA can be used for reliably estimating delays (see also the Supplemental Material, Section E). The instantaneous delays computed with the STFT showed a Gaussian distribution around the simulated value (except in those IMFs where the EFR amplitude was too small). However, the distribution of the instantaneous delays computed with CA was not Gaussian, as the simulated delay was mostly underestimated. For this method, using the mode across all instantaneous delays is convenient and leads to a robust estimation, since the mode is a value that belongs to the population. For the CA, even in the lowest pSNR scenario, only 8 out of the 50 repetitions showed the mode of instantaneous delays to be different from the simulated value.

Importantly, the computed delay depends on the overlapping of the analysis windows for the estimation of instantaneous delays. In our data, for a time step between consecutive windows of up to 100 ms (90% overlapping for 1-s windows), estimates using the mode are still similar to the simulated delay, while for higher time steps (up to 200 ms), the absolute errors can be of 8 to 15 ms. This suggests that when small delays are expected, the CA should be used with a high temporal overlapping and when delays are expected to be higher than 50 ms, the best option is to use the weighted mean or the median of the instantaneous delays computed with STFT. Additionally, the low relative error (lower than 2%) of the EFR estimated with CA when delays were 10 ms (Fig. 5) suggests that there is no need to correct the EFR by an estimated delay if this is expected to be small. Nevertheless, since the CA showed a biased EFR amplitude of more than 20% for delays of 50 ms (Fig. 4, bottom row), a delay-correction ensures a better reconstruction of the EFR with latencies higher than 40 ms.

Finally, we discuss the influence of the EFR amplitude on the reliability of the estimated delay. As described above, very small responses in a particular range of IMFs will lead to less reliable estimations of the instantaneous delays, which will influence the final estimation of the response delay. Physiologically, very low response amplitudes can be obtained when the EFR is elicited by near-threshold acoustic stimuli [3] and in early stages of maturation, in both humans and animal models [3], [24]. Therefore, it is important to establish the number of frequencies with supra-threshold amplitude necessary for reliably estimating the amplitude and latency of the EFR with the CA method introduced here. A preliminary study on this issue (see Supplemental Material, Section G) showed that, for an accurate estimation of the delay, the STFT needs more data points than the CA method, whose estimations with the mode and median were accurate for bandwidths as low as 17 Hz and 23 Hz, respectively. However, results also suggest that both methods are robust for the lack of data and can be used in a wide range of normal and pathological cases. In practice, a preliminary estimate of the EFR might be helpful to evaluate if there are enough IMFs with high amplitudes to support a good estimation of the response delay for a subsequent delay-correction of the EFR.

C. Analysis of EEG Recordings in a Rat and a Human Baby

Although the EFR estimated with the STFT, CWT and CA from real recordings showed the same general shape as the one found with the FA method, there are important differences regarding the absolute amplitude and the level of smoothness of the whole curve (Fig. 6 and 7). The differences in amplitudes might be explained by different normalization procedures of the methods. According to our simulations, we were able to recover the correct amplitudes with the CWT and CA methods. The STFT (equivalent to FA) was consistently biased toward lower amplitudes, however, in a commercial system the FA method can be calibrated to compensate this bias and offer the correct absolute EFR amplitudes. The different smoothness is more difficult to explain given that the original estimation of the EFR with the FA was not available. It might be possible that the FA was smoothed in a stronger way than the simple 7-point moving-average smoother applied with the methods presented in this study to facilitate the evaluation of variability and local maxima.

Another explanation for the differences found may arise from the fact that the neural generators of the EFR might not follow exactly the model assumed. For instance, the biological system, instead of responding to each IMF in the modulation chirp, might respond with equal amplitude to preferred frequency bands. In that scenario, time-frequency methods still seem a better option than FA, since they allow the exploration of the signal's energy before selecting ad hoc the relevant time-frequency pairs to extracting the EFR. Finally, we should recall that the EEG reflects the activity of many different brain sources that may generate non-linear responses, which are difficult to disambiguate from the macroscopic signal. Future research should address this topic with tailored experiments.

In the analysis of the real data, only the EFR estimated by the CA showed an increase in amplitude when it was corrected by the delay, which suggest that there was indeed a nonzero delay between the stimulus and the recorded signal. Interestingly, the delays estimated in the real datasets (24 ms in the rat and 34 ms in the human newborn) were higher than those usually handled in the literature [2], [3]. This can be explained by physiological differences between the subjects used in the studies and the specific experimental setup used for the EFR acquisition. Another aspect is the different methodologies used for the estimation of latencies (phase-based vs. amplitude-based estimation methods), considering the potential misleading results when using mathematical measures derived from phase estimates in real signals [22], [23]. Moreover, the EFR has been suggested to have different latencies for different IMF ranges [2], [3], while here we considered that the latency was the same for all IMFs. Finally, the noisy nature of the signals might also influence the estimation of delays with any

method. In any case, these discrepancies need to be confirmed in future studies with larger datasets.

V. Conclusions

In this work, we compared the estimation of the EFR with the use of explicit time-frequency methods and with a new method introduced here: the Chirp Analyzer (CA). Instead of using a Fourier basis, the CA uses the same linear chirp that modulates the carrier stimulus as the reference signal for the estimation of the EFR. Therefore, considering that the physiological response will closely follow the chirp modulation function, the CA allows a better match between the recorded signal and the reference function than the Fourier Analyzer (FA), which directly impact the estimation of the response parameters.

Using controlled simulated responses, the CA showed to be able to estimate EFR amplitudes without the typical bias obtained when using the STFT (equivalent to FA but for the whole time-frequency map). The CA was also more robust to noise than the CWT, but it should be used cautiously when neural delays higher than 50 ms are expected. Consequently, we proposed an amplitude-based methodology for estimating the apparent latency of the EFR, which proved to be reliable when using the STFT and the CA methods, as assessed using simulated responses. The estimation of the EFR amplitudes, especially those obtained with the CA, should be corrected when possible, by using the estimated latency.

Results in real data were consistent with findings using simulated data. Although a thorough validation should be carried out with larger datasets, the methods explored here are promising tools for a proper characterization of the EFR and might contribute to improve and standardize the methodologies currently available. This is important for extending the use of this kind of response in audiology. Importantly, our results will only hold for EFRs elicited by stimulus modulated by chirps and the CA reduces to the FA method in the common case of stimulation with sinusoidal modulated tones. However, since the strategy followed for the implementation of the CA is very general and can be applied to almost all types of oscillatory brain responses, this study opens the possibility of implementing more complex experiments to evaluate brain responses to non-stationary stimuli, including speech and music.

Acknowledgment

The authors gratefully acknowledge E. Mijares-Nodarse, T. Diez-Aldama, A. Torres-Fortuny, and G. Savío for supporting the acquisition of the electrophysiological recordings used in this study.

References

[1] J. Artieda, M. Valencia, M. Alegre, O. Olaziregi, E. Urrestarazu, and J. Iriarte, "Potentials evoked by chirp-modulated tones: A new technique to evaluate oscillatory activity in the auditory pathway," *Clin. Neurophysiol.*, vol. 115, pp. 699–709, Mar. 2004.

[2] D. W. Purcell, M. S. John, B. A. Schneider, and T. W. Picton, "Human temporal auditory acuity as assessed by envelope following responses," *J. Acoust. Soc. Amer.*, vol. 116, no. 6, pp. 3581–3593, 2004.

[3] P. Prado-Gutierrez, E. Mijares, G. Savio, M. Borrego, E. Martínez-Montes, and A. Torres, "Maturational time course of the envelope following response to amplitude-modulated acoustic signals in rats," *Int. J. Audiol.*, vol. 51, no. 4, pp. 309–316, 2012.

[4] S. J. Aiken and T. W. Picton, "Envelope following responses to natural vowels," *Audiol. Neurotol.*, vol. 11, no. 11, pp. 213–232, 2006.

[5] V. Easwar, D. W. Purcell, S. J. Aiken, V. Parsa, and S. D. Scollie, "Effect of stimulus level and bandwidth on speech-evoked envelope following responses in adults with normal hearing," *Ear Hear*, vol. 36, no. 6, pp. 619–634, 2015.

[6] V. Easwar, A. Banyard, S. Aiken, and D. Purcell, "Phase delays between tone pairs reveal interactions in scalp-recorded envelope following responses," *Neurosci. Lett.*, vol. 5, no. 665, pp. 257–262, 2018.

[7] S. J. Aiken and T. W. Picton, "Envelope and spectral frequency-following responses to vowel sounds," *Hearing Res.*, vol. 245, nos. 1–2, pp. 35–47, Nov. 2008.

[8] B. Boets, J. Wouters, A. van Wieringen, and P. Ghesquiere, "Auditory processing, speech perception and phonological ability in pre-school children at high-risk for dyslexia: A longitudinal study of the auditory temporal processing theory," *Neuropsychologia*, vol. 45, no. 8, pp. 1608–1620, 2007.

[9] K. Lehongre, F. Ramus, N. Villiermet, and D. A. L. S. Giraud, "Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia," *Neuron*, vol. 72, no. 6, pp. 1080–1090, 2011.

[10] V. Easwar, A. Banyard, S. J. Aiken, and D. W. Purcell, "Phase-locked responses to the vowel envelope vary in scalp-recorded amplitude due to across-frequency response interactions," *Eur. J. Neurosci.*, vol. 48, no. 10, pp. 3126–3145, 2018.

[11] V. Easwar, S. Scollie, and D. Purcell, "Investigating potential interactions between envelope following responses elicited simultaneously by different vowel formants," *Hear Res.*, vol. 1, no. 380, pp. 35–45, 2019.

[12] V. Easwar, S. Scollie, S. Aiken, and D. Purcell, "Test-retest variability in the characteristics of envelope following responses evoked by speech stimuli," *Ear Hear*, vol. 41, no. 1, pp. 150–164, 2020.

[13] P. Menell, K. I. McAnally, and J. F. Stein, "Psychophysical sensitivity and physiological response to amplitude modulation in adult dyslexic listeners," *J. Speech Lang Hear Res.*, vol. 42, pp. 797–803, Aug. 1999.

[14] B. Boashash, *Time Frequency Signal Analysis and Processing*. London, U.K.: Elsevier, 2003.

[15] H. Tang, J. Brock, and B. W. Johnson, "Sound envelope processing in the developing human brain: A MEG study," *Clin. Neurophysiol.*, vol. 127, no. 2, pp. 1206–1215, 2016.

[16] N. Weisz and C. Lithari, "Amplitude modulation rate dependent topographic organization of the auditory steady-state response in human auditory cortex," *Hearing Res.*, vol. 354, pp. 102–108, Oct. 2017.

[17] K. V. Nourski *et al.*, "Temporal envelope of time-compressed speech represented in the human auditory cortex," *J. Neurosci.*, vol. 49, pp. 15564–15574, Dec. 2009.

[18] R. E. Millman, G. Prendergast, M. Hymers, and G. G. Green, "Representations of the temporal envelope of sounds in human auditory cortex: Can the results from invasive intracortical 'depth' electrode recordings be replicated using non-invasive MEG virtual electrodes," *NeuroImage*, vol. 64, pp. 96–185, Jan. 2013.

[19] D. W. Purcell and M. S. John, "Evaluating the modulation transfer function of auditory steady state responses in the 65 Hz to 120 Hz range," *Ear Hear*, vol. 31, no. 5, pp. 667–678, 2010.

[20] D. Regan, *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. New York, NY, USA: Elsevier, 1989, pp. 70–98, 112–123, and 273–275.

[21] S. J. Kiebel, C. Tallon-Baudry, and K. J. Friston, "Parametric analysis of oscillatory activity as measured with EEG/MEG," *Hum Brain Mapp.*, vol. 26, no. 3, pp. 170–177, 2005.

[22] E. Martínez-Montes, E. R. Cuspineda-Bravo, W. El-Deredy, J. M. Sánchez-Bornot, A. Lage-Castellanos, and P. A. Valdés-Sosa, "Exploring event-related brain dynamics with tests on complex valued time–frequency representations," *Stat Med.*, vol. 27, no. 15, pp. 2922–2947, 2008.

[23] R. Sameni and E. Seraj, "A robust statistical framework for instantaneous electroencephalogram phase and frequency estimation and analysis," *Physiol. Meas.*, vol. 38, no. 12, p. 2141, 2017.

[24] E. Mijares-Nodarse, M. C. Pérez-Ábalo, A. Torres-Fortuny, M. Vega-Hernández, and A. Lage-Castellanos, "Maturational changes in the human envelope-following responses," *Acta Otorrinolaringol Esp.*, vol. 63, no. 4, pp. 258–264, 2012.