Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Cortical tracking of prosody after stroke and in aging: preliminary evidence from magnetoencephalography

Giada Antonicelli[1,2], *Nicola Molinaro*[1,3], *Patricia De La Riva*[4], *Raquel Laspiur*[5], *Arantza Lopez De Turiso*[4], *Maddi Carrera*[1], *Simona Mancini*[1,3]

[1]Basque Center on Cognition Brain and Language (BCBL), Donostia-San Sebastián, Spain
[2] University of the Basque Country (UPV/EHU), Bilbao, Spain
[3]Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[4]Hospital Universitario de Donostia-San Sebastian, Donostia-San Sebastián, Spain
[5]Biogipuzkoa Health Research Institute, Donostia-San Sebastián, Spain
`g.antonicelli@bcbl.eu, s.mancini@bcbl.eu`

Abstract

Evidence exists that brain-damaged but also healthy aging people often experience difficulties in linguistic (LP) and emotional prosody (EP) and exhibit poorer neural synchronization to the speech input (cortical tracking of speech, CTS) in the delta frequency band. Using a cross-sectional design with left-hemisphere (LH) and right-hemisphere (RH) stroke survivors, young (18-30 y.o) and old (35-80 y.o.) control participants (YC, OC, respectively), we ask whether: 1) CTS is anomalous after stroke and in healthy aging, 2) correlates with prosody interpretation, and 3) LP and EP processing are segregated in the brain. Participants listen to Spanish sentences in EP, LP, and neutral prosody conditions. We measure CTS using the multivariate Temporal Response Function (mTRF) approach. Preliminary data from 8 YC, 4 people with LH damage (LHD), and 3 OC show that, relative to controls, LHD have lower task accuracy, and their CTS is lower in temporal areas in the delta/theta bands but higher in the delta band over parietal sensors. So far, we confirmed that delta/theta CTS is anomalous and prosody comprehension is harder after stroke, while no effects of prosody type were observed. A tentative interpretation can be that since LHD experience a deficit in early input-driven processes, they need to recruit more top-down cognitive resources, which does not lead to control-like task performance.

Index Terms: stroke, aging, emotional prosody, linguistic prosody, prosody impairment, magnetoencephalography, multivariate temporal response function

1. Introduction

1.1. Prosody processing: accounts and evidence

A large part of our everyday communication takes place through spoken language. Research has devoted increasing attention to prosody, acknowledging its crucial role in conveying meaning and emotions. Prosody is typically divided into two broad subcategories: affective or emotional prosody (EP), which contributes to express the mental state of the speaker, and linguistic prosody (LP), which indicates syntactic boundaries, word-category distinctions, information structure and illocutive force [1].

Prosody processing relies on the same structures as speech processing, namely the superior temporal sulcus and gyrus (STS/STG) and the inferior frontal gyrus (IFG) [2], [3], [4], [5], [6], [7], [8], [9]. Nonetheless, the distribution of such brain function is still debated.

A traditional idea is that EP processing is right-lateralized, while LP processing is left-lateralized (function-dependent lateralization hypotheses, [10], [7], [11], [12], [13]. However, meta-analyses of stroke survivors' data found no hemispheric asymmetry for LP/EP [15], [16]. Other accounts focus on the physical properties of the input (cue-dependent lateralization hypotheses, [7], [11]). For example, the acoustic-lateralization hypothesis maintains that slow acoustic changes, such as pitch fluctuations, are processed in the right hemisphere, while faster ones, associated with syllable-level modulations, are more left-lateralized [10], [14].

Evidence from brain-damaged speakers shows that lesions to the bilateral frontotemporal network can hinder specific stages of prosody processing [4]. These findings informed multi-step models, which, although originally developed for EP comprehension and discrimination, may prove useful to describe speech, LP, and EP processing in an integrated way. According to such accounts, at the first stage, acoustic features are perceived, extracted and roughly analyzed in the superior temporal sulcus and gyrus (STS/STG). Once 'percepts' are formed, they are fed to frontal areas -such as the inferior frontal gyrus (IFG) for higher-order top-down processing [3], [4], [7], [8], [9].

We argue that such heterogeneous findings might be due to two main reasons. First, the relative weight of acoustic and categorical (type of prosody) aspects of prosody have not been assessed *together*. Second, to our knowledge, there are no brain imaging studies that compare EP and LP directly. Therefore, we contrasted EP and LP prosody in post-stroke survivors with lesions in either RH or LH, using magnetoencephalography (MEG).

10.21437/SpeechProsody.2024-5

1.2. A novel approach to the study of speech processing: cortical tracking of speech (CTS)

Most of the evidence on prosodic processing presented to date has focused on the temporal course and structural underpinnings of prosody function. However, recent research has developed a valuable approach to investigate the process of speech comprehension in both the frequency and time domains, namely cortical tracking of speech (CTS). CTS refers to the synchronization between the speech signal and the oscillatory activity of neurons, which is most robust in the delta (0-4Hz) and theta (4-8Hz) frequency bands across frontotemporal areas. In a typical CTS paradigm, participants listen to a long stream of speech while their brain signal is recorded via electroencephalography (EEG) or magnetoencephalography (MEG). Several studies established a link between comprehension (understanding of the linguistic message)/intelligibility (noisiness of speech acoustics) and CTS [17], [18], [19]. However, some reported a dissociation between the delta and theta bands, which might be due to functional reasons. Delta oscillations have been attributed a role in top-down mechanisms [20], intonation processing [21] and sentence parsing [22], while theta would be involved in syllable-level decoding/encoding [23], and executive mechanisms, such as attention and working memory [20].

CTS measures have also uncovered relevant aspects of speech processing in people with aphasia. In the studies to date, both people with logopenic primary progressive aphasia and post-stroke aphasia have exhibited reduced CTS in the delta band at the acoustic (envelope), (sub)-lexical segmentation (phoneme and word onsets), and word (word surprisal and frequency) level [24], [25]. Interestingly, increased CTS in the theta band has been observed [26] that correlated with aphasia severity but also with better behavioral performance [25], which has led to interpret this activity as reflective of compensatory mechanisms. Similar effects have also been reported in older as compared to young adults [27].

In sum, the literature suggests a link between CTS in delta/theta bands and successful speech comprehension. This makes CTS a good candidate to sensitively measure prosody processing and characterize its unfolding over time across the cortical surface.

1.3. Present study

The literature reviewed above provides mixed findings on LP/EP processing lateralization. Moreover, virtually no study has assessed the effect of stroke or aging on cortical tracking of intonation. Mindful of such gaps, we designed an experiment that asks the following research questions (RQs) and tests their relative predictions (Ps):
RQ1) Is CTS in the delta/theta bands different after stroke than in healthy aging, and in older than young adults?
P1) CTS in the delta band after a brain lesion and in aging is reduced in the delta band and anomalous in theta.
RQ2) Does CTS correlate with accuracy in prosody interpretation?
P2) CTS anomalies do not necessarily lead to worse prosody interpretation due to compensatory mechanisms.
RQ3) Are LP and EP processing segregated in the brain?
P3) LP and EP are not segregated but at least partially dissociable, be it in brain activity or behavior.

To address these questions, we recruited stroke survivors with LH or RH lesions as well as healthy participants from different age groups (RQ1). During MEG recording, we exposed them to sentences with 6 possible intonations belonging to baseline neutral, linguistic and emotional categories and tracked prosody interpretation abilities through an intonation-facial expression matching task (RQ2). We measured CTS in the delta/theta bands and will perform source reconstruction to detect possible dissociations between LP and EP (RQ3).

This study presents prospective clinical relevance. CTS could represent a fine-grained measure to detect speech and prosody comprehension impairments. Better screening tools can reflect into treatments that are more tailored to individual patients' needs. For example, if CTS is lower in temporal areas, the therapy could target acoustic perception via feature enhancement.

2. Methods

2.1. Participants

We aim to recruit two clinical groups, one with damage in the left hemisphere (LHD, N=30) and one with damage in the right hemisphere (RHD, N=30). Two additional control groups will be also recruited: a young adults' group (YC, N=30), and a group composed of older adults whose participants will be matched with stroke survivors in age range, education range and sex (OC, N=30). Young and older adults' groups will not be matched in education but we will control for this variable statistically (see also "Statistical analysis" for details). Here we report the data from 8 YC (Mean_{age}=24.5, Standard $\text{Deviation}_{\text{age}}$=6.1, henceforth M and SD), 4 LH stroke survivors (M_{age}=56, SD_{age}=18.4), and 3 OC (M_{age}=50, SD_{age}=18.3). All subjects had normal or corrected-to-normal hearing and vision. Control participants had no history of neurological disorders. Stroke survivors with first ever haemorrhagic or ischemic stroke, and with no history of neuropsychiatric disorder or neurodegenerative condition were included. Further, we excluded from OC and YC anyone who had advanced training in music or acting because they might have an advantage in intonation perception/interpretation. Since we are conducting this experiment in the Basque Country, most of our participants are Spanish-Basque bilinguals with Spanish as their dominant language.

2.2. Stimuli and rationale

We prepared a set of 185 sentences and recorded them from a Spanish-Basque bilingual actress in 5 intonations: angry and happy (EP condition), question and order (LP condition), and neutral declarative (baseline condition, B). To norm our stimuli, we then ran an online experiment with 130 Spanish speakers who rated each sentence by dragging a slider towards the word that best matched the speaker's intention or state of mind. For the MEG experiment we selected the 48 items that were most consistently (>70%) classified as expected across participants. We added flat contour to the B condition by levelling the neutral intonation to its average fundamental frequency (pitch). By comparing the neutral and flat contours we are thus able to isolate the effect of intonation. The final design consisted of 3 conditions, comprising 2 intonations each. For the task of intonation-facial expression matching, we created a set of six pictures, 5 of human faces (matching with natural intonations) and one of a robot (matching with the flat intonation). Before the experiment, participants were familiarized with each picture and were told which intention or emotion each of them expressed, in order to rule out that low accuracy was due to the misinterpretation of facial expressions.

2.3. Procedure

The experiment is structured in multiple sessions whose order can change to adjust to participants' needs. They include the MEG experiment, a magnetic resonance scan, and thorough linguistic and cognitive assessments. Here we only report the procedure and results of the MEG session.

2.3.1. MEG session

For MEG recording we employed a 306-sensor (204 planar gradiometers and 102 magnetometers; arranged in a helmet configuration) Elekta Neuromag® device with 16 digital trigger lines and 8 auxiliary analog input channels. Subjects were in a sitting position. MEG recording took place while subjects binaurally listened to a total of 288 sentences of 2-4 seconds duration each. After each sentence they were presented with three pictures, one was the correct face, one was from the same condition as the correct face, and one from a different condition (e.g., question, order, and happy). They had to choose which face best fit the speaker's intention/emotion via button press. Stimuli presentation was performed using the Psychophysics toolbox-3 for MATLAB. All subjects saw the same list of stimuli in a random order.

2.4. Preprocessing and statistical analyses

2.4.1. MEG signal preprocessing

After manual bad channels marking, the MEG signal was filtered from artifacts as well as internal and external noise sources using the MaxFilter™ software. Any subsequent pre-processing and analysis steps were performed in Python using the MNE [30] and Eelbrain [31] packages. We applied ICA to the signal band-pass filtered between 0.5-120Hz. Components that reached maximal correlation with EOG and ECG channels were automatically removed from the signal.

2.4.2. CTS and statistics

We measured cortical tracking using the multiple temporal response function approach (mTRF, [32]). For each subject and condition, we fit several forward models with the MEG signal (filtered at 0.5-4Hz, for the delta band, and at 4-8Hz, for the theta band) as the dependent variable and acoustic (envelope, pitch contour, spectrogram, envelope edges and spectrogram edges) and linguistic factors (word onset and phoneme onset) as predictors. Both the MEG signal and the predictors were down-sampled to 100Hz for computational reasons. Envelopes and spectrograms were computed after applying a gammatone filter to the original audio. This transformation resembles the one that takes place in the cochlea. Word and phoneme onset times were extracted via forced alignment using the online tool WebMAUSBasic [33] (https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic). The prediction accuracy of each model is the Pearson's correlation between the predicted and the observed neural signal. We used mass permutation-based ANOVAs to compare different mTRF models and assess the effect of experimental manipulations. Task accuracy was analysed with generalized linear mixed-effect models (glmer function from lm4 R package). In a final step, linear mixed effects models will be employed to integrate random factors and behavioral variables. Unfortunately, we still do not have enough data to perform such an analysis. The data we present here refers to magnetometer sensors.

3. Preliminary results

LHD had overall lower task accuracy (M=64.32%, SD=20%) than YC (M=90.23%, SD=11%) and OC (M=90.16%, SD=5%). The generalized linear mixed-effect model accuracy~condition*group+(1|subject) found that both the main effects and the interaction were significant. Pairwise post-hoc t-tests showed that responses in B were more accurate than in LP (except for OC, where no difference was found) and EP, and in LP than EP (except for LHD, where no difference was found), and in the neurotypical groups than in LHD. The spectrogram and the spectrogram of acoustic edges proved to be the best predictors when compared to the maximal model (the one including all the aforementioned predictors). In the delta band, CTS~condition*group ANOVAs found a main effect of group (F_{max}=14.29, p=0.003). According to pairwise t-tests (Fig.1), CTS was larger in YC than OC in right temporal sensors (RTS), and in OC than in LHD in left temporal ones (LTS). In the theta band the effect of group was marginal (F_{max}=9.02, p=0.054) but at visual inspection (Fig.2), CTS was larger in OC than in YC and LHD over LTS, and in YC than in OC over RTS. This lack of statistical significance might be due to the small sample size of OC and LHD.

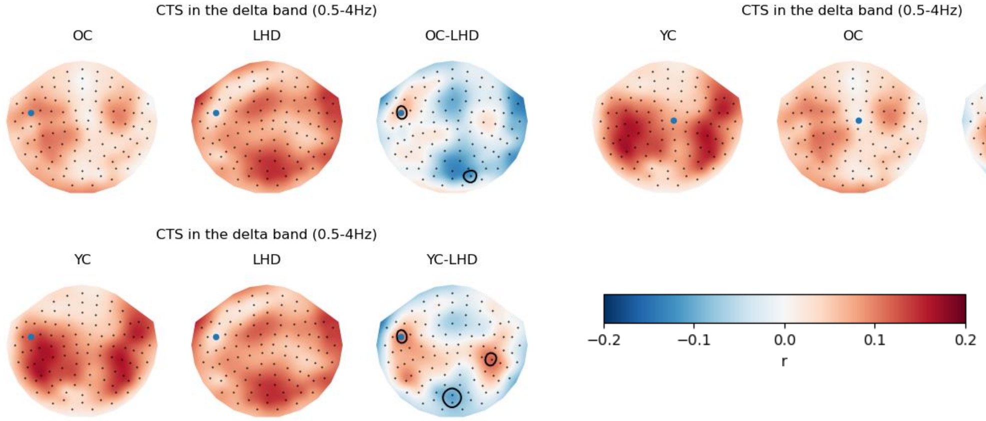

Figure 1: Topographical maps of between-group comparisons of CTS in the delta band. The maximal model included spectrogram, edges onsets spectrogram, word onset, phoneme onset, envelope, edges onsets envelope, pitch contour as predictors. The final model, here depicted, only featured spectrogram and edges onsets spectrogram. Blue dots highlight the first significant cluster.

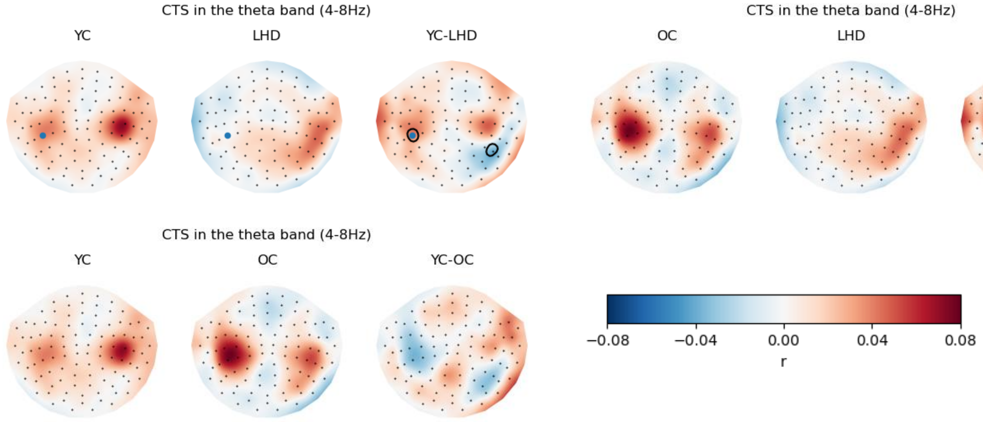

Figure 2: Topographical maps of between-group comparisons of CTS in the theta band. The final model, here depicted, only featured edges onsets spectrogram (further information in Figure 1 caption).

4. Discussion

In this study, we aimed to assess the effect of brain lesions and aging on CTS (RQ1) and prosody comprehension (RQ2) and test locational accounts of EP and LP processing (RQ3). Relative to RQ1 and RQ2, in LHD we observed anomalous CTS and less accurate behavior relative to controls. Like in previous literature [24], [25], LHD showed reduced CTS in the delta band over temporal sensors. Moreover, their CTS was stronger over parietal sensors in delta frequencies and reduced in temporal areas in the theta range. Regarding RQ3, we did not find any effect of condition on CTS, but we did on behaviour, with higher accuracy in the baseline than LP and EP conditions. Discrepancies in topographical maps suggest that prosody tracking might be not only quantitatively but also qualitatively different across groups. From the perspective of multi-stage models [4], [11], [8], [9], it could be that controls rely on more efficient early-stage, bottom-up processing that quickly analyses perceptual units and classifies them into task-oriented categories (here, intonation contours), making CTS less necessary in final bottom-up evaluation. The LHD group, on the other hand, might suffer some impairment in perceptual organization. Higher parietal CTS could index the attempt to compensate for that by recruiting additional cognitive resources. Crucially, significant CTS reduction in LHD is observed in the lesioned hemisphere and the only LHD that responded more than 65% correct was also the one showing a CTS in LTS similar to controls.

5. Limitations and conclusions

We underline once more that our results are only preliminary. The sample size is still small, and more fine-grained statistical analyses are necessary. We plan to recruit more participants, add linguistic features to the mTRF model, perform source reconstruction and integrate experimental and screening data using linear mixed-effect models. Results so far confirmed that CTS is anomalous in stroke survivors and this correlates with behavior, while no difference between LP and EP was observed.

6. Acknowledgements

This research is supported by the Basque Government through the BERC 2022-2025 program and Funded by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010/AEI/10.13039/501100011033 and through project CEX2020-001010-S-21-5 funded by the Basque Government. S.M acknowledges funding from the Spanish Ministry of Science and Innovation (PID2020-113945RB-I00)

7. References

[1] Bögels, S. (2011). *The role of prosody in language comprehension: when prosodic breaks and pitch accents come into play*. Doctoral thesis. Radboud University Repository.

[2] Bhaya-Grossman, I., & Chang, E. F. (2022). Speech computations of the human superior temporal gyrus. *Annual review of psychology*, *73*, 79-102.

[3] Sheppard, S. M., Stockbridge, M. D., Keator, L. M., Murray, L. L., Blake, M. L., Right Hemisphere Damage working group, & Evidence-Based Clinical Research Committee. (2021). The Company Prosodic Deficits Keep Following Right Hemisphere Stroke: A Systematic Review. *Journal of the International Neuropsychological Society*, 1-16.

[4] Sheppard, S. M., Meier, E. L., Durfee, A. Z., Walker, A., Shea, J., & Hillis, A. E. (2021). Characterizing subtypes and neural correlates of receptive aprosodia in acute right hemisphere stroke. *Cortex*, *141*, 36-54.

[5] Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, *8*(5), 393-402.

[6] Hickok, G., & Poeppel, D. (2016). Neural basis of speech perception. In Hickok, G., & Small, S. (Eds.). (2016). *Neurobiology of language*. (pp. 299-310). Academic Press.

[7] Witteman, J., Van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, *50*(12), 2752-2763.

[8] Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, *5*(3), 108-125.

[9] Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, *11*(4), 277-281.

[10] Paulmann, S. (2016). The neurocognition of prosody. In Hickok, G., & Small, S. (Eds.). (2016). *Neurobiology of language*. (pp. 1109-1120). Academic Press.

[11] Witteman, J., van IJzendoorn, M. H., van de Velde, D., van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: a meta-analysis of the lesion literature. *Neuropsychologia*, *49*(13), 3722-3738.

[12] Diehl, J. J., & Paul, R. (2009). The assessment and treatment of prosodic disorders and neurological theories of prosody. *International journal of speech-language pathology*, *11*(4), 287-292.

[13] Peppé, S. J. (2009). Why is prosody in speech-language pathology so difficult?. *International journal of speech-language pathology*, *11*(4), 258-271.

[14] Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816-847.

[15] Ukaegbe, O. C., Holt, B. E., Keator, L. M., Brownell, H., Blake, M. L., Lundgren, K., ... & Evidence-Based Clinical Research Committee. (2022). Aprosodia Following Focal Brain Damage: What's Right and What's Left?. *American Journal of Speech-Language Pathology*, 1-16.

[16] Stockbridge, M. D., Sheppard, S. M., Keator, L. M., Murray, L. L., Blake, M. L., Right Hemisphere Disorders working group, & Evidence-Based Clinical Research Committee. (2021). Aprosodia subsequent to right hemisphere brain damage: a systematic review and meta-analysis. *Journal of the International Neuropsychological Society*, 1-27.

[17] Lizarazu, M., Lallier, M., & Molinaro, N. (2019). Phase−amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the New York Academy of Sciences*, *1453*(1), 140-152.

[18] Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS biology*, *16*(3), e2004473.

[19] Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in psychology*, *2*, 130.

[20] Keitel, A., Ince, R. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage*, *147*, 32-42.

[21] 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. *European Journal of Neuroscience*, *50*(11), 3831-3842.

[22] Glushko, A., Poeppel, D., & Steinhauer, K. (2022). Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Scientific reports*, *12*(1), 1-18.

[23] Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience*, *19*(1), 158-164.

[24] Kries, J., De Clercq, P., Gillis, M., Vanthornhout, J., Lemmens, R., Francart, T., & Vandermosten, M. (2023). Exploring neural tracking of acoustic and linguistic speech representations in individuals with post-stroke aphasia. *bioRxiv*, 2023-03.

[25] Quique, Y. M., Gnanateja, G. N., Dickey, M. W., Evans, W. S., & Chandrasekaran, B. (2023). Examining cortical tracking of the speech envelope in post-stroke aphasia. *Frontiers in Human Neuroscience*, *17*.

[26] Dial, H. R., Gnanateja, G. N., Tessmer, R. S., Gorno-Tempini, M. L., Chandrasekaran, B., & Henry, M. L. (2021).

Cortical tracking of the speech envelope in logopenic variant primary progressive aphasia. *Frontiers in human neuroscience*, *14*, 597694.

[27] Gillis, M., Kries, J., Vandermosten, M., & Francart, T. (2023). Neural tracking of linguistic and acoustic speech representations decreases with advancing age. *Neuroimage*, *267*, 119841.

[28] Ansorena, X., Carreiras, M., Hernandez, M., Baena, I. & Mancini, S. (2022). Aphasia Cognitive Screening in Spanish (ACS.esp): a new digital test to assess language in aphasia. In *Proceedings of Science of Aphasia*; September 12th-15th 2022; Bordeaux, France. 65.

[29] Cuetos Vega, F., & González-Nosti, M. (2009). BETA: Batería para la evaluación de los trastornos afásicos. Madrid: Instituto de Orientación Psicológica EOS.

[30] Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience*, *7*, 267. https://doi.org/10.3389/fnins.2013.00267

[31] Brodbeck, C., Das, P., Gillis, M., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., & Simon, J. Z. (2023). Eelbrain, a Python toolkit for time-continuous analysis with temporal response functions. *eLife*, *12*, e85012. Advance online publication. https://doi.org/10.7554/eLife.85012

[32] Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in human neuroscience*, *10*, 604.

[33] Schiel, F. (2015). A statistical model for predicting pronunciation. *International Congress of Phonetic Sciences*.