



The impact of temporal synchronisation imprecision on TRF analyses

Sara Carta^{a,b,1}, Anthony M.A. Mangiacotti^{c,d,1}, Alejandro Lopez Valdes^{e,f,g,h},
Richard B. Reilly^{e,g,h,i}, Fabia Franco^c, Giovanni M. Di Liberto^{a,b,g,*}

^a ADAPT Centre, Trinity College, The University of Dublin, Ireland

^b School of Computer Science and Statistics, Trinity College, The University of Dublin, Ireland

^c Department of Psychology, Middlesex University, London, United Kingdom

^d FISPPA Department, University of Padova, Padova, Italy

^e Trinity Centre for Biomedical Engineering, Trinity College, The University of Dublin, Ireland

^f Global Brain Health Institute, Trinity College, The University of Dublin, Ireland

^g Trinity College Institute of Neuroscience, Trinity College, The University of Dublin, Ireland

^h School of Engineering, Trinity College, The University of Dublin, Ireland

ⁱ School of Medicine, Trinity College, The University of Dublin, Ireland

ARTICLE INFO

Keywords:

EEG
Auditory perception
Temporal response function
Signal processing
Speech
Music
Imagery

1. Introduction

Recent research in cognitive science has seen the rapid increase in the use of linear modelling to investigate links between behaviour, cognition, and disease. The study of sensory processing has been particularly impacted by such methodologies, with one notable approach being the temporal response function (TRF), which describes the temporal relationship between selected features of a sensory stimulus and the corresponding neural signals (Ding and Simon, 2012; Lalor et al., 2006). TRF analyses have been used to investigate sensory processing with various neural recording technologies, such as non-invasive electro- and magneto-encephalography (EEG/MEG; (Brodbeck et al., 2018b, Liberto et al., 2015a), electrocorticography (ECoG; (Mesgarani et al., 2014), and functional magnetic resonance imaging (fMRI; (Valente et al., 2014, Santoro et al., 2017)). In that context, TRF analyses are ground-breaking in that they enable the study of sensory perception in realistic scenarios, involving stimuli such as natural speech, music, and cartoons. This has brought up new opportunities for investigating cognition in cohorts of participants that could find traditional paradigms

uncomfortable and challenging (e.g., excessive distress), such as children with neurodevelopmental deficits and older adults with neurocognitive impairment (Liberto et al., 2018, Meyer et al., 2021, Mesik et al., 2021, Alickovic et al., 2020, Broderick et al., 2021). Translational research typically faces cohort-specific experimental constraints, involving limitations on the recording time and favouring devices that are portable and with a reduced number of sensors and capabilities, at the cost of recording quality. A particular issue corresponds to the temporal synchronisation between the sensory events and the neural recording, which may present varying degrees of imprecision in different devices and experimental setup. Another scenario that we consider is when the temporal imprecision is due to the task itself, with variability within or between participants, or both, as with auditory imagery tasks. This study answers the question of how imprecisions in the EEG temporal synchronisation can impact TRF analyses, providing a novel method to identify and estimate the extent of the problem. Finally, we present a case study on a pilot investigation on older participants with neurocognitive impairment in care-homes, where the EEG temporal synchronisation was particularly problematic.

* Corresponding author at: ADAPT Centre, Trinity College, The University of Dublin, Ireland.

E-mail address: gdiliber@tcd.ie (G.M. Di Liberto).

¹ These authors contributed equally

<https://doi.org/10.1016/j.jneumeth.2022.109765>

Received 22 July 2022; Received in revised form 17 November 2022; Accepted 2 December 2022

Available online 5 December 2022

0165-0270/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Investigating sensory processing with technologies such as EEG requires the precise temporal alignment of the sensory stimulus and the recorded neural signal (Woodman, 2010). This is a crucial processing operation when the investigation involves measuring the neural response to a given sensory input or class of sensory inputs. The information required for the temporal alignment is generally recorded as temporal triggers, indicating an identification code of the sensory stimulus and a timestamp indicating when it occurs (e.g., start) in the neural recording. Temporal triggers are recorded via wired setups in traditional research laboratory settings, which is a low-latency reliable solution. Recent technological developments have led to a variety of new solutions (Hairston et al., 2014; Bilucaglia et al., 2020; Ries et al., 2014). Portable EEG systems, for example, perform the temporal synchronisation of multiple streams of information, such as the neural signal and temporal triggers, by using wireless protocols, like the Lab Streaming Layer (Kothe et al., 2014). While these solutions are becoming both widely accepted for scientific investigations and crucial for out-of-lab applications, it is expected that they lack precision compared with typical wired laboratory settings. As such, it is important that the researchers are aware of 1) the magnitude and characteristic of the temporal imprecision with their recording setup; and 2) the impact of that temporal imprecision on the analysis they are conducting. While the effect of temporal imprecisions has been studied for traditional event related potentials (ERPs; (Hairston, 2012), there is no investigation on how that impacts TRF analyses to date. The distinct analysis methodologies between ERPs (time-locked averaging) and neural tracking (e.g., multiple lagged regression), as well as the fundamentally different sensory stimuli employed (repeated discrete events vs. continuous sensory stimuli) lead to different brain activations (Bonte et al., 2006). These fundamental differences raise the need for a dedicated assessment of the consequences that temporal desynchronisation bring upon TRFs. It is the goal of this investigation to determine the extent to which this desynchronisation may be problematic, offering important insights to the prompt identification of potential undesired synchronisation effects and proposing a method to mitigate their impact.

Another important scenario leading to temporal imprecisions involves tasks that inherently present temporal imprecision between the participants' perception or actions and the neural recording. One remarkable example is auditory imagery tasks, where the paradigm as well as the participant skills impact the precise synchronisation between the expected and actual timing of the imagery action (e.g., Martin et al., 2014). Previous studies demonstrated several strategies that can guide the participants with their task, guaranteeing some level of synchronisation. Part of these approaches led to controlled but less naturalistic imagery tasks (e.g., auditory imagery during silent gaps), while others developed a more realistic music imagery task by using a metronome, but only focusing with expert musicians who could appropriately perform the task with precise timing. The present investigation quantifies the minimum synchronisation precision that is required to perform TRF analyses on neural recordings during such tasks. Furthermore, it is important to highlight that speech listening experiments are also impacted by temporal imprecisions, with the notable example of word comprehension, where the exact timing of word comprehension is unknown and typically only studied in relation to word onsets (Kutas and Federmeier, 2011; Broderick et al., 2018b).

A TRF analysis captures an input-output relationship that is experiment-specific, meaning that it may reflect different neural underpinnings (e.g., evoked responses, neural oscillations) in distinct experiments (Obleser and Kayser, 2019). Hence, considerations on the neural basis of a TRF should be confined to each experiment. In general, the intuition is that TRFs estimate the systematic reaction in the neural activity corresponding to a specific event (e.g., auditory input). In other words, the TRF is an estimate of the system's impulse response (Crosse et al., 2016a; Crosse et al., 2021). In the context of rapidly changing stimuli, such as speech and music, the systematic stimulus to neural signal relationship exhibits a phenomenon which is referred to as

cortical entrainment in the broad sense, where the neural signal tracks a given property of the sensory input (a.k.a. neural tracking) (Obleser and Kayser, 2019). TRFs presented the field with a new opportunity of measuring the neural substrate of auditory perception based on EEG recordings where participants listen to continuous sounds such as natural speech (e.g., audio-stories) (Ding and Simon, 2012; Lalor and Foxe, 2010) and music (Liberto et al., 2020). Low-frequency cortical signals measured with EEG were shown to encode the hierarchical processing of speech, tracking information from acoustic-phonetic to semantic levels (Liberto et al., 2015, Liberto et al., 2021c, Brodbeck et al., 2018a, Broderick et al., 2018b, Teoh et al., 2019). This neural tracking phenomenon reflects attention (e.g., auditory selective attention, where stronger tracking is measured for attended vs. ignored sounds; (Ding and Simon, 2014, O'sullivan et al., 2014), comprehension, as well as the level of consciousness of the participants (Legendre et al., 2019). Crucially, those measurements can be obtained with a limited number of electrodes and short experimental time (Di Liberto and Lalor, 2017; Jessen et al., 2019), which makes this methodology particularly suitable for field experiments and for studying vulnerable groups.

The present investigation focusses on a particular TRF implementation based on envelope tracking and lagged linear regression (mTRF-Toolbox; (Crosse et al., 2016b). This method was proven effective in applied settings involving realistic tasks, such as watching a cartoon, and in various applied cohorts, such as infants (Attaheri et al., 2022; Kalashnikova et al., 2018; Jessica Tan et al., 2022), children with dyslexia (Liberto et al., 2018), and older adults (Brodbeck et al., 2018b, Broderick et al., 2021). While the present study is methodological and informs us on TRF analyses in general, we discuss the case of speech and music to provide a clear focus. Specifically, we quantified the impact of temporal imprecision on the sound envelope TRF of publicly available datasets. The study proposes a methodology to assess the negative impact of temporal imprecision in the neural data by applying a TRF-based re-alignment attempting to recover the correct TRF, by fixing the N1 latency to a predefined value. Finally, we present the results of correcting for temporal imprecision in newly recorded data involving older participants (>80 yrs) with neurocognitive impairment in care-home settings. Due to older adults' comorbidities, such as mild cognitive impairment, physical disabilities or psychiatric disorders, some participants are unable to access outdoor laboratory services (Brunnhuber et al., 2014), thus limiting the possibility to promote their participation in research. Recent developments in EEG technology resulted in the availability of various portable EEG systems tailored for out-of-lab scenarios, including care-homes (e.g., Nielsen Telemedical, mBrainTrain, Emotiv, Neurosky, g.tec, BrainVision). While the development of such devices has been largely promoted targeting brain computer interface systems for the entertainment field, recent studies have found that they are sufficiently reliable to be used in applied settings for scientific or clinical purposes (Badcock et al., 2013; Badcock et al., 2015; Sintotskiy and Hinrichs, 2020). Challenges of working with older adults living in care-homes, such as high participant dropout rate (due to participants' frail condition and fatigue, experimental location, duration, and encumbrance of the EEG setup) can potentially be overcome by using rapid experimental procedures involving portable devices with few electrodes and a rapid (e.g., dry-electrodes) and comfortable (e.g., wireless) setup. While it is crucial to estimate the trigger synchronisation precision of any given EEG recording set up, the present study discusses a limit-case where cohort-specific requirements and technical limitations led to the acquisition of data with particularly low temporal synchronisation. Finally, we discuss how the validation and mitigation strategies proposed in this study can be used to detect whether temporal synchronisation was an issue or not in a given dataset.

2. Part I: Investigating temporal imprecision on the TRF analysis

In this part of the study, we assessed the impact of synchronisation imprecision by simulating a progressively larger synchronisation jitter

on two publicly available EEG datasets. We then proposed a realignment approach to detect and estimate the negative impacts of the temporal imprecision.

3. Datasets

3.1. Dataset 1: Speech listening task

3.1.1. Participants and experimental design

Data from this experiment was part of a set of studies examining how human cortical signals track the envelope and phonemic content of speech (Liberto et al., 2015, Broderick et al., 2018b, Crosse et al., 2016b, O’Sullivan et al., 2014). 19 participants (13 male) aged between 19 and 38 years participated in the first experiment. All participants were native English speakers, and reported normal hearing, normal or corrected-to-normal vision, and no history of neurological disease. The experiment was conducted in a single session for each participant. EEG data were recorded as participants listened to a single professional audiobook version of a popular mid-20th century American work of fiction. The audio stimuli were presented in 20 trials (chapters), each of about 180 s, preserving the storyline, with neither repetitions nor discontinuities, and with an average speech rate of ~210 words/min.

3.1.2. EEG data acquisition

128-channel EEG data (plus two mastoid channels) were acquired at a rate of 512 Hz using an ActiveTwo system (BioSemi). Triggers indicating the start of each trial were sent by the stimulus presentation computer and included in the EEG recordings to ensure synchronisation. Testing was carried out in a dark, sound-attenuated room and participants were instructed to maintain visual fixation on a crosshair centred on the screen for the duration of each trial, and to minimise eye blinking and all motor activities. Participants were free to have breaks in-between trials for as long as they needed. Stimuli were presented at a sampling rate of 44100 Hz using Sennheiser HD650 headphones and Presentation software from Neurobehavioral Systems (<http://www.neurobs.com>). The original dataset is available on Dryad (Broderick et al., 2018a). In the present study we used a standardised version of the dataset according to the Continuous-event Neural Data structure (CND; <https://cnsworkshop.net>).

3.2. Dataset 2: Monophonic music listening task

3.2.1. Participants and experimental design

Data from this experiment was part of studies examining how human cortical signals track the acoustics and melodic expectations of music (Liberto et al., 2020, Liberto et al., 2021b). All participants reported normal hearing and no history of neurological disease and were paid for participating. 20 participants (10 female) aged between 23 and 42 years (median = 29) participated in this experiment. Ten of them were highly trained musicians with a degree in music and at least 10 years of active music experience, whereas the other participants had no musical training. The experiment was conducted in a single session. EEG data were recorded as participants listened to monophonic MIDI versions of 10 music pieces from Bach’s monodic instrumental corpus. Stimuli were partitioned into short snippets of about 150 s. The selected melodies were originally extracted from violin (partita Bach Works Catalog BWV 1001, presto; BWV 1002, allemande; BWV 1004, allemande and gigue; BWV 1006, loure and gavotte) and flute (partita BWV 1013 allemande, corrente, sarabande, and bourrée angloise) scores and were synthesised by using piano sounds with MuseScore 2 software, each played with a fixed rate (between 47 and 140 bpm). Each 150 s piece, corresponding to an EEG trial, was presented three times throughout the experiment, adding up to 30 trials that were presented in a random order. At the end of each trial, participants were asked to report on their familiarity with the piece (from 1 = unknown to 7 = know the piece very well). This rating could consider both their familiarity with the piece on first

occurrence in the experiment as well as the build-up of familiarity across repetitions. Participants reported repeated pieces as more familiar (paired *t*-test on the average familiarity ratings for all participants across repetitions: rep2 > rep1, $p = 6.910^{-6}$; rep3 > rep2, $p = 0.003$, Bonferroni correction). No significant difference emerged between musicians and non-musicians on this account (two-sample *t*-test, $p = 0.07, 0.16, 0.19$ for repetitions 1, 2, and 3, respectively).

3.2.2. EEG data acquisition

64-channel EEG data (plus two mastoid channels) were acquired at a rate of 512 Hz using an ActiveTwo system (BioSemi). Triggers indicating the start of each trial were sent by the stimulus presentation computer and included in the EEG recordings to ensure synchronisation. The study was undertaken in accordance with the Declaration of Helsinki and was approved by the CERES Committee of Paris Descartes University (CERES 2013–11). The participants provided their written informed consent to participate in this study. Testing was carried out at École Normale Supérieure (Paris, France) in a dark, electrically shielded, sound-proof room. Participants were instructed to maintain visual fixation on a crosshair centred on the screen for the duration of each trial, and to minimise eye blinking and all motor activities. Participants were free to have breaks in-between trials for as long as they needed. Stimuli were presented at a sampling rate of 44100 Hz using Sennheiser HD650 headphones and Presentation software from Neurobehavioral Systems (<http://www.neurobs.com>). The original dataset is available on Dryad (Di Liberto et al., 2021d). In the present study we used a standardised version of the dataset according to the Continuous-event Neural Data structure (CND; <https://cnsworkshop.net/resources.html>).

3.3. EEG signal analysis

3.3.1. Pre-processing

The same pre-processing procedure was used across all datasets. First, the broadband amplitude envelope of the speech signal (**Env**) was calculated using the Hilbert transform of the acoustic audio signal. Then, EEG data were analysed offline using MATLAB software (The Mathworks Inc.), digitally filtered between 1 and 8 Hz using a Butterworth zero-phase filter (low- and high-pass filters both with order 2 and implemented with the function *filtfilt*). Signals were down-sampled to 128 Hz and re-referenced to the average of the mastoid channels. To identify channels with excessive noise, the time series were visually inspected, and the standard deviation of each channel was compared with that of the surrounding channels. Channels contaminated by noise were recalculated by spline interpolating the surrounding clean channels in EEGLAB (Delorme and Makeig, 2004).

3.3.2. Objective auditory processing assessment

System identification was used to compute the channel-specific mapping between **Env** and the pre-processed EEG data. This method allows us to estimate TRFs describing the spatio-temporal dynamics that underlie the speech-neural coupling (Ding et al., 2014; Lalor et al., 2009). We used multivariate regularised lagged linear regression (Crosse et al., 2016a; Crosse et al., 2021) to estimate a subject-specific filter describing how the brain transforms acoustic envelopes into the corresponding neural response (forward model; Fig. 1-left). The TRF takes into consideration multiple time-lags between stimulus and neural signal, providing us with patterns of model weights interpretable in both space (scalp topographies) and time (speech-EEG latencies). Leave-one-out cross-validation (across trials) was used to assess how well the model can predict unseen data. This was quantified by calculating Pearson’s correlation between the recorded signals and the corresponding predictions for each scalp electrode (EEG prediction correlations; Fig. 1-right). Prediction correlations were calculated per trial. Then, such correlations were averaged, providing us with a single correlation value for a given subject. Prediction correlations were estimated for regularisation parameters $\lambda = [10^{-6}, 10^{-4}, \dots, 10^4]$.

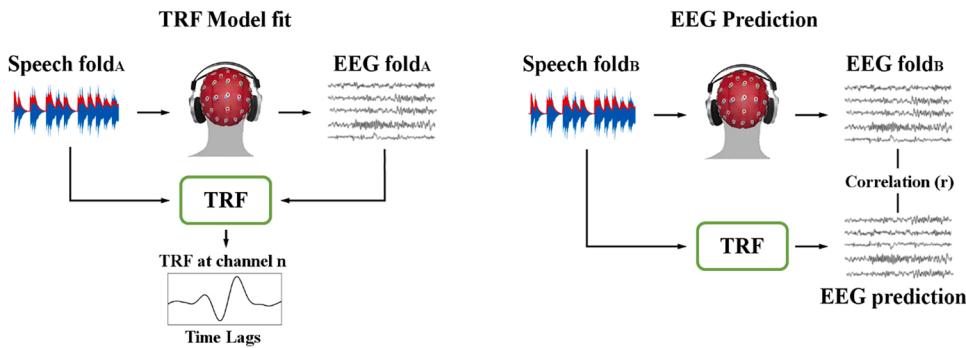


Fig. 1. : The Temporal Response Function Framework. Lagged linear regression is run to estimate how a given stimulus feature (e.g., sound envelope) is transformed into the electrical neural activity recorded with EEG. First, this mapping is estimated on a training portion of the data. Next, the quality of the estimated model is evaluated by predicting a separate portion of data (test set) and by calculating the prediction correlation. The procedure is then iterated for all possible test-sets (cross-validation).

The lambda value corresponding to the highest prediction correlation was selected. In all datasets, the time-lag window $[-100, 500]$ ms was used to fit the TRF models which were selected to capture the expected neural responses at latencies between about 0 and 250 ms (Liberto et al., 2015) as well as the signal-to-noise ratio of such responses when compared with latencies that should not reflect significant speech-EEG interactions (such as TRF responses at negative lags). Note that this approach provides two complementary modes to objectively assess the neural tracking of the speech envelope: The first view involves studying the regression weights, which allows us to assess the spatio-temporal dynamics of the envelope-EEG coupling. A second view involves measuring the EEG prediction correlations for each EEG channel to determine the reliability of the model and the strength of the neural tracking.

3.3.3. Simulating synchronisation imprecision

Datasets 1 and 2 were collected with wired, laboratory-grade Bio-Semi ActiveTwo systems with a sampling frequency of 512 Hz, thus guaranteeing high signal quality and temporal precision. Specifically, triggers were delivered via the parallel port in Dataset 1 and via a custom-made trigger-box in Dataset 2. Both solutions guarantee high precision for the trigger timing (< 1 ms). It is important to note that this precision relies on the appropriate implementation of the stimulus presentation, especially at the level of software in the presentation PC (e.g., the audio file must be fully buffered before sending the trigger).

All datasets used a single temporal trigger to mark the beginning of relatively long (> 1 min) EEG trials. To simulate temporal imprecision, which could be due to factors such as imprecise triggers or temporal imprecision in the task (e.g., auditory imagery), a random jitter was applied to produce a progressively more pronounced misalignment between the sound stimulus and the EEG signal for each trigger (hence, for each trial). We simulated temporal jitters spanning from minimal to unacceptably high (compared to the width of typical sound envelope TRF components): $M = [5, 10, 25, 50, 100, 150, 200, 400, 800, 1600]$ ms. For each trial and M_i , a different shift was applied between stimulus and EEG from a uniform distribution spanning from $-M_i$ to M_i . TRF analyses were conducted for the original dataset as well as each of the datasets with simulated temporal imprecision. This procedure was carried out for both Datasets 1 and 2. Please note that jittering the EEG triggers also serves as a simulation for other types of imprecision, for example due to the particular task rather than technological limitations or issues.

3.3.4. TRF-based temporal re-alignment procedure

A re-alignment procedure was devised and applied on the datasets with simulated temporal imprecision. It is important to note that the main goal of the proposed re-alignment strategy is not to correct for poorly synchronised data, but instead to determine whether the dataset presents (unexpected) substantial temporal misalignments or not.

The procedure consists of fitting a TRF model for each trial and by re-aligning the EEG data by forcing the dominant negative component (the

N1) to emerge at a fixed, arbitrary time t_{N1} . Indeed, this imposes the strong assumption that an N1 component emerges in all experimental trials TRFs, that such component occurs at a precise sound-EEG latency, and that the component has the largest negative weight. This assumption may not always be appropriate, depending on the goals of the experiment. The choice of re-aligning by focusing on the N1 component was based on the typical shape of auditory TRF measured with EEG, which includes two dominant positive components (P1 and P2), and one dominant negative component (N1) (Liberto et al., 2015). The presence of two dominant positive components at different latencies implies that the largest positive TRF weight could correspond to either component, thus causing significant uncertainty in the alignment, probably aggravated by inter-participant variability. Instead, a single dominant negative correlation was expected for the time-latency corresponding to the N1 component, making this the most reliable choice for precise sound-EEG alignment. According to previous research on Dataset 1 (Liberto et al., 2015), the N1 component was expected at a sound-EEG latency of 78 ms. As such, the alignment based on the largest negative correlations between Env and EEG was assigned to a fixed value of 78 ms.

The re-alignment procedure consists of fitting un-regularised forward TRFs by using the lagged regression procedure in the *mTRF-Toolbox* (*mTRFtrain* function) for each experimental trial. Note that this TRF fit was only performed for determining the re-alignment shift and was distinct from the TRF analysis used subsequently to assess the relationship between stimulus and EEG. The choice of un-regularised regression in the TRF fit for the re-alignment procedure is motivated by computational reasons, along with a potentially detrimental smoothing effect that regularisation may cause on the TRF per se. The TRF lag window was arbitrarily set to $[-0.2, 0.2]$ seconds, meaning that the algorithm could tolerate a maximum temporal misalignment of $[-0.2-t_{N1}, 0.2-t_{N1}]$ seconds (where t_{N1} is the time-lag of the N1 TRF component). Next, the largest negative TRF weight was identified, by computing all channels' global field power and assigning to it the polarity of channel FCz, and the data was shifted to make that negative component emerge at 78 ms. Fitting a TRF will likely cause side artifacts, which may in turn determine multiple negative peaks. To solve this issue in our re-alignment procedure, we removed from the search the initial and final five samples of our regression window. It is important to note that our results should be interpreted solely in terms of the shape and topographical distribution of the TRFs and the EEG prediction correlation while, again, they cannot be interpreted on the absolute latency values. Please also note that the procedure tends to inflate the N1 component, as the alignment is performed on that component. We also advise to assess the results with metrics such as the overall TRF shape or the N1-P2 magnitude, rather than the N1 magnitude, which is inflated by construction as the alignment is performed on that TRF component.

To assess the significance of the re-alignment procedure, two control analyses were conducted. First, the TRF procedure was re-run after randomising the trial indices of the stimulus envelopes, producing a

mismatch control (MM). A second control was instead carried out by time-reversing the stimulus envelope (REV). In these control conditions, re-alignment was applied following the usual procedure of shifting the largest negative components to the reference 78 ms latency. Since, in these cases, the TRF was not systematically shifted according to certain trigger ranges, but it was either mismatched, by randomly shuffling trial indices, or temporally reversed (by time-reversing the stimulus envelope), the N1s were found at random latencies, as expected, and their latency and standard deviation were similar to those of the largest trigger jitters.

4. Results

4.1. Datasets 1: Speech listening task

Forward TRFs were fit to assess how strongly EEG signals encode speech envelope information and to determine the temporal dynamics of that encoding. Then, we determined how precisely that encoding could be retrieved when considering simulated scenarios with progressively less precise temporal synchronisation or when applying a control Mismatch of the trials. As a first step, we successfully replicated the TRF results obtained in the original paper corresponding to Dataset 1 (Di Liberto et al., 2015 Fig. 2A-left, zero jitter results), which is also in line with the TRF shape and latency in other previous related studies using different datasets (Vanthornhout et al., 2019). The TRF analysis conducted after applying jittering confirmed that larger temporal

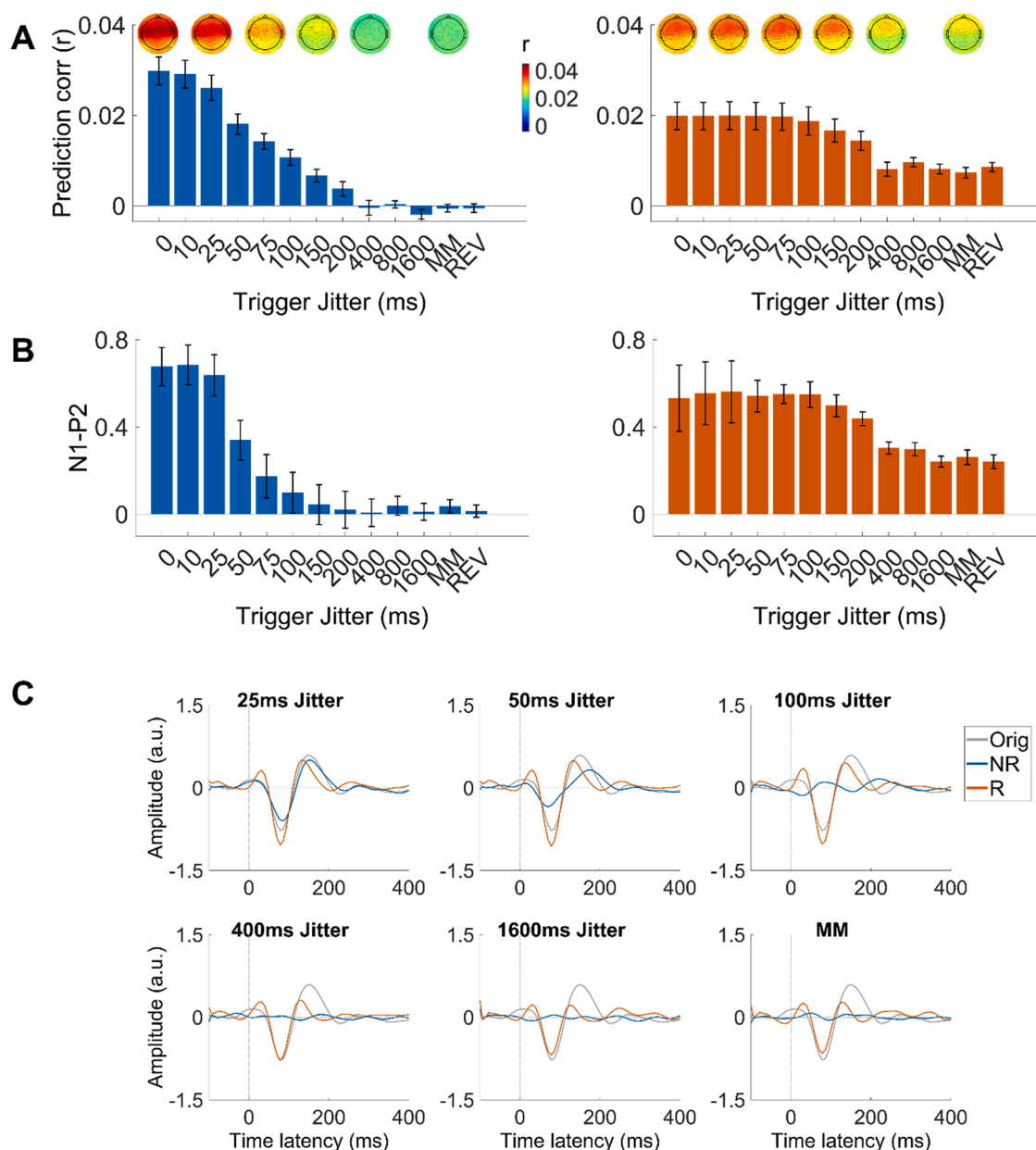


Fig. 2. : Impact of synchronisation imprecision on the envelope TRF when considering a natural speech listening task (Dataset 1). (A,B) EEG prediction correlation (avg across all EEG channels) and N1-P2 magnitude (FCz) for progressively larger temporal imprecision before (left) and after (right) re-alignment. In addition to jitters of ± 10 ms, 25, 50, 75, 100, 150, 200, 400, 800, and 1600 ms, results are shown for the control conditions trial mismatch (MM) and stimulus time-reversed (REV). The topographies indicate the EEG prediction correlations across all scalp channels for selected jitters. (C) TRF weights at FCz for selected trigger jitters and control MM condition, before (NR) and after (R) re-alignment.

imprecisions progressively impact both TRF shape and EEG prediction correlations (ANOVA: $F(10,198) = 34.8, p = 1.8 \times 10^{-38}$; note that error bars refer to the standard error across individual participants; Fig. 2A-left), with significance against the mismatch control condition (MM) for jitters up to 200 ms (post-hoc FDR-corrected t -test, $p < 0.001$ for jitters up to 100 ms, and $p = 0.002, 0.465, 0.695, 0.465$ for jitters of 150, 200, 400, 800, and 1600 ms respectively).

Imprecise synchronisation was also expected to impact the shape of the retrieved TRF. We studied this effect by measuring the impact of the jitter on the N1-P2 TRF amplitude (P2 minus N1, or N1-P2 amplitude; Fig. 2B-left) as well as by assessing the correlation between the TRFs for the original (zero jitter) and jittered EEG data (Supplementary Figure 1). For simplicity, we focussed on the TRFs at the electrode FCz, which was expected to be particularly relevant (e.g., Di Liberto et al., 2015). The N1-P2 amplitude was also computed on channel FCz. Note that the main result does not change when considering other adjacent channels (e.g., Cz, FT7, FT8). Fig. 2B shows that the N1-P2 amplitude is affected by the reduced synchronisation (ANOVA: $F(10,198) = 8.11, p = 6.1 \times 10^{-11}$; Fig. 2B-left), with significance against MM for jitters up to 25 ms (post-hoc FDR-corrected t -test, $p = 0.04, 0.05, 0.05, 0.95, 1, 1, 1, 0.7, 1, 1, 1$ for jitters of 0, 10, 25, 50, 75, 100, 150, 200, 400, 800, 1600 ms respectively). The reduced N1-P2 amplitude that we observed could also be the result of opposing polarities cancelling each other when combining different trials, but this effect is limited by the small number of trials. On the other hand, it is also possible that these components are shifted in time and emerge at other relevant latencies.

The detailed TRFs are depicted in Fig. 2C for selected jitters. An additional analysis was carried out to more directly assess the accuracy of the re-alignment procedure by measuring the distance between the jitters and the shifts identified and used for the re-alignment (Supplementary Figure 2). Results were consistent with the patterns in Fig. 2A, B. Specifically, the larger the inaccuracy of the re-alignment, the smaller the resulting TRF prediction scores ($r = -0.76, p = 0.002$). In other words, an accurate re-alignment improves the final EEG prediction correlations for the realigned TRFs.

The re-alignment procedure attempted to restore the TRFs in case of large temporal imprecisions. Improved TRF scores (N1-P2 magnitude and prediction correlations) were expected only when considering scenarios with a substantial temporal imprecision. Fig. 2A indicates that the larger jitters impact the EEG prediction correlations even after re-alignment (ANOVA: $F(10,198) = 3.2, p = 8.0 \times 10^{-4}$, Fig. 2A-right), however with significance persisting up to 400 ms (post-hoc FDR-corrected t -test against MM, $p < 0.01$ for jitters up to 100 ms, $p = 0.0244, 0.120, 0.557, 0.407$ for jitters of 200, 400, 800, and 1600 ms respectively), hence larger than without re-alignment. When looking at the prediction correlation scores, the re-alignment strategy appears particularly beneficial for larger trigger jitters, especially after 100 ms. On the other hand, for the shorter jitters up to 50 ms, applying a re-alignment lowers the EEG prediction correlations. To evaluate this interesting pattern, we tested the impact of Trigger Jitters on prediction correlations scores. We compared, for each jitter, the prediction correlations before and after re-alignment, by using a Wilcoxon signed rank test (with FDR-correction for multiple comparisons). This revealed significant decreases of prediction correlations for re-alignment applied to small jitters and a significant increase for larger jitters (0 ms: $p = 0.010$, 10 ms: $p = 0.010$, 25 ms: $p = 0.007$, 50 ms: $p = 0.840$, 75 ms: $p = 0.040$, 100 ms: $p = 0.004$, 150 ms: $p = 0.001$, 200 ms: $p = 0.001$, 400 ms: $p = 0.009$, 800 ms: $p = 0.001$, 1600 ms: $p = 0.001$, MM: $p = 0.001$). Regarding the TRF shape, note that the re-alignment procedure is based on the N1 TRF component, as such it is important to clarify that the re-aligned TRFs are indeed expected to exhibit an N1 component by construction. A successful re-alignment is also expected to produce the other relevant envelope TRF components, especially the P2. The N1-P2 TRF amplitude showed that, in fact, the re-alignment procedure could recover the TRF shape for jitters up to 200 ms and that, in fact, jitter did not significantly affect the TRF shape overall

(Fig. 2B-right; ANOVA: $F(10,198) = 1.45, p = 0.16$). Finally, we would like to highlight that the re-alignment “forces” the time latency of the N1 TRF component to a pre-defined value, which should generally refer to the known literature. Since the current analysis gave us access to the original N1 latency, we used that exact value (78 ms). Nevertheless, in general, time-latencies in the TRF after the re-alignment procedure should be interpreted on their relative (latency between components and duration of a component) but not absolute values (i.e., the re-aligned N1 occurs at 78 ms because we decided so; while the P2 latency in relation to the N1 latency is informative). To clarify this concept, the re-alignment in the next section is performed based on the same N1 latency of 78 ms, which should lead to some differences between original and re-aligned TRFs in terms of absolute time-latencies.

4.2. Dataset 2: Monophonic music listening task

To test the generalisability of the above findings, the previous analysis procedure was performed again on Dataset 2. Dataset 2 involved music listening, which is an interesting additional challenge for the re-alignment procedure, as the stimulus includes inherent temporal regularities that may affect the algorithm. We successfully replicated the TRF results in the original paper corresponding to Dataset 2 (Di Liberto et al., 2020) (Fig. 3A-left, zero jitter results). We found that larger jitters progressively degraded both TRF shape and EEG prediction correlations (ANOVA: $F(10, 209) = 49.1, p = 1.8 \times 10^{-49}$; Fig. 3A-left), with significance against the mismatch control condition (MM) for jitters up to 200 ms (post-hoc FDR-corrected t -test, $p < 0.05$). As for the previous dataset, we tested how prediction correlations are impacted by trigger jitters. We compared, for each jitter, the prediction correlations before and after re-alignment (Wilcoxon sign rank test, FDR corrected) and, consistently with the previous dataset, we observed a significant decrease and increase in prediction correlations for re-alignments applied to small and large jitters respectively (0 ms: $p = 7.1 \times 10^{-4}$, 10 ms: $p = 7.1 \times 10^{-4}$, 25 ms: $p = 7.6 \times 10^{-4}$, 50 ms: $p = 0.030$, 75 ms: $p = 0.560$, 100 ms: $p = 0.250$, 150 ms: $p = 0.020$, 200 ms: $p = 0.002$, 400 ms: $p = 7.1 \times 10^{-4}$, 800 ms: $p = 0.002$, 1600 ms: $p = 7.1 \times 10^{-4}$, MM: $p = 0.040$). We then assessed the impact of imprecise synchronisation on the shape of the retrieved TRF (Fig. 3B-left and Supplementary Figure 1). Similar to what found for Dataset 1, the N1-P2 amplitude was found to be affected by the reduced synchronisation (ANOVA: $F(10, 209) = 13.1, p = 1.1 \times 10^{-17}$; Fig. 3B-left), with significance for jitters up to 25 ms (post-hoc FDR-corrected t -test, $p < 0.05$ for jitters up to 25 ms, all other jitters had average N1-P2 amplitude lower than baseline). The detailed TRFs are depicted in Fig. 3C for selected jitters.

Fig. 3A-right indicates no significant effect of jitter on the EEG prediction correlations after re-alignment is performed (ANOVA: $F(10,209) = 1.3, p = 0.21$; Fig. 3A-right). Regarding the TRF shape, the N1-P2 TRF amplitude showed that the re-alignment procedure could recover the TRF shape for all jitters, even though the jitter had a significant impact on the N1-P2 TRF amplitude (Fig. 3B-right; ANOVA: $F(10,209) = 2.0, p = 0.03$; post-hoc FDR-corrected t -test, $p < 0.05$ for all jitters). As for the previous dataset, the accuracy of the re-alignment procedure was also assessed by calculating the distance between jitters and re-alignment shifts. Results were consistent with the patterns in Fig. 3A,B and Supplementary Figure 2. Specifically, the larger the inaccuracy of the re-alignment, the smaller the resulting TRF prediction scores ($r = -0.62, p = 0.010$).

5. Part II: Use-case study demonstrating the detection of temporal imprecision.

Here we present a case study demonstrating the effectiveness of the temporal re-alignment procedure on the TRF analysis of a challenging EEG dataset recorded out-of-lab, in care-home settings.

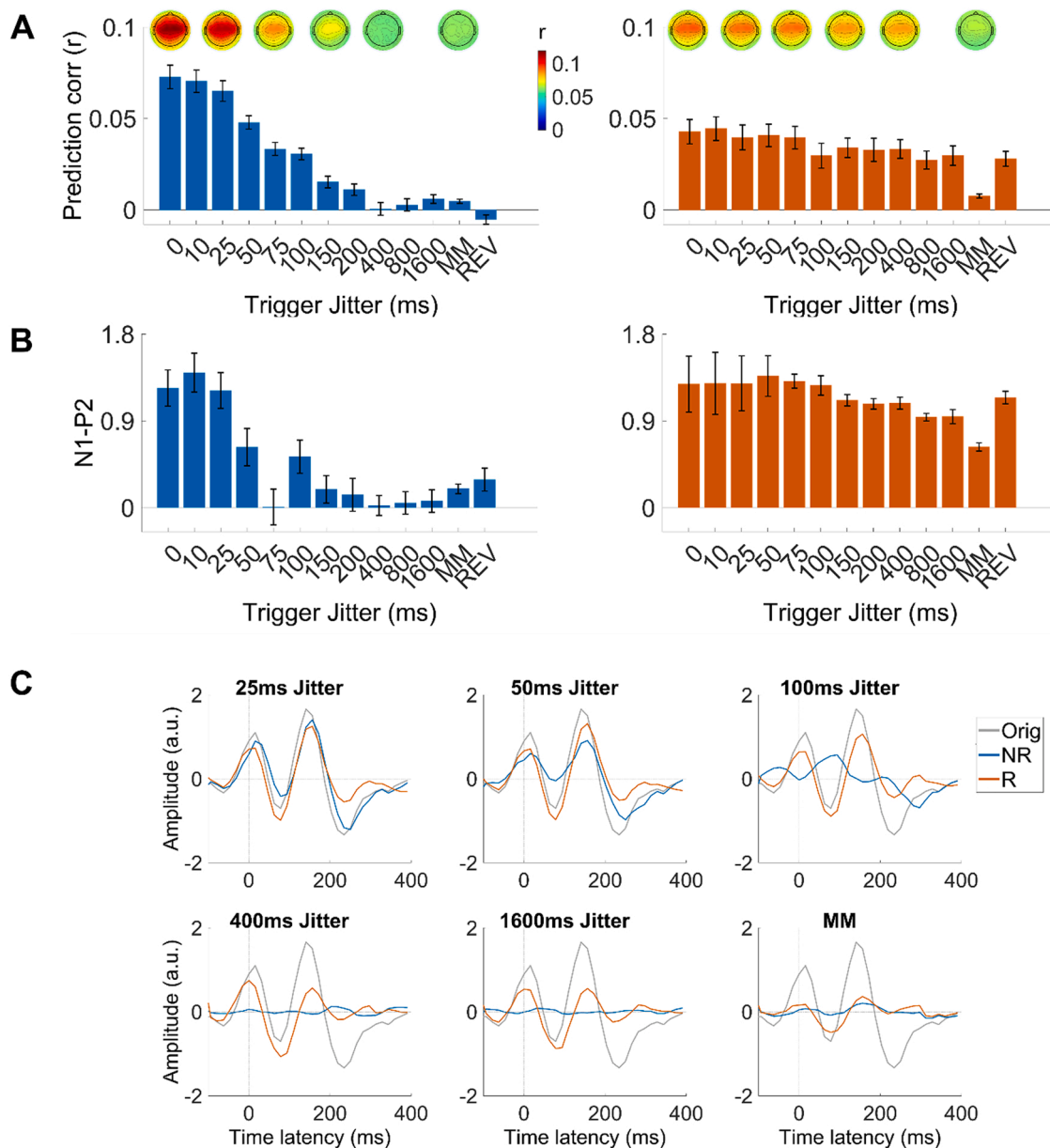


Fig. 3. : Impact of synchronisation imprecision on the envelope TRF when considering a music listening task (Dataset 2). (A,B) EEG prediction correlation (avg across all EEG channels) and N1-P2 magnitude (FCz) for progressively larger temporal imprecision before (left) and after (right) re-alignment. In addition to jitters of ± 10 ms, 25, 50, 75, 100, 150, 200, 400, 800, and 1600 ms, results are shown for the control conditions trial mismatch (MM) and stimulus time-reversed (REV). The topographies indicate the EEG prediction correlations across all scalp channels for selected jitters. (C) TRF weights at FCz for selected trigger jitters and control MM condition, before (NR) and after (R) re-alignment.

6. Dataset

6.1. Participants

Participants were a subsample of a larger pre/post 4-month randomised control trial (RCT) study studying the impact of speech and music therapy on older individuals with cognitive impairment (Mangiacotti et al., 2019). 54 older adults with mild to moderate cognitive impairment (see *Neuropsychological assessment* section) living in 5 different MHA care homes (Methodist Homes, UK), were randomly assigned to either a music therapy group (MT) or a storytelling group (ST). Participants from the MT and ST groups were matched by age, gender, health level, education, cognitive reserve and cognitive level (Tucker and Stern, 2011). Participants attended one-to-one 40 min of the assigned therapy. Participants were selected according to the following inclusion

criteria: i) aged 65 or older; ii) presence of mild to moderate cognitive impairment (Mini-mental State Examination, MMSE = 18–23); iii) no hearing impairment that would negatively interfere with participation in the proposed activities. Exclusion criteria were: a) presence of severe motor deficits; b) previous participation in a cognitive assessment within the last three months prior to the start of the study.

Participants completed a “paper and pencil” neuro-psychological test battery before the start of the interventions (pre: pre-test phase) and one week after the end of the 4-month intervention process (post: post-test phase). Tests were selected to assess cognitive functioning, with a specific focus on attention and executive functions (Mangiacotti et al., 2019). Analyses on the full sample ($N = 43$) indicated significant improvement in almost all cognitive tests for MT, while significant effects emerged for ST only in the verbal fluency test (see *Neuropsychological assessment* section) (Mangiacotti, 2020).

A total of $N_1 = 9$ (5 male) agreed to take part in the current feasibility study involving EEG recordings, 5 from the ST and 4 from the MT group (see Table 1). Information about the project was given in a meeting with staff and participants' family members, before the beginning of the recruitment process. All participants gave written informed consent, in accordance with the Declaration of Helsinki.

6.2. Neuropsychological assessment

Several neuropsychological tests were conducted with the participants, such as the Phonemic Verbal Fluency, Cumulative Illness Rating Scale, the Cognitive Reserve Index Questionnaire, and the Mini-mental State Examination. The present analysis will report only the results of the *Phonemic Verbal Fluency* test (Machado et al., 2009), which investigates lexical skills as well as the ability to organize an adequate verbal search strategy. Participants were asked to orally produce as many words as they could, which begin with a given letter, within a minute of time. The score was calculated as the average number of words found for each letter (three in total). This test was selected as it is a "fast, user-friendly speech quality assessment tool" (Opasso et al., 2016).

6.3. EEG data acquisition

Participants listened to excerpts from an audiobook ("The old man and the sea", by Ernest Hemingway) while limiting motor movements. This experimental design is based on previous studies on speech perception that used a larger set of the same auditory stimuli (Liberto et al., 2015, Liberto et al., 2019, Broderick et al., 2018a). While the original design involved about 60 min of audiobook listening (Di Liberto and Lalor, 2017), older adults were expected to agree to this EEG experiment only for much shorter durations. Here, we invited participants to undergo two sessions of a 3-min version of the audiobook listening EEG experiment which, considering the time required to position and remove the EEG equipment, required about 20 min in total per session. The second session was run after a 4-month intervention therapy.

EEG signal was recorded with a BioRadio™ system (Cleveland Medical Devices Inc.; Cleveland, OH). BioRadio is a portable wearable biomedical device for recording human physiological signals. Data is transmitted from the EEG unit to the acquisition Windows laptop wirelessly via Bluetooth. The signal is then recorded by BioCapture Software (Vilber Lourmat, France). Eight single-ended electrodes were placed on scalp locations that were most relevant for envelope tracking measurements based on previous studies: Cz, Pz, F3, F4, FC3, FC4, CP3, CP4 (Liberto et al., 2015). Left and right mastoids (A1, A2) were used respectively as reference and ground electrodes with the goal of maximizing the EEG responses to the auditory stimuli (Luck, 2005). Given the limited number of EEG channels and the short duration of the experiment, channel impedance was carefully monitored throughout the

experiment and no electrodes had to be excluded.

The same laptop was used both to present the audio stimuli and to record the EEG data (Fig. 4). A single 3-min audio file was presented with two Logitech Z506 loudspeakers placed in front of the participant at a one meter distance. The audio file was constructed as follows: 5 s silence, 1 s tone at 440 Hz, 5 s silence, 3 min audiobook. The experimenter played the audio file and waited for the expected acoustic cue (note A at 440 Hz) to press a manual trigger button connected to the BioRadio, providing us with a first, approximate synchronisation between the EEG and the audio stimulus. Tests occurred in a comfortable and well-lit room, acoustically attenuated and free of potentially interfering or disturbing events (e.g., other people walking, telephone ringing). All caregivers were notified of the experiment and a notable sign "do not disturb/enter" was placed outside the room.

6.4. Qualitative evaluation of the assessment

At the end of each EEG recording session, participants were asked to comment on the EEG assessment by answering verbal open-ended questions such as "Did you enjoy the story?" and "Did the EEG electrodes bother you?". This form of simple assessment was chosen both to minimise participant burden with further quantitative evaluation and to give them more freedom to express themselves. The answers were manually transcribed by the examiner. A qualitative content analysis of the responses was performed (Biasutti, 2013) with the aid of a software for qualitative text analysis ATLAS.ti (Scientific Software Development GmbH). The following response categories were obtained from the 'categorization' procedure: "Assessment experience", "Symptoms". Frequencies of the obtained categories were computed. Concerning to the first question "Did you enjoy the story?", all the participants (100%) reported positive comments saying that they "enjoy/liked" the story, two of which added that the narrator's voice was "monotonous" and another two reported that the story was "a little long". As to the second question "Did the electrodes bother you?", all the participants reported

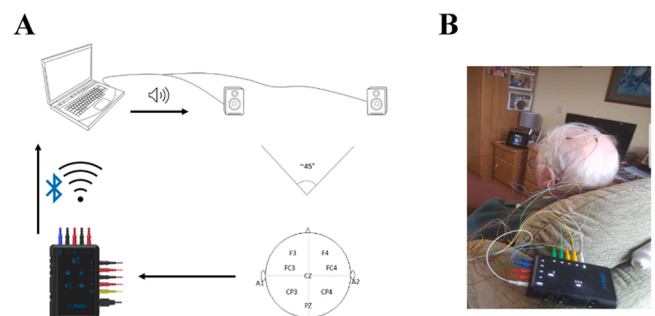


Fig. 4. : Layout (A) and photo (B) of the experimental setup for the use-case study.

Table 1

Participants demographic data. Note: CRIQ (Cognitive Reserve Index Questionnaire; (Nucci et al., 2012); CIRS (Cumulative Illness Rating Scale; (Linn et al., 1968); MMSE (MiniMetal State Examination; (Folstein et al., 1983).

ID	Group	Gender	Age	Education	CriQ	CIRS	MMSE
P1	ST	m	86	10	119	19	22
P2	ST	m	80	15	120	4	18
P3	ST	m	82	9	107	12	24
P4	ST	f	81	20	123	2	22
P5	ST	f	96	12	120	7	18
P6	MT	f	86	9	110	4	24
P7	MT	f	82	14	94	5	17
P8	MT	m	88	14	154	5	19
P9	MT	m	88	18	154	6	20
Mean		m= 55.6% f= 44.4%	87.56	12.11	118.67	7.11	20.22
SD			4.33	3.37	21.63	5.25	2.39

that electrodes did not bother them.

6.5. EEG data pre-processing

The use-case dataset was pre-processed as Datasets 1 and 2. The only difference compared to Datasets 1 and 2 is that signals were down-

sampled to 125 Hz (rather than 128 Hz) and that data was already referenced to the mastoid channels at the recording stage (rather than being re-referenced offline). Due to the limited number of EEG channels, signals were manually inspected for excessive noise (standard deviation calculated on a 10-s moving window). No channels were removed because of excessive noise across all nine participants.

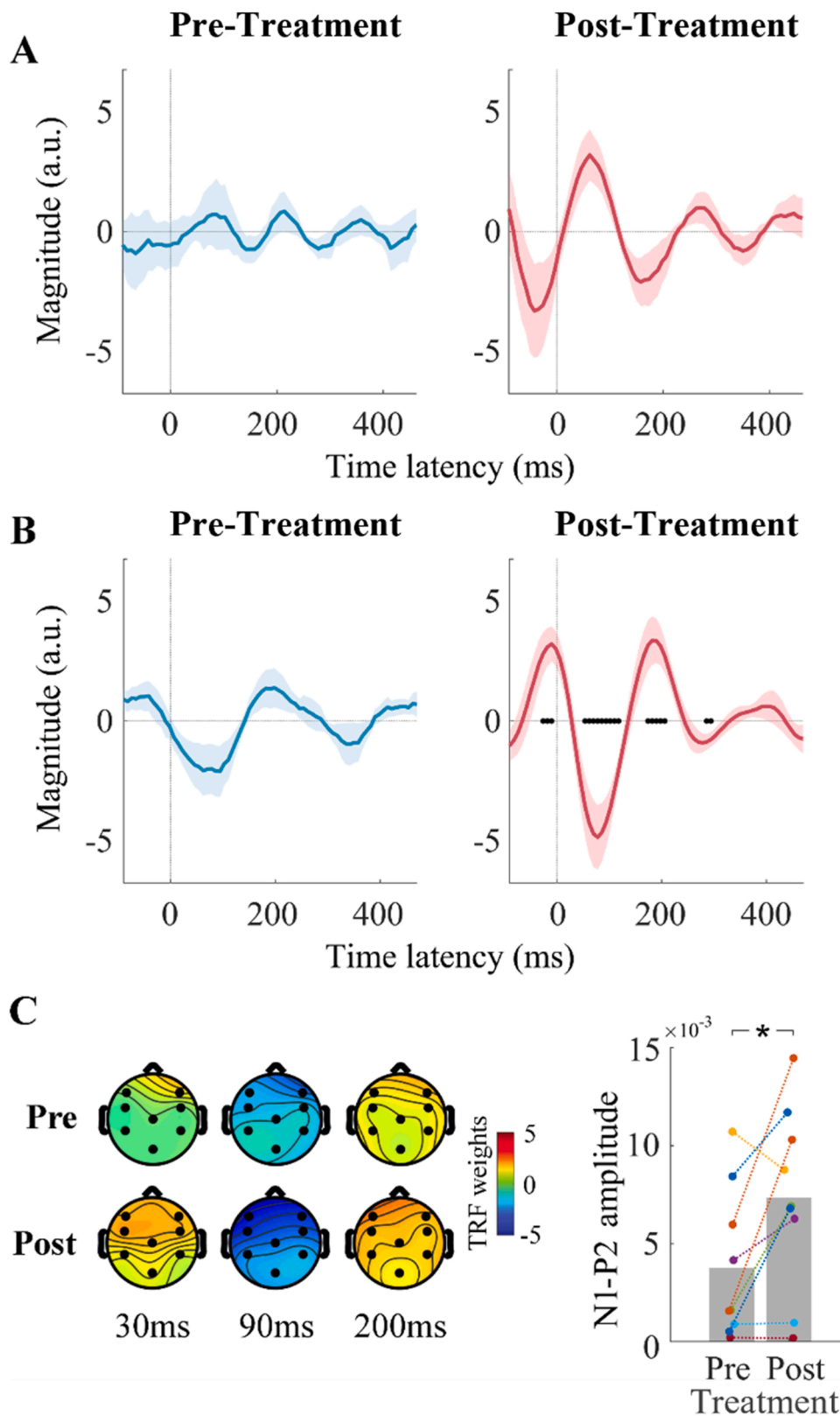


Fig. 5. (A) TRF weights calculated at channel Cz with the original EEG event triggers for the pre and post therapy recording sessions. Shaded areas indicate the standard error of the mean across participants. (B) TRF weights calculated at Cz after re-alignment. Black dots indicate TRF weights significantly different from zero (FDR-corrected Wilcoxon test, $p < 0.05$). (C) Left: Topographies of the TRF weights at selected time-latencies calculated on the re-aligned EEG data. Right: N1-P2 amplitude calculated on the re-aligned EEG data.

6.6. Objective auditory processing assessment

The analysis procedure was conducted as for Datasets 1 and 2.

7. Results

7.1. Cognitive psychological tests

Pre/post intervention improvement of the performance in the *Phonemic Verbal fluency test* was found in both MT and ST groups in the full sample of 43 participants (Mangiacotti, 2020). Results from the *pre* ($M = 5.61$, $SD = 1.87$) and *post* ($M = 5.84$, $SD = 1.94$) *t*-test analysis on the verbal fluency test in the sub-sample used here ($N = 9$, ST and MT combined) confirmed a significant improvement in this executive function at post-test with medium effect size, $t(8) = -2.391$, $p = 0.044$, $d = 0.434$ (Cohen's *d*).

7.2. EEG analysis

Forward TRF weights averaged across all EEG channels are shown in Fig. 5 for the first and second EEG sessions (*pre* and *post* rehabilitation therapy, respectively). TRF calculated by using the original triggers did not lead to responses with any significant component in either session (Fig. 5A). Instead, TRFs after applying the re-alignment procedure used in Part I showed components that were significantly different from zero across participants in the *post-treatment* but not *pre-treatment* session ($p < 0.05$, Wilcoxon test across participants with FDR-correction over the time-latency dimension; Fig. 5B). In line with the simulation results, which indicated that the re-alignment procedure was effective for jitters up to 200 ms, the average re-alignment latencies in this use-case were found within 200 ms ($83 \text{ ms} \pm 16 \text{ ms}$ (SEM), for the *pre-treatment* TRFs and $90 \text{ ms} \pm 11 \text{ ms}$ (SEM), for the *post-treatment* TRFs). This provides us with a numerical indication that the trigger imprecision was below ~ 100 ms in this experiment. Please note that the emergence of an N1 component is not surprising, since the data were aligned based on the N1 component itself. Interestingly, significant P1 and P2 components were measured in the *post* session ($p < 0.05$, Wilcoxon test). The topographical distributions of the TRF weights suggests that the increase in neural tracking comes from all scalp locations, especially from centro-frontal areas (Fig. 5C-left). The same effect also emerged as a robust increase in the N1-P2 amplitude difference from the *pre* to the *post* session (Fig. 5C-right; FDR-corrected Wilcoxon, $p = 0.027$).

In the *pre* and *post* sessions respectively, five and seven participants showed EEG prediction correlations that were greater than chance ($p < 0.01$; the chance level was calculated by predicting the EEG on mismatched, circularly shifted folds from the same participant; $N = 100$ shuffles). In addition to confirming that the TRF models are reliable as they can predict neural data, this result reveals an increasing trend from *pre* to *post* sessions (seven participants), which is consistent with the behavioural test (i.e., verbal fluency scores).

8. Discussion

This study investigated the impact of temporal imprecisions in the context of neural tracking of continuous sensory information, involving perception tasks and stimuli such as speech and music. We did so by carrying out TRF analyses on three different datasets, while also devising a proof-of-concept re-alignment methodology to detect temporal imprecisions, recovering both EEG prediction correlations and the TRF response shape. While this issue was previously investigated in the context of ERP analyses (Hairston, 2012), it is important to clarify that typical ERP paradigms differ from TRF experiments. The first difference is the distinct analysis methodologies used for calculating ERPs (time-locked averaging) and neural tracking (e.g., multiple lagged regression). Second, ERPs typically involve discrete events (but see (Khalighinejad

et al., 2017)), while neural tracking is measured based on continuous sensory stimuli, which likely leads to different brain activations (Bonte et al., 2006). Third, the possibility of using natural stimuli, such as speech and music, allows for experiments in previously unexplored scenarios (e.g., realistic driving; (Liberto et al., 2021a) and participant cohorts (e.g., older individuals in care-homes). These reasons make the present investigation necessary.

TRF analyses offer several views on neural activations, such as latency of the TRF components, spatial localisation of the TRF components, shape of the TRF curves (similar to ERPs), as well as EEG prediction correlations and their topographical distribution. Temporal imprecision was shown to impact some of those views more than others. In the speech listening experiment (Part I), the TRF shape was shown to be particularly sensitive to synchronisation imprecision, with N1-P2 amplitudes collapsing with trigger jitters from 50 ms, while the EEG prediction correlations were more robust and were significant up to 200 ms. Similar effects were observed for the music listening experiment (Part II). Nevertheless, the novel temporal re-alignment procedure introduced here was proven effective in recovering both the TRF shape and prediction correlations, providing a valuable approach to assess neural tracking from a dataset that would be otherwise inoperable. It is important to clarify that the proposed re-alignment procedure should not be used as a standard pre-processing step for improving data quality. Instead, it offers an extra tool for determining if temporal imprecision could be degrading TRF results (as opposed to, for example, low SNR). Future studies could also investigate other potential uses for this methodology. For example, TRF methods (as for ERPs) assume the time-invariance of the target system (e.g., same N1 latency within and across trials) and the detection of optimal re-alignment shifts could serve as a way to quantify the (in)validity of that assumption. Another limitation of the present study that future work could address is also our assumption that a trigger misalignment will be constant within trials, which, despite being plausible, will likely not be the case for all scenarios.

The effect of manipulating the data at the pre-processing stage (e.g., filtering (De Cheveigné and Nelken, 2019) must be carefully considered before drawing any conclusions on the results. Here, it is important to highlight two main limitations of the present approach that may otherwise mislead the experimenter. First, while the procedure can recover the TRF shape, a somewhat reliable N1 TRF component is expected (by definition), as the re-alignment relies on that component. Second, the re-alignment prevents us from drawing any conclusion on the absolute TRF latency, as such values were forced to a time-value that was pre-defined and obtained from the literature. For these reasons, appropriate control analyses must be carried out to robustly assess neural tracking with the TRF analysis (Crosse et al., 2021). Here we presented two such controls. First, we shuffled the indices of the stimulus trials, producing a mismatch between stimulus and EEG. A second control consisted of simply time-reversing each stimulus trial, which produced a different type of stimulus-EEG mismatch. In principle, a meaningful TRF model (and EEG prediction) can be derived only when stimulus and EEG correspond. Indeed, this was shown to be true on the non-realigned analyses of Dataset 1, where EEG predictions were extremely small for both MM and REV. It is noted that the re-alignment is causing an unwanted side-effect in Dataset 1 for the MM condition, possibly due to the regular reading pace of the speaker in the audiobook used across trials. Interestingly, this issue was less prominent in Dataset 2, which involved a variable tempo across the musical trials. In general, while the reversal profoundly disrupts the relationship between the speech envelope and EEG, music could be problematic if the control condition has the same tempo (e.g., another piece with same tempo or a time-reversal). While the re-alignment window may not have been large enough for this issue to arise in Dataset 2, this risk should be considered when using rhythmic stimuli such as music. Regarding the MM condition, it should be noted that the re-alignment procedure itself determines an inflated N1 component. While this does not raise particular issues in

our study, since all the considered scenarios (e.g., the various jitters) would have been affected by the same N1 inflation, it is an important point to consider.

Within the boundaries of these limitations, re-aligning the TRFs to the N1 component proved to be a feasible way to recover effects where synchronisation imprecision obscured them. Indeed, one could choose to set the N1 latency to another value without changing the effectiveness of the approach, for example based on the relevant literature for a particular task. This approach enabled the objective assessment of speech and music therapy effectiveness in a small cohort of patients with mild cognitive impairment, based on measures taken with simple, portable, and affordable set-ups (8 sensors) in out-of-lab settings. Such objective assessments play a crucial role in the validation of intervention efficacy and could strengthen the evidence-based implementation of support strategies for vulnerable populations. Specifically, the present use-case study offers a novel validation of the natural speech TRF approach in a challenging scenario and population, which will facilitate further, larger-scale studies. It is crucial to note that the re-alignment procedure cannot guarantee the appropriate correction of the temporal imprecision. Instead, that analysis could determine that the re-alignment improves the TRF analysis, which confirms that the alignment was imprecise to the point of corrupting the TRF results (without re-alignment). While this was expected here as the dataset selected was a limit-case for temporal misalignment, we propose that the re-alignment procedure could be used on other datasets to detect potential issues with the temporal alignment between stimuli or actions and the neural data. A re-alignment that improves the TRF model would reflect a poor temporal synchronisation, while the same procedure would instead either not affect or even damage a precisely aligned dataset (as showed for Datasets 1 and 2).

In conclusion, we have shown the effect that synchronisation imprecision has on different representations of the forward TRF analysis, particularly, on EEG prediction correlations, topographical TRF representations, and shape of the TRFs. It is noteworthy that forward TRF analysis seems to be able to cope with small jitters of up to 50 ms without complete compromise of EEG prediction correlations and 25 ms without complete compromise of N1-P2 amplitudes. When we state that the forward TRF is not excessively degraded when considering up to 50 ms jitters, we mean that it is only for jitters larger than 50 ms that the TRF prediction correlations exhibit drop of over 50% compared to the original TRF, despite maintaining its statistical significance for jitters up to 150 ms. However, the presence of synchronisation imprecisions challenges the interpretability of the TRFs. The proposed TRF re-alignment restored EEG correlation predictions, N1-P2 amplitudes, and relative component latencies across a longer range of synchronisation imprecision, but it cannot recover the interpretability of absolute TRF latency values. Instead, the improved TRF results after re-alignment can be taken as evidence for a substantial issue with the temporal alignment in the dataset. While this should not happen, and it is important that each experimenter ensures the precise temporal synchronisation of their device (triggers) and experiment (task), there are scenarios where technical issues (e.g., mistakes in the presentation script; substantial loss of packages with a wireless connection) or inherent properties of a task (e.g., auditory imagery) may challenge the precise temporal synchronisation. In those cases, the proposed procedure can be run to detect whether such imprecisions are a problem for the TRF analysis or not.

Ethics approval information

- Datasets 1 and 2: All procedures were undertaken in accordance with the Declaration of Helsinki and were approved by the Ethics Committees of the School of Psychology at Trinity College Dublin, and the Health Sciences Faculty at Trinity College Dublin.
- Case study: Ethical approval was granted by Middlesex University Psychology Research Ethics Committee (application number: ST020-

2018) and the study was conducted following MHA's directors and psychological expert approval at the care-homes.

Funding

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at Trinity College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme. This work was supported by the Science Foundation Ireland Career Development Award 15/CDA/3316. A.M. was supported by SEMPRES (n. 102399), MHA (n. 102334), and a Dunhill Medical Trust RPF2006\241 grants. S.C. was supported by a research grant from the William Demant Foundation (21-0628).

CRediT authorship contribution statement

Sara Carta: Formal analysis, Methodology, Writing – original draft. **Anthony M. A. Mangiacotti:** Conceptualization, Data curation, Formal analysis, Writing – original draft. **Alejandro Lopez Valdes:** Conceptualization, Methodology, Writing – original draft. **Richard B. Reilly:** Funding acquisition, Writing – review & editing. **Fabia Franco:** Conceptualization, Data curation, Funding acquisition, Writing – review & editing. **Giovanni M. Di Liberto:** Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft.

Declarations of interest

none.

Data Availability

Datasets 1 and 2 are publicly available. The use-case study involves patients data that cannot be publicly shared.

Acknowledgments

We would like to thank music therapists Dr Ming Hung Hsu and MT Clare Barone, and all the staff of MHA care homes that have helped with participants recruitment and management in the use-case study, together with SEMPRES and MHA for their financial contributions. We would also like to thank all the participants that took part in this research and the authors of the publicly available datasets. We thank the CNSP initiative (Cognition and Natural Sensory Processing; <https://cnsppworkshop.net>), for sharing the standardised datasets and analysis code used for this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.jneumeth.2022.109765](https://doi.org/10.1016/j.jneumeth.2022.109765).

References

- Alickovic, E., Lunner, T., Wendt, D., Fiedler, L., Hietkamp, R., Ng, E.H.N., Graversen, C., 2020. Neural Representation Enhanced for Speech and Reduced for Background Noise With a Hearing Aid Noise Reduction Scheme During a Selective Attention Task. *Front. Neurosci.* 14.
- Attaheri, A., Choidealbha Á, N., Di Liberto, G.M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., Goswami, U., 2022. Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *Neuroimage* 247, 118698.
- Badcock, N.A., Mousikou, P., Mahajan, Y., Lissa, D.E., Thie, J. P., McArthur, G., 2013. Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. *PeerJ* 2013.

- Badcock, N.A., Preece, K.A., Dewit, B., Glenn, K., Fieder, N., Thie, J., 2015. & GENEVIEVE, M. 2015. Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. *PeerJ* e907-e907.
- Biasutti, M., 2013. Improvisation in dance education: teacher views. *Res. Dance Educ.* 14, 120–140.
- Bilucaglia, M., Masi, R., Stanislao, G.D., Laureanti, R., Fici, A., Circi, R., Zito, M., Russo, V., 2020. ESB: a low-cost EEG synchronization box. *HardwareX* 8, e00125.
- Bonte, M., Parviainen, T., Hytönen, K., Salmelin, R., 2006. Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cereb. Cortex* 16, 115–123.
- Brodbeck, C., Presacco, A., Anderson, S., Simon, J.Z. Over-representation of speech in older adults originates from early response in higher order auditory cortex. *Acta Acustica united with Acustica*, 2018a. S. Hirzel Verlag GmbH, 774–777.
- Brodbeck, C., Hong, L.E., Simon, J.Z., 2018b. Rapid transformation from auditory to linguistic representations of continuous speech. *Curr. Biol.* 28 (3976–3983), e5.
- Broderick, M., Anderson, A., D.I. Liberto, G., Crosse, M, Lalor, E., 2018a. Data from: electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Dryad Digital Repository. Published online February 23, 2018.
- Broderick, M.P., Anderson, A.J., Liberto, D.I., Crosse, M. J, G.M., Lalor, E.C., 2018b. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr. Biol.*
- Broderick, M.P., Liberto, D.I., Anderson, G.M., Rofes, A, A.J., Lalor, E.C., 2021. Dissociable electrophysiological measures of natural language processing reveal differences in speech comprehension strategy in healthy ageing. *Sci. Rep.* 11, 4963.
- Brunnhuber, F., Amin, D., Nguyen, Y., Goyal, S., Richardson, M.P., 2014. Development, evaluation and implementation of video-EEG telemetry at home. *Seizure* 23, 338–343.
- Crosse, M.J., Liberto, D.I., Bednar, A, G.M., Lalor, E.C., 2016a. The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.* 10.
- Crosse, M.J., Di Liberto, G.M., Lalor, E.C., 2016b. Eye can hear clearly now: Inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *J. Neurosci.* 36.
- Crosse, M.J., Zuk, N.J., Liberto, D.I., Nidiffer, G.M., Molholm, S, A.R., Lalor, E.C., 2021. Linear modeling of neurophysiological responses to speech and other continuous stimuli: methodological considerations for applied research. *Front. Neurosci.* 15, 705621-705621.
- De Cheveigné, A., Nelken, I., 2019. Filters: when, why, and how (Not) to use them. *Neuron* 102, 280–293.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.
- Di Liberto, G.M., Lalor, E.C., 2017. Indexing cortical entrainment to natural speech at the phonemic level: methodological considerations for applied research. *Hear. Res.* 348, 70–77.
- Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. USA* 109, 11854–11859.
- Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8.
- Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage* 88, 41–46.
- Folstein, M.F., Robins, L.N., Helzer, J.E., 1983. The mini-mental state examination. *Arch. Gen. Psychiatry Arch. Gen. Psychiatry*.
- Hairston, W.D., 2012. Accounting for timing drift and variability in contemporary electroencephalography (EEG) systems. *Army Res. Lab Aberdeen Proving Ground*.
- Hairston, W.D., Whitaker, K.W., Ries, A.J., Vettel, J.M., Bradford, J.C., Kerick, S.E., McDowell, K., 2014. Usability of four commercially-oriented EEG systems. *J. Neural Eng.* 11, 046018.
- Jessen, S., Fiedler, L., Münte, T.F., Obleser, J., 2019. Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. *NeuroImage* 202, 116060-116060.
- Jessica Tan, S.H., Kalashnikova, M., Liberto, D.I., Crosse, M. J, G.M., Burnham, D., 2022. Seeing a talking face matters: The relationship between cortical tracking of continuous auditory-visual speech and gaze behaviour in infants, children and adults. *NeuroImage* 256, 119217.
- Kalashnikova, M., Peter, V., Liberto, D.I., Lalor, E. C, G.M., Burnham, D., 2018. Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Sci. Rep.* 8.
- Khalighinejad, B., Cruzatto D.A.S.I.L.V.A., G., Mesgarani, N., 2017. Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech. *The Journal of Neuroscience*.
- Kothe, C., Medine, D., Boulay, C., Grivich, M., Stenner, T. 2014. Lab streaming layer. URL <https://github.com/scn/labstreaminglayer>.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647.
- Lalor, E.C., Foxe, J.J., 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur. J. Neurosci.* 31, 189–193.
- Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., Mcdarby, G., Foxe, J.J., 2006. The VESPA: a method for the rapid estimation of a visual evoked potential. *NeuroImage* 32, 1549–1561.
- Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J., 2009. Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiol.* 102, 349–359.
- Legendre, G., Andriillon, T., Koroma, M., Kouider, S., 2019. Sleepers track informative speech in a multitalker environment. *Nat. Hum. Behav.*
- Liberto, D.I., O'sullivan, J. A, G.M., Lalor, E.C., 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.* 25.
- Liberto, D.I., Peter, G.M., Kalashnikova, V., Goswami, M., Burnham, D, U., Lalor, E.C., 2018. Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *NeuroImage*, NIMG 17–29, 70–79.
- Liberto, D.I., Wong, G.M., Melnik, G, A, D., De Cheveigné, A., 2019. Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. *NeuroImage* 196, 237–247.
- D.I. Liberto, G.M., Pelofi, C., Bianco, R., Patel, P., Menhta, A.D., Herrero, J.L., Shamma, S.A., Mesgarani, N., Mehta, A.D., Herrero, J.L., D.E. Cheveigné, A., Shamma, S. A. & Mesgarani, N. 2020. Cortical encoding of melodic expectations in human temporal cortex. *eLife*, 9.
- Liberto, D.I., Barsotti, G.M., Vecchiato, M., Ambeck-Madsen, J, G., Del Vecchio, M., Avanzini, P., Ascari, L., 2021a. Robust anticipation of continuous steering actions from electroencephalographic data during simulated driving. *Sci. Rep.* 11, 23383.
- Liberto, D.I., Marion, G, G.M., Shamma, S.A., 2021b. The music of silence: Part II: music listening induces imagery responses. *J. Neurosci.* 41, 7449.
- Liberto, D.I., Nie, G.M., Yeaton, J., Khalighinejad, J., Shamma, S. A, B., Mesgarani, N., 2021c. Neural representation of linguistic feature hierarchy reflects second-language proficiency. *NeuroImage* 227, 117586-117586.
- Liberto, D.I., Pelofi, G.M., Bianco, C., Patel, R., Mehta, P., Herrero, A.D., De Cheveigné, J. L., Shamma, S. A, A., Mesgarani, N., 2021d. Cortical encoding of melodic expectations in human temporal cortex. *Dryad*.
- Linn, B.S., Linn, M.W., Gurel, L.E.E., 1968. Cumulative illness rating scale. *J. Am. Geriatr. Soc.* 16, 622–626.
- Luck, S.J. 2005. An introduction to the event-related potential technique.
- Machado, T.H., Fichman, H.C., Santos, E.L., Carvalho, V.A., Fialho, P.P., Koenig, A.M., Fernandes, C.S., Lourenço, R.A., Paradela, E.M.D.P., Caramelli, P., 2009. Dados normativos de idosos saudáveis na prova de fluência verbal fonêmica - FAS. *Dement. e Neuropsychol.* 3, 55–60.
- Mangiaccotti, A., Hsu, M.H., Barone, C., Di Liberto, G., V.A.N. Puyvelde, M., Biasutti, M. & Franco, F. 2019. P3–465: Effects of music therapy in elderly care: cognition, biomarkers, mood and social behaviour. *Alzheimer's & Dementia*, 15, P1144-P1144.
- Mangiaccotti, A.M.A. 2020. The Music Cognitive Test: Validation, training and applicability of a music-based cognitive tool for music therapists - a multidisciplinary study. University of Padua.
- Martin, S., Brunner, P., Holdgraf, C., Heinze, H.-J., Crone, N.E., Rieger, J., Schalk, G., Knight, R.T., Pasley, B.N., 2014. Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front. Neuroeng.* 7.
- Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010.
- Mesik, J., Ray, L., Wojtczak, M., 2021. Effects of age on cortical tracking of word-level features of continuous competing speech. *Front. Neurosci.* 15.
- Meyer, L., Lakatos, P., He, Y., 2021. Language dysfunction in Schizophrenia: assessing neural tracking to characterize the underlying disorder(s)? *Front. Neurosci.* 15.
- Nucci, M., Mapelli, D., Mondini, S., 2012. Cognitive Reserve Index questionnaire (CRIq): A new instrument for measuring cognitive reserve. *Aging Clin. Exp. Res.* 24, 218–226.
- Obleser, J., Kayser, C., 2019. Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*. Elsevier Ltd.
- Opaso, P.R., Barreto, S.D.S., Ortiz, K.Z., 2016. Phonemic verbal fluency task in adults with high-level literacy. *Einstein* 14, 398–402.
- O'sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A., Lalor, E.C. 2014. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, bht355-bht355.
- Ries, A.J., Touryan, J., Vettel, J., McDowell, K., Hairston, W.D., 2014. A comparison of electroencephalography signals acquired from conventional and mobile systems. *J. Neurosci. Neuroeng.* 3, 10–20.
- Santoro, R., Moerel, M., Martino, D.E., Valente, F., Ugurbil, G., Yacoub, E, K., Formisano, E., 2017. Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. *Proc. Natl. Acad. Sci. USA* 114, 4799–4804.
- Sintotskiy, G., Hinrichs, H., 2020. In-ear-EEG—a portable platform for home monitoring. *J. Med. Eng. Technol.* 44, 26–37.
- Teoh, E.S., Cappelloni, M.S., Lalor, E.C., 2019. Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *Eur. J. Neurosci.* 50, 3831–3842.
- Tucker, A.M., Stern, Y., 2011. Cognitive reserve in aging. *Curr. Alzheimer Res.* 8, 354–360.
- Valente, G., Castellanos, A.L., Vanacore, G., Formisano, E., 2014. Multivariate linear regression of high-dimensional fMRI data with multiple target variables. *Hum. Brain Mapp.* 35, 2163–2177.
- Vanthornhout, J., Decruy, L., Francart, T., 2019. Effect of task and attention on neural tracking of speech. *Front. Neurosci.* 13, 977–977.
- Woodman, G.F., 2010. A brief introduction to the use of event-related potentials in studies of perception and attention. *Atten., Percept. Psychophys.* 72, 2031–2046.