



**HAL**  
open science

# Behavioral and Neurophysiological Representations of Speech Phonemic Units

Adrielle De Carvalho Santana

► **To cite this version:**

Adrielle De Carvalho Santana. Behavioral and Neurophysiological Representations of Speech Phonemic Units. Cognitive Sciences. Université Grenoble Alpes [2020-..]; Universidade federal de Minas Gerais, 2020. English. NNT : 2020GRALS036 . tel-03247462

**HAL Id: tel-03247462**

**<https://theses.hal.science/tel-03247462>**

Submitted on 3 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# BEHAVIORAL AND NEUROPHYSIOLOGICAL REPRESENTATIONS OF SPEECH PHONEMIC UNITS

Doctoral thesis presented and defended on December 16, 2020, for a Doctoral degree  
at Programa de Pós-Graduação em Engenharia Elétrica,  
Universidade Federal de Minas Gerais  
and École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement  
Université Grenoble Alpes

by

Adrielle de Carvalho Santana

## Defense committee composition:

<i>Reviewers:</i>	Antônio M. F. L. M. de Sá	Professor, UFRJ, LAPIS, Brazil
	Sophie Dufour	Researcher, LPL (UMR 7309), France
<i>Examiners:</i>	Jean-Luc Schwartz	Research Director, GIPSA-Lab (UMR 5216), France
	Adriano V. Barbosa	Professor, UFMG, CEFALA, Brazil
<i>Supervisors:</i>	Rafael Laboissière	Researcher, LPNC (UMR 5105), France
	Hani C. Yehia	Professor, UFMG, CEFALA, Brazil

# BEHAVIORAL AND NEUROPHYSIOLOGICAL REPRESENTATIONS OF SPEECH PHONEMIC UNITS

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES**

**préparée dans le cadre d'une cotutelle entre  
l'Université Grenoble Alpes et l'Universidade Federal  
de Minas Gerais**

Spécialité: **Sciences Cognitives, Psychologie & Neurocognition et  
Génie Électrique**

Arrêté ministériel : le 6 janvier 2005 – 25 mai 2016

Présentée par

**Adrielle DE CARVALHO SANTANA**

Thèse dirigée par **Rafael Laboissière** et **Hani Camille Yehia**

préparée au sein du **Laboratoire de Psychologie et NeuroCognition  
(LPNC)** et **Centro de Estudos da Fala, Acústica, Linguagem e Música  
(CEFALA)**

dans l'École Doctorale Ingénierie pour la Santé, la Cognition et  
l'Environnement et le *Programa de Pós-Graduação em Engenharia  
Elétrica*

## **Représentations comportementales et neurophysiologique des unités phonémiques de la parole**

Thèse soutenue publiquement le **16 décembre 2020**, devant le jury composé de :

**M. Antônio M. F. L. M. DE SÁ**

Professeur, UFRJ, LAPIS, Brésil, Rapporteur

**Mme. Sophie DUFOUR**

Chargé de recherche, LPL (UMR 7309), France, Rapporteuse

**M. Jean-luc SCHWARTZ**

Directeur de recherche, GIPSA-Lab (UMR 5216), France, Président

**M. Adriano V. BARBOSA**

Professeur, UFMG, CEFALA, Brésil, Membre

**M. Rafael LABOISSIÈRE**

Chargé de recherche, LPNC (UMR 5105), France, Directeur de thèse

**M. Hani C. YEHIA**

Professeur, UFMG, CEFALA, Brésil, Directeur de thèse



**Adrielle de Carvalho Santana**

**BEHAVIORAL AND NEUROPHYSIOLOGICAL  
REPRESENTATIONS OF SPEECH PHONEMIC UNITS**

A thesis submitted to the Université Grenoble Alpes and the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of doctor in Cognitive Sciences, Psychology and Neurocognition and in Electrical Engineering in the field Computer and Telecommunications Systems.

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Programa de Pós-Graduação em Engenharia Elétrica

UNIVERSITÉ GRENOBLE ALPES

École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement

Supervisor: Dr. Hani CAMILLE YEHA / Dr. Rafael LABOISSIÈRE

Belo Horizonte - MG - Brasil

February 24, 2021

S232b

Santana, Adrielle de Carvalho.

Behavioral and neurophysiological representations of speech phonemic units [recurso eletrônico] / Adrielle de Carvalho Santana. - 2021. 1 recurso online (342 f. : il., color.) : pdf.

Orientadores: Hani Camille Yehia, Rafael Laboissière.

Tese (doutorado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Apêndices: f. 255-342.

Bibliografia: f. 235-254.

Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia elétrica - Teses. 2. Eletroencefalografia - Teses. 3. Atenção - Teses. 4. Percepção da fala - Teses. I. Yehia, Hani Camille. II. Laboissière, Rafael. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título.

CDU: 621.3(043)


**"Behavioral and Neurophysiological Representations  
of Speech Phonemic Units"**

**Adrielle de Carvalho Santana**


Doctoral Thesis submitted to the Examining Board appointed by the Collegiate of the Graduate Program in Electrical Engineering of the School of Engineering of the Federal University of Minas Gerais, as a requirement for obtaining the degree of Doctor of Electrical Engineering.

Approved on December 16th, 2020.

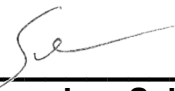
Por:

  
Prof. Dr. Hani Camille Yehia  
DELT (UFMG) - Orientador

  
Dr. Rafael Laboissière  
Laboratoire de Psychologie et NeuroCognition  
Université Grenoble Alpes (CNRS) - Orientador

  
Prof. Dr. Adriano Vilela Barbosa  
DELT (UFMG)

  
Prof. Dr. Antônio Maurício Ferreira Leite Miranda de Sá  
Instituto Luiz Coimbra de Pós Graduação e Pesquisa de Engenharia  
(UFRJ)

  
Dr. Jean-Luc Schwartz  
Laboratoire Grenoble Images Parole Signal Automatique (GIPSA-Lab)  
Université Grenoble Alpes (CNRS)

  
Dr. Sophie Dufour  
Laboratoire Parole et Langage  
Aix-Marseille Université (CNRS)

# Acknowledgements

I am very grateful to my family for all the support they have always given me in relation to my academic and professional training. I would like to thank professor Hani Camille Yehia for accepting to be my advisor, for the acquisition of the equipment necessary for the development of my work, for helping me deal with all the bureaucratic issues at UFMG and for the support during all the stages of the doctorate. I am very grateful to professor Rafael Laboissière for also accepting to be my advisor in the cotutelle agreement with the Université Grenoble Alpes (UGA), for the great reception given to me while I lived in Saint-Martin-d'Hères, for the idea to work with phonemic categorization, the idea to develop a specific regression technique to the problem of this work, for helping me deal with all the bureaucratic issues at UGA and for the support during the first year of the cotutelle and this year of 2020. I also want to thank all the teachers I had at UFMG who helped me in several stages of my work with knowledge of the neuroscience's area, the speech processing area (both new to me) and also statistics, signal processing and electronics. To all the volunteers who participated in my experiment, thank you very much for the patience and kindness. Without your brains, this work will not be possible. A special thanks to my friends Adriana, Marcos and Lucy (who I knew in France) and to all my friends and colleagues from CEFALA laboratory, UFMG and UFOP for all the support during the hard times and also for the wonderful moments we had together. I hope we can have many amazing new moments together in the future.

*“The human understanding resembles not a ‘dry light’, but admits a tincture of the will and passions, which generate their own system accordingly: for man always believes more readily that which he prefers. He, therefore, rejects difficulties for want of patience in investigation; sobriety, because it limits his hope; the depths of nature, from superstition; the light of experiment, from arrogance and pride, lest his mind should appear to be occupied with common and varying objects; paradoxes, from a fear of the opinion of the vulgar; in short, his feelings imbue and corrupt his understanding in innumerable and sometimes imperceptible ways.”*

(Francis Bacon, *Novum Organum*, 1620)<sup>1</sup>

---

<sup>1</sup> Source: <https://www.physicssayswhat.com/famous-quotes/>



# Resumo

O potencial evocado auditivo (PEA) é uma resposta neuroelétrica a um estímulo auditivo que reflete as atividades de um conjunto de neurônios ao longo das vias do sistema auditivo. Este biopotencial é utilizado no auxílio ao diagnóstico de transtornos auditivos e no estudo do processamento auditivo no cérebro humano. Assim, é interessante se trabalhar com estímulos mais complexos, tais como a fala, cujos parâmetros acústicos apresentam uma variação em tempo e frequência mais rica que os cliques ou tons utilizados nos exames audiométricos tradicionais. Uma das formas de se analisar o processamento da fala pelo cérebro humano é por meio do estudo da percepção categórica (PC) de fonemas que consiste em mapear mudanças contínuas dos sons em unidades perceptuais discretas durante uma identificação fonêmica. O objetivo deste trabalho é investigar os correlatos neurais da percepção categórica de fonemas em Português Brasileiro pela análise do PEA levando em conta as características acústicas dos fonemas, amplitude e latência das respostas, regiões corticais envolvidas, o grau de atenção à tarefa acústica (passiva ou ativa) e as características físicas ou psicofísicas da resposta. Um experimento foi realizado com tarefas que envolveram a categorização ativa e passiva de fonemas pertencentes ao longo de dois *continua* diferentes: um baseado em variações do *voice onset time* (VOT), e outro baseado em variações das frequências formantes. Os PEAs foram adquiridos via eletroencefalografia (EEG). A análise dos PEAs foi realizada nos domínios do tempo e do tempo-frequência em conjunto com dados comportamentais obtidos das curvas psicométricas dos participantes. No domínio do tempo foram analisadas as amplitudes e latências dos componentes N1 e P2 dos PEAs. No domínio tempo-frequência, os dados foram representados por meio de coeficientes da transformada wavelet discreta. Para extrair as representações física e psicofísica do processo de categorização, propusemos uma técnica de regressão, chamada *regression on low-dimension spanned input space* (RoLDSIS), que nos permite trabalhar com uma pequena quantidade de observações em um espaço de características muito grande. Modelos de efeitos mistos foram ajustados aos coeficientes de regressão da RoLDSIS e às amplitudes e latências das componentes N1 e P2. Os resultados mostraram que a percepção categórica é afetada pela característica acústica e pela tarefa e que é codificada em torno da latência N1 (e permanece nas latências tardias - P2) pelas bandas theta, alpha, beta e gamma. Vimos que cada banda de frequência e latência parecem codificar diferentes aspectos do som para o processamento da fala. Observou-se que participantes que apresentam comportamentalmente uma PC mais forte apresentam maior diferença entre a representação neural física e psicofísica dos estímulos. Esta diferença foi mais pronunciada para a característica acústica VOT do que para os formantes e para as tarefas ativas do que para as passivas. Mostrou-se também que a PC ocorre quando não há atenção à tarefa auditiva, mas apenas para a característica acústica baseada em formantes. Diferenças inter-hemisféricas também foram observadas, com atividade mais forte no hemisfério esquerdo. Também foram observadas diferenças entre as regiões corticais frontais e temporais codificadas

por ritmos de baixa frequência com mais atividade na região temporal. Na banda gama, não observamos diferença significativa entre a atividade nas regiões frontal e temporal. Nossos resultados mostraram que as estruturas da região temporal também podem realizar alguma categorização além do processamento das características acústicas físicas dos sons. Também mostramos como a característica e a tarefa acústicas reconfiguram dinamicamente a rede da fala o que deve ser levado em consideração por um modelo neurobiológico para a percepção da fala. Este estudo comparou diversos fatores relacionados à percepção categórica de fala no português brasileiro usando um protocolo reproduzível desenvolvido para o estudo e avaliação da percepção categórica fonêmica, e confirmou muitos dos resultados encontrados na literatura para outras línguas.

**Palavras-chave:** Resposta evocada auditiva de longa latência, eletroencefalografia, *voice onset time*, frequências formantes, atenção, transformada wavelet discreta, percepção categórica e regressão linear.

# Résumé

Le potentiel évoqué auditif (PEA) est une réponse CP neuroélectrique à un stimulus auditif qui reflète les activités d'un ensemble de neurones le long des voies du système auditif. Ce biopotential est utilisé pour aider au diagnostic des troubles auditifs et à l'étude du traitement auditif dans le cerveau humain. Ainsi, il est intéressant de travailler avec des stimuli plus complexes, comme la parole, dont les paramètres acoustiques montrent une variation de temps et de fréquence plus riche que les clics ou les tonalités utilisés dans les tests audiométriques traditionnels. L'un des moyens d'analyser le traitement de la parole par le cerveau humain consiste à étudier la perception catégorielle (PC) des phonèmes qui consiste à cartographier les changements continus des sons sur des unités perceptives discrètes lors de l'identification phonémique. L'objectif de ce travail est d'étudier les corrélats neuronaux de la perception catégorielle des phonèmes en portugais brésilien en analysant les PEAs en tenant compte des caractéristiques acoustiques des phonèmes, de l'amplitude et de la latence des réponses, des régions corticales impliquées, du degré de attention à la tâche acoustique (passive ou active) et aux caractéristiques physiques ou psychophysiques de la réponse. Une expérience a été menée avec des tâches qui impliquaient la catégorisation phonémique active et passive le long de deux continuums différents : l'un basé sur les variations du *voice onset time* (VOT), et l'autre basé sur les variations des fréquences des formants. Les PEAs ont été acquis par électroencéphalographie (EEG). L'analyse des PEAs a été réalisée dans les domaines temps et temps-fréquence en conjonction avec des données comportementales obtenues à partir des courbes psychométriques des participants. Dans le domaine temporel, les amplitudes et latences des composantes du PEAs N1 et P2 ont été analysées. Dans le domaine temps-fréquence, les données ont été représentées au moyen de coefficients d'ondelettes discrets. Pour extraire les représentations physiques et psychophysiques du processus de catégorisation, une technique de régression a été proposée, appelée *regression on low-dimension spanned input space* (RoLDSIS), qui permet de travailler avec une petite quantité d'observations dans un espace de caractéristiques de grande dimension. Des modèles à effets mixtes ont été ajustés aux coefficients de régression RoLDSIS et aux amplitudes et latences N1 et P2. Les résultats ont montré que la perception catégorielle est affectée par la caractéristique acoustique et par la tâche et qu'elle est observée dès la latence N1 (et reste dans les latences tardives - P2) par les activités des bandes thêta, alpha, bêta et gamma. Nous avons vu que chaque bande de fréquence et latence semble coder différents aspects du son pour le traitement de la parole. Il a été observé que les participants qui présentaient une PC comportementalement plus forte avaient une plus grande différence entre leur représentation neuronale physique et psychophysique des stimuli. Cette différence était prononcée pour la queue acoustique VOT que pour les formants et pour les tâches actives que pour les tâches passives. Il a également été montré que la PC se produit lorsqu'il n'y a pas d'attention à la tâche auditive mais uniquement pour le signal acoustique basé sur les formants. Des différences

hémisphériques ont été observées, avec une activité plus forte dans l'hémisphère gauche. Des différences ont également été observées entre les régions corticales frontales et temporales codées par des rythmes à basse fréquence avec plus d'activité au niveau de la région temporale. Dans la bande gamma, nous n'avons observé aucune différence significative entre l'activité dans les régions frontale et temporale. Nos résultats ont montré que les structures de régions temporelles peuvent également effectuer une certaine catégorisation en plus du traitement des caractéristiques acoustiques physiques des sons. Nous montrons également comment le signal acoustique et la tâche reconfigurent dynamiquement le réseau de parole qui doit être pris en compte par un modèle neurobiologique de perception de la parole. Cette étude a comparé différents facteurs liés à la perception catégorielle de la parole en portugais brésilien à l'aide d'un protocole reproductible développé pour l'étude et l'évaluation de la perception phonémique catégorielle, et a confirmé de nombreux résultats trouvés dans la littérature pour d'autres langues.

**Mots-clés :** Réponse tardive évoquée auditive, électroencéphalographie, *voice onset time*, fréquences des formants, attention, transformation en ondelettes discrète, perception catégorielle et régression linéaire.

# Abstract

The auditory evoked potential (AEP) is a neuroelectric response to an auditory stimulus that reflects the activities of a set of neurons along the pathways of the auditory system. This biopotential is used to aid in the diagnosis of hearing disorders and in the study of auditory processing in the human brain. Thus, it is interesting to work with more complex stimuli, such as speech, whose acoustic parameters show a richer variation in time and frequency than the clicks or tones used in traditional audiometric tests. One of the ways to analyze speech processing by the human brain is through the study of categorical perception (CP) of phonemes which consists of mapping continuous changes in sounds onto discrete perceptual units during phonemic identification. The objective of this work is to investigate the neural correlates of categorical perception of phonemes in Brazilian Portuguese by analyzing the AEPs taking into account the acoustic characteristics of the phonemes, the amplitude and the latency of the responses, the cortical regions involved, the degree of attention to the acoustic task (passive or active) and the physical or psychophysical characteristics of the response. An experiment was carried out with tasks that involved the active and passive phonemic categorization along two different continua: one based on variations of the voice onset time (VOT), and another based on variations of the formant frequencies. AEPs were acquired via electroencephalography (EEG). The analysis of the AEPs was performed in time and time-frequency domains in conjunction with behavioral data obtained from the participants' psychometric curves. In the time-domain, the amplitudes and latencies of the AEP components N1 and P2 were analyzed. In the time-frequency domain, data were represented by means of discrete wavelet coefficients. To extract the physical and psychophysical representations of the categorization process, a regression technique was proposed, called regression on low-dimension spanned input space (RoLDSIS), that allows working with a small amount of observations in a large dimensional feature space. Mixed-effects models were fitted to the RoLDSIS regression coefficients and to the N1 and P2 amplitudes and latencies. The results showed that the categorical perception is affected by the acoustic characteristic and by the task and that it is observed as early as in the N1 latency (and remains in late latencies - P2) by the theta, alpha, beta and gamma band activities. We saw that each frequency band and latency seems to code different aspects of the sound for the speech processing. It was observed that participants who presented behaviorally stronger CP had a larger difference between their physical and psychophysical neural representation of the stimuli. This difference was pronounced for the VOT acoustic cue than for the formants and for active tasks than for the passive ones. It was also shown that the CP occurs when there is no attention to the auditory task but only for the formant-based acoustic cue. Hemispheric differences were observed, with stronger activity at the left hemisphere. Differences were also observed between frontal and temporal cortical regions coded by low-frequency rhythms with more activity at the temporal region. In the gamma band we observed no significant difference between the activity

at the frontal and temporal regions. Our results showed that temporal region structures may also perform some categorization besides the processing of physical acoustic characteristics of the sounds. We also show how the acoustic cue and task dynamically reconfigure the speech network which should be taken into account by a neurobiological model for speech perception. This study compared different factors related to categorical speech perception in Brazilian Portuguese using a reproducible protocol developed for the study and the evaluation of phonemic categorical perception, and confirmed many of the results found in the literature for other languages.

**Keywords:** Auditory evoked late response, electroencephalography, voice onset time, formant frequencies, attention, discrete wavelet transform, categorical perception and linear regression.

# List of Figures

Figure 1 – Physical and psychophysical (categorical) neurophysiological axes for the /da/–/ta/ continuum. . . . .	36
Figure 2 – Structure of the human ear. . . . .	39
Figure 3 – Structures of the cochlea. . . . .	40
Figure 4 – Auditory pathway from the inner ear to the primary auditory cortex. . . . .	41
Figure 5 – Hypothetical articulatory/acoustic relation. . . . .	43
Figure 6 – Dual-stream model for the speech processing. . . . .	44
Figure 7 – Averaging. . . . .	46
Figure 8 – AELR waveform showing major waves at typical latencies including P1, N1, P2, N2 and P3. . . . .	48
Figure 9 – Relation between inter-stimulus interval and amplitude of the N1-P2 complex. . . . .	53
Figure 10 – Electric field pattern in response to EPSP and IPSP. . . . .	71
Figure 11 – Meninges. . . . .	72
Figure 12 – Room layout for the acquisition protocol adopted. . . . .	73
Figure 13 – Electrode placement scheme. . . . .	75
Figure 14 – Hand-made cap used in the experiments. . . . .	77
Figure 15 – RHD headstage with the RHD2132 amplifier chip. . . . .	80
Figure 16 – RHD2132 amplifier chip IC. Internal Circuit Diagram. . . . .	81
Figure 17 – RHD2000 system with the (a) interface board, the (b) SPI cable and the (c) RHD 32-channel headstage. . . . .	81
Figure 18 – RHD2000 System Development Board. . . . .	82
Figure 19 – Interface to the RHD2000 system. . . . .	83
Figure 20 – Melon head model for artifact detection in the RHD2000 system. . . . .	85
Figure 21 – Power supply adapted for the RHD2000. . . . .	86
Figure 22 – Cable to separate audio channels. . . . .	87
Figure 23 – Common GND system causing ground loop. . . . .	87
Figure 24 – A noise voltage enters the amplifier if the circuit has more than one GND point. . . . .	88
Figure 25 – Trigger signal on second stimulus audio channel for the [da] syllable. . . . .	88
Figure 26 – Optocoupler circuit and voltage divider. . . . .	89
Figure 28 – Coherent average of 350 stimulus repetitions /da/ in human experiment. . . . .	90
Figure 27 – Coherent average of 400 stimulus repetitions /ba/ melon model. . . . .	90
Figure 29 – Speech Production Model. . . . .	91
Figure 30 – Spectrogram of the syllable /da/. . . . .	92
Figure 31 – F1×F2 chart of Brazilian Portuguese tonic oral vowels. Large font: women; small font: men. . . . .	93
Figure 32 – Microphone Brüel & Kjaer for stimuli acquisition. . . . .	94

Figure 33 – Samples of the stimuli continuum based on formant variation. . . . .	95
Figure 34 – Samples of the stimuli continuum based on VOT variation. . . . .	96
Figure 35 – Earphone ATH-ANC33iS Audio-Thecnica <sup>®</sup> . . . . .	98
Figure 36 – Frequency response of the Earphone ATH-ANC33iS Audio-Thecnica <sup>®</sup> (averaged and compensated). . . . .	98
Figure 37 – Psychometric curve of the /pa/-/pɛ/ continuum for a representative participant. . . . .	101
Figure 38 – Frequency response of the 60Hz notch filter applied in the data. . . . .	105
Figure 39 – Bootstrap analysis for the participant 5, formant-active experimental condition, stimulus 1. . . . .	110
Figure 40 – Values of the magnitude N1–P2 obtained through direct measurement at the AELR averages and through the bootstrap technique for all cases. . . . .	110
Figure 41 – Values of the latency difference T2–T1 obtained through direct measurement at the AELR averages and through the bootstrap technique for all cases. . . . .	111
Figure 42 – Linear, ambiguity and psy-phy contrasts. . . . .	112
Figure 44 – Mean response time of all participants for the identification of the stimuli in the active task of the formant continuum. . . . .	117
Figure 43 – Mean response time of all participants for the identification of the stimuli in the active task of the VOT continuum. . . . .	117
Figure 45 – Projections of the mean ISI for active and passive tasks onto the curve given by Hall III (2015) to compute the mean increase in N1-P2 from the passive to active task. . . . .	118
Figure 46 – Average psychometric curves obtained from all participants results for the VOT (left) and the formant (right) continua. . . . .	120
Figure 47 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the VOT-act condition in the temporal cortex. . . . .	122
Figure 48 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the VOT-pass condition in the temporal cortex. . . . .	123
Figure 49 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the Form-act condition in the temporal cortex. . . . .	123
Figure 50 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the Form-pass condition in the temporal cortex. . . . .	124
Figure 51 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the VOT-act condition in the frontal cortex. . . . .	124
Figure 52 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the VOT-pass condition in the frontal cortex. . . . .	125
Figure 53 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the Form-act condition in the frontal cortex. . . . .	125
Figure 54 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the Form-pass condition in the frontal cortex. . . . .	126



Figure 55 – Grand averages of the auditory long latency potentials for Fz electrode at the frontocentral region for the VOT-act condition. . . . .	127
Figure 56 – Grand averages of the auditory long latency potentials for Fz electrode at the frontocentral region for the VOT-pass condition. . . . .	127
Figure 57 – Grand averages of the auditory long latency potentials for Fz electrode at the frontocentral region for the Form-act condition. . . . .	128
Figure 58 – Grand averages of the auditory long latency potentials for Fz electrode at the frontocentral region for the Form-pass condition. . . . .	128
Figure 59 – Representation of the complete mixed-effects model including the fixed factors.	131
Figure 60 – Density distribution of the results of all the divisions of the N1-P2 values, for all factors, at the active task by the passive task ones. . . . .	132
Figure 61 – Representation of the complete mixed-effects model including the fixed factors.	133
Figure 62 – Representation of the factor “feature” over the N1-P2 magnitude. . . . .	134
Figure 63 – Representation of the factor electrode over the variable N1-P2. . . . .	136
Figure 64 – Representation of the factor stimulus over the variable N1-P2. . . . .	136
Figure 65 – Representation of the interaction factor feature:stimulus over the variable N1-P2. . . . .	138
Figure 66 – Representation of the interaction factor feature:type over the variable N1-P2.	139
Figure 67 – Representation of the interaction factor feature:electrode over the variable N1-P2. . . . .	141
Figure 68 – Representation of the interaction factor type:electrode over the variable N1-P2.	141
Figure 69 – Representation of the complete mixed-effects model including the fixed factors.	143
Figure 70 – Representation of the factor type over the variable N1. . . . .	144
Figure 71 – Representation of the factor electrode over the variable N1. . . . .	145
Figure 72 – Representation of the factor stimulus over the variable N1. . . . .	145
Figure 73 – Representation of the interaction factor feature:stimulus over the variable N1.	146
Figure 74 – Representation of the interaction factor type:electrode over the variable N1. . . . .	147
Figure 75 – Representation of the interaction factor electrode:stimulus over the variable N1.	148
Figure 76 – Representation of the complete mixed-effects model including the fixed factors.	150
Figure 77 – Representation of the factor feature over the variable P2. . . . .	151
Figure 78 – Representation of the factor type over the variable P2. . . . .	151
Figure 79 – Representation of the factor electrode over the variable P2. . . . .	152
Figure 80 – Representation of the factor stimulus over the variable P2. . . . .	153
Figure 81 – Representation of the interaction factor feature:electrode over the variable P2.	154
Figure 82 – Representation of the interaction factor type:electrode over the variable P2. . . . .	155
Figure 83 – Representation of the complete mixed-effects model including the fixed factors.	157
Figure 84 – Representation of the factor feature over the variable T1. . . . .	158
Figure 85 – Representation of the factor electrode over the variable T1. . . . .	158
Figure 86 – Representation of the factor stimulus over the variable T1. . . . .	159

Figure 87 – Representation of the interaction factor feature:type over the variable T1. . . . .	161
Figure 88 – Representation of the interaction factor feature:electrode over the variable T1.	162
Figure 89 – Representation of the interaction factor type:electrode over the variable T1. . . . .	163
Figure 90 – Representation of the interaction factor feature:stimulus over the variable T1.	164
Figure 91 – Representation of the interaction factor type:stimulus over the variable T1. . . . .	165
Figure 92 – Representation of the complete mixed-effects model including the fixed factors.	166
Figure 93 – Representation of the factor feature over the variable T2. . . . .	167
Figure 94 – Representation of the factor electrode over the variable T2. . . . .	168
Figure 95 – Representation of the factor stimulus over the variable T2. . . . .	169
Figure 96 – Representation of the interaction factor feature:type over the variable T2. . . . .	170
Figure 97 – Representation of the interaction factor type:electrode over the variable T2. . . . .	171
Figure 98 – Representation of the interaction factor electrode:stimulus over the variable T2.	172
Figure 99 – Analysis of the effect of the factor feature over the $\Delta$ ERP computed for the N1-P2 magnitudes. . . . .	173
Figure 100–Analysis of the effect of the factor feature over the $\Delta$ ERP computed for the N1 amplitude. . . . .	174
Figure 101–Analysis of the effect of the factor electrode over the $\Delta$ ERP computed for the P2 amplitude. . . . .	175
Figure 102–Relation of the factors feature, type and stimulus for the P2 amplitude values considering stim1, stim3 and stim5 used to compute the $\Delta$ ERP variable. . . . .	176
Figure 103–Graphic representation of the RoLDSIS technique (see section 7.1.1 for details).	183
Figure 104–Results of the phonemic identification task for a representative participant. . . . .	184
Figure 105–Direction obtained for the RoLDSIS procedure for a representative participant.	186
Figure 106–Projections of ERPs for stimuli stim1, stim2, stim3, stim4, and stim5 onto the axis found by the RoLDSIS procedure, for a representative participant. . . . .	187
Figure 107–Population scatter plot of the slope of psychometric curve and the angle between the psychophysical directions and the physical directions. . . . .	189
Figure 108–Bootstrap results of the RoLDSIS procedure. . . . .	191
Figure 109–Cross-validation errors for the proposed regression method (RoLDSIS) and the methods of regularized linear regression for physical (left panel) and psychophysical (right panel) attributes. . . . .	193
Figure 110–Scalograms of the regression results. . . . .	195
Figure 111–Definition of the ROIs as dependent variables for the models. Each ROI corresponds to specific a frequency band and specific a time interval. . . . .	200
Figure 112–Relation between the physical and psychophysical direction vector angle and the slope ( $\beta$ ) of the psychometric curve of all participants for the VOT-active experimental condition. . . . .	202

Figure 113–Relation between the physical and psychophysical direction vector angle and the slope ( $\beta$ ) of the psychometric curve of all participants for the VOT-passive experimental condition. . . . .	203
Figure 114–Relation between the physical and psychophysical direction vector angle and the slope ( $\beta$ ) of the psychometric curve of all participants for the Form-active experimental condition. . . . .	204
Figure 115–Relation between the physical and psychophysical direction vector angle and the slope ( $\beta$ ) of the psychometric curve of all participants for the Form-passive experimental condition. . . . .	205
Figure 116–Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the VOT continuum active task, psychophysical response. Scalograms are organized according with the electrodes position in the scalp. . . . .	205
Figure 117–Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. . . . .	206
Figure 118–Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the VOT continuum passive task, psychophysical response. Scalograms are organized according with the electrodes position in the scalp. . . . .	206
Figure 119–Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the VOT continuum passive task, physical response. Scalograms are organized according with the electrodes position in the scalp. . . . .	207
Figure 120–Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the Formants continuum active task, psychophysical response. Scalograms are organized according with the electrodes position in the scalp. . . . .	207
Figure 121–Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the Formants continuum active task, physical response. Scalograms are organized according with the electrodes position in the scalp. . . . .	208
Figure 122–Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the Formants continuum passive task, psychophysical response. Scalograms are organized according with the electrodes position in the scalp. . . . .	208
Figure 123–Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the Formants continuum passive task, physical response. Scalograms are organized according with the electrodes position in the scalp. . . . .	209

Figure 124–Representation of the complete mixed-effects model for the discrepancy computed for the ROI-1 including the fixed factors: feature, type and electrodes.	210
Figure 125–Representation of the factor electrode over the discrepancy computed for all ROIs. . . . .	212
Figure 126–Representation of the factor feature over the discrepancy computed for all ROIs.	217
Figure 127–Representation of the factor type over the discrepancy computed for all ROIs.	220
Figure 128–Representation of the factor interaction feature-electrode over the discrepancy computed for the ROI-2, ROI-4 and ROI-6. . . . .	223
Figure 129–Representation of the factor interaction feature-type over the discrepancy computed for the ROI-3 and ROI-8. . . . .	225
Figure 130–Representation of the factor interaction type-electrode over the discrepancy computed for the ROI-3, ROI-7 and ROI-8. . . . .	227
Figure 131–Relation between ABR waveform latency and the structures involved in its origin. . . . .	256
Figure 132–Latencies band for each ABR component. . . . .	257
Figure 133–AMLR waveform showing major peaks at typical latencies including Na, Pa, Nb, and Pb. . . . .	258
Figure 134–Effect of stimulus type and rate of presentation at the AMLR waveform. . .	259
Figure 135–Signals from each channel, filtered with a 6th order butterworth filter (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 39 Hz. . . . .	263
Figure 136–Signals from each channel, filtered with a 6th order butterworth filter (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz. . . . .	264
Figure 137–Signals from each channel, filtered using the DWT band elimination approach (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz. . . . .	265
Figure 138–Signals from each channel, filtered using the DWT band elimination approach (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz. . . . .	266
Figure 139–Signals from each channel, filtered using the DWT band elimination approach combined with the wavelet thresholding technique (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz. . .	268
Figure 140–Signals from each channel, filtered using the DWT band elimination approach combined with the wavelet thresholding technique (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz. . .	269
Figure 141–Mean SNR for all tested frequencies and methods for one representative participant in the VOT active experiment. . . . .	270
Figure 142–Linear Regression. . . . .	272

Figure 143–Correlation of the regression predictions with the physical response for the normal (top) and randomized (bottom) input matrices of the VOT-active condition in the temporal cortex. Responses are separated by side (left or right) and the Pearson’s correlation coefficient is showed as the variable R2.	281
Figure 144–Average projections of stimuli 4 and 2 on the physical and psychophysical axis of all subjects for the VOT-active condition.	283
Figure 145–Elastic Net coefficients for the VOT-active condition for physical left and right and psychophysical left and right responses.	284
Figure 146–Comparison between the DWT and the discrete Fourier transform (DFT). The transforms are applied in a square wave with 512 samples as example. (A) DFT transforms the 512 samples of the time domain signal into 512 coefficients in frequency domain. (B) DWT transforms the 512 samples of the time domain signal into 512 wavelet coefficients (‘a’: approximation coefficients; ‘d’: detail coefficients) obtained through a recursive filtering and downsampling process. (C) DFT sinusoidal basis function is infinite in time while (D) wavelet basis functions are limited and localized in time. CREDITS: McKay et al. (2013) (adapted)	286
Figure 147–Scale and shift of mother wavelets filtering a signal (left). A DWT scalogram (right).	287
Figure 148–DWT of a chirp signal (signal with variable frequency over time) decomposed in 4 wavelet levels.	288
Figure 149–Representation of the complete mixed-effects model for the discrepancy computed for the ROI-2 including the fixed factors: feature, type, electrodes and regression.	290
Figure 150–Representation of the complete mixed-effects model for the discrepancy computed for the ROI-3 including the fixed factors: feature, type, electrodes and regression.	291
Figure 151–Representation of the complete mixed-effects model for the discrepancy computed for the ROI-4 including the fixed factors: feature, type, electrodes and regression.	293
Figure 152–Representation of the complete mixed-effects model for the discrepancy computed for the ROI-5 including the fixed factors: feature, type, electrodes and regression.	294
Figure 153–Representation of the complete mixed-effects model for the discrepancy computed for the ROI-6 including the fixed factors: feature, type, electrodes and regression.	296
Figure 154–Representation of the complete mixed-effects model for the discrepancy computed for the ROI-7 including the fixed factors: feature, type, electrodes and regression.	297

Figure 155–Representation of the complete mixed-effects model for the discrepancy computed for the ROI-8 including the fixed factors: feature, type, electrodes and regression. . . . .	299
Figure 156–Graphical representation for normality verification. . . . .	303
Figure 157–Graphical representation for normality and homocedasticity verification. . .	303
Figure 158–Graphical representation for normality verification. . . . .	304
Figure 159–Graphical representation for normality and homocedasticity verification. . .	305
Figure 160–Graphical representation for normality verification. . . . .	306
Figure 161–Graphical representation for normality and homocedasticity verification. . .	306
Figure 162–Graphical representation for normality verification. . . . .	307
Figure 163–Graphical representation for normality and homocedasticity verification. . .	308
Figure 164–Graphical representation for normality verification. . . . .	309
Figure 165–Graphical representation for normality and homocedasticity verification. . .	309
Figure 166–Graphical representation for normality verification. . . . .	310
Figure 167–Graphical representation for normality and homocedasticity verification. . .	311
Figure 168–Graphical representation for normality verification. . . . .	312
Figure 169–Graphical representation for normality and homocedasticity verification. . .	312
Figure 170–Graphical representation for normality verification. . . . .	313
Figure 171–Graphical representation for normality and homocedasticity verification. . .	314
Figure 172–Graphical representation for normality verification. . . . .	315
Figure 173–Graphical representation for normality and homocedasticity verification. . .	315
Figure 174–Graphical representation for normality verification. . . . .	316
Figure 175–Graphical representation for normality and homocedasticity verification. . .	317
Figure 176–Graphical representation for normality verification. . . . .	318
Figure 177–Graphical representation for normality and homocedasticity verification. . .	318
Figure 178–Graphical representation for normality verification. . . . .	319
Figure 179–Graphical representation for normality and homocedasticity verification. . .	320
Figure 180–Graphical representation for normality verification. . . . .	321
Figure 181–Graphical representation for normality and homocedasticity verification. . .	321
Figure 182–Prediction error of linear regression for overdetermined cases. Traditional least squares regression applied to the linear model relating ERP feature vectors and either physical (left) or psychophysical (right) attributes. The RMS prediction error is shown in the vertical axis. The number of trials per observation, varying from 1 to 6, is shown in the horizontal axis. Dots and vertical bars represent, respectively, the means and standard deviations obtained for the 11 participants. . . . .	324
Figure 183–Psychometric curve for the participant 1 for the VOT continuum. . . . .	327
Figure 184–Psychometric curve for the participant 2 for the VOT continuum. . . . .	327
Figure 185–Psychometric curve for the participant 3 for the VOT continuum. . . . .	328

Figure 186–Psychometric curve for the participant 4 for the VOT continuum. . . . .	328
Figure 187–Psychometric curve for the participant 5 for the VOT continuum. . . . .	329
Figure 188–Psychometric curve for the participant 6 for the VOT continuum. . . . .	329
Figure 189–Psychometric curve for the participant 7 for the VOT continuum. . . . .	330
Figure 190–Psychometric curve for the participant 8 for the VOT continuum. . . . .	330
Figure 191–Psychometric curve for the participant 9 for the VOT continuum. . . . .	331
Figure 192–Psychometric curve for the participant 10 for the VOT continuum. . . . .	331
Figure 193–Psychometric curve for the participant 11 for the VOT continuum. . . . .	332
Figure 194–Psychometric curve for the participant 1 for the Formant continuum. . . . .	333
Figure 195–Psychometric curve for the participant 2 for the Formant continuum. . . . .	333
Figure 196–Psychometric curve for the participant 3 for the Formant continuum. . . . .	334
Figure 197–Psychometric curve for the participant 4 for the Formant continuum. . . . .	334
Figure 198–Psychometric curve for the participant 5 for the Formant continuum. . . . .	335
Figure 199–Psychometric curve for the participant 6 for the Formant continuum. . . . .	335
Figure 200–Psychometric curve for the participant 7 for the Formant continuum. . . . .	336
Figure 201–Psychometric curve for the participant 8 for the Formant continuum. . . . .	336
Figure 202–Psychometric curve for the participant 9 for the Formant continuum. . . . .	337
Figure 203–Psychometric curve for the participant 10 for the Formant continuum. . . . .	337
Figure 204–Psychometric curve for the participant 11 for the Formant continuum. . . . .	338

# List of Tables

Table 1 – Discrete wavelet transform levels and frequencies used in time domain analysis.	106
Table 2 – Discrete wavelet transform levels and frequencies used in time-frequency domain analysis. . . . .	197



# Acronyms

ABR	auditory brainstem response . . . . .	46
A/D	analog/digital . . . . .	78
AELR	auditory evoked late response . . . . .	34
AEP	auditory evoked potential . . . . .	29
AMLR	auditory middle-latency response . . . . .	46
BCI	brain-computer interface . . . . .	179
CEFALA	Centro de Estudos da Fala, Acústica, Linguagem e Música . . . . .	72
CM	cochlear microphonics . . . . .	79
CV	consoant-vowel . . . . .	194
CP	categorical perception . . . . .	29
DWT	discrete wavelet transform . . . . .	32
CWT	continuous wavelet transform . . . . .	234
ECoG	electrocorticography . . . . .	70
ERP	event-related potential . . . . .	29
EEG	electroencephalography . . . . .	32
EP	evoked potential . . . . .	45
FFR	frequency following response . . . . .	50
fMRI	functional magnetic resonance imaging . . . . .	70
FPGA	field programmable gate array . . . . .	83
HDLSS	high dimension low sample size . . . . .	179
IC	integrated circuit . . . . .	79
IFG	inferior frontal gyrus . . . . .	214
ISI	inter-stimuli interval . . . . .	45
LASSO	least absolute shrinkage and selection operator . . . . .	35
LDA	linear discriminant analysis . . . . .	190
MEG	magnetoencephalography . . . . .	70
MSE	mean squared error . . . . .	192
NIRS	near infrared spectroscopy . . . . .	178
PAMR	post-auricular muscle response . . . . .	78
PC	principal component . . . . .	190
PCA	principal component analysis . . . . .	54
PCR	principal component regression . . . . .	325
PET	positron emission tomography . . . . .	70
PM	premotor cortex . . . . .	214
PoA	place of articulation . . . . .	29
pSTG	posterior superior temporal gyrus . . . . .	29

RMS	root mean square . . . . .	323
ROI	region of interest . . . . .	199
ROIs	regions of interest . . . . .	196
RoLDSIS	regression on low-dimension spanned input space . . . . .	35
RT	response time . . . . .	116
SNR	signal-to-noise ratio . . . . .	47
SOA	stimulus onset asynchrony . . . . .	53
SPI	serial peripheral interface . . . . .	82
SPLS	sparse partial least squares . . . . .	180
STG	superior temporal gyrus . . . . .	30
STS	superior temporal sulcus . . . . .	30
STRAIGHT	speech transformation and representation using adaptive interpolation of weighted spectrum . . . . .	93
UPS	uninterruptible power supply . . . . .	85
VOT	voice onset time . . . . .	30

# Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>28</b>
1.1	Background	28
1.2	Motivation	31
1.3	Objective	33
1.3.1	Specific objectives	33
1.4	Methodology	34
1.5	Dissertation organization	37
<b>2</b>	<b>AUDITORY EVOKED POTENTIALS</b>	<b>38</b>
2.1	Auditory pathway	38
2.1.1	Speech perception models	42
2.2	Event-related potential	44
2.3	Auditory evoked late response - AELR	48
2.3.1	AELR acquisition	50
2.3.2	Factors affecting latency and amplitude of AELR components	51
2.3.3	Remarks about AELR analysis	54
<b>3</b>	<b>PHONEMIC CATEGORICAL PERCEPTION</b>	<b>58</b>
3.1	Categorical perception	58
3.2	Speech categorization	61
3.2.1	Experience with the categorized sound	62
3.2.2	Laterality and brain auditory processing regions	63
3.2.3	Attention and cortical region	66
3.2.4	Final considerations	68
<b>4</b>	<b>MATERIALS AND METHODS</b>	<b>70</b>
4.1	Brain potentials	70
4.2	Acquisition scheme	72
4.3	Acquisition	74
4.3.1	Electrodes	74
4.3.2	Artifacts	78
4.3.3	RHD2000 system	79
4.3.4	Circuit artifact treatment	84
4.4	Speech stimuli	91
4.5	Acquisition procedure	97
4.5.1	Psychometric curve plot	100

4.5.2	Active stage . . . . .	101
4.5.3	Passive stage . . . . .	102
4.5.4	Comment . . . . .	103
	<b>5 TIME DOMAIN PROCESSING . . . . .</b>	<b>104</b>
5.1	Filtering . . . . .	104
5.1.1	Notch filter . . . . .	104
5.1.2	Discrete wavelet transform filtering . . . . .	105
5.2	Data preprocessing . . . . .	106
5.3	Time domain analysis . . . . .	108
5.3.1	Processing for time domain analysis . . . . .	109
5.3.2	$\Delta$ ERP test . . . . .	112
5.3.3	Averaging test after bad trials removal . . . . .	113
5.3.4	Comment . . . . .	114
	<b>6 RESULTS OF THE TIME-DOMAIN ANALYSIS . . . . .</b>	<b>115</b>
6.1	Response times . . . . .	116
6.2	Grand averages and psychometric curves . . . . .	119
6.3	Analysis of mean latencies and amplitudes of AELR waves . . . . .	126
6.3.1	N1-P2 magnitude analysis . . . . .	130
6.3.2	N1 analysis . . . . .	142
6.3.3	P2 analysis . . . . .	149
6.3.4	T1 analysis . . . . .	156
6.3.5	T2 analysis . . . . .	166
6.3.6	$\Delta$ ERP . . . . .	172
	<b>7 REGRESSION ON LOW-DIMENSION SPANNED INPUT SPACE</b>	
	– ROLDSIS . . . . .	178
7.1	Background . . . . .	178
7.1.1	The RoLDSIS technique . . . . .	180
7.2	Example of application of RoLDSIS . . . . .	183
7.2.1	Projections onto physical and psychophysical directions . . . . .	185
7.3	Assessment of the RoLDSIS technique . . . . .	188
7.3.1	Relationship between $\Phi$ and $\Psi$ divergence and the degree of categorization . . . . .	188
7.3.2	Bootstrap analysis . . . . .	189
7.3.3	Comparison with regularized linear regression procedures . . . . .	191
	<b>8 TIME-FREQUENCY DOMAIN ANALYSIS . . . . .</b>	<b>196</b>
8.1	Data processing . . . . .	197
8.1.1	Resampling of the ERP signals . . . . .	197
8.1.2	ROI selection and mixed-effects models . . . . .	198

8.2	Results	201
8.2.1	Relationship between $\Phi$ and $\Psi$ divergence and the degree of categorization	201
8.2.2	Mean scalograms	204
8.2.3	Mixed effects models for selected ROIs	209
8.2.3.1	Electrode	211
8.2.3.2	Feature	216
8.2.3.3	Type	219
8.2.3.4	Feature-electrode	222
8.2.3.5	Feature-type	225
8.2.3.6	Type-electrode	226
	<b>9 CONCLUSION</b>	<b>229</b>
9.1	General discussion	229
9.2	Availability of the code	233
9.3	Future extensions	234
	<b>Bibliography</b>	<b>235</b>
	<b>APPENDIX A MIDDLE AND SHORT LATENCY EVOKED RESPONSES</b>	<b>255</b>
A.1	Auditory brainstem response - ABR	255
A.2	Auditory middle-latency response - AMLR	257
	<b>APPENDIX B SIGNAL FILTERING TESTS</b>	<b>261</b>
B.1	Butterworth filter	262
B.2	Wavelet filtering with DWT	264
B.3	Wavelet thresholding plus band filtering	266
B.4	Final considerations	269
	<b>APPENDIX C REGRESSION TECHNIQUES</b>	<b>271</b>
C.1	Regularization and characteristics selection	272
C.1.1	Ridge Regression	273
C.1.2	LASSO	274
C.2	Elastic Net	275
C.3	Principal Component Regression	276
C.4	Partial Least Squares Regression	276
C.5	Sparse Partial Least Squares	277
C.6	Assembly of the EN input matrix and response vectors	278
C.7	Testing Elastic Net Responses	282
	<b>APPENDIX D DISCRETE WAVELET TRANSFORM</b>	<b>285</b>

<b>APPENDIX E</b>	<b>COMPLETE ANOVA AND RANOVA TABLES . . .</b>	<b>289</b>
<b>APPENDIX F</b>	<b>ASSUMPTIONS . . . . .</b>	<b>301</b>
<b>APPENDIX G</b>	<b>THE NEED FOR AVERAGING ACROSS TRIALS .</b>	<b>323</b>
<b>APPENDIX H</b>	<b>PSYCHOMETRIC CURVES . . . . .</b>	<b>326</b>
<b>APPENDIX I</b>	<b>SCRIPTS . . . . .</b>	<b>339</b>

# Chapter 1

## INTRODUCTION

The transformation of speech acoustic waves into meaningful messages by the human brain is a complex process. Its comprehension demands measurements of physical and psychophysical phenomena which must be modeled and interpreted. This is a hard task, as neurophysiological measurements are difficult to carry out, inherently noisy and result from the combination of several components. However, it is a necessary step to understand hearing and speech disorders, as well as to develop new technologies for man-machine interfaces.

In this study, auditory event related potentials are measured for single syllables varying continuously between the formant frequencies of two vowels and between the voice onset times of two consonants. These measurements are performed in two different attention conditions. Then, amplitude and latency of the potentials are analyzed in the time domain. Next, classic regression analysis is used to develop a novel method to understand how these potentials are related to physical characteristics of speech sounds and to psychophysical characteristics associated with phonemic categorization. This method is then used to explore the contents of the measured event related potentials in the time-frequency domain.

### 1.1 Background

According to ([Anderson and Kraus, 2011](#)), in the auditory system, around 30,000 fibers generate impulses at a rate of 0 to 300 spikes per second or 9 million of data bits per second, considering 1 bit per spike. The human nervous system has mechanisms that enable the incredible task of

processing and interpreting these data in real time. It is believed that the way it does this is by detecting patterns in the acoustic stimuli so that it is possible to differentiate between, for example, a specific person's speech in a dialogue and the sound of a piano that is played at the same time and in the same environment where the dialogue takes place. Thus, the detected "pattern" will be the information effectively used by the nervous system to identify the auditory stimulus. Several studies relate a categorical perception (CP) process (associated with the pattern detection) to this identification (Alho et al., 2014, Bidelman et al., 2014, Bidelman and Walker, 2017, Picton et al., 1999, Husain et al., 2006, Chevillet et al., 2013, Feldman et al., 2009, Liégeois-Chauvel et al., 1999).

Humans perceive the world by categorizing sensory inputs. This is the case, for example, of color perception, facial emotions and speech (Harnad, 1987). In the case of speech, continuous changes in sounds are mapped onto discrete perceptual classes, in a process called phonemic categorization (Bidelman et al., 2013). This phenomenon has been studied for more than 60 years in behavioral experiments. Their neurophysiological mechanisms have been better investigated recently due to improvements in technologies for brain signal acquisition, the availability of computational power and the development and application of techniques of machine learning to big data processing.

The categorical perception of phonemes is performed by different cortical areas depending on some factors. One of them is the attention to the auditory stimulus. Temporal and frontal cortical regions are involved in the phonemic processing of speech streams and, more specifically, in its categorization in early stages of the auditory stimulus processing (Chang et al., 2010, Bouton et al., 2018, Alho et al., 2016, Möttönen et al., 2014, Bidelman and Walker, 2017).

Some studies investigating the auditory categorical perception present different results depending on the cortical area considered and the attention to the auditory task. For example, Bidelman and Walker (2017) noted that the neural coding of the categorical effect requires attention to the auditory task (active task) with measurements of event-related potentials (ERPs) made at the frontocentral cortical region in the time-frame of the N1-P2 complex (between 100 and 200 ms) which are common waves in auditory evoked potential (AEP) studies. Chang et al. (2010), in turn, reported a categorical effect 110 ms after stimulus onset in the posterior superior temporal gyrus (pSTG) in a task that did not require attention (passive task). The work of Bouton et al. (2018) also showed an early categorical coding (90–120 ms) but in a categorization task with attention with a repetition of the effect at a higher latency (175 ms) in the pSTG. All these studies used stimuli differing in their spectral acoustic cue. Bidelman and Walker (2017) used a variation of the first formant frequency value, while Bouton et al. (2018) varied the second formant and Chang et al. (2010) continuously varied the place of articulation (PoA), which causes changes in the formant transitions at the beginning of the syllable.



[Alho et al. \(2016\)](#) reconciled part of these results by showing that the phoneme category selectivity (sensitivity to acoustic variations between phonetic categories) in the lower left frontal cortical areas (which are part of the speech-motor structures) with short latency (115 to 140 ms) occurs only when there is attention to the auditory task. Furthermore, they also report a broad acoustic-phonetic selectivity (with sensitivity to acoustic variations within and between phonetic categories) in areas of the lateral temporal lobe (auditory structures), regardless of attention. This result was corroborated by [Möttönen et al. \(2014\)](#) who showed that an interaction of temporal and frontal auditory structures occurs regardless of attention in higher latencies ( $> 170$  ms) while an early auditory-motor relation depends on attention and is left lateralized. Both studies used stimuli differing in the PoA. These results show that categorical coding happens in both active and passive auditory tasks, but with different latencies. So, the early latency observed by [Chang et al. \(2010\)](#) in a passive task is still an issue even if we take into account the measurement techniques used by each author.

These works show an integration of auditory and motor information for phonemic identification which is described by [Hickok and Poeppel \(2007\)](#) as a sensorimotor integration through the dorsal auditory stream in the dual-stream model of speech processing that they propose. Through this sensorimotor integration, speech sounds are mapped onto the motor representations likely to have produced them. This is corroborated by the work of [Myers \(2014\)](#), that conclude that the grained acoustic-phonetic details are processed at the superior temporal gyrus (STG)/superior temporal sulcus (STS) and then projected to prefrontal regions where the categorical codes can be consulted.

[Myers et al. \(2009\)](#) also showed the same differences between frontal and temporal regions regarding the phoneme category selectivity for a /da-/ta/ voice onset time (VOT) based continuum. VOT is a temporal acoustic cue that can be defined as the interval between the occlusion release and the voicing arising from the vibration of the vocal folds ([Lisker and Abramson, 1964](#)). It can be used to distinguish between voiced and unvoiced consonants. [Myers et al. \(2009\)](#) performed a task that required attention to the auditory stimuli but they did not analyze latencies or results to a passive auditory experimental conditions.

Based on the results of [Myers et al. \(2009\)](#), [Alho et al. \(2016\)](#), and [Möttönen et al. \(2014\)](#), concerning phoneme category selectivity (when attention is involved) one could think that VOT and formants based continua evoke similar effects in the temporal and frontal cortices. However, those different acoustic cues seem to be processed differently. [Altmann et al. \(2014a\)](#) reported that categorical effects are less pronounced in vowels than in consonants, which can be interpreted as spectral vs. temporal acoustic characteristics, respectively. The authors applied an active discrimination task and their analysis was focused on the temporal region.

Differences can be also observed regarding the laterality of this categorical processing. Several studies investigating the laterality of categorical perception report a better perception of the temporal detail of the stimulus by the left hemisphere contrasting with a better perception of spectral details by the right hemisphere (Abrams et al., 2008, Zatorre and Belin, 2001, Bouton et al., 2018, Obleser et al., 2008, Liégeois-Chauvel et al., 1999, Hickok and Poeppel, 2007). Thus, categorization of stimuli characterized by rapid transitions, such as VOT, would be more precisely performed by the left hemisphere while stimuli characterized by slow transitions, such as formant frequencies, would be better processed by the right hemisphere. Some studies suggest that speech processing occurs with different temporal resolutions in each hemisphere (Boemio et al., 2005, Zatorre and Belin, 2001, Hickok and Poeppel, 2007, Giraud and Poeppel, 2012). The right hemisphere would have a selectivity for long-term integration while the left hemisphere would be less selective, working with different temporal resolutions (Boemio et al., 2005). This leads to a second interpretation in which the left hemisphere would categorize stimuli, in general, better than the right hemisphere (Hickok and Poeppel, 2007).

These studies show that the way some regions of the cortex are involved in the categorical perception of an acoustic stimulus depends on the degree of attention to the auditory task and the acoustic cue being categorized. Understanding how different acoustic cues influence categorical perception in different auditory related cortical regions and in different hearing conditions is important to deepen the knowledge about the operation of the auditory system.

The work presented in this dissertation is placed in this specific scientific context, with the motivations and goals described below.

## 1.2 Motivation

Clicks and sine tones are commonly used in audiometric examinations that evoke the auditory brainstem response (ABR) used to investigate some basic auditory response patterns. However, such stimuli are not valid for predicting the human auditory response to speech or music. Indeed, human speech is characterized by complex sounds composed of rich harmonic structures, modulations of amplitude dynamics and fast spectrotemporal fluctuations (Skoe and Kraus, 2010). The AEP obtained from a complex stimulus, such as speech, has been increasingly used by neuroscientists to investigate hearing processes, since such responses reflect the behavior of the human auditory system to everyday sounds.

Studies on categorical perception of speech sounds use a continuum of stimuli where the acoustic cue being investigated is varied continuously between two extremes. In brain imaging studies of

the categorical perception of speech, the set of stimuli used must be reduced, mostly because, like in the case of electroencephalography (EEG), many repetitions of the same stimuli must be used in order to obtain an averaged response. The choice of stimuli is thus made a priori by the investigators (see [Bidelman et al. \(2013\)](#), [Bidelman and Walker \(2017\)](#) and [Chang et al. \(2010\)](#), for instance). In particular, some of these studies ([Bidelman et al., 2013](#), [Bidelman and Walker, 2017](#)) chose some stimuli in the middle of the continuum to represent the ones having ambiguous responses between the two extreme phonemes. However, participants vary in the position of the categorical boundary along the continuum. Hence, some stimuli that are considered ambiguous by the investigators may be perceived by some participants as being a member of a clearly perceived category. In this dissertation, we took special care in the choice of stimuli that were used in the acquisition of neurophysiological responses by designing a custom procedure based on the psychometric response of the participant. This is a novelty of this dissertation in respect to previous studies.

Another important point that motivates the present dissertation work is the use of continua that respect the natural transition between exemplars that are close to each other in the phonetic feature space. For instance, the /u/ to /a/ transition used in the continuum in the works of [Bidelman and Walker \(2017\)](#), [Bidelman \(2015\)](#), [Bidelman et al. \(2013\)](#) and [Bidelman et al. \(2014\)](#), cannot be considered as a natural transition between these two phonemes, because, for the English-speaking participants involved in those studies, there are other phonemes in the continuum, like the /ʊ/ (as in “book”), the /o/ (as in “boat”), and the /ɔ/ (as in “bought”). Therefore, in these studies cited before, even though the participants were forced to choose between phonemes /a/ and /u/, the intermediate stimuli may have undergone some kind of categorization in one of the intermediate phonemes. Hence, the processing and conclusions in those papers concerning the ambiguous stimulus may be compromised.

Regarding the detection of neurophysiological features in the auditory evoked responses that may be related to the categorization processes, many of the previous studies in the literature involved essentially the temporal characteristics of neuronal responses. In this dissertation, we aimed at introducing an innovative technique of time-frequency analysis, based on the discrete wavelet transform (DWT), which is a parsimonious way of describing the data, associated with a new regularization technique for linear regression, applied to high dimensional data with few data exemplars. This technique shed light into the time unfolding of the categorization process in different cortical areas and allowed its localization in the time-frequency domain. Specifically, we tried to dissociate the part of the AEP which is related to the physical processing of the speech sound from the part related to the psychophysical processing related to the categorical perception.

Regarding auditory attention, it can be selectively directed to a variety of acoustic features,

for instance, spatial location, auditory pitch, frequency, intensity, duration, characteristics of individual voices, etc. This suggests that there may be multiple neural loci for auditory attention processing (Tsunada and Cohen, 2014). Thus, it is important to design a controlled experiment that can focus on the acoustic feature of interest. In our case, an experiment that evaluate the effects of attention on the categorical perception of different acoustic cues: VOT and formant frequencies.

Studies of categorical perception differ in several aspects such as the acoustic cue analyzed, the brain potential measurement systems, the auditory task, the number of trials, the brain regions measured, among others. This makes it difficult to compare studies regarding the effect of attention and acoustic cue on the categorical perception. This motivated us to perform a study that aims to compare different dimensions related to categorical speech perception, with speech stimuli from Brazilian Portuguese, unifying those dimensions in one experiment, thus, reducing the results variability and making it possible to cross those dimensions in a more complete analysis.

## 1.3 Objective

The aim of this work is to investigate the neural correlates of categorical perception of human speech sounds, specifically of Brazilian Portuguese phonemes, evaluating the late auditory event related potentials in the scope of the acoustic characteristic of the stimulus (VOT or formant frequencies), brain cortical regions involved in speech perception (temporal, frontal, left, right) and the degree of attention to the task, using EEG data. In this context, we also seek to extract and analyze characteristics of AEPs which are related to physical speech features and those related to psychophysical categorical processing.

### 1.3.1 Specific objectives

- Development of an experimental protocol that provides the necessary data for the analysis of the dimensions addressed in this work, but short enough to avoid tiring the experiment participant and therefore preventing this factor from influencing the results.
- Development of a method to obtain the physical and psychophysical attributes from ERP signals to identify and analyze the time and frequency characteristics of CP.

- Development of analysis techniques in the time and time-frequency domains to determine relations between the [AEP](#) to different acoustic cues ([VOT](#) or formant frequencies) and categorical perception for the tasks and cortical areas analyzed.
- Comparison of the obtained results to those found in the literature, in order to verify similarities and differences, especially concerning the acoustic cue used, tasks (passive or active), cortical area measured and the use of Brazilian Portuguese phonemes.

## 1.4 Methodology

We choose to work with the auditory evoked late response ([AELR](#)) because it has been shown that the acoustic-phonetic transformation that occurs during categorical perception of speech stimuli lies between the [ERP](#) latencies of N1 and P2 waves ([Bidelman et al., 2013](#), [Alho et al., 2016](#), [Möttönen et al., 2014](#), [Chang et al., 2010](#), [Chevillet et al., 2013](#)). Several temporally overlapping, spatially distributed neural sources contribute to electric potentials recorded on the scalp in the latency region of those waves. Their magnitudes are influenced by different factors ([Crowley and Colrain, 2004](#)). The latencies and amplitudes of N1 and P2 waves are reported in the literature as being related to categorical coding ([Bidelman et al., 2013](#), [Bidelman and Walker, 2017](#), [Alho et al., 2016](#), [Möttönen et al., 2014](#), [Chang et al., 2010](#)), and to the perception of speech sound cues ([Picton et al., 1999](#), [Altmann et al., 2014a](#), [Manca et al., 2013](#), [Tremblay et al., 2001](#)). Furthermore, some studies suggest that the neural correlates of categorical perception emerge around N1 and are fully manifested at P2 ([Bidelman et al., 2013](#), [Bidelman and Lee, 2015](#)).

There is still controversy in the literature as to what would be the smallest unit of speech representation in the human brain. Would it be a phoneme, a diphone, a syllable or a word? To cope with this problem, monosyllabic speech stimuli differentiated by a single (Brazilian Portuguese) phoneme were used in this study. Thus, the results obtained can be compared to those of authors who performed tests with English-speaking participants, allowing us to draw conclusions about differences in categorical perception among speakers of these languages. Short speech stimuli are also important to avoid temporal superposition of event related potentials.

The investigation proposed here is performed by the analysis of the [ERP](#) collected from normal hearing participants in two different experiments. In the first one, a continuum is generated between phonemes /da/ and /ta/, differing only in a temporal characteristic, namely the [VOT](#). In the second experiment, we considered a continuum between phonemes /pa/ and /pɛ/ differing only in a spectral characteristic, namely the vowel formant frequencies. [EEG](#), which is a non-invasive

measurement technique, is used for signal acquisition.

In each experiment, passive (not involving attention to the sounds) and active (involving attention) auditory tasks were applied while measurements were performed using electrodes placed over the temporal and frontocentral areas of the brain. This should not prevent a comparison of the results with those of the literature obtained for other languages, given that the processing of different acoustic cues in the early hearing stages seems to be dependent on the physical characteristics of the syllables regardless of speech contents (Liégeois-Chauvel et al., 1999, Husain et al., 2006, Mirman et al., 2004, Holt et al., 2004) and that semantic processing is expected to occur later in the AELR, around P3 and N400 waves (Hall III, 2015). Time and time-frequency domain analyses were conducted over the AELR acquired.

For the time-domain analysis, EEG data were baseline corrected and then filtered using DWT. N1 and P2 peaks of the AELR as well as their latencies, T1 and T2, and N1-P2 difference were obtained for each participant, electrode, stimulus and experimental condition to be analyzed through a contrast analysis of mixed-effects models. For the time-frequency domain analysis, the averaged responses to each stimulus were represented as DWT coefficients.

Next, in order to extract the physical and psychophysical representations of the categorical coding we developed a technique, called regression on low-dimension spanned input space (RoLDSIS), to work with a small amount of observations in a high dimensional space. RoLDSIS does not lead to loss of dimensions and brings out more interpretable results than techniques such as least absolute shrinkage and selection operator (LASSO), ridge regression, sparse partial least squares or Elastic Net. RoLDSIS. Each AELR signal can now be represented as points in a high-dimensional space resulting from the DWT.

The main hypothesis of the present study is that it should be possible to identify separate axes in this neurophysiological space described by the DWT coefficients. These axes are aligned with physical characteristics of the stimuli or with categorical psychophysical responses. Figure 1 illustrates this hypothesis. It shows four stimuli in a phonemic continuum, in this example between the syllables /da/ and /ta/, differing by VOT. The physical values of VOT is represented in the horizontal axis. The vertical axis shows the probability of /ta/ responses in an identification task. The first (blue) and last (red) stimuli would be unambiguously identified as either /da/ or /ta/. However, the central stimuli (yellow and green), which are close to each other in terms of physical characteristics, would be represented rather distantly from each other in the psychophysical domain. Notice that we have chosen the intermediate stimuli here so that they lie on the boundaries of the transition portion of the psychometric curve. We hypothesize that the two axes of this figure correspond to axes in the DWT coefficient space onto which the projections are homomorphic either to the physical characteristics of the stimuli or to the psychophysical

response of the participant.

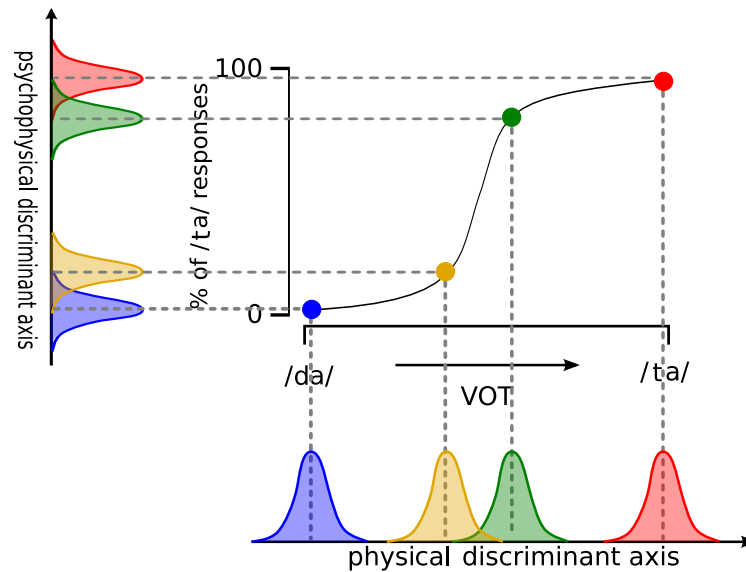


Figure 1 – Physical and psychophysical (categorical) neurophysiological axes for the /da-/ta/ continuum. The physical values of VOT is represented in the horizontal axis. In the vertical axis, is represented the probability of /ta/ responses in an identification task.

Regression coefficients were analyzed for each participant, electrode, regression response type (physical or psychophysical), continuum (VOT or formants) and type of task (active or passive) through a contrast analysis of mixed-effects models. This allowed us to draw conclusions about brain oscillations involved in different stages of the speech processing as well as the time-frame of the observed effects.

A psychometric curve was obtained for each participant and for each continuum. This curve relates the physical characteristics of the stimuli (e.g.: VOT or formant frequencies) with the way they are perceived behaviorally. With the direction vectors composed of regression coefficients for the physical and psychophysical attributes we also performed an analysis of the correlation between the angle between these vectors and the maximum slope of the psychometric curve. A steeper psychometric curve indicates stronger categorical coding (Bidelman and Walker, 2017) so that the results of this analysis can also allow us to draw conclusions about the categorical perception in each experimental condition.

## 1.5 Dissertation organization

This study involve subjects from the areas of statistics, signal processing, phonetics and neuroscience. For this reason, some sections explain basic concepts so that readers from a given area can become familiar with contents that are well known in another area.

The Chapter 2 describes auditory evoked potentials and defines the related terms used throughout the text. Chapter 3 presents a review of the literature about categorical perception focusing on the dimensions to be analyzed in this study, that is, attention, acoustic cue and cortical regions involved in the CP. Chapter 4 describes the materials and methods used for the acquisition of ERPs. Chapter 5 explains how the acquired potentials were processed from cleaning to the generation of the DWT coefficients and how N1 and P2 amplitude and latency analysis was performed. Chapter 6 presents the results of the time-domain analysis. Chapter 7 presents a novel regression technique, called RoLDSIS, with an example of its use and a comparison of the results with other regularized linear regression techniques. Chapter 8 presents the processing and the results of the time-frequency domain analysis. Finally, Chapter 9 summarizes and contextualizes the main results obtained.

Appendix A detail the short and medium latency auditory evoked potentials (not used in the analysis). Appendix B presents tests performed for choosing the best filtering technique. Appendix C shows the tests performed with other regression techniques. Appendix D contains an introduction to the discrete wavelet transform. Appendix E presents the complete results of the ANOVAs of the mixed-effects models obtained in the time-domain analysis. Appendix F contains the tests of the ANOVA assumptions. Appendix G presents a discussion about the need for averaging the epochs in ERP analysis. Appendix H contains the psychometric curves for all the 11 participants for both VOT and Formants continua. Appendix I have list of the main scripts developed during this work with a brief description of each of them<sup>1</sup>.

---

<sup>1</sup> Scripts are available on line at <https://github.com/Adrielle-Santana/ThesisScripts> and <https://github.com/RoLDSIS/code>



## Chapter 2

# AUDITORY EVOKED POTENTIALS

This chapter provides an introduction about auditory evoked potentials, specifically, the auditory evoked late response to base the reader with concepts that will be deepened throughout the text.

### 2.1 Auditory pathway

This section presents a brief description of the auditory system to base future references to some auditory structures in this work.

The sound is a mechanical wave composed of air compressions and rarefactions. When we hear a sound, this wave is directed by our outer ear to our tympanum (eardrum or tympanic membrane). This causes the vibration of this structure which is transmitted to small bones in our middle ear. Those bones (*Malleus*, *Incus* and *Stapes*), transmit the movement to the oval window (which is linked with the end of the *Stapes*) in the inner ear. This structure of the ear is illustrated in Figure 2.

The inner ear is composed by a structure named *cochlea*. The movement of the oval window is transmitted to the liquid that fills the cochlea that causes the movement of a structure there named *basilar membrane*. Only a specific region of the membrane that is influenced by the frequency of the sound will move. We call this frequency distribution along the basilar membrane of cochlear tonotopy. The structures of the inner ear are illustrated in Figure 3. The movement of the basilar membrane causes structures named *hair cells* to slide under the tectorial membrane. The hair

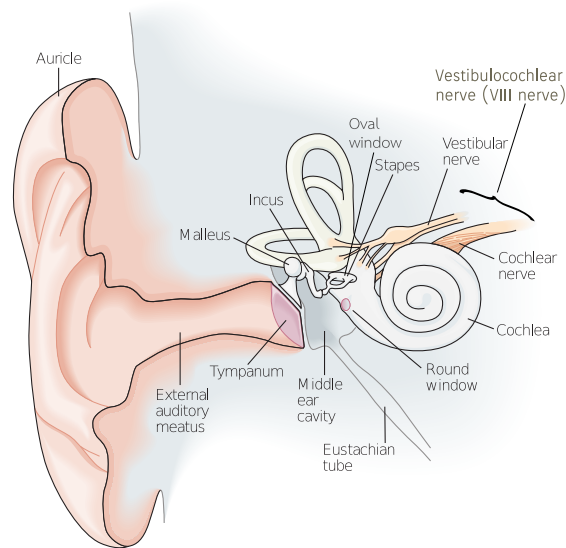


Figure 2 – Structure of the human ear.  
CREDITS: [Kandel et al. \(2012\)](#) (adapted)

cells are sensor neurons and this mechanical movement causes the opening of channels in the hair cells where an exchange of ions happens with the environment of the cochlea. This generates an ionic action potential which will travel through the auditory pathway from the cochlea to the primary auditory cortex (A1 area).

In Figure 4 is illustrated the central auditory pathway with the projections of neurons from cochlea that compose part of the cranial nerve VIII. The auditory nuclei after the cochlea is the cochlear nuclei (ventral and dorsal). It can be seen in the figure that after those nuclei the auditory pathway is divided into four main ascending parallel paths that simultaneously work by extracting different acoustic information such as location, intensity and sound temporal and spectral structures.

Axons from the dorsal cochlear nucleus make synapses with the superior olivary nucleus in both hemispheres. Other axons go directly to the contralateral inferior colliculus. Axons from the ventral and dorsal cochlear nuclei compose the intermediate acoustic stria that also goes directly to the contralateral inferior colliculus. Axons from the ventral nucleus compose the trapezoid body (or ventral acoustic stria) that goes to the ipsilateral superior olivary nuclei. Observe that the olivary nuclei receive inputs from the cochlear nuclei bilaterally. Thus, some functions of those nuclei include spatial location (inferred through the intensity and time lag that the sound reach each ear).

Axons from the superior olivary nuclei compose the lateral lemniscus. Part of them make synapses with the lateral lemniscus nucleus and part goes to the inferior colliculus. From the inferior colliculus the signal follows to the medial geniculate nucleus (NGM) in the thalamus.

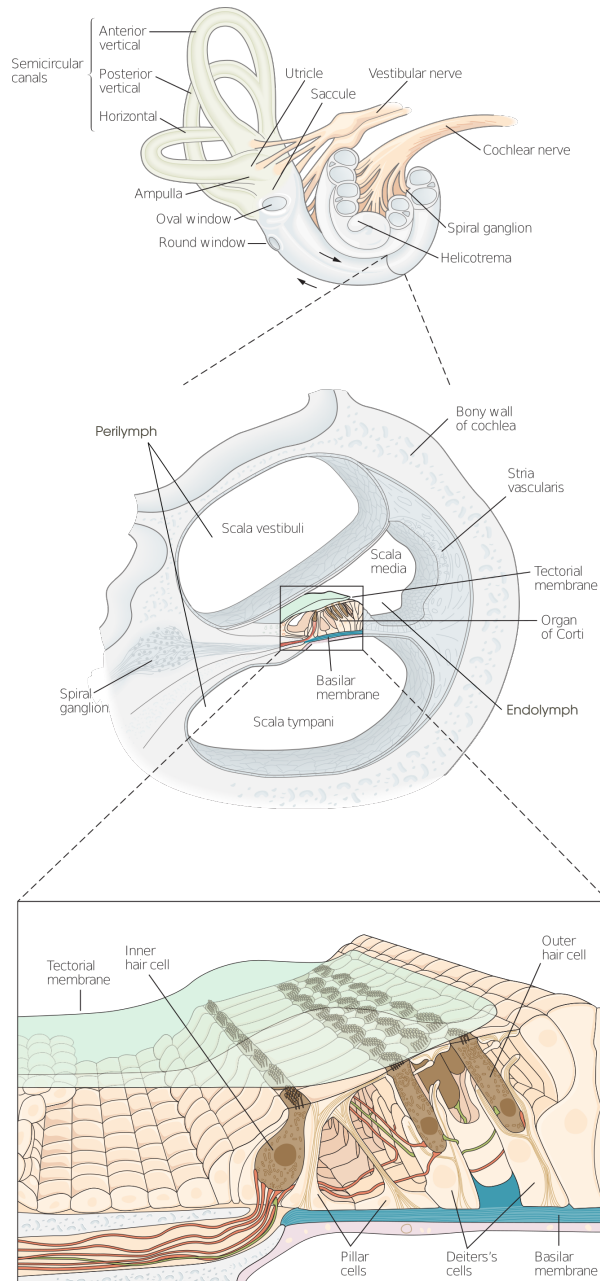


Figure 3 – Structures of the cochlea.  
CREDITS: [Kandel et al. \(2012\)](#) (adapted)

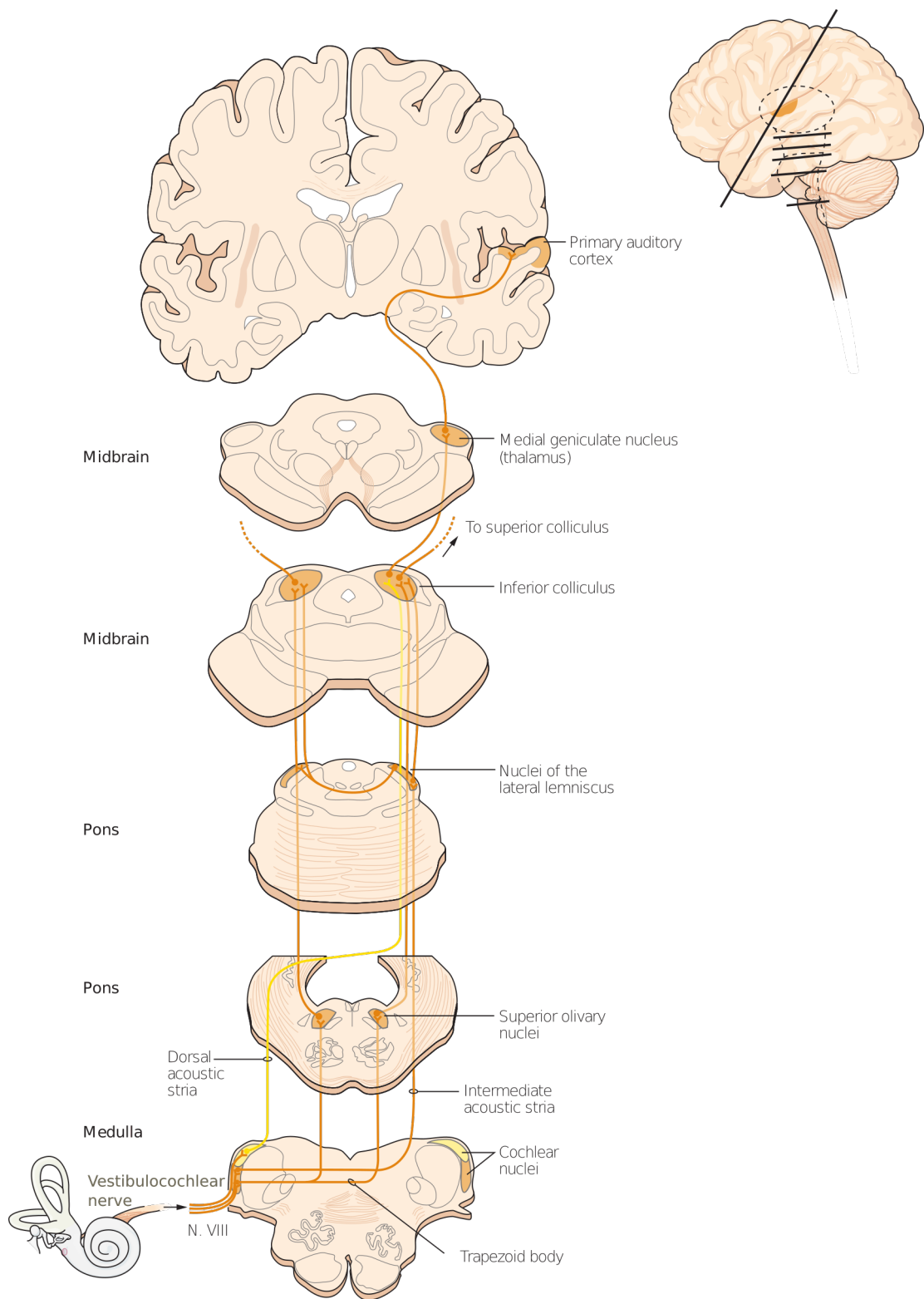


Figure 4 – Auditory pathway from the inner ear to the primary auditory cortex.  
 CREDITS: [Kandel et al. \(2012\)](#) (adapted)

Projections of NGM axons are then projected to the primary auditory cortex (A1). Observe that the cochlear nuclei receive inputs from the ipsilateral ear while all the other nuclei in the ascending auditory pathway receive inputs from both ears.

From the auditory cortex, the processing of the acoustic information is a topic still under investigation, but several works in the literature suggest models to try to explain how this processing “from sound to meaning” happens.

### 2.1.1 Speech perception models

The classical model of language organization in the brain propose separated neurobiological processes for speech perception and speech production with the first occurring in the Wernicke’s area and the latter in the Broca’s area. However, as speech present high acoustic variability, it is better addressed by models of speech perception that include the speech production system as “motor commands are likely as or nearly as variable as the acoustic signals themselves” (Skipper et al., 2017).

In the motor theory of speech perception proposed by Liberman and Mattingly (1985) the sounds provide only the information that the brain need to perceive the “gestures” and the gestures are “represented in the brain as invariant motor commands that call for movements of the articulators”. Thus, there is a perception-production link that is the requirement for the speech recognition. In the quantal theory os speech Stevens (1972) theorizes that the relation between an articulatory parameter and a acoustic parameter can be described by the model in the Figure 5. It can be seen that in regions I and III, large changes in the articulatory parameter results in little changes in the acoustic parameter while in region II, small changes in the articulatory parameter results in large changes in the acoustic parameter. Then, it is hypothesized that the anatomy and physiology of our production system assumes discrete states based on internalized perceptual representations of the acoustic parameter, that is, perceptual models of the acoustic cue modulate the motor commands for the articulators.

However, other theories supported by clinical observations show that the motor system helps but is not “required” for speech perception as the dual-stream model (relative preservation of speech perception in patients with damage to Broca’s area) or that other areas, besides Broca’s, are also involved in speech production and participates in the speech perception together with dynamic and predictive mechanisms.

The dual-stream model for the speech processing proposed by Hickok and Poeppel (2007) is

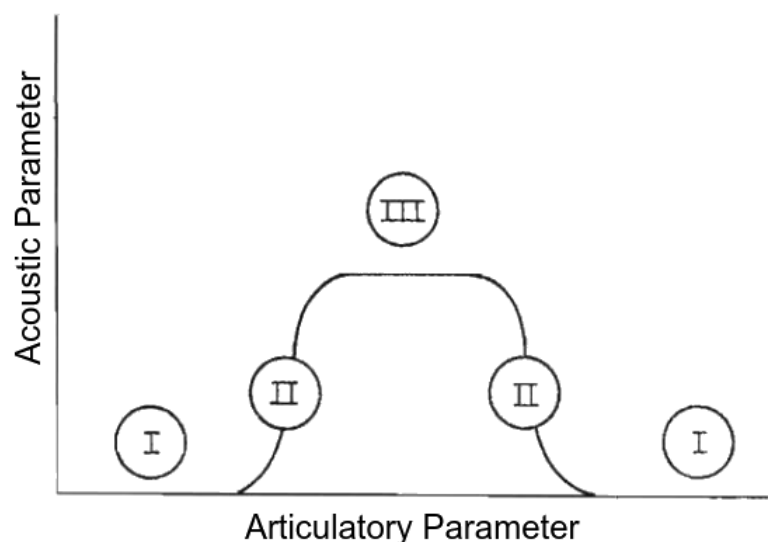


Figure 5 – Hypothetical articulatory/acoustic relation. In regions I and III, large changes in the articulatory parameter results in little changes in the acoustic parameter while in region II, small changes in the articulatory parameter results in large changes in the acoustic parameter.

CREDITS: [Stevens \(1972\)](#) (adapted)

illustrated in Figure 6. This model is very cited by the scientific community, including in works that investigate the categorical speech perception. In this model the first step of the cortical speech processing is a spectro-temporal analysis carried out at the auditory cortices (primary and secondary) bilaterally (region in green representing the dorsal surface of the superior temporal gyrus (STG)). Phonological-level processing and representation involve the posterior half of the STS bilaterally (region in yellow). Then, the system diverges in a dorsal (blue) and ventral (pink) pathway. The former would “[...] map sensory or phonological representations onto articulatory motor representations”, and the latter would “[...] map sensory or phonological representations onto lexical conceptual representations” ([Hickok and Poeppel, 2007](#)). The ventral stream include the posterior middle and inferior portions of the temporal lobes while the dorsal stream include: parieto-temporal boundary (area Spt in the Sylvian fissure); Broca’s region, anterior inferior temporal sulcus (aITS); anterior middle temporal gyrus (aMTG); posterior inferior frontal gyrus (pIFG); premotor cortex (PM).

But the dual-stream model have some problems as pointed out by [Skipper et al. \(2017\)](#). First, the speech production system includes more areas besides the posterior ventral frontal regions as, for example, the cerebellum and the basal ganglia. Second, both dorsal and ventral stream regions are possibly involved in the speech production and not only the dorsal stream. Third, dorsal stream regions also contribute to speech recognition. Fourth, the static architecture of the dual-stream model do not account for the nature of the language networks which dynamically reconfigure as a function of context.

The analysis-by-synthesis (AxS) model of speech ([Bever and Poeppel, 2010](#)) is not a neurobio-

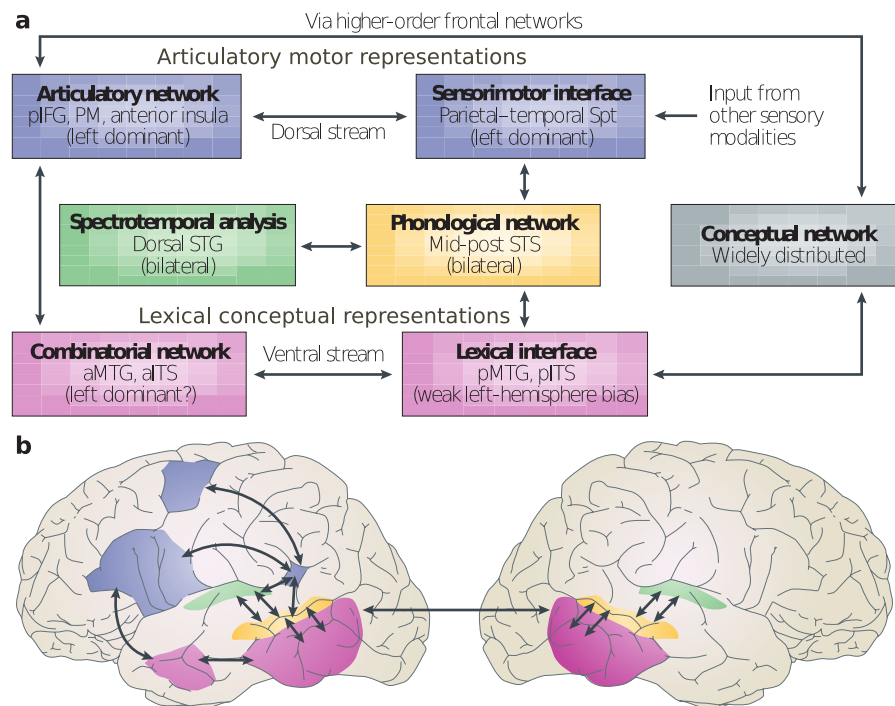


Figure 6 – Dual-stream model for the speech processing. Regions in yellow represent the posterior half of the STS involved in phonological-level processing. Regions in green represent dorsal surface of the superior temporal gyrus (STG) involved in the spectrotemporal analysis. The ventral stream, in pink, include the posterior middle and inferior portions of the temporal lobes involved in the mapping of sensory or phonological representations onto lexical conceptual representations. The dorsal stream is involved in the mapping of sensory or phonological representations onto articulatory motor representations and include the area Spt, Broca’s region, PM, aITS, pIFG and aMTG.

CREDITS: [Hickok and Poeppel \(2007\)](#) (adapted)

logical model but it accounts for the problems in the previously presented models. This model suggest that the speech perception network is more distributed in the brain and relies on motor system structures which provide “production-based constraints on the interpretation of acoustic patterns as needed” in a dynamic, constructive and predictive (dependent of context) way ([Skipper et al., 2017](#)). Finally, it is still necessary a neurobiological model for speech perception that involve a large distributed set of brain regions dynamically recruited for computations related to speech production. This model have to be able to predict when those regions will be engaged in perception what their computational role in this processing.

## 2.2 Event-related potential

An [EEG](#) signal is a measurement of ionic currents in the neurons when there are synaptic excitation in the dendrites. These currents generate an electric field in the scalp, which is seized

by EEG systems (Sanei and Chambers, 2007). In an EEG the brain signals are measured using different types of small metal plates called electrodes, which are attached over the scalp. One of the methods used to acquire an EEG signal (and the one used in this work) is the referential method. In this method, the signal of one electrode is used as reference for all the other ones. Thus, the signal of a given electrode is obtained by subtracting its measure from the one from the reference electrode. This results in specific polarities for the signal components which depend on the position of this reference electrode. This position depends on the signal one wants to measure. ERP are cortical electrophysiological responses to stimuli that can be sensory (external) or cognitive (internal). Those responses are a portion of an EEG signal which “reflects oscillatory brain activity that is time- and phase-locked to stimulus onset” (Key, 2016). See more details of the ERP acquisition and electrode positioning in sections 2.3.1 and 4.3.

An ERP is time-locked with the stimulus onset if its components, in the time domain, are lined up in time with one or more stimulus-related component(s). If a stimulus-related component interacts with ongoing or spontaneous activity in the brain in such a way that dominant brain oscillations (such as theta, alpha, beta, gamma, etc) experience a phase-change, so that they acquire a degree of phase-locking to the stimulus (phase reset), and then we say that the ERP is phase-locked with the stimulus onset (David et al., 2007, Roach and Mathalon, 2008). When the ERP is time- and phase-locked with the stimulus onset we have an evoked potential (EP), whilst if the ERP is time- but not phase-locked with the stimulus onset we have an induced potential (Key, 2016, Roach and Mathalon, 2008).

EPs can be of auditory, visual, somatosensory, etc, types. In this work, we will focus on AEPs. The waveform of an AEP is obtained through the averaging process illustrated in Figure 7. This process can be described in some steps: (i) repeat a given stimulus  $M$  times for the participant while an EEG signal is recorded; (ii) select signal windows  $y_i$  of the same length time-locked with the stimulus onset, which are called *epochs*; (iii) average all epochs together to obtain the AEP waveform to that specific stimulus (illustrated as the signal in blue in the figure). In Figure 7 the vertical lines represent an auditory stimuli, which is repeated in this case, and the rectangles in red represent the epochs that will be used to obtain the average. The time between the end of a stimulus and the beginning of the next one is the inter-stimuli interval (ISI). The latency, represented in the abscissa of the blue signal graph, is the AEP elapsed time beginning at the stimulus onset. Observe that the resulting AEP has the length of the epoch, so that choosing the correct window length is an important factor to take into account depending on the analysis intended. It is worth pointing out that each epoch is composed of the stimulus ERP as well as of the neuron spontaneous firing that will be occurring even without the stimulation.

The averaging is a technique which allows to attenuate the signal noise from the EEG signal (considering that a noise with a near Gaussian distribution tends to zero in such averaging



process) whilst it increases the parcel of the **AEP** that is time-locked with the stimulus onset. This way, the resulting averaged signal has characteristics close to those of the expected cortical response (it is assumed that the response to the stimulus is basically the same across repetitions of the stimulus). The main limitation of the **EEG** is its poor spatial resolution which hinders the link between the electrical activity recorded on the scalp with the specific brain sources. This happens because we have much more brain sources than electrodes to cover the scalp in an **EEG** acquisition.

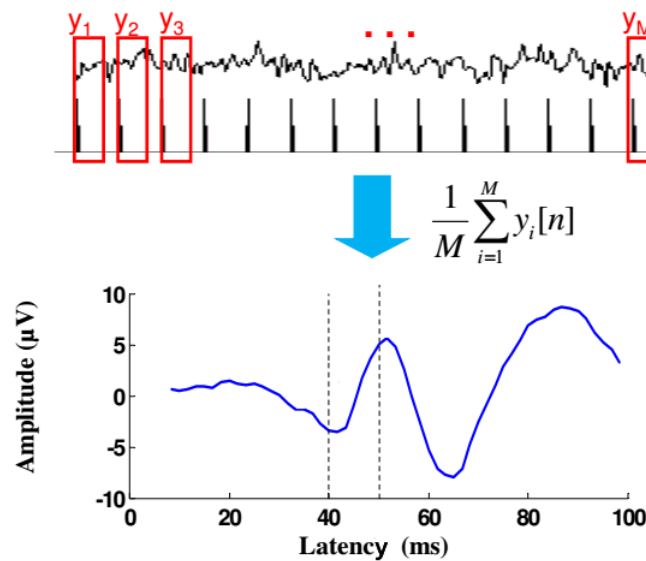


Figure 7 – Averaging.  
CREDITS: (Melges, 2013) (adapted)

Regions of an **AEP** can be classified according to its latency into:

- Short latency: below 10 milliseconds;
- Middle-latency: between 15 and 50 milliseconds;
- Long latency: between 50 and 400 milliseconds.

The waveforms of each of those regions can be named as: auditory brainstem response (**ABR**) for the short latency region signal, auditory middle-latency response (**AMLR**) for the middle-latency signal and **AE LR** for the long latency signal. Each of those signals are composed by specific components associated, in the literature, to different brain neural generators (groups of neurons or nuclei) along the auditory pathway which, consequently, are related to different stages of the auditory stimulus processing going from the perception of physical properties of the stimulus (initial part of the pathway) to its psychophysical processing (final part of the pathway).

involving cortical areas). For instance, the waves I, II and III of the [ABR](#) presumably arise from auditory pathways ipsilateral to the ear stimulated being waves I and II related to responses of the brainstem portion of the VIII cranial nerve ([Hall, 2007](#)). [ABR](#) wave V reflects activity in midbrain auditory structures ([Hall, 2007](#)). Those waves arise around 5.5 ms after stimulus onset. For the [AMLR](#) waves Pa and Pb, it is suggested that the former is the result of activity in the thalamus and primary auditory cortex (around 25 ms after stimulus onset) while the latter depends on the activity in the secondary auditory regions (50 ms after stimulus onset). For the [AELR](#) the structures involved in the generation of the waves are not totally clear in the literature, but we will cover what we could review in the literature so far in the next sections.

As we are interested in the study of the [CP](#) including its physical and psychophysical characteristics, we have to select the latency that will contain the relevant information for our analysis of the [AEP](#). Psychophysical processing seems to occur in longer latencies, but it is not discarded the possibility that can exist [CP](#) in the initial stages of the auditory processing. The “sensory gating”, perceived in the P50 (Pb) component of the [AMLR](#), indicates the brain ability to detect repeated stimuli, showing some psychophysical processing in thalamus and primary auditory cortex, relevant in the language processing. See details about the [ABR](#) and [AMLR](#) including the “sensory gating” explanation at the Appendix [A](#).

Recent works showed the emergence of the [CP](#) around 100 ms after stimulus onset (considering the use of syllables) ([Chang et al., 2010](#), [Bouton et al., 2018](#), [Alho et al., 2016](#)), that is, in the latency of the [AELR](#). Waves in this latency include the P50 component and also have higher amplitudes than those of the short and middle latency responses, improving the signal-to-noise ratio ([SNR](#)) and, thus, requiring less repetitions for their acquisition using an [EEG](#). Also, differences in the shape, amplitudes, latencies and even the emergence of some [AELR](#) waves can be observed for different syllables, training conditions and attention ([Tremblay et al., 2003a, 2001](#), [Tremblay and Kraus, 2002](#), [Bidelman and Walker, 2017](#)), showing a psychophysical processing of speech information, which does not happen for clicks used in [ABR](#) and some [AMLR](#) studies. Furthermore, due to the length of a speech () syllable (stimulus desirable in our investigation), it is not possible to use this kind of stimulus to work with [ABR](#) and [AMLR](#), because, to obtain those signals with good quality, it is required hundreds of repetitions of the stimulus for the averaging. This would make the experiment very long and tiresome for the participant, influencing in the results. This way, the aim here will be on the acquisition and analysis of [AELRs](#).

## 2.3 Auditory evoked late response - AELR

**AELR** can be used in the investigation of the representation of speech characteristics in the central auditory nervous system. Speech sounds are effective in eliciting **AELR** unlike of what happens in the acquisition of **ABR** and **AMLR**, where clicks and tones are commonly used. Several studies analyze the behavior of **AELR** for different speech signals, whether they are vowels, syllables, words, and their synthetic or natural variations, as well as their relationship to stimulus physical characteristics and cognitive processing of acoustic information (Hall, 2007, Ostroff et al., 1998, Näätänen and Picton, 1987, Kaukoranta et al., 1987, Tremblay et al., 2003a, Liégeois-Chauvel et al., 1999, Bidelman and Walker, 2017, Bidelman et al., 2013, Altmann et al., 2014b).

Figure 8 illustrates the main waves of an **AELR**. The nomenclature of the waves that make up the **AELR** was proposed by Williams et al. (1962). Peaks are labeled by polarity being the letter 'P' for positive and 'N' for negative deflections (which depends on the electrode position in the acquisition scheme, see section 4.3 for details of **AEP** acquisition). They are also labeled by latency (as in 'P200' that indicates a positive deflection at around 200 ms after the stimulus onset) or sequential number (e.g., 'P1' indicating the first positive peak after stimulus onset). Shorter latencies indicate rapid processing while larger amplitudes can indicate increased brain activity (more neural generators compounding the **ERP** peak at a certain latency).

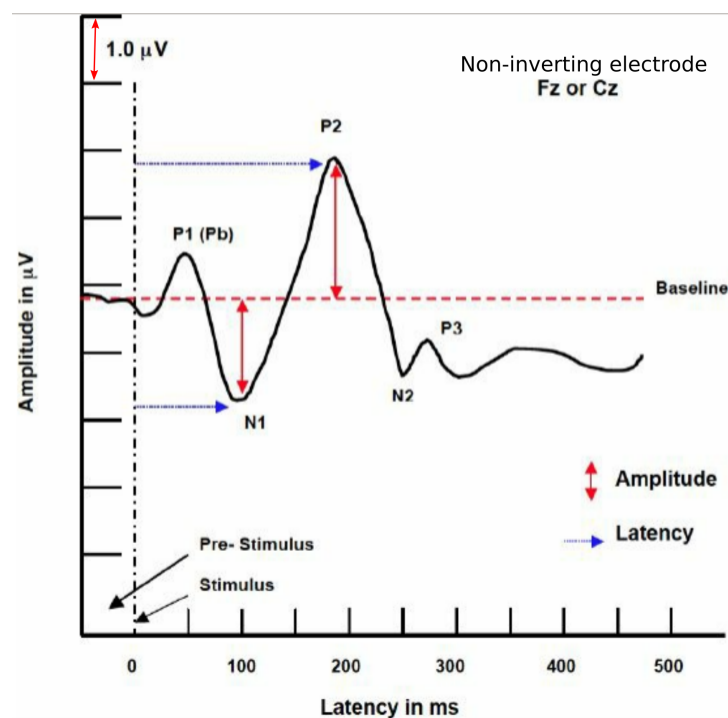


Figure 8 – AELR waveform showing major waves at typical latencies including P1, N1, P2, N2 and P3. CREDITS: Hall III (2015) (adapted)

The component P1 of the [AELR](#), known as Pb or P50, is present in the [AMLR](#) as seen in Appendix A and its latency is around 40 to 50 ms after the stimulus. The next [AELR](#) wave is the N1 trough that can occur between 75 and 150 ms after the stimulus. Such variability in its occurrence happens according to the electrode positioning and task-related variables of the acquisition. The N1 is a wave complex composed by the N1 subcomponents (N1a ( 70 ms), N1b ( 100 ms), N1c ( 140 ms)) and other negative waves occurring in the same time frame (as the Nd). The detection of such components depends on certain stimulus, task, electrode position and participant conditions. Wave P2 is a peak that occurs between 160 and 200 ms. The second trough, N2, occurs after P2 with a latency of approximately 275 ms and may be present or not in measurements, being better visualized in children.

The [AELR](#) P3 (P300) wave can occur at a latency of 250 to 400 ms and is generally obtained with the oddball paradigm where a deviant stimulus is presented randomly in the middle of sequences of repetitive stimuli. The P3b wave overlaps with the P3 in latency and it is a wave related to the attention and categorization.

The [AELR](#) N400 wave occurs at approximately 400 ms latency and is related to the semantic content of the stimulus such as the meaning of words. It is especially evident when the word has some anomaly ([Hall, 2007](#)).

The Nd is a complex wave that usually begins at a latency of about 150 ms and persists after presentation of a stimulus. However it can coincide with the earliest portion of the N1 wave depending on stimulus parameters such as the [ISI](#) or the participant attention to the stimuli. The Nd wave is not the same as the N1, but contributes to the recording of negative voltage amplitude of the N1 component within the same latency region. It is related to attention, recognition and memory. Its generators (that is, the group of neurons whose firings results in the [AELR](#) wave or wave complex measured) would be in the frontal lobe and it tends to last longer than N1 in passive listening and if choosing a long enough [ISI](#) (over 1.25 s). The presence of Nd in an [AELR](#) overlaps N1 and P2 causing an increase in N1 and a decrease in P2 ([Hall, 2007](#)). Thus, it can be said that the amplitude of P2 would be reduced with attention since this is a factor on which Nd depends.

[AELR](#) is best visualized when the acoustic stimulus has a longer duration (unlike [ABR](#) and [AMLR](#)). These include human speech stimuli as well as longer rise/fall time in tones ranging from 10 to 20 ms and also longer plateau, up to hundreds of milliseconds ([Hall \(2007\)](#) apud [Onishi and Davis \(1968\)](#), [Rothman \(1970\)](#), [Ruhm and Jansen \(1969\)](#), [Skinner and Jones \(1968\)](#)). The N1-P2 complex can be elicited either by amplitude or frequency modulation of tones or by acoustic manipulations of human speech sounds, showing a superior processing of the auditory system (that is, its high capacity to encode and integrate complex information for perception)

in detecting these variations (Hall (2007) apud Kaukoranta et al. (1987), Näätänen and Picton (1987), Ostroff et al. (1998)). However, the N1-P2 complex is larger for speech than for tones and its latency is also higher (Hall (2007) apud Čeponienė et al. (2001), Tiitinen et al. (1999)).

Speech processing can also be more lateralized than tone processing, with a greater amplitude on the left hemisphere and more symmetrical behavior in case of tones (Hall (2007) apud Szymanski et al. (1999)). Thus, given the goal proposed in this work, using Brazilian portuguese phonemes, the AELR becomes an interesting signal to be used for the analysis.

### 2.3.1 AELR acquisition

Early latency ERPs are related to the response of structures at the beginning of the auditory pathway which closely represents the patterns of the sound stimulus as in the ABR and frequency following response (FFR) (Smith et al., 1975, Krishnan, 2002). Thus, the polarity of the stimulus (that is, if it is shifted in  $180^\circ$ ) is important in those cases but not for high latency responses as in the AELR which represent the responses of structures in the end of the auditory pathway.

With an EEG signal analysis time of at least 500 ms it is possible to measure up to the N400 component of the AELR. A pre-stimulus of at least 100 ms is advisable to perform the baseline correction of the AELR. The baseline correction is generally applied to each epoch before the averaging. It consists in selecting a segment of the epoch, generally before the stimulus onset, whose mean value is defined as the new zero point of the epoch. The correction is performed by subtracting this mean value from each sample of the epoch.

To increase the quality of the AELR signal it can be filtered with a 0.1 to 30 or 100 Hz bandpass filter since the AELR is composed of low frequency waves. The 60 Hz notch filter is not recommended to avoid distortion near the frequencies of interest of the AELR signal but, as those frequencies are below 30 Hz (Savers et al., 1974), a narrow-band notch filter may not be a problem. For the averaging, 50 to 200 epochs (from repetitions of the same stimulus) are enough for a suitable AELR quality. A presentation rate of at least 1.1 stimuli/second is recommended to achieve AELR waves with good amplitudes (more visible).

The electrodes for the acquisition of AELR can be placed at different places on the scalp, but the waves N1 and P2 have greater amplitude when the non-inverting electrode is positioned at electrode Cz considering the international system 10–20. The N1 component is composed of generators located approximately in the posterior superior temporal plane and near parietal lobe regions there (Hall, 2007). Generators of the P2 component include the planum temporale and

the association cortex (area 22) (Godey et al., 2001). Recent studies using MEG showed that the N1 sources lay in the posterior part of auditory cortex, the planum temporale, whereas the center of activity for P2 lay in anterior auditory cortex, the lateral part of Heschl's gyrus (Ross and Tremblay, 2009). It is suggested to place the inverting electrode in the mastoid or earlobe on the same side of the stimulated ear or in both earlobes connecting the electrodes in this case (Hall, 2007). In this study, since the stimulation was binaural and the signal from both hemispheres was desired, the electrodes were positioned at points TP9 and TP10 and not linked together. AELR from frontocentral sites were also acquired using the signals from the electrodes at the positions FP1, F3, F7, FP2, F4, F8 and Fz. The frontocentral site is also related to the categorical perception of phonemes in the literature (Bidelman et al., 2013, Bidelman and Walker, 2017). The point Fpz was used as ground.

The complete description of the AELR acquisition procedure using the EEG technique is given in section 4.

### 2.3.2 Factors affecting latency and amplitude of AELR components

The N1 and P2 components of the AELR seem to encode different properties of the acoustic stimuli and its perception. In fact, Crowley and Colrain (2004) showed the existence of a functional independence of both N1 and P2 waves in the stimulus encoding. In what follows, we discuss some stimuli and acquisition parameters that affect the latency and amplitude of the N1 and P2 waves.

In a passive listening experiment using naturally pronounced syllables with different acoustic characteristics (/bi/, /pi/, /si/, /shi/), Tremblay et al. (2003a) showed that evoked AELRs vary across syllables but not across participants.

The VOT is a speech acoustic cue that represents the time between the release of the plosive consonant and the beginning of voicing distinguishing, for example, voiced from unvoiced consonants. It directly affects the latency of the AELR component N1 (Tremblay et al., 2003b) and the N1 and P2 magnitudes (Simos et al., 1997, Horev et al., 2007). In (Hall (2007) apud Steinschneider et al. (1999)), differences in AELR N1-P2 wave amplitudes were observed, being larger for voiced (/ba/, /ga/, and /da/) than for unvoiced (/pa/, /ka/, and /ta/) sounds. However, this result seems to be related to the VOT of each syllable. In the works of Simos et al. (1998) and Steinschneider et al. (1999) the authors worked with positive VOTs for the voiced and voiceless consonants being the short VOT for voiced and longer ones for voiceless consonants. Furthermore in Tremblay et al. (2003a), the authors showed that the voiced consonant syllable

/bi/ (VOT = 5.1 ms) evoked a greater ERP than the voiceless consonant syllable /pi/ (VOT = 65.4 ms). [Korczak and Stapells \(2010\)](#) also observed this result for the /da/ and /ta/ syllables (larger positive VOTs for /ta/).

Stimulus duration influences differences in the [AELR](#). Using tones, Davis and colleagues observed that variations in tone rise/fall time and plateau influence the [AELR](#) amplitude and latency ([Onishi and Davis, 1968](#), [Davis and Zerlin, 1966](#)). Basically short tone rise/fall times (3 ms) lead to shorter latency [AELRs](#) and their amplitude vary in this case with the duration of the tone plateau. For long tone rise/fall times (30 ms) or long tone plateau, the [AELR](#) amplitude and latency remains stable. Such time thresholds vary from person to person as well as age ([Hall \(2007\)](#) apud [Onishi and Davis \(1968\)](#)).

It has been suggested that the process generating the N1 wave an “onset detector” that may contribute to the conscious perception of sounds, rather than being directly involved in sound identification ([Näätänen, 1990](#), [Näätänen et al., 2011](#)). However, [Alain et al. \(1997\)](#) argue that the potential subcomponents of the N1 wave are probably not only onset detectors but also represent neural processes related to signal duration. The authors used tones of different durations and observed that components of N1 are influenced in different ways by changing the stimulus duration, which also led them to the perception of different scalp distributions for waves N1 and P2. Then, [Alain et al. \(1997\)](#) showed that [AEP](#) components generally increased in amplitude and decreased in latency with increasing signal duration. In an experiment with tones of different duration with fixed rise/fall time (4 ms) [Eddins and Peterson \(1999\)](#) observed that the latencies of the N1-P2 complex decrease as the stimulus duration increases.

Stimulus intensity also directly influences [AELR](#) latency and amplitude, in particular the amplitude of the N1-P2 complex measured from the N1 trough to the P2 peak ([Davis and Zerlin, 1966](#), [Onishi and Davis, 1968](#), [Picton et al., 1977](#)). In general for [AEPs](#), increasing intensity leads to increased amplitude and decreased latency. However, [Adler and Adler \(1989\)](#) showed that this enhancement is not linear and is different for N1 and P2 since such waves have different neural generators ([Crowley and Colrain, 2004](#)). The authors, observed that after a certain level of sound intensity (up to 70 dB) the amplitude and latency of N1 falls while that for P2 the amplitude continues to grow (until it saturates at each person’s limit) and only its latency falls. All these effects are influenced by stimulus duration, frequency, presentation rate, age and even gender. More details can be found at [Hall \(2007\)](#) and [Hall III \(2015\)](#).

Increases in the [ISI](#) lead to increased [AELR](#) amplitudes ([Davis et al., 1966](#), [Davis and Zerlin, 1966](#)). For example, with the use of tones in two different tasks with [ISI](#) variations, [Pereira et al. \(2014\)](#) concluded that N1 and P2 amplitudes increased in conditions with longer [ISIs](#), regardless of task. The curve illustrated in [Figure 9](#), show how the the N1-P2 magnitude increases with the

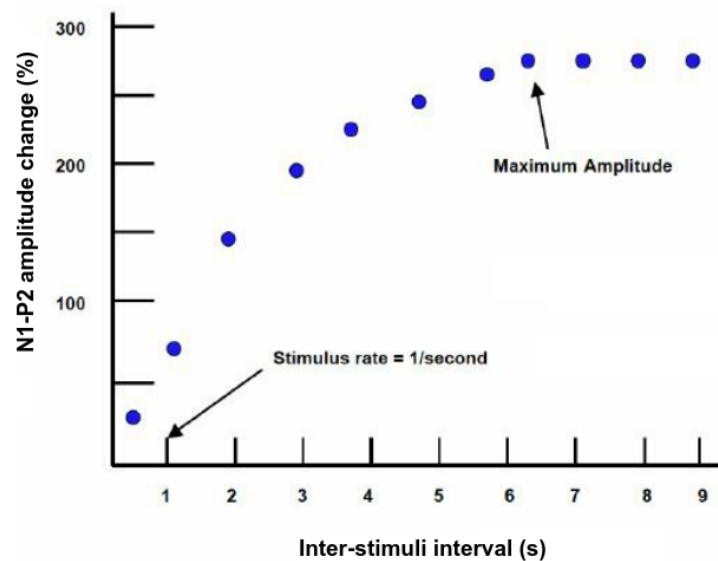


Figure 9 – Relation between inter-stimulus interval and amplitude of the N1-P2 complex.  
Hall III (2015) (adapted)

ISI (Hall III, 2015).

The stimulus onset asynchrony (SOA) is the time between the onset of one stimulus and the next (i.e., includes the stimulus duration time itself). Larger SOAs produce higher N1-P2 amplitude without affecting latency, but there is a limit after which increases in the SOA will not result in increases in the AELR amplitude. The time analysis of an AELR (see section 2.3.1) is much longer than the refractory period required for repolarization of neurons, yet there is a “refractory” period of the N1-P2 complex that takes into account other factors such as latency and memory (Umbricht et al., 2004). Thus, a shorter SOA does not respect this “refractory period” of the AELR and therefore leads to smaller amplitudes in the evoked response.

Roth et al. (1976) reported an increase in P2 amplitude and stabilization of the N1 amplitude for an increase in ISI from 0.75 to 1.5 s. The visualization of Nd is highly dependent on ISI since times shorter than 1.25 s lead to overlap of this component with N1. ISI also influences AELR in different ways depending on the intensity of the stimulus, whether the stimulation is mono or binaural and also according to some brain injury or hearing problems of the participant. Specifically for the N1 and P2 components, some studies show that the ISI had little effect on the latency of this AELR components (Hall (2007) apud Davis et al. (1966) and Hari et al. (1982) and Rothman et al. (1970)).

The repetition or not of the stimulus is also a factor that influences the composition of the AELR. The Pb wave, which is the first peak in the AELR, is an indicator of sensory gating (see Appendix A). This phenomenon is defined as the change in the amplitude of P1 when, in a sequence of two stimuli, the second is different or equal to the first. If it is different, the amplitude of P1 at



the second stimulus tends to be greater than that of the first one. If equal, the amplitude of P1 at the second stimulus is smaller. Thus, it is already possible to observe a detection of “equal” or “different” at latency of only 50 ms after the stimulus. The effect of habituation can also be seen in decreasing N1 amplitude (Hall (2007) apud Crowley and Colrain (2004)).

The N1 wave amplitude and latency is affected by the frequency of the stimulus. In Picton et al. (1978) the N1 amplitude was greater for signals with low frequency than those with higher frequency. The increase in the frequency of the stimulus also lead to a decrease in the latency of the N1 wave (Crottaz-Herbette and Ragot, 2000, Verkindt et al., 1995, Gordon et al., 2008). Bertrand et al. (1991) showed that these changes in N1 are due to changes in the orientation of the brain sources in the auditory cortex which respond to each frequency. This is related to (and an evidence of) the tonotopic organization in the auditory cortex.

Auditory training also influences P1, N1 and P2 wave amplitudes according to studies of Tremblay and Kraus (2002), Tremblay et al. (2001), Ross et al. (2013). These studies showed that training to improve VOT detection in syllables resulted in AELRs with higher N1, higher P2 and lower P1 amplitudes.

### 2.3.3 Remarks about AELR analysis

Each wave classified as an AELR can be seen as an independent auditory evoked response that differs from the others in its components, distribution of its neural generators, and sensitivity to factors such as stimulus type, auditory task, intensity, duration, and repetition rate (Hari et al., 1982, Lehtonen, 1973, Hari et al., 1979a,b, Crowley and Colrain, 2004, Hall III, 2015). The N1 wave, besides being affected by those factors, also tends to have greater amplitude (more negative) as the listener’s attention to the stimulus increases. However, it is not known whether this is an effect of N1 purely or if it is an effect of Nd wave overlap.

The amplitudes and latencies of the N1 and P2 waves are the result of the overlapping contribution of evoked potentials from multiple cortical and subcortical generators. In the work of Silva et al. (2020), subcomponents of N1 and P2 waves were obtained by using the principal component analysis (PCA) and a speech continuum with vowels varying between /i/ and /e/. The authors showed that different subcomponents of the same wave responded to different characteristics of the continuum or of the task, revealing that distinct underlying processes are at work during speech sound perception. Furthermore, Giard et al. (1994) showed by the current field analysis (fields responsible for generating the potential detected by EEG electrodes (da Silva and Rotterdam, 2012)) that neurons generators of such currents at the latency between 65 and

140 ms necessarily lie in and outside the auditory cortex. The authors showed that for the N1 wave there is a dissociation between the purely sensory component from other exogenous components generated by parallel processes in the composition of the wave after an acoustic stimulus. [Bruneau et al. \(1997\)](#) observed that the N1 wave amplitude in frontal and temporal sites changes with age. This shows differences with age in the neural generators contributing to auditory evoked potentials recorded in the N1 latency range.

The Nd wave is also known as “processing negativity” ([Näätänen and Michie, 1979](#)) and is a broad component that ultimately influences the amplitudes of other AELR waves. Its generation depends on factors such as the ISI used, memory, recognition (of speaker) and level of attention of the participant to stimulation ([Hall \(2007\)](#) apud [Hillyard et al. \(1973\)](#)). Processing negativity is also influenced by stimulus-related factors such as its relevance, its probability of occurrence, and differentiation between stimuli. Its amplitude is larger (more negative) as the stimuli is less likely to be presented, or when differentiation between stimuli is more difficult or also in discriminatory tasks ([Hall \(2007\)](#) apud [Alho et al. \(1990\)](#)).

The Nd is also generated from components that can run in short or long latencies, having different neural generators and varying with task and stimulus characteristics as mentioned before ([Hansen and Hillyard, 1980](#)). The Nd can be differentiated from N1 in three ways ([Hall \(2007\)](#) apud [Hansen and Hillyard \(1980\)](#), [Hillyard et al. \(1978\)](#), [Näätänen \(1975\)](#), [Näätänen and Michie \(1979\)](#), [Okita \(1979\)](#)):

- Its duration tends to persist beyond the N1 latency.
- By subtracting N1 of a detected stimulus from that of an undetected stimulus, the resulting wave is Nd.
- Nd is generated in the frontal lobe region while N1 is generated in the superior temporal gyrus in the auditory cortex.

The N400 wave that occurs at approximately 400 ms latency and can be elicited by paradigms that involve comparing related versus unrelated words in a sentence. In this scenario, the evoked response arises when the unexpected or semantically incongruent sentence is presented ([Hall, 2007](#)). It is therefore a response linked to stimulus semantics.

The P2 wave can take different shapes because, similarly to N1, it is composed of different generators. It can have one or multiple peaks. Some factors that influence this wave have been discussed in subsection 2.3.2. In addition, the age of the participant affects its latency as well as

the other major waves of [AELR](#) (higher for children and infants than for the elderly). The P2 wave can also be influenced by the Nd wave, which can reduce the its amplitude.

The N2 wave is influenced by the intensity of the stimulus, whether or not it is expected (its negativity is greater when an unexpected stimulus occurs), by the difficulty in differentiating stimuli (ambiguous stimuli), and by attention ([Hall, 2007](#)).

The [AELR](#) also varies with the age and sleep state of the participant. Latency, amplitude and even existence of some [AELR](#) waves, such as the N1 and P2, changes greatly with the age and just reach adult-like values around 16 to 18 years after the participant's birth ([Hall III \(2015\)](#) apud [Sharma et al. \(1997\)](#), [Ponton et al. \(2000\)](#), [Ptok et al. \(2004\)](#)). In the case of the influence of the person sleep state, wave N1 tends to have a decrease in amplitude (becomes less negative) as the participant becomes drowsy and sleepy, while wave P2 has an increase in amplitude ([Campbell and Colrain, 2002](#), [Näätänen and Picton, 1987](#)). Different explanations for this effect have been proposed, but that advocated by [Campbell and Colrain \(2002\)](#) mentions that these differences are due to decreased attention to the stimulus. This may be related to the disappearance of wave Nd which would result in the effect observed on waves N1 and P2. Furthermore, [Crowley and Colrain \(2004\)](#) observed also an interaction between the factors sleep and age in the P2 amplitude.

Amplitude of the P3b wave is affected by the ambiguity of the stimulus. For instance, [Scharenborg et al. \(2019\)](#) showed that the P3b amplitude decreased from a pretest to a posttest in an active categorization task involving ambiguous and unambiguous stimuli. The authors suggested that in the pretest, participants did not notice the difference between an ambiguous and a non-ambiguous stimulus so that they were easy to categorize resulting in a bigger amplitude for P3b even for the ambiguous stimuli. As the participants learn to perceive the difference between stimuli at the posttest, they began to perceive the ambiguous stimuli and this resulted in a reduction in the P3b amplitude.

The handedness of the participants was analyzed by [Alexander and Polich \(1997\)](#) who observed that the N1, N2 and P3 amplitudes were not affected by the handedness of the participants, while the latency was shorter for left-handed participants. Otherwise, amplitude of the P2 wave were smaller for the left-hand participants while latency did not change with the handedness.

Since [AELR](#) is sensitive to stimuli and participant factors, the waves at this latency reflect several participant-related characteristics in addition to the previously noted factors as attention and sleepiness. Thus, [AELRs](#) can be used in clinical applications not only for the analysis of the functioning of the highest level (end part of the auditory pathway) nuclei of the participant's auditory system, but also in the study of the effects of drugs and alcohol on the body, effects of

head injuries, epilepsy, schizophrenia, tinnitus, among others (see more details of those studies in [Hall \(2007\)](#), [Hall III \(2015\)](#)).

## Chapter 3

# PHONEMIC CATEGORICAL PERCEPTION

This chapter presents a literature review about categorical perception focusing on the definition and the dimensions to be analyzed in this work which include attention, acoustic cue and cortical sources of the auditory evoked potentials.

### 3.1 Categorical perception

Humans perceive the world by categorizing their sensory inputs. This has been demonstrated by studies about the perception of colors, facial expressions of emotions and also speech ([Harnad, 1987](#)). In speech, the large time and frequency variations of the sounds are mapped onto discrete perceptual linguistic representations that can be phonemes, syllables or words, in a process known as speech categorization. “The categorical perception refers to the ability to discriminate between – but not within – category differences along a stimulus continuum” ([Hary and Massaro, 1982](#)).

Human speech provides much more information to the listener than just the content of the participant being spoken. Through speech, it is possible to infer various pieces of information about the speakers, such as their location in space, gender, age, geographical region where they live and even their mood. Speech perception relies not only on the available acoustic information, but also on the visual information and the underlying linguistic context. Under

normal circumstances, speech perception is possible from the acoustic information alone, and therefore many studies in speech perception are based only on the acoustic information and on isolated syllables or words produced out of context.

In this work, phonemes are used as acoustic units. This is because phonemes i) are short enough to avoid overlaps in the evoked responses and ii) make it possible to analyze specific acoustic characteristics such as **VOT** and formant frequencies. A phoneme is defined as the smallest linguistic particle that distinguishes a word from another in a given language. A phoneme is not a single sound, but a class that includes many physically different sounds but that are not different enough to change the meaning of a given word.

Considering that the perception of a phoneme is a necessary step for the understanding of human speech, many studies involving the acquisition of auditory evoked potentials use this linguistic unit as a stimulus. However, there is controversy in the scientific community as to the definition of the smallest linguistic representation that can be categorized by the human brain. As this definition is not the focus of the present study, we decided to use monosyllabic phonemes that, in Brazilian Portuguese, represent words with known meaning:

- /pa/ ⇒ pá (shovel)
- /pɛ/ ⇒ pé (foot)
- /da/ ⇒ dá (give)
- /ta/ ⇒ tá (okay)

It is not possible to map a specific acoustic characteristic onto the identity of a phoneme since there are several sources of variability in the acoustic characteristics of sounds, such as speaker anatomy, phonetic context or presence of noise and reverberation in the environment. This makes it difficult to state that perception of a given phoneme occurs according to the presence or absence of a given characteristic (Holt and Lotto, 2010). Thus, speech acoustic stimulation is a signal composed of continuous variables which make up physically different signals, but which can be mapped onto the same class (such as a phoneme) in a process of auditory categorical perception.

Even the categorical speech perception studies suffers from some variability since the mapping of phonemes onto their categories is based on each person's perceptual limits, which change according to their experience with the language, the context, the surrounding noise, age, among other factors. This can determine which acoustic characteristics of speech are used by the listener to distinguish between two phonemes (Holt and Lotto, 2010, Nittrouer, 2004, Iverson et al.,

2003) as well as the perception of some new phonemes (a factor that often hinders the learning of a new language).

In the literature on speech perception, the terms “identification” and “categorization” sometimes are mixed. Identification refers to the differentiation between objects according to their unique characteristics so that, for example, two utterances of the syllable /ba/ can be discriminated according to their acoustic characteristics. In turn, categorization requires a certain generalization of the physical characteristics of objects, differentiating them according to the classes to which they belong, so the two /ba/ syllables of the previous example would be classified as belonging to the same category (Holt and Lotto (2010) apud Palmeri and Gauthier (2004)).

Thus, the speech perception considered in this work (with a binary identification task) comes closer to the term “categorization”, but according to Holt and Lotto (2010), this does not imply that such perception is categorical. This happens because within a category, a stimulus can be perceived as representing “better” a particular syllable than another categorized in that same class, i.e., acoustic details of the stimulus are perceived beyond the category to which the phoneme was classified and this may change the way a word is perceived.

Another factor that alters categorical speech perception is the context in which a given stimulus is presented. Thus, identical signals can be categorized into different classes depending on the other acoustic stimuli presented before or after them, the speed of articulation, the presence of nonspeech sounds between target stimulus presentations and the lexical and semantic context in which the experiment occurs (Holt and Lotto (2010) apud Holt (2005), Magnuson et al. (2003), Mann and Repp (1980), Borsky et al. (1998)).

In the view of what has been exposed, the cues for categorical perception change according to the task and the context suggesting that they are dynamically determined over the course of speech. This shows how flexible categorical perception is, explaining the ability of humans to understand speech even in the presence of noise or distortion in the harmonic or spectral envelope structure of acoustic stimuli (Davis et al., 2005). A good example is our ability to hold a phone conversation which has a limited bandwidth of 300 to 3000 Hz.

From the foregoing, it can be concluded that categorical speech perception depends on a wider range of factors than those generally captured in laboratory-controlled tasks. According to Holt and Lotto (2010), studies involving speech perception suggest that the cognitive and perceptual processes involved in categorizing speech and its dynamic perception, may not be one and the same. The present work takes into account the influence of attention as a cognitive factor that may affect phonemic categorization.

Since monosyllabic words with meaning in Brazilian Portuguese are used, a semantic processing can be present in the AEP, but longer latencies (where semantic content seems to be processed) will not be the focus of the analysis. The focus will be in the AELR before the P3 peak. This is a latency range where seems to occur a dissociation between brain responses that reflect changes in stimulus acoustic characteristics and those that indicates true internalized percepts (Bidelman et al., 2013), where an early phoneme categorization may be performed. Specifically, some studies suggested that the neural correlates of categorical perception emerge around the time-frame of N1 and are fully manifested by P2 (Bidelman et al., 2013, Bidelman and Lee, 2015).

## 3.2 Speech categorization

During a conversation, many of the sounds we hear, albeit acoustically different, can be understood as a /b/. With a small change in its acoustics (such as a step in a VOT continuum), this sound can still be categorized as a /b/ or it can change categories and be perceived as a /p/. Thus, we see that acoustic changes of the same magnitude make us perceive physically different sounds as belonging to the same phonemic class (such as the /b/ class) of the original one or shift to a different class. So, perceptually, would a listener better distinguish between sounds belonging to different phonemic classes or would that make no difference since the physical acoustic distance between classes is the same?

One of the first studies about speech categorization was performed by Liberman et al. (1957) that tried to answer that question. The authors conducted two behavioral experiments using a continuum with 14 synthetic syllables composed by the stops b, d and g followed by a steady-state part with approximately the sound of the vowel e (as in *gate*). In the first experiment, participants had to label the stimuli as belonging to the /b/, /d/ or /g/ classes. The second experiment was a discrimination ABX test. From the first experiment, the authors obtained categorization curves relating the stimuli with the percentage of participants that classified it as belonging to a given class. These curves were very similar to auditory psychometric curves which relate the stimuli (or stimuli cues) to the probability that they belong to a given phoneme class. These curves usually have a transition interval in the region separating the phoneme classes. The larger the slope of the curve in this transition, the better the phonemic categorization.

Liberman and colleagues (Liberman et al., 1957) concluded that a listener would better distinguish between sounds that fall on opposite sides of the phonemic categorization boundary than between those that fall into the same category. This study showed that speech is perceived categorically. Several studies followed trying to understand how this categorical perception



occurs in the brain, the location of the activity and what extrinsic parameters modulate this categorization (experience, attention, etc).

### 3.2.1 Experience with the categorized sound

The effect of experience with the sounds categorized were analyzed in some studies. In a mismatch negativity study (MMN), [Cheour-Luhtanen et al. \(1996\)](#) demonstrated that premature infants born from 30 to 35 weeks are capable of discriminating vowels, showing that humans learn to discriminate some sounds while still in the womb. Full-term newborns are capable of discriminating frequency tones and duration ([Čeponien et al., 2002](#)). Some effects can be observed even when they are sleeping ([Martynova et al., 2003](#)). [Eimas et al. \(1971\)](#) studied the ability of infants aged between 1 and 4 months to categorize plosive consonants /b/ and /p/. The authors observed that babies were able to perceive the acoustic variations even though they have little exposure to speech sounds and no experience with the production of such sounds.

The treatment of different acoustic cues in the early stages of the acoustic processing seems to be dependent on the physical characteristics of the syllables regardless of the speech content ([Liégeois-Chauvel et al., 1999](#), [Husain et al., 2006](#), [Mirman et al., 2004](#), [Holt et al., 2004](#)). [Liégeois-Chauvel et al. \(1999\)](#) demonstrated in a passive listening task that the processing of the VOT occurs also for non-speech sounds. This suggests that this processing depends on the physical characteristic of the syllables regardless of speech content. This result is corroborated by [Husain et al. \(2006\)](#), who concluded that it was the fast-slow temporal dimension of the sounds rather than the speech-nonspeech dimension that affected the left hemisphere lateralization of the response during the categorization task.

It is worth pointing out that the previous knowledge of the sound is important so that the stimulus representation in the cortex is more stable and its effective classification can be performed. This is evidenced in the work of [Strait et al. \(2010\)](#), where people who have an active relationship with music (played some musical instrument) showed a better ability to detect and code the sound patterns. Thus, the better the pattern identification by the nervous system, the more stable is the stimulus representation in the cortex and, consequently, more resistant will be the degradation of this information by a noisy communication channel/environment ([Anderson and Kraus, 2011](#)).

Depending on the language, a stimuli with, for example, a positive VOT of up to 20 ms can be perceived as a /ba/ ([Rufener et al., 2019](#)) or as a /ta/ (as in this dissertation). Thus, the perception of lead (negative VOT), short or long-lag voicing (positive VOT) in a phoneme varies with the native language ([Holt et al., 2004](#)). On the other hand, [Tremblay et al. \(2001\)](#), [Ross et al. \(2013\)](#)

showed that the perception of this time characteristic can be learned by training. Related to this, the work of [Duncan \(2019\)](#) showed that bilinguals have a double phonemic boundary which changes according to the language they speak. The bilingual speakers also presented a better categorical perception than monolinguals.

The sensitivity to temporal differences in the acoustic stimuli is an important aspect of categorical speech perception. Intracranial recordings from primary and secondary auditory cortices showed that syllables with a larger VOT (40 to 80 ms) evoked components time-locked to the consonant release and voicing onset while syllables with shorter VOT (20 ms) evoked a diminished AEP to voicing onset and that was morphologically similar to the AEP to a syllable with a 0 ms VOT ([Steinschneider et al., 1999](#)). The temporal distribution of such difference matched the behavioral sound discrimination data, indicating a categorical shift in the consonant identification from voiced to unvoiced stop consonants as the VOT changed. The N1 peak seems to encode this sensitivity to the VOT as demonstrated in the work of [Simos et al. \(1997\)](#).

Categorical perception of temporal acoustic cues can also be improved artificially. For instance, in the study of [Rufener et al. \(2019\)](#), the temporal precision of the auditory system in people with developmental dyslexia was improved through transcranial alternating current stimulation (tACS) and transcranial random noise stimulation (tRNS). This is related to the result of studies of nonspeech sounds categorization where the participants learn to categorize sounds they do not know previously, producing categorization functions (psychometric curves) similar to those produced with speech sounds, for VOT and formant frequency variations on the stimuli ([Mirman et al., 2004](#), [Holt et al., 2004](#)).

### 3.2.2 Laterality and brain auditory processing regions

Several studies investigating the laterality of categorical perception report a better perception of temporal details of the stimulus by the left hemisphere contrasting with a better perception of spectral details by the right hemisphere ([Abrams et al., 2008](#), [Zatorre and Belin, 2001](#), [Bouton et al., 2018](#), [Obleser et al., 2008](#), [Liégeois-Chauvel et al., 1999](#)). Thus, categorization of stimuli characterized by rapid transitions, such as VOT, would be more precisely done by the left hemisphere while stimuli characterized by slow transitions, such as their formant frequencies, would be better processed by the right hemisphere. However, some studies suggest that processing occurs with different temporal resolutions in each hemisphere. The right hemisphere would have a selectivity for long-term integration while the left hemisphere would be less selective, working with different temporal resolutions ([Boemio et al., 2005](#)). This leads to a second interpretation in which the left hemisphere would categorize stimuli, in general, better than the right hemisphere

(Hickok and Poeppel, 2007).

Studies have shown that the right pSTG, the left STG, and the STS in the auditory cortex are involved in the phonemic processing of speech streams (Bouton et al., 2018, Chang et al., 2010, Zatorre and Belin, 2001, Altmann et al., 2014b). Chang et al. (2010) demonstrated the neural correlates of phonemic categorization in the left STG around 110 ms after the stimulus onset using stimuli with spectral variations. Altmann et al. (2014b) reported that categorical effects are less pronounced in vowels than in consonants, which can be interpreted as spectral vs. temporal acoustic characteristics respectively. The authors observed categorical activity in the left STG and STS and posterior Heschl's gyrus. Zatorre and Belin (2001) and Bouton et al. (2018) showed that this categorization is lateralized with spectral variations being perceived by the right pSTG and temporal variations by the left STG and STS.

Zatorre and Belin (2001) worked with pure tones with variations in their temporal and frequency (spectral) presentation. The authors concluded that both hemispheres are sensitive to temporal and spectral differences, but the major difference is in temporal resolution that appears to be greater in the left and spectral in the right hemisphere. Obleser et al. (2008) confirms the processing of the temporal detail of sounds by the left hemisphere and the spectral detail by the right hemisphere in an active task where words had to be classified as comprehensible or not. In a study using sentences presented to children, Abrams et al. (2008) concluded that rapid transitions in the frequency (as in the PoA) and in time (as in the VOT) would be better represented in the left hemisphere. Slow temporal characteristics of speech (as the formant envelope) would be better represented in the right hemisphere. The authors reported that the right hemisphere followed the stimuli envelope with 100% accuracy and had a response amplitude 33% greater when compared to the left hemisphere. Thus, right hemisphere would have a selectivity for long-term integration, while the left hemisphere would be less selective, working at different temporal resolutions (Boemio et al., 2005). This suggests that, in general, the left hemisphere would categorize stimuli better than the right one (Hickok and Poeppel, 2007).

Hickok and Poeppel (2007) relates this fast and slow processing with the brain oscillations. The authors reported that fast information is processed bilaterally (gamma waves, 30–100 Hz) and slow information at the right hemisphere (theta waves, 4–8 Hz) with the phonological lexical information being represented bilaterally. Bouton et al. (2018) also reported gamma waves in the left hemisphere for coding VOT, specifically in the left STG and STS. They also reported beta waves (13–30 Hz) encoding the formant frequency information (F2 variation) in the right hemisphere pSTG. Bidelman (2014) related these brain oscillations with phoneme ambiguity. The authors noticed increased gamma wave amplitude for ambiguous sounds and larger beta wave amplitudes for clearer non-ambiguous phonemes. They reported that beta waves may encode the degree of relationship between speech and the internal representation of the phoneme.

Hickok and Poeppel (2007) mention the existence of a bilateral ventral cortical stream (upper and medial temporal lobe) that would be responsible for speech processing and comprehension. One of the reasons that led them to conclude about the bilaterality is that injury to the left hemisphere does not lead to large deficits in speech recognition, although computational differences exist between hemispheres (brain computation). The authors also report a dorsal stream (superior temporal gyrus, anterior inferior temporal sulcus, anterior middle temporal gyrus, posterior inferior frontal gyrus and premotor cortex) dominant in the left hemisphere, which would be involved in the auditory-motor (sensorimotor) integration, participating in the speech perception process. This is hypothesized due to the fact that problems of speech production follow injury in the temporal and frontal dorsal regions. Hickok and Poeppel (2007) concluded that both streams share the left STG.

Giraud and Poeppel (2012) theorize about oscillation-based operations for speech perception, particularly the delta-theta and gamma oscillations. Thus, speech would be analyzed in parallel (at the left auditory cortex) in two different time scales: at a syllabic (slow) and a phonemic (fast) rate, through theta-gamma nesting. The authors report a frequency ratio of 4 in this nesting where “[...] about 4 cycles of the higher frequency occur during one cycle of the lower one”. According to Zatorre et al. (2002) gamma activity dominates the left auditory cortex for fast computations whilst theta activity dominates the right auditory cortex. These asymmetric oscillatory properties would be related to cytoarchitectonic differences related to the pyramidal cells (Giraud and Poeppel, 2012). This analysis at slow rates at the right auditory cortex allows a better analysis of steady-state speech signals such as vowels. That is also important for paralinguistic processes such as speaker identification. However, analysis of vowels at fast rates (gamma) by the left auditory cortex is sufficient for vowel identification in speech processing (Giraud and Poeppel, 2012). Also, in regions of the motor cortex involved in tongue, lip and hand control, the theta and gamma activity seems to be left dominant (Morillon et al., 2010).

The theta oscillations, besides participating in the spectrotemporal analysis as reported in the model presented by Hickok and Poeppel (2007), is also reported in the literature as being correlated with speech clarity (Etard and Reichenbach, 2019), increased cognitive demand (INANAGA, 1998) and auditory attention (Viswanathan et al., 2019). Stronger theta band activity was also observed in the connectivity of frontal structures and auditory cortices during intelligible speech but not during unintelligible speech, showing a top-down modulation (Park et al., 2015). This possibly works by facilitating target processing by enhancing the neural response to the speech signal. However, Strauß et al. (2014) showed that theta power was highest for the ambiguous (i.e. the most difficult) case. On the other hand, the alpha band brain activity is usually suppressed with increasing attention and cognitive demands but it is also linked to speech perception. Weisz and colleagues (Weisz et al., 2011) observed that challenges to the auditory system arising from signal degradation trigger increased alpha power. Foxe et al. (1998)

reported an enhancement in alpha activity related to selective attention to auditory cues, which is in line with a role of enhanced alpha oscillations in inhibiting task-irrelevant information (such as unintelligible speech). This was observed during early stimulus encoding around 100-200 ms after stimulus onset [Weisz et al. \(2011\)](#). [Luo et al. \(2005\)](#) observed enhanced alpha activity during categorization of non-speech sounds compared to speech sounds. The authors also observed more alpha activity dominating the auditory cortex at the temporal region and beta activity in the frontal cortex.

There is also stronger activity in the gamma band when the speech stream is attended to compared to when it is ignored ([Viswanathan et al., 2019](#)). It is also involved in the integration of top-down (frontal-temporal) and bottom-up (temporal-frontal) processes during local synchronization in sensory and motor cortices ([Bidelman, 2017](#), [Giraud and Poeppel, 2012](#), [Bouton et al., 2018](#)).

### 3.2.3 Attention and cortical region

Attention affects the way the auditory processing takes place in different cortical regions. Studies show that attention works as a top-down process of modulation of categorical perception mechanisms in the auditory system. Attention is known to enhance neuronal selectivity and responsiveness ([Spitzer et al., 1988](#)).

Auditory attention can be selectively directed to a rich variety of acoustic features including spatial location, auditory pitch, frequency or intensity, tone duration, timbre, speech versus nonspeech streams, and characteristics of individual voices. Given the multiplicity of acoustic dimensions to which we can attend and the richly interconnected auditory processing networks, there are likely to be multiple neural loci for auditory attention. [Tsunada and Cohen \(2014\)](#) showed that neurons differentially compute categorical information along the ascending auditory system in the ventral pathway, in different local microcircuits.

Studies report that attention tunes the responses to task-relevant feature values, acting as a band-pass filter which dynamically reallocates cortical resources depending upon task demands and underlines the flexibility in auditory processing ([Fritz et al., 2007](#), [Schröger et al., 2015](#)). In an experiment of selective attention with rhesus macaques using acoustic stimuli that varied along spectral and temporal feature dimensions, [Downer et al. \(2020\)](#) showed that ensembles of neurons at the primary auditory cortex (A1) exhibit enhanced encoding of attended sound features (which was not observed at the single neuron level) and suppressed the distractor feature. In a previous study from the same authors, they explain that the effect of attention in perception also depends on the tuning to the distractor feature as well on a mechanism by which

A1 simultaneously enhances relevant information and suppresses irrelevant information (Downer et al., 2017).

Atiani et al. (2014) showed that the task modulations that occurs at A1 also occurs in the cortical belt areas (dorsal posterior ectosylvian gyrus - dPEG) with much larger magnitude which seems to be driven largely by a selective increase in dPEG response firing rates to target tones during behavior (study performed in ferrets in an auditory discrimination task). Together, Downer et al. (2017, 2020) and Atiani et al. (2014) works suggest a mechanism in which distractor suppression and selective target enhancement are controlled by top-down circuits that could gradually extract behaviorally relevant sensory features through a hierarchy of brain areas (Ahissar et al., 2009).

According to Bidelman et al. (2013), the AELRs in the N1 wave latency during passive listening reflect the physical properties of the speech vowels they used in their experiment, but not the categorical information. In this same study the authors have found categorical effects on the time-range of the P2 wave (150–200 ms) during an active labeling task, i.e., with attention to the auditory stimulus. In this experiment, measurements were performed in the temporal cortex and it was used a synthetic continuum between /u/ and /a/ created with the variation of the first formant frequency (F1). In another study using this same continuum, Bidelman and Walker (2017) measured the event-related potentials (ERP) with electrodes on the frontocentral area where the authors report that categorical effects in the auditory ERPs were most prominent. The authors performed both a passive listening task and an active categorization task and noted that the neural coding of the categorical effect requires attention in the auditory task, by analyzing the N1-P2 wave complex and its relation with the ambiguity of the stimulus token. This result contrasts with that found by Chang et al. (2010), who observed that there is a passive categorization within the first 110 ms after the stimulus using a /ba/-/da/-/ga/ continuum with measurements performed at the pSTG.

In these three studies, the continuum consisted of spectral variations of the stimuli. The first and second ones used a variation of the first formant frequency value while the third study varied in the PoA, which is characterized by changes in the format transitions at the beginning of the syllable. However, it is worth questioning the perceptual effects of a transition between the vowels /u/ and /a/ which is not normal even in English (language of the participants of those two studies which used this continuum).

Alho et al. (2016) reconcile those results by showing that the phoneme category selectivity (sensitivity to acoustic variations between phonetic categories) in the left lower frontal cortical areas (which are part of the speech-motor structures) with short latency (115 to 140 ms) occurs only when there is attention to the auditory task. Furthermore, they also report a broad acoustic-phonetic selectivity (with sensitivity to acoustic variations within and between phonetic

categories) in areas of the lateral temporal lobe (auditory structures), regardless of attention. This result is corroborated by Möttönen et al. (2014) who showed that an interaction of temporal and frontal auditory structures occurs regardless of attention in higher latencies ( $> 170$  ms) while an early auditory-motor integration are dependent on attention and is also left lateralized. The authors showed the relevance of this integration for speech perception once the disruption at the articulatory motor cortex (through transcranial magnetic stimulation) affects the sensory processing of speech. The work of Chevillet et al. (2013) also showed that the sensorimotor integration occurs regardless of attention at the P2 wave latency.

Both Möttönen et al. (2014) and Chevillet et al. (2013) used stimuli differing in the PoA in their studies. Their results show an integration of auditory and motor information for early phonemic categorization process through the dorsal auditory stream when there is attention, as described by Hickok and Poeppel (2007). Through this sensorimotor integration, speech sounds are mapped onto the motor representations likely to have produced them. Möttönen et al. (2014) also reported that this integration is relevant for speech production where the dorsal stream is involved in the prediction of sensory consequences of articulatory movements. Thus, the auditory speech signals are transformed to motor models, which in turn affect sensory processing.

Myers et al. (2009) showed the same differences reported by Alho et al. (2016) between frontal and temporal cortices in regard to the phoneme category selectivity for a /da/-/ta/ VOT based continuum. The authors performed a task that required attention to the auditory stimuli but they did not analyze latencies or results to a passive auditory condition.

### 3.2.4 Final considerations

The studies discussed in section 3.2.3 suggest that the way some regions of the cortex are involved in the categorical perception of an acoustic stimulus depends on the degree of attention to the auditory task. However, how this varies with the acoustic cue is less clear. Understanding how different acoustic cues influence categorical perception in different auditory related regions and in different hearing conditions (active or passive) is important to deepen the knowledge about the operation of the auditory system. Thus, it would be possible, for example, to improve clinical methods of assessment and treatment of hearing disorders related to the perception of different acoustic characteristics, as temporal and spectral ones, which compound the speech sounds.

In view of what has been exposed here, the investigation of the categorical perception of speech sounds can be performed focusing on different dimensions, such as: categorization in active or

passive listening (Bidelman and Walker, 2017, Alho et al., 2014, Möttönen et al., 2014), cortical laterality of the speech processing (Liégeois-Chauvel et al., 1999, Hickok and Poeppel, 2007, Altmann et al., 2014b), brain oscillations and time frame involved in the perception of speech stimuli (Bidelman, 2015, Bouton et al., 2018, Giraud and Poeppel, 2012, Chang et al., 2010) and neuronal processing of spectral and temporal characteristics of speech (Altmann et al., 2014b, Zatorre and Belin, 2001, Obleser et al., 2008). Thus, we aim to investigate the neural correlates of categorical perception of human speech sounds by evaluating the auditory evoked late responses for variations in the temporal (VOT) and spectral (formant frequencies) acoustic characteristics of the stimuli, taking into account the degree of attention to the auditory task and the cortical regions.



## Chapter 4

# MATERIALS AND METHODS

In this chapter the materials and methods for the acquisition of ERPs are detailed including concepts about [EEG](#) acquisition, measurement system used, stimuli continua and the acquisition protocol.

### 4.1 Brain potentials

A variety of techniques exist for brain activity measurement such as [EEG](#), electrocorticography ([ECoG](#)), magnetoencephalography ([MEG](#)), positron emission tomography ([PET](#)) and functional magnetic resonance imaging ([fMRI](#)). However, [MEG](#), [PET](#) and [fMRI](#) are expensive techniques that also demand special facilities if compared to the [EEG](#) or [ECoG](#). Also, for the [PET](#) and the [fMRI](#) techniques, measurements depend on the participant's metabolic processes and, consequently, are associated with large time constants and therefore have low temporal resolution ([Schalk and Mellinger, 2010](#)).

Systems based on [EEG](#) and [ECoG](#) are simpler and cheaper and offer higher time resolution. The main disadvantages of [EEG](#) are its low spatial resolution and its susceptibility to several artifacts. [ECoG](#) based systems, although less prone to artifacts ([Schalk and Mellinger, 2010](#)), require an invasive procedure for electrode placing. In this work, [EEG](#) was the method used for the acquisition of brain signals.

The pyramidal cells present in the cortex communicate with each other through synapses. In

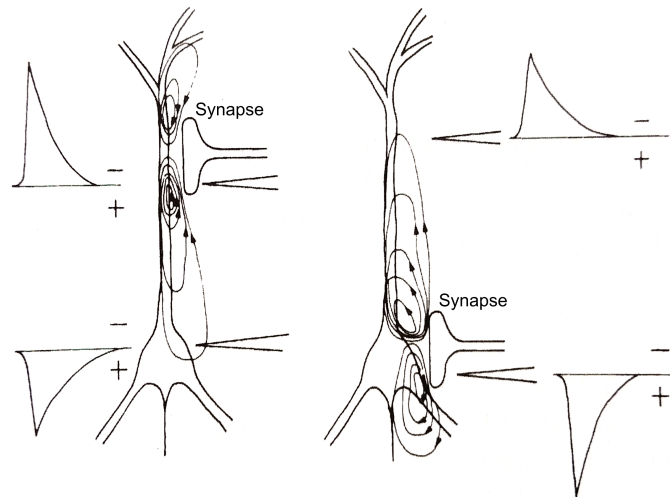


Figure 10 – Electric field pattern in response to EPSP and IPSP.  
 CREDITS: (da Silva and Rotterdam, 2012) (adapted)

axodendritic synapses, there is a release of neurotransmitters in the synaptic cleft which results in a polarity change in the synapse area. Thus, because there is a predominance of one electric charge (a graded potential) in the synapse area, with the consequent accumulation of the opposite charge in another part of the neuron body, an electric dipole is formed in the neuron. The small electric field from this dipole is captured by EEG, but, are necessary several of those dipoles in several near neurons, organized approximately parallel and occurring in a synchronized way to be possible to obtain an electric field strong enough for noticeable EEG record.

Figure 10 illustrates this process in the case of an incoming Excitatory Post Synaptic Potential (EPSP) (left) and an Inhibitory Post Synaptic Potential (IPSP) (right). The dipole, created by either an EPSP or an IPSP, generates an ion flow both inside and outside the neuron. It has been shown that the electric field generated by this ion flow inside the neuron is the main contributor to the signal measured by EEG (da Silva and Rotterdam, 2012). Glial cells also seem to contribute in part of the signals measured. It was suggested that electrical activity of neurons are likely to depolarize glial cells in distributed brain areas so that these cells may play a role in coupling brain regions (Turbes, 1996). In a study using human astrocytes (one kind of glial cell) in mice, showed that these cells enhanced both activity-dependent plasticity and learning in the animals making the synapses more efficient and complex (Han et al., 2013).

The brain potentials measured by EEG are attenuated by several tissue layers on their way from their generation in the neuron until the EEG electrode on the scalp. Figure 11 illustrates these layers. Besides the attenuation, the tissues also result in brain potential dispersion since they present different physical properties depending on the region measured. It is possible to conclude that the EEG technique in addition to being quite susceptible to artifacts, is used to measure signals that reflect a sum of graded potentials generated by a set of neurons that can be oriented

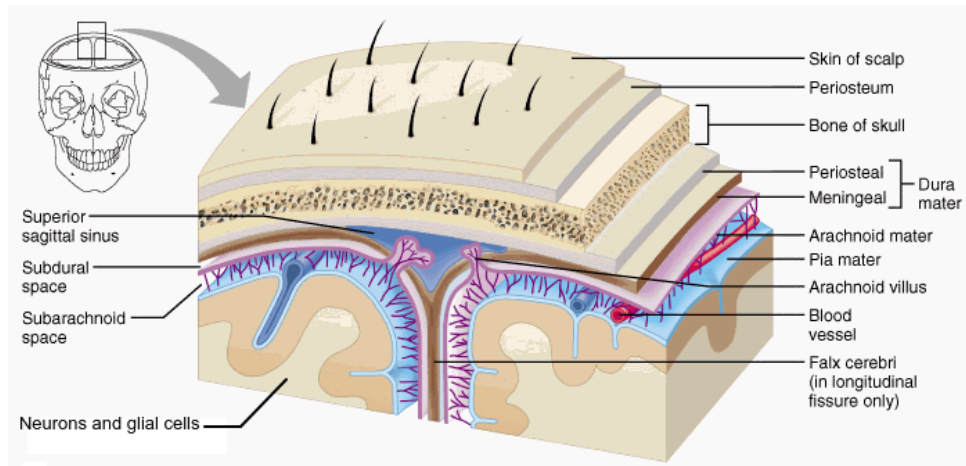


Figure 11 – Meninges.

CREDITS: (Cummings, 2001) (adapted)

in a way so that the resulting potential may be pointing to a place on the scalp where there is no electrode or where is not possible to place one.

EEG measurement systems date back to the end of the 20th century and continue to evolve. Current systems are equipped with signal processing tools, sensitive and accurate electrodes and enough memory for long acquisitions. The EEG enabled the development of clinical, experimental and computational studies which have contributed to the discovery, recognition, diagnosis and treatment of a large amount of physical and neurological disorders in the human brain (Sanei and Chambers, 2007).

## 4.2 Acquisition scheme

The experiment consisted of three main auditory tasks, one task involving the identification of phonemes to obtain behavioral results and two tasks performed with simultaneous EEG recording: an active forced choice task involving the identification of phonemes and a passive task where the participants watched muted videos while the acoustic stimuli were played. A trigger signal is also recorded and it is used to synchronize the beginning of each stimulus heard by the participant with that of the EEG signal. More details of the acquisition procedure are given in Section 4.5.

Signal acquisition was performed inside a sound-proof booth at the Centro de Estudos da Fala, Acústica, Linguagem e Música (CEFALA) laboratory at UFMG. This room is air-conditioned and illuminated with LED lamps so as not to generate electromagnetic interference in the measured signals. Figure 12 shows the layout of the room where the acquisitions took place and



Figure 12 – Room layout for the acquisition protocol adopted. (a) Computer for signals acquisition. (b) Computer for stimuli generation and responses reception in all stages of the protocol. (c) Monitor for participant interaction. (d) Computer to display videos in the passive task. (e) Cable for audio channels separation. (f) Battery to power the RHD2000. (g) RHD2000 evaluation system. (h) RHD2000 amplifier, electrodes connector and optocoupler circuit. (i) Electrodes. (j) Chair for the participant. (k) Chair for the researcher. (l) Keyboard for participant interaction.

the equipment used. In total, three computers were used, two notebooks and one Apple MAC Mini. One of the notebooks is a Sony Vaio VPCSB Core i5 with 8 GB of RAM, and was used to run all scripts related to the experiment tasks, generating stimuli and receiving responses via external keyboard when needed. It runs a Linux operating system (Mint 19 Cinnamon). The second notebook is a DELL Vostro 5470 Core i7 with 8 GB of RAM, which was used exclusively for connecting to the RHD2000 interface board for configuring and recording of the EEG signals. In order to acquire the signals, it was necessary to run Intan Technologies' own software which worked on the Windows 8.1 operating system and therefore was the system used on that computer for the acquisitions. This computer was also used later in signal processing, but now running the Ubuntu 18.04.2 LTS operating system. The MAC Mini was only used to display the videos in the passive stage of the experiment.

All the written words and symbols in the participant's visual field were covered including the keyboard itself avoiding eye contact with any visual stimuli that could evoke signs related to language processing.

## 4.3 Acquisition

### 4.3.1 Electrodes

The ions flow inside and outside the neuron generate an electric field in the scalp, which is measured using the EEG technique (Sanei and Chambers, 2007). The EEG signals are detected by small metal plates placed on the scalp. These plates, which can be of different types, are called electrodes and are usually attached to the scalp by means of a conductive gel made specifically for this end. The electrodes can be classified as disposable (with or without gel) or reusable, metal (gold, silver, brass or stainless steel) or saline (Ag-AgCl), surface (applied to the skin) or depth (nail type applied in the brain) and active (with the amplifier circuit in the body of the electrode) or passive (amplifier circuit away from the electrode).

Distortions in the EEG signal can happen if the skin-electrode impedance is not controlled. This impedance is a result of the electrode material, the area of the surface contact (skin, muscle, mucosa) and any material in between (oil, dirt, fluids, etc.) (ASHA, 1987). Some EEG acquisition systems have impedance monitors that facilitate the verification of this variable at each acquisition. It is recommended that the skin-electrode impedance be lower than 5 k $\Omega$  and do not vary more than 1 k $\Omega$  between electrodes (Sanei and Chambers, 2007).

The naming of the electrodes and their placing on the scalp during an EEG session are defined by the International 10–20 system (Jasper, 1958). This system is based on the division of the scalp in arcs, using some scalp points as reference: the nasion (Ns), inion (In) and pre-auricular points (A1 and A2) (see Figure 13) (Schalk and Mellinger, 2010). The intersection of the longitudinal (Ns-In) and lateral (A1-A2) lines defines a point named Vertex (Cz) (Schalk and Mellinger, 2010).

Using the points Ns, In, A1, A2 and Cz as anatomic reference, the electrodes are positioned with distances corresponding to 10% or 20% of the total distance between two reference points as illustrated in Figure 13 (a) and (b). Based on the 10-20 pattern shown in figures 6a and 6b, the American Electroencephalographic Society defined the scheme illustrated in Figure 13 (c), to be used when a greater amount of electrodes (75) is required (Nomenclature, 1991).

The naming of the points illustrated in Figure 13 (c) derives from the names of scalp regions where these points are located:

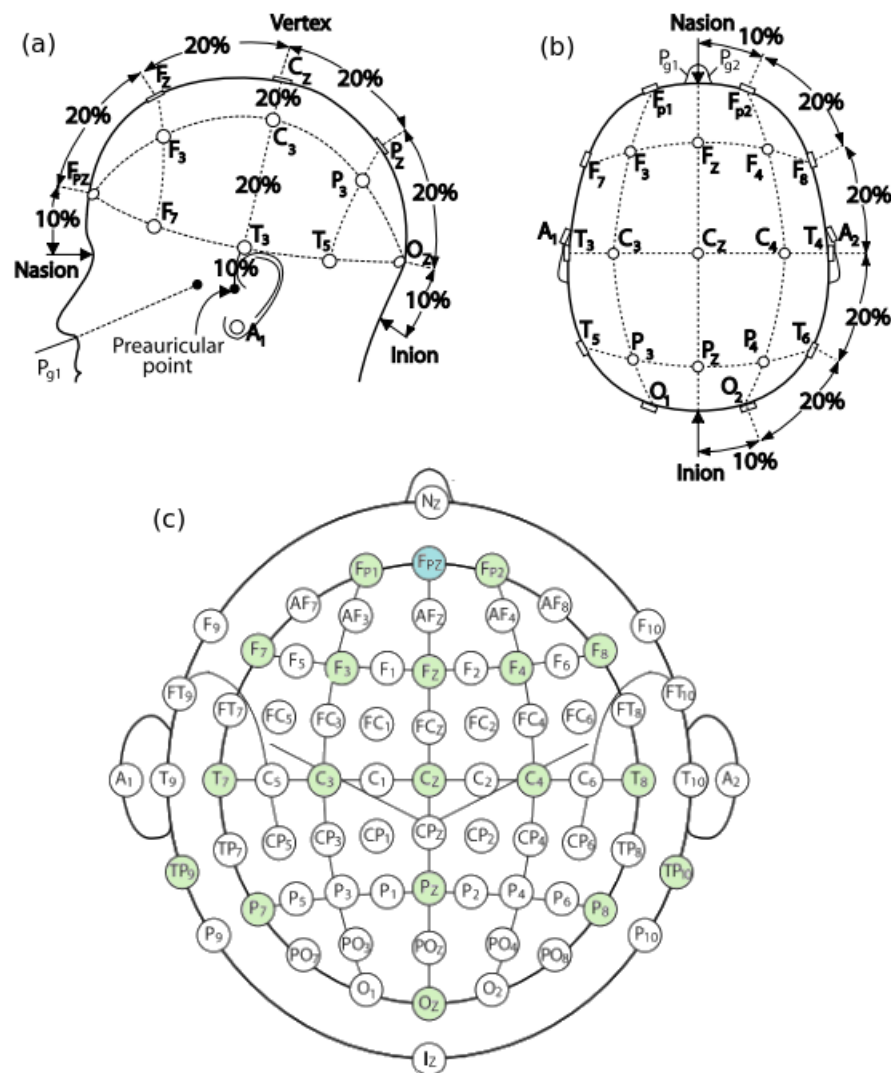


Figure 13 – Electrode placement scheme: (a) and (b) original 10–20 scheme (CREDITS: [Schalk and Mellinger \(2010\)](#) apud [Jasper \(1958\)](#)); (c) adaptation of 10–20 system for placing of up to 75 electrodes. The electrodes used in this work were colored in green and blue. (CREDITS: [Schalk and Mellinger \(2010\)](#) apud [Society \(1991\)](#)) (adapted)

- P → Parietal;
- F → Frontal;
- T → Temporal;
- C → Central;
- O → Occipital;
- A → Auricular.

There are two main methods to acquire EEG: differential and referential. They differ according to the source of the two input signals from the differential amplifier. This amplifier is an electronic device that performs the difference of two input signals and amplify the resulting signal. This device inputs are named inverting and non-inverting. The difference between them is that the

signal at the inverting input is inverted in the output. In the differential method both inputs of the differential amplifier come from electrodes positioned in places meant for signal acquisition, with the resulting signal being the difference between them. In the referential method, one of the differential amplifier inputs come from one or more electrodes of reference (REF) (Sanei and Chambers, 2007).

Ground (GND) and reference electrodes can be placed on the Vertex, or in points interconnected on the ears (earlobes or the mastoids); when both ears are physically interconnected with an electrical wire, or on the ipsilateral ear, or on the contralateral ear, or on the nose (Sanei and Chambers (2007) apud Bickford (1987)) or yet on FPz point above Nasion.

Schalk and Mellinger (2010) describe a simple procedure for electrode positioning using a cap and based on Figure 13 (c):

- Locate points Ns, In, A1 and A2 and, based on them, define the Cz point. Measure the distances Ns-In and A1-A2 and then use the 10-20 system to calculate the position of the remaining points;
- Define the FPz and Oz points on the scalp;
- Identify the Cz point on the cap and then put the cap on the participant, aligning the cap's and the participant's Cz points;
- Keeping the Cz point fixed, adjust the cap in a way that the longitudinal line (Ns-In) stays in the middle, FP1-FP2 stays in the horizontal at the same level of the Fpz point and O1-O2 also stays in the horizontal at the same level of the Oz point;
- Finally, place the electrodes REF and GND.

In this work the referential method was adopted due to the characteristics of the equipment used (see subsection 4.3.3). The data acquisitions were performed with gold electrodes. The cap used was hand-made with a thick elastic band as illustrated in Figure 14.



Figure 14 – Hand-made cap used in the experiments.

Nineteen electrodes (18 channels and the GND) were used for data acquisition. The electrodes were placed according to the 10-20 system at the following points: FP1, FP2, FPz (GND), F3, F7, F4, F8, Fz, C3, C4, Cz, T3, T4, T5, T6, TP9, TP10, Pz and Oz. The concentration of electrodes in the frontocentral areas are due to the observations that the categorical effects in the auditory evoked potentials are more visible at frontocentral scalp locations (Bidelman and Walker, 2017). Electrodes placed in the temporal region are recommended to capture AELR, AMLR and ABR in audiology (Hall, 2007). For the measurement of the AELR it is recommended to place the non-inverting electrodes on Fz or Cz points with inverting electrodes at A1 and A2 linking the earlobes. For the AMLR it is recommended to place the non-inverting electrodes on C3 and C4 sites with inverting electrodes at A1 and A2 (linked earlobes also), although the Pb component of the AMLR is often observed with the non-inverting electrodes at Fz or Cz (Nelson et al., 1997).

In this work, a Cz non-inverting electrode was used with respect to TP9 and TP10 instead of A1 and A2, respectively, since it is much easier to place the electrodes at these sites. This was not considered a problem because the generators of interest are captured by this setup, which also solved a potential problem that could be caused by small earlobes during the placement of the electrodes. Also, the TP9 and TP10 electrodes were not linked, as suggested by Hall (2007), because it is important to separate the signal related to each hemisphere for the investigation of laterality of the categorical effect desired in this work. This link would equalize the voltage at these sites. Due to the small amount and, consequently, the sparse distribution of electrodes, it is also not recommended to use the average reference because the calculated neutral reference point will not be exactly neutral, introducing distortions in the measured potentials.



### 4.3.2 Artifacts

Artifacts are internal (from the participant) or external (from the environment) disturbances that affect the EEG signal. The signal measured by the electrode in the EEG acquisition need to be conditioned before it can be analyzed. This conditioning basically consists of an amplification of the signal, a low-pass analog filtering followed by its analog/digital (A/D) conversion and finally its digital filtering. Basic types of filters include the highpass, lowpass, bandpass or notch, depending on the desired frequency range in the analysis and also on the artifacts that can be present in the signal. Understanding the cause of the artifacts helps in their identification, in the choice of the removal method and also in the adoption of measures to avoid their occurrence.

The interference noise from the electrical grid is an artifact that can be present in the EEG signal. This artifact can be easily detected in the frequency domain, as it consists of a peak around the frequency of 60 Hz (or 50 Hz depending on the country) and smaller peaks in their harmonics. Its removal can be performed by a notch filter around this frequency or by a comb filter to remove both the fundamental frequency (60 or 50 Hz) and its harmonics.

Artifacts due to eye blinking generate a signal in the AEP that can be up to ten times higher in amplitude than the cortical signals and that last from 200 to 400 ms (Sanei and Chambers, 2007). In Shoker et al. (2005) a hybrid technique based on Blind Source Separation (BSS) and Support Vector Machine (SVM) is applied with a reported accuracy of 99% in the detection of independent components related to the eye blink artifact.

Another commonly observed artifact is related to eye movement. Differently from the eye blinking artifact, where only the cornea moves, the eye movement artifact involves the movement of both the retina and the cornea. Its effect is similar to that of the eye blinking artifact, with the difference that its amplitude is larger and its frequency is smaller. The eye moves even if closed during the acquisition, what can generate significant artifacts.

There are also artifacts related to muscle movements. These movements can be either voluntary (e.g., jaw and eyebrow) or involuntary (e.g., heart beat). Some EEG acquisition devices have specific filters for the treatment of the heart beat artifact which can be better detected by the electrodes in the occipital region (Sanei and Chambers, 2007). Another muscle artifact that needs to be addressed in the acquisition of auditory evoked potentials is the post-auricular muscle response (PAMR), which is a muscle potential evoked by sound. It is detected in the post-auricular muscle located behind the external ear and it is synchronized with the auditory stimulus.

In the case of the voluntary movements, the treatment becomes more complex since they manifest themselves in an inconstant way in the EEG signal as high frequency noise (Schalk and Mellinger, 2010). There are some strategies that can be used with the subject in an attempt to avoid this type of artifact. For example, in order to avoid jaw movement, the subject can be asked to maintain his mouth slightly open during the acquisition (Schalk and Mellinger, 2010).

Artifacts due to interferences of the stimulation system can also occur if the circuitry of the system is not well shielded. This artifact has a waveform very similar to the waveform of the stimulus in the EEG signal, so there is no significant latency between the recorded potential and the stimulus. This behavior is similar to that observed in cochlear microphonics (CM).

The CM is an artifact whose waveform follows that one of the stimulus. It is a potential generated by the cochlea hair cells (Skoe and Kraus, 2010). To distinguish the CM from the auditory evoked potential, one can verify the response latency, since the CM occurs almost simultaneously with the stimulus. In turn, the auditory evoked potential presents typical wave latencies.

Another observation that helps to distinguish a CM from the desired auditory evoked potential is that an increase in the stimulus frequency does not affect the CM but changes the shape and latency of the evoked potential. Another point is that these two potentials are different in the maximal response amplitude. While the CM grows linearly with an increase in the stimulus intensity, the evoked potential does not exceed a maximal limit (Skoe and Kraus, 2010).

The ERP is obtained by repeating the same stimulus several times and averaging the AEP to all repetitions. Then, the CM can be minimized by using stimuli with alternate polarities, i.e., by adding a 180° lag to half of the stimuli that will be repeated. This way, when one performs the averaging, the effect of the CM, that follows very well the stimulus waveform, is canceled, while the evoked potential, which is not affected by the phase shift in the stimulus, will remain in the averaged response. See Appendix G for more details about the averaging of EEG signals.

### 4.3.3 RHD2000 system

The EEG acquisition system used in this work was the RHD2000 made by the Intan Technologies, LLC (Technologies, 2014). The RHD2000 is a modular system with integrated circuits (IC) and an integrated development environment for bio-signals acquisition. The main component of the RHD2000 system used in this work is the integrated circuit (IC) RHD2132 amplifier chip illustrated in Figure 15. It consists of an amplifier and an A/D converter with 32 acquisition channels. In the image, the IC is assembled in a printed circuit board (the headstage), but it is

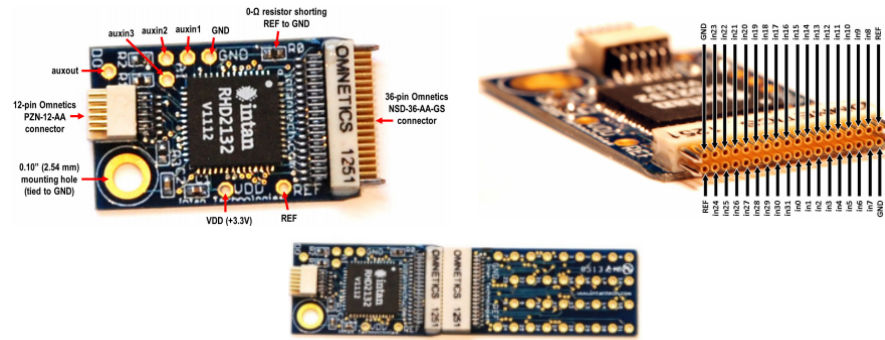


Figure 15 – RHD headstage with the RHD2132 amplifier chip. The headstage also contains a few support components (resistors and capacitors), a 12-pin Omnetics connector and a 36-pin Omnetics strip connector. At the bottom figure, the amplifier is connected to a 36-pin electrode adapter board that facilitates the soldering of the electrode connectors.

CREDITS: (Technologies, 2014) (adapted)

also sold separately so that the user can assemble its own circuit. The dimensions of this IC are  $4.8 \times 4.2$  mm.

The circuit of the RHD2132 amplifier chip is illustrated in Figure 16. This IC enables the acquisition of bio-potentials following the referential method, since all 32 channels available on the IC have a single port-linked reference. Its A/D converter is 16 bits providing for the representation of signal amplitude information. Each input channel is sampled at 30 kHz (kSamples/s), resulting in an aggregate rate of 1.05 MHz for the 35 multiplexed channels: 32 amplifier inputs from the electrodes and 3 auxiliary inputs that can be used to set amplifier bandwidth if the on-chip bandwidth registers are not used. The total gain of amplifier is 192 (considering the midband region). The A/D converter of the RHD2000 series amplifiers operate with a total voltage range of 0–2.45 V. This, combined with the 16-bit resolution, provides  $\mu\text{V}$  measurements without the need for large amplifier gains.

The step size of the quantizer of the A/D converter is given by Equation 4.1,

$$\frac{2.45\text{V}}{2^{16\text{bits}}} = 37.4\mu\text{V} \quad (4.1)$$

which divided by the gain of the amplifier results in the minimum voltage of  $0.195 \mu\text{V}$ , measured by the system. In short, observing the Figure 16 we can see that the analog input voltages (In0 - In31) go through two amplification stages (gains of 96 and 2), which gives an aggregate gain of 192. After amplification, the analog signals are multiplexed and fed to the 16-bit A/D converter where they are sampled (with a sampling frequency up to 30 kHz) and then each sample is quantized in one of the  $2^{16}$  quantization levels (with the difference between levels being of 37.4

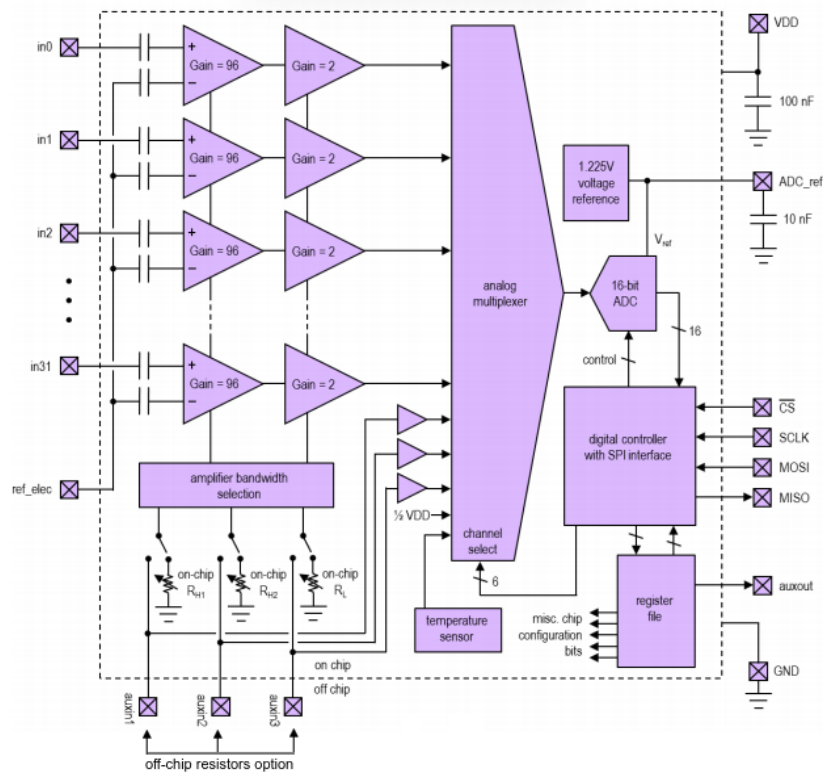


Figure 16 – RHD2132 amplifier chip IC. Internal Circuit Diagram.  
CREDITS: (Technologies, 2012) (adapted)

$\mu\text{V}$ ).

The RHD 32-channel headstage can be attached to the RHD2000 interface board sold by Intan Technologies as illustrated in Figure 17.

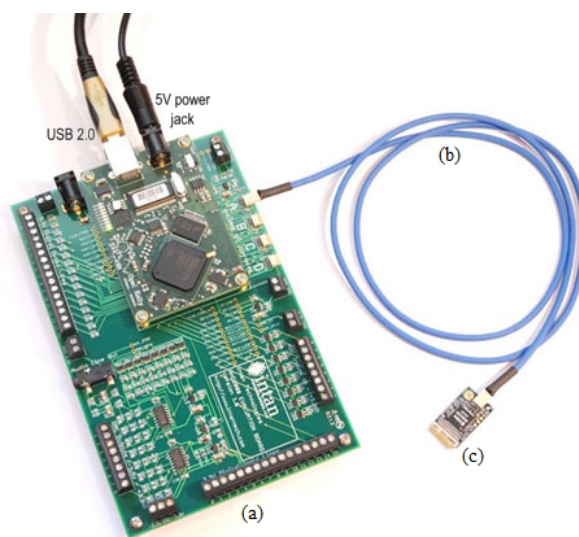


Figure 17 – RHD2000 system with the (a) interface board, the (b) SPI cable and the (c) RHD 32-channel headstage.

CREDITS: (Technologies, 2014) (adapted)

The interface board supports connecting up to four RHD2000 headstages via serial peripheral interface (SPI) cables (see Figure 17 (b)) allowing to work with up to 256 channels (considering the use of RHD 64-channel headstages) with a rate of up to 30 kHz per channel. The system consists of an USB interface board, Figure 17 (a), which enables the system to communicate with a computer; and one to four headstages with amplifier chips, Figure 17 (c). The boards are connected through SPI cables with 12-pin Omnetics connectors. Figure 18 illustrates the interface board.

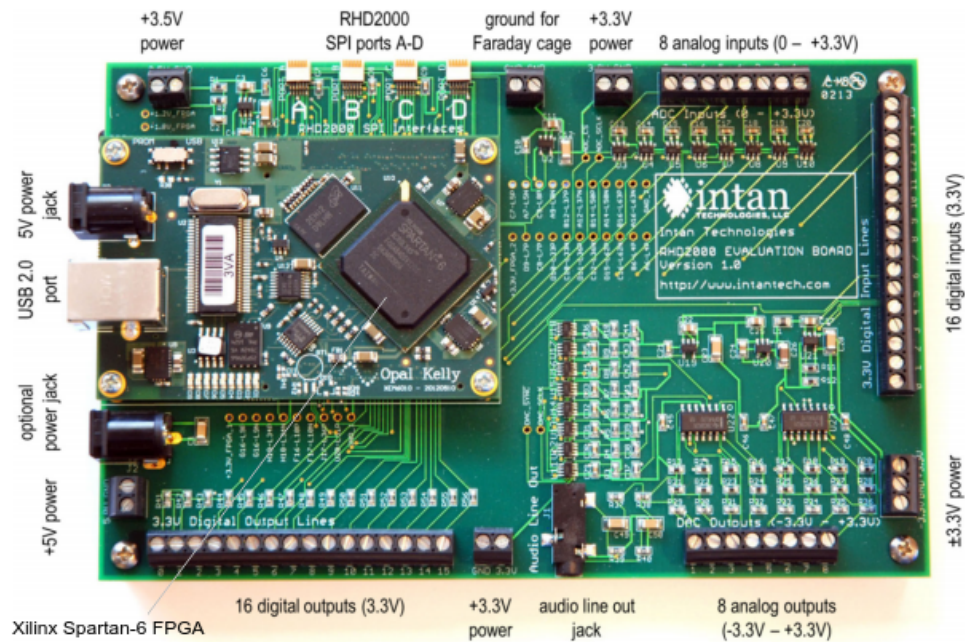


Figure 18 – RHD2000 System Development Board.  
CREDITS: (Technologies, 2014)

The interface board is supplied with 5 Vdc voltage from a source connected to the AC network. In this project, the original power supply was replaced by a 12 Vdc battery connected to a voltage regulator circuit for a 5 Vdc output. This was done in order to decouple the power supply from the power grid and therefore to eliminate a source of interference (the network noise). The interface board contains a Xilinx Spartan-6 FPGA that controls the amplifier cards and communicates with the host computer through an Opal Kelly XEM6010 USB/FPGA interface. The interface board has 16 digital inputs, 16 digital outputs, 8 analog inputs, 8 analog outputs, a standard P2 female audio output plug (which can be connected to speakers) and a Faraday cage ground terminal for conductive shields that improve 50/60Hz noise rejection (Technologies, 2014).

The availability of analog inputs in the development board is important in our work as it enables the simultaneous acquisition of the evoked potential sync signal (the trigger). The availability of this signal makes it possible to calculate the coherent average later in the signal processing

step. Unfortunately, this analog input presented a serious cross-talk problem with the electrode channels, inserting both noise and the trigger signal itself into the EEG measurements.<sup>1</sup>

The RHD2000 maker provides a software that is open source and provided free of charge. The software is used to perform parameter configurations such as sampling rate, upper and lower cutoff frequencies, filter configuration, offset treatment, signal visualization, impedance tests, among others. The maker of the RHD2000 system also provides a MATLAB toolbox for use with RHD2000 interface board, a LabView library, Python scripts, MATLAB functions for reading and importing data, Windows drivers, Verilog code for the field programmable gate array (FPGA), among others. The software interface is shown in Figure 19.

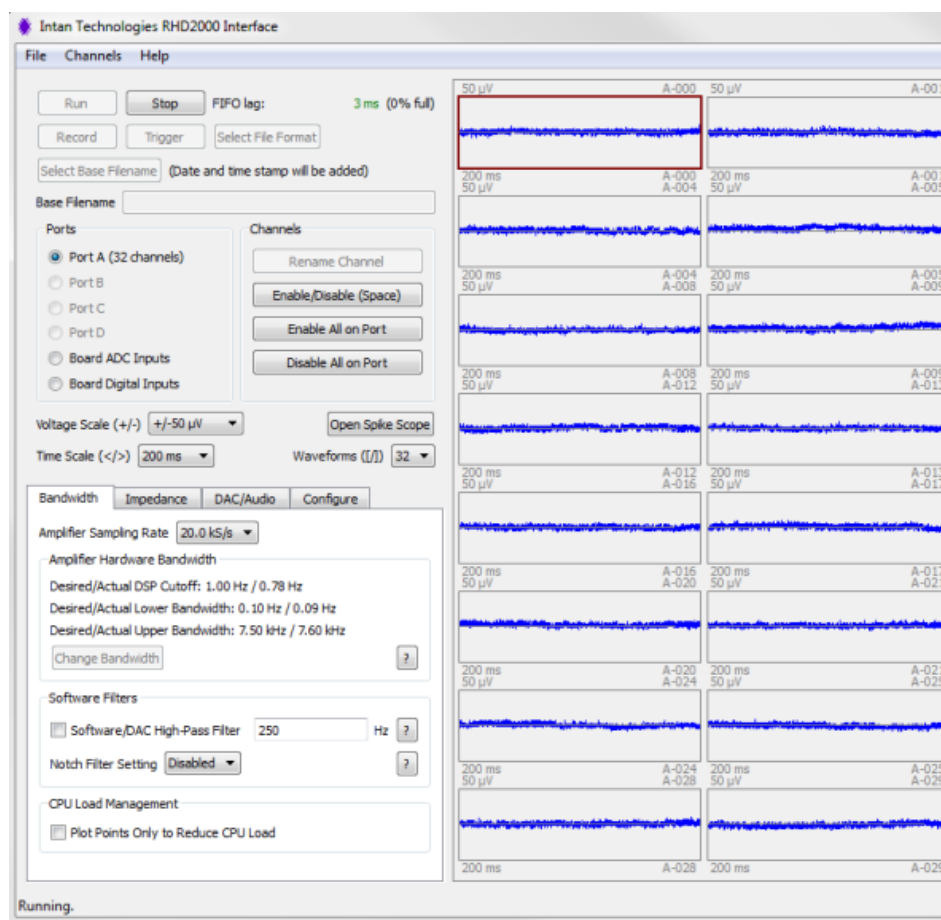


Figure 19 – Interface to the RHD2000 system.

CREDITS: (Technologies, 2014)

The RHD2000 software allows the configuration of analog filters of the types low-pass, high-pass and a digital notch filter (the latter for 50 or 60 Hz). Since the filters are configured digitally through software, the bandwidth may be changed on the fly. Registers available on the amplifiers allows the configuration of on-chip resistors that set the upper and lower bandwidth of the

<sup>1</sup> More details of the other features of the amplifier cards are available in (Technologies, 2012) and (Technologies, 2014).

amplifiers. The low cut-off frequency can range from 0.1 Hz to 500 Hz, whereas the high cut-off frequency can range from 100 Hz to 20 kHz. In this work, a sampling rate of 5 kHz was used, which is more than enough to record the major known brain oscillations in addition to the first two formants of the pronounced vowels and the  $F_0$ . The default value of the software for the amplifiers bandwidth was kept (between 0.1 Hz and 7500 Hz). The digital notch filter, configured to 60 Hz, was included just during acquisitions (to improve signals visualization) but was not kept in the recorded EEG signal. We decided to implement all the need post-acquisition filters in order to have a better control of the signal processing performed.

### 4.3.4 Circuit artifact treatment

As discussed in section 4.3.2, the artifacts can be internal (participant) or external (environmental). Understanding the origin of artifacts helps their identification, the choice of the removal method and also the adoption of measures to prevent their occurrence. Participant-related artifacts are those related to blinking and eye movement and also those of the participant's other muscular movements such as jaw, eyebrows and heart. Such artifacts can be attenuated with the use of pattern recognition techniques or simple amplitude-based removal, since some of them generate a signal with amplitudes up to ten times greater than the cortical signals (Sanei and Chambers, 2007).

Regarding external artifacts, some problems specific to the RHD2000 system used in this work were noted. The entire circuit with both boards was not shielded from external electromagnetic fields. Also, the common mode rejection rate of the amplifiers (82 dB at 1 kHz) was not high enough, as excessive signal noise was observed. Furthermore, trigger signal interference, used to synchronize the signal being played to the individual with the EEG signal, was also observed in the signals measured. To investigate these noise sources and to elaborate measures to mitigate them, it was decided to perform tests on a head model represented by a melon, as illustrated in Figure 20.

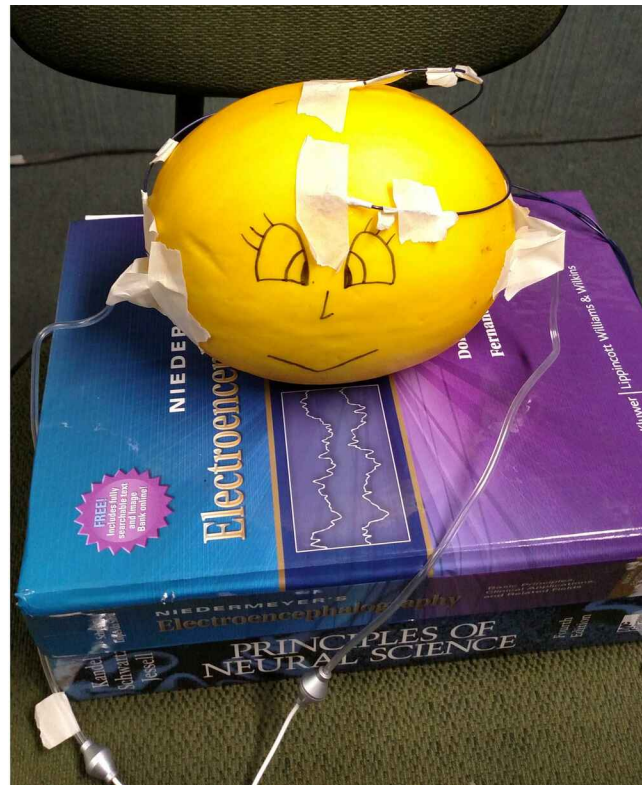


Figure 20 – Melon head model for artifact detection in the RHD2000 system.

This model provides low “electrode-shell” impedance ( $< 5 \text{ k}\Omega$ ), does not include movement artifacts in the tests, and is close to the size of a human head, allowing electrode spacing close to the practical case. In addition, it enables to speed up tests, since there is no need to schedule volunteers or worry about the participant’s fatigue after hours of testing. Detection and treatment of artifacts appearing in the signal was greatly facilitated as responses to the auditory evoked potentials of the signal were not present. This model allowed tests to be performed until problems were mitigated, before moving on to the tests with human participants.

Several tests were performed in an attempt to detect possible sources of noise. The first step was to replace the power supply (provided by Intan) with a battery used in uninterruptible power supply (UPS) systems. These tests showed that the power supply provided with the RHD2000 system inserted strong power grid artifacts into the 60 Hz measured signals and their harmonics. By switching to the battery, these artifacts were visibly reduced. The battery was regularly recharged from an external source available in the laboratory.

The notebook computers used in the experiments also needed to work on battery power alone, disconnected from the power grid. One notebook was used to acquire the RHD2000 readings through a USB interface whereas a second notebook was used to generate the tasks, record the responses in the tasks involving identification of phonemes and provide the auditory stimuli and



trigger signals for the tasks where EEG acquisitions were performed. A third computer, used for video viewing, was a desktop and therefore had to be connected to the power grid. This, however, had no influence on the measurements since this computer had no physical connection to the measurement or stimulation circuits. A small voltage regulator circuit was built in a printed circuit board to convert the 12 V battery to the 5 V required for RHD2000. Figure 21 illustrates this new power source. A 60Hz digital notch filter available on the RHD2000 system was also used during acquisitions to mitigate power grid noise that still corrupts the signal, but with a much lower intensity now.

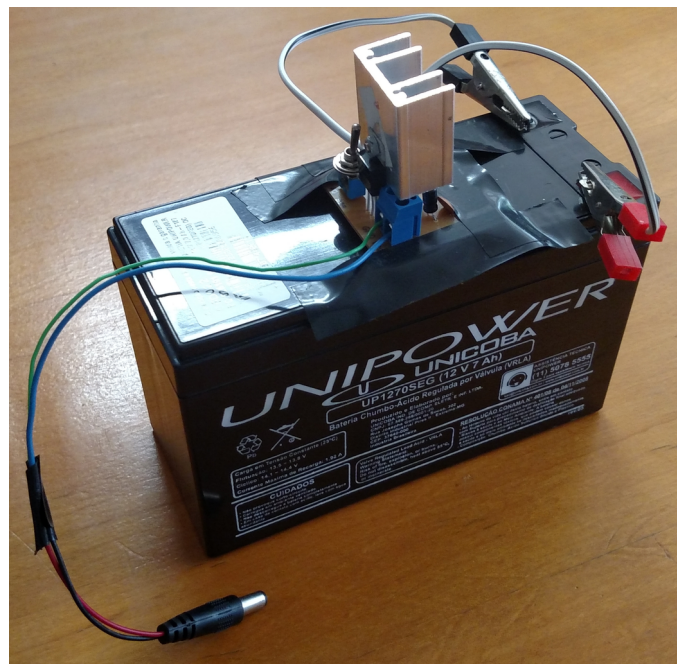


Figure 21 – Power supply adapted for the RHD2000.

After this change in the power supply, the negative battery terminal became the circuit's GND (ground) reference. Thus, any interference influencing the GND enters the signal. We believe, therefore, that the trigger started to enter the EEG signal due to a ground loop issue, even using a cable, illustrated in Figure 22, to split the audio channels (one channel for the trigger and the other for the stimulus that goes to the participant's earphones). According to Vivace (2019) (own translation),

“The ground loop occurs when there is more than one ground path, generating undesirable currents between these points and causing possible measurement errors, malfunctions, intermittences and even equipment burnout.”

The RHD2000 circuit board has several analog and digital non-amplified I/O pins for several uses. Initially, one of its analog inputs was used to receive the trigger signal, which is just a



Figure 22 – Cable to separate audio channels.

sound signal sent out by one of the notebook computers, through its sound card. Throughout the control board several GND points are available for use with such I/O pins. There are two more GND terminals available for use in the amplification board in addition to the four points soldered on the chip. This board receives power (therefore GND) from the control board and also provides the digitized and amplified measurements to it, through the SPI cable mentioned before. Finally, the two circuits form a common GND system of the type illustrated in Figure 23.

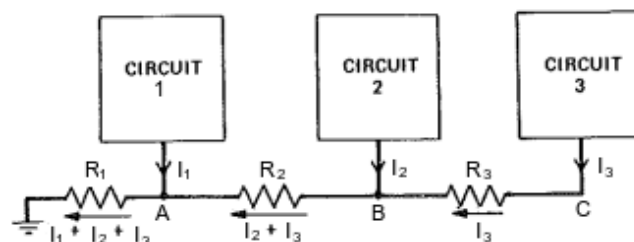


Figure 23 – Common GND system causing ground loop.  
 CREDITS: (Ott, 1976) (adapted)

At high frequencies, such as those used in this work, the impedances of the GND connections increase (as a result of inductance increases) so that two physically separated GND points will rarely be at the same potential. Increasing inductances can also result in inductive couplings between GND conductors. Depending on the frequency, these conductors may begin to function as antennas radiating noise.

Since GND points may not be at the same potential, any differences in potential between the circuit's GNDs ( $V_G$ ) will enter the signal ( $V_S$ ). Figure 24 illustrates how this might be happening on the amplifier. Note that the  $V_G$  voltage, which is a noise, links with the input signal that should be the only one entering the amplifier.

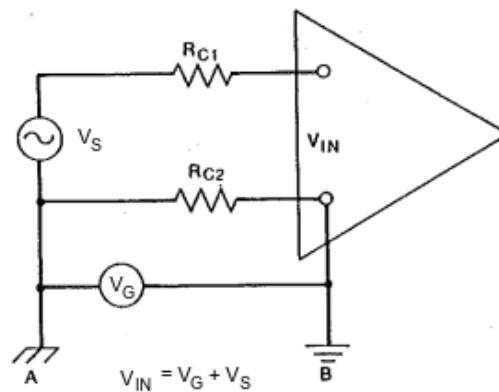


Figure 24 – A noise voltage enters the amplifier if the circuit has more than one GND point.  
CREDITS: (Ott, 1976) (adapted)

The trigger signal used in this work is a square wave, lasting half the stimulus time and occupying the second audio channel of this stimulus as illustrated in Figure 25 of the stimulus [da]. All stimuli were generated this way. A cable is used to separate channels on the notebook's sound output. It separates the stimulus channel, that goes into the earphones, from the channel with the trigger signal. A second stereo sound cable receives the stimulus signal and duplicates it to its second channel so that the earphones connected to that cable presents the stimulus on its two channels.

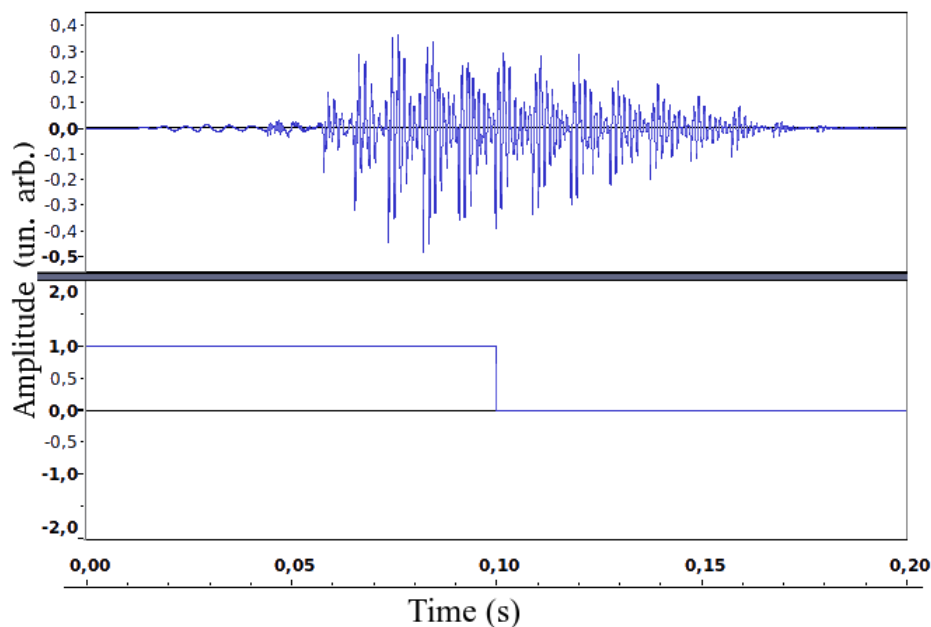


Figure 25 – Trigger signal on second stimulus audio channel for the [da] syllable.

Since changing the headstage board (where the RHD2132 amplifier chip is placed) is not an option, all that can be done to minimize the GND effects is to rely on the common mode rejection rate of the amplifiers and use GNDs as close as possible to each other in the case where the

amplifier board is connected to the trigger. To address the problem of the distance between the amplifier's GND and trigger signal's GND, and at the same time reduce the amplitude of the latter signal (thus reducing field effects on the electrodes), we decided to connect the trigger signal directly to one of the EEG electrode inputs. In order to do this, the trigger signal, which comes from the notebook's sound card circuit, was isolated from the acquisition card using an TIL111 optocoupler powered by the amplifier board (reducing GND distances once again). In addition, the signal amplitude was reduced with a simple voltage divider built with resistors. The entire circuit of the optocoupler was welded on a board as shown in Figure 26.

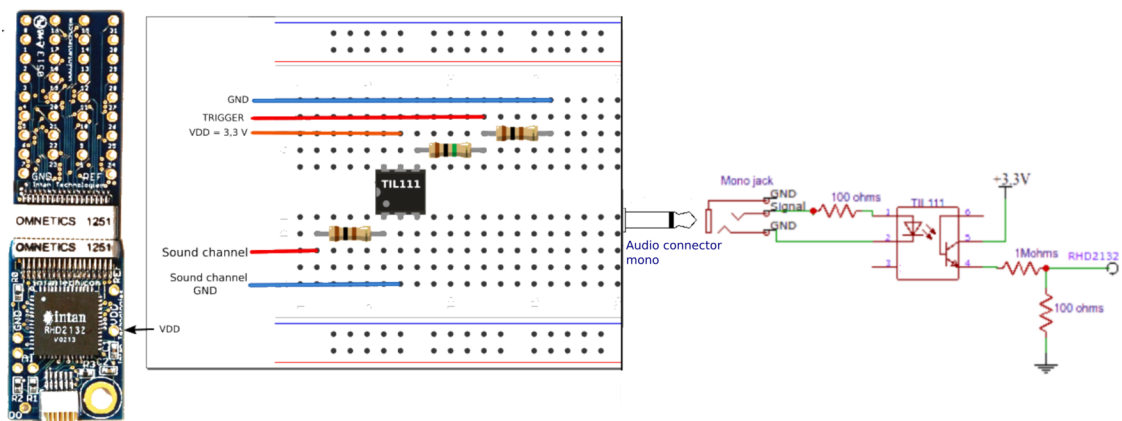


Figure 26 – Optocoupler circuit and voltage divider.

Several other tests were conducted in an attempt to reduce artifact interference with the EEG signal, but many made no difference or made things worse. One attempt involved shielding the electrode cables and grounding the shield on the amplifier board's GND. Such shielding led to a greater interference signal and therefore such approach was aborted. However, improvements were observed by braiding and stretching, as much as possible, the electrode wires. The amplifier and digitizer board, optocoupler circuit as well as the electrode and trigger connections to the amplifier board were all placed in a metal container that acted as a Faraday cage. The RHD2000 circuit board was also placed in a plastic container wrapped in aluminum foil. Although this arrangement did not reduce the noise significantly, we decided to keep it for protection and mobility of the circuits. Lights at the place of the experiment, formerly reactor fluorescent lights, were replaced by LED lights.

After each improvement, tests were performed on the melon with 400 syllable /ba/ repetitions. The result presented in Figure 27 shows the average magnitude (RMS) of the noise, which was of 0.2267  $\mu\text{V}$  at left electrode and 0.3298  $\mu\text{V}$  at the right electrode. The difference between the electrode magnitudes is probably due to the difference in the electrode-shell impedances of each one.

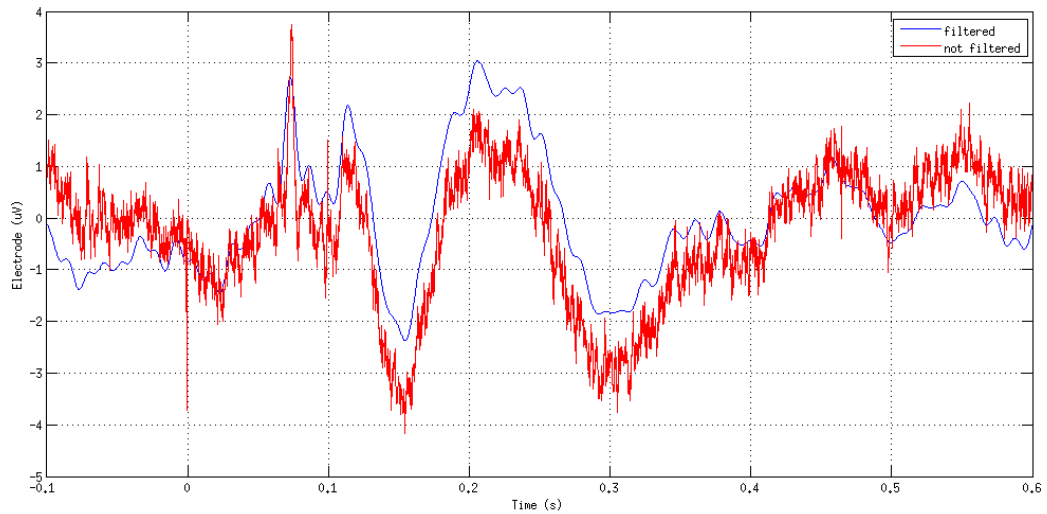


Figure 28 – Coherent average of 350 stimulus repetitions /da/ in human experiment.

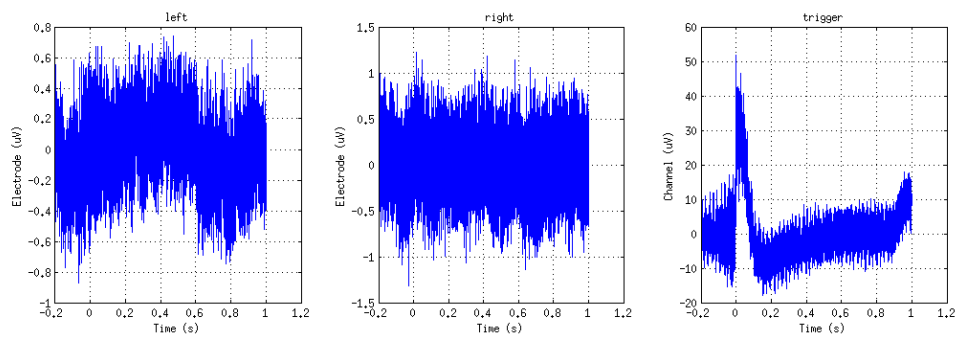


Figure 27 – Coherent average of 400 stimulus repetitions /ba/ melon model.

Once the trigger was no longer interfering with the EEG signal and the noise magnitude was considered acceptable, we started tests on humans. Figure 28 illustrates the result of a pilot test performed on a volunteer using 350 syllable /da/ repetitions (details about the stimulus are presented in the 4.4 section). The averaged signal (in red) has a baseline correction and its filtered version with 1-80 Hz bandpass filter is shown in blue. The signal to noise ratio was 10.84 dB. We believed that this relationship could be improved with a greater number of stimulus repetitions but, due to the time limitation of the experiment (avoiding fatigue effects on the participant), this increase is not possible. Moreover, the filtered signals allows us to perform the desired analysis in this work.

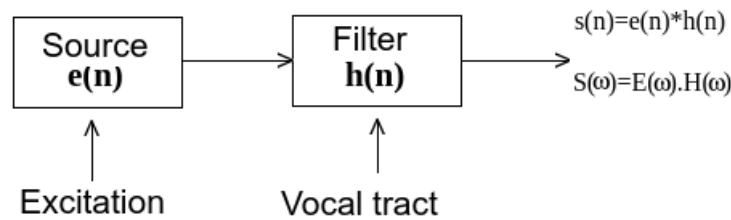


Figure 29 – Speech Production Model.  
CREDITS: [Alcaim and Oliveira \(2011\)](#) (adapted)

## 4.4 Speech stimuli

Two stimuli continua were used in the data acquisitions conducted in the study. The first continuum was based on the variation of the **VOT** whereas the second one was based on the variation of the spectral characteristics of the syllable vowel (formant frequencies). **VOT** is the time between the completion of occlusion release in the occlusive consonants and the voicing (or vocal fold vibration) ([Silva et al. \(2019\)](#) apud [Lisker and Abramson \(1964\)](#)). This value may be negative for voiced consonants, when voicing occurs before the end of the occlusive release, which is characteristic of voiced occlusive consonants. The **VOT** can be positive for voiceless occlusive consonants, where voicing occurs after the occlusive release ([Silva et al., 2019](#)). The spectral characteristics of the second continuum stimuli are their formant frequencies.

To better understand formants, consider a simple model of speech production. A voice signal can be modeled by a source-filter system. A voiced signal  $s(n)$  is given by the temporal convolution of an excitation signal  $e(n)$ , which would physically represent the vibrations of the vocal cords; with the impulse response of a filter  $h(n)$  which represents the vocal apparatus, modulating the excitatory signal. Figure 29 illustrates the source-filter model described.

Considering the  $S(\omega)$ ,  $E(\omega)$  and  $H(\omega)$  are the frequency response of  $s(n)$ ,  $e(n)$  and  $h(n)$ , respectively, the magnitude spectrum of this model can be given by ([Alcaim and Oliveira, 2011](#)):

$$|S(\omega)|_{dB} = 10 \log_{10} |S(\omega)| = 20 \log_{10} (|E(\omega)| \cdot |H(\omega)|) = |E(\omega)|_{dB} + |H(\omega)|_{dB}. \quad (4.2)$$

In speech production, excitation signal harmonics (fundamental speech frequency –  $F_0$ ) close to the resonant frequencies of the vocal tract are emphasized. These resonances are typical of each person giving their particular speech characteristics. Formants are the fundamental frequency harmonics in the natural resonances of the vocal apparatus ([Huang et al., 2001](#)). Figure 30 illustrates a spectrogram (which is a map time vs. frequency) obtained for the syllable /da/ in the

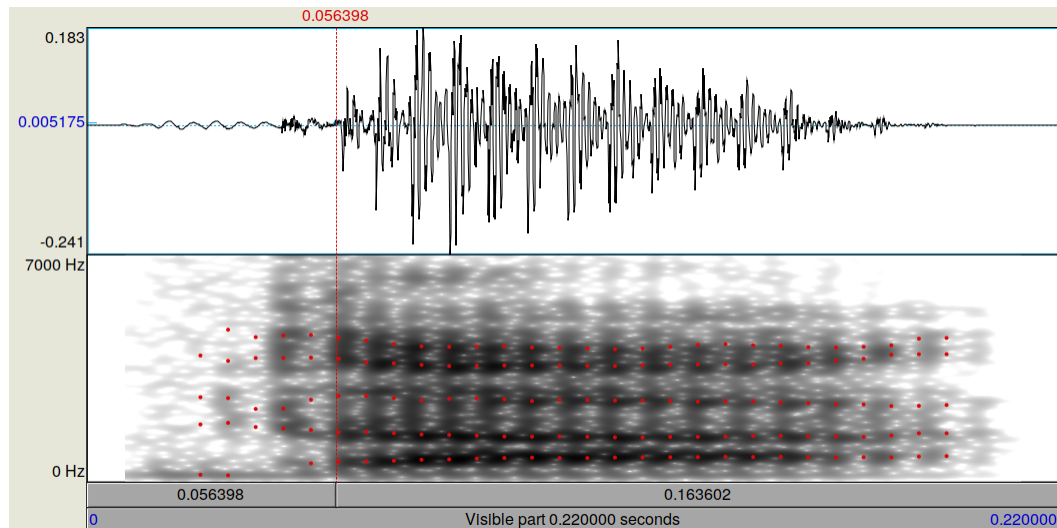


Figure 30 – Spectrogram of the syllable /da/.

Praat software (Boersma and Weenink, 2018). The formants of the vowel “a” of this syllable can be visualized in black horizontal lines, marked by the software with red dots, showing a relative accumulation of energy at these frequencies.

Each continuum was obtained differently but both consisted of 200 stimuli. This amount of stimuli was used to allow large accuracy in the psychometric curve obtained for each participant. This curve relates the physical characteristics of the stimuli (e.g.: VOT or formant frequencies) with the way they are perceived behaviorally. The continuum based on VOT variations (VOT continuum) was obtained between the syllables /da/ and /ta/, where at each step the alveolar voiced consonant /d/ becomes the alveolar unvoiced /t/ while the vowel /a/ is the same for the entire continuum (excerpted from the original syllable /ta/ with 160 ms of duration). Its first two glottal cycles were included in the morphing to compensate for the syllable /ta/ aspiration time. The formant-based continuum (formant continuum) was obtained between the syllables /pa/ and /pɛ/ where the unvoiced consonant /p/ was kept constant along the continuum while the vowel /a/ was varied at each step and throughout the continuum as it becomes the vowel /ɛ/.

The four original syllables, /da/, /ta/, /pa/ and /pɛ/ have a meaning in the Brazilian Portuguese. This “meaning” dimension will not be analyzed in this work and this is not expected to interfere with the categorical perception investigation in the latencies that will be analyzed as the literature showed (Liégeois-Chauvel et al., 1999, Husain et al., 2006, Mirman et al., 2004, Holt et al., 2004).

In the case of the formant continuum, the transition from /a/ to /ɛ/ is acceptable in Portuguese because it has no other intermediate sounds according to the Portuguese vowel acoustic space illustrated in Figure 31, where F1 and F2 are the first and second formant frequencies respectively.

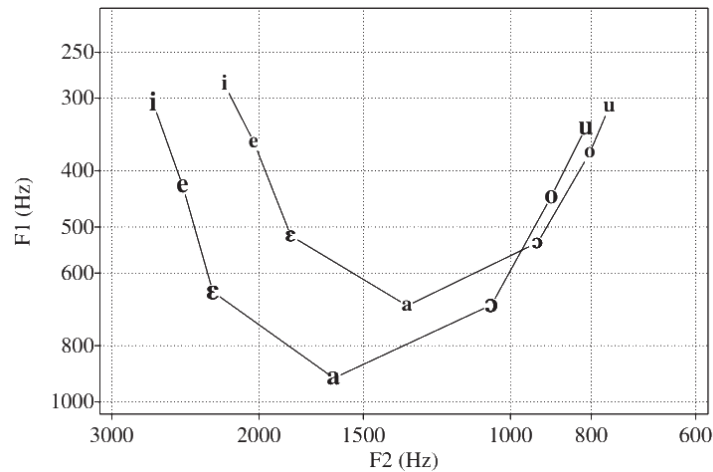


Figure 31 – F1 × F2 chart of Brazilian Portuguese tonic oral vowels. Large font: women; small font: men. CREDITS: [Escudero et al. \(2009\)](#) (adapted)

By convention, F1 × F2 graphs are always drawn with the axis configuration present in Figure 31.

The four original syllables, /da/, /ta/, /pa/ and /pε/ were pronounced by three different people, two males and one female, who were asked to maintain the same intonation and duration. We used a Brüel & Kjaer condenser microphone type 4165, with an incidence degree of 0°, illustrated in Figure 32. The recording was held inside a sound-proof booth. Audacity software ([Mazzoni and Dannenberg, 2000](#)) was used for recording with a sample frequency of 44100 Hz and considering one channel (mono). After visual and auditory inspection, also in Audacity, the clearest pronunciation was selected. The recording voice belonged to one of the men with an average fundamental frequency (F<sub>0</sub>) of 115 Hz. These natural stimuli were chosen instead of synthesized stimuli because they have complex spectral and temporal details that can provide important information for categorical perception. Furthermore, it has been shown that the P1-N1-P2 long latency evoked potentials are reliably evoked with natural speech stimuli ([Tremblay et al., 2003a](#)).

For the formant continuum, was used the speech transformation and representation using adaptive interpolation of weighted spectrum (**STRAIGHT**) toolbox for MATLAB ([Kawahara et al., 1999](#)). **STRAIGHT** is a software for speech analysis, modification, and synthesis based on the source-filter speech decomposition model working with one channel signal (mono). The continuum was generated between the vowels /a/ and /ε/, and later the consonant /p/ (extracted from the original /pa/ syllable) was added to the beginning of each of the 200 morphed vowels. As shown in [Kawahara and Matsui \(2003\)](#), where emotional aspects present in speech are morphed, **STRAIGHT** provides a natural sounding morphing for human speech. In this work, we performed a linear morphing with 50 anchor points (points representative of the speech signal) manually marked at points along the first 4 formants plus the fundamental frequency in the spectrograms of each vowel. For details of the use of **STRAIGHT** see [Kawahara \(2005\)](#).





Figure 32 – Microphone Brüel & Kjær for stimuli acquisition.

The frequency difference between the first and second formants in the continuum ( $F_2-F_1$ ) goes from 533 Hz for /pa/ to 1387 Hz for /pɛ/.  $F_1$  decreases from 758 Hz to 547 Hz and  $F_2$  increases from 1291 Hz to 1934 Hz. All morphed sounds had their energies normalized at the end of the process using a function of **STRAIGHT**. Each stimulus on this continuum lasts for 191 ms (fixed /p/ consonant **VOT** 15 ms and vowel 176 ms). Figure 33 illustrates the nonambiguous stimuli /pa/ and /pɛ/ numbered as stimulus 1 and 200 respectively and an ambiguous stimulus from the middle of this continuum (number 100).

The **VOT** continuum was constructed in R (**R Core Team, 2014**) using a script based on the script used in **Bellier et al. (2013)** where a continuum from /ba/ to /pa/ was created. The script works with a WAV format audio file containing the two syllables to be morphed. The script contains the syllable final duration value, and a CSV file (generated with the aid of PRAAT software (**Boersma and Weenink, 2018**) and a second python script). This file contains markers made for the two syllables (manually), necessary for the script to perform the morphing including: the start time of the syllables, the plosive release time and the time of the beginning of the voicing. Here, this script was used to create the /da-/ta/ continuum. The consonants of these syllables have the same point of articulation at the front of the vocal tract, but differ in **VOT**. The original voiced consonant syllable /da/ lasts for 386 ms with a negative **VOT** of 130 ms. This **VOT** was cut to 52 ms because we intended to have a final stimulus of around 200 ms duration which is related to the total time of the **EEG** experiments. The unvoiced consonant syllable /ta/ lasts for 160 ms with a positive **VOT** of 16 ms.

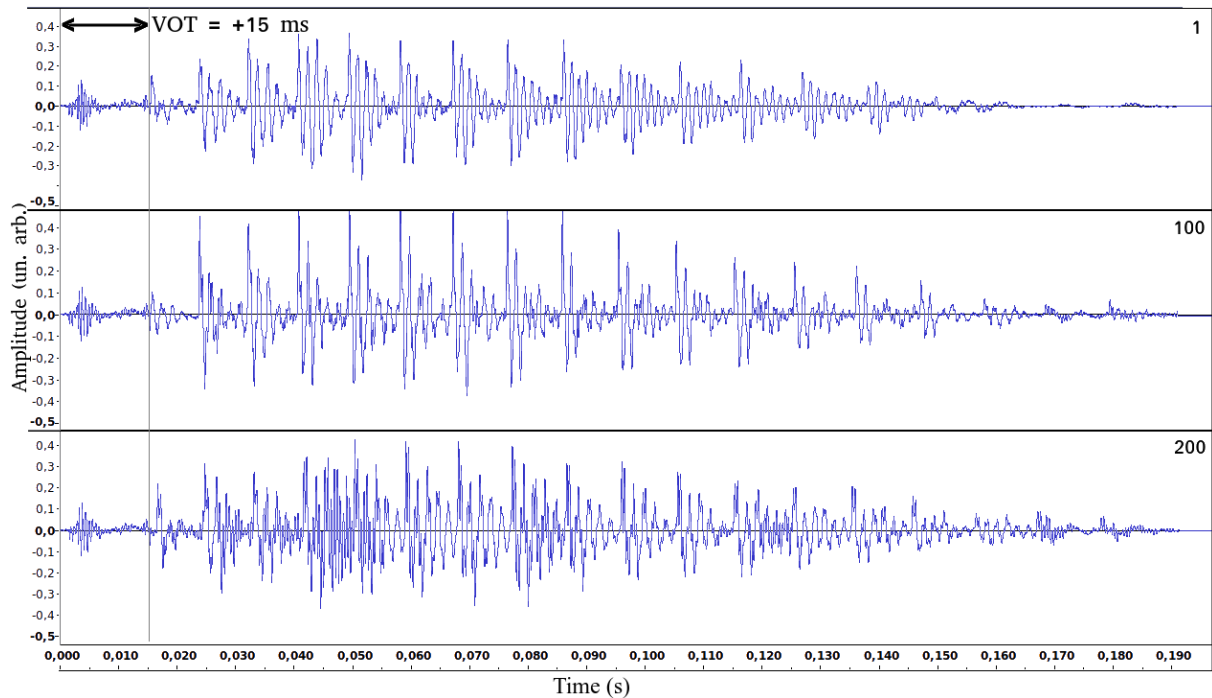


Figure 33 – Samples of the stimuli continuum based on formant variation. Stimulus /pa/ nonambiguous (above), /pɛ/ nonambiguous (below) and ambiguous (middle).

The script performed a morphing procedure to generate the 200 intermediary, synthetic stimuli of the continuum. These stimuli were created by continuously varying the onset of the voicing murmur of /da/, from  $-52$  ms to  $0$  ms. In all cases, the release burst is present, resulting in a  $+16$  ms VOT value for the extreme /ta/ syllable. Each stimulus was saved to an audio WAV file. The reference time was chosen to be the beginning of the stationary part of the vowel, such that the beginning of the WAV file corresponds to  $t = -74$  ms in the original stimuli. Thus, the stationary part of the vowel for all stimuli was temporally aligned, in relation to the beginning of the WAV file. Each stimulus on the continuum obtained was contained in a duration of 220 ms including silence before and after (variable according to the stimulus). Figure 34 illustrates the nonambiguous stimuli /da/ and /ta/ numbered as stimulus 1 and 200 respectively and an ambiguous stimulus from the middle of this continuum (stimulus number 100). The alignment point is shown by the red line in Figure 34.

The voiced part of the syllable /da/ is about 15 ms longer than the syllable /ta/ due to the way the continuum was generated. This reduction in syllable voicing duration /ta/ did not affect the syllable's intelligibility. Hillenbrand et al. (2000) showed that changes in vowel duration have little effect on the ability of the participants to identify them as long as the vowel spectral characteristics are preserved. In addition, in Brazilian Portuguese, the vowel /a/ after a voiced occlusive tends to be longer than that after an unvoiced occlusive (Melo et al., 2011, Lofredo-Bonato, 2008).

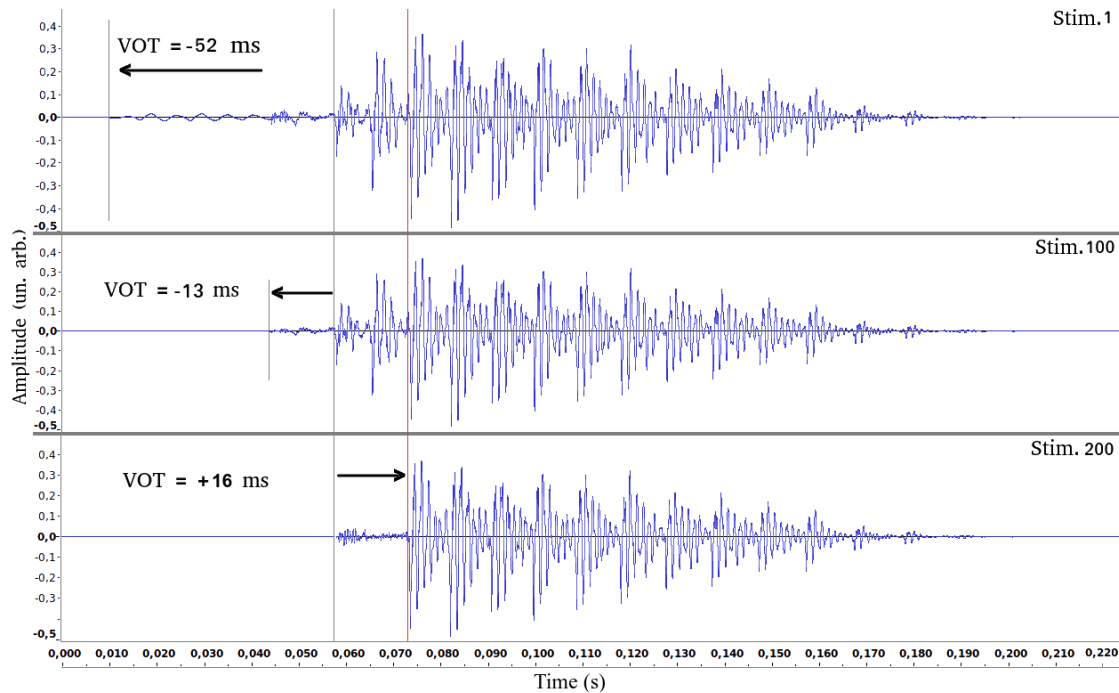


Figure 34 – Samples of the stimuli continuum based on VOT variation. Stimulus /da/ nonambiguous (above), /ta/ nonambiguous (below) and ambiguous (middle). The red line indicates the alignment point of common to all stimuli at the beginning of the stationary part of the vowel.

The **VOT** of the nonambiguous stimulus /da/ (first of the continuum) is the most negative with  $-52$  ms due to voicing prior to plosive release. The ambiguous stimulus in the middle of the continuum has a negative but short **VOT** with  $-13$  ms, and finally, the nonambiguous stimulus /ta/ presents a positive **VOT** of  $+16$  ms, indicating a small aspiration after plosive release. In the latter case, the occlusive consonant is considered to be non-aspirated according to the classification made by [Cho and Ladefoged \(1999\)](#) in which occlusions with **VOT** of 0 to 35 ms falls into this class.

It has been shown that the average **VOT** for voiced occlusive consonants /d/ followed by the vowel /a/ would be  $-89.15$  ms (in a range from  $-155.05$  to  $-50.20$  ms), with voicing beginning before the release of the plosive consonant ([Klein, 1999](#)). For unvoiced /t/ followed by the /a/ vowel, this average **VOT** would be  $+14.03$  ms, with voicing starting after the release of the stop consonant ([Klein, 1999](#)). As the **VOT** value of the consonant /d/ used our experiment was far from the reported average value (but inside the range reported in [Klein \(1999\)](#)), the intelligibility of the continuum was measured in pilot tests with three different people (two males and one female, being one of them musician). Their psychometric curves did not show large distortions and were within the expected shape for people with normal hearing, with inflection points near the center of the continuum and adequate slope. Thus, there was no reason to discard the continuum obtained in the experiments.

The morphing of the sounds was done in different ways for the three parts of the signal indicated

by the markers:

- In the period before the consonant release, the duration of the voicing murmur was continuously reduced in time, from its full duration for /da/ until it disappears for /ta/.
- The next 16 ms after the consonant release were linearly interpolated, such that the initial oscillations of the vocal folds for vowel /a/ were morphed into the air burst for /ta/.
- For the remainder of the stimulus (the vowel part) it was used the corresponding part of the signal from the /ta/ syllable (160 ms in the original version with all the glottal cycles as in the original /da/ syllable).

As the vocalic part is the same for all stimuli (being identical from the third glottal cycle) it is possible to rule out exogenous interpretation to the neurophysiological phonemic categorization not related to the VOT.

Half of the stimuli in both continua were generated with a lag of 180°, aiming to cancel out cochlear microphonics effects during acquisitions.

The earphones used in the EEG acquisitions of this work are illustrated in Figure 35. This is an active noise-canceling in-ear earphone powered by an AAA battery, model ATH-ANC33iS from Audio-Technica®. According to the manufacturer its frequency response ranges from 20 to 20 kHz, as shown in Figure 36, with active noise reduction up to 20 dB and 105 dB SPL sensitivity.

## 4.5 Acquisition procedure

Eleven healthy Brazilian participants, five males and six females, aging between 21 and 50 years old (average  $28 \pm 9$  years old) participated voluntarily in the experiment. All participants claimed to have normal hearing abilities and were right-handed with an average mark of 76.8 for the right hand according with Oldfield's laterality index (Oldfield, 1971, Cohen, 2008). All participants were previously informed about the procedures and tasks of the experiment and provided written informed consent to participate in the study. Their participant's data is preserved and they have the right to request the withdrawal of their experiment data at any time. The experiment has been carried out in accordance with the local ethical committee of the Federal University of Minas Gerais, Brazil (COEP-UFMG Brazil - Number: 3.660.444). The entire experiment was performed inside a sound-proof booth with an average background noise of 33 dB SPL.



Figure 35 – Earphone ATH-ANC33iS Audio-Thecnica<sup>®</sup>.

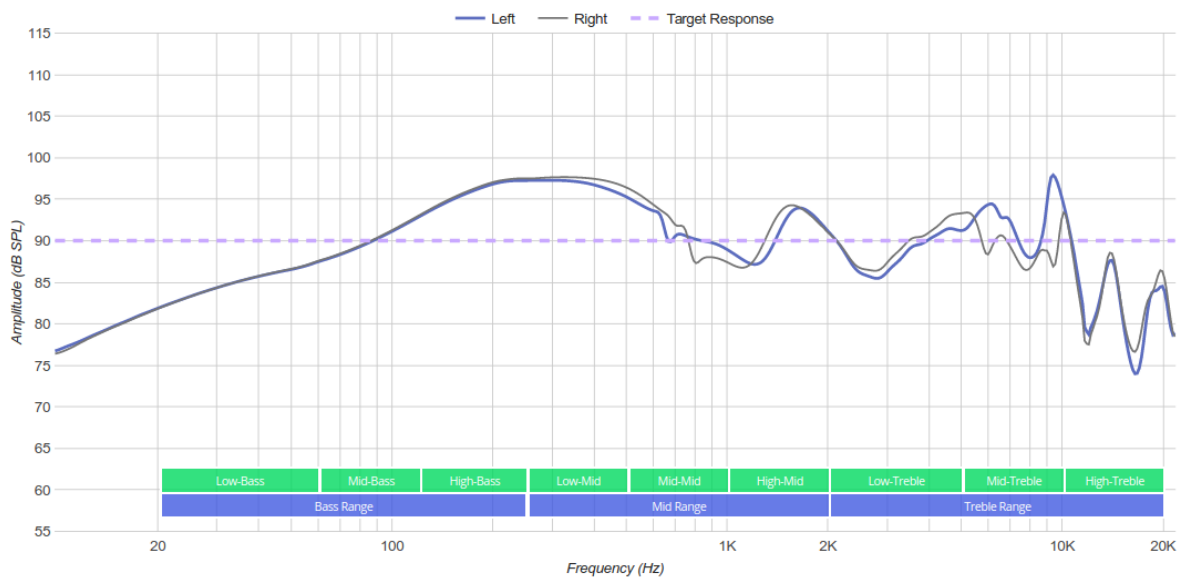


Figure 36 – Frequency response of the Earphone ATH-ANC33iS Audio-Thecnica<sup>®</sup> (averaged and compensated).

CREDITS: <https://www.rtings.com/headphones/reviews/audio-technica/ath-anc33is> (adapted)

The signal acquisition process adopted in this work consisted of acquisitions made on two distinct days, not necessarily consecutive ones, each taking approximately two hours, for a total of four hours per participant. On each day (of the two for each subject) a different continuum was used and the order of the continua was randomized among the participants, that is, for a given participant the continuum /da-/ta/ was the one used in their first day of acquisition and the /pa-/pɛ/ on the second day. For another participant /pa-/pɛ/ was presented on the first day and /da-/ta/ on the second. Each experiment day consisted of three stages: survey of the psychometric curve, passive hearing acquisition (passive stage) and acquisition with an active categorization of the stimuli (active stage). The active and passive stages were also randomized between participants and days of acquisition, that is, if a particular participant performed on the first day the passive stage followed by the active, on the second day this order was reversed. This procedure was applied to both continua in the same way. For the sake of brevity, the four acquisition conditions mixing continuum and task will be called from now on: VOT-act, VOT-pass, Form-act and Form-pass.

For the acquisition of [AMLR](#), about 100 sound repetitions are recommended, considering high intensities and quiet patients with normal hearing while for [AELR](#) 50 repetitions are recommended under the same conditions ([Hall, 2007](#)). [Bidelman and Walker \(2017\)](#) in the investigation of the neural correlates of categorical perception using a continuum between /u/ and /a/ used 200 repetitions of each sound. For a more robust response, more repetitions are desirable, however, given the participant's fatigue as well as the need to be as quiet as possible, it becomes necessary to limit the total time of the experiment. As will be detailed next, from the psychometric curve stage, 5 stimuli out of the 200 will be selected to be used in the passive and active stages. Thus, in this work, 350 repetitions were used in the passive stage, out of a total of 1750 stimuli, whereas 200 were used in the active stage, out of a total of 1000 stimuli.

The number of trials used in the active and passive stages were calculated based on the mean response time observed in [Bidelman and Walker \(2017\)](#), which was 500 ms. It was also considered an [ISI](#) greater than 1 second to obtain larger amplitude in waves N1 and P2, as well as the maximum number of repetitions and taking into account the effect of experiment fatigue on the participants ([Hall, 2007](#)). Based on the total acquisition time of the active stage, taking into account the previous observations, the number of stimuli in the passive stage was defined, in which it was possible to perform more repetitions in the same expected average time for the active stage, which is around 35 minutes

The presentation is binaural because it is more natural for the participants, and also because of the loudness summing effect that gives a perception of a stronger sound than the monaural presentation (estimated in more 6 dB) ([Skoe and Kraus, 2010](#)). A script to measure participants' hearing threshold was made with the initial idea of adding a fixed value, in dB, to each one's

threshold. However, after some testing with four different participants, the volume required for trigger activation was not reached, so this procedure to make the definition of the sound intensity was aborted. The sound intensity was then set at 80 dB SPL considering in the measurement an external acoustic meatus of 3 cm. This distance was maintained between the decibel meter and the earphone which were connected by a tube. Measurements were made with a Polimed PM-1900 decibel meter previously calibrated with a G.R.A.S. 42AB at 1 kHz and 114 dB SPL. This sound intensity was adopted because it is the minimum value required for the trigger circuit to detect the signal. The participants of the experiment reported no discomfort with such sound intensity.

After explaining the experiment to the participants, they were instructed to sit comfortably and then put on the earphones. Prior to each step, the procedures were briefly explained again. The three steps used in the acquisitions are described next.

### 4.5.1 Psychometric curve plot

In this stage the participant was presented with 200 stimuli from the continuum. Each stimuli was presented once in randomized order. The participants were asked to identify the heard syllable as either /pa/ or /pɛ/ (or /da/ and /ta/ for the other continuum), in a forced-choice task, using two specific keys in the keyboard. During the presentation of the acoustic stimulus a screen with a cross was displayed staying for the duration time of the stimuli plus a jitter normally distributed between 0 e 0.3 s. This delay helped to avoid temporal overlap between perceptual processes and motor effects due to button press. After that, the response screen is presented with the syllables /pa/ and /pɛ/, with one at the right and the other at the left of the screen, corresponding to the more right and the more left keys of the keyboard, respectively. The syllables also changed positions at the screen randomly, so that there was not a specific key for each syllable at the keyboard.

For each stimulus, the participant had 2.5 s to respond, otherwise the response was discarded and the next stimulus was presented. The time of one second plus a jitter evenly distributed between 0 and 0.5 s was applied after the participant's response before the new stimulus. At every 20 ranked stimuli, the participant is given time to rest before starting another series of 20. This visual and interactive part of the experiment was programmed in python using the features of the *Expyriment* (Krause and Lindemann, 2014) library.

Psychometric curves were fitted to the responses of each participant using R's *glmRob* function (R Core Team, 2014) for a binomial distribution with a treatment for outliers. Stimuli corresponding

to 0%, 5%, 50%, 95% and 100% of the theoretical psychometric curve of each participant were selected for the next stage with the EEG acquisition. As the stimuli are chosen based on the perception of the participant, the stimuli which are really ambiguous for the participant are selected, differently from what is done in many studies about categorical speech. These stimuli will be referenced as *stim1*, *stim2*, *stim3*, *stim4* and *stim5* from now on. The values 5% and 95% were chosen because they consist of positions close to the categorical transition point (center of the psychometric curve corresponding to 50%) and with some distance from the original stimuli at the edges of the curve. Using these criteria, different values could have been used, such as 10% and 90%, for example. A script in R contains the function for the psychometric curve generation and selection of five stimuli whose numbering (in the range of 1 to 200) is recorded in a .dat file. Those numbers (being a different set for each participant) are used in the following steps. A similar script is used for the /da/ and /ta/ stimulus. Figure 37 shows the psychometric curve obtained for one representative participant for the /pa/-/pɛ/ continuum.

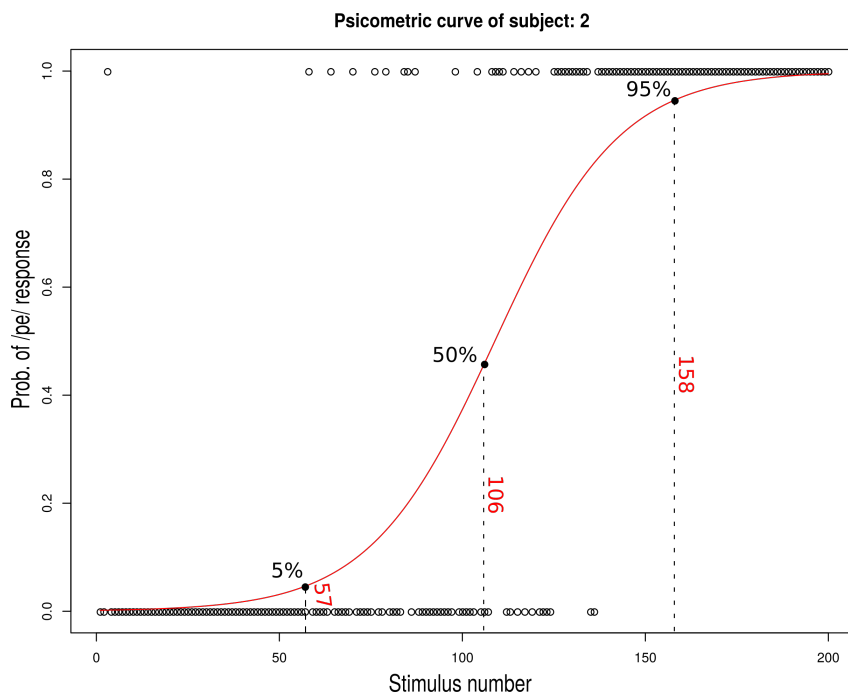


Figure 37 – Psychometric curve of the /pa/-/pɛ/ continuum for a representative participant.

## 4.5.2 Active stage

In the active phase of the experiment, the five selected stimuli on the psychometric curve were repeated at random 1000 times in 5 blocks of 200 presentations. During the rest time between the blocks, chosen by the participants themselves, the experimenter talked a little with them about random topics to help the ears rest from the repeated stimuli heard in each block. After each stimulus was presented, the participant had to classify it into /pa/ or /pɛ/ (or /da/ or /ta/



in the other experiment) using two specific keys on the keyboard, i.e., each key now assigned to a specific syllable, differently from what happened in the task for the psychometric curve. This avoids eye contact with written letters which could evoke potentials related to language processing. Recent studies with fMRI show that some brain areas involved in auditory language processing are also involved in reading (Moerel et al., 2012). Some participants chose to keep their eyes closed during the experiment. Apparently this helped them focus on the sound and classify it better as some of them reported.

Participants were instructed to respond as quickly as possible so that the experiment would not be too long and tiring. However, no time limitation was imposed for giving the response after the presentation of the stimulus. After the response, a random time was given before the presentation of the new stimulus avoiding a possible imposition of rhythm by the participant himself. This time was calculated based on the participant's response time. If the time was less than 1 second the waiting time was calculated according to the first item in Equation 4.3, if the participant took more than one second to answer, the calculation is given by the second item. Both conditions depend on the participant's response time (variable *response.time*) and a uniformly distributed jitter with an average value of 300 milliseconds (to avoid adaptation effects).

$$sleep.time = \begin{cases} (1.5 - response.time) + jitter, & \text{if } response.time < 1 \\ 0.9 + jitter, & \text{if } response.time > 1 \end{cases} \quad (4.3)$$

### 4.5.3 Passive stage

In the passive stage, each of the 5 selected stimuli from the psychometric curve was repeated 350 times at random in a total of 1750 stimuli. The stimuli were repeated at random and, while listening, the participants were distracted from the acoustic stimuli by being presented with a series of silent videos. The videos had an average duration of 5 minutes. At the end of each video, the participants must press the space bar on the keyboard in front of them to start the next video. This was done to ensure that the participants had some activity to perform from time to time to prevent them from becoming drowsy and sleepy. The videos had varied themes but lacked any linguistic stimulation either by letters or symbols, lip reading or gestures. Some of the themes in the videos include cityscapes, recipe making, painting and drawing, recyclable bottle crafts, aquatic animals, developing plants, among others.

After starting the playing of each stimulus, the algorithm waited for around 1 sec plus a uniformly distributed jitter with an average value of 150 ms. This avoids adaptation effect that can elicit

unwanted evoked responses and is long enough to avoid the overlap of mid and long latency auditory evoked responses.

#### **4.5.4 Comment**

This chapter described the experiments carried out to acquire physical and psychophysical data. Based on the auditory perception and phonemic categorization theory presented in previous chapters, these data are processed and analyzed in the next chapters.

# Chapter 5

## TIME DOMAIN PROCESSING

In this chapter we explain how the acquired potentials were processed from cleaning to the generation of [DWT](#) coefficients and how amplitude and latency analysis of N1 and P2 ERPs were performed through multiple comparison of mixed-effects models and contrasts.

### 5.1 Filtering

#### 5.1.1 Notch filter

A notch filter attenuates a specific signal frequency and is used to attenuate the line noise artifact, 60 Hz in Brazil, that could contaminate an [EEG](#) acquisition. As filters generally do not perform ideally, some frequencies close to the notch frequency are attenuated too. An increase in filter order may decrease the width of this attenuation, but it is worth analyzing if the content of these frequencies is relevant since higher order filters may become unstable.

In the experiments carried out, all channels were filtered with a 60 Hz notch filter and filtered again (by reversing the previous filtered signal) to cancel phase distortions. The notch filter is an infinity impulse response (IIR) filter with a bandwidth of 4 Hz whose frequency response is shown in [Figure 38](#).

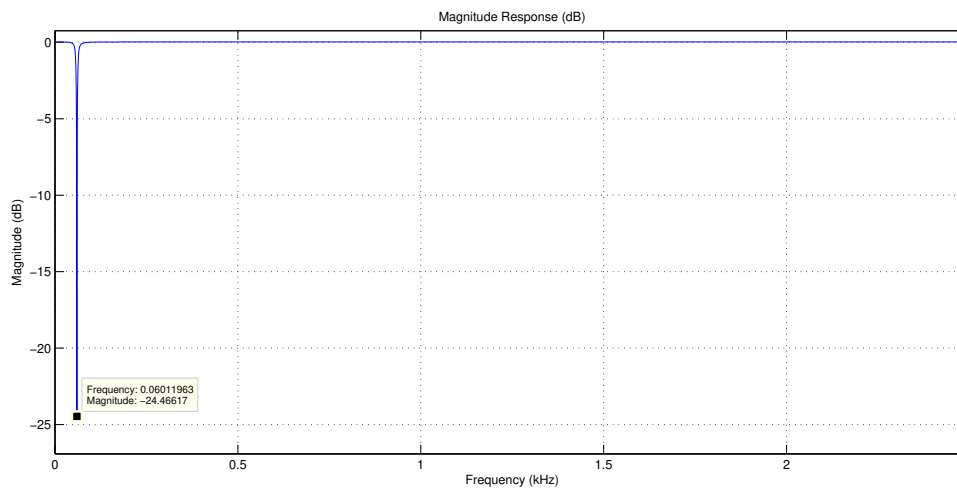


Figure 38 – Frequency response of the 60Hz notch filter applied in the data.

### 5.1.2 Discrete wavelet transform filtering

The **DWT** was used in this work for time-frequency representation of the data and for filtering. The use of **DWT** provides a non-overlapping representation that is statistically manageable, since the wavelet basis functions are orthogonal. This guarantees that the wavelet coefficients are nearly uncorrelated. This could also be attained with the Fourier Transform, but the **DWT** allows a parsimonious representation of the data on the time-frequency domain.

The **DWT** routine was implemented in the software R ([R Core Team, 2014](#)) using the package *wavelets*. Appendix D presents an introduction to **DWT**. The mother wavelet used was a Daubechies orthonormal compactly supported wavelet of length 8, least asymmetric family. This wavelet present a phase response that closely resembles a linear phase filter which makes it easier to align the coefficients with the original sinal in the time-domain. The length is enough to extract the main features of interest in our investigation and the reconstructed signal (using the inverse **DWT**) was very similar to the original one. For the time-frequency domain analysis, nine frequency levels of decomposition were used: eight of detail coefficients (with the highest frequency one being the W1) and one of approximation coefficients (lowest frequency level being at V8). This was the default number of levels in the package *wavelets* and we kept it because it was enough to provide the frequency we chose to perform the filtering of the signal as will be described below. Considering the 5 kHz sampling rate used in the acquisitions of this work, the frequencies covered in each band were organized as in Table 1.

Table 1 – Discrete wavelet transform levels and frequencies used in time domain analysis.

Band	W1	W2	W3	W4	W5
Frequency	2500 ~1250	1250 ~625	625 ~312.5	312.5 ~156.25	156.25 ~78.13
Band	W6	W7	W8	V8	
Frequency	78.13 ~39.06	39.06 ~19.53	19.53 ~9.77	9.77 ~0	

The **DWT** filtering consists of the decomposition of the signal into discrete frequency bands (i.e, the **DWT** of the signal), removal of the frequency bands that should be filtered out of the signal and then the computation of the inverse discrete wavelet transform (**iDWT**) to obtain the filtered time domain signal. The removal step is accomplished by simply zeroing the wavelet coefficients of the bands.

For the frequency domain analysis the cutoff frequency was the W5 band (156 Hz) which is enough to include the main brain oscillations that are often referenced in studies of speech categorization ([Bouton et al., 2018](#), [Bidelman, 2014](#)) and also the F0 of the speaker who provided the acoustic stimuli used in this work. The **SNR** obtained after filtering was 12 dB. For the time domain analysis the cutoff frequency was the W7 band (39 Hz) which is enough to allow the computation of the magnitudes and latencies of the main **AELR** peaks analyzed in this work, namely N1 and P2 waves. Here, the **SNR** was 13 dB.

Tests with other two filtering approaches were performed and are detailed in the Appendix B. One is a Butterworth filter of order 6 and the other is the combination of the **DWT** filtering with the wavelet thresholding approach. There was no significant difference between the tested methods. Thus, in the time-frequency and time domain analysis of this work the wavelet filtering without thresholding was adopted for simplicity considering that we are already performing the wavelet transform of the signals for the frequency domain analysis.

## 5.2 Data preprocessing

After the acquisitions, four **EEG** (.rhd extension) files were generated for each participant, one per experimental condition: **VOT**-pass, **VOT**-act, Form-pass and Form-act. Those files contained the signals from the electrode channels and the trigger signal channel. Each electrode channel was filtered with a notch filter of 60 Hz and then referenced to the channel at point Cz. For each condition, a file (in Python) was generated with information of the acquisition including the number of the 5 stimuli presented at each trial. With that information and the trigger signal it was possible to identify each trial and assemble matrices with epochs going from –150 ms to

850 ms in a total of 5000 samples (1 s sampled at 5 kHz). The matrices of epochs were organized according to the EEG signals for each participant, continuum, stimulus, electrode and task type. The first 150 ms of each epoch constitute a pre-stimulus period (EEG signal before the acoustic stimulus), which was used to perform the baseline correction for each epoch individually.

Trials which were over a given threshold were eliminated (around 91.5% of the trials were kept).<sup>1</sup> This threshold value varied among participants and acquisition condition (different impedances in different days of acquisition). Values range between 45  $\mu\text{V}$  and 90  $\mu\text{V}$ . The determination of the threshold was performed by visual inspection of the signals. Under no circumstances more than a third of the trials were eliminated during the bad trials removal as recommended by Picton et al. (2000). Even in the worst cases, this value was, on average, within one fifth of the total number of trials.

Different procedures were adopted following the processing for the time or time-frequency domain. For the time domain analysis, after the baseline correction, the number of samples for each epoch was changed in order to exclude the samples from the pre-stimulus (now the first sample corresponds to the time 0 s).<sup>2</sup> We worked with 2048 samples which correspond to approximately 410 ms of signal. This is enough to cover the time frame, reported in the literature, where the categorical perception is expected to occur (Bidelman and Walker, 2017, Bidelman et al., 2013, Bouton et al., 2018).

The DWT was applied to each epoch resulting in a vector of wavelet coefficients with the same amount of samples of the epoch (in this case 2048). Thus, we now have DWT matrices for each participant, continuum, electrode, stimulus and task type. For example, for participant 9, VOT continuum, TP10 electrode, active task, the stimulus 3 has 185 valid trials (after removal of bad trials) which resulted in a DWT matrix of  $185 \times 2048$  discrete wavelet coefficients.

Different processing steps were applied to the obtained DWT matrices in accordance with the domain of analysis: time or time-frequency domain.

R (R Core Team, 2014) and MATLAB<sup>®</sup> (MATLAB, 2014) were used in this processing, including the R package “wavelets” (Aldrich, 2013).

<sup>1</sup> For participant 2 and 8, the amount of trials kept in the VOT active acquisition was around half of the intended 200 due to technical issues during the acquisition, however the SNR of the averaged signal was good enough for us to keep these data.

<sup>2</sup> Exceptionally, for the graphic representation of the grand averages, the baseline samples were maintained for a better visualization of the plot.

### 5.3 Time domain analysis

The time domain analysis of the processed responses consisted of verifying the general characteristics of the grand average AELR for each condition, and analyzing the effects of different factors over the amplitudes of N1, P2 and N1-P2 complex extracted from the AELR for each participant. Those factors include type of task (passive/active), continuum (VOT/Formants), stimulus and electrode.

In this work, we examined specifically the amplitudes and latencies of the N1 and P2 waves that compound the AELRs. N1 and P2 and their latencies are reported in the literature as being related to categorical coding (Bidelman et al., 2013, Bidelman and Walker, 2017, Alho et al., 2016, Möttönen et al., 2014, Chang et al., 2010), and to the perception of speech sound cues (Picton et al., 1999, Altmann et al., 2014a, Manca et al., 2013, Tremblay et al., 2001). Some studies suggested that the neural correlates of categorical perception emerge around the time-frame of N1 and are fully manifested by P2 (Bidelman et al., 2013, Bidelman and Lee, 2015).

Graphs were obtained for the average amplitude and latency values for N1, P2 and N1-P2 to stimulus, electrode and acquisition condition, encompassing the responses of all participants. The amplitude of the N1-P2 complex was also evaluated in relation to each stimulus in each condition and also with respect to the maximum slope of the psychometric curve ( $\beta$ ). The acoustic-phonetic transformation, necessary to generate the categorical perception of an acoustic stimulus, seems to occur between the latency of the N1 and P2 waves that compose the event-related potential (ERP) (Bidelman et al., 2013, Alho et al., 2016, Möttönen et al., 2014, Chang et al., 2010, Chevillet et al., 2013). Many studies use the N1-P2 complex in their analysis, but it is important to evaluate the N1 and P2 waves separately because several temporally overlapping, spatially distributed neural sources contribute to scalp recorded potentials in the latency region of those waves, and their magnitudes are influenced by different factors (Crowley and Colrain, 2004). For example, Silva et al. (2020) showed that different subcomponents of the same wave (N1 or P2) responded to different characteristics of the continuum (/i/-/e/) or of the task, revealing that distinct underlying processes are at work during speech sound perception.

All the analyses were performed for the temporal electrodes (Tp9 and Tp10 referenced at the non-inverting Cz electrode) and also for the frontal electrodes (F7, F8 and Fz also referenced at Cz). The temporal electrodes reflect the activity near the primary and secondary auditory cortices whilst frontal and frontocentral electrodes reflect activity at the motor cortex, which is related to the speech processing through the sensorimotor integration of auditory and motor structures (Alho et al., 2016, Möttönen et al., 2014, Chevillet et al., 2013) and also is related to the categorical perception of speech (Bidelman and Walker, 2017).

### 5.3.1 Processing for time domain analysis

DWT transformed epochs from each participant were averaged by stimulus, electrode, continuum and task and then filtered at 39.06 Hz in accordance with the description presented in section 5.1.2. For example, for participant 1, in the acquisition condition VOT-act at the temporal left scalp region (TP9 electrode), 5 AELR signals were obtained, one for each stimulus. Then, the inverse DWT was applied to the averaged signals to obtain the representation in the time-domain. From each average, the AELR peak amplitudes N1 (latency between 90 and 150 ms) and P2 (latency between 160 and 200 ms) and their latencies T1 and T2 were obtained.

In some cases, N1 or P2 presented a double peak/valley making it difficult to reliably identify the correct peak/valley and their latencies. To deal with this problem, a bootstrap technique was used to obtain the peaks N1 and P2 and the latencies T1 and T2. It was implemented as follows: the epochs matrix for each participant, stimulus, electrode, continuum and task was re-sampled with replacement 200 times to obtain 200 averages of N1, P2, T1 and T2. The median values of these averages were then used to identify N1, P2, T1 and T2 (the median is more robust to outliers than the mean).

Figure 39 illustrates the bootstrap applied to participant 5, formant-active experimental condition, stimulus 1. A double valley can be seen for N1. The bootstrap measurements were divided among these two valleys. This is not a big problem for the measurement of the magnitude of this valley because both are very close in value so that the median of the 200 values is similar to the direct measurement. This happens for the majority of the averages as can be seen in Figure 40 that shows that the direct measurement and the bootstrap measurement for the N1–P2 magnitude are positively correlated with a correlation coefficient of  $r = 0.997$  ( $p < 0.001$ ). However, for the latency measurements this difference is significant. It can be seen in Figure 41 that the bootstrap and direct measurements are correlated with a coefficient of  $r = 0.9$  ( $p < 0.001$ ) but some measurements are too different leading to possibly wrong conclusions about the latency. Specifically for the latencies, when dealing with double or multiple peaks, the bootstrap measurement results in the median latency of the 200 runs which is a value more acceptable than the latency of one of the peaks alone. For the case in Figure 39, the latency computed for N1 was located between the valleys.

After N1, P2, T1 and T2 were measured for all averages (and the N1–P2 computed), mixed-effects models (MEM) were used to analyze the effects of stimuli, electrode, task and continuum over those measurements. They were implemented in R using the packages “lme4”, “lmerTest” and “emmeans” (Kuznetsova et al., 2017, Lenth, 2020). A suggested reference about mixed-effects models using R that present the general concepts about this kind of analysis with examples is



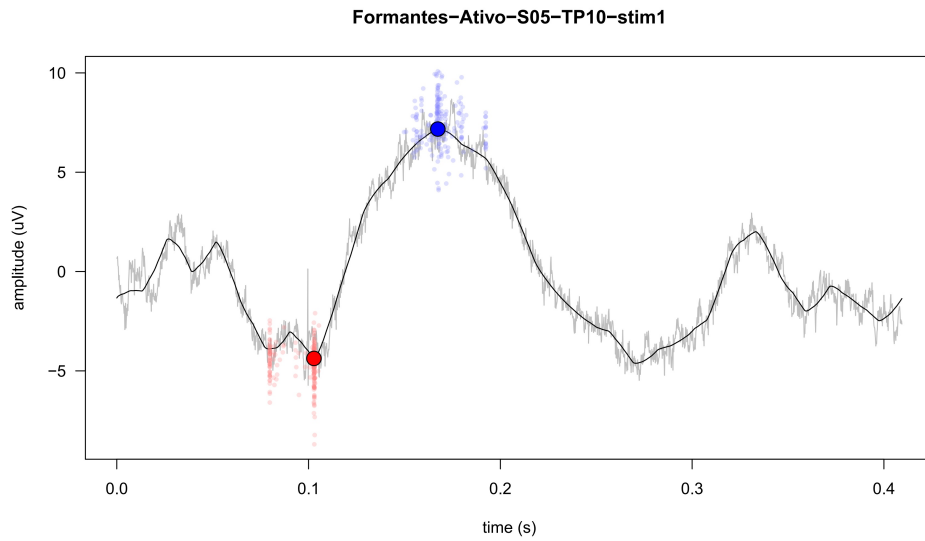


Figure 39 – Bootstrap analysis for the participant 5, formant-active experimental condition, stimulus 1. The epochs filtered average is represented in black continuous line. In gray continuous line is represented the non filtered version of this same signal. Small dots in red and blue indicates the 200 values of N1 and P2 obtained by the bootstrap. Big red and blue dots represent the direct measurement value.

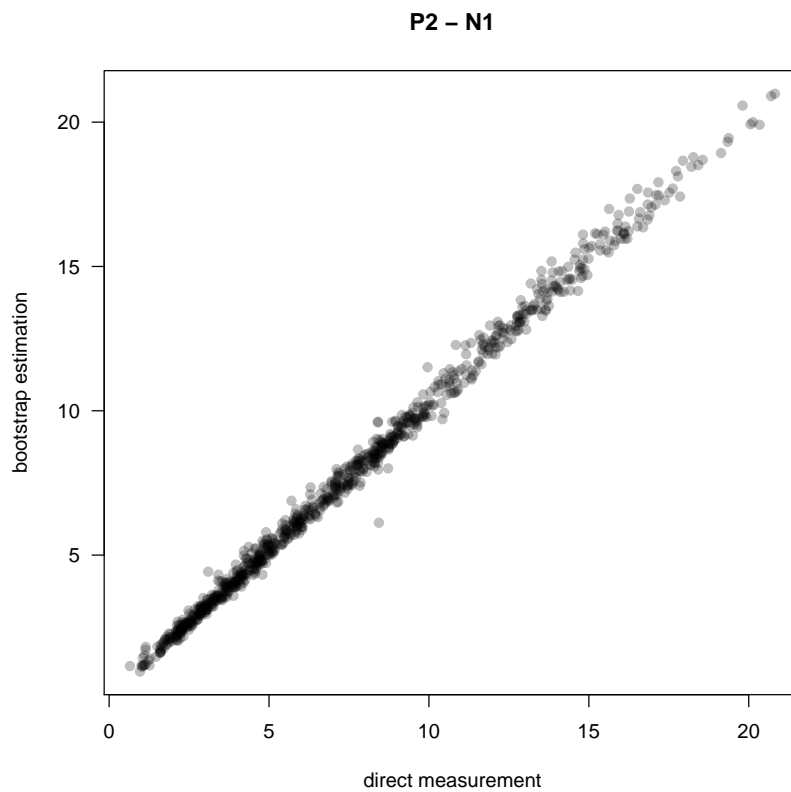


Figure 40 – Values of the magnitude N1-P2 obtained through direct measurement at the AELR averages and through the bootstrap technique for all cases.

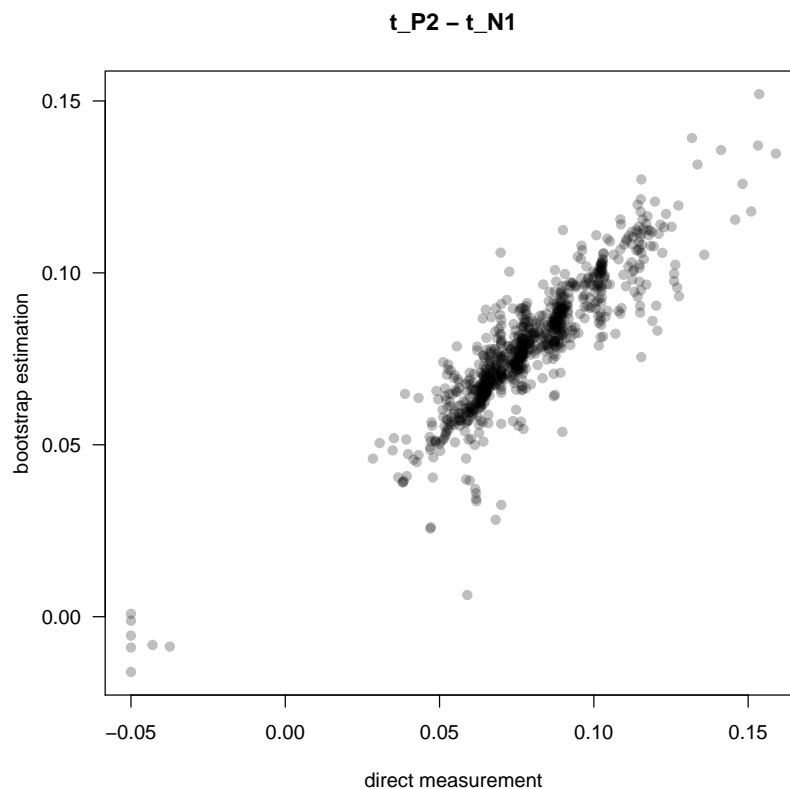


Figure 41 – Values of the latency difference T2–T1 obtained through direct measurement at the AELR averages and through the bootstrap technique for all cases.

Bates (2010). ANOVA (analysis of variance) and RANOVA (analysis of variance for random effects) were used to evaluate the models and find the factors (and factor interactions) with significant effects over the dependent variable analyzed (N1, P2, T1, T2 and N1–P2 values). The assumptions of normality, heterocedasticity and no correlation of the fixed effect predictors were analyzed at the Appendix F. In general, MEMs are quite robust to violations of those assumptions to some level. For the factors with more than two levels and factor interactions, a contrast analysis was applied as a post hoc test to evaluate which and how levels and levels interactions influence the variable analyzed. According to Abdi and Williams (2010) “[...]a contrast expresses a specific question about the pattern of results of an ANOVA” and “[...]corresponds to a prediction precise enough to be translated into a set of numbers called *contrast coefficients* which reflect the prediction.” Thus, for each factor or factor interaction we designed a contrast to express the research hypothesis about the observed effect and test it.

For example, for the contrast to evaluate the effect of laterality for the factor electrode, considering the electrode order [F7, F8, Fz, Tp10, TP9], the contrast [1/2, -1/2, 0, -1/2, 1/2] was used. So, the left electrodes have equal values and also the right ones. The central electrode Fz does not have contribution for the laterality analysis so its value was set to 0. In all contrasts, the sum of the contrast coefficients has to be zero. To evaluate the effects of the stimulus factor, three contrasts

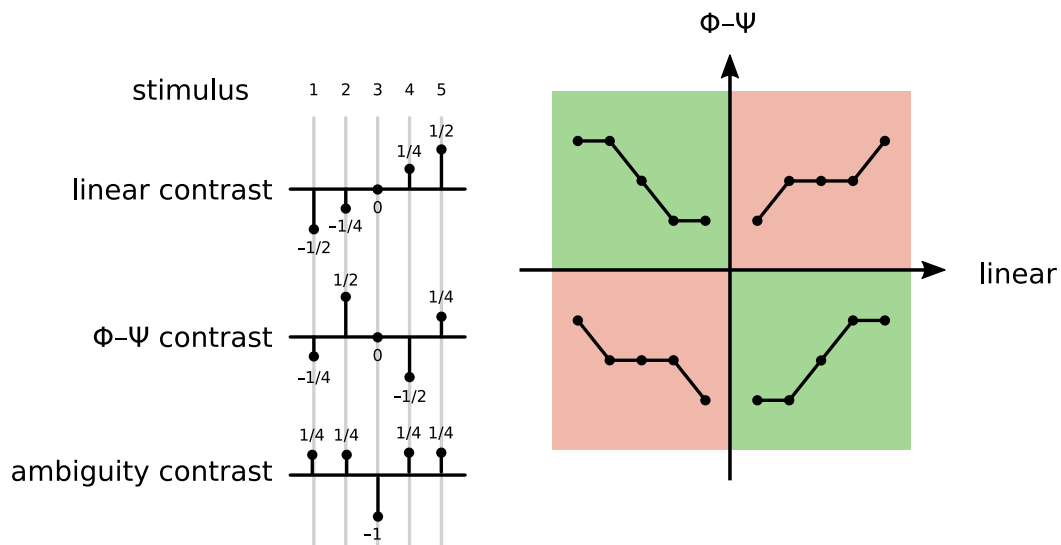


Figure 42 – Linear, ambiguity and psy-phy contrasts.

were used: linear, ambiguity and psy-phy. Those contrasts are depicted in Figure 42. The linear contrast tests whether there is a linear behavior of the dependent variable from stimulus 1 to stimulus 5. The ambiguity contrast tests whether there is a different behavior of the dependent variable for the ambiguous stimulus (stim 3) compared to the remaining unambiguous ones. The psy-phy contrast tests whether the behavior of the dependent variable is more psychophysical, that is, with values for stimuli 1 and 2 closer to each other as well as stimuli 4 and 5; or more physical, that is, with values for stimuli 2, 3 and 4 closer to each other. These three contrast are orthogonal to each other. More details about contrasts can be found in [Abdi and Williams \(2010\)](#).

Grand averages were also computed across participants for each stimulus, continuum, scalp region and task. For the frontal region averages, the electrode F7 signal was used to represent the frontal left region, F8 the frontal right region and Fz the frontocentral region. For the temporal region, electrodes TP9 and TP10 were used for the temporal left and temporal right cortical regions, respectively. R software ([R Core Team, 2014](#)) and MATLAB<sup>®</sup> were used for processing ([MATLAB, 2014](#)).

### 5.3.2 $\Delta$ ERP test

Considering the case of formant based stimuli, [Bidelman and Walker \(2017\)](#) derived a variable from the listeners' N1–P2 magnitude named “ $\Delta$ ERP”. It is computed as the difference between the N1–P2 amplitudes evoked by stim1 and stim5 (original unambiguous) and the stimulus stim3 (ambiguous) calculated as:  $\Delta\text{ERP} = \text{mean}(\text{stim1}, \text{stim5}) - \text{stim3}$ . This variable measures the degree to which neural responses differentiated stimuli with well defined categories (stim1

and stim5) from those heard ambiguously (stim3). [Bidelman and Walker \(2017\)](#) showed that as participants categorized better, the  $\Delta$ ERP increased. In general they noticed, for passive and active tasks, that the N1–P2 amplitude of the [AELR](#) from stim3 was smaller than that of stim1 and stim5. The authors report that the smaller N1–P2 for stim3 may be due to an “inhibition or top-down gating on sensory coding with the increased listening effort for perceptually ambiguous speech” ([Bidelman and Walker \(2017\)](#) apud [Knight et al. \(1999\)](#)). The authors performed measurements in the frontocentral region with a continuum based in formant variations. In this work, we verified if this effect also occurs for our data and expanded the analysis for N1 and P2 separately and also for the case of [VOT](#) stimuli.

### 5.3.3 Averaging test after bad trials removal

As described in Chapter 4, the number of trials adopted for the passive and active steps were different. As some comparisons of the results in each step should be made, it was sought to analyze if the smaller number of trials to perform the average in the active stage would affect the comparison of the waves of this stage with those of the passive stage.

A test was performed equaling the number of trials for each participant in each acquisition condition. Thus, for each participant, there are twenty averages to be performed (four conditions, each with five stimuli). The minimum number of trials used in the four conditions was defined from the number of trials remaining in the active conditions after the removal of bad trials (as they already have fewer trials than the passive ones). After performing this procedure for all participants, the N1-P2 amplitude value was obtained for each one of the 5 stimuli averages applied under the four conditions. For example, for participant 1, [VOT](#)-active condition, the stimulus 1 average was calculated using the minimum number of trials defined, and with this [AELR](#), the amplitude N1–P2 was calculated in  $\mu$ V.

In a second test, the number of trials was evenly matched between the stimuli of each condition but no longer between conditions. This ensures that within the condition, the N1-P2 amplitude calculated for each stimulus was not affected by the number of trials used in the mean, which could affect the interpretation of results, for example by concluding that the amplitude N1–P2 of an unambiguous stimulus is larger or smaller than that of an ambiguous stimulus. This equality of trials between stimuli was also performed in the first test.

In both tests, it was observed that N1-P2 amplitudes, on the average of the participants, are higher for active than for passive conditions and that such amplitudes are higher for [VOT](#) than for formant stimuli. These tests showed that the fact that there were fewer trials to perform the

averages in the active conditions did not affect the relationship of the average amplitudes between the conditions. Thus conclusions can be drawn by comparing them in the analysis performed. The tests also showed that the amount of bad trials removed from each stimulus did not affect significantly the N1–P2 amplitudes.

The larger number of trials used in our experiment is similar to what is recommended for [AELR](#) (considering controlled environment, high intensity and quiet participants, ([Hall, 2007](#))), absorbs the difference between the number of trials used in each condition as well as the trials lost in the removal of artifacts for each stimulus. Thus, the remaining number of trials was maintained for each stimulus, condition and participant after the removal of the bad trials even if there was difference in the number of trials for each average performed.

### **5.3.4 Comment**

In this chapter, we described the processing carried out to analyze the ERPs acquired. In the next chapter, the results obtained will be presented and interpreted.

## Chapter 6

# RESULTS OF THE TIME-DOMAIN ANALYSIS

This chapter presents the results for the analysis in the time domain, performed on the grand averages (across participants) and the mean values of amplitudes and latencies of the N1 and P2 peaks. As explained in Chapter 5, these peaks were obtained from the bootstrap technique.

Those results are presented and commented in relation to the factors in the model: laterality (left/right), attention (active/passive tasks), cortex auditory region (temporal/frontal), and acoustic cue (VOT/Formants). For the mixed-effects models, the we considered the following factors: type (active/passive), feature (VOT/formants), electrode (F7, F8, Fz, TP10, TP9) and stimulus (stim1, stim2, stim3, stim4, stim5).

Data from some electrodes of specific participants in specific experimental conditions were discarded due to the bad quality of the signal. Fortunately, the mixed-effects models used in our analyses are robust to the elimination of such a small amount of data. There was eight bad cases in a total of 220 (2 types \* 2 features \* 5 electrodes \* 11 participants) cases. They are listed below:

- participant 4:
  - Formant-passive electrode Fz
- participant 5:
  - VOT-active electrode F7

- VOT-active electrode F8
- VOT-passive electrode F8
- participant 11:
  - Formant-passive electrode F7
  - Formant-passive electrode F8
  - VOT-active electrode F7
  - VOT-active electrode F8

## 6.1 Response times

Figures 43 and 44 show the mean response time (RT) of all participants at the active task for each phonemic continuum. Considering that each continuum have 200 stimuli, the abscissa axis represents the stimulus number for the mean of the stim2, stim3 and stim4. The stim1 and stim5 are always the first and last stimuli of the continua.

Note that for both VOT and Formants cases, the ambiguous stimulus (stim3), was the one with greater RT. This is expected, because stim3 is the stimulus that should be the more difficult to identify. Following this same logic, stim2 and stim4 should have RTs higher than stim1 and stim5, respectively. This is true for the VOT case. However, for the Formants case, there almost was no difference in RT for stim1 and stim2.

The syllable /da/ was identified faster than the syllable /ta/ in the VOT continuum. In the formant continuum syllable /pɛ/ was identified faster than /pa/.

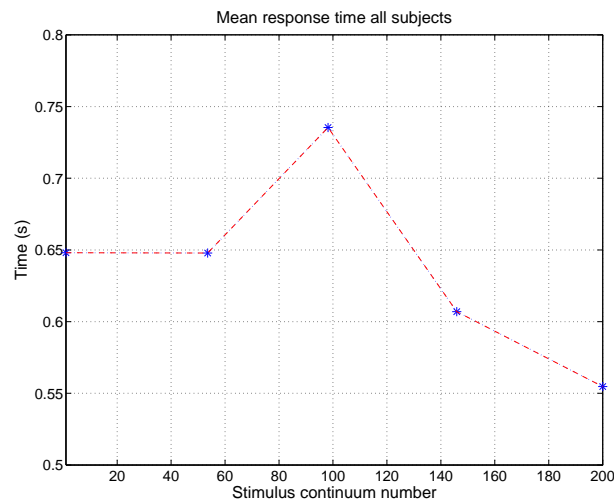


Figure 44 – Mean response time of all participants for the identification of the stim1, stim2, stim3, stim4 and stim5 in the active task of the formant continuum. In the abscissas axis is represented the number of the mean stimuli stim2, stim3 and stim4 of all participants considering the continuum of 200 stimuli.

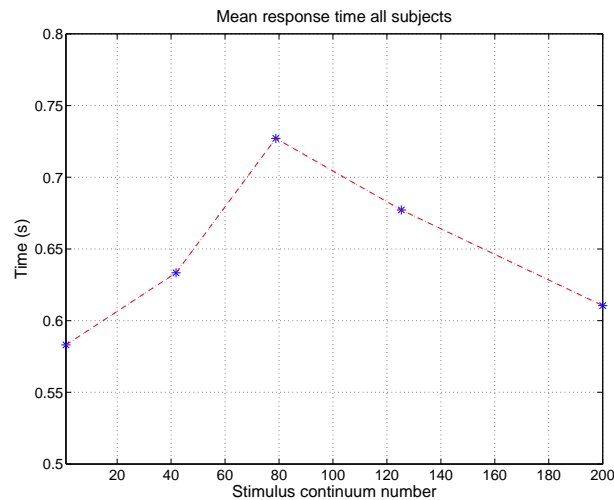


Figure 43 – Mean response time of all participants for the identification of the stim1, stim2, stim3, stim4 and stim5 in the active task of the VOT continuum. In the abscissas axis is represented the number of the mean stimuli stim2, stim3 and stim4 of all participants considering the continuum of 200 stimuli.

For each continuum, a Wilcoxon signed rank paired test was performed using the R software to compare the RTs of stim3 with those of stim1 and stim5 (alternative hypothesis:  $\text{stim3} > \text{mean}(\text{stim1/5})$ ). As expected, the median difference was greater than zero, meaning that, in general, the RT of stim3 was significantly higher than that for the unambiguous stimuli (stim1 and stim5) and for both VOT continuum ( $V = 65, p < 0.0001$ ) and formant continuum ( $V = 55, p = 0.026$ ).



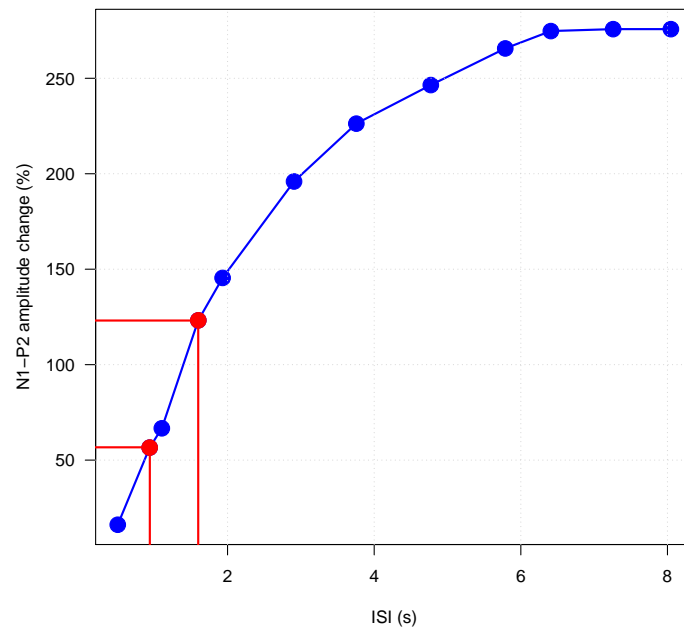


Figure 45 – Projections of the mean ISI for active and passive tasks onto the curve given by Hall III (2015) to compute the mean increase in N1-P2 from the passive to active task. Both projections are represented by red points.

This indicates that, in general, stim3 is more difficult to categorize than stim1 and stim5.

For the active task, the RT influenced the mean interstimulus interval and consequently the magnitude of the AELR. Considering the average RT across participants of  $\sim 646$  ms, for the VOT continuum, the ISI jittered between 1280 and 1880 ms. For the formant continuum, with the average RT of  $\sim 639$  ms, this variation was between 1309 and 1909 ms. In comparison, with the logic implemented for the passive task, the ISI varied randomly (with uniformly distributed jitters) between 780 and 1080 ms for the VOT continuum and 809 and 1109 ms for the formant continuum. Thus, for the active task, the mean ISI (for both VOT and Formant continua) was 1600 ms, while for the passive task it was of 945 ms. This results in an ISI, for the active task, that is 1.7 times greater than that for the passive task.

The curve illustrated in the Figure 9 in Section 2 shows how the ISI influences the N1-P2 magnitude. By projecting the mean ISI for active and passive tasks onto this curve, it is possible to have an approximation of the increase of the N1-P2 amplitude from the passive to active task. These projections are represented by the red points in Figure 45. Using those points, we compute an increase of 2.17 for the N1-P2 amplitude. However, there are also other factors that influence the increase of the N1-P2 besides the ISI as shown in Chapter 2.

In a MEG experiment, Bouton et al. (2018) showed that the acoustic cue was encoded (perceived)

in the left pSTG 150 ms after stimulus onset. This time delay correspond to a latency between the N1 and P2 peaks. As perception is a necessary step before identification, which must precede the motor planing and motor action, then the action of pressing the button in the active task should not influence the response at the latency of the N1 and P2 waves. Furthermore, the average RT reported before is well above the P2 latency for both continua showing that certainly there was no overlap of the responses in our experiment.

## 6.2 Grand averages and psychometric curves

In Figure 46 (left), the mean psychometric curves for the VOT and formant continua are illustrated. On the curve for the formant continuum, the mean value of stim3 was close to the center of the continuum (stimulus 98 from a 200 stimuli continuum). For VOT, there was a shift of the curve to the left, and, on average, the stimulus 79 was classified as the most ambiguous. This stimulus presents a VOT of  $-26$  ms showing that, for Brazilian Portuguese, for a voiced syllable out of context, if the VOT is not negative enough, that is, if the murmur before the release of the stop consonant is not long enough, it is hard to identify the phoneme as voiced.

In Brazilian Portuguese, the average VOT for the voiced dental consonant /d/ followed by the vowel /a/ would be  $-89.15$  ms (in a range from  $-155.05$  to  $-50.20$  ms); with voicing beginning before the release of the plosive consonant. For unvoiced /t/ followed by the /a/ vowel, this average VOT would be  $+14.03$  ms, with voicing starting after the release of the stop (Klein, 1999). This explain the displacement of the VOT average psychometric curve to the left because the continuum begins with the minimum VOT (for stim1) with  $-52$  ms, which is near the minimum value in the range reported in Klein (1999) ( $-50.20$  ms).

Furthermore, since the common vowel /a/ used in the continuum belonged to the original syllable /ta/, we believe that the transitory part of this vowel could influence the tendency to perceive the syllable /ta/ overall in the continuum. The participant may use acoustic cues of the transitory part of the vowel to identify the stimulus instead of just the pre-voicing of the consonant. We believe that this did not affect our analysis in any way. In fact, the stim1, stim2, stim3, stim4 and stim5 used in the analysis were chosen according to their position obtained from the individual psychometric curve. This should not be affected by the location of the 50% threshold in the continuum. Besides, all syllables in both continua have a quite natural sound.

For the formant continuum, it was used the F2-F1 formant frequencies differences from each stimulus. Figure 46 (right) shows the mean F2-F1 difference for all the participants.

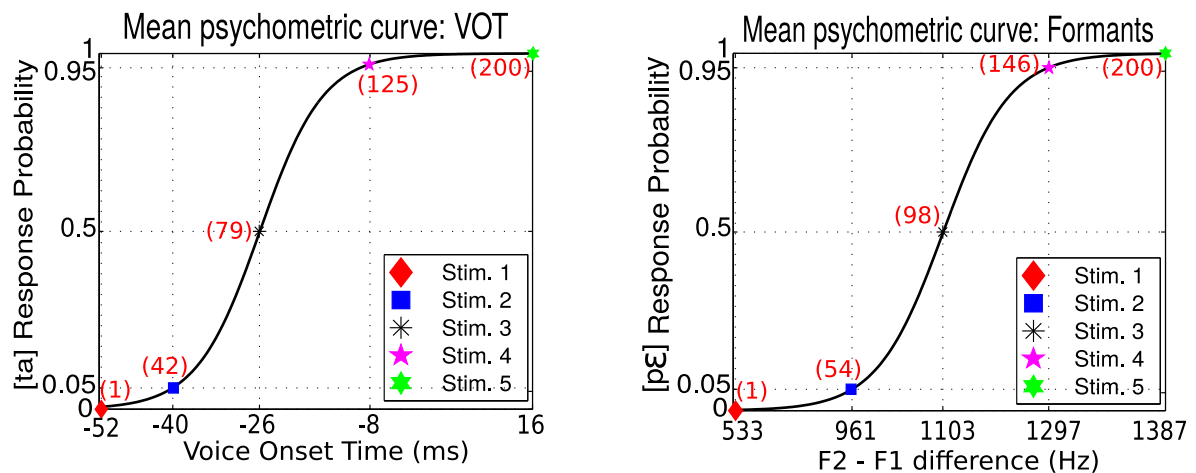


Figure 46 – Average psychometric curves obtained from all participants results for the VOT (left) and the formant (right) continua. Mean VOT for each stimuli are showed in the abscissas axis as well as the mean formant frequencies difference (F2-F1). Number of the stimulus at each continuum (between 1 and 200) are displayed in red.

For a given stimulus, electrode and experimental condition, the AELRs obtained for all participants were averaged together producing what is called the grand average. Figures 47, 48, 49, and 50 illustrate the grand average of the AELR for the left and right hemispheres under the four conditions tested. It is evident, by visual inspection, that the grand average amplitudes of the active conditions have greater values than those of the passive conditions. It can also be seen that, for the same task, the VOT continuum signals have greater amplitudes than those of the formant continuum. This can be related to the stimulus duration that is greater for VOT than for Formants. AELR components generally increased in amplitude with increasing signal duration (Alain et al., 1997).

The N1 wave is often described to be an “exogenous” response, being sensitive to physical characteristics of the sound and reflecting the detection of acoustic changes as the onset of sound (Wagner et al., 2013, Picton, 2013). For example, Toscano et al. (2010) showed, in an oddball (active) task, that the N1 component was affected by VOT but not by phonological category, reflecting the coding of a physical acoustic property by this component. Furthermore, (Horev et al., 2007) showed that the VOT value had a significant effect on N1 latency and also on N1 and P2 amplitudes.

It can be observed that for VOT-passive and VOT-active conditions, there is a lag between the plots that follows the sequence of the stimuli, the /da/ syllable average being the one with the lowest latency. On the other hand, the syllable /ta/ has the highest latency. It is important to mention that the lags observed between each grand average are not the same as the stimuli that evoked them, but are rather related to their perception. With respect to the VOT-passive

condition, we observe a larger amplitude of the N1 wave for stim3, stim4 and stim5 than for stim1 and stim2, indicating a shift in the perception of the consonant in this case. For the VOT-active condition, the N1 amplitude seems to follow the VOT magnitude, but we also note a distance between stim3, stim4 and stim5, on one side, and stim1 and stim2, on the other side, in terms of N1 amplitudes, as in the VOT-passive condition. We also observed a different behavior for the P2 amplitude between the passive and the active conditions. This shows that, at this latency, attention plays a role that can be related to the perception of the VOT.

Curves for the right hemisphere were similar to those at the left for all conditions. ERP amplitudes (in special P1 and P2 peaks) on the left hemisphere were larger than those on the right hemisphere for the active conditions. As stated before, amplitudes for active conditions are greater than those for passive conditions. This can be due to the fact that the ISI for the active tasks were greater than those for the passive ones as discussed before. It is known that longer latency ERPs are dependent on longer refractory times and vice versa (see discussion in Chapter 2).

For the Form-passive and Form-active conditions, there is no lag between the stimuli plots, indicating that the spectral variation performed to go from /a/ to /ɛ/ did not affect the latencies. However, the P2 wave has greater amplitude for stim5 than for the others, while the ambiguous stimulus stim3 has a small P2 amplitude in the Form-pass condition. This suggests that the P2 amplitude codes in some way the perception of ambiguity. This can be related to the work of Ross et al. (2013) and Tremblay et al. (2014), which showed that an improved speech perception after training was associated with an increase in the P2 so that the P2 magnitude is larger for known stimuli than for the ambiguous ones.

In some cases we observed a double N1 peak. This pattern was observed by Sharma et al. (2000) for longer VOT values. Han (2010) analyzed this result and concluded that the first N1 peak is due to the onset of the consonant and the second to the onset of the vowel. However, this would be true only for positive VOT values. As we can observe in Figures 54 and 56 (stim5) there is an aspiration between the release of the plosive consonant and the onset of the vowel. But for the effect observed at Figure 47 (stim1), which presents a large negative VOT, the interpretation can be similar but considering that the first N1 peak is evoked by the prevoicing and the second one by the release of the plosive consonant. Simos et al. (1998) reported that all formants in their work seems to contribute in the N1 wave so that there may be a short window of temporal integration involved in the generation of this response justifying the fact that the temporal distance between the beginning of the prevocing in the /da/ syllable and the release of the plosive is greater than the actual distance between the N1 peaks in the AELRs evoked by this stimulus.

Sharma et al. (2000) and Steinschneider et al. (1999) showed that the double N1 peak does not occur for small VOT values (<30 ms and <20 ms, respectively). However it is important to

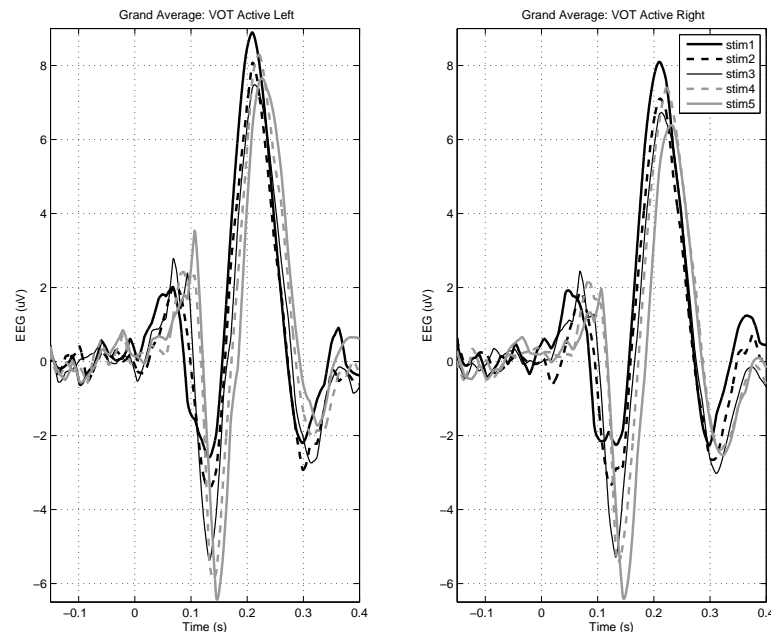


Figure 47 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the VOT-act condition in the temporal cortex.

notice that, in their experiments, they used stimuli in English, for which a positive **VOT** around 20 ms corresponds to a voiced consonant. Thus, we should take into account this difference in perception of voiceless consonants in Brazilian Portuguese whose aspiration is much smaller than their equivalent in English so that a little aspiration as our 16 ms for the /ta/ syllable may be perceived and enough to generate de double N1 peaks.

Figures 51, 52, 53 and 54 show the grand average of the **AELR** for the left and right hemispheres under the four conditions tested for the frontal cortex signals (F7 and F8 electrode sites). As in the temporal cortex data, **ERP** amplitudes (in special of P1 and P2 peaks) at the left hemisphere averages were larger than those of the right hemisphere for the active conditions. Furthermore, it can be observed that the amplitudes of all the grand averages, for all conditions, are smaller in the frontal cortex than those in the temporal cortex. For instance, the N1-P2 amplitude in condition **VOT**-active of the stim5 at the temporal left region is 2.06 times greater than the correspondent value at the frontal left region. As also observed at the temporal region, amplitudes for active conditions are greater than those for the passive conditions. Furthermore, the **VOT** continuum signals also have greater amplitudes than those of the formant continuum but this difference is not so pronounced as at the temporal region.

Figures 55, 56, 57, and 58 show the grand average of the **AELR** for the frontocentral region measured with the Fz electrode. As observed in the temporal and frontal regions, amplitudes

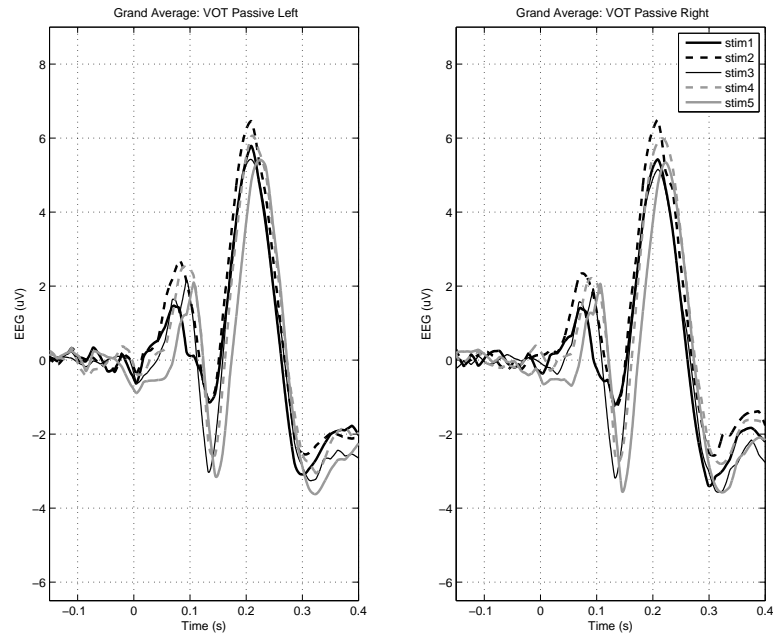


Figure 48 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the VOT-pass condition in the temporal cortex.

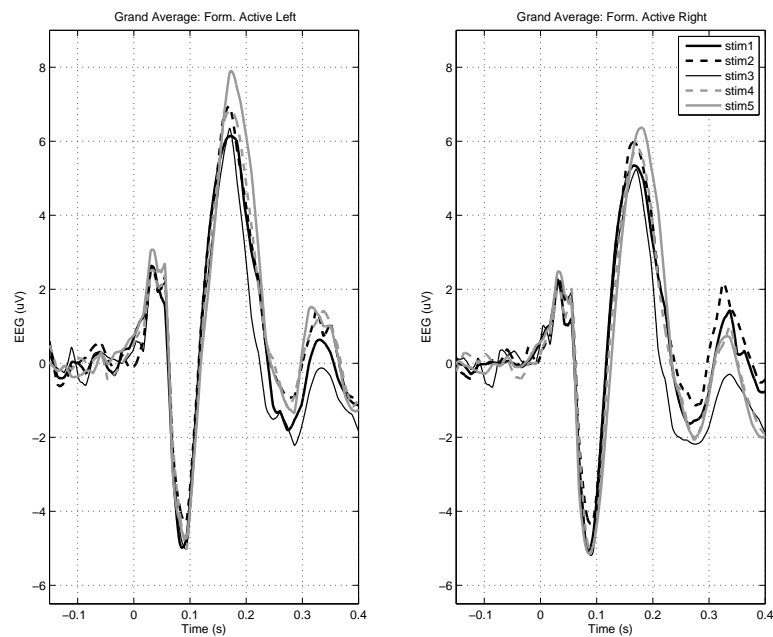


Figure 49 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the Form-act condition in the temporal cortex.

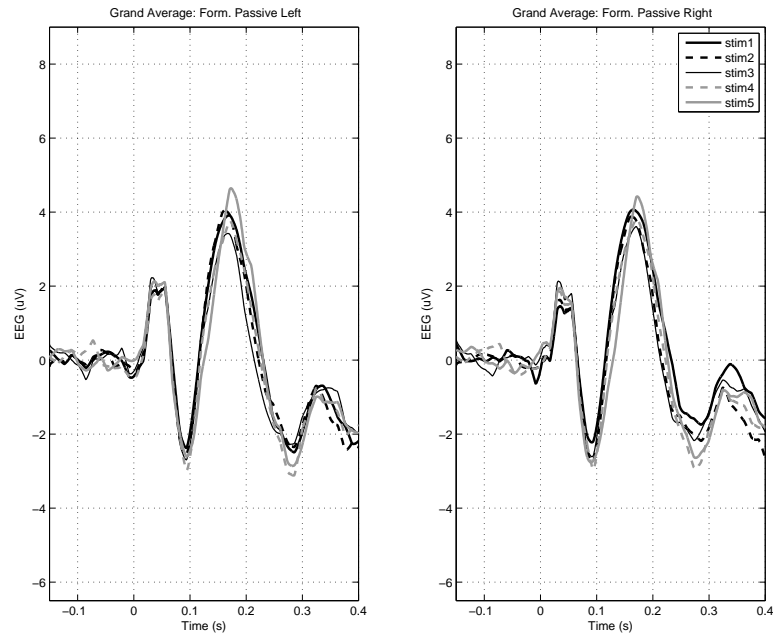


Figure 50 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the Form-pass condition in the temporal cortex.

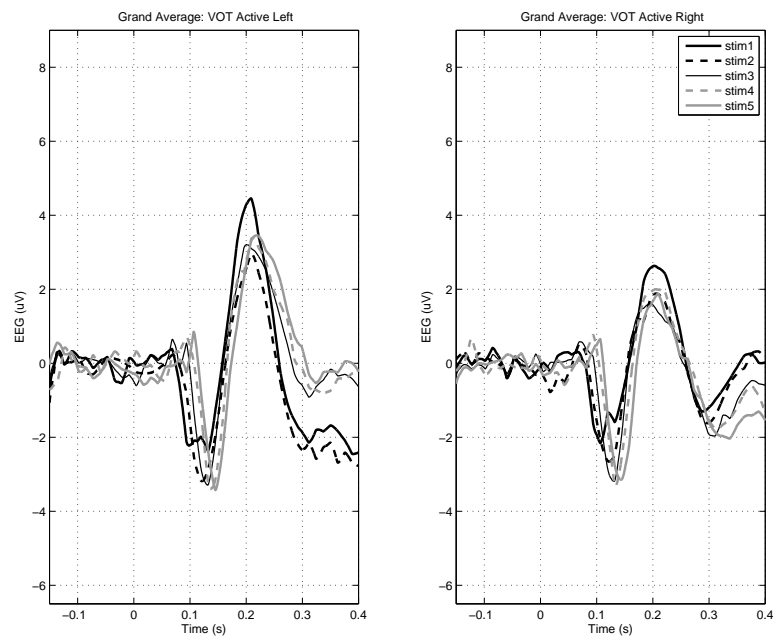


Figure 51 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the VOT-act condition in the frontal cortex.

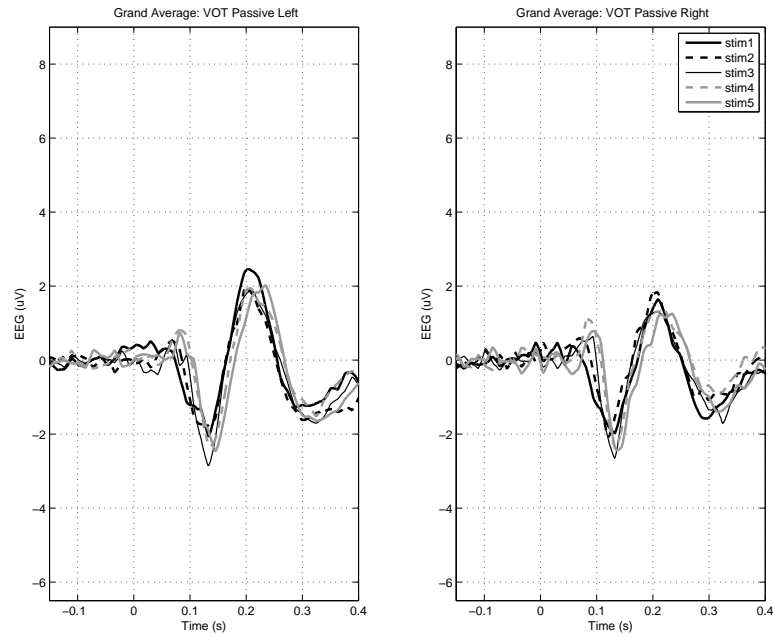


Figure 52 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the VOT-pass condition in the frontal cortex.

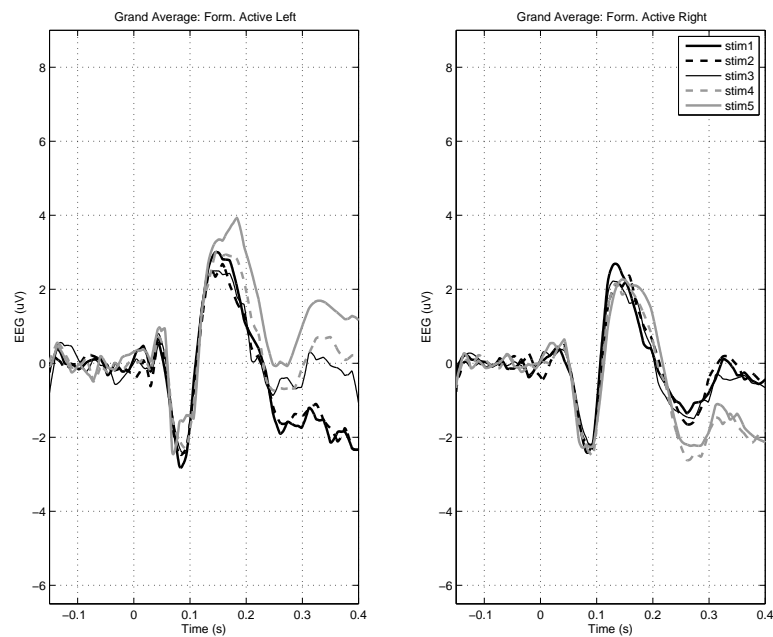


Figure 53 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the Form-act condition in the frontal cortex.



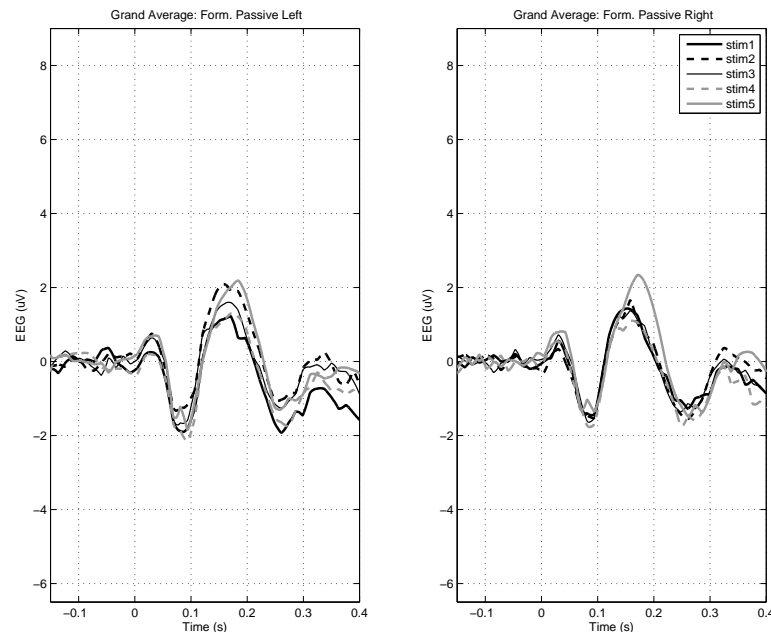


Figure 54 – Grand averages of the auditory long latency potentials for the left and right hemispheres for the Form-pass condition in the frontal cortex.

for active conditions are greater than those for the passive conditions, considering the same continuum. The VOT continuum signals also have greater amplitudes than those of the formant continuum, as observed before, but just for the active task. The amplitudes for all conditions are smaller than those observed before for the temporal and frontal regions. For instance, using the same VOT-active N1-P2 amplitude of the stim5 compared before, this value for the frontal left region is 2.13 times greater than the same one here for the frontocentral region. Thus, it is 4.39 times smaller than the correspondent value at the temporal left region.

### 6.3 Analysis of mean latencies and amplitudes of AELR waves

To make a better evaluation of the effects of the experimental factors on the amplitudes and latencies of the AELR, we fitted a mixed-effects model to the data. The continuum, the stimulus, the task and the cortical region were considered as fixed factors, while the participants was taken as a random effect. The dependent variables used in this analysis are those found by the bootstrap technique.

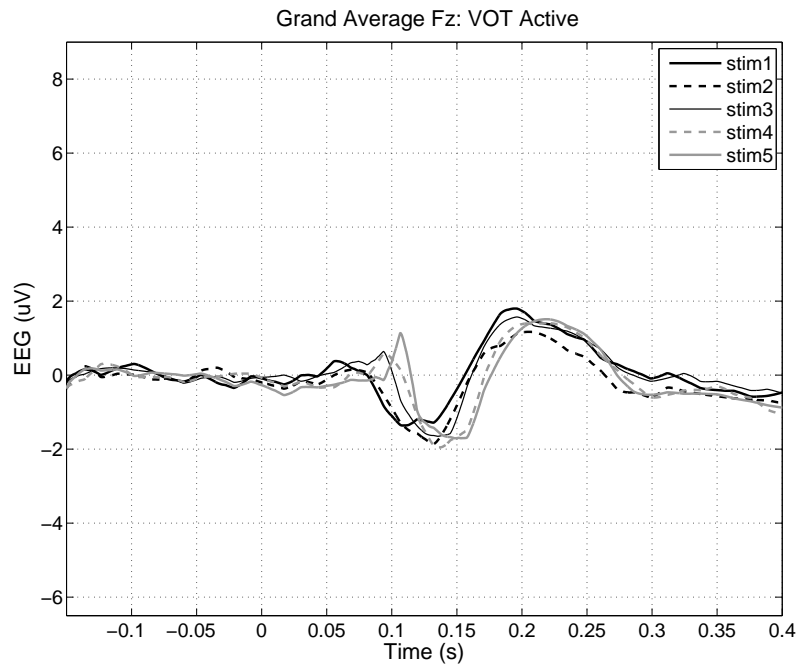


Figure 55 – Grand averages of the auditory long latency potentials for Fz electrode at the frontocentral region for the VOT-act condition.

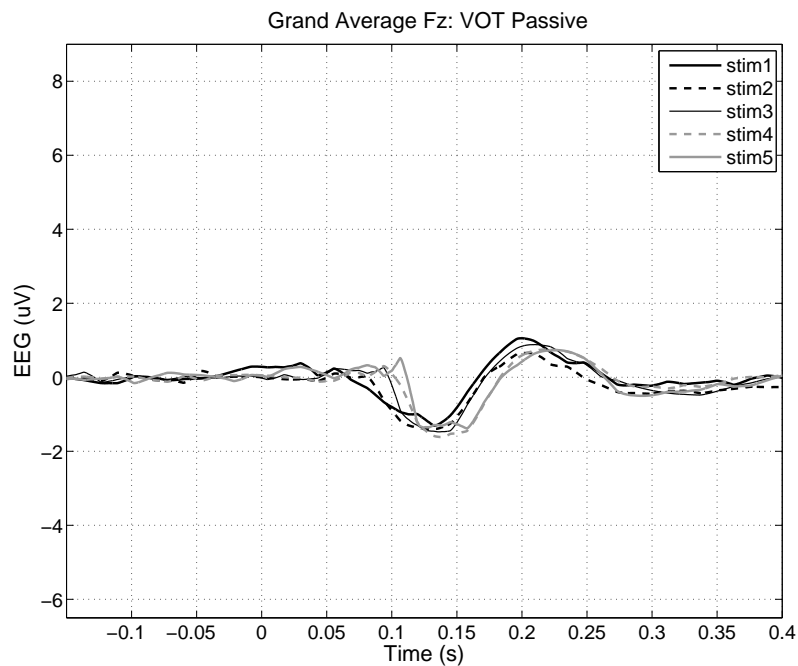


Figure 56 – Grand averages of the auditory long latency potentials for Fz electrode at the frontocentral region for the VOT-pass condition.

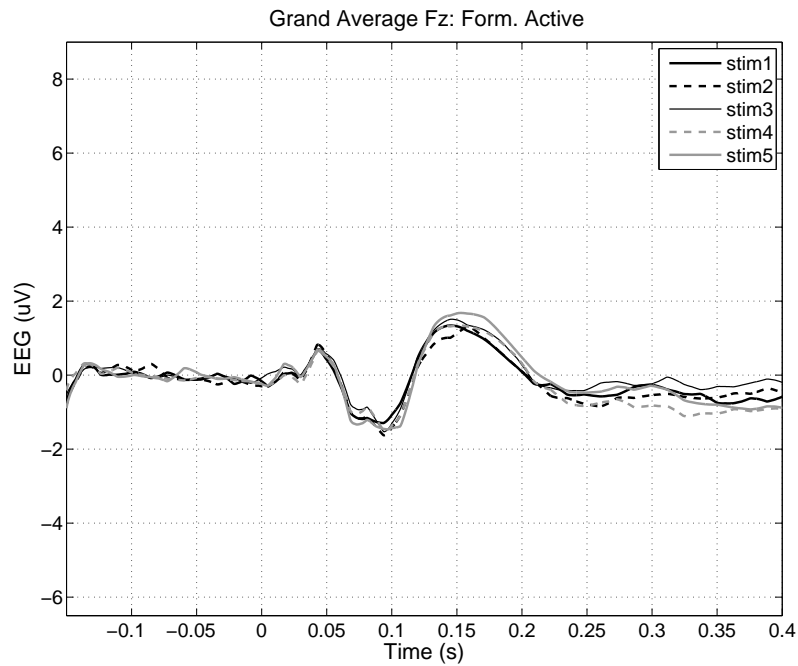


Figure 57 – Grand averages of the auditory long latency potentials for Fz electrode at the frontocentral region for the Form-act condition.

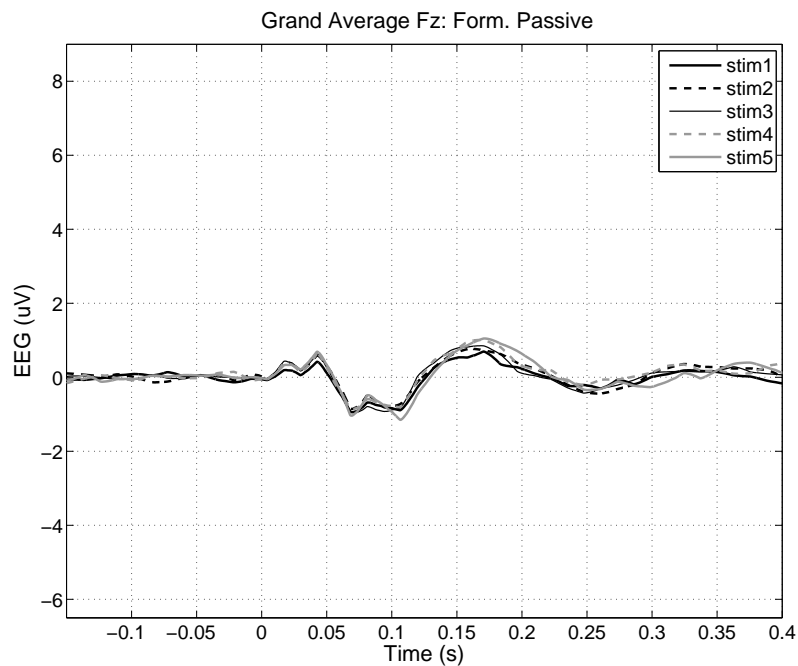


Figure 58 – Grand averages of the auditory long latency potentials for Fz electrode at the frontocentral region for the Form-pass condition.

With this analysis we want to evaluate if our data can confirm some hypothesis that we rised based in the literature review regarding the cortical regions involved in the categorical speech processing, the influence of the acoustic cue in the evoked responses, the influence of attention in these responses and if they somehow reflect the perception of ambiguous stimuli.

Regarding the cortical region, our literature review lead us to assume that:

- There is a left hemisphere dominance for speech processing ([Hickok and Poeppel, 2007](#), [Boemio et al., 2005](#));
- A spectrotemporal analysis of the acoustic cue happens at the temporal region ([Hickok and Poeppel, 2007](#));
- Generators of N1 and P2 are more laterally localized ([Hall, 2007](#), [Woldorff et al., 1993](#), [Godey et al., 2001](#), [Ross and Tremblay, 2009](#)).

Regarding the acoustic cue we assume that:

- Formants and VOT evoke different behaviours in N1 and P2 generators;
- N1 is sensitive to VOT variations ([Steinschneider et al., 1995](#), [Eggermont, 1995](#));
- Stimuli are processed differently when there is attention to the task ([Möttönen et al., 2014](#), [Alho et al., 2016](#)).

Regarding the influence of attention we speculate that:

- Attention influences the generators recruited to process stimuli ([Hillyard et al., 1973](#));
- Attention influences more left hemisphere generators than right hemisphere ones;
- Attention influences the speed of stimuli processing ([Möttönen et al., 2014](#), [Alho et al., 2016](#)).

Regarding the ambiguity of the stimuli we think that the effect of the ambiguity will be reflected in the amplitude of the ERP ([Bidelman and Walker, 2017](#), [Rao et al., 2010](#)).

### 6.3.1 N1-P2 magnitude analysis

The first analysis was for the N1-P2 magnitude. This measurement is frequently used in speech research and is also immune to baseline variations among participants. The complete model is represented in Figure 59, whose R formula is:

$$P2.N1 \sim \text{feature} * \text{type} * \text{electrode} * \text{stimulus} + (1|participant).$$

A multiplicative effect can be observed in Figure 59 for the active values in relation to the passive ones for all electrodes, features and stimuli. This can happen due to the ISI but also to other factors, as commented in Section 6.1 and in Chapter 2. As this increase occurs in all active cases, it is difficult to compare the responses for the factor type (active/passive) because it will be significant in all comparisons performed. Thus, it is important to eliminate this multiplicative effect. In order to do this, we divided all measurements of N1-P2 in the active task by those in the passive task. The density distribution of the results of all those divisions is shown at the Figure 60. The median value is indicated by the red line in the figure. The values of the N1-P2 amplitude for the active task are, in average, 1.52 times greater than those for the passive cases.

By dividing all active task N1-P2 measurements by 1.52, we correct this multiplicative effect and then it is possible to include the factor type in the model. It is worth to mention that this correction is made only because the multiplicative effect was observed in all features, electrodes and stimuli. If not, the multiplicative effect will be really an effect of the factor analyzed and it will not be possible to eliminate this effect by the correction proposed. The representation of the corrected full model is shown in Figure 61.

After performing the ANOVA and RANOVA using the mixed-model formula presented before using the functions “anova” and “ranova” in R software, the following ANOVA table was obtained:

Type III Analysis of Variance Table with Satterthwaite's method							
	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
feature	82.3	82.31	1	1026.1	49.7300	3.238e-12	
type	0.2	0.18	1	1026.2	0.1057	0.7451196	
electrode	6090.3	1522.58	4	1026.1	919.8922	< 2.2e-16	
stimulus	107.7	26.93	4	1026.0	16.2719	6.257e-13	
feature:type	13.5	13.55	1	1026.0	8.1839	0.0043122	
feature:electrode	43.3	10.82	4	1026.0	6.5348	3.414e-05	
type:electrode	23.0	5.75	4	1026.1	3.4734	0.0079133	
feature:stimulus	31.5	7.87	4	1026.0	4.7519	0.0008404	

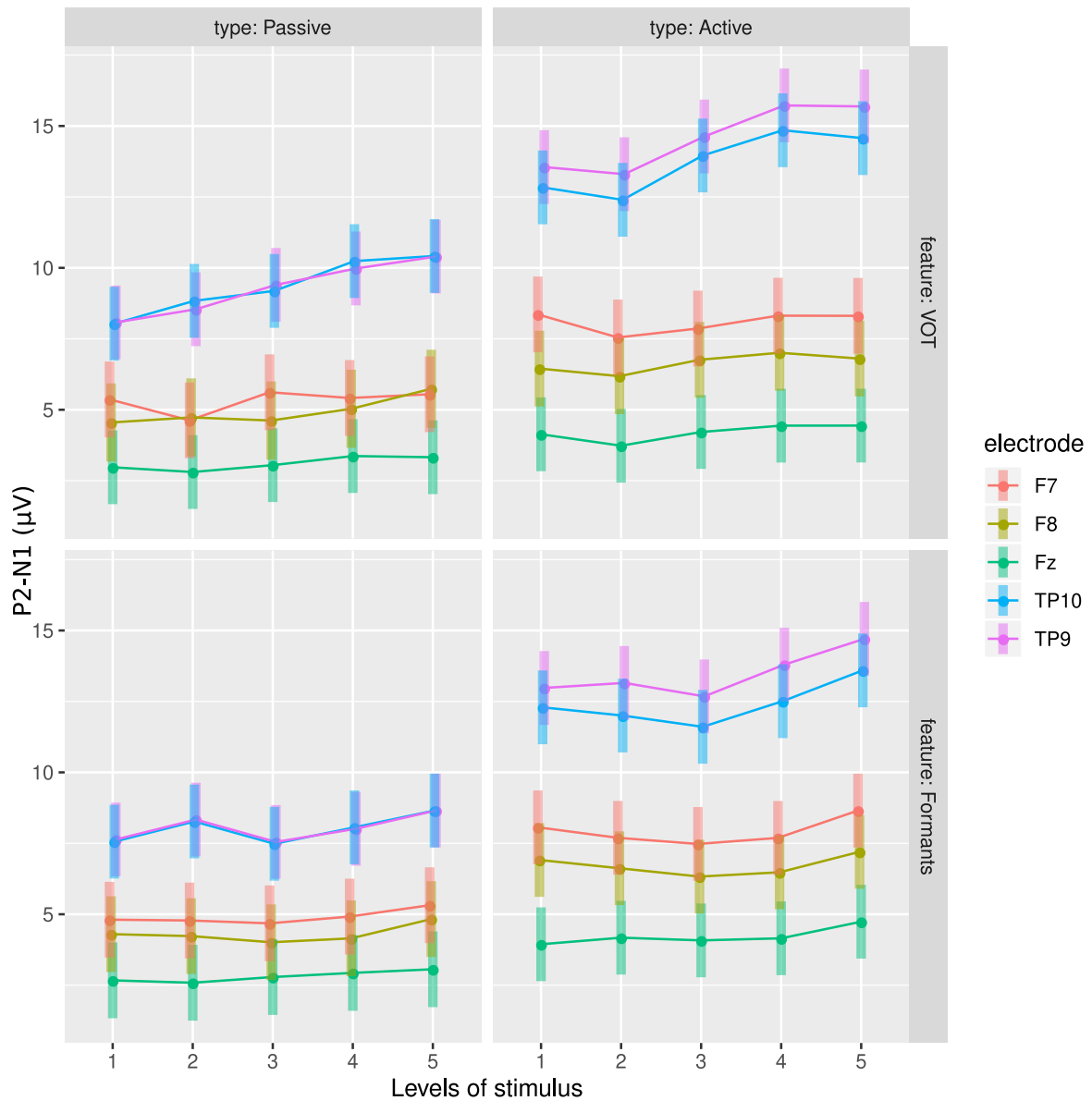


Figure 59 – Representation of the complete mixed-effects model including the fixed factors: feature, type, stimulus and electrodes. Predicted values for N1-P2 are computed for all 11 subjects. A multiplicative effect can be observed for the active values in relation to the passive ones for all electrodes, features and stimuli.

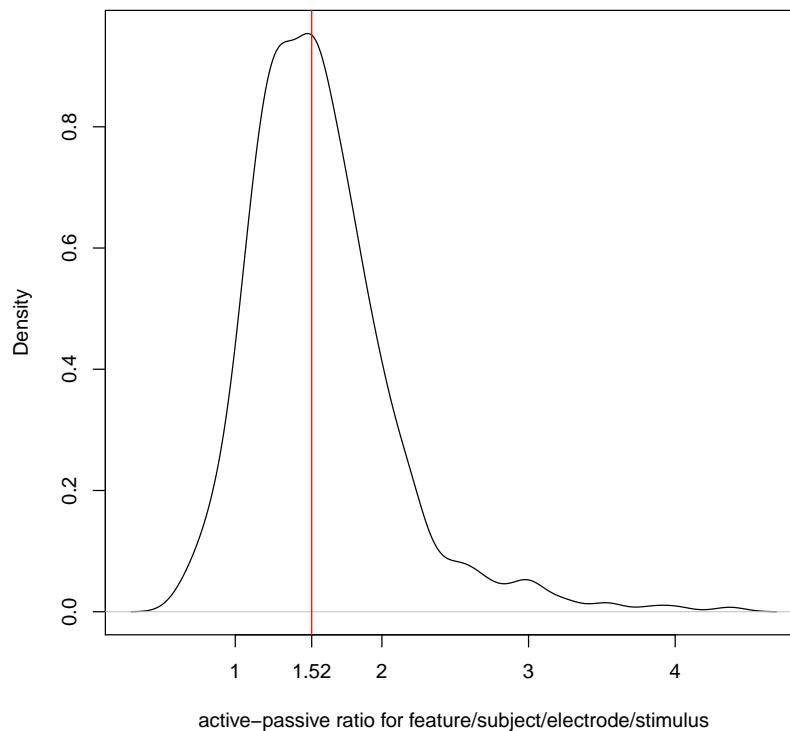


Figure 60 – Density distribution of the results of all the divisions of the N1-P2 values for features, stimuli, electrodes and participants at the active task by the passive task ones. The red line indicates the median of the distribution.

ANOVA-like table for random-effects: Single term deletions

Model:

```
T1 ~ feature + type + electrode + stimulus + (1 | participant) +
  feature:type + feature:electrode + type:electrode + feature:stimulus +
  type:stimulus + feature:type:electrode
```

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	34	2825.5	-5583.0			
(1   participant)	33	2724.2	-5382.3	202.65	1	< 2.2e-16 ***

The assumptions of normality, heterocedasticity and no correlation of the fixed effect predictors were analyzed at the Appendix F. The ANOVA table above shows that the factors feature, electrode, stimulus, and the interaction factors feature:type, feature:electrode, feature:stimulus, and type:electrode had significant effects over the N1-P2 magnitude. The random factor “participant” also presented significant effect here and for all dependent variables analyzed in this section. From now on, the ranova results will not be shown as we are interested in the analysis of the effects on the group, not on individual participants (but the random factor will continue to be

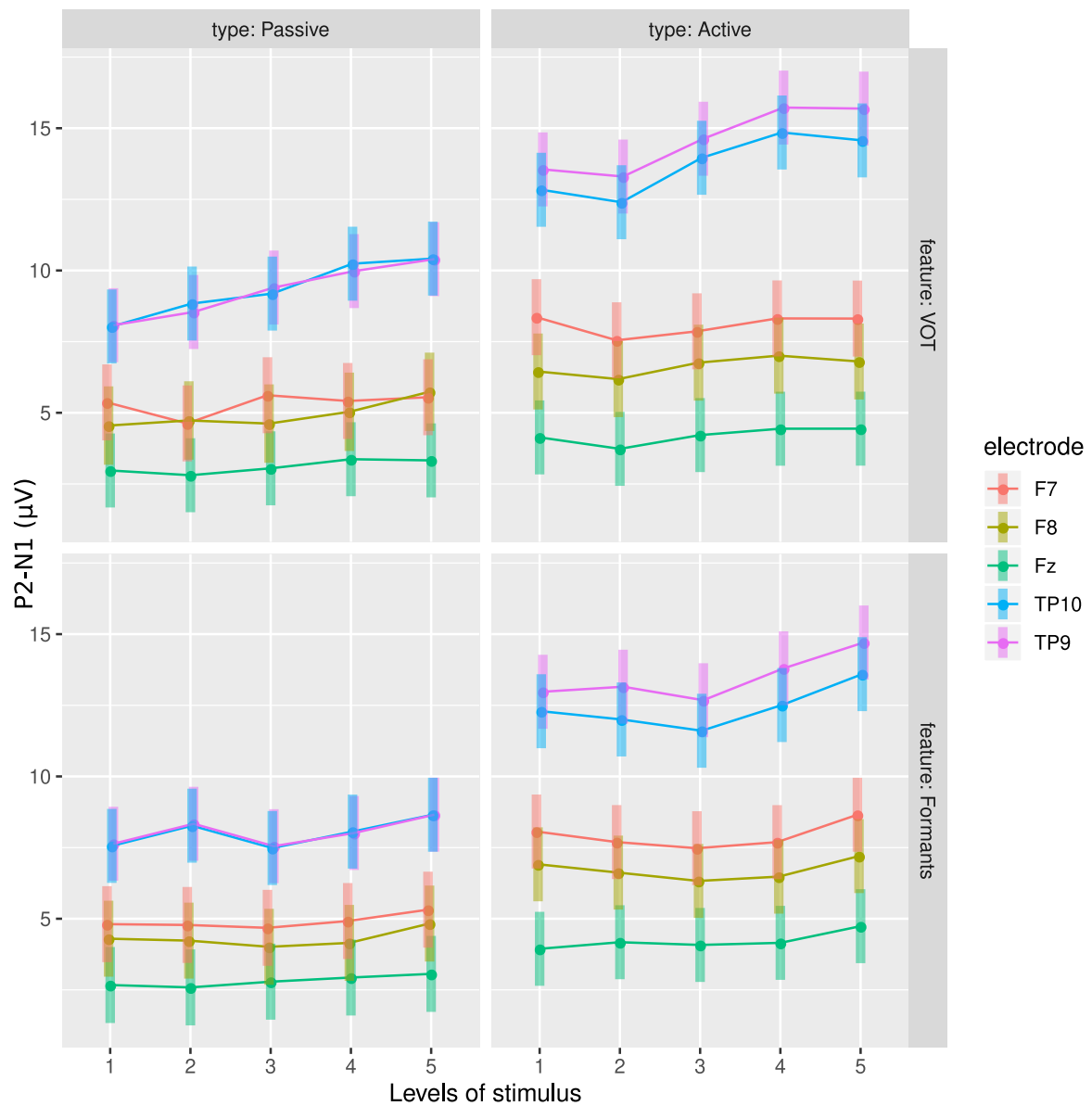


Figure 61 – Representation of the complete mixed-effects model including the fixed factors: feature, type, stimulus and electrodes. Predicted values for N1-P2 are computed for all 11 participants. The active measurements of N1-P2 were corrected in all electrodes, features and stimuli so that multiplicative effect can not be observed anymore.



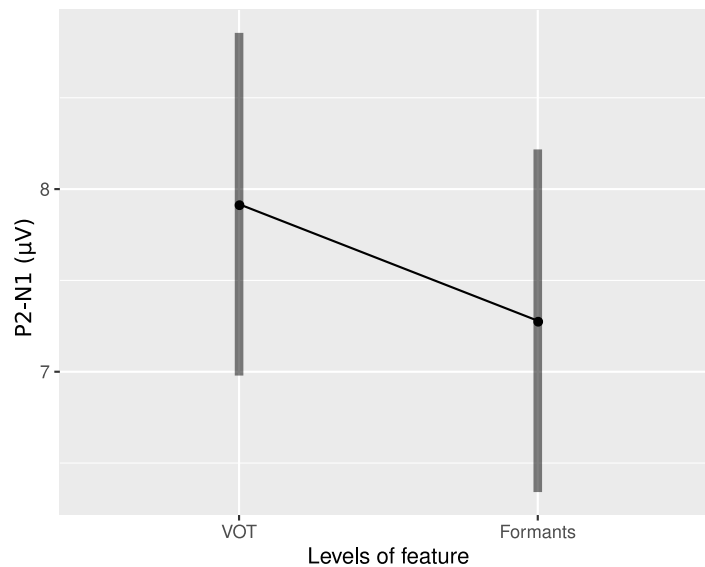


Figure 62 – Representation of the factor “feature” over the N1-P2 magnitude.

considered in the models). Each significant fixed factor was analyzed in a contrast test. The contrast analysis was performed using Kenward-Roger’s method for the degrees of freedom (Kenward and Roger, 1997).

The results for the “feature” factor is shown in Figure 62. It can be observed that, in average, the N1-P2 amplitude for the VOT continuum is greater than that for the formant continuum. Simos et al. (1998) and Steinschneider et al. (1999) worked with positive VOT for the voiced and voiceless consonants. Their results show that the AELR decreases with increasing VOT. Thus, as in our data, the shortest positive VOT occurs in the VOT continuum (near 0 ms for stim4 for some participants). This can explain the greater amplitudes observed in comparison with the formant continuum. Furthermore in Tremblay et al. (2003a), the authors show that the voiced consonant syllable /bi/ (VOT = 5.1 ms) evoked a greater AELR than the voiceless consonant syllable /pi/ (VOT = 65.4 ms). Korczak and Stapells (2010) also show this result for the /da/ and /ta/ syllables (larger positive VOT for /ta/). Another factor here that may lead to a greater amplitude for VOT responses is the length of the speech stimuli which is larger for stim1 and stim2 of the VOT continuum (discounting here the silence part of the other three stimuli which make then became smaller than stim1). It is also possible to consider the influence of the silence part of stimuli stim2 to stim5 in the VOT continuum which increase the time of the ISI for those stimuli. Thus, with a larger ISI the response to those stimuli tend to be also greater. This silence does not exist in the formant continuum (both can be seen in chapter 4).

The contrast analysis for this factor is:

Results are averaged over the levels of: type, electrode, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.559	0.0793	1026	-7.052	<.0001

The results for the “electrode” factor are shown in Figure 63. The contrast test performed is shown next and was performed considering the contrasts medial-lateral, frontal-temporal and left-right. All contrasts showed significant effects over the N1-P2 magnitude. The medial-lateral contrast indicates that the Fz electrode have smaller N1-P2 amplitudes than the F7, F8, TP9 and TP10 electrodes which are more laterally located at the scalp. This result makes sense considering the literature about speech processing and categorization, that usually consider studies with more lateral regions, as the inferior frontal gyrus (IFG), pSTG, STS and auditory cortices (A1 and A2) (Bouton et al., 2018, Alho et al., 2014, Altmann et al., 2014a).

The results for the frontal-temporal contrast show that frontal electrodes F7 and F8 have smaller amplitudes than temporal electrodes TP9 and TP10. This can be due to the position of the auditory cortices at the temporal region where the first stage of processing of the speech stimuli is performed and also to the location of some generators for the N1 and P2 components (Ross et al., 2013, Leaver and Rauschecker, 2010, Naatanen et al., 1992). The results for the left-right contrast show that the left hemisphere has greater amplitudes than the right one. The right hemisphere is assumed to have a selectivity for long-term integration while the left hemisphere is assumed to be less selective, working in different temporal resolutions (Boemio et al., 2005). This leads to the conclusion that the left hemisphere would categorize the stimuli, in general, better than the right (Hickok and Poeppel, 2007). This can explain the greater amplitudes at the left hemisphere for both continua we used.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-3.974	0.0985	1026	-40.347	<.0001
frontal - temporal	-3.916	0.0890	1026	-43.982	<.0001
left - right	0.495	0.0887	1026	5.580	<.0001

The results for the “stimulus” factor are shown in Figure 64. Here, there is the combination of

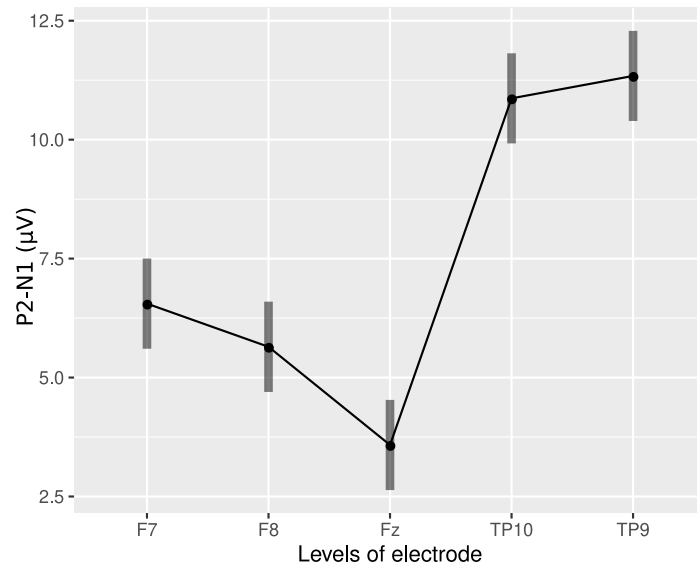


Figure 63 – Representation of the factor electrode over the variable N1-P2.

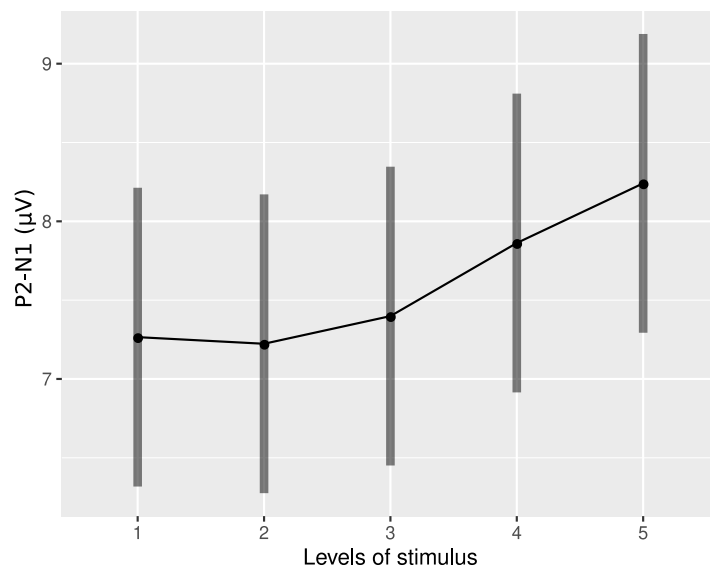


Figure 64 – Representation of the factor stimulus over the variable N1-P2.

two factors related to the continuum so that the increase of the N1-P2 magnitude with the stimuli sequence can be better understood through the analysis of the interaction feature:stimulus that can be seen at Figure 65. The linear and ambiguity contrasts described the significant effects.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, electrode

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
----------	----------	----	----	---------	---------

linear	1.0606	0.1397	1026	7.591	<.0001
psy - phy	-0.0447	0.0699	1026	-0.639	0.5227
ambiguity	0.1983	0.0988	1026	2.007	0.0450

The results for the interaction factor “feature:stimulus” are shown in Figure 65. As can be seen, in both continua, the N1-P2 amplitude increases with the stimulus sequence. Also, the pattern of variation with the stimuli is significantly different between VOT and Formants. The contrast shows that the increase and this interaction effect can be described by a linear function and also a psy-phy behavior. The linear contrast evidences by the difference in the slope of the curves as we go from the VOT to the Formants feature. The psy-phy contrast indicate the significant difference in the behavior of the dependent variable as we compare the features, with the behavior of the VOT stimuli being more psychophysical and the Formants more physical. As the effects observed in this interaction may be artificially induced due to the way that we choose to organize our continuum (from /da/ to /ta/ and from /pa/ to /pɛ/), it is worth to analyze each curve separately.

The increase observed for the VOT stimuli can be explained by the VOT that decreases (in module) from stim1 to stim4 then increases from stim4 to stim5, considering the average VOT of the 11 participants show in Figure 46. This effect was already explained in the analysis of the factor feature. As reported by Simos et al. (1998), the amplitude differences between VOT = 0 ms and VOT = 20 ms and also between VOT = 40 ms and VOT = 60 ms are not significant. We suggest that this may be also valid for negative VOTxs, as the psy-phy contrast was not significant to describe the effect of the factor stimulus in Figure 64 (which would indicate significant differences between stim1-stim2 and stim4-stim5). Thus, stim1 and stim2 have smaller amplitudes than stim4 and stim5 due to the VOT that is greater, in module, for the first two stimuli.

To explain the increase in amplitude for the formant continuum it is worth to explain the inhibitory formant frequency principle. This principle attests that there is an inhibitory interaction between the formant frequencies of the vowels so that the closer they are to each other, the smaller the amplitude of the evoked potential Ohl and Scheich (1997). This effect on the N1 and P2 amplitudes is demonstrated in studies with vowels in different languages like German, Italian, Korean and also Russian (Obleser et al., 2003, Manca et al., 2013, Shestakova et al., 2004, Kim et al., 2018) . We also observed this effect for our formant continuum data with the N1-P2 amplitude of the syllable /pa/ being smaller than that for the /pɛ/ syllable because the F2-F1 of the vowel /a/ is smaller than that of the /ɛ/ vowel. Furthermore, as F1 and F2 formant frequency distances changes with language, the effects of the vowel heard, over the AELR amplitudes, are different, but the formant frequency principle is still valid for the Brazilian Portuguese.

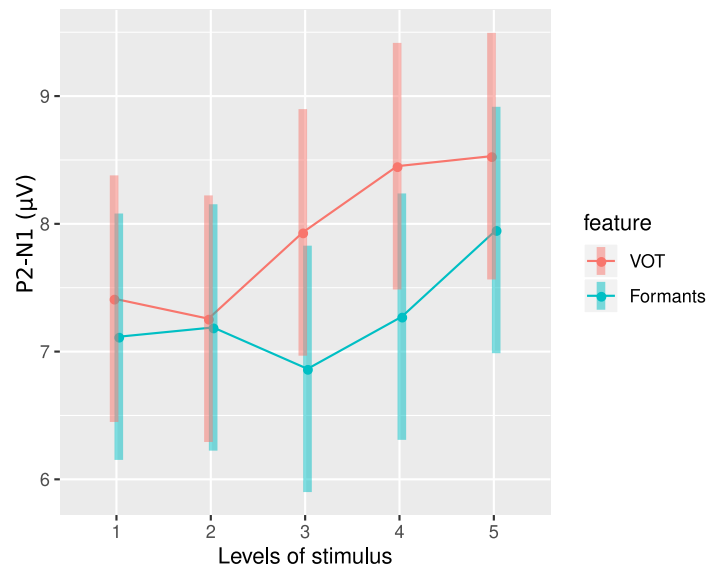


Figure 65 – Representation of the interaction factor feature:stimulus over the variable N1-P2.

This result confirm our assumption that Formants and VOT evoke different behaviors in N1 and P2 generators, that is, depending on the acoustic cue, the way how the stimuli is processed is different indicating that, probably, different generators are being recruited.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: type, electrode

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : linear	-0.732	0.279	1026	-2.620	0.0089
VOT - Formants : phy - psy	0.374	0.140	1026	2.679	0.0075
VOT - Formants : ambiguity	0.416	0.198	1026	2.106	0.0355

Figure 66 shows the significant effect on the interaction factor “feature:type” on the N1-P2 amplitude ( $p = 0.0043$ ). While the amplitude is greater for the passive task in the Formants case, its value decreases for the VOT continuum. Observing the grand averages, this result for VOT seems not to make sense but it is important to remember that a correction of the multiplicative effect was performed, so that the amplitudes observed in the active grand averages are not the same anymore. It is known that attention increases the component Nd (processing negativity) (Alho et al., 1986a,b), as commented at the Chapter 2. This component overlaps with N1 and P2 components making N1 amplitude to increase and P2 to decrease. But, as commented in

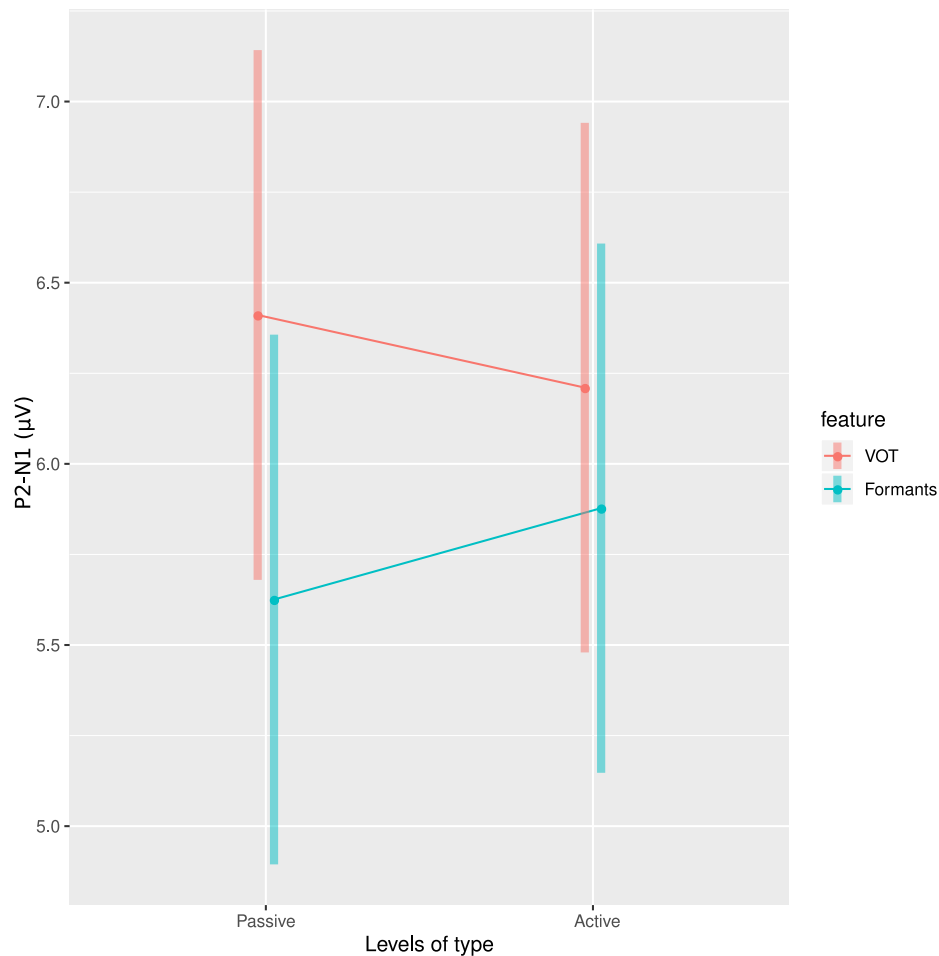


Figure 66 – Representation of the interaction factor feature:type over the variable N1-P2.

Chapter 2, N1 and P2 amplitudes also changes with other factors and those changes are different for each component as they have different cortical generators.

The N1 component vary greatly with the VOT as can be seen in the grand averages. For instance, Toscano et al. (2010) showed, with an oddball (active) task, that the N1 component was affected by the VOT but not by phonological category, reflecting the coding of a physical acoustic property by this component (for VOT a continuum). Thus, summing up the Nd effect over the N1 and P2 components resulted in a smaller N1-P2, in average, for the VOT-active than for the VOT-passive condition, which does not occurred for the formant continuum. This is probably due to a more stable N1 amplitude, across stimuli, for this continuum. This result is interesting because, in general, the literature indicates an increase in the AELR from passive to active tasks (Davis, 1964). This is probably due to the fact the many studies about this topic used tones or formant based stimuli in the experiments (Davis, 1964, Mast and Watson, 1968). In Möttönen et al. (2014), the authors used syllables varying in the place of articulation. In their results there is also a decrease in the N1-P2 amplitude as they go from the passive to the active task for both left and right hemispheres.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: electrode, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : Passive - Active	0.453	0.158	1026	2.861	0.0043

In Figure 67 we observe the significant interaction effect of “feature” and “electrode” on the N1-P2 amplitude. It is possible to notice that as we go from VOT to the formant feature, the difference in amplitude increases between electrodes. This difference is more pronounced between temporal electrodes than frontal ones. This observation is in agreement with the contrast results that show significant effects between frontal-temporal electrodes ( $p < 0.0001$ ) and medial-laterally located electrodes ( $p = 0.0109$ ). The effect observed for the factor feature has been commented above, but now we also see that this effect is more pronounced at the temporal region. This may be due to the location of the primary and secondary auditory cortices at this region, which are the first cortical structures known to process the acoustic stimuli. Generators for N1 and P2 components are also located in this region (Ross et al., 2013, Leaver and Rauschecker, 2010, Naatanen et al., 1992).

This factors interaction together with the effect observed for the factor electrode, confirm the assumptions made previously for the cortical region. We can see here that a larger activity at the temporal region and that the VOT and Formants seems to be processed differently at this region. We also saw a left hemisphere dominance with larger activity at this hemisphere than in the right one.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: type, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : medial - lateral	0.5021	0.197	1026	2.551	0.0109
VOT - Formants : frontal - temporal	0.7729	0.178	1026	4.352	<.0001
VOT - Formants : left - right	0.0666	0.177	1026	0.375	0.7075

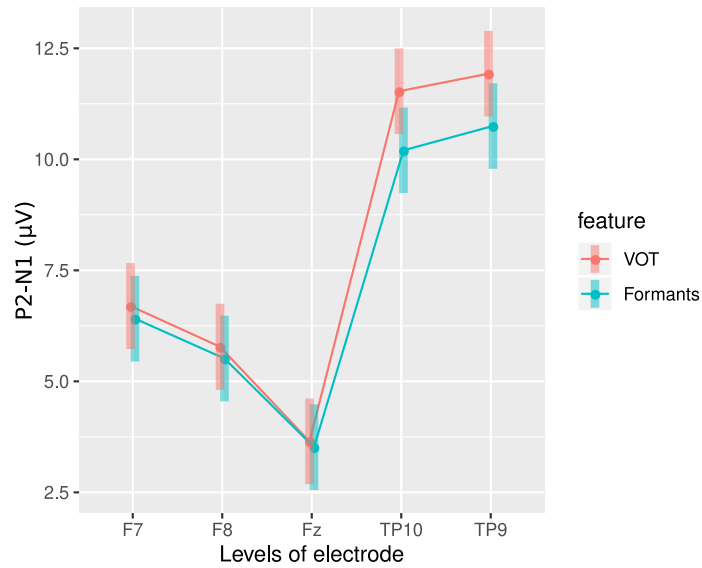


Figure 67 – Representation of the interaction factor feature:electrode over the variable N1-P2.

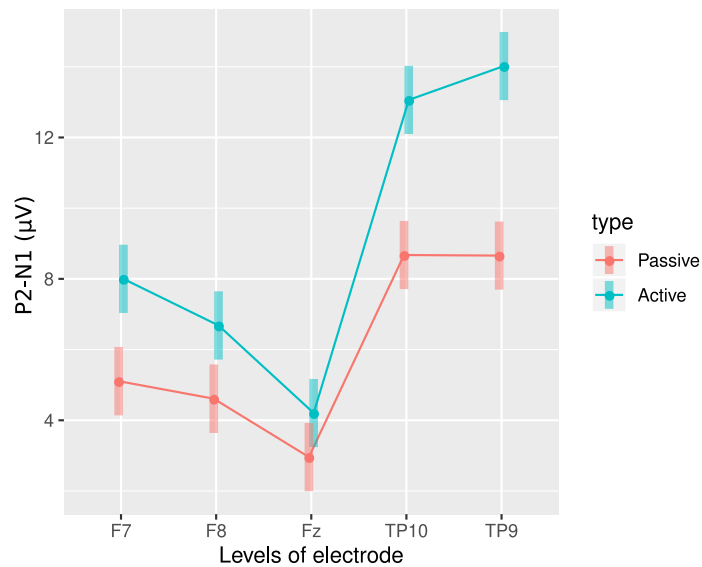


Figure 68 – Representation of the interaction factor type:electrode over the variable N1-P2.

Figure 68 shows the interaction between the task “type” and the “electrodes”. This interaction factor is significant between hemispheres as we go from the passive to the active task ( $p = 0.0044$ ). This shows that the difference in amplitude observed between left and right localized electrodes (already commented in the factor electrode before) is even greater for the active than for the passive task. In fact, for the temporal electrodes the N1-P2 magnitude is the same for both left and right electrodes in the passive task while, in the active task, the TP9 electrode presents greater magnitude than TP10. It is also interesting to notice that the right electrodes, F8 and TP10, have a greater difference in the N1-P2 magnitude in relation to F7 and TP9 respectively, when comparing the active task with the passive one. This shows that the left hemisphere responses are stronger than the right and even stronger in the active task.



The contrast analysis for this interaction factor is:

Results are averaged over the levels of: feature, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active : medial - lateral	-0.250	0.197	1026	-1.270	0.2045
Passive - Active : frontal - temporal	-0.319	0.178	1026	-1.795	0.0730
Passive - Active : left - right	0.506	0.177	1026	2.854	0.0044

### 6.3.2 N1 analysis

For the following N1 and P2 peaks analysis, the multiplicative effect correction was not performed because, for those measurements, the baseline across participants, stimuli and electrodes varies. In some cases, the P2 have negative values and the N1 have positive values. Then, amplitudes differences computed between active and passive cases here will not take into account the baseline issue and the correction will be wrong.

The ANOVA result for the complete model (including only significant factors) is:

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	1.45	1.45	1	1015.2	0.6863	0.4076
type	610.24	610.24	1	1015.4	288.4898	< 2.2e-16
electrode	1057.76	264.44	4	1015.2	125.0132	< 2.2e-16
stimulus	103.13	25.78	4	1015.0	12.1884	1.108e-09
type:electrode	244.40	61.10	4	1015.2	28.8845	< 2.2e-16
feature:stimulus	57.21	14.30	4	1015.0	6.7619	2.265e-05
electrode:stimulus	66.72	4.17	16	1015.0	1.9714	0.0124

Interestingly, the “feature” factor was not significant in the N1 amplitude. A possible explanation for this is the differences between N1 values of each stimuli that were all included together for the computation of the feature effect. This way the smaller N1 values compensated the larger values and, in average, the resulting N1 for the VOT feature was not significantly different than that for the Formants (continuum where the N1 values doesn't vary significantly).

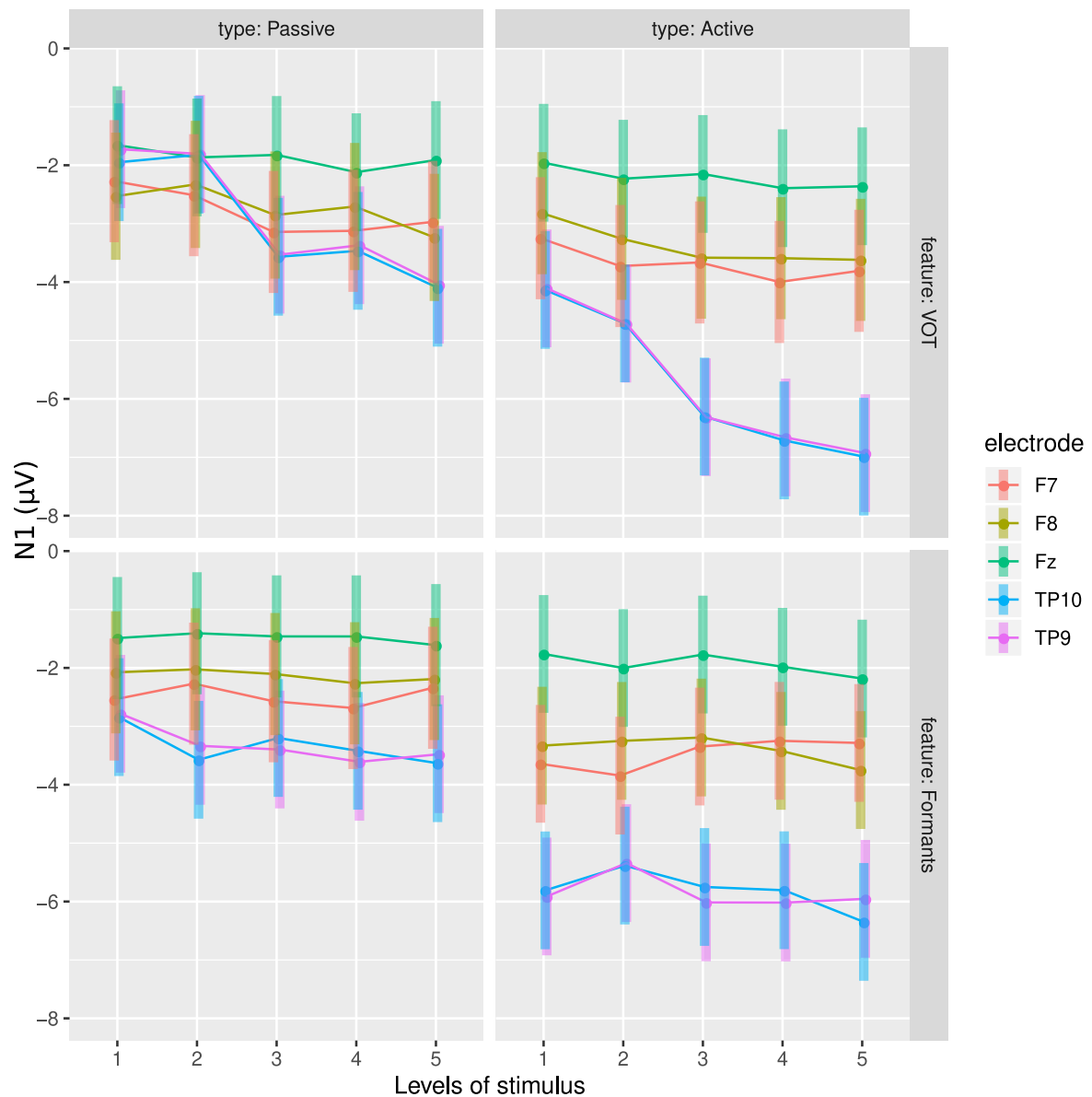


Figure 69 – Representation of the complete mixed-effects model including the fixed factors: feature, type, stimulus and electrodes. Predicted values for N1 are computed for all 11 participants.

The contrast for the factor “type” can be seen in Figure 70. The amplitude of N1 varies significantly being greater for the active task than for the passive task. This result is different than that found by [Bidelman and Walker \(2017\)](#), who did not show a significant difference between the N1 amplitudes for the active and passive task. However, [Woldorff et al. \(1993\)](#) observed an attention-related enhancement in the activity of a N1 generator located at the supratemporal plane lateral to Heschl’s gyrus. [Hillyard et al. \(1973\)](#) also observed an increase in N1 amplitude with attention and attributed this effect to the recruitment of additional neurons or to an improvement in the synchronicity of neuronal firing or to both. In our case, it is also important to consider the effect of the ISI difference between active and passive tasks which was not controlled in our experiment due to the randomization of the tasks across participants.

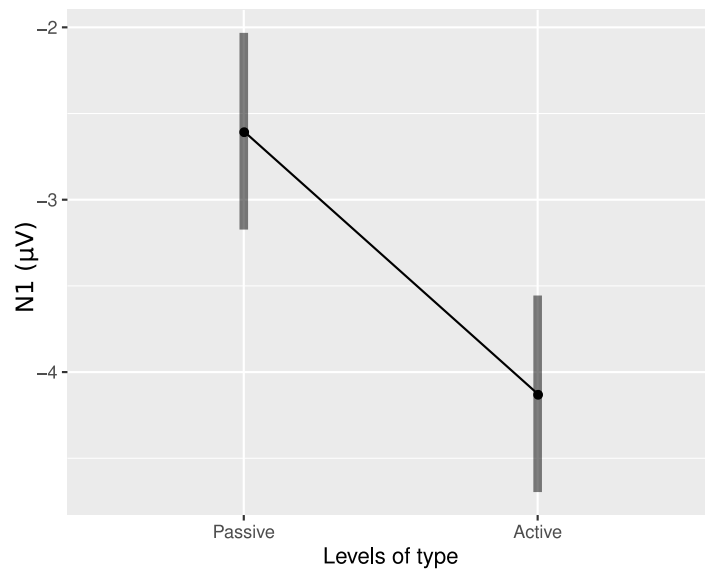


Figure 70 – Representation of the factor type over the variable N1.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, electrode, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	-1.52	0.0897	1015	-16.984	<.0001

The results for the “electrode” factor are illustrated in Figure 71. The medial-lateral and frontal-temporal contrasts showed significant effects on the N1 magnitude ( $p < 0.0001$ ). As the N1 amplitude generally is negative, this result is similar to the result found for this factor on the N1-P2 magnitude, considering the absolute value of N1. Thus, the conclusions made for the medial-lateral and frontal-temporal contrasts on the N1-P2 magnitude are valid here. The difference was the absence of the contrast left-right here showing that the N1 amplitude is similar on both hemispheres.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	1.8506	0.111	1015	16.622	<.0001

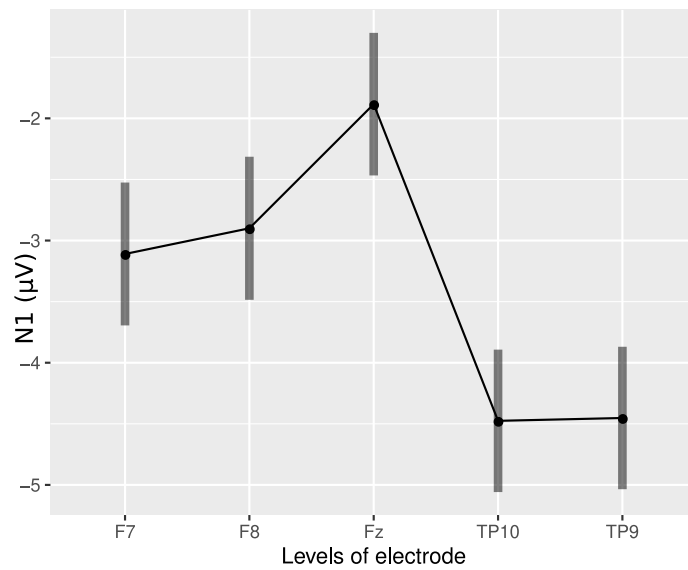


Figure 71 – Representation of the factor electrode over the variable N1.

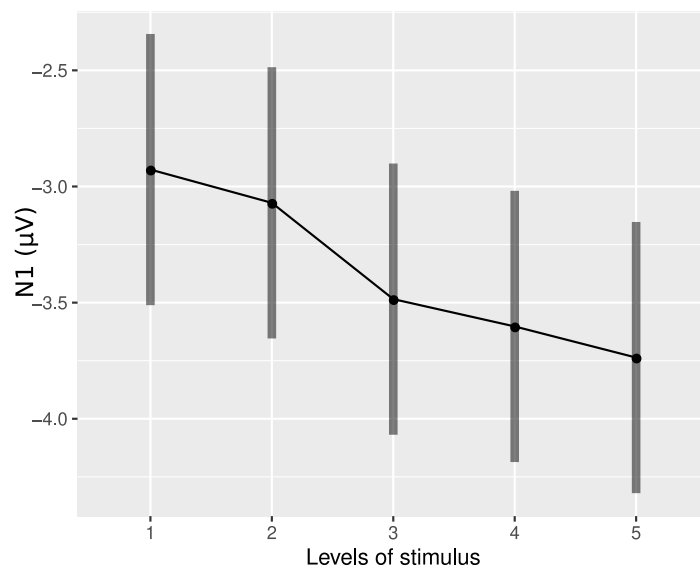


Figure 72 – Representation of the factor stimulus over the variable N1.

frontal - temporal	1.4591	0.101	1016	14.501	<.0001
left - right	-0.0935	0.100	1015	-0.933	0.3512

The results for the “stimulus” factor are shown in Figure 72. The results are similar to those for the N1-P2 magnitude analysis, considering the absolute values of the N1 amplitude. It is possible to see the increase from stim1 to stim5 (in modulus). The contrasts show that the effect can be described by a linear function. This result can be better understood by means of the interaction factor feature:stimulus. We saw in Chapter 2 that the VOT directly affects the latency of the AELR component N1 (Tremblay et al., 2003b), as well as the magnitudes of N1 and P2 (Simos et al., 1997, Horev et al., 2007).

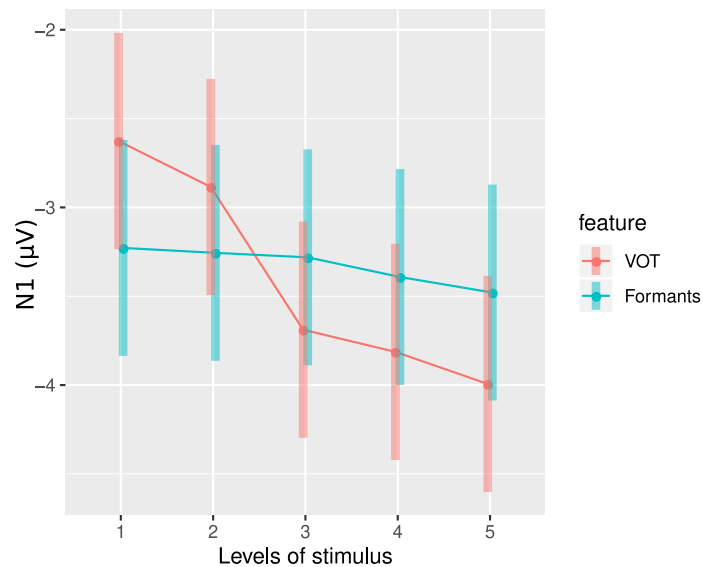


Figure 73 – Representation of the interaction factor feature:stimulus over the variable N1.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, electrode

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
linear	-1.0753	0.158	1015	-6.803	<.0001
phy - psy	0.0637	0.079	1015	0.806	0.4205
ambiguity	0.1508	0.112	1015	1.349	0.1775

The interaction factor “feature:stimulus” is illustrated in Figure 73. Considering N1 absolute values, it is possible to apply the same analysis made for the N1-P2 feature:stimulus interaction here, considering the VOT values and the inhibitory formant frequency principle as explanations for the differences and the increasing tendency observed in the sequence from stim1 to stim5. The difference here is that only the linear contrast was significant for the interaction factor ( $p < 0.0001$ ). We observe that a different pattern of variation with the stimuli between the VOT and the Formants cases. This suggests that much of the variation observed for the Formants feature in Fig. 65 may be due to the influence of the P2 amplitude.

This result confirm our assumption that the N1 component is sensitive to VOT variations.

The contrast analysis for this interaction factor is:

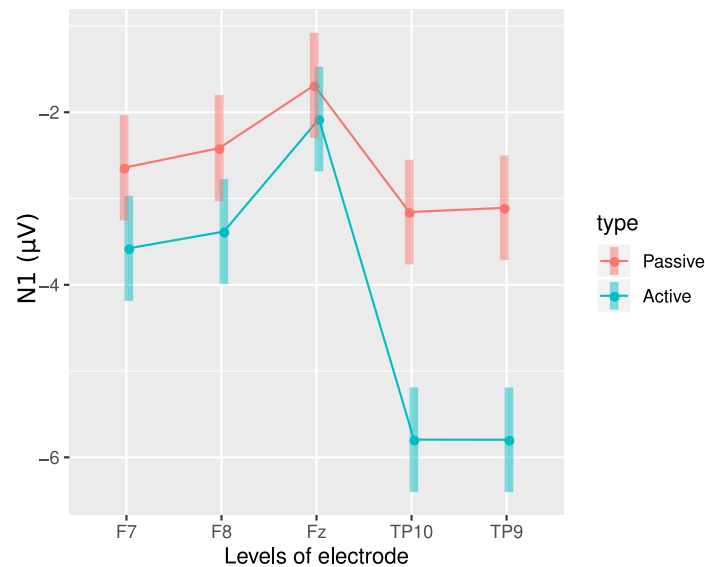


Figure 74 – Representation of the interaction factor type:electrode over the variable N1.

Results are averaged over the levels of: type, electrode

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : linear	1.512	0.316	1015	4.786	<.0001
VOT - Formants : phy - psy	-0.117	0.158	1015	-0.740	0.4594
VOT - Formants : ambiguity	-0.417	0.223	1015	-1.864	0.0625

Figure 74 shows the interaction between the task type and the electrodes. This interaction was significant for the medial-lateral contrast ( $p < 0.0001$ ) showing that the difference in the N1 amplitude between the passive and active responses for the Fz electrode is smaller than the average of this difference for the more lateral electrodes (F7, F8, TP9 and TP10). The frontal-temporal contrast has also a significant effect ( $p < 0.0001$ ). As can be seen in the figure, the difference in N1 amplitude as we go from passive to active task is greater for the temporal electrodes (TP9 and TP10) than for the frontal ones (F7 and F8). This result is in accordance with our assumption that attention influences the generators recruited to process stimuli, in this case the N1 generators are clearly influenced.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: feature, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

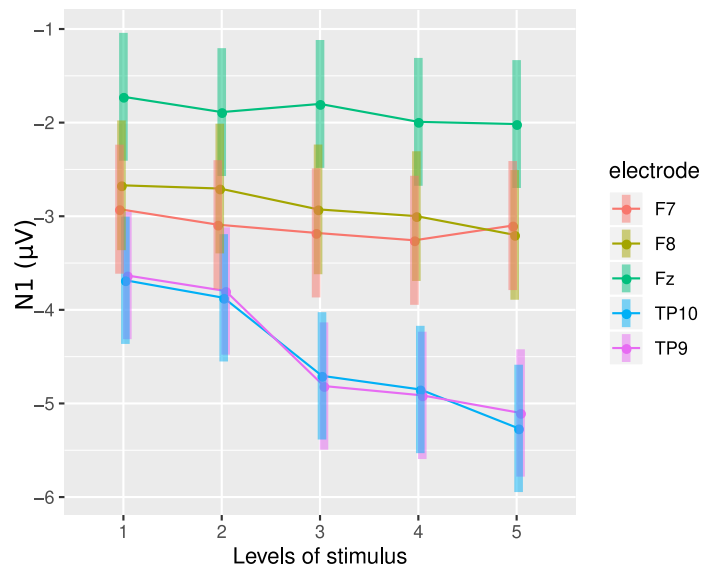


Figure 75 – Representation of the interaction factor electrode:stimulus over the variable N1.

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active : medial - lateral	1.41606	0.223	1015	6.362	<.0001
Passive - Active : frontal - temporal	1.71252	0.201	1016	8.517	<.0001
Passive - Active : left - right	-0.00875	0.200	1015	-0.044	0.9652

Figure 75 shows the interaction between electrodes and stimuli. We noticed that the contrasts tend to have a linear behavior and this is reflected in the significant effect of the medial-lateral contrast ( $p = 0.02$ ) and the frontal-temporal contrast ( $p < 0.0001$ ). The medial-lateral contrast indicates that there is a significant difference between the linear function that describes the behavior of stimuli for the medial electrode (Fz) and the function of the more lateral ones (F7, F8, TP9 and TP10). This is evident if we imagine that the average slope of the latter is greater than that of the first. The frontal-temporal contrast indicates that there is a significant difference in the linear behavior between the more frontal electrodes (F7 and F8) and the temporal ones (TP9 and TP10). This is evident in the figure as we can see a greater slope of the linear function that describes the behavior of the stimuli for the temporal electrodes than for those in the frontal electrodes. From feature:stimulus analysis, is possible to hypothesize that this may be due to influence of the VOT more that Formants variation.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: feature, type  
 Degrees-of-freedom method: kenward-roger  
 Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral : linear	0.9152	0.393	1015	2.329	0.0200
medial - lateral : phy - psy	-0.1054	0.196	1015	-0.536	0.5918
medial - lateral : ambiguity	-0.3190	0.278	1015	-1.148	0.2512
frontal - temporal : linear	1.5813	0.354	1015	4.465	<.0001
frontal - temporal : phy - psy	-0.1149	0.177	1015	-0.649	0.5165
frontal - temporal : ambiguity	-0.3101	0.250	1015	-1.238	0.2160
left - right : linear	0.2325	0.354	1015	0.657	0.5116
left - right : phy - psy	0.0606	0.177	1015	0.342	0.7321
left - right : ambiguity	0.1085	0.250	1015	0.433	0.6649

### 6.3.3 P2 analysis

Figure 76 presents the complete mixed-effects model representation for the P2 amplitude dependent variable. The result of the ANOVA applied to this model is presented below and show the significant factors and interactions: feature, type, electrode, stimulus, feature:electrode and type:electrode.

The ANOVA result for the complete model (including only significant factors) is:

Type III Analysis of Variance Table with Satterthwaite's method							
	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
feature	86.1	86.11	1	1031.1	37.2665	1.457e-09	
type	768.1	768.12	1	1031.2	332.4379	< 2.2e-16	
electrode	4467.9	1116.98	4	1031.1	483.4225	< 2.2e-16	
stimulus	46.8	11.69	4	1031.0	5.0586	0.0004865	
feature:electrode	164.2	41.06	4	1031.0	17.7689	4.069e-14	
type:electrode	103.3	25.83	4	1031.1	11.1786	7.012e-09	

The results for the “feature” factor are shown in Figure 77. It can be observed that, in average, the P2 amplitude for the VOT continuum is greater than that for the formant continuum. The same observations made for this feature in the case of the N1-P2 magnitude are valid here. As the factor feature did not present significant effect on the N1 amplitude, this shows the P2 peak is the one that more strongly influences this factor in the N1-P2 amplitude.

The contrast analysis for this factor is:



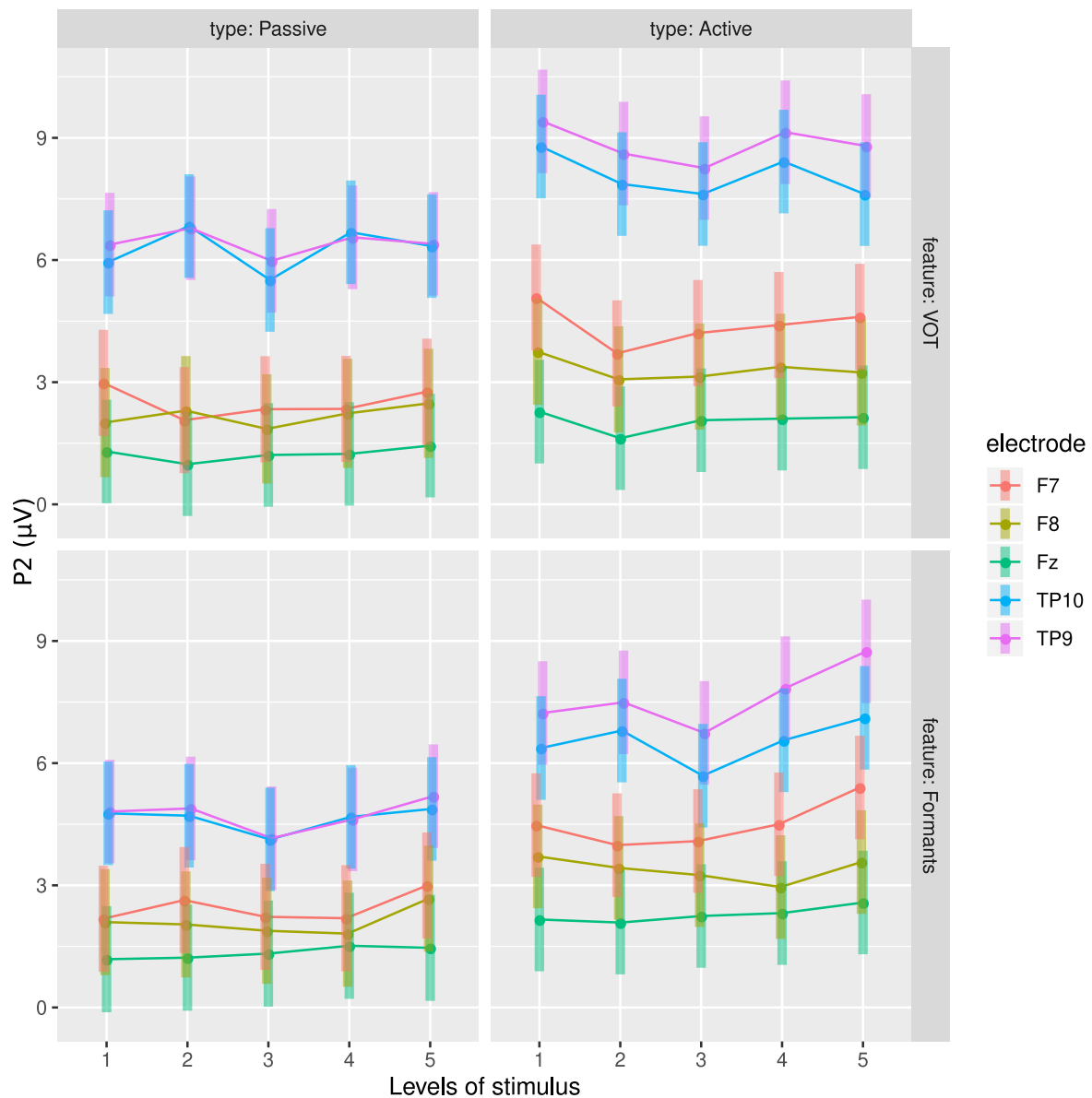


Figure 76 – Representation of the complete mixed-effects model including the fixed factors: feature, type, stimulus and electrodes. Predicted values for P2 are computed for all 11 participants.

Results are averaged over the levels of: type, electrode, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.572	0.0936	1031	-6.105	<.0001

The “feature:stimulus” interaction factor had no significant effect on the P2 amplitude. This suggests that the variation of P2 across stimuli for the VOT continuum is not significantly different from the variation observed for the formant continuum. This shows that the VOT values do not affect the P2 in the same way as it affects N1. Thus, the variations observed in P2 amplitude

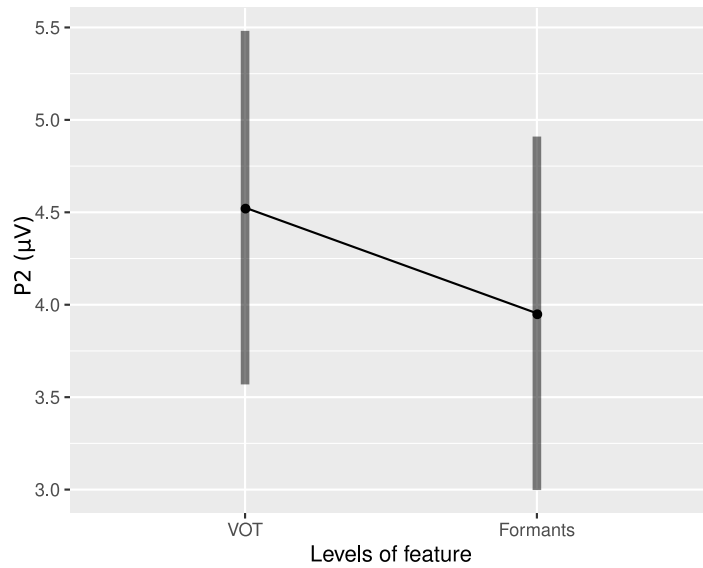


Figure 77 – Representation of the factor feature over the variable P2.

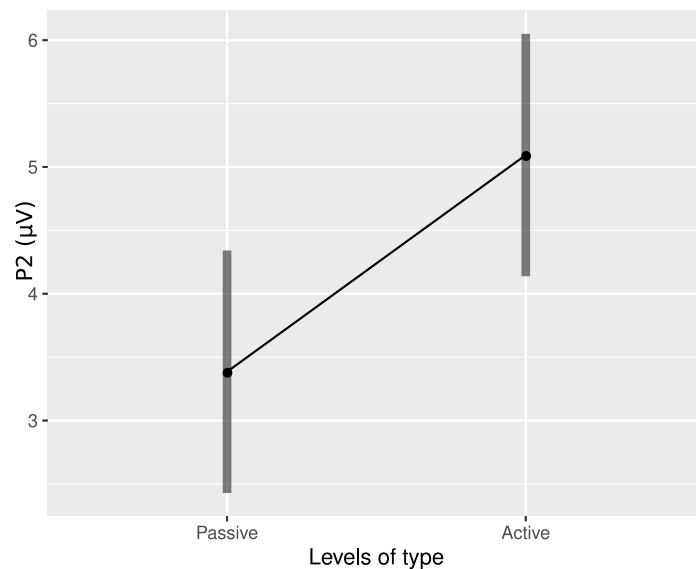


Figure 78 – Representation of the factor type over the variable P2.

may be related to a higher level processing of speech, which is not related to a spectrotemporal processing of physical characteristics of the stimuli.

Figure 78 shows a significant effect of the factor “type” on the P2 amplitude, which increase from the passive to active task ( $p < 0.0001$ ). This can be due to the greater *ISI* of the active task than the passive one as it was not possible to control this variable in our experiment.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, electrode, stimulus

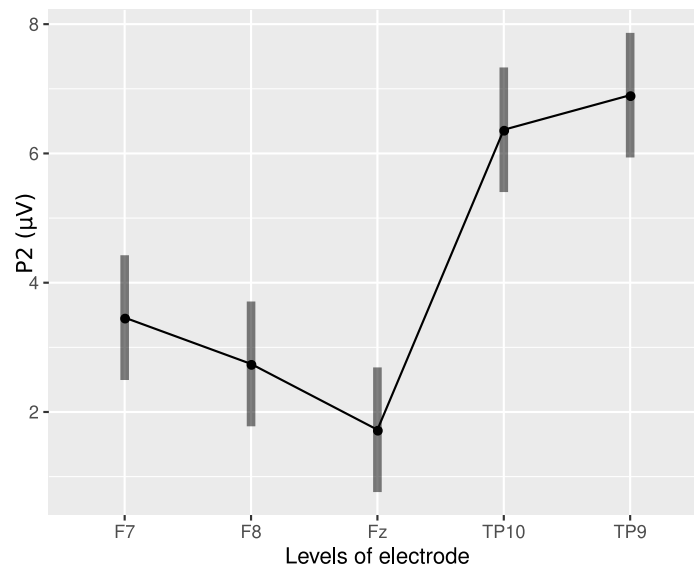


Figure 79 – Representation of the factor electrode over the variable P2.

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	1.71	0.0938	1031	18.233	<.0001

The results for the electrode factor are shown in Figure 79. The tests for the medial-lateral, frontal-temporal and left-right contrasts are shown below. All contrasts showed significant effects over the P2 magnitude and the same considerations already made for this factor and contrast for the N1-P2 magnitude are valid here.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-3.143	0.116	1031	-27.002	<.0001
frontal - temporal	-3.532	0.105	1031	-33.572	<.0001
left - right	0.626	0.105	1031	5.970	<.0001

Figure 80 shows an interesting result for the “stimulus” factor. The contrast presented a significant ambiguity effect for this result ( $p = 0.0002$ ). In the work of [Bidelman and Walker \(2017\)](#) the authors observed that their ambiguous stimulus (equivalent to our stim3), in a formant based

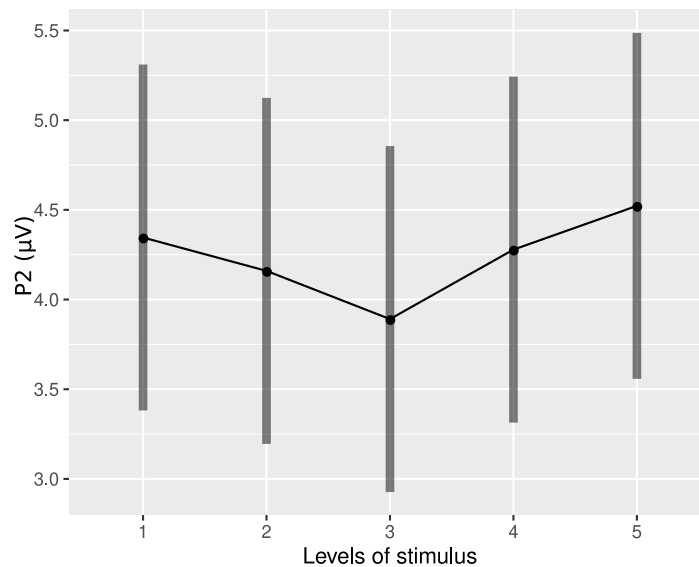


Figure 80 – Representation of the factor stimulus over the variable P2.

continuum, presented a smaller N1-P2 magnitude than the unambiguous stimuli (equivalent to our stim1 and stim5). What we observe here is that this effect occurs for both [VOT](#) and Formants continuum, and for both active and passive tasks. We also show that the stim2 and stim4, which are not so unambiguous as stim1 and stim5 but also not so ambiguous as stim3, have a P2 amplitude located between the other stimuli. This shows that P2 may codify the ambiguity of the stimulus or the effort for its perception. For instance, in a study using tones, [Rao et al. \(2010\)](#) reported a greater P2 amplitude for the stimuli easier to classify (equivalent to our stim1 and stim5) than for the more difficult ones (stim2, stim3 and stim4).

Some differences between our result and that of [Bidelman and Walker \(2017\)](#) that can explain those different results are related to the stimuli continuum and the dependent variable. First, the continuum /u/-/a/ used by the [Bidelman and Walker \(2017\)](#) can be problematic even for the English listeners, as already discussed in the Introduction of this thesis. Second, as the authors used the N1-P2 amplitude, this possible detection of ambiguity performed by the P2 component may be masked by the influence of the N1 component.

This effect confirm our assumption that the ambiguity is reflected in the [ERP](#) amplitude, but, as we can see, just in the P2 component.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, electrode  
 Degrees-of-freedom method: kenward-roger  
 Confidence level used: 0.95

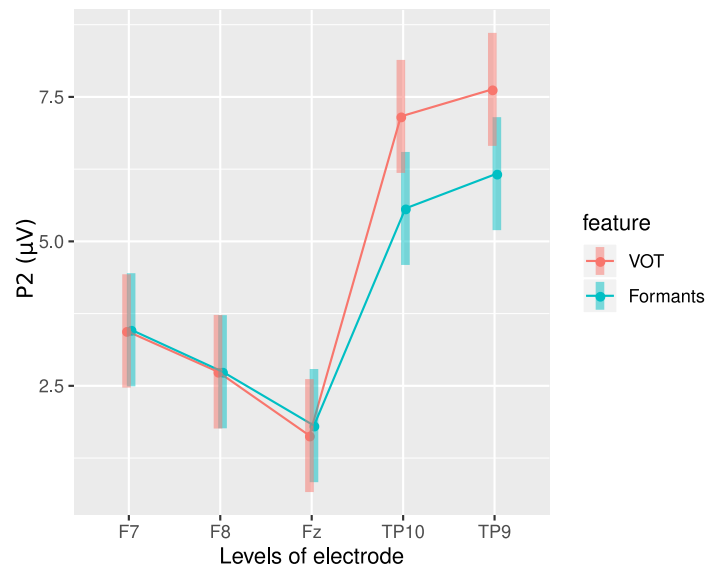


Figure 81 – Representation of the interaction factor feature:electrode over the variable P2.

contrast	estimate	SE	df	t.ratio	p.value
linear	0.2357	0.1651	1031	1.428	0.1536
phy - psy	-0.0152	0.0825	1031	-0.184	0.8543
ambiguity	0.4354	0.1167	1031	3.731	0.0002

Figure 81 shows the effect of the interaction factor “feature:electrode” on the P2 amplitude. All considerations made for this interaction factor for the N1-P2 dependent variable are valid here.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: type, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : medial - lateral	0.9305	0.233	1031	4.001	0.0001
VOT - Formants : frontal - temporal	1.5362	0.210	1031	7.322	<.0001
VOT - Formants : left - right	0.0767	0.209	1031	0.366	0.7144

Figure 82 shows an interaction between the task and the electrodes. This interaction has a significant effect for the medial-lateral contrast ( $p < 0.0001$ ). This indicates that the difference in the P2 amplitude between the passive and active responses for the Fz electrode is smaller than the average of this difference for the lateral electrodes (F7, F8, TP9 and TP10). The frontal-temporal contrast was also significant ( $p = 0.0021$ ). As can be seen in the figure, the difference in P2

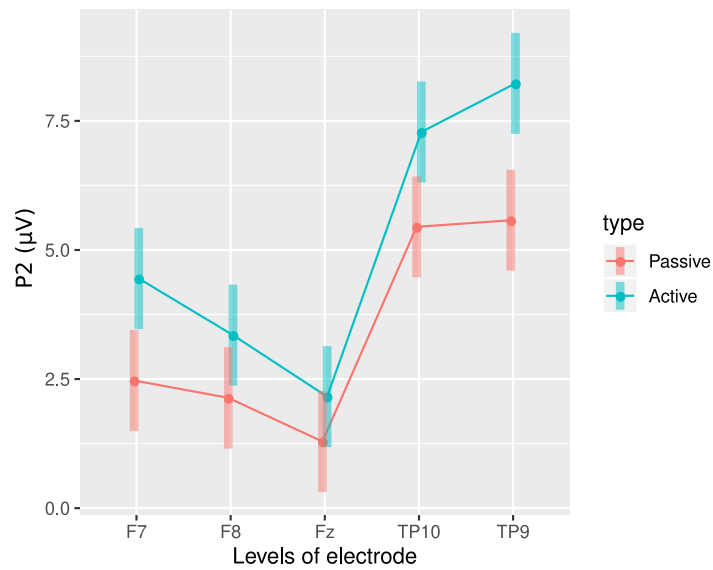


Figure 82 – Representation of the interaction factor type:electrode over the variable P2.

amplitude as we go from the passive to the active task is greater for the temporal electrodes (TP9 and TP10) than for the frontal ones (F7 and F8). This interaction is also significant between hemispheres as we go from the passive to the active task ( $p = 0.0002$ ). This shows that the difference in amplitude observed between left and right electrodes is even greater for the active than for the passive task.

This result confirm our assumption that attention influences the generators recruited to process stimuli and also that this effect is more accentuated for the P2 generators at the left hemisphere than for the right hemisphere ones.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: feature, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active : medial - lateral	-1.050	0.233	1031	-4.513	<.0001
Passive - Active : frontal - temporal	-0.650	0.210	1031	-3.091	0.0021
Passive - Active : left - right	0.787	0.209	1031	3.757	0.0002

### 6.3.4 T1 analysis

Figure 83 presents the complete mixed-effects model representation for the T1 dependent variable (N1 latency). The result of the ANOVA applied to this model is presented next. It shows significant effects for the main factors feature, electrode, and stimulus and for the interaction factors feature:type, feature:electrode, type:electrode, feature:stimulus, type:stimulus and feature:type:electrode.

The N1 and P2 waves amplitudes and latencies are the result of the overlapping contribution of evoked potentials from multiple cortical and subcortical generators. Giard et al. (1994) showed by the current field analysis that generators of such currents at the latency between 65 and 140 ms necessarily lie in and outside the auditory cortex. They showed that, for the N1 wave, there is a dissociation between the purely sensory component from other exogenous components generated by other parallel processes. Thus, it is expected some latency differences for the N1 and P2 components for different stimuli, tasks and scalp locations (electrodes).

The ANOVA result for the complete model (including only significant factors) is:

Type III Analysis of Variance Table with Satterthwaite's method								
	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)		
feature	0.47605	0.47605	1	1018.1	4256.3213	< 2.2e-16		
type	0.00030	0.00030	1	1018.2	2.7185	0.0994976		
electrode	0.00275	0.00069	4	1018.1	6.1507	6.845e-05		
stimulus	0.02209	0.00552	4	1018.0	49.3656	< 2.2e-16		
feature:type	0.00044	0.00044	1	1018.0	3.9462	0.0472450		
feature:electrode	0.00184	0.00046	4	1018.0	4.1111	0.0026078		
type:electrode	0.00118	0.00029	4	1018.1	2.6310	0.0330685		
feature:stimulus	0.01507	0.00377	4	1018.0	33.6941	< 2.2e-16		
type:stimulus	0.00122	0.00031	4	1018.0	2.7331	0.0278887		
feature:type:electrode	0.00213	0.00053	4	1018.0	4.7644	0.0008224		

The results for the “feature” factor on the T1 value are shown in Figure 84. Its effect is significant ( $p < 0.0001$ ). However, it is not possible to compare the responses between continua because the VOT continuum latencies are affected by the the VOT at each stimulus but, in general this result in a latency, in average, greater than that of the formant continuum responses.

The contrast analysis for this factor is:

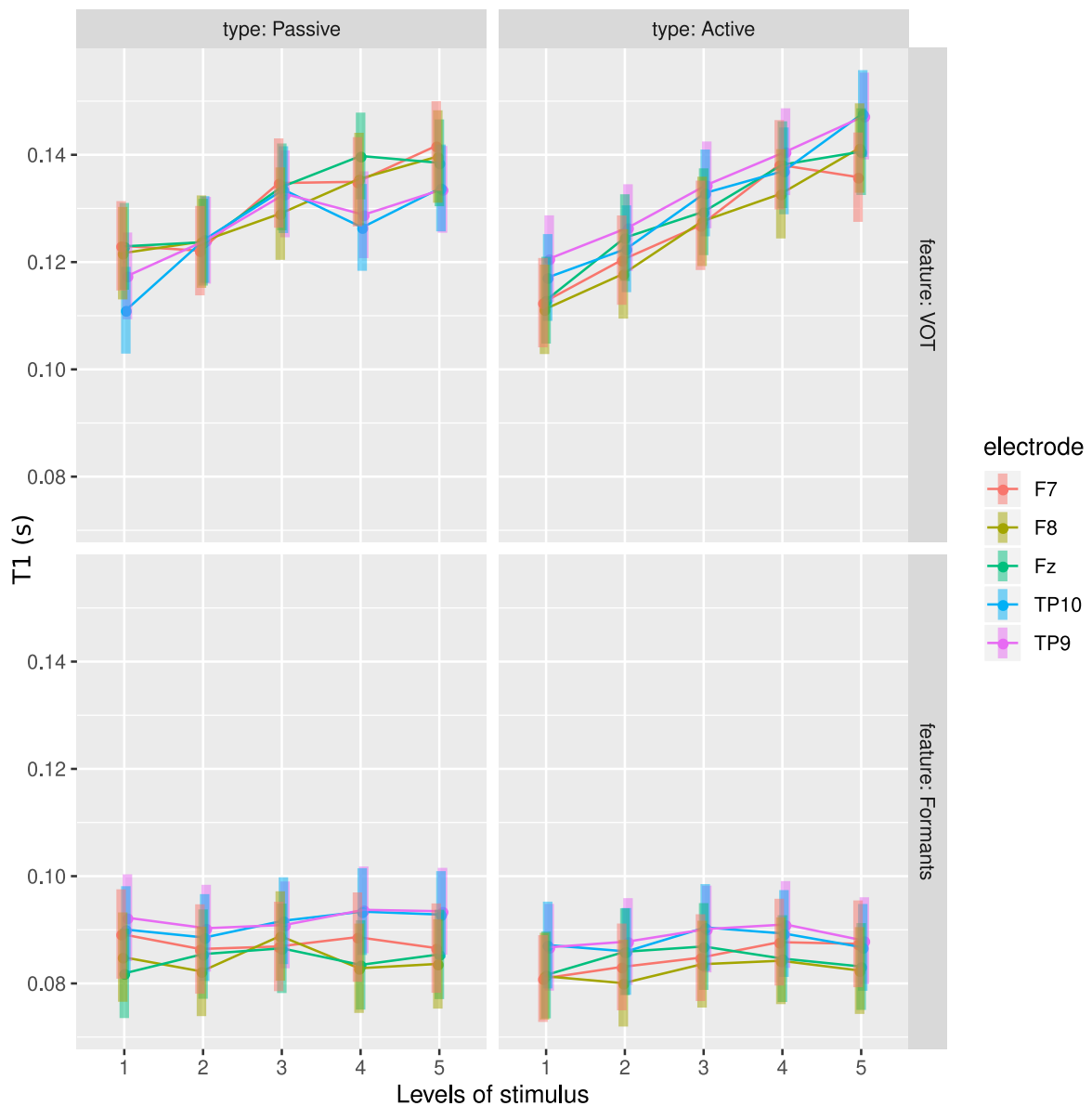


Figure 83 – Representation of the complete mixed-effects model including the fixed factors: feature, type, stimulus and electrodes. Predicted values for T1 are computed for all 11 participants.

Results are averaged over the levels of: type, electrode, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0425	0.000652	1018	-65.240	<.0001

The results for the “electrode” factor on the T1 value is shown in Figure 85. Its effect is significant for the frontal-temporal contrast ( $p < 0.0001$ ) and the left-right contrast ( $p = 0.0255$ ). This result shows that the N1 latency smaller for the frontal electrodes, in comparison with the temporal ones. The N1 wave happens first at the right hemisphere than in the left. It is important



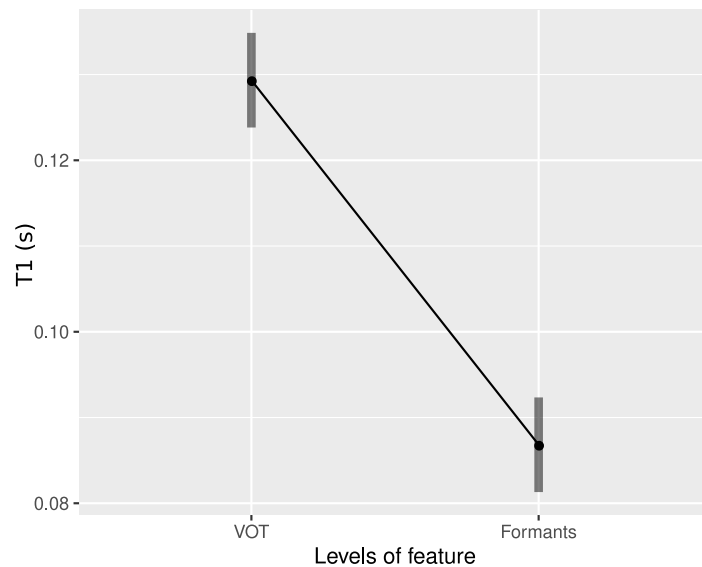


Figure 84 – Representation of the factor feature over the variable T1.

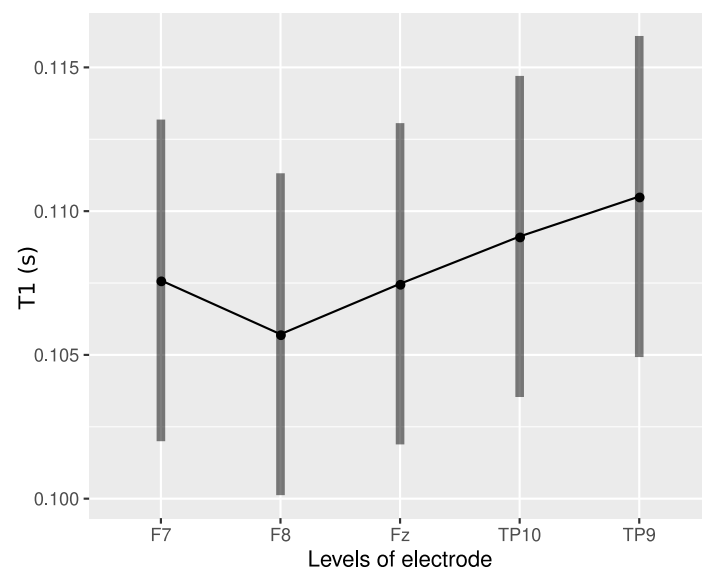


Figure 85 – Representation of the factor electrode over the variable T1.

to note that asymmetric amplitudes and latencies for the N1 component may not be directly associated with asymmetric generators inside the brain. Those differences may be due to different functions of those generators. The literature has confirmed that N1 wave receives contributions from multiple neural generators in the frontal and temporal lobes (Näätänen and Picton, 1987, Picton et al., 1999) with the primary generators specifically localized in the STG, Heschl's gyrus, and the planum temporale (Godey et al., 2001, Woldorff et al., 1993). This suggests that the different latencies are probably due to the location and participation of different generators in the processing of different characteristics of our experiment (stimuli, task, presentation rate, etc).

The contrast analysis for this factor is:

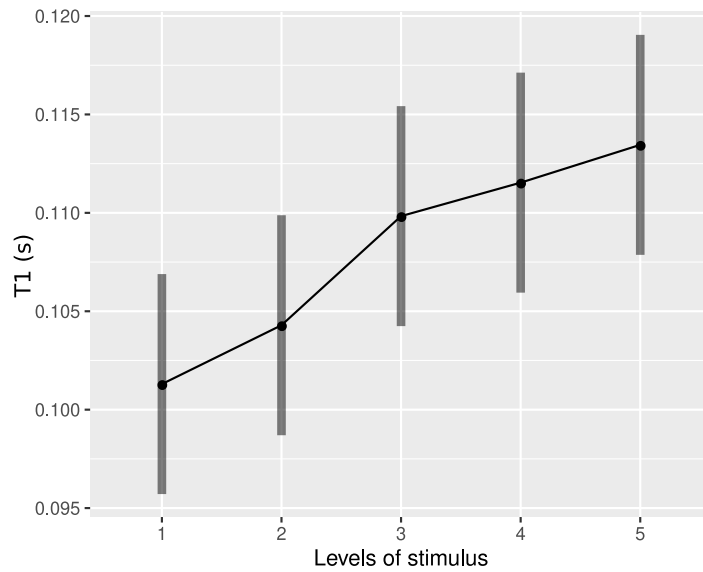


Figure 86 – Representation of the factor stimulus over the variable T1.

Results are averaged over the levels of: feature, type, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.000761	0.000810	1018	-0.940	0.3476
frontal - temporal	-0.003158	0.000732	1018	-4.314	<.0001
left - right	0.001631	0.000729	1018	2.237	0.0255

The results for the “stimulus” factor on the T1 value are shown in Figure 86. Significant effects were found for the linear contrast ( $p < 0.0001$ ) and the ambiguity contrast ( $p = 0.0072$ ). The increasing tendency observed from stim1 to stim5 may be due to the influence of the VOT response latencies, as was visually observed in the VOT AELR grand averages. For the formant continuum, it is possible to see that the N1 component do not present the same displacement in time as in the VOT case. Thus, we conclude that the latency increasing tendency observed is due to the VOT continuum AELR influence. This is better visualized by the feature:stimulus factor interaction.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, electrode

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
----------	----------	----	----	---------	---------

linear	0.015776	0.001149	1018	13.734	<.0001
phy - psy	-0.000584	0.000574	1018	-1.017	0.3095
ambiguity	-0.002189	0.000812	1018	-2.695	0.0072

The results for the interaction factor “feature:type” on the T1 value is shown in Figure 87. It has a significant effect ( $p = 0.0472$ ). It can be observed that there is not a significant difference in the N1 latency between active and passive tasks for the VOT feature but there is for the Formants case (N1 latency smaller for the active task than for the passive task). The processing negativity (Nd) component, discussed in Chapter 2, is considered an “attentional trace”, made of multiple components (Giard et al., 1988), which are influenced by several factors such as the stimulus relevance, the probability of stimulus presentation, measured region, the frequency difference between stimuli, and it also varies in latency and amplitude (Hall III, 2015).

The Nd can be related or not to the additional attention mechanism proposed by Ahveninen et al. (2011), in which through a “tuning”, attention would modulate feature selectivity of auditory neurons, in addition to AELR gain. This study concludes, for the N1 time-frame (50 – 150 ms), that “[...] a simple gain model alone cannot explain auditory selective attention. In nonprimary auditory cortices, attention-driven short-term plasticity retunes neurons to segregate relevant sounds from noise.” Thus, this mechanism may explain the difference in the N1 amplitude between the active and passive tasks (see Figure 70) and may explain the latency difference observed here for the formant continuum which have formant frequency variations. In this case, the tuning of different neurons that code different formant frequencies explains the shift in time of the N1 component between active and passive tasks (the tuning would also occur for the passive listening but have wider frequency tuning in this case (Ahveninen et al., 2011)). A shift in the  $Nd_1$  component of Nd is reported by Giard et al. (1988) for different tone frequencies. The latency of the VOT feature is greatly dictated by the VOT of the stimuli and, as the vowel is the same for all stimuli in this continuum, the frequency variation is minimal and the tuning effect only affects the amplitude of the N1 response.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: electrode, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : Passive - Active	-0.00259	0.0013	1018	-1.987	0.0472

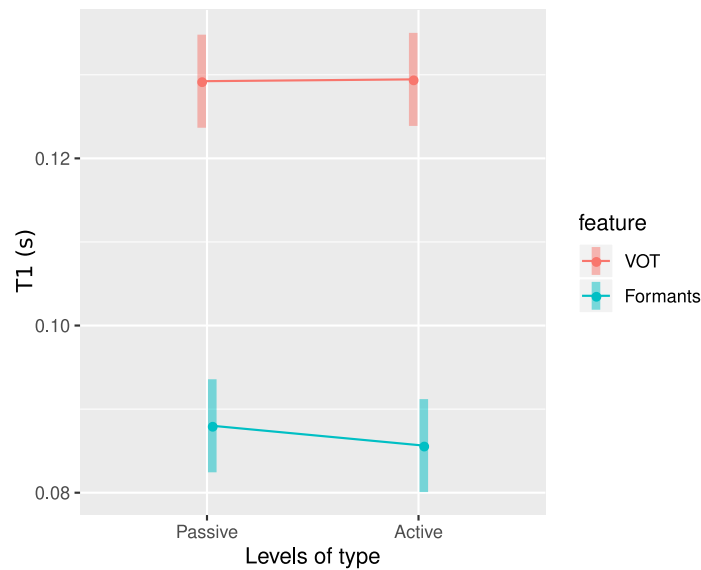


Figure 87 – Representation of the interaction factor feature:type over the variable T1.

The results for the interaction factor “feature:electrode” on the T1 value are shown in Figure 88. Significant effects were found for the medial-lateral contrast ( $p = 0.008$ ) and from frontal-temporal contrast ( $p = 0.0046$ ). For the medial-lateral contrast it is possible to see at the figure that the N1 latency is greater than the average of the lateral electrodes, in the VOT case, and that the contrary occurs for the Formants case. For the frontal-temporal contrast, it is possible to observe that, for the formant continuum, the N1 latencies for the frontal electrodes are, in average, smaller than those for the temporal region. This effect is not so pronounced for the VOT continuum. Those differences indicate that different neural generators process each acoustic cue. Specifically for the consonants (for instance, the /t/ of the VOT continuum and the /p/ of the formant continuum), [Obleser et al. \(2005\)](#) and [Pisoni and Remez \(2005\)](#) showed that different neural mechanisms support perception and discrimination of the place of articulation. So, consequently, this influences the latency. In general, it seems that formant information is first processed at frontal regions. For VOT this is not so clear, as frontal and temporal regions (TP10 specifically) present similar latencies showing that maybe different characteristics of the acoustic cue are processed simultaneously by those different cortical regions.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: type, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : medial - lateral	-0.004303	0.00162	1018	-2.659	0.0080
VOT - Formants : frontal - temporal	-0.004148	0.00146	1018	-2.841	0.0046

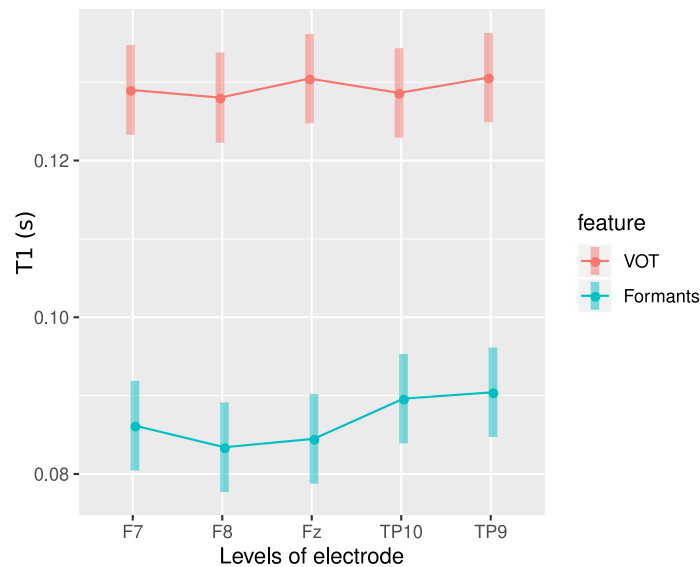


Figure 88 – Representation of the interaction factor feature:electrode over the variable T1.

VOT - Formants : left - right      0.000311 0.00146 1018 0.213 0.8313

The results for the interaction factor “type:electrode” on the T1 value are shown in Figure 89. As we go from the passive to the active task, a significant effect can be observed for frontal-temporal contrast ( $p = 0.0014$ ). It is possible to observe in the figure that the N1 latency of the response at the frontal electrode for the active task is smaller than that for the passive task. This observation is the contrary for the temporal electrodes, that is, the passive task present a N1 response faster than the active one.

Previous studies (discussed at Chapter 3) showed that, although the motor contribution to speech processing seems to occur automatically, early sensorimotor (temporal-frontal) interactions are dependent on attention (Möttönen et al., 2014, Alho et al., 2016, Chevillet et al., 2013). This integration seems to occur around N1 latency when attention is paid to the task and around P2 latency even when not paying attention. It is possible to see this effect at the frontal electrodes but not at the right ones. According to Möttönen et al. (2014), it would exist a top-down and bottom-up processing interaction between temporal (sensory/bottom) and frontal (motor/top) areas so that “[...] the auditory speech signals are transformed to motor models, which in turn affect sensory processing”. In addition they say that “[...] attention can facilitate the generation of motor models and enhance their specificity.”

Thus, we can speculate that the increase in the latency at the temporal region, with attention, may be related to some effect of this top-down processing. For example, the frontal generators involved in the speech processing when there is attention are different from those involved in passive speech processing, explaining the difference in latency. Those generators would be

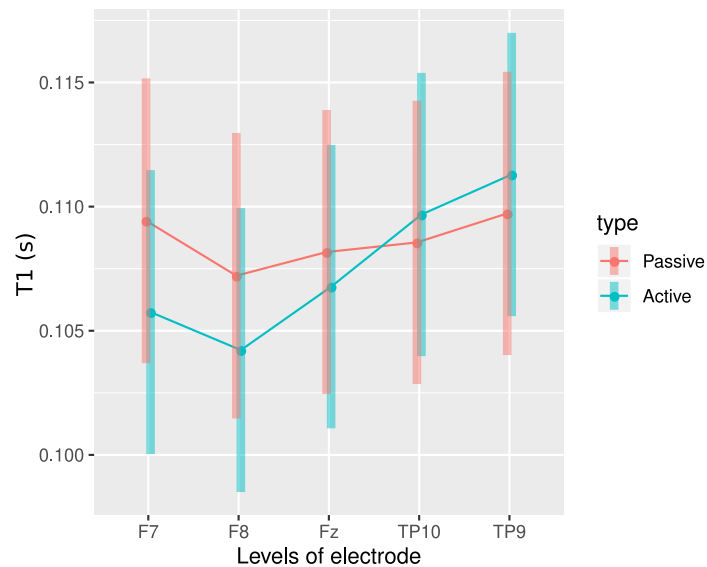


Figure 89 – Representation of the interaction factor type:electrode over the variable T1.

located with a greater distance from the temporal generators so that the top-down communication will delay the temporal response.

Another possibility (related to the former one) is that the N1 generators (or N1 components around the brain) involved are just narrowed by attention so that less neurons participates and, their firing and location pattern are such that it results in a larger potential for the active task, as observed in Figure 70. Those generators may be located in such a way that they are closer to the frontal electrodes and more distant of the temporal electrodes, justifying the difference in latency. As multiple generators are involved in the N1 component for the passive case, the combination of their potentials results in the observed latencies, that are not so different between the frontal and the temporal regions. This possibility would be also valid for the P2 latency. It would be interesting to investigate if this effect happens for other continua as well.

Then, this result confirm our previous assumption that attention may influence the speed of stimuli processing and we can suggest also the this may reflect the recruitment of different generators for this early processing at the frontal region or that attention direct a different pattern of firing of the generators involved in the speech processing.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: feature, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
----------	----------	----	----	---------	---------

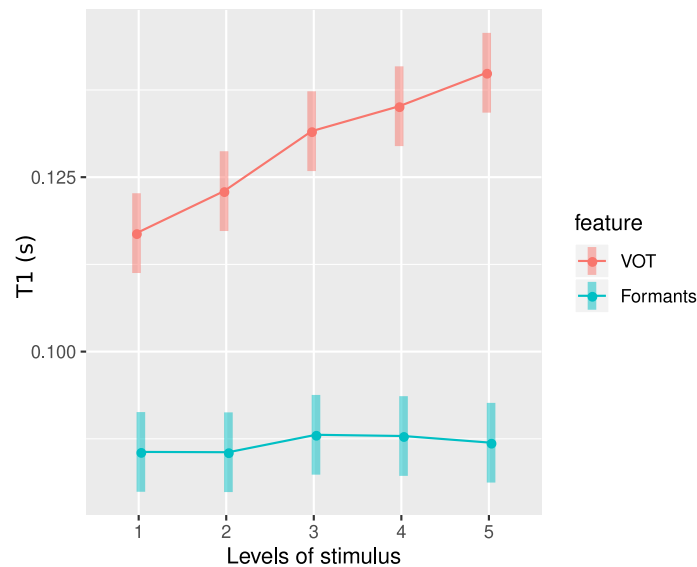


Figure 90 – Representation of the interaction factor feature:stimulus over the variable T1.

Passive - Active : medial - lateral	-0.000400	0.00162	1018	-0.247	0.8047
Passive - Active : frontal - temporal	-0.004683	0.00146	1018	-3.201	0.0014
Passive - Active : left - right	-0.000118	0.00146	1018	-0.081	0.9353

The result for the interaction factor “feature:stimulus” on the T1 value are shown in Figure 90. Significant effects were found for the linear contrast ( $p < 0.0001$ ). As already commented for the stimulus factor effect, it can be observed here that VOT response latencies may explain this result. This is consistent with the significant effect for the linear contrast.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: type, electrode  
 Degrees-of-freedom method: kenward-roger  
 Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : linear	-0.026590	0.00230	1018	-11.576	<.0001
VOT - Formants : phy - psy	-0.000492	0.00115	1018	-0.429	0.6683
VOT - Formants : ambiguity	0.001247	0.00162	1018	0.768	0.4429

The result for the interaction factor “type:stimulus” on the T1 value are shown in Figure 91. Significant effects were found for the linear contrast ( $p = 0.0027$ ). The increasing tendency in the latency, influenced by the VOT continuum (see feature:stimulus interaction for T1) is observed clearly for the active task. However, for the passive task, there is a decrease in the latency of the

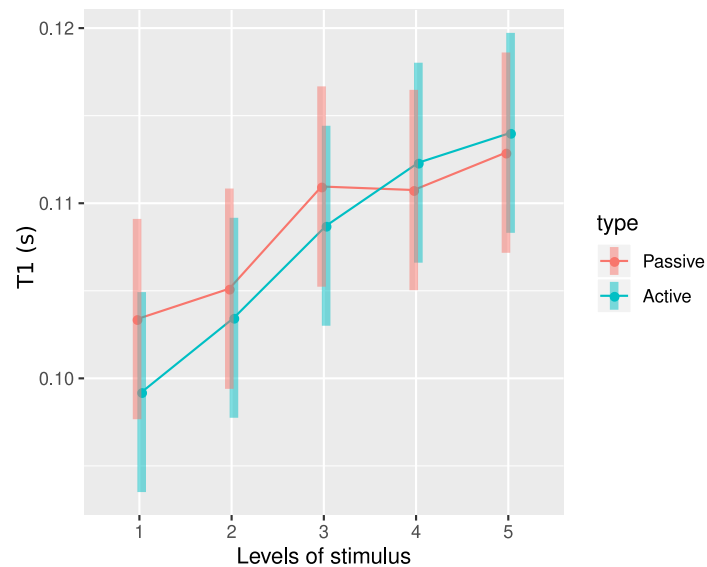


Figure 91 – Representation of the interaction factor type:stimulus over the variable T1.

stim4 and stim5 in comparison with the active task. This causes the linear difference observed between the curves. As this result is a combination of both VOT and formant stimuli, it is not possible so say for sure if the effect was due to the shift in the perception of the syllable /pɛ/ of the formant continuum or the /ta/ of the VOT continuum. However, observing the grand averages for the passive task, it is possible to see that the N1 variation in time is greater for stim4 and stim5 at the VOT continuum than at the formant continuum. Thus, again, the observed effect on the stimuli latencies for the passive task may be due to the influence of the VOT continuum stimuli.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: feature, electrode

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active : linear	0.006910	0.00230	1018	3.008	0.0027
Passive - Active : phy - psy	-0.000285	0.00115	1018	-0.248	0.8039
Passive - Active : ambiguity	0.001454	0.00162	1018	0.895	0.3710



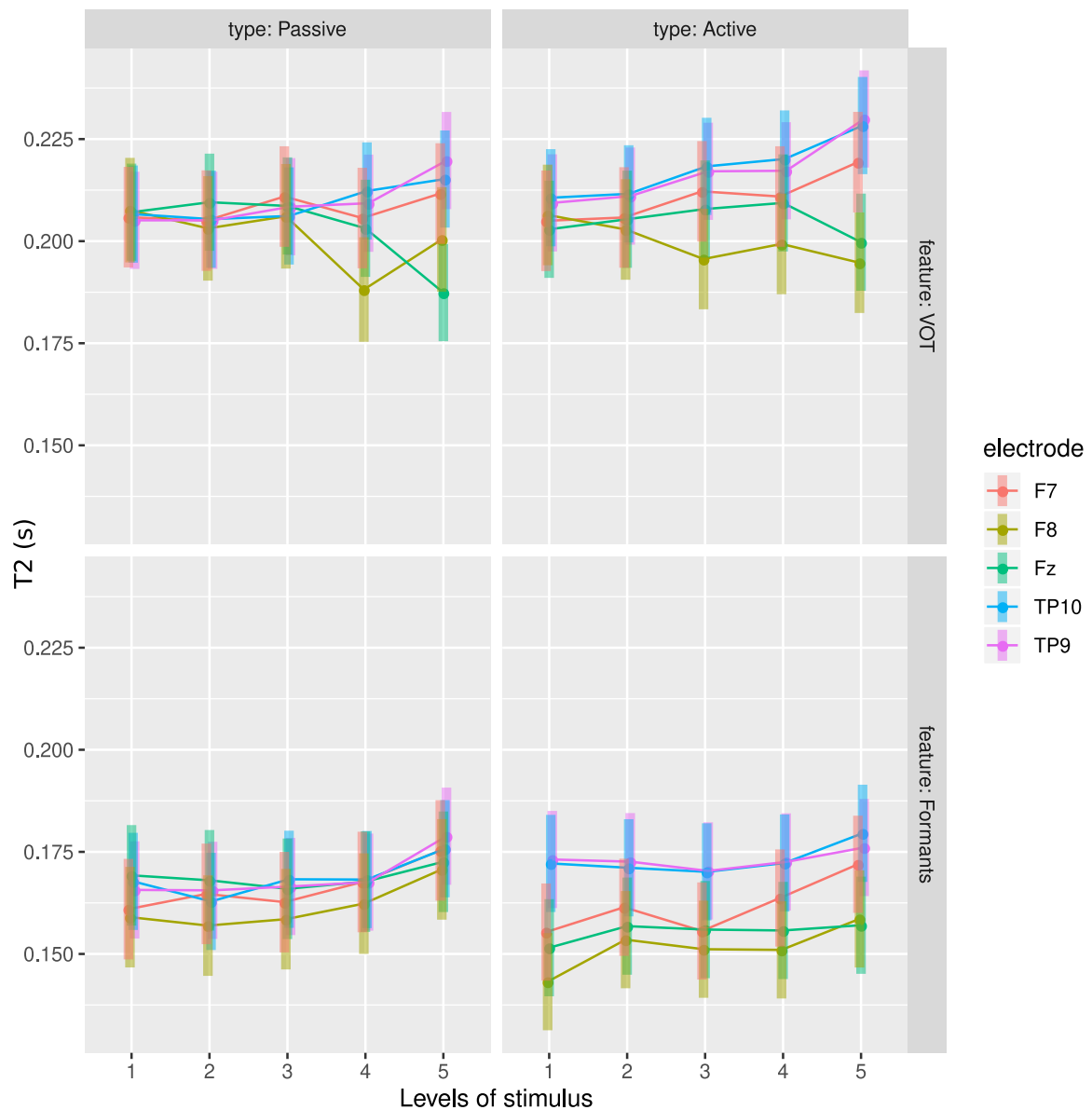


Figure 92 – Representation of the complete mixed-effects model including the fixed factors: feature, type, stimulus and electrodes. Predicted values for T2 are computed for all 11 participants.

### 6.3.5 T2 analysis

Figure 92 presents the complete mixed-effects model representation for T2 (P2 latency). The result of the ANOVA applied to this model is presented below. Significant effects were found for the main factors, feature, electrode, and stimulus and the interaction factors feature:type, type:electrode and electrode:stimulus.

The ANOVA result for the complete model (including only significant factors) is:

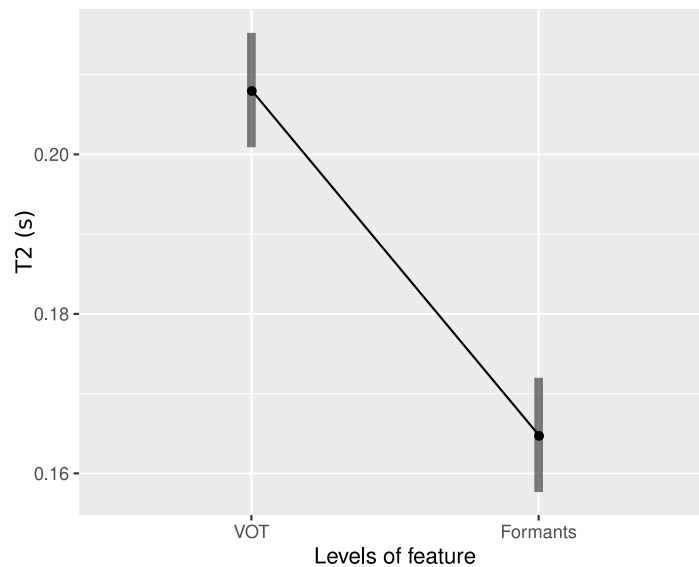


Figure 93 – Representation of the factor feature over the variable T2.

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.49348	0.49348	1	1018.2	1751.3698	< 2.2e-16
type	0.00000	0.00000	1	1018.4	0.0002	0.9876706
electrode	0.02920	0.00730	4	1018.2	25.9062	< 2.2e-16
stimulus	0.00639	0.00160	4	1018.0	5.6713	0.0001624
feature:type	0.00396	0.00396	1	1018.0	14.0656	0.0001865
type:electrode	0.00748	0.00187	4	1018.2	6.6374	2.837e-05
electrode:stimulus	0.00789	0.00049	16	1018.0	1.7511	0.0331385

The results for the “feature” factor on the T2 value is shown in Figure 93. It has a significant effect ( $p < 0.0001$ ). As happened for this factor on the T1 value, it is not possible to compare the responses between continua and, again, it is possible to observe that the latencies are greater, in average, for VOT than for Formants.

The contrast analysis for this factor is:

Results are averaged over the levels of: type, electrode, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0432	0.00103	1018	-41.849	<.0001

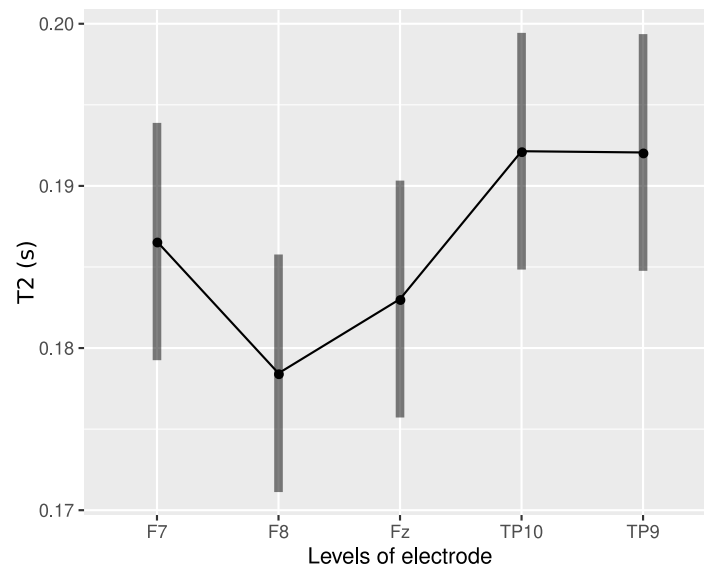


Figure 94 – Representation of the factor electrode over the variable T2.

The results for the “electrode” factor on the T2 value is shown in Figure 94. Significant effects were found for medial-lateral contrast ( $p < 0.0001$ ), the frontal-temporal contrast ( $p < 0.0001$ ), and the left-right contrast ( $p < 0.0001$ ). This same result was observed for T1 with the difference that, here, the latency difference between temporal electrodes is not significant. Generators of the P2 component include the planum temporale and the association cortex (area 22) (Godey et al., 2001). Thus, as for the T1 values, the effect we observed between electrode latencies are probably due to the location and participation of different generators in the processing of different characteristics of our experiment.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.00428	0.00128	1018	-3.332	0.0009
frontal - temporal	-0.00959	0.00116	1019	-8.261	<.0001
left - right	0.00402	0.00116	1018	3.479	0.0005

The results for the “stimulus” factor on the T2 value is shown in Figure 95. Significant effects were found for the linear contrast ( $p < 0.0001$ ). As observed for T1, there is a latency increase from stim1 to stim5, but this increase is not so pronounced here. As discussed for the T1, this effect is greatly due to the VOT continuum AELR but, as it is not so pronounced as for the

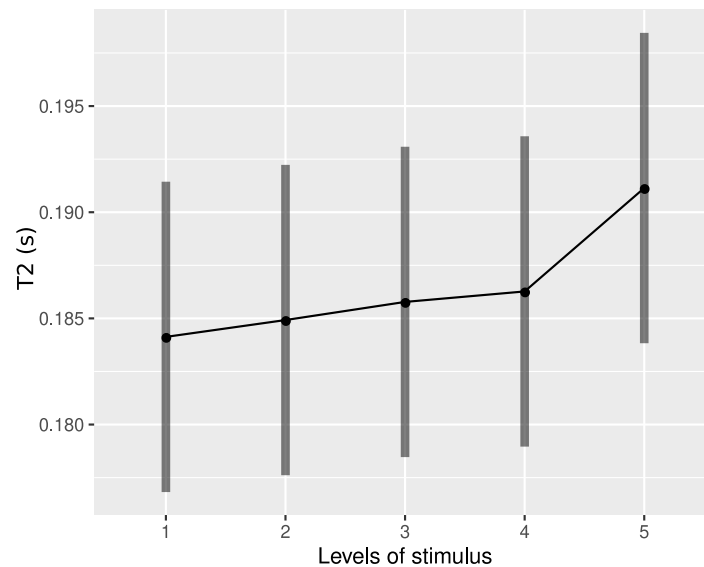


Figure 95 – Representation of the factor stimulus over the variable T2.

T1, there was not an interaction effect feature:stimulus for the T2 measurement. This result is interesting because it shows that P2 is not strongly affected by the VOT as N1. This smaller influence of the VOT over the P2 wave was also observed for the amplitude analysis.

The contrast analysis for this factor is:

Results are averaged over the levels of: feature, type, electrode

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
linear	0.007684	0.001824	1018	4.212	<.0001
phy - psy	0.001079	0.000912	1018	1.183	0.2370
ambiguity	0.000839	0.001290	1018	0.650	0.5156

The results for the factor interaction “feature:type” on the T2 value is shown in Figure 96. It has a significant effect ( $p = 0.0002$ ). For the formant continuum case, the shorter latency observed for the active task can be explained in the same way as it was for T1. However, P2 present a variation for the VOT feature between tasks. Here, for the Formants case, there is an increase in the latency for the active case. This may have to do with the shape of the Nd component that will be different for each continuum and would change differently the P2 amplitude and latency when both components overlap.

This result, together with this same factors interaction observed for the T1 dependent variable,

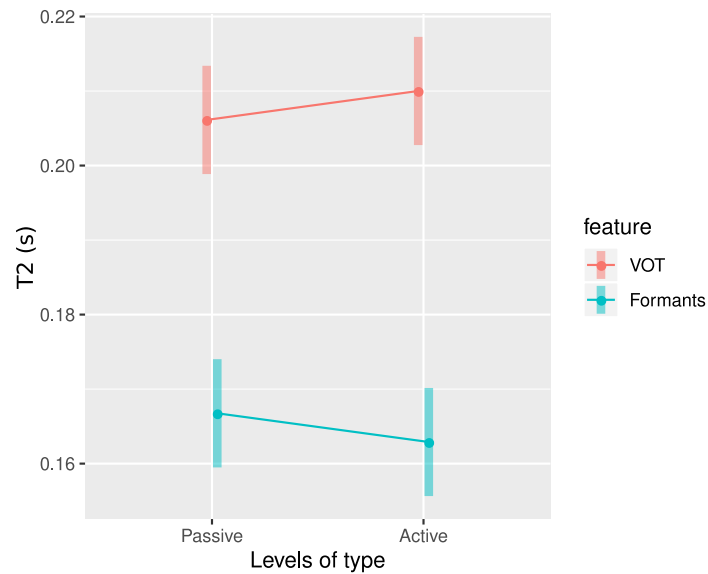


Figure 96 – Representation of the interaction factor feature:type over the variable T2.

shows that stimuli are processed differently when there is attention to the task and that formants and VOT evoke different behaviors in N1 and P2 generators as we assumed previously.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: electrode, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : Passive - Active	-0.00774	0.00206	1018	-3.750	0.0002

The results for the interaction factor “type:electrode” on the T2 value is shown in Figure 97. Significant effects were found for medial-lateral contrast ( $p < 0.01$ ) and the frontal-temporal contrast ( $p < 0.0001$ ). The explanation for the observed effects is similar to the case of this interaction factor on the T1 variable. The difference here is that the medial-lateral contrast was also significant but the same explanation can be applied.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: feature, stimulus

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

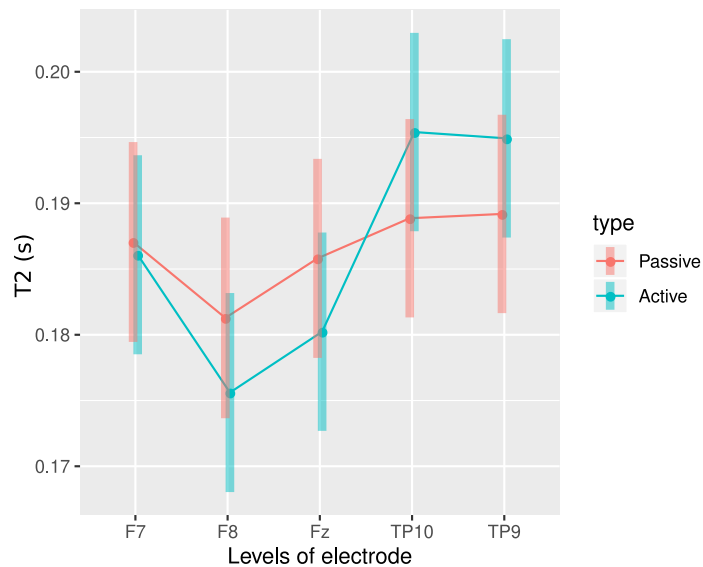


Figure 97 – Representation of the interaction factor type:electrode over the variable T2.

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active : medial - lateral	-0.00699	0.00257	1018	-2.720	0.0066
Passive - Active : frontal - temporal	-0.00947	0.00232	1018	-4.082	<.0001
Passive - Active : left - right	0.00195	0.00231	1018	0.843	0.3997

The results for the interaction factor “electrode:stimulus” on the T2 value are shown in Figure 98. Significant effects were found for the linear contrast ( $p = 0.0011$ ). The medial-lateral effect is illustrated by the curves for the medial electrode Fz and the general behavior of the lateral electrodes F7, F8, TP9 and TP10. It is possible to see a decrease in the latency from stim3 to stim5 for electrode Fz, while there is a general increase for the other electrodes. Consequently, the shape of the curves puts in evidence the linear effect observed. Observing Figure 92, it is possible to see that, in general, electrodes F7, F8, TP9 and TP10 have an increasing tendency in P2 latency for both Formants and VOT cases, while the decreasing tendency observed for the Fz electrode is due to the influence of the VOT continuum only. This shows that for the processing (perception) of syllables similar to /ta/, the P2 latency at medial regions of the scalp is smaller than for the syllables similar to /da/ while for lateral regions this effect is reversed. Considering the electrodes individually, F7 and F8 have very different behaviors with lower latency for F8. This may have to do with the location of generators that process a given characteristic of the stimuli or task in this experiment.

The contrast analysis for this interaction factor is:

Results are averaged over the levels of: feature, type

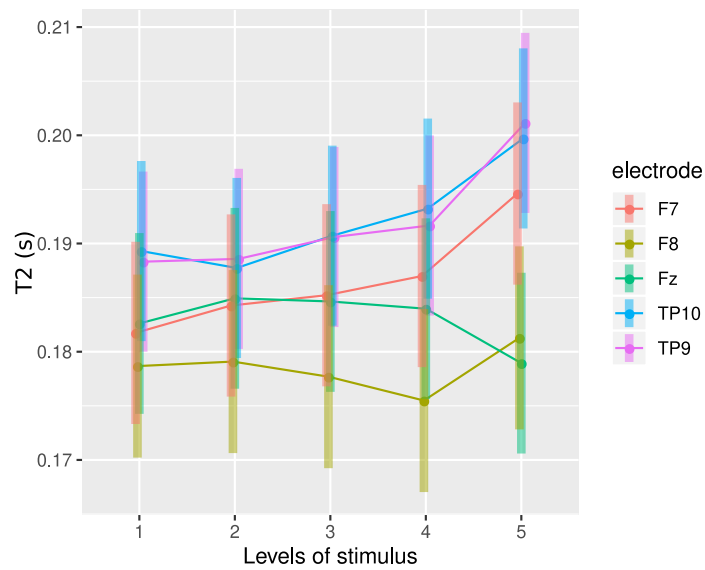


Figure 98 – Representation of the interaction factor electrode:stimulus over the variable T2.

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral : linear	-0.014788	0.00453	1018	-3.262	0.0011
medial - lateral : phy - psy	-0.001895	0.00227	1018	-0.836	0.4034
medial - lateral : ambiguity	-0.003584	0.00321	1018	-1.118	0.2638
frontal - temporal : linear	-0.006233	0.00409	1018	-1.525	0.1275
frontal - temporal : phy - psy	0.001394	0.00204	1018	0.682	0.4953
frontal - temporal : ambiguity	-0.000487	0.00289	1018	-0.168	0.8663
left - right : linear	0.007321	0.00409	1018	1.791	0.0736
left - right : phy - psy	0.000603	0.00204	1018	0.295	0.7678
left - right : ambiguity	0.000402	0.00289	1018	0.139	0.8893

### 6.3.6 $\Delta$ ERP

Testing how the  $\Delta$ ERP variable, computed for N1, P2 and N1-P2, varies with the same fixed factors analyzed before, some effects were found and are reported as follows. It is important to remember that for N1-P2 magnitude and P2 amplitude, a positive value for the dependent variable is in accordance with the result reported by [Bidelman and Walker \(2017\)](#), according to which the unambiguous stimuli stim1/stim5 have greater amplitudes, in average, than the ambiguous stimulus stim3. For the N1, the reverse is true, so negative values are expected for the dependent variable.

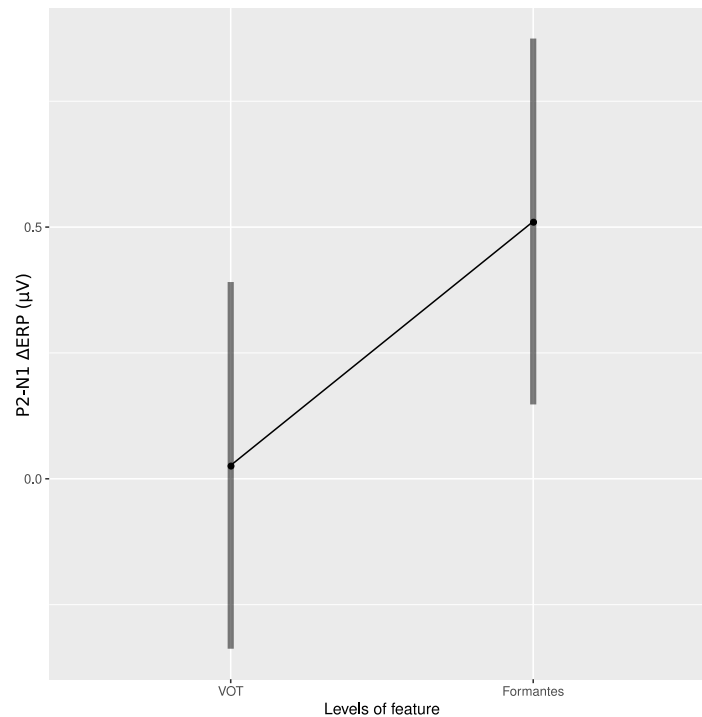


Figure 99 – Analysis of the effect of the factor feature over the  $\Delta$ ERP computed for the N1-P2 magnitudes. The formants feature presents, in average, a positive value for the variable that is significantly different that the value for the VOT feature which is near zero.

For the N1-P2 magnitude, only the feature factor has a significant effect ( $p < 0.0001$ ). Figure 99 shows that for formants the  $\Delta$ ERP have, in average, a positive value. This is consistent with the results of [Bidelman and Walker \(2017\)](#), who also worked with this feature. However, for **VOT**, the result is close to zero, showing that there is not a significant difference between the ambiguous and unambiguous stimuli in this case.

The ANOVA result for the complete model (including only significant factors) is:

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	12.41	12.41	1	200.2	13.8	0.0002638

For the N1 amplitude, an effect was also observed only for the feature factor ( $p = 0.0015$ ). Figure 100 shows that the expected negative value for this dependent variable occurs for the Formants case, indicating that N1 is greater for the unambiguous stimuli than for the ambiguous one. However, for the **VOT** case, the value is positive. This happens because N1 is strongly affected by the **VOT** of the stimulus. As can be seen in Figures 47, 48, 51, and 52, the value of N1 for stim1 is smaller than that for stim3 and stim5, independently of the task. Also, the value for stim3 is closer to that for stim5 than it is for the stim1. Thus, when we compute the mean



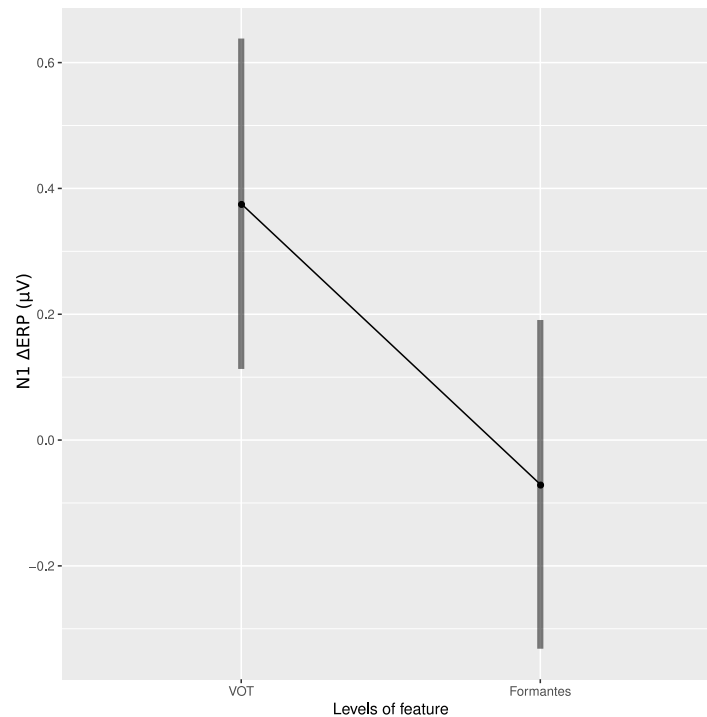


Figure 100 – Analysis of the effect of the factor feature over the  $\Delta$ ERP computed for the N1 amplitude. The formants feature presents, in average, a negative value for the variable that is significantly different that the value for the VOT feature which positive. For interpretation is important to point out that N1 usually has negative values.

value between stim1 and stim5, the resulting N1 value is smaller than that for stim3 because of the small value of stim1. In conclusion, even if the amplitude N1 conveys information on the ambiguity, as can be concluded for the Formants case (see Figures 53 and 54), this is not observed in the the VOT case, because this feature affects N1 much more. It is known that changes in voice onset time are evident by the time they reach auditory cortex and are reflected in amplitude and latency of the N1 response (Steinschneider et al., 1995, Eggermont, 1995).

It is important to notice that the increase in the ISI from stimulus stim1 to stim5, for the VOT continuum, can also cause the differences in N1 amplitude observed here.

The ANOVA result for the complete model (including only significant factors) is:

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	10.53	10.53	1	200.15	10.351	0.001509

For the P2 amplitude, effects were observed only for the electrode factor ( $p = 0.0254$ ). Figure 101 shows that the value of  $\Delta$ ERP increases significantly from medial to temporal electrodes ( $p <$

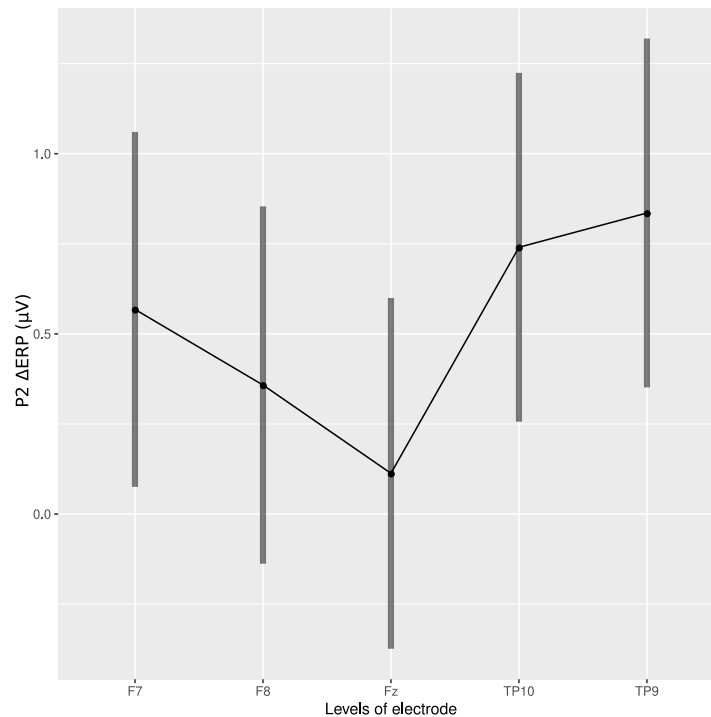


Figure 101 – Analysis of the effect of the factor electrode over the  $\Delta$ ERP computed for the P2 amplitude. It can be observed that the variable is smaller at the medial region (Fz) than at more lateral regions (F7-F8, TP9-TP10).

0.01). The P2 component of the AELR is related to auditory object representation in the literature (Tremblay et al., 2014, Ross et al., 2013) and the areas in the anterior superior temporal plane have been shown to be responsive to auditory objects (Ross et al., 2013, Leaver and Rauschecker, 2010). This fact may explain why amplitudes are greater, in general, at more lateral regions than in medial ones for P2 and N1-P2. As expected, the  $\Delta$ ERP have positive values for all electrodes and the feature factor have no significant effect. This absence of result for the feature factor is interesting and suggests that there was not significant difference between P2  $\Delta$ ERP of both features.

The effect of stim1, stim3 and stim5 for the VOT and the Formants cases on  $\Delta$ ERP can be visualized in Figure 102. As can be seen, the value for stim3 is smaller than those for stim1 and stim5 for both Formants and VOT cases. This explains the positive results obtained for  $\Delta$ ERP and the absence of effect for the feature factor. This shows that the VOT physical characteristic do not affect P2 as it affects N1, and suggests that a higher-level processing of speech occurs at that latency. For instance, Ross et al. (2013) reports that the object representation, coded by the P2 amplitude, is important for learning and “[...] allows the listener to access details in the sensory representation, which in turn permits the correct identification of phonetically similar objects and potentially even categorical perception”. Besides that, there is no significant effect for the factor type, probably because  $\Delta$ ERP is a difference value and it is not affected by the greater amplitudes in the active task, in comparison to those for the passive task.

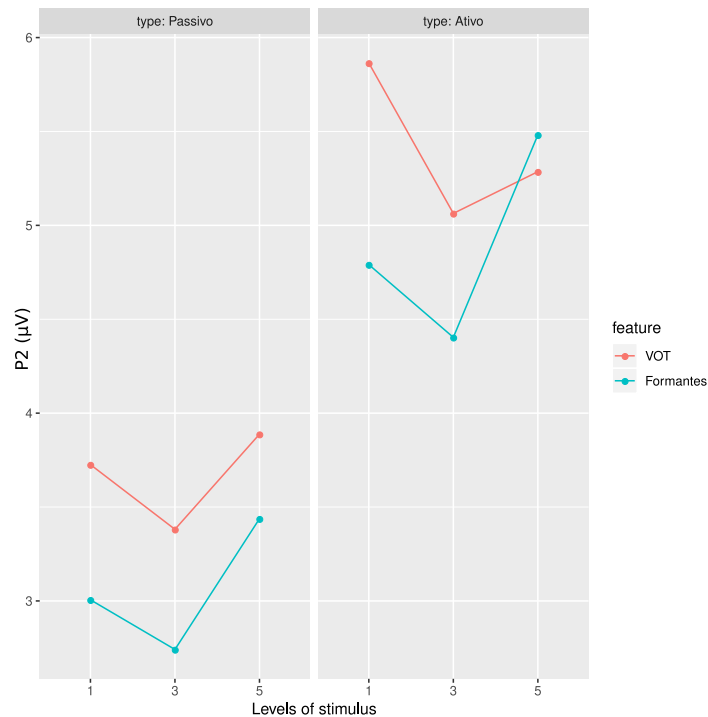


Figure 102 – Relation of the factors feature, type and stimulus for the P2 amplitude values considering stim1, stim3 and stim5 used to compute the  $\Delta$ ERP variable.

The ANOVA result for the complete model (including only significant factors) and the contrast for the electrode factor is:

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
electrode	14.759	3.6898	4	197.21	2.8404	0.02544

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.512	0.195	197	-2.629	0.0093
frontal - temporal	-0.325	0.176	198	-1.845	0.0665
left - right	0.153	0.176	197	0.869	0.3860

In conclusion, [Ross et al. \(2013\)](#) reports that that the N1 component (its amplitude and latency) seems to reflect physical characteristics of the stimuli processed at the auditory cortex but that the perception of VOT, which is necessary to categorize phonemes, depends on further processing and is also influenced by experience. These authors also show that the categorical perception may occur at the P2 latency. Other works also report that the N1 component are more affected by acoustic stimuli characteristics than by their category and this has been verified for both VOT and formant continua ([Toscano et al., 2010](#), [Bidelman et al., 2013](#)). The effect of the feature factor on  $\Delta$ ERP computed with N1, suggests different treatments for ambiguous and unambiguous

stimuli and also a difference in amplitudes of stim1 and stim2 related to stim3, stim4 and stim5 (for the VOT continuum). This can also be concluded from a visual inspection of the grand averages. This indicates a possible categorical coding happening at N1 latency. [Bidelman and Walker \(2017\)](#) reports that the neurophysiological underpinnings of categorization of speech sounds are present in the cortex around 175 ms after stimulus onset, in the time frame of the N1-P2 complex. However, our results show that different processing events are coded by the N1 and P2 components and they would not be observed if we used just the N1-P2 magnitude in our analysis.

## Chapter 7

# REGRESSION ON LOW-DIMENSION SPANNED INPUT SPACE – RoLDSIS

In the preceding chapters we presented time-domain analyses of the ERPs. In this chapter we present the results of time-frequency analysis using a novel regression technique called [RoLDSIS](#). The present chapter is adapted from the contents of a recent publication ([Santana et al., 2020](#)).

We describe here in detail the [RoLDSIS](#) technique. We then illustrate its use for the specific case of the [VOT](#) feature and the TP9 electrode data. We compared the results with other regularized linear regression techniques. We showed that the angle between the obtained psychophysical and physical directions (obtained from RoLDSIS) correlates with the slope of the psychometric curve.

### 7.1 Background

Functional brain imaging experiments are currently used in studies that aim to identify the neurophysiological correlates of perception. In these experiments, it is assumed that a given perceptual stimulus evokes a specific pattern of neuronal activity in the central nervous system. This activity can be captured through a variety of measurements, like electric potentials in [EEG](#) and [ECoG](#), magnetic fields in [MEG](#), blood flow changes in near infrared spectroscopy ([NIRS](#)), or haemodynamic response in [fMRI](#). The recorded signals are usually represented in time and frequency (through spectro-temporal analysis, like Fourier or wavelet transforms), as well as in

the physical space (EEG or MEG sensors, or fMRI voxels).

These measurements represent the evoked response in the brain and can be mathematically represented as vectors in an  $\mathbb{R}^N$  space, where  $N$  is the total number of *features* used to represent the EEG measurements. Each feature corresponds to a discrete point in time, frequency, and spatial domains. The dimension of this representation space is usually very high. For instance, consider an EEG experiment with 64 electrodes in which the ERP lasts for 0.5 s and is represented in the time-frequency domain by a spectrogram with ten binned frequency bands and sampled in time every 1 ms. This would result in a representation space containing  $64 \times 500 \times 10 = 320,000$  features. Such high dimensions are not uncommon in brain imaging studies.

In EEG experiments, the ERP evoked by the stimulus corresponds to electric potential fluctuations which are very small in comparison with the ongoing, background electric activity measured on the scalp. In order to obtain reliable measures of the ERP for each stimulus, it is necessary to average the responses across a large amount of trials. Depending on the desired SNR, several hundreds, or sometimes thousands of trials are required to obtain reliable ERPs Luck (2014). This requirement imposed by the SNR is also critical in other cases, such as in studies of epileptic seizures, in which measurements may take up days in order to detect epileptogenic zones Gajic et al. (2014), Birjandtalab et al. (2017), and in brain-computer interface (BCI) systems that rely on a small amount of EEG observations for inferring the intention of the user (Tu et al., 2014, Sturm et al., 2016). At any rate, due to time limitations in recording the data for a single participant, typical EEG experiments involve a limited amount of *observations*, which are the ERPs for each stimulus.

In this thesis, we are interested in the neurophysiological correlates of perception, in the context of such EEG experiments that fall in the high dimension low sample size (HDLSS) case. We will assume that each stimulus  $i$  used in the experiment can be characterized by a scalar *attribute*  $y_i \in \mathbb{R}$ . We also assume that this attribute has a functional relationship with the evoked response  $\mathbf{x}_i$ , written as  $y = f(\mathbf{x})$ . We will consider here the simplest, linear approximation for this relationship, the affine transformation:

$$y = a + \mathbf{b}^\top \mathbf{x}. \quad (7.1)$$

The vector  $\mathbf{b} \in \mathbb{R}^N$  represents the *neurophysiological axis* related to the attribute  $y$  or, in other words, how the features in  $\mathbf{x}$  must be combined in order to yield the value associated with the stimulus attribute  $y$ . The vector  $\mathbf{b}$  and the scalar constant  $a$  must be inferred from  $M$  pairs of observations  $\{\mathbf{x}_i, y_i\}$ .

Since the affine relationship is only an approximation to the real world, the  $M$  observations are related through the equation:

$$y_i = a + \mathbf{b}^\top \mathbf{x}_i + \varepsilon_i, i = 1, \dots, M, \quad (7.2)$$

where the error  $\varepsilon_i$  is assumed to be independent and normally distributed. The technique for solving this problem is called regression and its goal is to minimize the quadratic error function

$$E(a, \mathbf{b} | \{\mathbf{x}_i, y_i\}) = \sum_{i=1}^M \varepsilon_i^2 = \sum_{i=1}^M (y_i - a - \mathbf{b}^\top \mathbf{x}_i)^2. \quad (7.3)$$

When  $M < N$ , the problem is underdetermined, meaning that there is an infinite number of values for  $a$  and  $\mathbf{b}$  that yield an optimal solution. Techniques of regularization or variable selection, such as **LASSO**, Ridge regression (Friedman et al., 2010a), and sparse partial least squares (**SPLS**) (Chun and Keleş, 2010), can be used to obtain a well-posed problem, formulated as:

$$\min_{\{a, \mathbf{b}\}} [E(a, \mathbf{b} | \{\mathbf{x}_i, y_i\}) + \lambda P(\mathbf{b})], \quad (7.4)$$

where  $\lambda$  is a regularization parameter and  $P$  is a penalty function for the regression coefficients in vector  $\mathbf{b}$ . In general, the parameter  $\lambda$  cannot be determined *a priori* and must be inferred from the data, using some kind of cross-validation procedure. This is possible when there is an abundant number of pairs of observations  $\{\mathbf{x}_i, y_i\}$  in order to feed the cross-validation procedure.

As described above, in **EEG** experiments, the observations are extremely scarce. For this specific case, we propose the application of a projection technique, reminiscent of the dimension reduction methods described in (James et al., 2013), that avoids the problem of specifying regularization parameters when the number of observations is very small, called **RoLDSIS**.

### 7.1.1 The RoLDSIS technique

The main idea behind the application of this technique is to assume that the neurophysiological axis  $\mathbf{b}$  is restricted to the  $(M - 1)$ -dimensional subspace spanned by the  $M$  linearly independent

points  $\mathbf{x}_i$

$$\mathbf{x}_i = [x_{i1} \ x_{i2} \ \dots \ x_{iN}]^T, \quad i = 1, \dots, M, \quad (7.5)$$

where  $N$  is the number of features contained in each grand-averaged response,  $M$  is the number of observations available and  $^T$  denotes transpose.

This subspace is described by an origin point  $\mathbf{x}_0 \in \mathbb{R}^N$ , that is a linear combination of  $\mathbf{x}_i$ , and by an  $N \times (M - 1)$  matrix  $\mathbf{V}$ , whose columns represent the vectors of an orthonormal basis of the  $(M - 1)$ -dimensional subspace spanned by the  $M$  points  $\mathbf{x}_i$  which should not be collinear.  $\mathbf{x}_0$  can be taken as the mean of  $\mathbf{x}_i$

$$\mathbf{x}_0 = \mathbf{m} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_i, \quad (7.6)$$

and principal component analysis (PCA) can be used to obtain  $\mathbf{V}$  as

$$\mathbf{V} = (\mathbf{X} - \mathbf{m})\mathbf{U}\mathbf{S}^{-\frac{1}{2}}, \quad (7.7)$$

where

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_M], \quad (7.8)$$

and  $\mathbf{U}$  and  $\mathbf{S}$  are the result of singular value decomposition

$$\mathbf{U}\mathbf{S}\mathbf{U}^T = (\mathbf{X} - \mathbf{m})^T(\mathbf{X} - \mathbf{m}). \quad (7.9)$$

Any point  $\mathbf{x} \in \mathbb{R}^N$  is projected onto the spanned subspace through the transformation

$$\mathbf{z} = \mathbf{V}^T(\mathbf{x} - \mathbf{m}), \quad (7.10)$$

where  $\mathbf{z}$  is an  $(M - 1)$ -dimensional vector from which the components of  $\mathbf{x}$ , contained in the spanned subspace, are determined

$$\mathbf{x} = \mathbf{V}\mathbf{z} + \mathbf{m} + \mathbf{e}, \quad (7.11)$$

where  $\mathbf{e}$  is the part of  $\mathbf{x}$  that is not contained in the spanned subspace. For the particular case of  $\mathbf{x}_i$ ,  $i = 1, \dots, M$ , which are contained in the spanned subspace,  $\mathbf{e} = 0$ ,

$$\mathbf{z}_i = \mathbf{V}^T(\mathbf{x}_i - \mathbf{m}), \quad (7.12)$$

$$\mathbf{x}_i = \mathbf{V}\mathbf{z}_i + \mathbf{m}. \quad (7.13)$$

It is now possible to express the equation

$$y = a + \mathbf{b}^T \mathbf{x}, \quad (7.14)$$



which has  $N + 1$  unknowns, as

$$y = a + \mathbf{b}^\top(\mathbf{V}\mathbf{z} + \mathbf{m}), \quad (7.15)$$

$$= a + \mathbf{b}^\top\mathbf{m} + \mathbf{b}^\top\mathbf{V}\mathbf{z}, \quad (7.16)$$

$$= c + \mathbf{d}^\top\mathbf{z}, \quad (7.17)$$

which has  $M$  unknowns: the scalar  $c = a + \mathbf{b}^\top\mathbf{m}$  and the  $(M - 1)$  components of the vector  $\mathbf{d} = \mathbf{V}^\top\mathbf{b}$ . Now, the  $M$  pairs of observations  $\{\mathbf{x}_i, y_i\}$  can be used to define a linear system with  $M$  unknowns and  $M$  equations:

$$y_i = c + \mathbf{d}^\top\mathbf{z}_i, \quad i = 1, \dots, M, \quad (7.18)$$

which can be exactly solved to find  $c$  and  $\mathbf{d}$ . Next, it is possible to use

$$\mathbf{d} = \mathbf{V}^\top\mathbf{b}, \quad (7.19)$$

$$c = a + \mathbf{b}^\top\mathbf{m} \quad (7.20)$$

to determine  $a \in \mathbb{R}$  and  $\mathbf{b} \in \mathbb{R}^N$  restricted to the  $(M - 1)$ -dimensional subspace spanned by the  $M$  points  $\mathbf{x}_i$

$$\mathbf{b} = \mathbf{V}\mathbf{d}, \quad (7.21)$$

$$a = c - \mathbf{d}^\top\mathbf{V}^\top\mathbf{m} \quad (7.22)$$

Finally, the original observations  $\mathbf{x}_i$  can be projected onto the normalized neurophysiological axis  $\mathbf{b}$  given by,

$$\hat{\mathbf{b}} = \mathbf{b}/\|\mathbf{b}\|, \quad (7.23)$$

yielding the representations

$$\tilde{\mathbf{x}}_i = \mathbf{m} + \hat{\mathbf{b}}[\hat{\mathbf{b}}^\top(\mathbf{x}_i - \mathbf{m})], \quad (7.24)$$

where  $\mathbf{m}$  is the mean of the  $\mathbf{x}_i$  observations. The properties of the projections  $\tilde{\mathbf{x}}_i$  can then be further analyzed in order to infer the underlying brain states related to the physical or psychophysical measurements  $y_i$ . Fig. 103 illustrates the RoLDSIS technique for the case of  $M = 3$  observation points  $\mathbf{x}_i \in \mathbb{R}^3$  spanning a two-dimensional subspace. This figure also illustrates an example

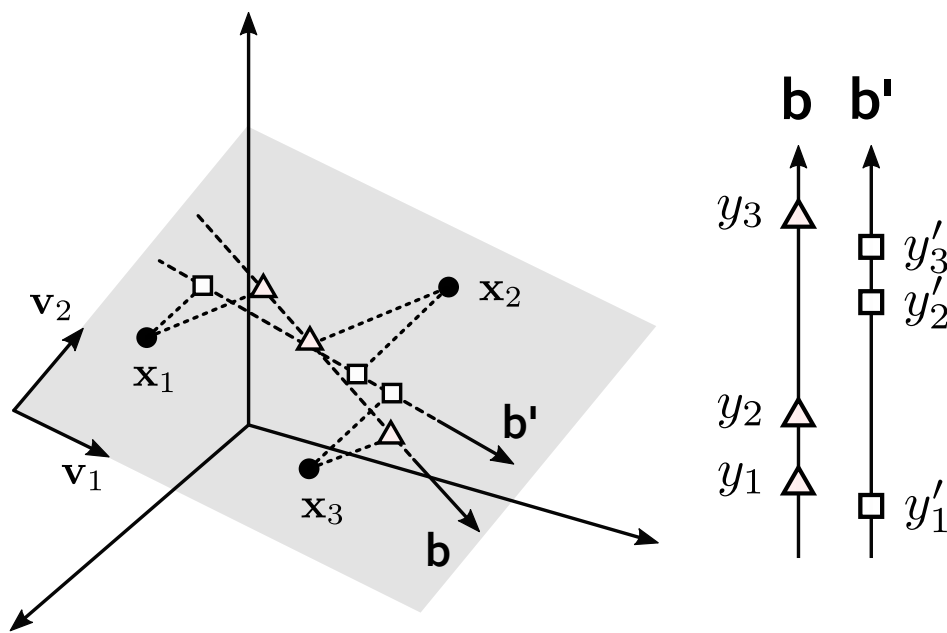


Figure 103 – Graphic representation of the RoLDSIS technique (see section 7.1.1 for details).

with two different outcomes ( $\mathbf{b}$  and  $\mathbf{b}'$ ) using three observation points. The RoLDSIS results in two different neurophysiological axes, one for each outcome, where the distances between the projections follow those of the outcomes  $y_i$  and  $y'_i$  (with an error). The angle between those two axes can be used as a measure of the difference between the neurophysiological representation of both outcomes.

## 7.2 Example of application of RoLDSIS

To show how to apply and interpret the results of the RoLDSIS we applied it to the temporal left signal (TP9 electrode) of a representative participant for the VOT-active condition considering the psychophysical response ( $y = [0, 0.05, 0.5, 0.95, 1]$ ) and the physical response ( $y = [1, 45, 88, 134, 200]$ ). The Figure 104 illustrates the psychometric curve for the VOT continuum, for the representative participant used to illustrate this example <sup>1</sup>.

The DWT predictors matrix was already presented at the Section 5.2. As we saw in the Chapter 5, the DWT yields a set of 2048 coefficients, organized in blocks. The first block contains the so-called *approximation coefficients* (V) and comprise a low-pass filtered representation of the signal. The remaining blocks contain the *detail coefficients* (W), which comprises the high frequency information. These coefficients are obtained by convolving the signal with a band-

<sup>1</sup> Values of the physical response are represented in terms of the VOT but correspond to the stimuli numbered between 1 and 200 in the physical response vector.

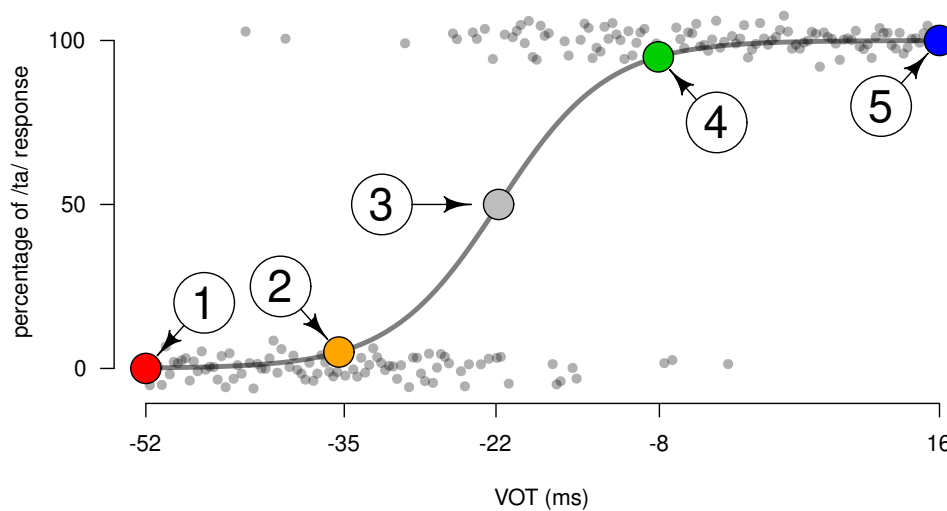


Figure 104 – Results of the phonemic identification task for a representative participant. Responses to the 200 stimuli, each one for a specific value of voice onset time (along the horizontal axis) are shown as gray dots around 0.0 (for /da/ responses) and around 1.0 (for /ta/ responses). Vertical jitter has been added for the sake of clarity. The gray curve is the theoretical psychometric response fitted to the data. Choices of stimuli stim1, stim2, stim3, stim4, and stim5, corresponding to 0%, 5%, 50%, 95% and 100% of /ta/ responses, respectively, are shown by colored dots on the psychometric curve. The VOT values for the stimuli are indicated in the horizontal axis.

pass filter based on the *mother wavelet* McKay et al. (2013). We used as mother wavelet the Daubechies orthonormal compactly supported wavelet of length 8, from the least asymmetric family, available in the package wavelets of the R software Aldrich (2013). Only the DWT coefficients corresponding to the low frequency bands (W5, W6, W7, W8 and V8 - between 0 and 156 Hz) were retained, resulting in a feature vector of length 128. This range of frequencies covers the bands  $\theta$ ,  $\alpha$ ,  $\beta$ , and  $\gamma$ , which are of interest in brain electrophysiology studies of speech perception Giraud and Poeppel (2012), Bidelman (2015). The feature vectors were averaged across trials for each participant and each stimulus. The RoLDSIS technique was then applied to these averaged observations  $\mathbf{x}_i, i = 1, \dots, 5$ .

Of course, regularization techniques cannot be applied in this case due to the small amount of observations (sample size) which make difficult to perform the cross-validation to tune the regression parameters. There is not enough observations to separate a training and test group for the cross-validation. However, it is possible to apply the RoLDSIS technique to this data set.

As explained before, each stimulus  $i$  is associated both with a specific physical attribute  $\phi_i$  (the VOT value for the associated stimulus) and with a specific psychophysical attribute  $\psi_i$  (the proportion of /ta/ response for the associated stimulus, obtained from the psychometric curve). For the psychophysical attribute, we used the proportions of /ta/ responses corresponding to the

selected stimuli  $\psi_1 = 0.0$ ,  $\psi_2 = 0.05$ ,  $\psi_3 = 0.5$ ,  $\psi_4 = 0.95$ , and  $\psi_5 = 1.0$ . Those proportions were selected so that stim1 and 2 would be closer to each other as well as stim4 and 5, while these four stimuli would be distant of the ambiguous stim3. For the physical attributes, we used the VOT of the selected stimuli. The first ( $\phi_1$ ) and last ( $\phi_5$ ) values were equal for all participants and corresponded to the /da/ and /ta/ stimuli at the beginning and at the end of the continuum (−52 ms and +16 ms, respectively). The other three values varied for each participant, since the psychometric curve is idiosyncratic. For instance, for the participant whose psychometric curve is depicted in Figure 104, the physical attributes were  $\phi_2 = -35$  ms,  $\phi_3 = -22$  ms, and  $\phi_4 = -8$  ms.

We hypothesize the following linear relationships, for  $i = 1, \dots, 5$ :

$$\phi_i = a_\Phi + \mathbf{b}_\Phi^\top \mathbf{x}_i, \quad (7.25)$$

$$\psi_i = a_\Psi + \mathbf{b}_\Psi^\top \mathbf{x}_i. \quad (7.26)$$

We assume that  $\mathbf{b}_\Phi$  and  $\mathbf{b}_\Psi$  are unit vectors (*ie*  $\|\mathbf{b}\| = 1$ ). Since  $\mathbf{x}_i \in \mathbb{R}^{128}$ , vectors  $\mathbf{b}_\Phi$  and  $\mathbf{b}_\Psi$  have 127 free coefficients to be determined. Considering also the scalar parameters  $a_\Phi$  and  $a_\Psi$ , each equation above results in a system of 5 linear equations with 128 unknowns.

The solution can be found using the RoLDSIS technique. The 128 coefficients of each one of the vectors  $\mathbf{b}_\Phi$  and  $\mathbf{b}_\Psi$  can be represented in the form of a scalogram, which is a time-frequency representation, similar to the one used in (Bertrand et al., 1994). This is depicted in Fig. 105. In the scalograms, the magnitude of each coefficient is encoded by the color saturation, such that the paler the color, the closer the coefficient is to zero. The sign of the coefficient is encoded by the color, red and blue meaning negative and positive values, respectively. The  $\mathbf{b}$  vectors can then be transformed into the time domain using the inverse DWT. The associated time profiles for  $\mathbf{b}_\Phi$  and  $\mathbf{b}_\Psi$  are shown on the top of the respective scalograms in Figure 105.

### 7.2.1 Projections onto physical and psychophysical directions

The vectors  $\mathbf{b}_\Phi$  and  $\mathbf{b}_\Psi$  obtained by the RoLDSIS procedure can be interpreted as specific directions in the space of features. These directions would then represent a sort of “canonical” representation of the neuronal activity that is associated with variation in the stimulus attribute, either physical or psychophysical. The varying response along these directions can be represented in the time domain as in Fig. 106. Each curve in Figure is obtained by projecting the original point  $\mathbf{x}_i$  onto the respective direction and by using the inverse DWT to obtain the associated time profile.

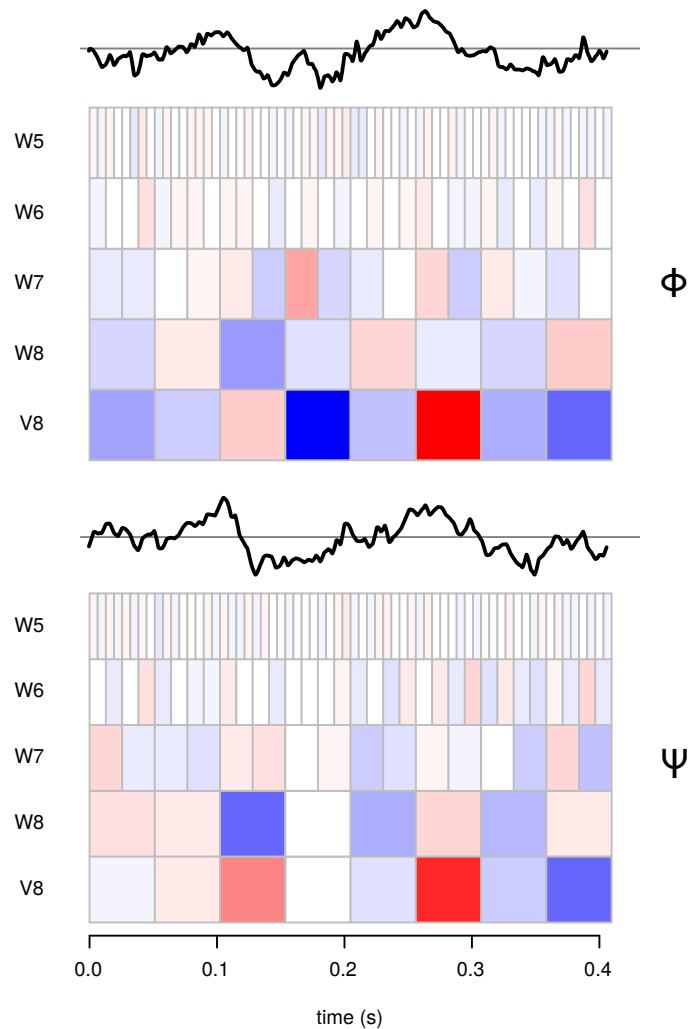


Figure 105 – Direction obtained for the RoLDSIS procedure for a representative participant (the same as in Fig. 104). The results of the RoLDSIS for the physical and psychophysical observations are shown in the top and bottom panels, respectively. In each panel, the time-domain representation of the optimal direction vector, obtained by applying the inverse DWT on the RoLDSIS result is shown by the black line, which is at top of the scalogram (time/frequency representation) of this direction vector. The amplitudes of the DWT coefficients are represented in a color scale, negative values in blue and positive values in red. The more saturated the color in a cell, the higher is the magnitude of the DWT coefficient associated with that cell. Frequency bands of the DWT are shown in increasing order from bottom to top (V8: 0–9.76 Hz, W8: 9.76–19.5 Hz, W7: 19.5–39.1 Hz, W6: 39.1–78.1 Hz, W5: 78.1–156 Hz).

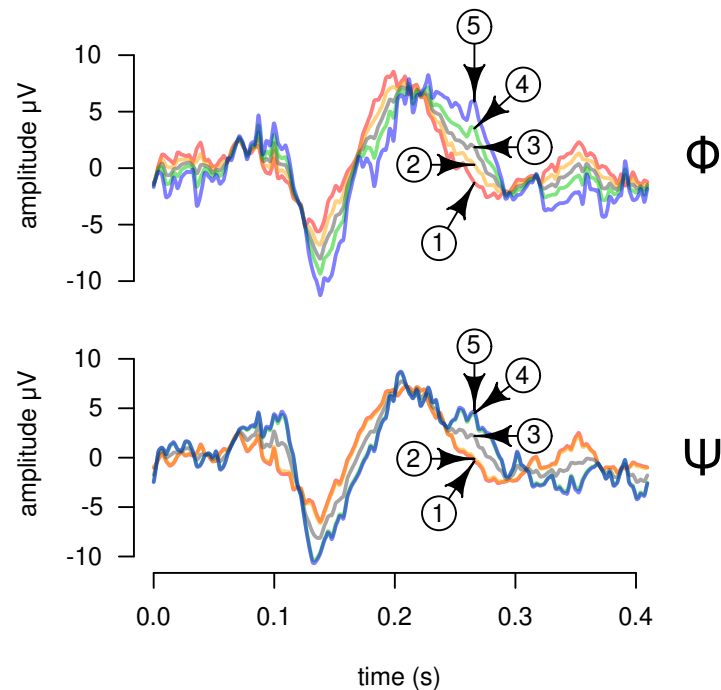


Figure 106 – Projections of ERPs for stimuli stim1, stim2, stim3, stim4, and stim5 onto the axis found by the RoLDSIS procedure, for a representative participant (the same as in Fig. 104). The responses projected onto the physical and psychophysical axes are shown in the top and bottom panels, respectively. Each projection, represented in the time domain, is drawn with a different color and indicated by the corresponding stimulus number. Note that, for the psychophysical case, the signals for stimuli stim1 and stim2, and for stimuli stim4 and stim5 are almost identical.

As can be observed in the figure, the signals resulting from the projections on a given neurophysiological axis reflect the values of the attribute associated with that axis. For instance, for the  $\mathbf{b}_\Psi$  axis, projections of stimuli stim1 and stim2 are almost indistinguishable. This also happens with stimuli stim4 and stim5. This mimics the values of the  $\psi$  attribute which are 0.0, 0.05, 0.5, 0.95, and 1.0. An equivalent result can be observed for projections on the  $\mathbf{b}_\Phi$  axis, where stimuli stim2 and stim4 are closer to stimulus stim3 than to stim1 and stim5, respectively. This mimics the values of the VOT of those stimuli (see the abscissa of the plot in Fig. 104) which are the values of the  $\phi$  attribute.

Another interesting observation concerning the projections on the neurophysiological axes is that the separation between the projections stimuli stim1 and stim5 varies with time. This variation in time is typically different between the  $\mathbf{b}_\Phi$  and  $\mathbf{b}_\Psi$  axes. For instance, in the example shown in Fig. 106, the projections of the five stimuli collapse to the same value around  $t = 180$  ms for  $\mathbf{b}_\Phi$ , while stimuli stim1 and stim5 are well apart at that instant for  $\mathbf{b}_\Psi$ . A more precise analysis of the differences between the  $\mathbf{b}_\Phi$  and  $\mathbf{b}_\Psi$  axes can be found by inspecting the scalograms representing them (Figure 105). Indeed, we can see that the effects described above are due to the wavelet coefficients in bands V8 and W7 around  $t = 180$  ms. These wavelet coefficients have stronger

loadings for the  $\mathbf{b}_\Phi$  axis, in comparison with the  $\mathbf{b}_\Psi$  axis.

These differences in the loadings for the  $\mathbf{b}_\Phi$  and  $\mathbf{b}_\Psi$  axes indicate different neurophysiological representations for the stimuli attributes. In our data, we observed that the RoLDSIS loadings are participant-specific, which indicates idiosyncratic ways of VOT processing and phonemic categorization. However, at the population level, the loadings are concentrated at specific regions of the time-frequency domain (see Figure 110). Our results are compatible with evidence reported elsewhere (Bidelman et al., 2013, Bouton et al., 2018, Alho et al., 2014, Chang et al., 2010), in terms of neurophysiological correlates of phonemic categorization. For instance, Bouton and colleagues (Bouton et al., 2018) observed that the tracking of a specific acoustic cue happens in the time interval 95–120 ms and again around 175 ms. Chang and colleagues (Chang et al., 2010) showed that maximal consonant categorization happens in the STG around 110 ms. Also, previous studies show the importance of theta oscillations (our V8 DWT band), beta oscillations (W8 and W7 bands) and low-gamma oscillations (W6 band) in speech processing (Giraud and Poeppel, 2012, Bidelman, 2015), which is also shown in our results. In sum, these findings corroborates the usefulness of RoLDSIS for the identification of neurophysiological correlates of speech perception.

## 7.3 Assessment of the RoLDSIS technique

### 7.3.1 Relationship between $\Phi$ and $\Psi$ divergence and the degree of categorization

In order to assess the relevance of the results obtained by the RoLDSIS technique, we computed the angle between the obtained physical and psychophysical directions. The value of this angle is specific for each participant and represent the separation between the neuronal representations for the two attributes. The minimal value for this angle is  $0^\circ$ , which corresponds to indistinguishable physical and psychophysical representations. When the two directions are orthogonal, the angle attains its maximal value of  $90^\circ$ . We investigated the relationship between this angle and the degree of categorization, which corresponds to the maximal slope of the psychometric curve fitted to the participant's responses in the identification task (see Figure 104). The psychometric curve is described by the sigmoid function  $p(t) = 100/[1 + e^{\beta(t-t_0)}]$ , where  $t$  is the VOT,  $p(t)$  is the probability of choosing /ta/ for VOT  $t$ , and  $t_0$  corresponds to the value of  $t$  at the point of inflection of the curve. The maximal slope of the psychometric curve happens at  $t = t_0$  and is equal to  $100\beta/4$  (in %/ms units). A large value of  $\beta$  indicates a stronger categorical perception by the participant (Bidelman and Walker, 2017).

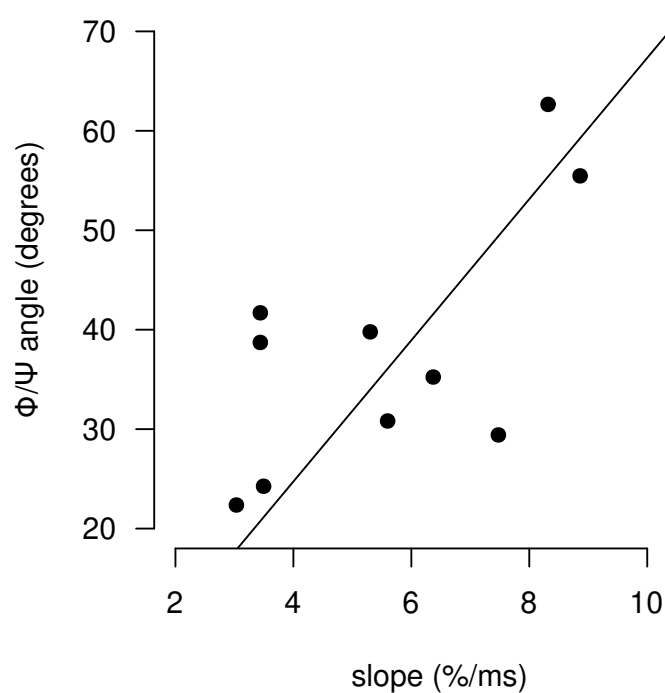


Figure 107 – Population scatter plot of the slope of psychometric curve and the angle between the psychophysical directions and the physical directions. Each point represents a participant. The horizontal and vertical axes represent, respectively, the slope of the fitted psychometric curve at 50% and the angle between the physical and the psychophysical directions obtained by the RoLDSIS procedure. The black line corresponds to the correlation line.

The results for the 11 participants are shown in Figure 107. The angle is significantly correlated with the slope in the population (Pearson's  $r = 0.67$ ,  $t[9] = 2.68$ ,  $p < 0.05$ ). Then, in this analysis we showed that there is a relationship between the categorical ability of the participant and the way how they represent the physical and psychophysical attributes neurophysiologically which is captured by RoLDSIS. The results of this analysis for all experimental conditions, using the sample frequency of 3584 Hz is shown and discussed in section 8.

### 7.3.2 Bootstrap analysis

The separation between the  $\Phi$  and  $\Psi$  axes, expressed by the angle between these two directions (see Fig. 107) could be simply the result of a statistical fluke. In order to assess this issue, we ran a bootstrap procedure. Each one of the obtained 128-dimensional DWT vectors is a trial associated with a stimulus. For each participant and each stimulus, we have a matrix of those vectors as we had multiple repetitions of the same stimulus in the experiment. Lets  $\mathbf{X}_i$  be this matrix of vectors associated with a given stimulus  $\text{stim}_i$ , where  $i=1, \dots, 5$ .

The bootstrap procedure was executed 100 times for each participant and outcome (physical



or psychophysical) in order to obtain the distribution of regression vectors  $\mathbf{b}$ . In each run  $k$ ,  $k = 1, \dots, 100$ , for each stimulus  $i$ ,  $i=1, \dots, 5$ , a matrix  $\mathbf{X}\mathbf{b}_i^k$  was obtained by sampling, with replacement, trials from the matrix  $\mathbf{X}_i$ . The amount of trials in  $\mathbf{X}\mathbf{b}_i^k$  is equal of that of  $\mathbf{X}_i$ . After that, each matrix  $\mathbf{X}\mathbf{b}_i^k$  was averaged resulting in five vectors  $\mathbf{x}\mathbf{b}_i^k$ , each one associated with a stimulus  $\text{stim}_i$ . RoLDSIS was then applied using those five vectors and the outcome  $y$  evaluated, resulting in a regression vector  $\mathbf{b}\mathbf{b}_k$ .

Each regression vector obtained through RoLDSIS represents a direction in the 128-dimensional space. The analysis of directional data depends on different methodologies than those usually applied for Cartesian data even for the computation of a basic statistic as the mean. In our case, the axes are unit vectors, lying on an 127-dimensional hypersphere, and can thus be represented by 127 spherical coordinates (the analogous of azimuth and elevation in a 3D sphere) (Miller, 1964).

PCA is a dimension reduction technique that works by projecting the high-dimensional data in a direction so that the variance of the data is maximized in the linear subspace spanned by a small number of latent components (Hu and Zhang, 2019). Using an orthogonal transformation, PCA works by converting a data set with correlated dimensions (features) into a set of linearly uncorrelated (orthogonal) dimensions called principal components (principal component (PC)). A PC that explains only a small amount of the variance of the original data set can be omitted as it represents the loss of a small amount of information. It is possible to construct as many PCs as the amount of dimensions in the data. The PCs are generated so that the amount of variance explained by them decreases as we go from the first to the last PC. Thus, usually the first and second PCs alone can explain a larger percentage of the variance.

A linear discriminant analysis (linear discriminant analysis (LDA)) was performed over the projections of the spherical representation of the regression vectors over the two first principal components (PC1 and PC2), for a representative participant data. LDA works by finding the linear transformation that maximizes the ratio between the inter-class variance and the intra-class variance. The resulting LDA separatrix defines the linear decision boundary that optimally separates the  $\Phi$  and the  $\Psi$  points. The reliability of the RoLDSIS procedure is assessed by the amount of LDA misclassifications, which has a median value of 7 across participants (minimum value 0, maximum value 63). Fig. 108 shows the results of this PCA–LDA procedure for a representative participant.

The Fig. 108 illustrates the result for the LDA analysis of the bootstrapped data. Observe that together, PC1 and PC2 explain 35% of the variance in the data. It was shown that both  $\Phi$  and  $\Psi$  neurophysiological axes are different. The LDA successfully discriminated between both classes and presented only 4 misclassifications in this example for the representative participant. This

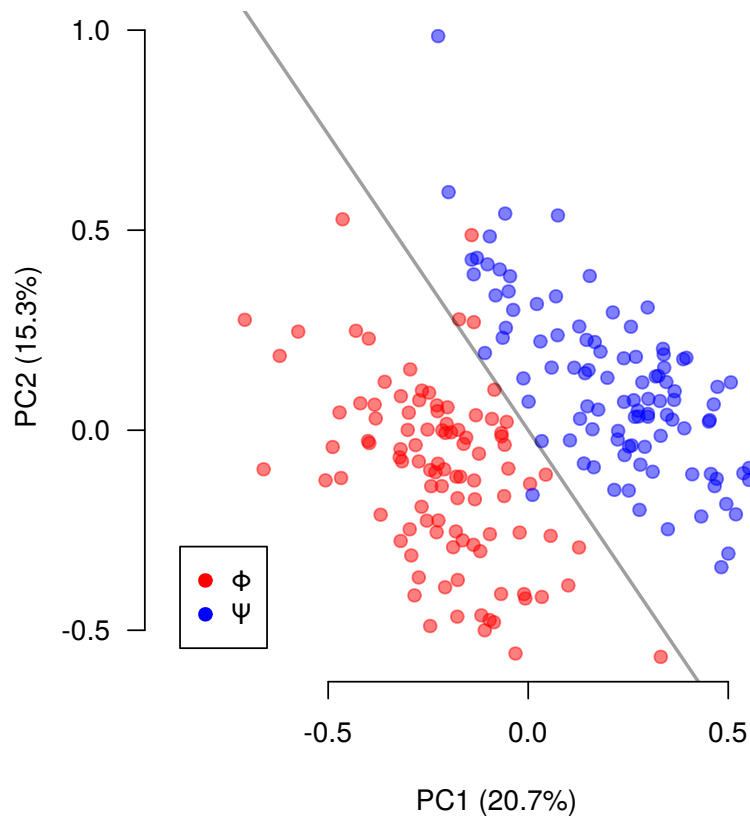


Figure 108 – Bootstrap results of the RoLDSIS procedure. The samples obtained by the bootstrap procedure on the RoLDSIS for both physical ( $\Phi$ ) and psychophysical ( $\Psi$ ) are shown in red and blue points, respectively, for a representative participant (the same as in Figure 104). The horizontal and the vertical axes represent the first and second components of the PCA applied to RoLDSIS direction axis transformed into spherical coordinates. The percentage of variance explained by this two PCs are indicated in the axes labels. The gray line corresponds to the LDA separatrix.

result shows that **RoLDSIS** is capable of finding separable  $\Phi$  and  $\Psi$  neurophysiological axes validating the results observed in the projections and scalograms discussed before. However, as expected, this result varies from participant to participant. (It is important to note that this misclassification value may vary because the bootstrap sampling is performed at random so it is different at each run resulting in different sets of regression vectors.)

### 7.3.3 Comparison with regularized linear regression procedures

A legitimate question that may be asked at this point is how the RoLDSIS procedure compares with other regularized regression techniques. To make this comparison, we considered three popular regression techniques, namely LASSO, Ridge Regression and SPLS (Friedman et al., 2010a, Chun and Keleş, 2010) and performed  $k$ -folds cross-validation (CV) procedures, for values of  $k$  varying from 3 to 6. For a CV with  $k$  folds, we generated  $5 \times k$  points by randomly

partitioning the set of **DWT** vectors for each of the five stimuli into  $k$  sets with similar amount of trials. The **DWT** vectors are then averaged inside each set. Each fold contained five **DWT** vectors corresponding to the five stimuli. At each pass of the CV procedure, one of the folds is put apart as the test set, while the regression model is fitted to the remaining folds (the training set). In our specific implementation of the CV procedure, the  $k - 1$  points, in the training set, corresponding to each stimulus were averaged, resulting in a set of five points, to which the model was fitted. Ridge regression and **LASSO** were implemented using the library “glmnet” from the software R which is based on the work of [Friedman et al. \(2010b\)](#), [Simon et al. \(2011\)](#). **SPLS** was implemented using the package “spls” ([Chung et al., 2019](#)).

The goal of the CV procedure is to select optimal values for the regularization parameters ( $\lambda$  for **LASSO** and Ridge Regression,  $\zeta$  and  $K$  for **SPLS**). Given a set of parameters, the model is fitted to training set and the global CV error is computed as the sum of the mean prediction squared errors (MSE) for the  $k$  test sets. The mean-squared error (MSE) is computed comparing the predicted value of each stimulus with the expected value at the response vector and averaging this error, considering the 5 stimuli. This is described by the equation 7.27:

$$MSE = \frac{1}{5} \sum_{i=1}^5 (y_i - f(x_i))^2 \quad (7.27)$$

Where  $f(x_i)$  is the prediction for stimulus “ $i$ ” performed by the regression model and  $y_i$  is the expected response described at by the response vector “ $Y$ ”.

Using an optimization procedure, we found the optimal values of the regularization parameters that yield the minimum value of the CV error. Note that RoLDSIS has no regularization parameter, so that the optimization procedure described above does not apply to it. This procedure was applied to each one of the eleven participants. Figure 109 shows the population mean CV errors for each number of folds, as well as the 95% confidence intervals of the mean estimations. This analysis was performed for both physical and psychophysical responses.

In order to assess how differently the regression techniques perform on our data, we fitted a linear mixed model to the results, considering the number of folds as a continuous fixed factor, the regression technique as a fixed discrete factor, and the participant as a random factor. The mean squared error (**MSE**) values, which usually follow a  $\chi^2$  distribution, were transformed to normal ([Hawkins and Wixley, 1986](#)) and the resulting values were used as the dependent variable of the linear model. The results show a significant increase in **MSE** with the number of folds ( $F[1, 158] = 50.4, p < 0.001$  for  $\Phi$  and  $F[1, 158] = 32.2, p < 0.001$  for  $\Psi$ ). For the  $\Phi$  case, there was a significant effect for the method factor ( $F[3, 158] = 5.22, p < 0.01$ ), and

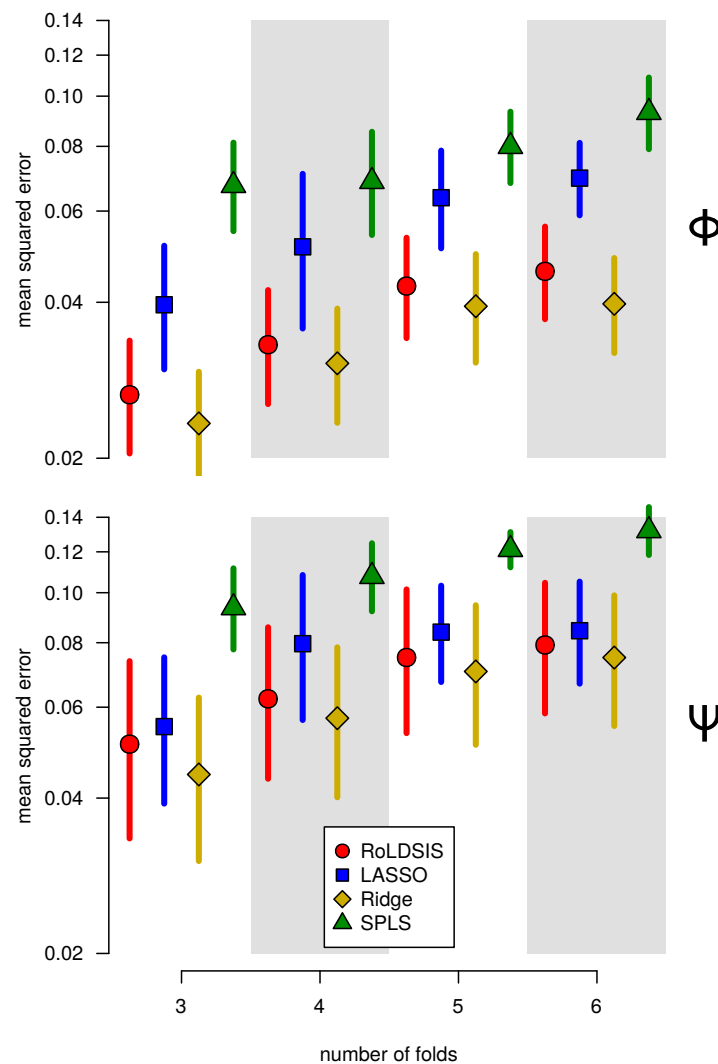


Figure 109 – Cross-validation errors for the proposed regression method (RoLDSIS) and the methods of regularized linear regression for physical (left panel) and psychophysical (right panel) attributes. Results for 3, 4, 5 and 6 folds are shown. The mean squared errors for the test set of the CV are shown with dots. Confidence intervals at 95% are represented by vertical bars.

multiple comparisons showed significant differences among all pairs of methods, besides the pair RoLDSIS and Ridge Regression. For the  $\Phi$  case, the method factor has a marginal effect ( $F[3, 158] = 2.47, p < 0.064$ ). In this later case, no significant differences were found among RoLDSIS, LASSO, and Ridge Regression, but SPLS was significantly different from the others.

As we illustrated in Fig. 105, the result of RoLDSIS can be useful for revealing the locations, in the time-frequency domain, associated with the stimulus attributes (physical and psychophysical in the present paper). Since the regression is obtained on an individual basis, the patterns of time-frequency distribution associated to the neurophysiological axis may differ from one participant to another. Therefore, it would be interesting to know whether there are global time-frequency patterns that would arise in the population.

This investigation will involve **RoLDSIS**, as well as the other three regression techniques considered in the previous section, and consists in the computation of the population-wide histogram of the neurophysiological axis in the time-frequency domain. For doing it, we first compute the squared value of the 128-dimensional axis **b** obtained by the regression technique. The squared value of a given wavelet coefficient can be assimilated to the importance (or the “energy”) of the neurophysiological axis at the associated time-frequency slot. The resulting values are then accumulated for all participants, separately for the physical and the psychophysical axes, and the square root was computed for each wavelet component. For the **RoLDSIS** technique, we used the grand average of the ERPs for each stimulus for doing the regression. For the other techniques, we used the regression result of the 3-fold consonant-vowel (**CV**) (see previous section).

The results are shown in Fig. 110 in the form of time-frequency scalograms. The darker a **DWT** component appears in a scalogram, the more important it will contribute to the associated neurophysiological direction across the population.

The darker a **DWT** component appears in the scalogram, the more frequently it appears in the neurophysiological direction found by the regression across the population. We can notice that the **RoLDSIS** procedure yields results similar to the Ridge Regression technique. The main difference is that Ridge’s solutions are more dispersed in the scalogram, whereas they appear more concentrated in the regions of the time-frequency space that should be related to the phonemic categorization process. The scalograms for **LASSO** looks like a chopped version of the scalogram for **RoLDSIS**. At least, **LASSO** seems to be capturing the important aspects of the neurophysiological axis, but since it also does feature selection, besides regularizing the regression, fewer **DWT** components appear in the scalogram. On the other hand, the **SPLS** technique produces a very dispersed scalogram, even though it captures the relevant components between 0.1 and 0.3 s in the W8 and V8 bands.

It must be emphasized that the results shown in Fig 110 for the **RoLDSIS** technique are consistent with the literature on neurophysiological correlates of speech categorization and processing of theta oscillations (our V8 **DWT** band), beta oscillations (W8 and W7 bands) and gamma oscillations (W6 and W5 bands) (Bouton et al., 2018, Giraud and Poeppel, 2012, Bidelman, 2015).

The prediction error obtained by **RoLDSIS**, in our data set, is comparable to those obtained with the Ridge Regression technique and performed better than **LASSO** and **SPLS**, showing that **RoLDSIS** is a suitable alternative for the processing of neurophysiological signals. **RoLDSIS** avoids the need for cross-validation, which implies the extraction of a large amount of observations from the data and, consequently, a decreased signal-to-noise ratio when averaging trials. For this reason, this technique may be appropriate for coping with extreme **HDLSS** problems as

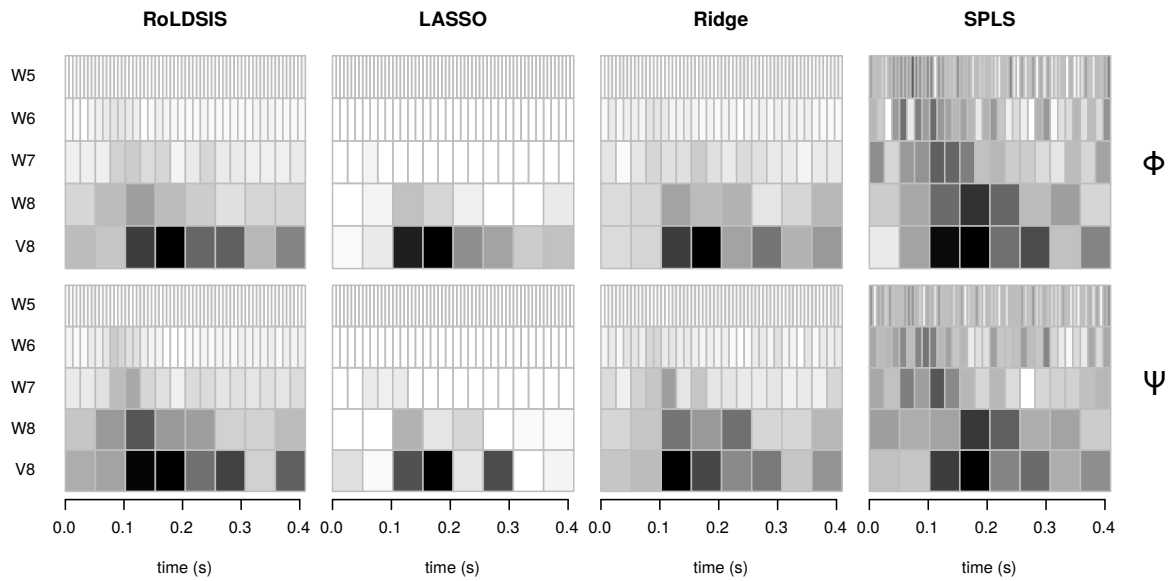


Figure 110 – Scalograms of the regression results. Scalograms for the root mean squared regression coefficients for each component of the DWT, across the population, are shown for the proposed regression method (RoLDSIS), for the methods of regularized linear regression and for physical (left panel) and psychophysical (right panel) attributes. Shades of gray represent the cumulative RMS (white for zero and black for the maximum value). Frequency bands of the DWT are the same as those in Fig. 105.

the one we deal in this work.

## Chapter 8

# TIME-FREQUENCY DOMAIN ANALYSIS

In Chapter 6, we presented the results of a mixed-effects model applied to the variables measured directly from the ERP waveform, namely the latencies and amplitudes of the N1 and P2 peaks. As we have shown, a substantial amount of information, regarding the underlying neurophysiological mechanisms of phonemic processing, can be obtained from these variables alone. However, the question is still open whether other properties of the ERP can convey further information. In particular, we are interested in the way the evoked responses are modulated in the time-frequency domain and how this modulation varies with the electrode position, the type of the task (active or passive), and the type of continuum used in the identification task (VOT or Formants).

In the present chapter, we show the results of mixed-effects models applied to the time-frequency domain representation of the average evoked responses. These time-frequency domain representations were obtained with the DWT, as described in Chapter 5. The responses for the five stimuli in each continuum, represented by the DWT coefficients, were then fed to the RoLDSIS procedure, as described in Chapter 7. The physical and psychophysical directions in the DWT coefficients space, obtained by RoLDSIS, were then used to compute the dependent variables in the mixed-effects model analysis. Since the RoLDSIS directions are represented by a relatively high-dimension vector, we need to define regions of interest (ROIs) that will be effectively used as dependent variables in the analysis.

In terms of time intervals, these ROIs were chosen to broadly cover “early” (close to N1, from 0 to 135 ms) and “late” (close to P2, from 135 to 275 ms) parts of the ERP, which can be associated with early and late neuronal processes associated with phonemic processing and

Table 2 – Discrete wavelet transform levels and frequencies used in time-frequency domain analysis.

Band	W1	W2	W3	W4	W5
Frequency	1792 – 896	896 – 448	448 – 224	224 – 112	112 – 56
Band	W6	W7	W8	W9	V9
Frequency	56 – 28	28 – 14	14 – 7	7 – 3.5	3.5 – 0

categorization. In terms of frequency intervals, the ROIs were chosen to cover roughly each of the classic electrophysiological bands, namely the theta, alpha, beta and gamma zones. In particular, we resampled our signals at 3584 Hz, such that the DWT levels will coincide with those frequency bands (see explanation below).

In sum, the fixed factors of the mixed-effects models were the type of the experiment (active or passive), the feature manipulated in the phonemic continuum (VOT or Formants) and the electrode (F7, F8, Fz, TP9, or TP10). The participants were considered as a random factor. This chapter presents the details of the data processing and the statistical analyses, as well as a discussion of the obtained results in the context of the literature in the domain.

## 8.1 Data processing

### 8.1.1 Resampling of the ERP signals

For the time-frequency analysis, the ERPs were resampled at 3584 Hz. This resampling was necessary so that the wavelet decomposition resulted in lower level bands with frequency range similar to those of traditional electrophysiological bands. The DWT was performed with ten frequency levels of decomposition, nine of detail coefficients (with the highest frequency one being the W1) and one of approximation coefficients (lowest frequency level being at V9). This amount of levels was enough to capture the relevant frequency bands for the analysis of brain potentials such as delta (0.5 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz), beta (14 – 30 Hz) and gamma (30 – 100 Hz) waves (Jensen et al., 2019). It is important to note that the boundaries of the electrophysiological bands change slightly according to different authors, so that the difference of 1 to 2 Hz from these bands to those obtained through our resample should not be a problem. Considering the 3.584kHz sampling rate, the frequencies of each band are shown in Table 2.

Each epoch was resampled to a sample frequency of 3584 Hz and, after that, the baseline correction was performed. The samples of the baseline part of the signal were maintained. This



same amount of samples was used in the **DWT** of each epoch, performed now with 10 levels (nine of detail and one of approximation coefficients) which resulted in 7 coefficients in the lower band (V9). The first and last coefficients were removed to avoid the influence of threshold effects in the analysis. This removal was performed in all other bands for the coefficients corresponding to the same time-frame of the ones removed from the V9 band. The threshold effect occurs when the shift parameter of the **DWT** is such that during the convolution process, the wavelet of a given band falls out of the interval defined by the signal size (see Appendix D for details). Periodic or symmetric extensions can be used to minimize this problem. In our case, we used the periodic extension, which can cause some distortion in the coefficients at the beginning and the end of each band. As we used the entire epoch (with one second of duration) for the **DWT**, including the prestimulus part used for baseline correction, the removed coefficients corresponded to around 286 ms of signal being 143 ms at the beginning and 143 ms at the end (each corresponding to the time-frame of one coefficient of the V9 level). Since the baseline part of all epochs corresponded to 150 ms, the resulting scalogram time scale goes from -7 ms to 707 ms, which covers easily the time frame necessary for our analysis. After the removal, the original 3584 coefficients were reduced to 2560.

The organization of the DWT matrices was the same of that of the time domain processing. The difference is that for the example of the participant 9 described in Chapter 7, the matrix now has dimensions  $185 \times 2560$ . The limitation to the W5 level (between 0 and 112 Hz) resulted in 160 coefficients so, for this example, the final matrix has dimensions  $185 \times 160$ .

### 8.1.2 ROI selection and mixed-effects models

Similar to the description in Chapter 7, here, the **RoLDSIS** was applied using the physical and psychophysical attributes for each subject, feature (**VOT** or formants), type of experiment (passive or active) and electrodes. The DWT matrices for each case were averaged and the resulting vector used in the regression. Vectors of the five electrodes evaluated (F7, F8, Fz, TP9 and TP10) were grouped resulting in only one big vector with 800 predictors for the regression (per stimulus). In doing that, we make sure that the regression captures the differences between the regions of the brain in response to each experimental condition for the physical and psychophysical analysis.

Similarly to the analysis performed for the time domain data (N1, P2, N1-P2, T1 and T2), we fitted mixed-effects models for the direction vectors computed by **RoLDSIS**, which was applied for both physical and psychophysical responses, for all eleven participants, both **VOT** and Formants continua and, both active and passive tasks resulting in 88 direction vectors.

Each direction vector contained the regression coefficients for the five electrodes that were separated for this analysis, so that we can evaluate the effect of the factor electrode and also represent this regression result in the scalograms. Thus, now, we have 440 ( $88 \times 5$ ) different results to evaluate. As each scalogram has 160 regression coefficients corresponding to weights for each wavelet coefficient, we decided to group some coefficients in **ROIs**, representing a wavelet frequency band and a time frame that we judged to be enough to draw significant conclusions for this study. The selection of the eight **ROIs** is depicted in Figure 111. This definition of **ROIs** was necessary because we observed that, for each participant, coefficients occur at different time or frequency band around a given region in the scalogram so that analyzing each coefficient separately could mask the effects of the different factors analyzed over the participants' group. The **ROIs** were selected according to the electrophysiological bands in the frequency axis and according with the N1 and P2 waves position in the time axis. These regions extend the time interval where phonemic categorization is expected to happen [Chang et al. \(2010\)](#), [Bidelman et al. \(2013\)](#), [Bouton et al. \(2018\)](#).

The band W5 was analyzed together with band W6 to encompass the gamma band, which has been previously considered by other authors. For instance, [Bidelman \(2015\)](#) observed effects in the 60-80 Hz band. [Bouton et al. \(2018\)](#) have also considered the broad gamma band of 40-110 Hz. A maximum speech-brain coherence was observed by [Giraud and Poeppel \(2012\)](#) around 30-70 Hz. The V9 band corresponding to the delta oscillation was not included, as this band is not typically used in the studies of speech categorization. We will reference each region of interest (**ROI**) according to the frequency band and time-frame that it represents. Thus, for example, the **ROI-5** represents an early-beta band while the **ROI-2** represents a late-theta band.

For each **ROI**, the dependent variable was obtained by computing the square of the sum of squared difference between the direction vectors related to the physical and psychophysical outcomes. This measurement can also be understood as the Euclidian distance between the physical and psychophysical direction vectors. This relation is represented in the Equation 8.1. As each coefficient in a direction vector is the result of the regression, it represents the weight given to the wavelet coefficient. Our relation represent the distance between the physical and psychophysical neural representations for a given experimental condition and cortical region (acoustic cue, task and electrode). Thus, we can say that this variable represents the “discrepancy” between the physical and psychophysical representations in a given **ROI** and we will use the term *discrepancy* for referring to this dependent variable.

$$discrepancy = \sqrt{\sum(\phi - \psi)^2} \quad (8.1)$$

Another interesting measure will be the sum of the difference of the squared direction vectors

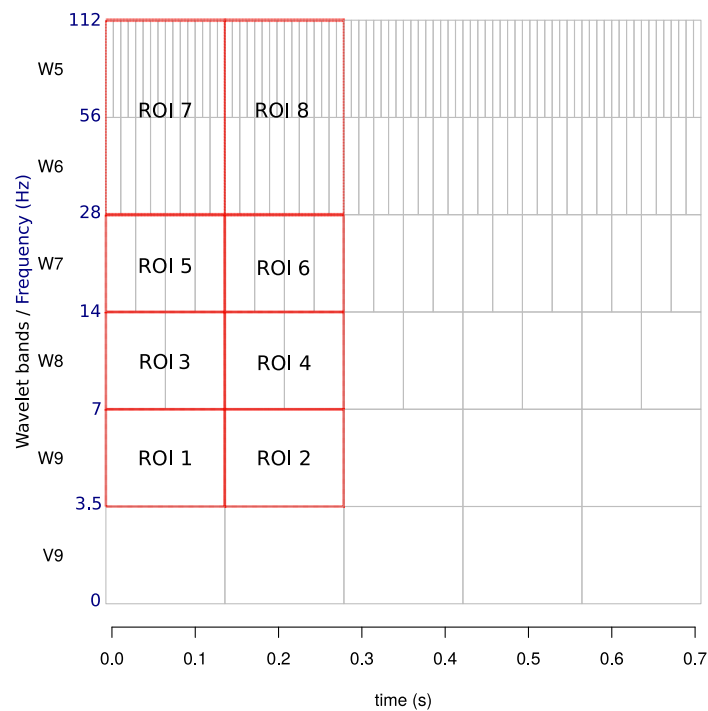


Figure 111 – Definition of the ROIs as dependent variables for the models. Each ROI corresponds to specific a frequency band and specific a time interval.

for the physical and psychophysical attributes. With this measure as dependent variable in the model, if its value is positive, it means that the level of the factor has a more physical representation at that ROI than a psychophysical one. On the other hand, if its value is negative, the level of the factor analyzed has a more psychophysical representation than a physical one. The psychophysical representation means that the response is more categorical following the psychophysical attributes whilst the a physical representation follow the physical attributes related to the stimuli (for each subject). After testing this possible measure, we did not observe significant effects for any factor or factor interaction. We believe that with data from more participants it will be possible to improve the power of the statistical test and then obtain significant effects using this dependent variable. Then, for the present work, we considered the discrepancy measure described before in the Equation 8.1.

Participants are considered to be a random effect in the model. The fixed factors analyzed were: feature (VOT, formants), type (active, passive) and electrode (F7, F8, Fz, TP9, TP10). Contrast analyses were performed for each factor and factor interactions (up to two factors) that presented significant effects. We analyzed models with the the physical and psychophysical responses separately and saw that the same effects were observed at the same ROIs. Then we decided to work with the distance between those neurophysiological axes to identify which ROI code the effects of this difference considering the feature, type and electrode factors.

The same R software libraries used in the mixed model analysis presented in Chapter 6 were

used here. For each ROI, a mixed-effects linear model was obtained and analyzed.

## 8.2 Results

### 8.2.1 Relationship between $\Phi$ and $\Psi$ divergence and the degree of categorization

As in Chapter 7, we found correlations between the categorical ability of the participant and the way they represent the physical and psychophysical attributes neurophysiologically. Previous studies, like the ones by Bidelman and colleagues, have tried to associate the degree of categorization with neurophysiological features extracted from ERP signals (Bidelman et al., 2013, Bidelman and Walker, 2017). However, to our knowledge, our study is the first one that attempts to associate stimulus attributes with the whole set of extracted features (thanks to the RoLDSIS technique) without ad-hoc definition of the neurophysiological correlates.

Figure 112 shows the scatter plot of the maximum slope of psychometric curve ( $\beta$ ) and the angle between the psychophysical directions and the physical directions (*phy* – *psy* angle) for the 11 participants for the VOT-active experimental condition. Each point represents a participant. The horizontal and vertical axes represent, respectively, the slope of the fitted psychometric curve at 50% and the angle between the physical and the psychophysical directions obtained by the RoLDSIS procedure. The black line corresponds to the correlation line. Pearson's correlation coefficient was computed using the *cor.test* function in the R software. The p-value of the correlation test indicates the probability that the correlation value found would be also observed even if the correlation of the data was zero (null hypothesis), thus, the smaller this p-value the better. The correlation coefficient of  $r = 0.629$  was significant for this case ( $p < 0.05$ ). Hence, the stronger the categorical perception, the more distinct are the physical and psychophysical internal representations.

One could argue that the observed effect is just artefactual because participants with a small  $\beta$  have stim1 closer to stim2 and stim4 closer to stim5 in the physical axis. This means that the physical and psychophysical attributes used in the regression are similar and, consequently, the direction vectors found by RoLDSIS should also be similar, resulting in a small angle. Conversely, larger values of  $\beta$  should naturally result in larger angles.

This artefactual interpretation can be ruled out by the analysis of the VOT-passive, whose results

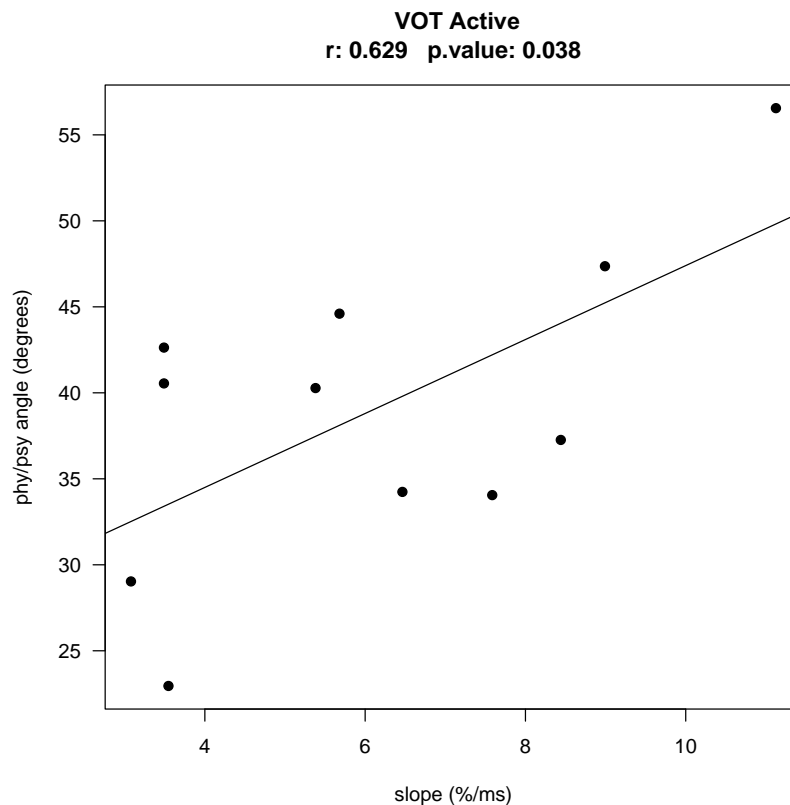


Figure 112 – Relation between the physical and psychophysical direction vector angle and the slope ( $\beta$ ) of the psychometric curve of all participants for the VOT-active experimental condition.

are shown in Figure 113. In this case, the correlation coefficient  $r = 0.299$  was not significantly different from zero ( $p > 0.37$ ). Note that, for both VOT-active and VOT-passive, the same set of stim1, . . . , stim5 was used. If the artefactual interpretation was true, then the  $\beta$  and the  $phy - psy$  angle should have been correlated in both cases.

Figures 114 and 115 show the population scatter plots for the physical and psychophysical direction vector angle against the maximum slope of the psychometric curve for both the Form-active and Form-passive conditions. The correlation coefficients were  $r = 0.788$  for the Form-active case and  $r = 0.794$  for the Form-passive being both significant ( $p < 0.01$ ). Here, differently from the VOT feature, the passive case presented a significant correlation with the slope, which suggests that the underlying processing may be different between the VOT and the Formants conditions. It can be also observed that the angles of the active task are greater than those observed for the passive task.

This result is also interesting because Bidelman and Walker (2017) used a vowel continuum (as ours) and found a correlation between time-domain features extracted from the ERP signals and  $\beta$ , even though they did not observe categorical perception for the passive listening cases. Now, there are differences between our procedures but, in general, our results showed that

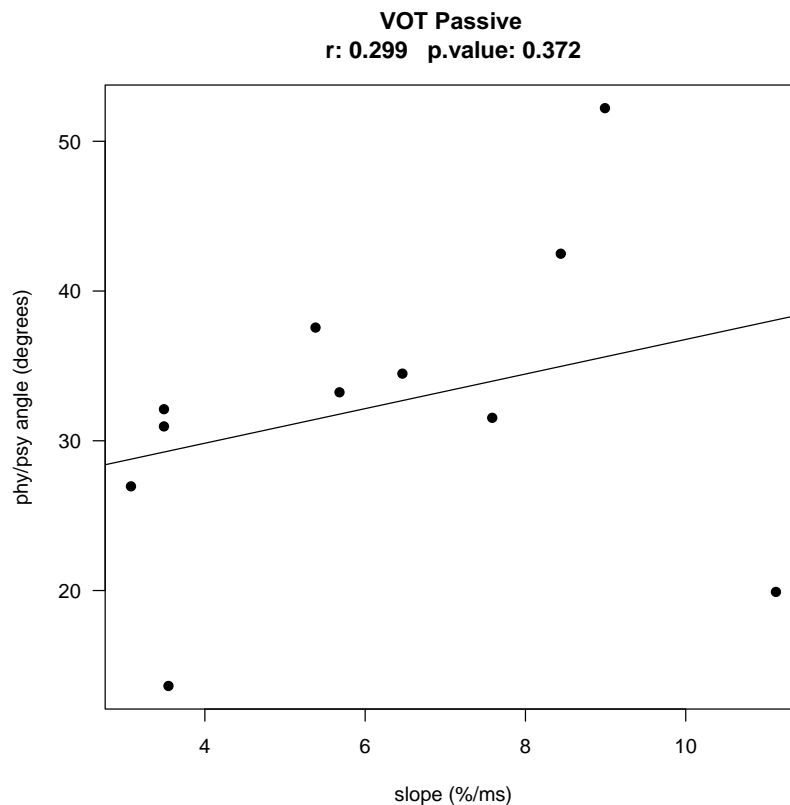


Figure 113 – Relation between the physical and psychophysical direction vector angle and the slope ( $\beta$ ) of the psychometric curve of all participants for the VOT-passive experimental condition.

there is categorical perception of formants for the passive listening task. The unnatural formant transitions between /u/ and /a/ used by [Bidelman and Walker \(2017\)](#) or the features they extracted from the [ERP](#) signals may have caused this difference in results between our study and theirs.

For the [VOT](#)-active, Form-active and Form-passive conditions, we observe a positive correlation between the angles and  $\beta$ . This suggests that participants which categorizes better, have a more distinct representation of the physical characteristics of the stimuli and its psychophysical perceptual characteristics.

It also is interesting to observe that angles for the [VOT](#)-active case are, in average, greater than those for the Form-active. This indicates that the distinction between physical and psychophysical representations of the stimuli should be greater for the [VOT](#) than for the formant acoustic cue, when the participant pays attention to the task. [Altmann and colleagues](#) observed that categorical effects are less pronounced for vowels (formants) than for consonants ([VOT](#)) ([Altmann et al., 2014a](#)). This is consistent with the fact that the observed angles were smaller for Formants than for [VOT](#).

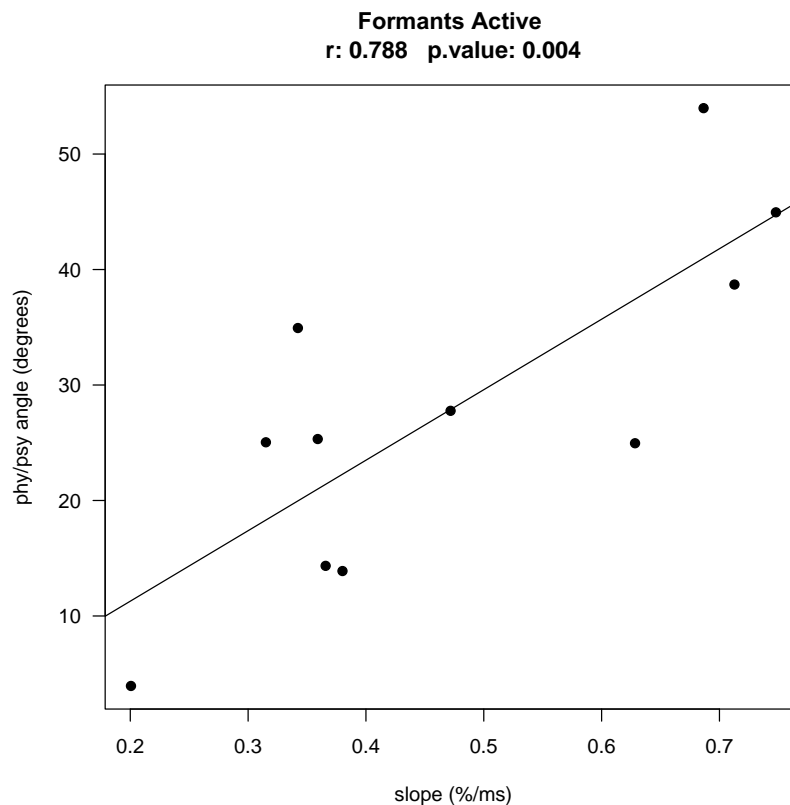


Figure 114 – Relation between the physical and psychophysical direction vector angle and the slope ( $\beta$ ) of the psychometric curve of all participants for the Form-active experimental condition.

## 8.2.2 Mean scalograms

Figures 116 to 123 present the scalograms with the sum of the absolute value of the regression coefficients (direction vector **b**) computed across the population, for each combination of experimental conditions. In each case, five scalograms are shown, that correspond to the five electrodes: F7 (top left), F8 (top right), Fz (top center), TP9 (bottom left) and TP10 (bottom right).

In general, we observe that the VOT scalograms related to the temporal electrodes responses present larger coefficients than those of the frontal electrodes. In particular, in all cases, the scalogram relative to the Fz electrode presented the smaller coefficients. This is in accordance with the amplitude of the ERPs in those regions, illustrated and commented in Section 6.2. This may be related to the physical and psychophysical aspects of the perception of the stimuli by the generators of those scalp regions. For Formants scalograms this is not clear. In fact, for the Formants-active case, it seems that the scalograms for the frontal electrodes, F7 and F8, present larger coefficients than the temporal ones. The Fz electrode still presents the smaller coefficients in active and passive cases here.

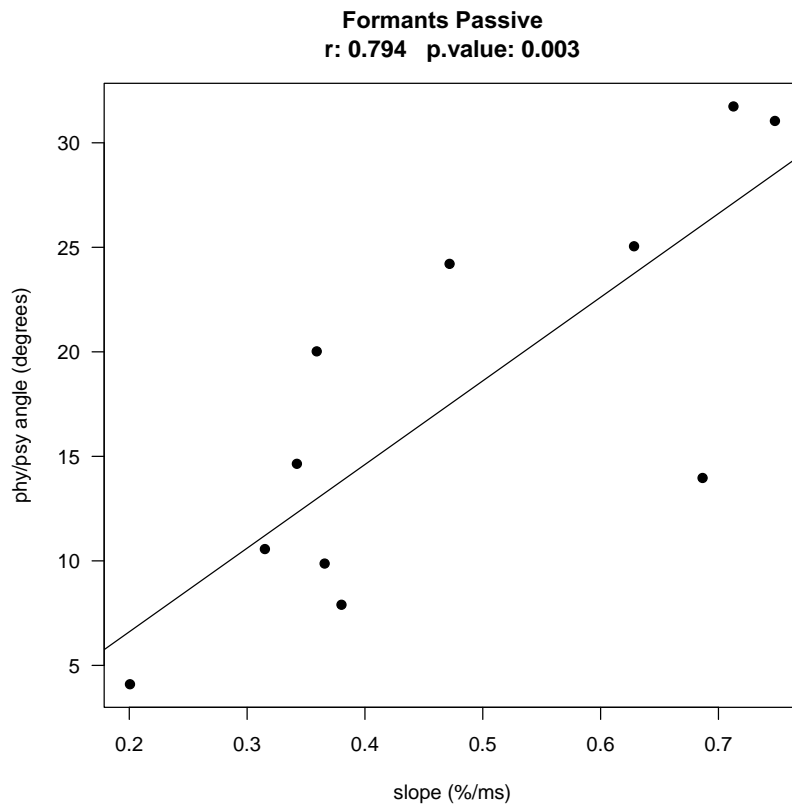


Figure 115 – Relation between the physical and psychophysical direction vector angle and the slope ( $\beta$ ) of the psychometric curve of all participants for the Form-passive experimental condition.

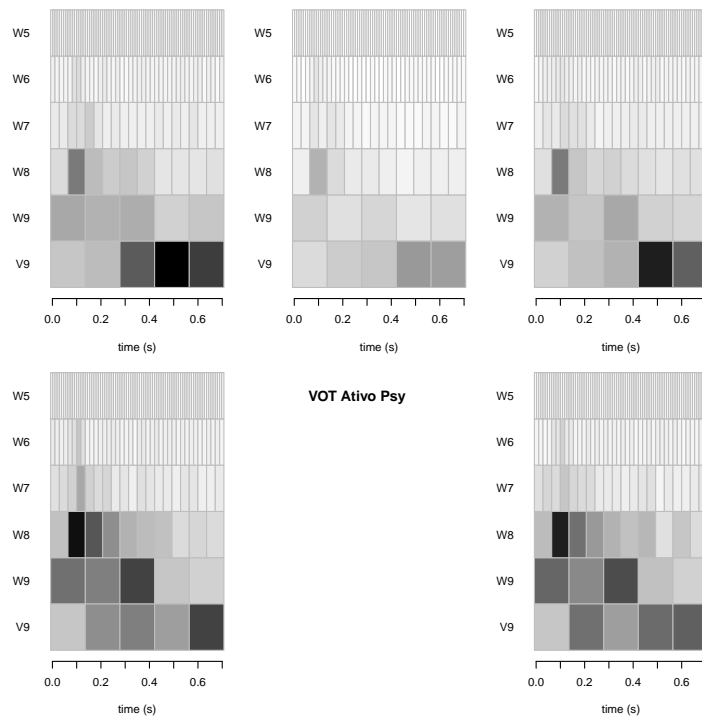


Figure 116 – Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the VOT continuum active task, psychophysical response. Scalograms are organized according with the electrodes position in the scalp.



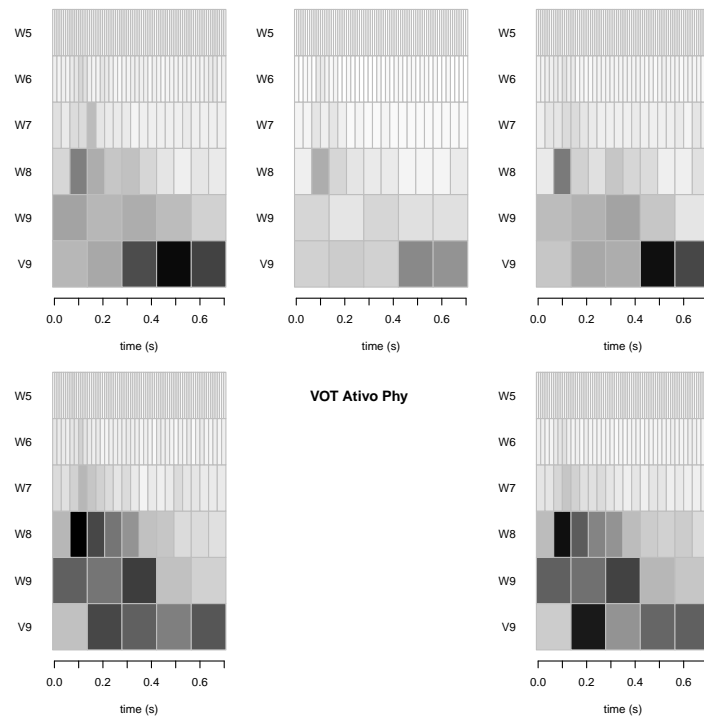


Figure 117 – Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the VOT continuum active task, physical response. Scalograms are organized according with the electrodes position in the scalp.

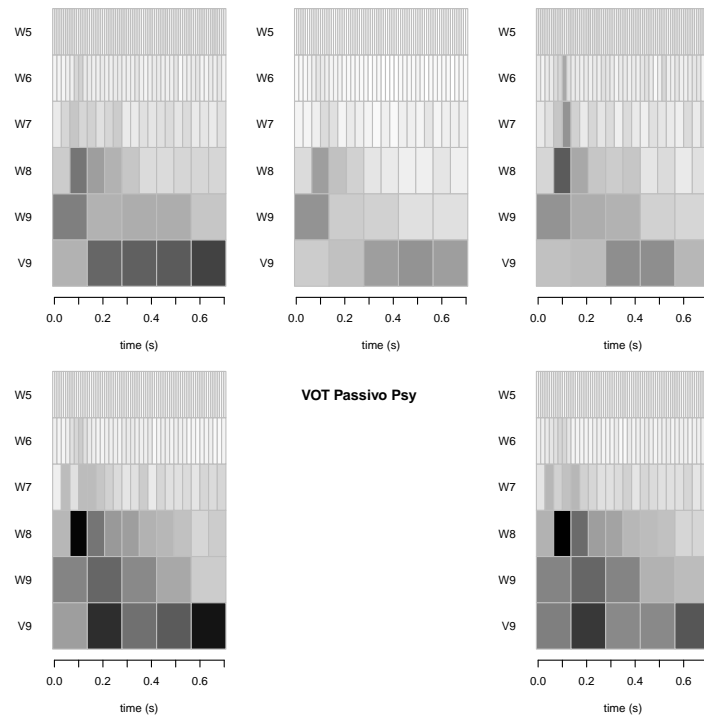


Figure 118 – Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the VOT continuum passive task, psychophysical response. Scalograms are organized according with the electrodes position in the scalp.

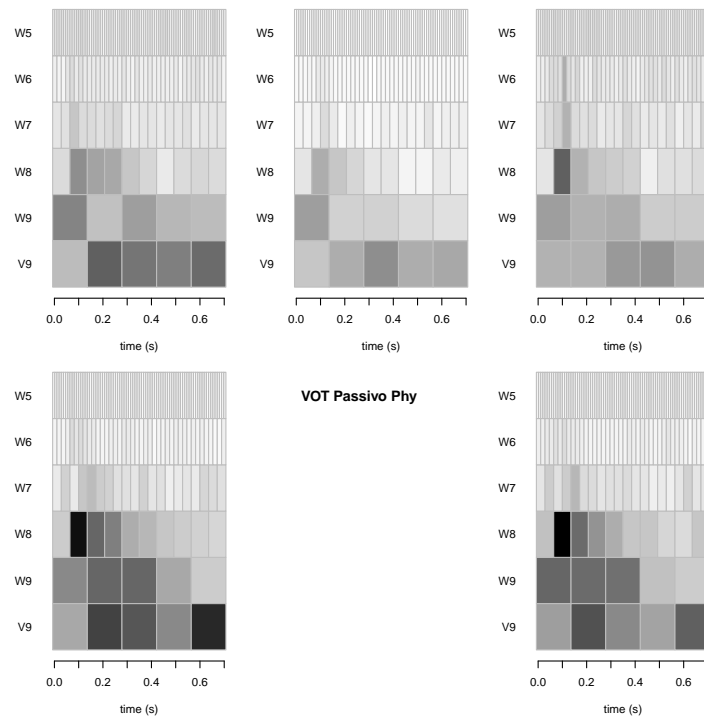


Figure 119 – Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the VOT continuum passive task, physical response. Scalograms are organized according with the electrodes position in the scalp.

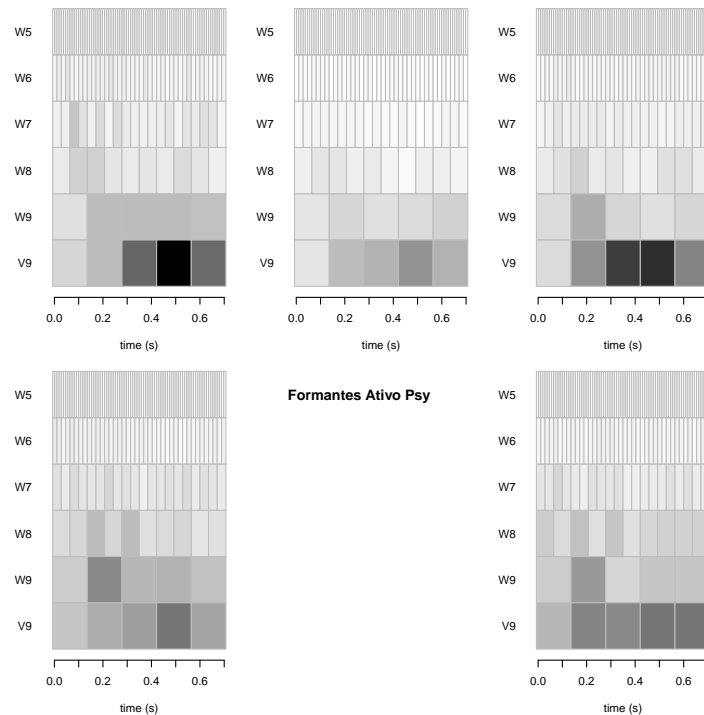


Figure 120 – Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the Formantes continuum active task, psychophysical response. Scalograms are organized according with the electrodes position in the scalp.

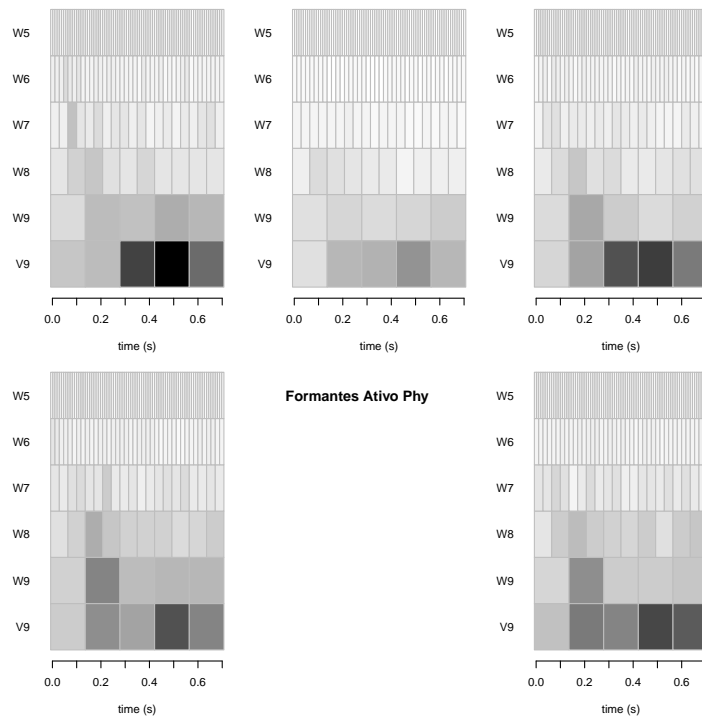


Figure 121 – Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the Formants continuum active task, physical response. Scalograms are organized according with the electrodes position in the scalp.

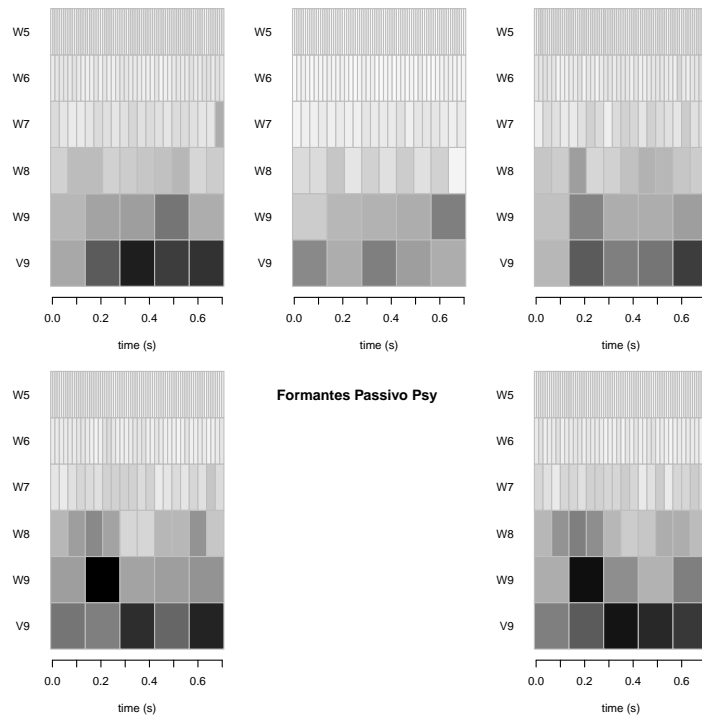


Figure 122 – Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the Formants continuum passive task, psychophysical response. Scalograms are organized according with the electrodes position in the scalp.

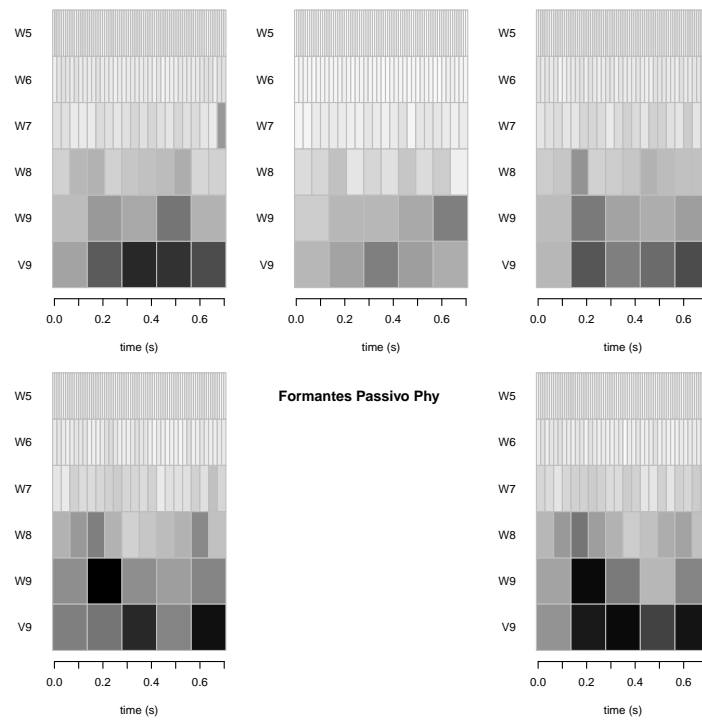


Figure 123 – Scalograms of the five electrodes direction vectors for the mean of the coefficients along the 11 participants. Results for the Formants continuum passive task, physical response. Scalograms are organized according with the electrodes position in the scalp.

From these scalograms, it is difficult to analyze the existence of significant effects of the task and acoustic cue on the physical and psychophysical differences of the neural responses at different electrodes. In order to investigate those effects, the analysis of mixed-effects models fitted to our regression results were performed.

### 8.2.3 Mixed effects models for selected ROIs

The results for the ROI-1 discrepancy are presented in the full model in Figure 124. The R formula used in the mixed-effects model was:

$$\text{roi1} \sim \text{feature} * \text{type} * \text{electrode} + (1|\text{participant}).$$

The “\*” symbol indicates that interaction factors are included in the model.

ANOVA and RANOVA table (obtained with the *anova* and *ranova*, respectively, in the R software) are shown below. The assumptions of normality, heterocedasticity and no correlation of the fixed effect predictors were analyzed at the Appendix F. Only the main factors and interaction factors

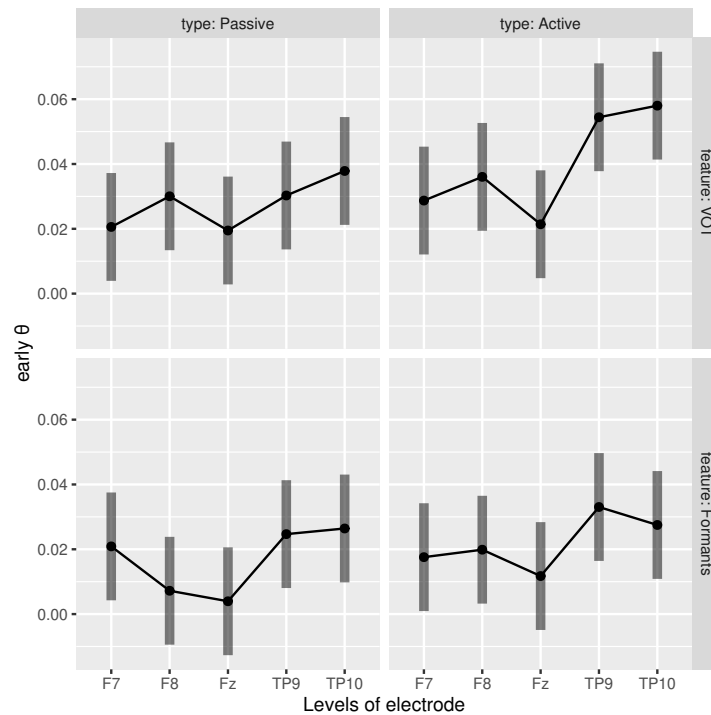


Figure 124 – Representation of the complete mixed-effects model for the discrepancy computed for the ROI-1 including the fixed factors: feature, type and electrodes.

that presented significant effects are shown <sup>1</sup>. The effect of the random factor *participant* was significant, which indicates that there exist inter-individual differences and they are important in the model, but they will not be analyzed as we are interested in the effects on the group.

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.0113965	0.0113965	1	190	17.1372	5.224e-05
type	0.0041576	0.0041576	1	190	6.2519	0.01325
electrode	0.0170077	0.0042519	4	190	6.3938	7.561e-05
feature:type	0.0006275	0.0006275	1	190	0.9436	0.33258
feature:electrode	0.0016927	0.0004232	4	190	0.6363	0.63719
type:electrode	0.0012711	0.0003178	4	190	0.4779	0.75196
feature:type:electrode	0.0016352	0.0004088	4	190	0.6147	0.65255

ANOVA-like table for random-effects: Single term deletions

Model:

<sup>1</sup> Some factors that are not significant may appear in the table because they appear in interaction factors that have significant effects.

```

r1.early.theta ~ feature + type + electrode + (1 | subject)
              npar logLik      AIC    LRT Df Pr(>Chisq)
<none>          9 458.58 -899.16
(1 | subject)   8 449.74 -883.49 17.672  1  2.625e-05

```

In Appendix E, the ANOVA and RANOVA tables for the remaining ROIs and the graphic representation of the associated complete mixed-effects models are presented. Below, each significant factor and interaction factors (involving up to two factors) are shown and discussed for all ROIs involved. Contrast analyses are performed for each significant factor or interaction factor in order to obtain an interpretation of the observed effects. 95% confidence intervals are represented by vertical bars in the figures.

### 8.2.3.1 Electrode

For the factor electrode were considered as levels of the factor the data from electrodes at the positions F7, Fz, F8, TP9 and TP10. Figure 125 illustrates the effects observed for this factor in all the ROIs.

Contrast analysis was performed using Kenward-Roger's method for the degrees of freedom (Kenward and Roger, 1997). We choose three contrasts: medial  $\times$  lateral electrodes [ $Fz - (F7 + F8 + TP9 + TP10) / 4$ ], frontal  $\times$  temporal electrodes [ $(F7 + F8) - (TP9 + TP10)$ ], and left  $\times$  right side [ $(TP9 + F7) - (TP10 + F8)$ ]. For each ROI where the factor electrode was significant, the contrast analysis was performed and the results are listed as follows:

#### ROI-1

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.015425	0.00429	203	-3.595	0.0004
frontal - temporal	-0.013914	0.00384	203	-3.625	0.0004
left - right	0.000245	0.00384	203	0.064	0.9491

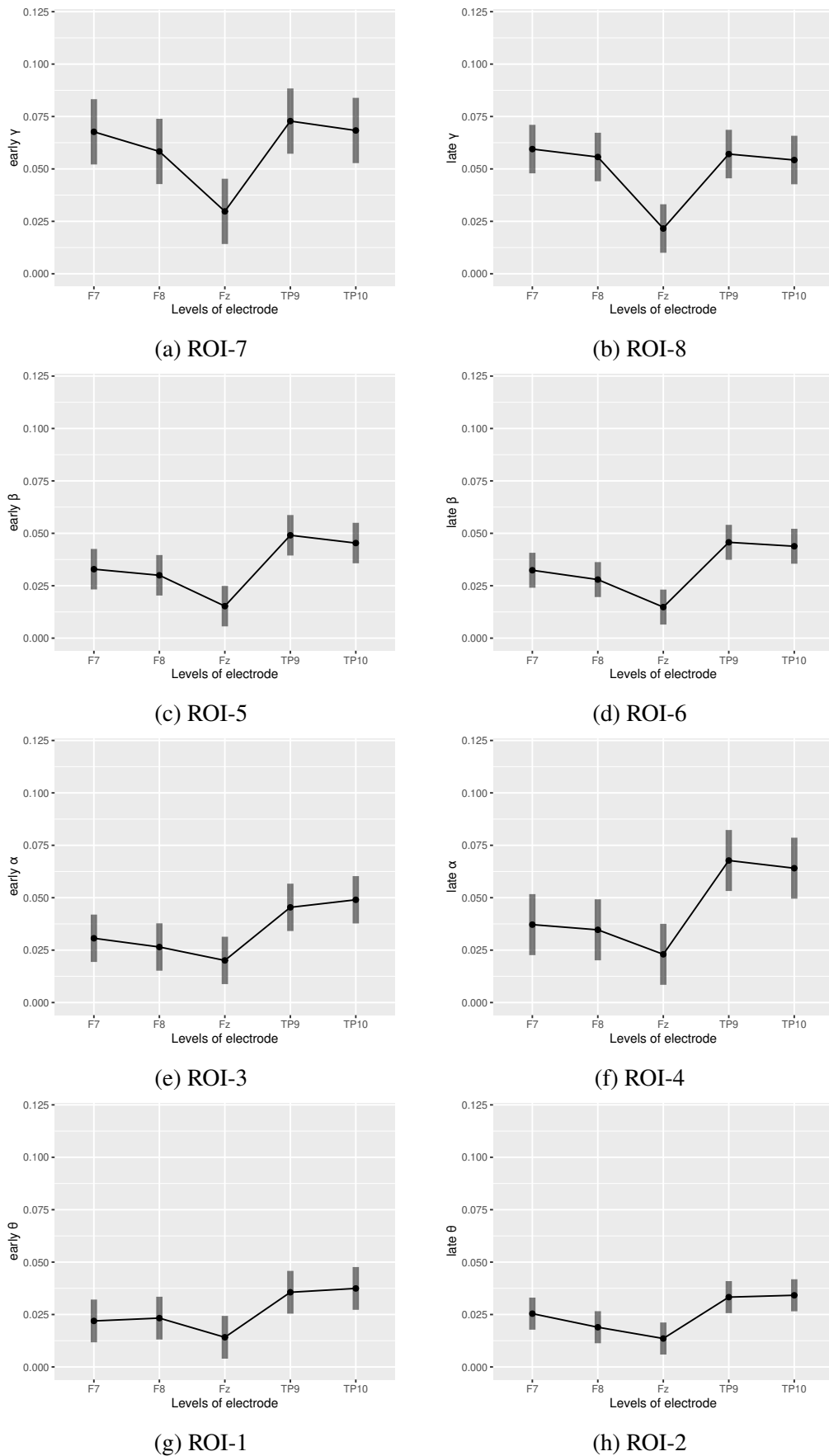


Figure 125 – Representation of the factor electrode over the discrepancy computed for all ROIs.

**ROI-2**

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.01440	0.00370	199	-3.888	0.0001
frontal - temporal	-0.01156	0.00331	199	-3.490	0.0006
left - right	0.00367	0.00331	199	1.109	0.2686

**ROI-3**

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.01780	0.00407	198	-4.371	<.0001
frontal - temporal	-0.01863	0.00364	198	-5.116	<.0001
left - right	0.00388	0.00364	198	1.067	0.2875

**ROI-4**

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.027918	0.00510	200	-5.476	<.0001
frontal - temporal	-0.029989	0.00456	200	-6.577	<.0001
left - right	-0.000596	0.00456	200	-0.131	0.8962

**ROI-5**

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.0241	0.00379	203	-6.348	<.0001
frontal - temporal	-0.0158	0.00339	203	-4.655	<.0001
left - right	-0.0004	0.00339	203	-0.118	0.9062

**ROI-6**

contrast	estimate	SE	df	t.ratio	p.value
----------	----------	----	----	---------	---------



medial - lateral	-0.02265	0.00304	199	-7.462	<.0001
frontal - temporal	-0.01461	0.00271	199	-5.383	<.0001
left - right	0.00128	0.00271	199	0.473	0.6365

**ROI-7**

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.03705	0.00577	199	-6.417	<.0001
frontal - temporal	-0.00755	0.00516	199	-1.463	0.1451
left - right	0.00244	0.00516	199	0.473	0.6369

**ROI-8**

contrast	estimate	SE	df	t.ratio	p.value
medial - lateral	-0.035046	0.00499	198	-7.029	<.0001
frontal - temporal	0.001911	0.00446	198	0.428	0.6688
left - right	0.000467	0.00446	198	0.105	0.9168

A medial-lateral contrast presented significant effect ( $p < 0.05$ ) in all ROIs with a negative estimate. This shows that the medial electrode (Fz) presents a smaller discrepancy than the lateral electrodes at temporal and frontal regions. This may be related to the location of the main structures related to speech processing which is more lateral in the cortex as the inferior frontal gyrus (IFG), premotor cortex (PM), pSTG, Heschl's gyrus, Auditory cortices, STS, etc.

By analyzing the frontal-temporal contrast, we can see that it was significant in all ROIs ( $p < 0.05$ ) in theta, alpha and beta bands with a negative estimate. The negative estimate indicates that the frontal electrodes (F7 and F8) presented a discrepancy value less than that of the temporal electrodes (TP9 and TP10).

We can see that the estimates for the frontal-temporal contrast are small for the ROI-1 and ROI-2 at the theta band indicating a smaller difference between the frontal and temporal discrepancy. Following, the estimate increase at the alpha band (ROI-3 and ROI-4) indicating an increase in the frontal-temporal difference. Then, the estimate decreases at the beta band showing a decrease in the frontal-temporal difference.

The enhancement in the alpha activity in auditory areas related to the processing of degraded or unintelligible speech sounds has been shown in the literature (Fuxe et al., 1998, Weisz et al., 2011, Luo et al., 2005). Challenges to the auditory system arising from signal degradation trigger increased alpha power as observed by Weisz et al. (2011) around 100-200 ms after stimulus onset. Luo et al. (2005) speculate that, during the categorization of speech or intelligible sounds, the auditory areas function for low-level sensory analysis because, for these sounds, the category representations already exist in our brain and then a high-level processing would be performed in the frontal areas. However, for non-speech or unintelligible sounds, the category needs to be learned, then, the auditory areas will be involved in the extraction of characteristics for the categorization in a bottom-up and top-down process together with the frontal cortex. Then, the alpha oscillation helps to select category relevant acoustic properties by inhibiting irrelevant information (such as noise). The larger discrepancy observed at the temporal electrodes can then be explained by an enhanced alpha activity coding the physical characteristics of the stimuli, probably related to the categorization of the more ambiguous speech sounds. Larger values of the discrepancy for the temporal electrodes in the ROI-4 indicates a large dissociation between the physical and psychophysical neural representation of the stimuli.

It has been demonstrated that the gamma band is related to the spectrotemporal analysis of stimuli and in the information transfer through synchronization of brain regions in a sensorimotor integration of frontal and temporal cortical regions (Hickok and Poeppel, 2007, Giraud and Poeppel, 2012). In Bouton et al. (2018) it was shown that the gamma activity tracked the acoustic cue (F2 formant frequency) in the pSTG at an early time-frame (95-120 ms post stimulus onset) while a beta activity coded the decision related to an identification task around 165 ms. The beta band is involved in the process of template matching (Shahin et al., 2009, Bidelman, 2015) coding of sound ambiguity and strength of categorical speech perception (Bidelman, 2015). In the dual-stream model of speech perception Hickok and Poeppel (2007) proposed that basically a spectrotemporal analysis is carried out in auditory cortices (in the dorsal surface of the superior temporal gyrus) which is coded by theta and gamma oscillations (right and left localized, respectively) in the supratemporal plane. Following that, the signals diverge into a dorsal pathway, which maps sensory or phonological representations onto articulatory motor representations, and a ventral pathway, which maps sensory or phonological representations onto lexical conceptual representations. The articulatory motor representations would be localized in more frontal areas including the posterior inferior frontal gyrus, premotor cortex and the anterior insula.

This result shows that both frontal and temporal areas are involved in the physical and categorical (psychophysical) coding of the stimuli with greater discrepancy in the temporal than in the frontal region for theta, alpha and beta bands indicating a better  $\phi - \psi$  dissociation. At the early and late gamma bands (ROI-7 and ROI-8) the discrepancy is greater for the lateral electrodes than for the

medial one. It is interesting to note that at this frequency band the difference between temporal and frontal electrodes is not significant showing that probably, similar processes are coded by gamma activity in those regions. This result may be related with spectrotemporal analysis and auditory-motor areas integration for speech processing, in which the gamma activity is expected to play a role.

### 8.2.3.2 Feature

For the factor feature we contrasted **VOT** against Formants. Figure 126 illustrates the effects observed for this factor in all ROIs.

The contrast analysis was performed using the Kenward-Roger method for degrees of freedom (Kenward and Roger, 1997). As we have only two levels, the contrast was simple and just analyzed the difference between both levels. The results of the contrast analysis for the ROIs where the feature factor was significant are listed as follows:

#### ROI-1

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0144	0.00343	203	-4.193	<.0001

#### ROI-2

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0136	0.00296	199	-4.596	<.0001

#### ROI-3

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0185	0.00326	198	-5.669	<.0001

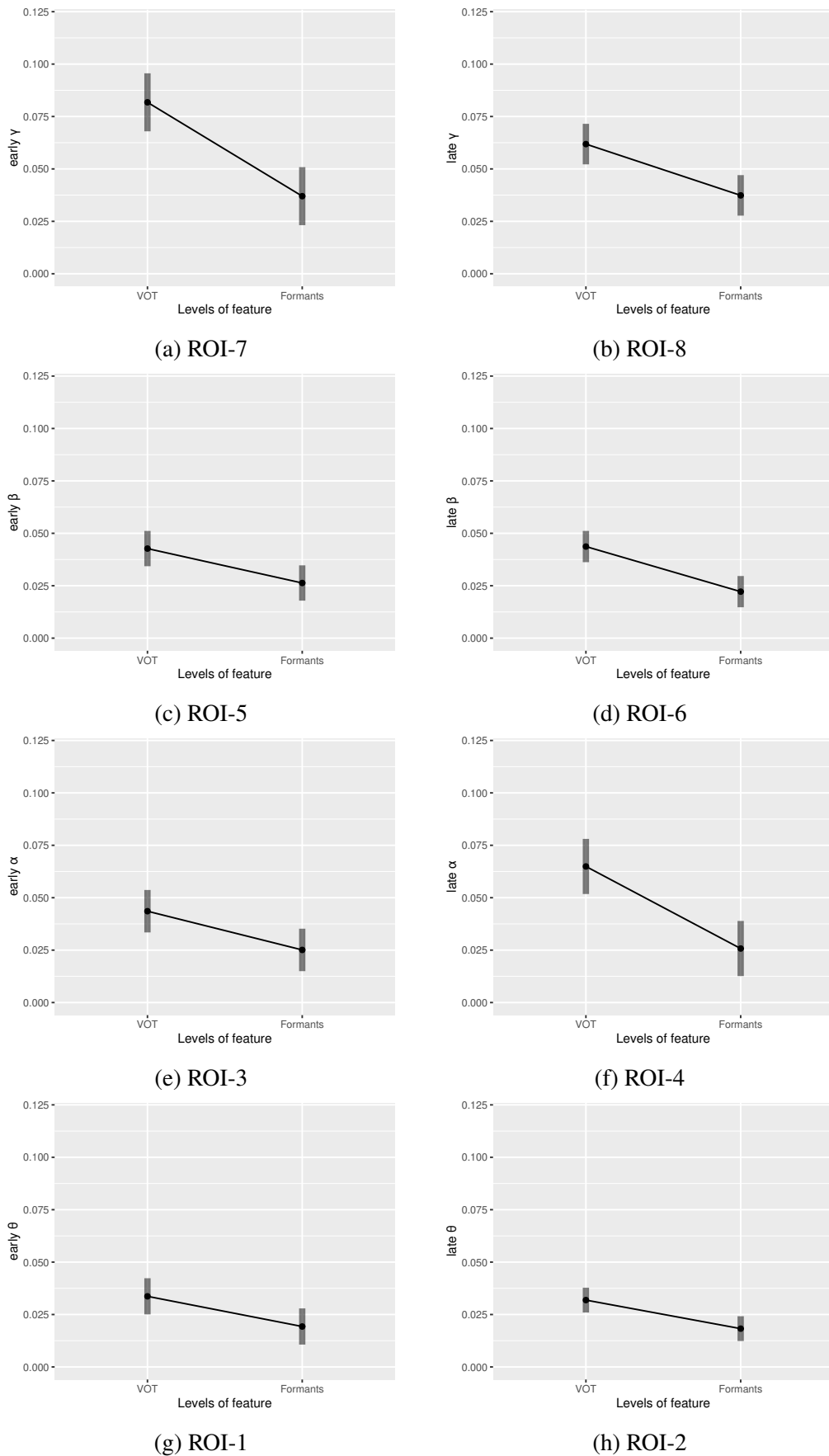


Figure 126 – Representation of the factor feature over the discrepancy computed for all ROIs.

**ROI-4**

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0392	0.00408	200	-9.599	<.0001

**ROI-5**

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0164	0.00303	203	-5.414	<.0001

**ROI-6**

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0215	0.00243	199	-8.857	<.0001

**ROI-7**

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0448	0.00462	199	-9.691	<.0001

**ROI-8**

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants	-0.0245	0.00399	198	-6.133	<.0001

The feature factor showed significant effect in all ROIs analyzed and all of them presented a negative estimate indicating that we observe a decrease in the value of the discrepancy as we compare the VOT with the Formants level (negative slope). Thus, this result indicates that for the VOT acoustic cue the dissociation between its physical and psychophysical neural representation is greater than for the formant-based acoustic cue. This shows that different neural mechanisms

and structures support perception and discrimination of different acoustic cues (Obleser et al., 2008, Pisoni and Remez, 2005). This is also in accordance with works that pointed out that there is a less pronounced categorical perception of formant-based stimuli compared with time-based stimuli as in the VOT continuum (Pisoni, 1973, Altmann et al., 2014b). This result is related to the analysis of the slope vs. angle presented in the Section 8.2.1, where we observed greater angles between the direction vectors for the VOT acoustic cue than for the Formants acoustic cue.

It is interesting to note in this analysis that the acoustic cue is differentiated by the auditory system as early as in the ROI-1 time-frame and at a low frequency band. In this case, probably the theta oscillation is participating in the spectrotemporal analysis of the pre-voicing in the VOT-continuum stimuli or of the formant envelop. It was discussed that a slower-rate integration would be performed by a theta sampling, dominant on the right hemisphere, that would allow, for example, a more accurate analysis of the formant envelop (Giraud and Poeppel, 2012)

### 8.2.3.3 Type

The factor type were analyzed contrasting the passive and active cases. Figure 127 illustrates the effects observed for this factor in ROIs 1, 2, 3, 5, 6, 7 and 8.

In the contrast analysis we have only two levels and the contrast analyzed the difference between both Passive and Active levels. The results of the contrast analysis for the all the ROIs analyzed are listed as follows:

#### ROI-1

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	0.00869	0.00343	203	2.533	0.0121

#### ROI-2

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	0.0117	0.00296	199	3.951	0.0001

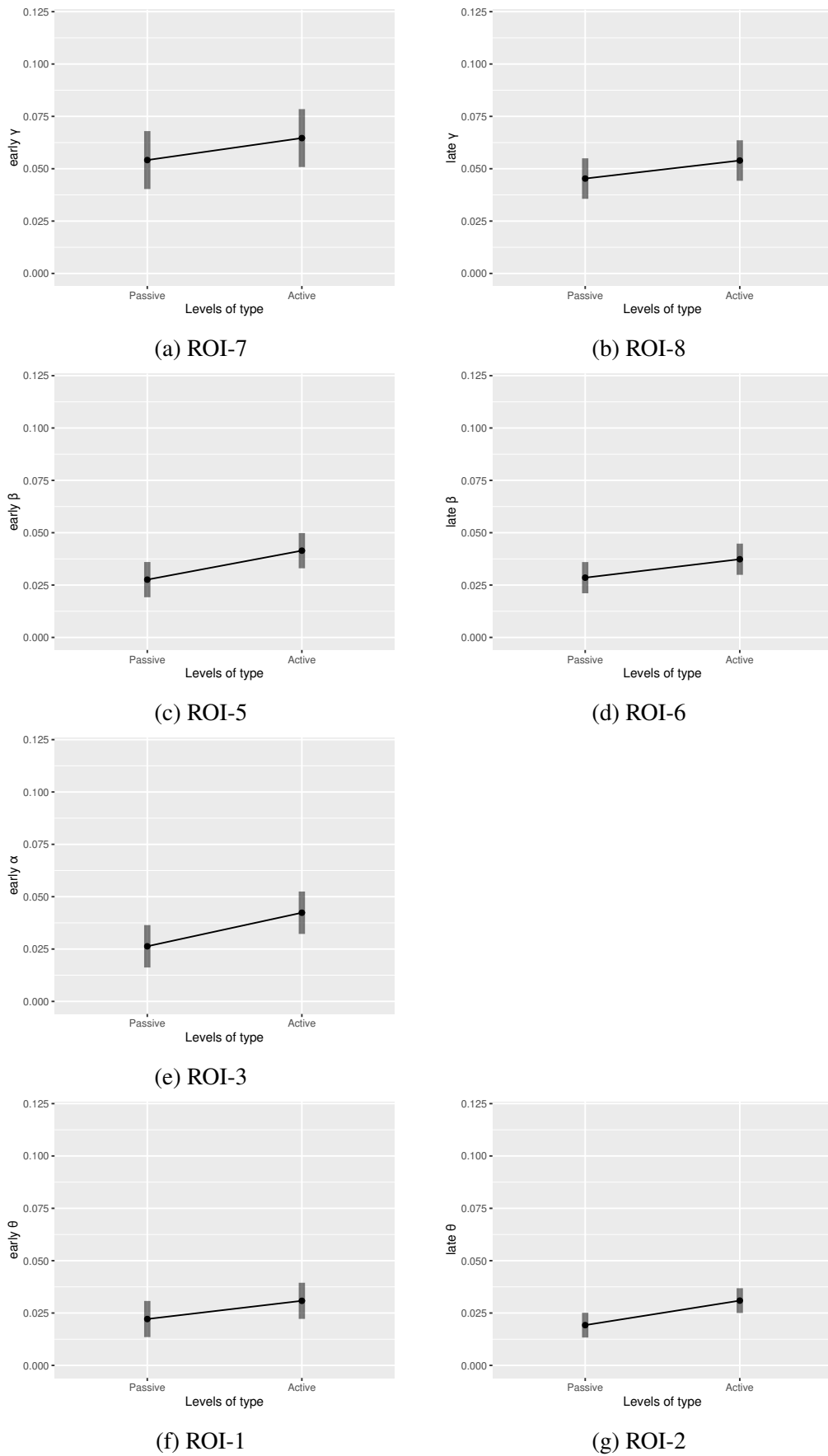


Figure 127 – Representation of the factor type over the discrepancy computed for all ROIs.

**ROI-3**

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	0.016	0.00326	198	4.920	<.0001

**ROI-5**

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	0.0138	0.00303	203	4.563	<.0001

**ROI-6**

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	0.0088	0.00243	199	3.626	0.0004

**ROI-7**

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	0.0105	0.00462	199	2.273	0.0241

**ROI-8**

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active	0.00861	0.00399	198	2.159	0.0320

In all the contrasts we can observe that the estimate has a positive value. This positive estimate indicates a increase in the discrepancy as we compare the Passive and Active task. Specifically for [ROI-7](#) and [ROI-8](#), at the gamma band, it is possible to see that the discrepancies have greater values than those of the other frequency bands. This result is also related to the analysis of the



slope vs. angle presented in the Subsection 8.2.1 where we observed greater angles between the direction vectors for the Active cases than for the Passive ones.

We saw in Section 3 that attention tunes the responses to task-relevant feature values, acting as a band-pass filter which dynamically reallocate cortical resources depending upon task demands and underlines the flexibility in auditory processing (Fritz et al., 2007, Schröger et al., 2015). We also saw that attention enhances relevant information and suppresses the irrelevant ones in the auditory cortex and belt areas (Downer et al., 2017, 2020, Atiani et al., 2014). The result in all ROIs where the type factor presented significant effect is in accordance with these findings. Specifically for ROI-7 and ROI-8 at early and late-gamma band, respectively, the result is in accordance with the work of Viswanathan et al. (2019) in which the authors showed that low-gamma band activity increased with attention to the speech stream.

It is interesting to note that the type factor was not significant for the late-alpha band (ROI-4). Foxe et al. (1998) reported a enhancement in alpha activity related to selective attention to auditory cues, which is in line with a role of enhanced alpha oscillations in inhibiting task-irrelevant information (such as noise). Through our results we can speculate that such enhancement for the stimuli categorization occurred at an early time-frame as we observed effect for the type factor in the early-alpha band (ROI-3). On the other hand, another possibility is that at a late time-frame the alpha enhancement would occur even without attention to the auditory task which will be in line with a late processing of speech occurring when the participants ignored the sounds during an experiment (Möttönen et al., 2014). In this case, the values of discrepancy for both Active and Passive conditions will be similar resulting in the non existence of significant effect for the factor type.

#### 8.2.3.4 Feature-electrode

The feature-electrode interaction analyzes how the discrepancy computed for each electrode change as we vary from the feature VOT to the feature Formants. This factor interaction presented significant effects for ROI-2, -4 and -6. Figure 128 illustrates these effects.

The contrast analysis included the difference VOT-Formants for each electrode contrast: medial-lateral, frontal-temporal and left-right. The contrasts results for each ROI where this factor interaction was significant are as follows:

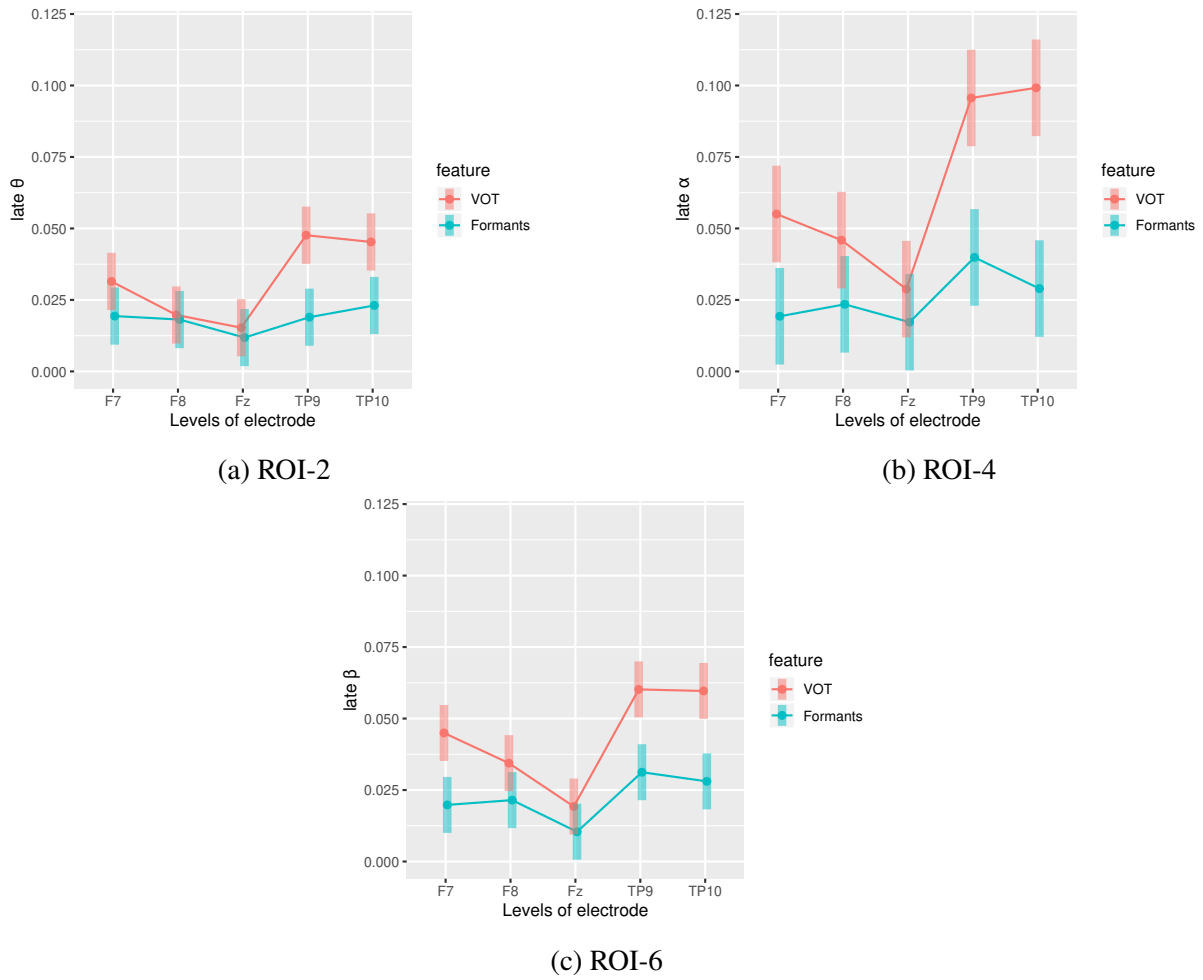


Figure 128 – Representation of the factor interaction feature-electrode over the discrepancy computed for the ROI-2, ROI-4 and ROI-6.

**ROI-2**

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : medial - lateral	0.01270	0.00741	199	1.715	0.0879
VOT - Formants : frontal - temporal	0.01858	0.00662	199	2.806	0.0055
VOT - Formants : left - right	-0.00203	0.00662	199	-0.306	0.7597

**ROI-4**

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : medial - lateral	0.0345	0.01020	200	3.383	0.0009
VOT - Formants : frontal - temporal	0.0339	0.00912	200	3.715	0.0003
VOT - Formants : left - right	-0.0139	0.00912	200	-1.522	0.1297

**ROI-6**

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : medial - lateral	0.01583	0.00607	199	2.608	0.0098
VOT - Formants : frontal - temporal	0.01124	0.00543	199	2.070	0.0398
VOT - Formants : left - right	-0.00744	0.00543	199	-1.370	0.1721

The feature-electrode factors interaction presented effects for the ROIs -2, -4 and -6. Note that those ROIs are at a late time-frame. Also, observe that for all electrodes and ROIs here, the discrepancy for the VOT acoustic cue was greater than for the Formants acoustic cue. This result is in line with the contrasts analyzed for the feature factor before.

For the contrast *VOT - Formants : frontal - temporal* the positive estimates show that the discrepancies between VOT and Formants are greater for the temporal electrodes than for the frontal ones. Then, it seems that different mechanisms are involved in the processing of each acoustic cue at the temporal and frontal regions. In the ROI-2 the processing at the frontal region seems to be similar for both acoustic cues suggesting a more high-level processing in this region coded by a theta activity while for the temporal electrodes the discrepancy values are quite different comparing acoustic cues suggesting a processing more related to the physical characteristic of the stimuli there. However, as the discrepancy is a measure of distance, this observations can be just a coincidence and would need further analyses to be validated. The result observed for the ROI-2 shows that the theta activity also code the VOT acoustic cue and present a more pronounced distinction between the physical and psychophysical neural representations of this acoustic cue than for the stimuli from the formant-based continuum. This result is interesting because the literature reports that slower-rate integration by a theta-dominant sampling would allow for a more accurate analysis of the formant envelop than rapidly varying acoustic cues as the VOT.

Regarding the contrast *VOT - Formants : medial - lateral*, we can see that the difference between the discrepancies of the VOT and Formants is greater for the lateral electrodes than for the medial one. This effect was significant only for ROI-4 and ROI-6. In both ROIs and for both continua, we can see that the discrepancy value for the medial region is smaller than for the other regions.



Figure 129 – Representation of the factor interaction feature-type over the discrepancy computed for the ROI-3 and ROI-8.

### 8.2.3.5 Feature-type

The interaction factor feature-type indicate how the type factor vary according to the feature (VOT or Formants). This interaction presented significant effects in the ROI-3 and -8 and is illustrated in Figure 129.

The contrast analysis took into account the interaction of the differences Passive-Active and VOT-Formants. The contrasts results for each ROI where this factor interaction was significant are as follows:

#### ROI-3

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : Passive - Active	0.0158	0.00652	198	2.431	0.0159

#### ROI-8

contrast	estimate	SE	df	t.ratio	p.value
VOT - Formants : Passive - Active	0.016	0.00798	198	2.001	0.0468

The feature-type interaction presented significant effects at the ROI-3 and ROI-8, both with positive estimates indicating that the difference in the discrepancy value between the acoustic cues decrease as we go from the Passive to the Active task. In both ROIs and tasks, the discrepancy value for the VOT acoustic cue was greater than for the Formants which is in line with the analysis performed for the factor *feature*.

For ROI-8, the difference VOT-Formants is smaller in the active task than in the passive task. This can indicate that, at the ROI-8 time-frame, the gamma band is involved in a processing that is similar for both features. This was also observed for the ROI-3. This can be related to the phonemic identification or categorical perception of the acoustic cue. The time-frame is in accordance with what is known from studies that involved experiments with active tasks (Alho et al., 2016, Bouton et al., 2018, Bidelman et al., 2013). In the passive task, it seems that other kind of processing, which is more influenced by the acoustic cue, is occurring at this time-frame given the significant difference between the discrepancy values for VOT and Formants. It can be related to some late spectrotemporal analysis, given that it is expected late speech processing when there is not attention to the speech sounds Möttönen et al. (2014), Alho et al. (2016), or it is possible that some categorical processing is occurring for some feature. In this case, this processing will involve different mechanisms or even generators for each feature given the big difference in the discrepancy values from the passive for the active case. It seems that in the ROI-8 the VOT discrepancy does not vary significantly with the task which will show that the difference between physical and psychophysical neural representations of the stimuli are not affected by the task for this acoustic cue. However, a statistical test for only the VOT feature will be necessary to validate this observation.

Bidelman (2017) reported that the categorical perception was predicted by early (150 ms) evoked alpha activity in an active identification task. We also saw that an enhancement in the alpha activity is related to selective attention and suppression of irrelevant acoustic cues for the processing of unintelligible speech (Foxe et al., 1998, Weisz et al., 2011, Luo et al., 2005). The increase in the discrepancy values for both continua comparing the passive with the active tasks can be observed in the Figure 129a and this can be related to the enhancement of the alpha activity.

### 8.2.3.6 Type-electrode

The type-electrode interaction analyzes the discrepancy computed for each electrode change as we vary from the passive to the active task. This interaction factor presented effects in ROIs 3, 7 and 8. Figure 130 illustrates these effects.

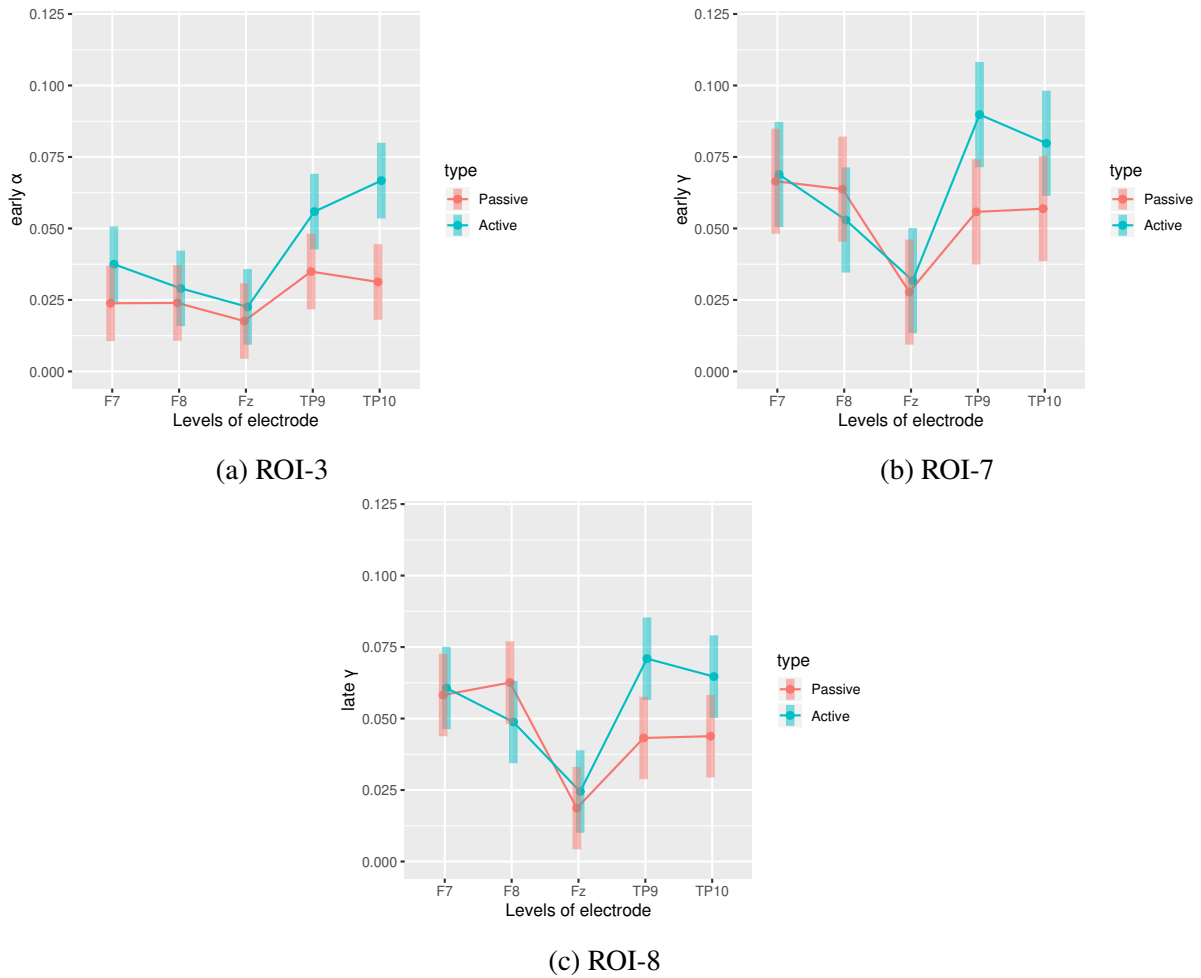


Figure 130 – Representation of the factor interaction type-electrode over the discrepancy computed for the ROI-3, ROI-7 and ROI-8.

The contrast analysis included the difference passive-active for each electrode contrast (medial-lateral, frontal-temporal, and left-right). The results for each ROI where this factor interaction was significant are the following:

**ROI-3**

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active : medial - lateral	-0.0138	0.00814	198	-1.699	0.0909
Passive - Active : frontal - temporal	-0.0188	0.00728	198	-2.587	0.0104
Passive - Active : left - right	0.0115	0.00728	198	1.578	0.1161

**ROI-7**

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active : medial - lateral	-0.00812	0.0115	199	-0.703	0.4827
Passive - Active : frontal - temporal	-0.03266	0.0103	199	-3.163	0.0018
Passive - Active : left - right	0.00100	0.0103	199	0.097	0.9230

**ROI-8**

contrast	estimate	SE	df	t.ratio	p.value
Passive - Active : medial - lateral	-0.00352	0.00997	198	-0.353	0.7245
Passive - Active : frontal - temporal	-0.02996	0.00892	198	-3.359	0.0009
Passive - Active : left - right	0.00468	0.00892	198	0.525	0.6005

The type-electrode interaction factor presented significant effect for the contrast *Passive - Active : frontal - temporal* in the ROIs -3, -7 and -8 as can be visualized in the Figure 130. The negative estimate indicates that the difference of the discrepancy between the passive and the active task is more pronounced in the temporal electrodes than in the frontal ones. We can see that the effect analyzed for the *type* factor seems to be more related to activity in the temporal regions than in the frontal ones. Thus, the distinction between the physical and psychophysical neural representation of the stimuli is greater at the temporal cortical regions than in frontal regions. This shows that some categorical perception seems to be coded by activity at the temporal regions and not only a spectrotemporal analysis of physical characteristics of the stimuli.

## Chapter 9

# CONCLUSION

In this chapter we present and discuss the main conclusions about the investigation of the neural correlates of speech sounds categorical perception performed by means of analysis of ERPs in the time and time-frequency domains and separating the physical and psychophysical aspects of the response. We also suggest some improvements and further investigations for the continuation of this work in the future.

### 9.1 General discussion

We saw that studies of categorical perception differ among themselves in several aspects as in the acoustic cue analyzed, the brain potential measurement systems, the auditory task, the number of trials, the brain regions measured, among others. Thus, it is difficult to compare studies regarding the effects of attention and acoustic cue on the categorical perception and separating the physical and psychophysical characteristics of these effects.

Then, we investigated the neural correlates of categorical perception of speech sounds, specifically of Brazilian Portuguese phonemes, evaluating **AELR** in the scope of the stimulus acoustic characteristic (**VOT** and formant frequencies). We studied the brain cortical regions involved in speech perception (temporal and frontal), manipulating the degree of attention to the identification task and using data acquired with the use of a non-invasive method. In our analysis, we propose to identify the physical and psychophysical responses in the **ERP**, in order to show how the modulations in the time and frequency characteristics of the ERPs can be related to the phonemic



categorical perception.

As one of our objectives, we developed and applied an experimental protocol with active and passive tasks where acoustic cues based on **VOT** and formant frequency variations were used to generate the phonemic continua. For each participant, a set of five stimuli (stim1, . . . , stim5) was selected for a phonemic identification experiment. Stim1 and stim5 are the extreme ends of the continuum and are more unambiguously identified, while stim3 is the more ambiguous one. Analysis were carried out in time and time-frequency domain using mixed-effects models, in order to study the effects in the different acoustic cues, stimuli, tasks and cortical areas.

In the time-domain analysis, we investigated the effects on the amplitude and latency of the N1 and P2 **ERP** components. For the analysis of the time-frequency characteristics of the **CP**, a regression technique named **RoLDSIS** was developed in order to identify the coefficients of the **DWT**, applied to the averaged **ERP**, that are related to the physical and psychophysical behaviors of each participant. We defined **ROIs** of regression coefficients related to each **DWT** coefficient in order to perform our analysis. For each **ROI** we measured the distance (discrepancy) between the physical and psychophysical neural representations of the stimuli computed through the regression coefficients. We observed significant effects in different regions in the time-frequency domain depending on the acoustic cue, task and electrode. We also computed the correlation between the angles between the physical and psychophysical vectors of coefficients returned by the **RoLDSIS** technique and the slope of the psychometric curve obtained for each subject and continuum (slope vs. angle).

The analysis of the slope vs. angle showed a different treatment for the **VOT** acoustic cue depending on the task with the **CP** occurring only in the Active task while there was no difference for the Formant continuum related to the task. This result is in accordance with the analysis of the factors *type* and *feature* where the **VOT** acoustic cue presented greater discrepancy values than the Formants which also occurred for the Active task discrepancies in relation to those of the Passive task. These effects were observed for the majority of the **ROIs** showing that the *phi* – *psi* distinction is coded by the theta, alpha, beta and gamma bands in different time-frames and it is affected by the task and acoustic cue.

The ambiguity seems to be coded by the P2 wave as can be seen in the analysis in Section 6.3.3. P2 is also the main wave in the time-frame of **ROIs** 2, 4, 6 and 8. We saw in Section 6.3.3 that there was no significant difference on the P2 behavior across stimuli comparing **VOT** and Formants. In Section 6.3.6, this was confirmed again for the P2 wave. This probably suggests that a higher level processing of speech occurs at that latency and is independent of the acoustic cue. It is interesting to note that the contrast frontal-temporal was significant for the factor feature-electrode in the P2 analysis and also in our late-bands **ROIs** 2, 4 and 6 (which encompass

the P2 time-frame) with the exception of the late-gamma band (ROI-8). For this latter ROI the frontal-temporal contrast was significant for the factor type-electrode as also observed in the P2 analysis. This suggests that generators of the P2 component may be affected by activity in the theta, alpha, beta and gamma bands and code both the perception of the acoustic cue and the task. In fact, Schabus (2001) demonstrated that the “P1-N1-P2” complex is the manifestation of oscillatory processes in the alpha and theta bands and related the “P1-N1” complex to sensory and early attentional processes, just as it is suggested for the alpha rhythms. The author also related an increase in the N1-P2 peak-to-peak amplitudes with an event-related theta synchronization for the “encoding of new information”.

In our latency investigation, in the feature-type analysis of Section 6.3.4 we observed an effect which was probably caused by a difference in the N1 latency for the Formants case. This may be explained by the tuning model proposed by Ahveninen et al. (2011), which explains how attention modulates the feature selectivity of auditory neurons at the N1 time-frame. This is also related to the study of Foxe et al. (1998) reported an enhancement in alpha activity related to selective attention to auditory cues, which is in line with a role of enhanced alpha oscillations in inhibiting task-irrelevant information (such as noise) (see Section 8.2.3.3). In our time-frequency domain analysis, we observed an effect of the feature-type interaction factor on ROI-3 (early-alpha band). Thus, it would be possible to associate at least one of the generators of the N1 wave to this tuning model and to an alpha band activity corroborating the observations in Schabus (2001) reported before in this section for the P1-N1 complex.

Regarding the brain region, in both time and in the time-frequency domains, in general, we observed stronger responses in the temporal region in comparison with the frontal region. In the time-frequency domain analysis we observed the effect of the electrode factor over all ROIs with greater values of discrepancy for the temporal electrodes than for the medial and frontal electrodes coded by activity in the theta, alpha and beta bands. This is perhaps not surprising, because the auditory cortices are located at the temporal area. However, for the gamma band in both early and late time-frames, there was no significant difference between the discrepancy obtained for the frontal and temporal regions indicating that probably, similar processes are coded by this rhythm at those regions. We also observed in our time-domain analysis, a left hemisphere dominance in our tasks (subsection 6.3.1, subsection 6.3.3, subsection 6.3.4). This can be related to the gamma band oscillation involved either in spectrotemporal analysis or in information transfer along the dorsal or ventral auditory streams at this hemisphere (Giraud and Poeppel, 2012, Hickok and Poeppel, 2007, Bidelman, 2017). Furthermore, this result shows that some categorical perception is coded by activity in the temporal region demonstrating that structures at this region are not only involved in the processing of acoustic characteristics of the sounds.

In our time-domain analysis we showed that **VOT** influences the N1 and N1-P2 amplitudes. This influence in the N1 amplitude does not occur in the Formants case. This suggests that **VOT** processing begins as soon as the latency of the N1 wave. In Section 8.2.3.2 we observed significant effect in the *feature* factor in all ROIs with the discrepancy being greater for **VOT** than for Formants. This evidences that the categorical perception (here observed through a greater distance between the physical and psychophysical neural representations of the stimuli) is more pronounced in time-based acoustic cues as the **VOT** than in vowels, as in our formant continuum (Altmann et al., 2014b, Pisoni, 1973). Still, we saw in the analysis performed in the Section 6.3.1 that the formant frequency principle on the N1-P2 amplitude is valid for Brazilian Portuguese, as was observed in other languages (Ohl and Scheich, 1997, Obleser et al., 2003, Manca et al., 2013, Shestakova et al., 2004, Kim et al., 2018). This is interesting because it can indicate that the generators involved in the processing of formant frequencies may be similar across languages.

The contrast analysis performed for the N1 and P2 latencies in Section 6 supports the fact that different neural mechanisms are involved in the perception and discrimination of different acoustic cues. Different latencies in the response may indicate (i) the firing of different generators or (ii) the firing of the same generators with a different pattern or (iii) both. Also in that section, we observed that the amplitude of responses were greater for the active task, in relation with the passive task. In the time-frequency analysis we saw a greater value of discrepancy for active than for passive tasks indicating a better distinction between the physical and psychophysical neural representations of the stimuli when there is attention to the sounds. This effect was observed for all time-frames and frequency bands with the exception of the late-alpha band (ROI-4). We suggest that this may be occurred or because the effect of the task was coded by the alpha band at an early time-frame or because there was an equal increase in the discrepancy value for the passive case.

From the ROI analyses, we observed that, for the same frequency band, the behavior of the discrepancy was different depending on the time-frame and the factor analyzed. This was also observed for the time-domain analysis of amplitudes and latencies of the ERP components. This emphasizes how the tasks and acoustic cues modulate the activity of each brain region. According to the literature, different neural mechanisms and structures are involved in the processing of acoustic cues for different auditory tasks (Obleser et al., 2008, Pisoni and Remez, 2005). Thus, the dual-stream model of speech perception seems to be too “static” and doesn’t encompass this dynamic reconfiguration of the speech processing networks that we observed here. In Skipper et al. (2017) the authors argue that the production system include more structures than those indicated in the dual-stream model. Besides, the authors also report that ventral-stream regions are also involved in speech production (not only dorsal-stream regions) and that, conversely, dorsal-stream regions are also involved in speech perception (not only ventral-stream ones).

In general, our experiments on Brazilian Portuguese confirmed several findings related to speech categorical perception found in the literature. This was the case for several effects observed in the amplitude and latency of the ERP waves and in the time-frequency analysis of the regression coefficients related to physical and psychophysical responses. Our results confirmed some observations from the literature in which the majority of the experiments encompassed formant-based acoustic cues and active identification tasks. We showed that the selection of the acoustic cue and task influence the brain oscillations, time-frame and brain regions (generators) that participate in the speech processing so that those factors should be taken into account when developing an experimental protocol.

We also showed that the mechanisms of CP are different depending on the acoustic cue. This result is also corroborated by correlations between the psychophysical-physical direction vector angles and the slope of the psychometric curve. We noted that the strength of CP is positively correlated with the psy-phy angles, what suggests that participants which categorized better present larger differences in the internal psychophysical and physical representation of the acoustic cues which was confirmed by the analysis of the feature and type factors in Sections 8.2.3.2 and 8.2.3.3 for all frequency bands.

Our results not only validate our regression technique but also shows that the behavioral result is correlated with information that is contained in the early stages of the speech processing (N1 time-frame) and coded by different brain oscillations which seems to be executing different stages of the speech processing. In this thesis, we demonstrate that these results can be obtained directly from EEG measurements. It also shows that this processing is similar for Brazilian Portuguese, English and French speakers, involving the same time-frame and brain oscillation bands.

## 9.2 Availability of the code

The scripts written in the context of the present thesis are available at <https://github.com/Adrielle-Santana/ThesisScripts> and <https://github.com/RoLDSIS/code>. A brief description of each script is available at the Appendix I.

### 9.3 Future extensions

Here are some suggestions for improvements of the work done in this thesis:

- Since monosyllabic words with meaning in Brazilian Portuguese were used, a semantic processing can be present in the [AEP](#) (involving the ventral auditory stream), but longer latencies (where semantic content seems to be processed) was not the focus of the analysis. Thus, as a future work, our data can be also used for an analysis of the effects of attention and acoustic cue in the semantic processing of the stimuli with a focus in longer latency waves as the N2 and N400.
- Work with the continuous wavelet transform ([CWT](#)) to improve the time and frequency resolution in the ROIs selection.
- Work with different continua to observe if the same general effects we observed in the time and time-frequency analysis remain, validating our experimental protocol and analysis methods.
- Consider the evaluation of other dimensions to be analyzed besides the acoustic cue and task. For example, the use of stimuli from different speakers, spatial location of the sound, context of the speech sound, etc.

# Bibliography

- Abdi, H. and Williams, L. J. (2010). Contrast analysis. *Encyclopedia of research design*, 1:243–251.
- Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *Journal of Neuroscience*, 28(15):3958–3965.
- Adler, G. and Adler, J. (1989). Influence of stimulus intensity on aep components in the 80-to 200-millisecond latency range. *Audiology*, 28(6):316–324.
- Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1515):285–299.
- Ahveninen, J., Hämäläinen, M., Jääskeläinen, I. P., Ahlfors, S. P., Huang, S., Lin, F.-H., Raij, T., Sams, M., Vasios, C. E., and Belliveau, J. W. (2011). Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proceedings of the National Academy of Sciences*, 108(10):4182–4187.
- Alain, C., Woods, D. L., and Covarrubias, D. (1997). Activation of duration-sensitive auditory cortical fields in humans. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 104(6):531–539.
- Alcain, A. and Oliveira, C. d. S. (2011). *Fundamentos do processamento de sinais de voz e imagem*, volume 66.
- Aldrich, E. (2013). *wavelets: A package of functions for computing wavelet filters, wavelet transforms and multiresolution analyses*. R package version 0.3-0.
- Alexander, J. E. and Polich, J. (1997). Handedness and p300 from auditory stimuli. *Brain and Cognition*, 35(2):259–270.
- Alho, J., Green, B. M., May, P. J., Sams, M., Tiitinen, H., Rauschecker, J. P., and Jääskeläinen, I. P. (2016). Early-latency categorical speech sound representations in the left inferior frontal gyrus. *Neuroimage*, 129:214–223.

- Alho, J., Lin, F.-H., Sato, M., Tiitinen, H., Sams, M., and Jääskeläinen, I. P. (2014). Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. *Frontiers in psychology*, 5:394.
- Alho, K., Paavilainen, P., Reinikainen, K., Sams, M., and Näätänen, R. (1986a). Separability of different negative components of the event-related potential associated with auditory stimulus processing. *Psychophysiology*, 23(6):613–623.
- Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., and Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 77(2):151–155.
- Alho, K., Sams, M., Paavilainen, P., and Näätänen, R. (1986b). Small pitch separation and the selective-attention effect on the erp. *Psychophysiology*, 23(2):189–197.
- Altmann, C. F., Uesaki, M., Ono, K., Matsushashi, M., Mima, T., and Fukuyama, H. (2014a). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia*, 64:13–23.
- Altmann, C. F., Uesaki, M., Ono, K., Matsushashi, M., Mima, T., and Fukuyama, H. (2014b). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia*, 64:13–23.
- Anderson, S. and Kraus, N. (2011). Neural Encoding of Speech and Music: Implications for Hearing Speech in Noise. In *Seminars in Hearing Proceedings of the Widex Pediatric Audiology Congress; Guest Editor, André M. Marcoux, Ph.D. Semin Hear*, number 2, pages 207–212. Thieme.
- Angelini, C. and Vidakovic, B. (2003). Some novel methods in wavelet data analysis: wavelet anova, f-test shrinkage, and  $\gamma$ -minimax wavelet shrinkage. *Wavelets and their Applications*, pages 31–45.
- ASHA, A. S.-L.-H. A. (1987). Short latency auditory evoked potentials.
- Atiani, S., David, S. V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S. A., and Fritz, J. B. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron*, 82(2):486–499.
- Bartlett, J. (2014). Robustness of linear mixed models. <https://thestatsgeek.com/2014/08/17/robustness-of-linear-mixed-models/>.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Bates, D. M. (2010). lme4: Mixed-effects modeling with r. <http://lme4.r-forge.r-project.org/book/>.

- Bellier, L., Mazzuca, M., Thai-Van, H., Caclin, A., and Laboissière, R. (2013). Categorization of speech in early auditory evoked responses. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pages 911–915.
- Bertrand, O., Bohorquez, J., and Pernier, J. (1994). Time-frequency digital filtering based on an invertible wavelet transform: an application to evoked potentials. *IEEE Transactions on Biomedical Engineering*, 41(1):77–88.
- Bertrand, O., Perrin, F., and Pernier, J. (1991). Evidence for a tonotopic organization of the auditory cortex observed with auditory evoked potentials. *Acta Oto-Laryngologica*, 111(sup491):116–123.
- Bever, T. G. and Poeppel, D. (2010). Analysis by synthesis: a (re-) emerging program of research for language and vision. *Biolinguistics*, 4(2-3):174–200.
- Bickford, R. D. (1987). Electroencephalography. In Adelman, G., editor, *Encyclopedia of Neuroscience*, pages 371–373. Birkhauser, Cambridge (USA).
- Bidelman, G. M. (2014). Objective information-theoretic algorithm for detecting brainstem-evoked responses to complex stimuli. *Journal of the American Academy of Audiology*, 25(8):715–726.
- Bidelman, G. M. (2015). Induced neural beta oscillations predict categorical speech perception abilities. *Brain and language*, 141:62–69.
- Bidelman, G. M. (2017). Amplified induced neural oscillatory activity predicts musicians' benefits in categorical speech perception. *Neuroscience*, 348:107–113.
- Bidelman, G. M. and Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage*, 120:191–200.
- Bidelman, G. M., Moreno, S., and Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, 79:201–212.
- Bidelman, G. M. and Walker, B. S. (2017). Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception. *European Journal of Neuroscience*, 45(5):690–699.
- Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, 40(4):2662–2673.
- Birjandtalab, J., Pouyan, M. B., Cogan, D., Nourani, M., and Harvey, J. (2017). Automated seizure detection using limited-channel eeg and non-linear dimension reduction. *Computers in biology and medicine*, 82:49–58.



- Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature neuroscience*, 8(3):389.
- Boersma, P. and Weenink, D. (2018). Praat: doing phonetics by computer [computer program].
- Borsky, S., Tuller, B., and Shapiro, L. P. (1998). “how to milk a coat:” the effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, 103(5):2670–2676.
- Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A. G., Seeck, M., Karkar, S., van de Ville, D., and Giraud, A.-L. (2018). Focal versus distributed temporal cortex activity for speech sound category assignment. *Proceedings of the National Academy of Sciences*, 115(6):E1299–E1308.
- Boutros, N. N. and Belger, A. (1999). Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. *Biological psychiatry*, 45(7):917–922.
- Boutros, N. N., Torello, M. W., Barker, B. A., Tueting, P. A., Wu, S.-C., and Nasrallah, H. A. (1995). The p50 evoked potential component and mismatch detection in normal volunteers: implications for the study of sensory gating. *Psychiatry Research*, 57(1):83–88.
- Bruneau, N., Roux, S., Guerin, P., Barthelemy, C., and Lelord, G. (1997). Temporal prominence of auditory evoked potentials (n1 wave) in 4-8-year-old children. *Psychophysiology*, 34(1):32–38.
- Campbell, K. B. and Colrain, I. M. (2002). Event-related potential measures of the inhibition of information processing: Ii. the sleep onset period. *International Journal of Psychophysiology*, 46(3):197–214.
- Castellanos, N. P. and Makarov, V. A. (2006). Recovering eeg brain signals: artifact suppression with wavelet enhanced independent component analysis. *Journal of neuroscience methods*, 158(2):300–312.
- Čeponien, R., Kushnerenko, E., Fellman, V., Renlund, M., Suominen, K., and Näätänen, R. (2002). Event-related potential features indexing central auditory discrimination by newborns. *Cognitive Brain Research*, 13(1):101–113.
- Čeponiené, R., Shestakova, A., Balan, P., Alku, P., Yiaguchi, K., and Naatanen, R. (2001). Children’s auditory event-related potentials index sound complexity and “speechness”. *International Journal of Neuroscience*, 109(3-4):245–260.
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature neuroscience*, 13(11):1428.

- Cheour-Luhtanen, M., Alho, K., Sainio, K., Rinne, T., Reinikainen, K., Pohjavuori, M., Renlund, M., Aaltonen, O., Eerola, O., and Näätänen, R. (1996). The ontogenetically earliest discriminative response of the human brain. *Psychophysiology*, 33(4):478–481.
- Chevillet, M. A., Jiang, X., Rauschecker, J. P., and Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Journal of Neuroscience*, 33(12):5208–5215.
- Cho, T. and Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27(2):207 – 229.
- Chun, H. and Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat. Soc. Series B Stat. Methodol.*, 72(1):3–25.
- Chun, H. and Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3–25.
- Chung, D., Chun, H., and Keles, S. (2019). *spls: Sparse Partial Least Squares (SPLS) regression and classification*. R package version 2.2-3.
- Cohen, A. and Kovacevic, J. (1996). Wavelets: The mathematical background. *Proceedings of the IEEE*, 84(4):514–522.
- Cohen, M. S. (2008). Handedness questionnaire. online.
- Crottaz-Herbette, S. and Ragot, R. (2000). Perception of complex sounds: N1 latency codes pitch and topography codes spectra. *Clinical neurophysiology*, 111(10):1759–1766.
- Crowley, K. E. and Colrain, I. M. (2004). A review of the evidence for p2 being an independent component process: age, sleep and modality. *Clinical neurophysiology*, 115(4):732–744.
- Cummings, B. (2001). Meninges.
- da Silva, F. L. and Rotterdam, A. V. (2012). Biophysical Aspects of EEG and Magnetoencephalogram Generation. In Schomer, D. L. and da Silva, F. L., editors, *Niedermeyer’s Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, chapter 5. Lippincott Williams & Wilkins.
- David, O., Harrison, L., and Friston, K. (2007). Neuronal models of eeg and meg. In *Statistical Parametric Mapping*, pages 414–440. Elsevier.
- Davis, H. (1964). Enhancement of evoked cortical potentials in humans related to a task requiring a decision. *Science*, 145(3628):182–183.
- Davis, H., Mast, T., Yoshie, N., and Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: recovery process. *Electroencephalography and clinical neurophysiology*, 21(2):105–113.

- Davis, H. and Zerlin, S. (1966). Acoustic relations of the human vertex potential. *The Journal of the Acoustical Society of America*, 39(1):109–116.
- Davis, M. H., Johnsruide, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134(2):222.
- Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455.
- Downer, J. D., Rapone, B., Verhein, J., O'Connor, K. N., and Sutter, M. L. (2017). Feature-selective attention adaptively shifts noise correlations in primary auditory cortex. *Journal of Neuroscience*, 37(21):5378–5392.
- Downer, J. D., Verhein, J. R., Rapone, B. C., O'Connor, K. N., and Sutter, M. L. (2020). An emergent population code in primary auditory cortex supports selective attention to spectral and temporal sound features. *bioRxiv*.
- Duncan, G. (2019). Bilinguals' double phonemic boundary: Not one of normal nature.
- Eddins, A. C. and Peterson, J. R. (1999). Time-intensity trading in the late auditory evoked potential. *Journal of Speech, Language, and Hearing Research*, 42(3):516–525.
- Eggermont, J. J. (1995). Representation of a voice onset time continuum in primary auditory cortex of the cat. *The Journal of the Acoustical Society of America*, 98(2):911–920.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968):303–306.
- Escudero, P., Boersma, P., Rauber, A. S., and Bion, R. A. (2009). A cross-dialect acoustic description of vowels: Brazilian and european portuguese. *The Journal of the Acoustical Society of America*, 126(3):1379–1393.
- Etard, O. and Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, 39(29):5750–5759.
- Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological review*, 116(4):752.
- Foxe, J., Simpson, G., and Ahlfors, S. (1998). Parieto-occipital–10 hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport*, 9(17):3929–3933.

- Friedman, J., Hastie, T., and Tibshirani, R. (2010a). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, 33(1).
- Friedman, J., Hastie, T., and Tibshirani, R. (2010b). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.
- Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Current opinion in neurobiology*, 17(4):437–455.
- Gajic, D., Djurovic, Z., Di Gennaro, S., and Gustafsson, F. (2014). Classification of eeg signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. *Biomedical Engineering: Applications, Basis and Communications*, 26(02):1450021.
- German-Sallo, Z. and Ciufudean, C. (2012). Waveform-adapted wavelet denoising of ecg signals. *Adv. Math. Computat. Methods*, 172175.
- Giard, M., Perrin, F., Echallier, J., Thevenet, M., Froment, J., and Pernier, J. (1994). Dissociation of temporal and frontal components in the human auditory n1 wave: a scalp current density and dipole model analysis. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(3):238–252.
- Giard, M., Perrin, F., Pernier, J., and Peronnet, F. (1988). Several attention-related wave forms in auditory areas: a topographic study. *Electroencephalography and Clinical Neurophysiology*, 69(4):371–384.
- Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, 15(4):511.
- Godey, B., Schwartz, D., De Graaf, J., Chauvel, P., and Liegeois-Chauvel, C. (2001). Neuro-magnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clinical neurophysiology*, 112(10):1850–1859.
- Gordon, K., Tanaka, S., Wong, D., and Papsin, B. (2008). Characterizing responses from auditory cortex in young people with several years of cochlear implant experience. *Clinical Neurophysiology*, 119(10):2347–2362.
- Hall, J. W. (2007). *New handbook of auditory evoked responses*. Pearson.
- Hall III, J. (2015). *eHandbook of auditory evoked responses: Principles, procedures & protocols*. Pretoria: Pearson.
- Han, M., Liu, Y., Xi, J., and Guo, W. (2006). Noise smoothing for nonlinear time series using wavelet soft threshold. *IEEE signal processing letters*, 14(1):62–65.
- Han, W. (2010). P1-n1-p2 complex and acoustic change complex elicited by speech sounds: current research and applications. *Audiology*, 6(2):121–127.

- Han, X., Chen, M., Wang, F., Windrem, M., Wang, S., Shanz, S., Xu, Q., Oberheim, N. A., Bekar, L., Betstadt, S., et al. (2013). Forebrain engraftment by human glial progenitor cells enhances synaptic plasticity and learning in adult mice. *Cell stem cell*, 12(3):342–353.
- Hansen, J. C. and Hillyard, S. A. (1980). Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and clinical neurophysiology*, 49(3-4):277–290.
- Hari, R., Kaila, K., Katila, T., Tuomisto, T., and Varpula, T. (1982). Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. *Electroencephalography and clinical neurophysiology*, 54(5):561–569.
- Hari, R., Sams, M., and Järvilehto, T. (1979a). Auditory evoked transient and sustained potentials in the human eeg: I. effects of expectation of stimuli. *Psychiatry Research*, 1(3):297–306.
- Hari, R., Sams, M., and Järvilehto, T. (1979b). Auditory evoked transient and sustained potentials in the human eeg: II. effects of small doses of ethanol. *Psychiatry Research*, 1(3):307–312.
- Harnad, S. (1987). *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, Cambridge.
- Hary, J. M. and Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Perception & Psychophysics*, 32(5):409–418.
- Hashimoto, I. (1982). Auditory evoked potentials from the human midbrain: slow brain stem responses. *Electroencephalography and clinical neurophysiology*, 53(6):652–657.
- Hawkins, D. M. and Wixley, R. A. J. (1986). A note on the transformation of chi-squared variables to normality. *Am. Stat.*, 40(4):296.
- Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, 8(5):393.
- Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, 108(6):3013–3022.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, 182(4108):177–180.
- Hillyard, S. A., Picton, T. W., and Regan, D. (1978). Sensation, perception and attention: Analysis using erps. In *Event-related brain potentials in man*, pages 223–321. Elsevier.
- Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, 16(4):305–312.
- Holt, L. L. and Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, 72(5):1218–1227.

- Holt, L. L., Lotto, A. J., and Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *The Journal of the Acoustical Society of America*, 116(3):1763–1773.
- Horev, N., Most, T., and Pratt, H. (2007). Categorical perception of speech (VOT) and analogous non-speech (FOT) signals: behavioral and electrophysiological correlates. *Ear and hearing*, 28(1):111–128.
- Hu, L. and Zhang, Z. (2019). *EEG Signal Processing and Feature Extraction*. Springer, Singapore.
- Huang, X., Acero, A., Hon, H.-W., and Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*, volume 1. Prentice hall PTR Upper Saddle River.
- Husain, F. T., Fromm, S. J., Pursley, R. H., Hosey, L. A., Braun, A. R., and Horwitz, B. (2006). Neural bases of categorization of simple speech and nonspeech sounds. *Human brain mapping*, 27(8):636–651.
- INANAGA, K. (1998). Frontal midline theta rhythm and mental activity. *Psychiatry and clinical neurosciences*, 52(6):555–566.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1):B47–B57.
- Jacobson, G. P., Newman, C., Privitera, M., and Grayson, A. (1991). Differences in superficial and deep source contributions to middle latency auditory evoked potential pa component in normal subjects and patients with neurologic disease. *Journal of the American Academy of Audiology*, 2(1):7–17.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer, New York.
- Jasper, H. H. (1958). The ten twenty system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–375.
- Jensen, O., Spaak, E., and Zumer, J. M. (2019). Human brain oscillations: from physiological mechanisms to analysis and cognition. *Magnetoencephalography: From signals to dynamic cortical networks*, pages 471–517.
- Jewett, D. L. and Williston, J. S. (1971). Auditory-evoked far fields averaged from the scalp of humans. *Brain*, 94(4):681–696.
- Kandel, E., Schwartz, J., Jessell, T., Siegelbaum, S., and Hudspeth, A. (2012). *Principles of neural science*, volume 5. McGraw-hill.

- Kaukoranta, E., Hari, R., and Lounasmaa, O. (1987). Responses of the human auditory cortex to vowel onset after fricative consonants. *Experimental Brain Research*, 69(1):19–23.
- Kaushik, G., Sinha, H., and Dewan, L. (2014). Biomedical signals analysis by dwt signal denoising with neural networks. *Journal of Theoretical & Applied Information Technology*, 62(1).
- Kawahara, H. (2005). *Audioty morphing using minimum STRAIGHT*. [https://github.com/HidekiKawahara/legacy\\_STRAIGHT/blob/master/doc/morphing-WithSTRAIGHTe.pdf](https://github.com/HidekiKawahara/legacy_STRAIGHT/blob/master/doc/morphing-WithSTRAIGHTe.pdf). accessed: 31-10-2020.
- Kawahara, H., Masuda-Katsuse, I., and De Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds. *Speech communication*, 27(3-4):187–207.
- Kawahara, H. and Matsui, H. (2003). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03)*, volume 1, pages I–I. IEEE.
- Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, pages 983–997.
- Key, A. P. (2016). Human auditory processing: Insights from cortical event-related potentials. *AIMS Neuroscience*, 3.
- Kim, C., Lee, S., Jin, I., and Kim, J. (2018). Acoustic features and cortical auditory evoked potentials according to emotional statuses of /u/, /a/, /i/ vowels. *Journal of audiology & otology*, 22(2):80.
- Klein, S. (1999). Estudo do VOT no português brasileiro. Master’s thesis, Universidade Federal de Santa Catarina - UFSC, Florianópolis, SC.
- Knight, R. T., Staines, W. R., Swick, D., and Chao, L. L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta psychologica*, 101(2-3):159–178.
- Korczak, P. A. and Stapells, D. R. (2010). Effects of various articulatory features of speech on cortical event-related potentials and behavioral measures of speech-sound processing. *Ear and hearing*, 31(4):491–504.
- Krause, F. and Lindemann, O. (2014). Expyriment: A python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, 46(2):416–428.
- Krishnan, A. (2002). Human frequency-following responses: Representation of steady-state synthetic vowels. *Hearing Research*, 166(1-2):192–201.

- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26.
- Leaver, A. M. and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *Journal of Neuroscience*, 30(22):7604–7612.
- Lehtonen, J. (1973). Functional differentiation between late components of visual evoked potentials recorded at occiput and vertex: Effect of stimulus interval and contour. *Electroencephalography and clinical neurophysiology*, 35(1):75–82.
- Leite, L. C. R., Francisco, M. d. V., Duarte, S. G., Garcia, C. F. D., and Bizinoto, S. N. (2013). Potencial evocado auditivo de tronco encefálico no prognóstico do coma superficial. *Revista CEFAC*, 15:1032–1039.
- Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.4.6.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, 54(5):358.
- Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36.
- Liégeois-Chauvel, C., De Graaf, J. B., Laguitton, V., and Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral cortex*, 9(5):484–496.
- Liegeois-Chauvel, C., Musolino, A., Badier, J., Marquis, P., and Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(3):204–214.
- Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3):384–422.
- Lofredo-Bonato, M. T. R. (2008). Vozes infantis: a caracterização do contraste de vozeamento das consoantes plosivas no Português Brasileiro na fala de crianças de 3 a 12 anos. *Revista da Sociedade Brasileira de Fonoaudiologia*, 13:304 – 304.
- Luck, S. (2014). *An Introduction to the Event-Related Potential Technique*. A Bradford Book. MIT Press, Massachusetts, USA.



- Luo, H., Husain, F. T., Horwitz, B., and Poeppel, D. (2005). Discrimination and categorization of speech and non-speech sounds in an meg delayed-match-to-sample study. *Neuroimage*, 28(1):59–71.
- Magnuson, J. S., McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of christmas past. *Cognitive Science*, 27(2):285–298.
- Mäkelä, J., Hämäläinen, M., Hari, R., and McEvoy, L. (1994). Whole-head mapping of middle-latency auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(5):414–421.
- Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.
- Manca, A. D., Grimaldi, M., and sul Linguaggio, C. d. R. I. (2013). Perception, production, articulation and imagery articulation of italian vowels: an erp study. In *Proceedings of the 9th Conference of Multimodalità e Multilinguallità: La sfida più avanzata della comunicazione orale*, pages 213–217.
- Mann, V. A. and Repp, B. H. (1980). Influence of vocalic context on perception of the [j]-[s] distinction. *Perception & Psychophysics*, 28(3):213–228.
- Martynova, O., Kirjavainen, J., and Cheour, M. (2003). Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neuroscience Letters*, 340(2):75–78.
- Mast, T. E. and Watson, C. S. (1968). Attention and auditory evoked responses to low-detectability signals. *Perception & Psychophysics*, 4(4):237–240.
- MATLAB (2014). *version 8.3.0.532 (R2014a)*. The MathWorks Inc., Natick, Massachusetts.
- Mazzoni, D. and Dannenberg, R. (2000). Audacity [software]. *The Audacity Team, Pittsburg, PA, USA*.
- McGee, T. and Kraus, N. (1996). Auditory development reflected by middle latency response. *Ear and hearing*, 17(5):419–429.
- McKay, J. L., Welch, T. D., Vidakovic, B., and Ting, L. H. (2013). Statistically significant contrasts between emg waveforms revealed using wavelet-based functional anova. *Journal of neurophysiology*, 109(2):591–602.
- Melges, D. B. (2013). *Processamento de Sinais Biomédicos*. Material de Disciplina, Universidade Federal de Minas Gerais - UFMG.
- Melo, R. M. et al. (2011). *Caracterização acústica do contraste de sonoridade das consoantes plosivas*. Master's thesis.

- Miller, K. (1964). *Multidimensional Gaussian Distributions*. SIAM series in applied mathematics. Wiley, New York.
- Mirman, D., Holt, L. L., and McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *The Journal of the Acoustical Society of America*, 116(2):1198–1207.
- Moerel, M., De Martino, F., and Formisano, E. (2012). Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *Journal of Neuroscience*, 32(41):14205–14216.
- Molenberghs, G. and Verbeke, G. (2000). *Linear mixed models for longitudinal data*. Springer.
- Morillon, B., Lehongre, K., Frackowiak, R. S., Ducorps, A., Kleinschmidt, A., Poeppel, D., and Giraud, A.-L. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences*, 107(43):18688–18693.
- Möttönen, R., van de Ven, G. M., and Watkins, K. E. (2014). Attention fine-tunes auditory–motor processing of speech sounds. *Journal of Neuroscience*, 34(11):4064–4069.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Myers, E. B. (2014). Emergence of category-level sensitivities in non-native speech sound learning. *Frontiers in neuroscience*, 8:238.
- Myers, E. B., Blumstein, S. E., Walsh, E., and Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, 20(7):895–903.
- Näätänen, R. (1975). Selective attention and evoked potentials in humans—a critical review. *Biological Psychology*, 2(4):237–307.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and brain sciences*, 13(2):201–233.
- Näätänen, R., Kujala, T., and Winkler, I. (2011). Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, 48(1):4–22.
- Näätänen, R. and Michie, P. T. (1979). Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biological psychology*, 8(2):81–136.
- Naatanen, R., Näätänen, R., et al. (1992). *Attention and brain function*. Psychology Press.
- Näätänen, R. and Picton, T. (1987). The n1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4):375–425.

- Nelson, M., Hall, J., and Jacobson, G. (1997). Factors influencing the auditory middle latency response pb component (pi). *Journal of the American Academy of Audiology*, 8:89–99.
- Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *The Journal of the Acoustical Society of America*, 115(4):1777–1790.
- Nomenclature, S. E. P. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of clinical Neurophysiology*, 8(2):200–2.
- Obleser, J., Eisner, F., and Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of neuroscience*, 28(32):8116–8123.
- Obleser, J., Elbert, T., Lahiri, A., and Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research*, 15(3):207–213.
- Obleser, J., Scott, S. K., and Eulitz, C. (2005). Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. *Cerebral Cortex*, 16(8):1069–1076.
- Ohl, F. W. and Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction. *Proceedings of the National Academy of Sciences*, 94(17):9440–9444.
- Okita, T. (1979). Event-related potentials and selective attention to auditory stimuli varying in pitch and localization. *Biological Psychology*, 9(4):271–284.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113.
- Onishi, S. and Davis, H. (1968). Effects of duration and rise time of tone bursts on evoked v potentials. *The Journal of the Acoustical Society of America*, 44(2):582–591.
- Ostroff, J. M., Martin, B. A., and Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear and hearing*, 19(4):290–297.
- Ott, H. W. (1976). *Noise reduction techniques in electronic systems*. John Wiley & Sons, USA.
- Palmeri, T. J. and Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5(4):291.
- Park, H., Ince, R. A., Schyns, P. G., Thut, G., and Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, 25(12):1649–1653.

- Pereira, D. R., Cardoso, S., Ferreira-Santos, F., Fernandes, C., Cunha-Reis, C., Paiva, T. O., Almeida, P. R., Silveira, C., Barbosa, F., and Marques-Teixeira, J. (2014). Effects of inter-stimulus interval (isi) duration on the n1 and p2 components of the auditory event-related potential. *International journal of psychophysiology*, 94(3):311–318.
- Picton, T. (2013). Hearing in time: evoked potential studies of temporal processing. *Ear and hearing*, 34(4):385–401.
- Picton, T., Alain, C., Woods, D. L., John, M., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., and Trujillo, N. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neurotology*, 4(2):64–79.
- Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., Miller, G., Ritter, W., Ruchkin, D., Rugg, M., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, 37(2):127–152.
- Picton, T. W., Woods, D. L., Baribeau-Braun, J., and Healey, T. M. (1977). Evoked potential audiometry. *J Otolaryngol*, 6(2):90–119.
- Picton, T. W., Woods, D. L., and Proulx, G. (1978). Human auditory sustained potentials. ii. stimulus relationships. *Electroencephalography and clinical Neurophysiology*, 45(2):198–210.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & psychophysics*, 13(2):253–260.
- Pisoni, D. B. and Remez, R. E. (2005). *The handbook of speech perception*. Wiley Online Library.
- Polyakov, A. and Pratt, H. (1994). Three-channel lissajous' trajectory of human middle latency auditory evoked potentials. *Ear and hearing*, 15(5):390–399.
- Ponton, C., Eggermont, J., Don, M., Waring, M., Kwong, B., Cunningham, J., and Trautwein, P. (2000). Maturation of the mismatch negativity: effects of profound deafness and cochlear implant use. *Audiology and Neurotology*, 5(3-4):167–185.
- Ptok, M., Blachnik, P., and Schönweiler, R. (2004). Late auditory potentials (nc-erp) in children with symptoms of auditory processing and perception disorder. with and without attention deficit disorder. *HNO*, 52(1):67–75.
- R Core Team (2014). R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2013.
- Rao, A., Zhang, Y., and Miller, S. (2010). Selective listening of concurrent auditory stimuli: an event-related potential study. *Hearing research*, 268(1-2):123–132.

- Roach, B. J. and Mathalon, D. H. (2008). Event-related eeg time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia bulletin*, 34(5):907–926.
- Ross, B., Jamali, S., and Tremblay, K. L. (2013). Plasticity in neuromagnetic cortical responses suggests enhanced auditory object representation. *BMC neuroscience*, 14(1):151.
- Ross, B. and Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing research*, 248(1-2):48–59.
- Roth, W., Krainz, P., Ford, J., Tinklenberg, J., Rothbart, R., and Kopell, B. (1976). Parameters of temporal recovery of the human auditory evoked potential. *Electroencephalography and clinical neurophysiology*, 40(6):623–632.
- Rothman, H. H. (1970). Effects of high frequencies and intersubject variability on the auditory-evoked cortical response. *The Journal of the Acoustical Society of America*, 47(2B):569–573.
- Rothman, H. H., Davis, H., and Hay, I. S. (1970). Slow evoked cortical potentials and temporal features of stimulation. *Electroencephalography and clinical neurophysiology*, 29(3):225–232.
- Rufener, K. S., Krauel, K., Meyer, M., Heinze, H.-J., and Zaehle, T. (2019). Transcranial electrical stimulation improves phoneme processing in developmental dyslexia. *Brain stimulation*.
- Ruhm, H. and Jansen, J. (1969). Rate of stimulus change and evoked response. 1. signal rise-time. *Journal of Auditory Research*, 9(3):211–216.
- Sanei, S. and Chambers, J. A. (2007). *EEG signal processing*. Wiley Online Library.
- Santana, A. C., Barbosa, A. V., Hani C, Y., and Laboissière, R. (2020). A dimension reduction technique applied to regression on high dimension, low sample size neurophysiological data sets. *BMC Neuroscience*. (in press).
- Savers, B. M., Beagley, H., and Henshall, W. (1974). The mechanism of auditory evoked eeg responses. *Nature*, 247(5441):481–483.
- Schabus, M. (2001). *Cognitive electrophysiology and attention*. University of Salzburg; online; accessed 13-12-2020.
- Schalk, G. and Mellinger, J. (2010). *A practical guide to brain-computer interfacing with BCI2000: General-purpose software for brain-computer interface research, data acquisition, stimulus presentation, and brain monitoring*. Springer London Dordrecht Heidelberg New York.
- Scharenborg, O., Koemans, J., Smith, C., Hasegawa-Johnson, M., and Federmeier, K. D. (2019). The neural correlates underlying lexically-guided perceptual learning. *Proc. Interspeech 2019*, pages 1223–1227.

- Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Alagade, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., and Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11(9):1141–1152.
- Schröger, E., Marzecová, A., and SanMiguel, I. (2015). Attention and prediction in human audition: a lesson from cognitive psychophysiology. *European Journal of Neuroscience*, 41(5):641–664.
- Shahin, A. J., Picton, T. W., and Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain and cognition*, 70(3):259–266.
- Sharma, A., Kraus, N., McGee, T. J., and Nicol, T. G. (1997). Developmental changes in p1 and n1 central auditory responses elicited by consonant-vowel syllables. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 104(6):540–545.
- Sharma, A., Marsh, C. M., and Dorman, M. F. (2000). Relationship between n1 evoked potential morphology and the perception of voicing. *The Journal of the Acoustical Society of America*, 108(6):3030–3035.
- Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., and Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. *Cognitive Brain Research*, 21(3):342–350.
- Shoker, L., Sanei, S., and Chambers, J. (2005). Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm. *IEEE Signal Processing Letters*, 12(10):721–724.
- Siddiqi, S. S., Gupta, R., Aslam, M., Abrar Hasan, S., Ahmad Khan, S., and Gandhi, R. (2013). Type-2 diabetes mellitus and auditory brainstem response. *Indian J Endocrinol Metab*, 17(6):1073–1077.
- Silva, D. M., Rothe-Neves, R., and Melges, D. B. (2020). Long-latency event-related responses to vowels: N1-p2 decomposition by two-step principal component analysis. *International Journal of Psychophysiology*, 148:93–102.
- Silva, T., Seara, I., Silva, A., Rauber, A., and Cantoni, M. (2019). *Fonética Acústica: OS SONS DO PORTUGUÊS BRASILEIRO*. CONTEXTO.
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization paths for cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(5):1–13.
- Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., and Papanicolaou, A. C. (1998). Meg correlates of categorical perception of a voice onset time continuum in humans. *Cognitive Brain Research*, 7(2):215–219.

- Simos, P. G., Molfese, D. L., and Brenden, R. A. (1997). Behavioral and electrophysiological indices of voicing-cue discrimination: Laterality patterns and development. *Brain and language*, 57(1):122–150.
- Skinner, P. H. and Jones, H. C. (1968). Effects of signal duration and rise time on the auditory evoked potential. *Journal of Speech and Hearing Research*, 11(2):301–306.
- Skipper, J. I., Devlin, J. T., and Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and language*, 164:77–105.
- Skoe, E. and Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and hearing*, 31(3):302–324.
- Smith, J. C., Marsh, J. T., and Brown, W. S. (1975). Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology*, 39(5):465–472.
- Society, A. E. (1991). American Electroencephalographic Society Guidelines for Standard Electrode Position Nomenclature. *J. Clin. Neurophysiol.*, 8:200–202.
- Spitzer, H., Desimone, R., and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240(4850):338–340.
- Steinschneider, M., Schroeder, C. E., Arezzo, J. C., and Vaughan, H. G. (1995). Physiologic correlates of the voice onset time boundary in primary auditory cortex (a1) of the awake monkey: temporal response patterns. *Brain and language*, 48(3):326–340.
- Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard III, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of neurophysiology*, 82(5):2346–2357.
- Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E., D.-J. and P. B., D., editors, *Human communication: A unified view*, pages 51–66. McGraw Hill.
- Strait, D. L., Ashley, R., Hornickel, J., and Kraus, N. (2010). Context-dependent encoding of speech in the human auditory brainstem as a marker of musical aptitude. *Baltimore, MD: Association for Research in Otolaryngology*.
- Strauß, A., Kotz, S. A., Scharinger, M., and Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *Neuroimage*, 97:387–395.
- Sturm, I., Lapuschkin, S., Samek, W., and Müller, K.-R. (2016). Interpretable deep neural networks for single-trial eeg classification. *Journal of neuroscience methods*, 274:141–145.

- Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (1999). Killer whale (orcinus orca) hearing: Auditory brainstem response and behavioral audiograms. *The Journal of the Acoustical Society of America*, 106(2):1134–1141.
- Technologies, I. (2012). RHD2000 Series Amplifier Arrays.
- Technologies, I. (2014). RHD2000 Evaluation System.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Tiitinen, H., Sivonen, P., Alku, P., Virtanen, J., and Näätänen, R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Cognitive Brain Research*, 8(3):355–363.
- Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological science*, 21(10):1532–1540.
- Tremblay, K., Friesen, L., Martin, B., and Wright, R. (2003a). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear and hearing*, 24(3):225–232.
- Tremblay, K., Kraus, N., McGee, T., Ponton, C., Otis, B., et al. (2001). Central auditory plasticity: changes in the n1-p2 complex after speech-sound training. *Ear and hearing*, 22(2):79–90.
- Tremblay, K., Ross, B., Inoue, K., McClannahan, K., and Collet, G. (2014). Is the auditory evoked p2 response a biomarker of learning? *Frontiers in systems neuroscience*, 8:28.
- Tremblay, K. L. and Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech, Language, and Hearing Research*.
- Tremblay, K. L., Piskosz, M., and Souza, P. (2003b). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology*, 114(7):1332–1343.
- Tsunada, J. and Cohen, Y. E. (2014). Neural mechanisms of auditory categorization: from across brain areas to within local microcircuits. *Frontiers in Neuroscience*, 8:161.
- Tu, Y., Hung, Y. S., Hu, L., Huang, G., Hu, Y., and Zhang, Z. (2014). An automated and fast approach to detect single-trial visual evoked potentials with application to brain–computer interface. *Clinical Neurophysiology*, 125(12):2372–2383.
- Turbes, C. (1996). Neurons’ and glia role in electroencephalogram–evoked potential (eeg-ep) dynamics. *Biomedical sciences instrumentation*, 32:107.
- Umbricht, D., Vyssotky, D., Latanov, A., Nitsch, R., Brambilla, R., D’Adamo, P., and Lipp, H.-P. (2004). Midlatency auditory event-related potentials in mice: comparison to midlatency auditory erps in humans. *Brain research*, 1019(1-2):189–200.



- Unser, M. and Aldroubi, A. (1996). A review of wavelets in biomedical applications. *Proceedings of the IEEE*, 84(4):626–638.
- Verkindt, C., Bertrand, O., Perrin, F., Echallier, J.-F., and Pernier, J. (1995). Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 96(2):143–156.
- Viswanathan, V., Bharadwaj, H. M., and Shinn-Cunningham, B. G. (2019). Electroencephalographic signatures of the neural representation of speech during selective attention. *Eneuro*, 6(5).
- Vivace (2019). Sobre loops de terra. online.
- Wagner, M., Shafer, V. L., Martin, B., and Steinschneider, M. (2013). The effect of native-language experience on the sensory-obligatory components, the p1–n1–p2 and the t-complex. *Brain research*, 1522:31–37.
- Weisz, N., Hartmann, T., Müller, N., and Obleser, J. (2011). Alpha rhythms in audition: cognitive and clinical perspectives. *Frontiers in psychology*, 2:73.
- Williams, H. L., Tepas, D. I., and Morlock, H. C. (1962). Evoked responses to clicks and electroencephalographic stages of sleep in man. *Science*, 138(3541):685–686.
- Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., and Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, 90(18):8722–8726.
- Zatorre, R. J. and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral cortex*, 11(10):946–953.
- Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in cognitive sciences*, 6(1):37–46.
- Zhang, H., Blackburn, T., Phung, B., and Sen, D. (2007). A novel wavelet transform technique for on-line partial discharge measurements. 1. wt de-noising algorithm. *IEEE transactions on dielectrics and electrical insulation*, 14(1):3–14.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

# APPENDIX A

## Middle and short latency evoked responses

This section presents the details about short latency evoked responses, represented by the Auditory Brainstem Response (ABR), and also about Middle-Latency Evoked Responses (AMLR).

### A.1 Auditory brainstem response - ABR

The Auditory Brainstem Response (ABR) is a short latency auditory evoked response, from structures from the human auditory system in the ascending direction of this system; making possible identify disorders in those structures only by the analysis of the obtained ABR. The ABR anatomy is quite complex, once multiple structures in the auditory pathway contribute to compound the ABR waveform. The waveforms that compose the ABR are denoted by roman numbers being the main ones (those analyzed in clinic inspection) those going from I to V. It is possible to find some researches that represent the ABR up until the wave VII. How bigger the wave latency (their roman number), more distant from the cochlea is the structure of the auditory pathway that generated it.

Figure 131 illustrates some auditory pathway structures related to their corresponding ABR waveform and latency. Each ABR component latency vary from person to person but, in average, they have the values illustrated.

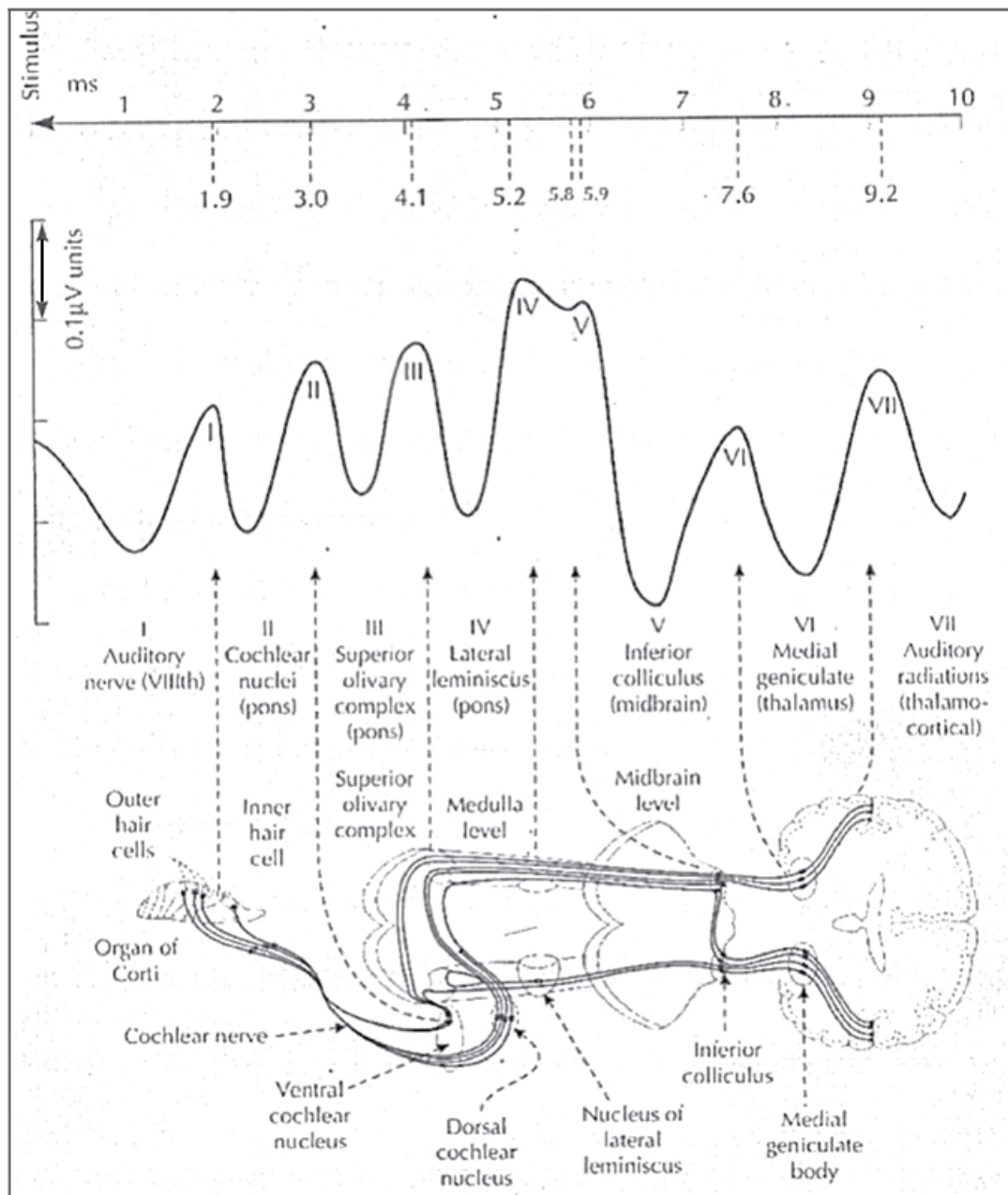


Figure 131 – Relation between ABR waveform latency and the structures involved in its origin.  
 CREDITS: (Siddiqi et al., 2013) (adapted)

The ABR waveforms latency also vary according with the stimuli intensity. As demonstrated by Hall (2007), if the stimuli intensity decrease, this increases the ABR components latency and at the same time decreases the response amplitude in a way that the component V is the only one visible with lower sound intensities. With 80 dB nHL all components can be visualized considering a participant with hearing inside the normal limits. The stimulation frequency leads to changes in the ABR component shape as demonstrated also in Hall (2007). Presentation frequencies in the order of 11,1 or 21,1 stimuli by second provide ABR signals easier to interpret. The fractional values for stimulation frequencies are chosen in a way that the division of the power grid frequency (60Hz in Brazil) by this frequency do not result in an integer, otherwise

the power grid artifact filtering will filter out also the intended signal. Figure 132 illustrates the average latencies band for each ABR component.

Waves	Suggested correspondent	Latency in adults (ms)
I	Part of the 8th nerve distal to the brainstem	1,5 a 1,9
II	Part of the 8th nerve proximal to the brainstem	2,5 a 3,0
III	Cochlear nucleus	2,5 a 4,1
IV	Superior olivar complex	4,3 a 5,2
V	Lateral lemniscus	5,0 a 5,9
VI	Inferior colliculus	
VII	Medial geniculate body	
INTER PEAKS	I – III	2,14
	III – V	1,89
	I – V	4,02

Figure 132 – Latencies band for each ABR component.  
 CREDITS: [Leite et al. \(2013\)](#) apud [Hall \(2007\)](#) (adapted)

A possible electrode placing for ABR acquisition consists in using the nasion as ground (GND), the mastoid as the electrode for the inverting input of the amplifier and the vertex for the electrode of the non-inverting input ([ASHA \(1987\)](#) apud [Jewett and Williston \(1971\)](#)).

## A.2 Auditory middle-latency response - AMLR

We listen with our brain, not with our ears. Thus ABR is a technique that attests to the functioning of subcortical structures, not reaching the cortex where the signal is effectively “heard”. The Auditory Middle-Latency Response (AMLR) would represent structure responses at the thalamus and cortex level.

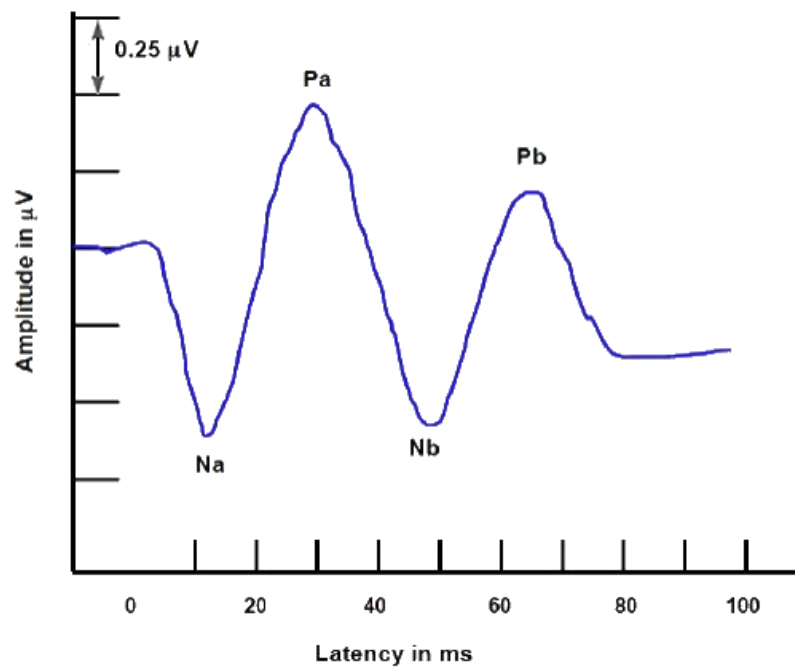


Figure 133 – AMLR waveform showing major peaks at typical latencies including Na, Pa, Nb, and Pb.  
CREDITS: Hall III (2015) (adapted)

The first component of the AMLR is called Na and can be seen in Figure 133. It occurs 10 to 15 ms after stimulation and reflects the activity of the medial geniculate body of the thalamus, portions of the lower colliculus and also from the Heschl's gyrus [Hall (2007) apud Mäkelä et al. (1994), Liegeois-Chauvel et al. (1994), Hashimoto (1982)]. The Pa component is a peak that can occur between 15 and 40 ms after the stimulus. It reflects activity of thalamic structures and from the primary auditory cortex [Hall (2007) apud McGee and Kraus (1996), Liegeois-Chauvel et al. (1994), Polyakov and Pratt (1994), Jacobson et al. (1991)].

The Pb component occurs between 50 and 60 ms after the stimulus and may also be called P50. It is also the wave called P1 in the AELR. This component would reflect the activity of the primary auditory cortex (Hall, 2007). In the two stimuli paradigm, with the second stimulus occurring within 500ms after the first, the P50 wave is attenuated on the second stimulus showing a habituating effect. In the oddball paradigm, where a different stimulus plays after a set of equal stimuli, P50 on the different stimulus may be equal to or greater than that in response to the first stimulus [Hall (2007) apud Boutros et al. (1995), Boutros and Belger (1999)]. Thus the P50 indicates a *sensory gating* because if the second stimulus is the same, the P50 is reduced showing that there was no detection of a “novelty” in the presented content, i.e., there is a higher level auditory processing here.

Intensity, duration, tone frequency and stimulus presentation rate directly affect AMLR behavior. AMLR amplitude and latency increase with increasing ramp up / down time of the tones. Latency

is lower for higher frequency tones, and in this case there is also a decrease in amplitude. But amplitude and latency also tend to decrease with increasing intensity (Hall, 2007). The Pb component, on the other hand, has an increase in amplitude as the tone duration increases, but an increase in the stimuli presentation rate decreases its amplitude (Nelson et al., 1997).

The most commonly used types of stimuli for AMLR are clicks and tone bursts, but speech can also be used, however, as can be seen in Figure 134 the use of low frequency tone bursts is better to highlight key components from AMLR. In addition, the stimulus presentation rate also influences the visibility of AMLR components. The latency of the Pa component is short for slow presentation rates (0.5 and 1 stimuli/sec) and its amplitude tends to be stable between 1 and 15 stimuli / sec, however, between 15 and 40 stimuli/sec its latency increases and the amplitude decreases. Furthermore, the Pb component tends to decay in amplitude and to have its latency increased at rates above 1 stimuli/sec (Hall, 2007).

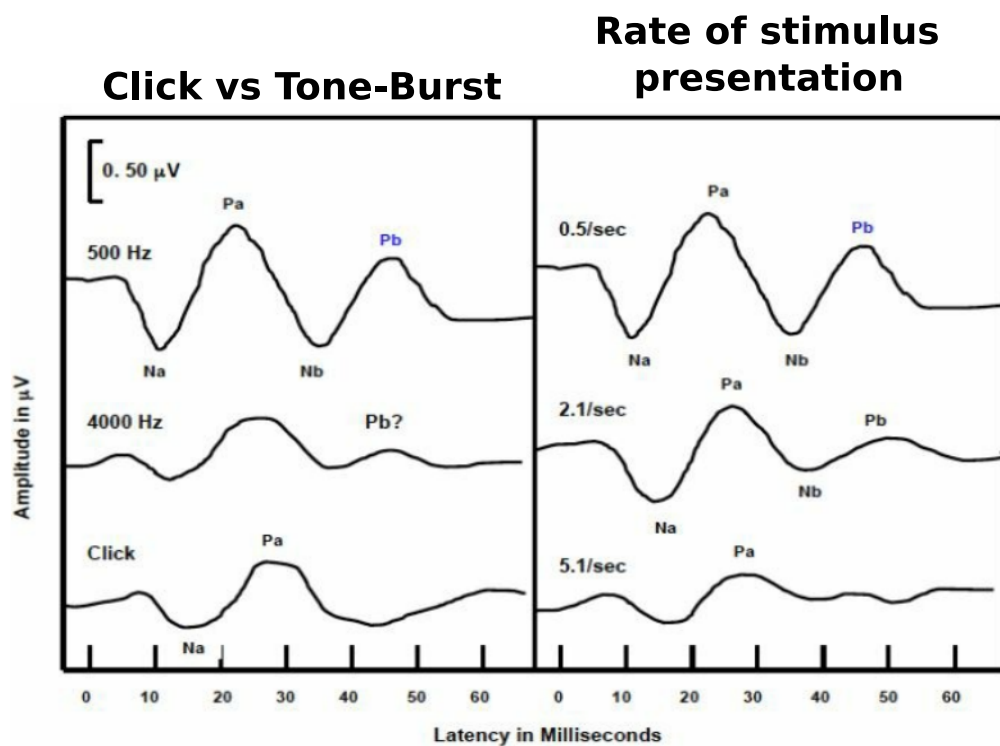


Figure 134 – Effect of stimulus type and rate of presentation at the AMLR waveform.  
CREDITS: Hall (2007) (adapted)

Acquisition of AMLR with EEG can be done with only four electrodes. A possible configuration in the international 10-20 system is to position the GND at Fpz point, the non-inverting electrode at Cz and the inverting electrodes at A1 and A2. It is suggested to connect the inverter electrodes to attenuate the response of the ipsilateral electrode to the stimulated ear, considering the use of AMLR for hearing tests in which the stimulation is monaural. However, in binaural stimulation

with investigations involving the study of evoked response laterality, this interconnection would not be adequate.

To perform the average, the SNR of the acquisition system must be analyzed. Considering high intensities ( $> 70$  dB nHL) and a person with normal hearing, about 100 repetitions of the stimulus are sufficient to obtain a good coherent average. The recommended analysis time for each average ranges from  $-10ms$  to  $100ms$ . Thus, the component Pb which is the last one occurring would be captured.

Stimulus polarity does not matter in AMLR since such responses are expected to be equal whatever the polarity because they reflect a higher order response in the auditory pathway. However, an artifact related to the postauricular muscles can be observed approximately 13 to 15ms after the stimulus and can be attenuated by reversing the polarity of half of the stimuli. For filtering of the acquired signal, it is recommended that it be made between 0.1 – 200 Hz. A notch filter to remove power grid interference (60Hz), for example, is not recommended because it is possible to remove important components around the frequency of 40Hz

AMLR is sensitive to several variables such as age, gender, sleep status, drug and anesthesia use, body temperature, neurological diseases such as Alzheimer's, Parkinson's, epilepsy, brain tumors, multiple sclerosis and schizophrenia, and also factors such as autism and down syndrome. Thus, with the advance of technology for ABR measurement in the 1970s, AMLR became less widely used to measure people's hearing thresholds. However, its sensitivity to the above factors allows its use for several other purposes related to them. In addition, AMLR has advantages over ABR for the possibility of using longer and more complex tones and sounds in its generation as well as the study of factors related to the thalamus and primary auditory cortex.

# APPENDIX B

## Signal filtering tests

According with what was shown in Section 5, it was used a filtering approach that use the **DWT**. Next, it will be shown some tests performed before this approach was adopted using normal filtering with a butterworth filter of 6th order, the adopted approach with **DWT** and a combination of the **DWT** approach with the wavelet thresholding technique which will be detailed.

After signals from all channels were filtered with a 60 Hz notch filter, they were averaged and referenced to the Cz electrode. After that, the signal from each channel was baseline corrected using the 150 ms baseline period. After that, three different tests were performed:

1. Filtering with butterworth filter;
2. Wavelet filtering by removing frequency bands;
3. thresholding and wavelet filtering by removing frequency bands;

For each test, four different cutoff frequencies were adopted: 39, 78.13, 156.25 and 312.5 Hz. Results will be shown for 39 and 156.25 Hz cutoff frequencies. All the signals used in those tests belong to only one participant in the active stage for the **VOT** based continuum. The **SNR** was used to measure the goodness of the filtering aproches compared. The **SNR** was calculated considering the 150 ms prestimulus time as the information for the noisy part. It was computed according the the equation **B.1**. Also, the mean **SNR** of all channels is represented in the same



figure together with the **SNR** for the non-filtered signal.

$$SNR = 20 \log_{10} \left( \frac{\sqrt{\frac{1}{n} \sum_{k=1}^n signal_k^2}}{\sqrt{\frac{1}{n} \sum_{k=1}^n noise_k^2}} \right)^2 \quad (\text{B.1})$$

## B.1 Butterworth filter

In the first test data were filtered with a 6th order butterworth filter and refiltered in the reverse direction to avoid phase distortions. Results for the 17 channels are plotted in Figures 135 and 136 together with the **SNR** for each channel. They are the results for the cutoff frequencies of 39 and 156.25 Hz. For the non-filtered signal the mean **SNR** was 9,47 dB. For the 39 Hz filtering the butterworth filter resulted in a **SNR** of 13.67 dB while for the 156.25 Hz the **SNR** was of 12.06 dB. To improve visualization, it was plotted in red the original signal together with the filtered signal in blue.

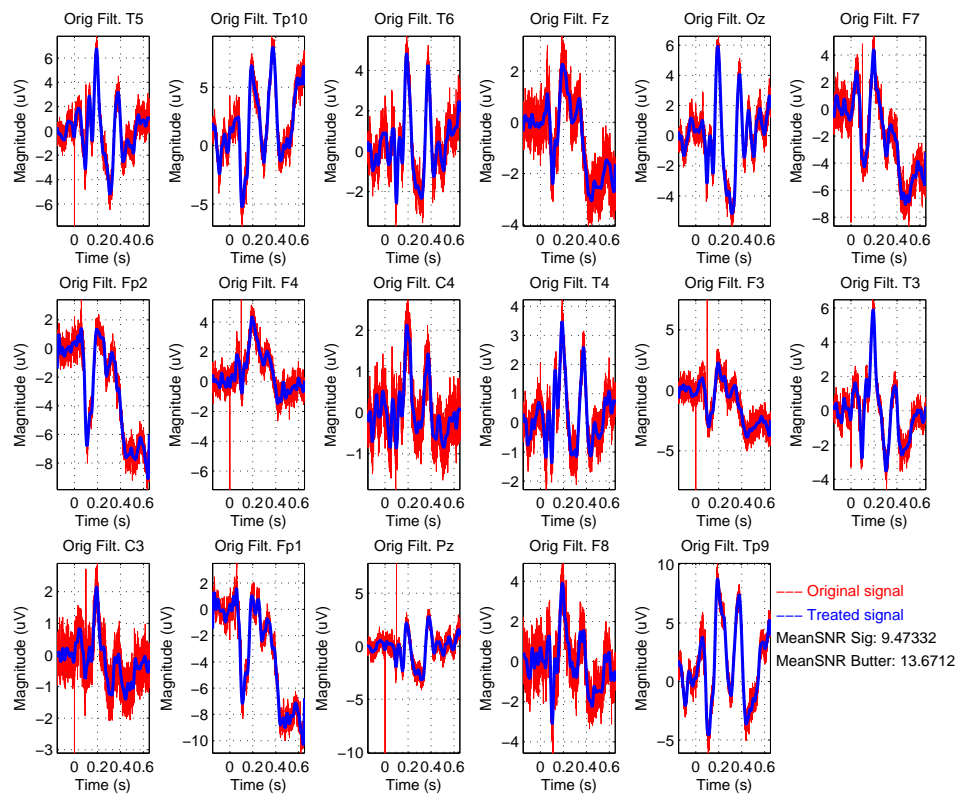


Figure 135 – Signals from each channel, filtered with a 6th order butterworth filter (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 39 Hz.

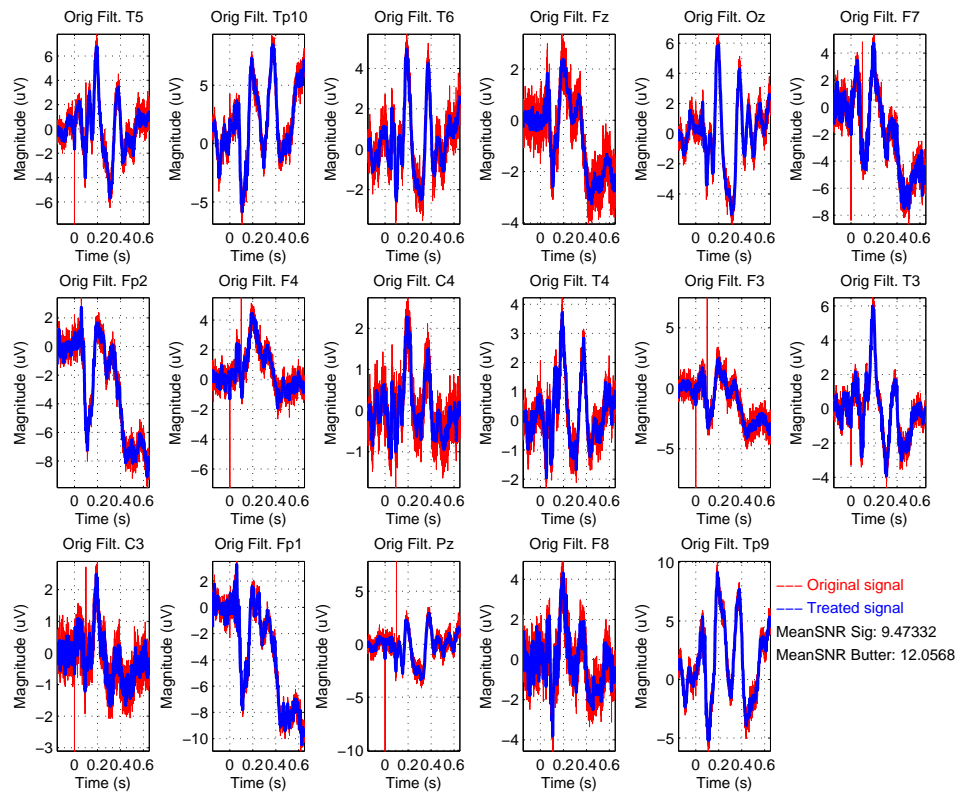


Figure 136 – Signals from each channel, filtered with a 6th order butterworth filter (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz.

## B.2 Wavelet filtering with DWT

The **DWT** wavelet filtering consists in the decomposition of each signal in discrete frequency bands, removal of the frequency bands that should be filtered out of the signal and then the computation of the inverse discrete wavelet transform to obtain the filtered time domain signal. Considering the four cutoff frequencies tested here, 39, 78.13, 156.25 and 312.5 Hz, they corresponds to eliminating all frequencies above the wavelet frequency bands W7, W6, W5 and W4, respectively.

Results for the 17 channels are plotted in Figures 137 and 138 together with the **SNR** for each channel. They are the results for the cutoff frequencies of 39 and 156.25 Hz. For the 39 Hz filtering the wavelet filtering resulted in a **SNR** of 12.87 dB while for the 156.25 Hz the **SNR** was of 12.04 dB.

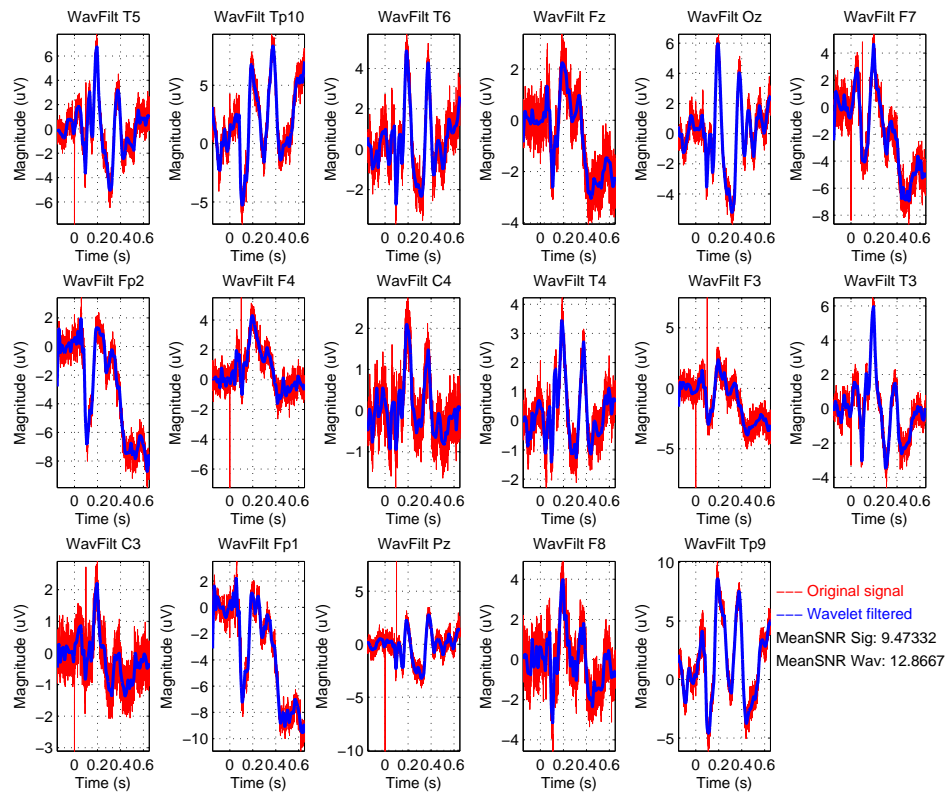


Figure 137 – Signals from each channel, filtered using the DWT band elimination approach (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz.

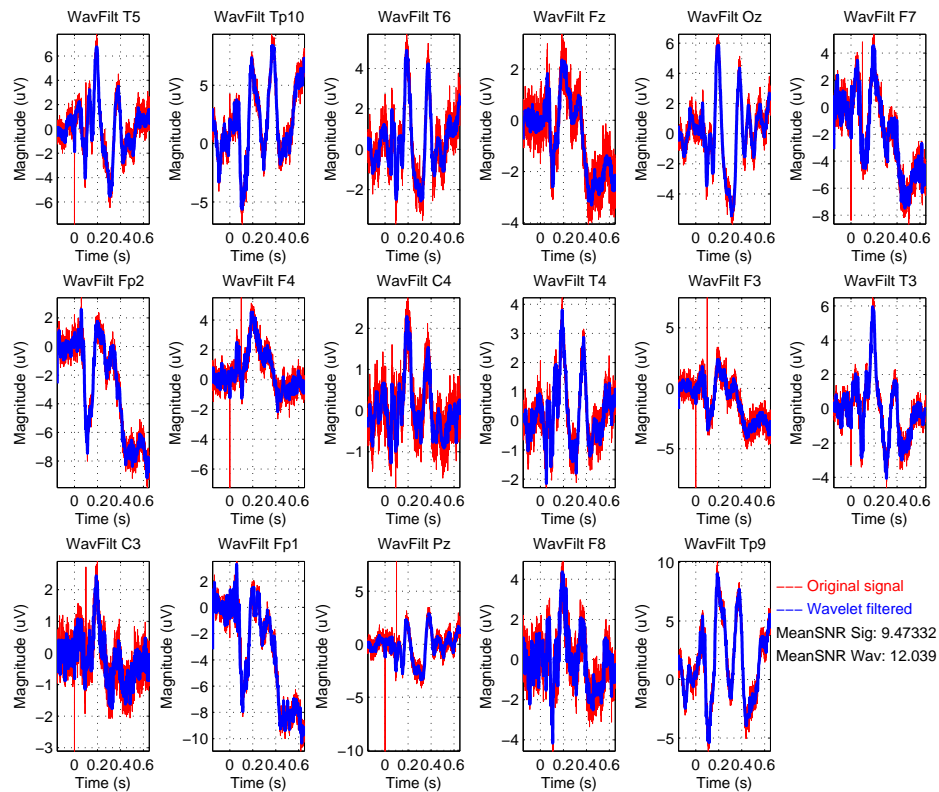


Figure 138 – Signals from each channel, filtered using the DWT band elimination approach (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz.

### B.3 Wavelet thresholding plus band filtering

The filtering involving the combination of the wavelet thresholding with a wavelet filtering through band elimination was also adopted. The first step is to perform the discrete wavelet transform of each trial, then it is applied the wavelet thresholding in each frequency band and then the wavelet filtering according with the explanation of the section 5.1.2.

Wavelet thresholding removes signal noise and can be performed in different ways depending on the choice of the analysis function, thresholding type and threshold value (German-Sallo and Ciufudean, 2012). This technique was first presented by Donoho and Johnstone (1994). In this work we used the universal thresholding T, which is proportional to the noise standard deviation

according with the following equation:

$$T = \sigma \sqrt{2 \ln M} \quad (\text{B.2})$$

where  $M$  is the signal size and  $\sigma$  is the noise variance given by,

$$\sigma = 1.4826 * \text{mad}(W1_{\text{coefs}}) \quad (\text{B.3})$$

where “mad” is the Median Absolute Deviation which is a robust measure of the variability of a univariate sample of quantitative data.  $\sigma$  is obtained using only the highest wavelet band (W1) coefficients where the noise is concentrated (represented as  $W1_{\text{coefs}}$  at the sigma equation). Using the variance approximation by the median is robust to the presence of outliers.

The thresholding can be soft or hard. The soft thresholding is also called wavelet shrinkage, as values for both positive and negative coefficients are being “shrunk” towards zero, in contrary to hard thresholding which either keeps or removes values of coefficients. The soft thresholding results in a smoother signal. The procedure to compute it consists of performing the following condition for each wavelet coefficient (Zhang et al., 2007, Castellanos and Makarov, 2006, Kaushik et al., 2014):

$$x_i = \begin{cases} \frac{x_i}{|x_i|}(|x_i| - T), & |x_i| \geq T \\ 0, & |x_i| < T \end{cases} \quad (\text{B.4})$$

As the soft thresholding can influence the magnitude of the reconstructed signal (Zhang et al., 2007) so one can choose to work with the hard thresholding obtained as follows:

$$x_i = \begin{cases} |x_i|, & |x_i| \geq T \\ 0, & |x_i| < T \end{cases} \quad (\text{B.5})$$

We choose to apply the hard thresholding in our tests.

However, the value obtained for the threshold seems to be too high and leads to the elimination of many wavelet coefficients. A solution to this problem is suggested in the work of Han et al.

(2006) where the threshold is adapted to the DWT level. This is performed by dividing  $T$  by  $\sqrt{j}$ , where  $j$  is the DWT level ( $j=1$  for  $W1$ ,  $j=2$  for  $W2$ , etc).

Results for the 17 channels are plotted in Figures 139 and 140 together with the SNR for each channel. They are the results for the cutoff frequencies of 39 and 156.25 Hz. For the 39 Hz filtering the wavelet filtering resulted in a SNR of 12.87 dB while for the 156.25 Hz the SNR was of 12.04 dB.

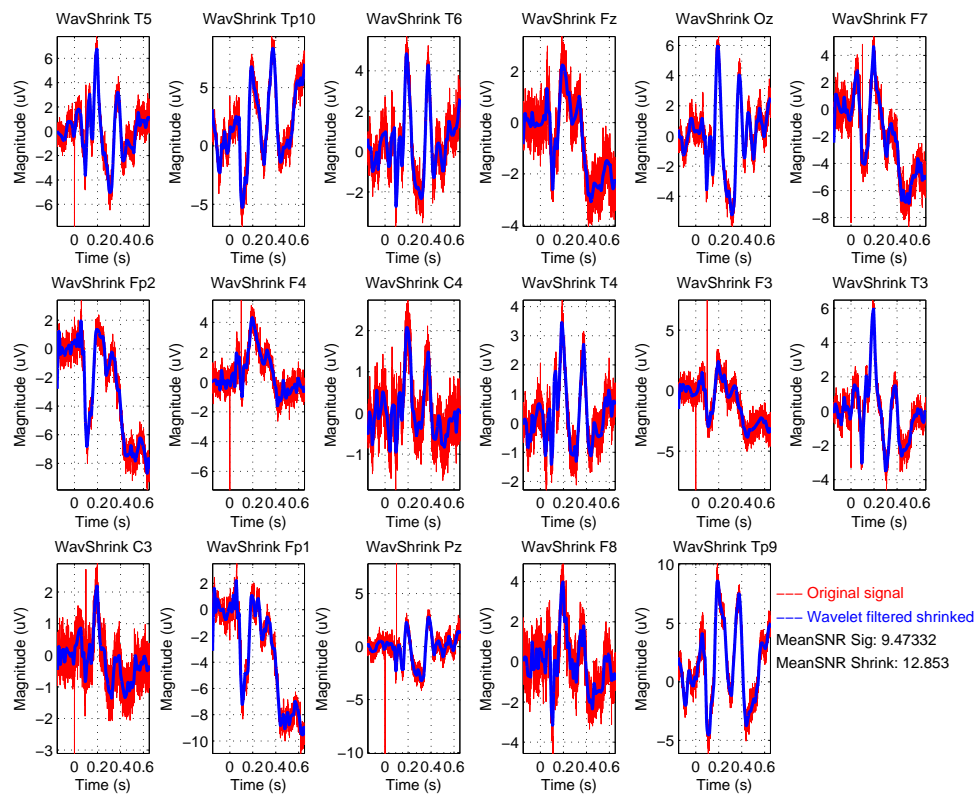


Figure 139 – Signals from each channel, filtered using the DWT band elimination approach combined with the wavelet thresholding technique (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz.

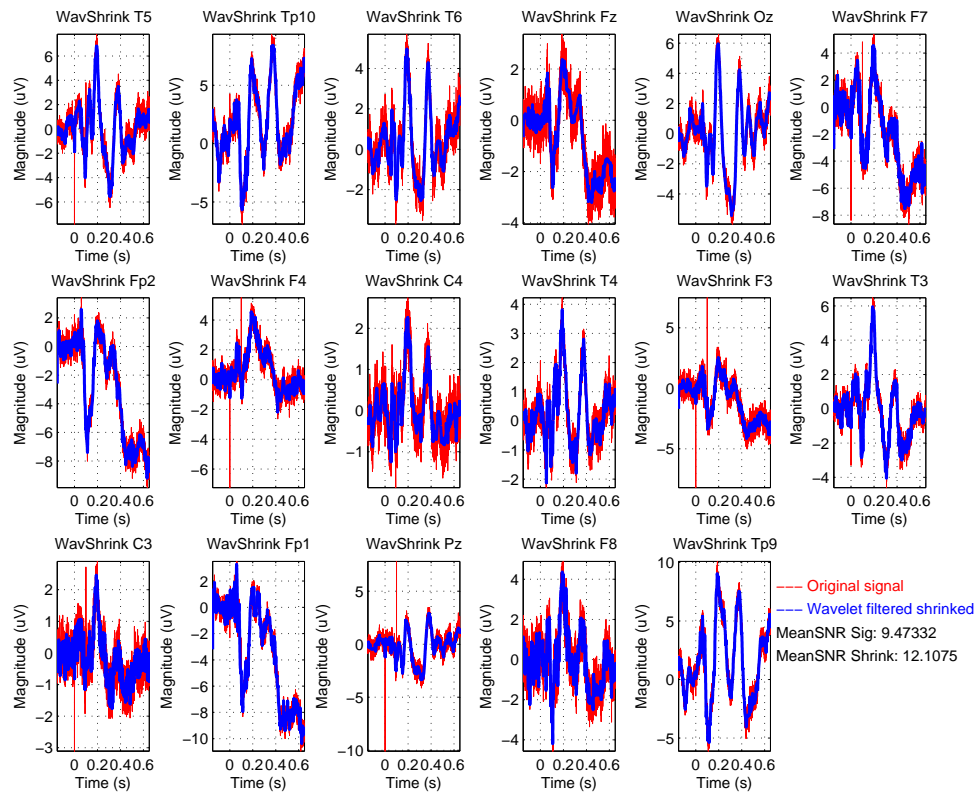


Figure 140 – Signals from each channel, filtered using the DWT band elimination approach combined with the wavelet thresholding technique (in blue) with their SNR values and their original version (in red). Cutoff frequency set to 156 Hz.

## B.4 Final considerations

Figure 141 presents a table with the mean SNR for all tested frequencies and methods. As concluded in 5 section, there is no significant difference between the tested methods and mainly between the wavelet filtering with and without the thresholding. Thus, in the frequency and time domains analysis it was adopted the wavelet filtering without thresholding for simplicity once we are already performing the wavelet transform of the signals for processing. For the frequency domain the cutoff frequency was defined in the W5 band with maximum frequency of 156.25Hz which is enough to include the main brain oscillations that partake in the speech categorical processing and also the F0 of the speaker who provided the stimuli used in this work. For the time domain analysis the cutoff frequency was defined in the W7 band with maximum frequency of 39.06Hz which is enough to allow the computation of the magnitudes and latencies of the main AELR peaks analyzed in this work, namely N1 and P2 waves.



Method/Frequency (Hz)	39,06	78,13	156.25	312.5
No method	9,47	9,47	9,47	9,47
Butterworth	13,67	13,05	12,06	11,49
Wavelet	12,87	12,63	12,04	11,37
Wavelet+thresholding	12,85	12,67	12,11	11,57

Figure 141 – Mean SNR for all tested frequencies and methods for one representative participant in the VOT active experiment.

It can be observed that the thresholding+DWT approach performs better than the other two approach at higher frequencies. This makes sense once the thresholding takes into account the noise in the highest bank of the wavelet decomposition, which tends to concentrate much of the noise in the signal, so that the shrinkage treats the noise in those higher frequencies better.

# APPENDIX C

## Regression techniques

When working with a large data set such as the auditory evoked potentials (AEP) of this work, there is often some problems in its processing and interpretation such as:

- Select from this mass of data the variables that are really relevant to explain the phenomenon.
- Overfitting when there is many or more predictors ( $p$ ) than observations/trials ( $n$ ).
- Correlated predictors.

To solve these problems and increase the computational efficiency in the data processing, it is worth to use a technique that reduces the dimensionality of the problem and selects the really relevant variables (AEPs) that explain the physical and psychophysical responses to the stimuli. Next, some techniques that were tested in this work are described, with a greater detail for the Elastic Net which was the more tested one because of its better responses. All tests and the final solution were performed on the R software ([R Core Team, 2014](#)). Other regression techniques were tested and a short description of them are made in this chapter. They are the Principal Component Regression, Partial Least Squares Regression and Sparse Partial Least Squares.

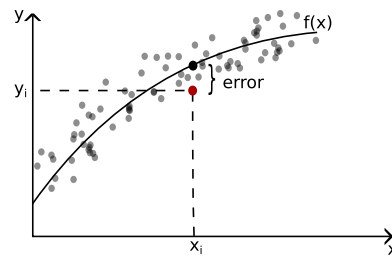


Figure 142 – Linear Regression.

## C.1 Regularization and characteristics selection

In a regression, we seek to find a model described by a function  $f(x)$  (which can be linear resulting in a linear regression) that relates the inputs ( $x_i$ ) with the outputs ( $y_i$ ) with the smallest possible error ( $\varepsilon$ ). The Figure 142 and the equation C.1 describes this idea.

$$y_i = f(x_i) + \varepsilon_i \quad (\text{C.1})$$

If  $f(x)$  was a straight line, the equation C.1 will be:

$$y_i = \omega_0 + \omega_1 x_i + \varepsilon_i \quad (\text{C.2})$$

where the variables  $\omega$  are the coefficients of the linear regression. However, since the regression is not perfect and contains errors, what is obtained is an estimate of the regression coefficients  $\hat{\omega}$ . Two common errors in regressions are bias and variance.

Consider that repeated regressions are performed with different observations (measurements) of the same phenomenon and different  $\hat{\omega}$  are obtained for each one. Performing the mean between the  $\hat{\omega}_i$  of each regression an average set of predictors is obtained. The difference between these predictors and those that would be obtained if it were possible to have a perfect regression is the bias. More complex models (such as those described by high-grade polynomials) tend to follow the perfect model well by reducing the bias.

The overfitting is a problem that occurs in pattern classification techniques characterized by statistical models that fit too well with training data but are bad at modeling new data not previously observed, i.e., they do not have good generalization. One cause of overfitting in linear regression techniques is the use of very complex models. Another cause is the use of a small

amount of observations ( $n$ ) in relation to the number of model variables or predictors ( $p$ ). In this case, an observed effect of overfitting is high values for the regression coefficients ( $\omega$ ).

Variance defines how much a set of regressions varies from the expected (perfect). It tends to be larger in more complex models since they tend to overfitting and thus differences between regressions for different sets of data obtained from observing the same phenomenon is large. In short, less complex models have high variance but less bias and vice versa. Therefore, it is important to find a balanced model that minimizes these two errors, and this is the Ridge Regression's proposal.

### C.1.1 Ridge Regression

At Ridge regression regularization technique, it is defined a cost function (to be minimized) that quantifies the quality of the fit and the magnitude of the regression coefficients. Residual Sum of Squares (RSS) is a function that can be used to mediate the quality of the fit so that the lower its value, the better the fit. The equation C.3 describes RSS in terms of the actual value of the output  $y$  and its estimated value by the model  $\hat{y}(\omega)$ .

$$RSS(\omega) = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (\text{C.3})$$

To measure the magnitude of the coefficients, you can use the sum of the squares of the coefficients given by your L2 Norm:

$$\sum_{j=0}^D \omega_j^2 = \|\omega\|_2^2 \quad (\text{C.4})$$

Thus, the Ridge regression cost function is given according to the equation C.5,

$$Cost(\hat{\omega}) = RSS(\hat{\omega}) + \lambda \|\omega\|_2^2 \quad (\text{C.5})$$

where the term  $\lambda$  It is a tuning parameter that balances the two terms of the cost function.

If  $\lambda = 0$ ,  $Cost = RSS(\hat{\omega})$ , which is the solution given by least squares regression characterized by complex models with low trend and high variance. Said, solution not regularized. If  $\lambda = \infty$ , there is two possibilities:

$$Cost = \infty, \text{ se } \hat{\omega} \neq 0$$

$$Cost = RSS(0), \text{ se } \hat{\omega} \approx 0 \text{ which is the acceptable solution in this case.}$$

In this case, there is a simple model that has low variance and high bias (underfitting).

If  $0 < \lambda < \infty$ , there is the objective of the Ridge regression. The best value of  $\lambda$  can be found through cross-validation techniques such as the k-fold..

### C.1.2 LASSO

In this work, there is the problem of working with more variables ( $p$ ) than observations/trials ( $n$ ). Ridge regression alleviates this problem, but the issue of selecting the variables that really matter for the interpretation of the results remains. One way to deal with this problem is through a penalty regression as in the Least Absolute Shrinkage and Selection Operator (LASSO) method. In Ridge regression, the regression coefficients tend to be reduced but are never zeroed. With LASSO, some  $\omega$  coefficients are zeroed, i.e., features are eliminated from the model.

In a simplistic solution, one might suggest working with a limit on Ridge regression coefficients by zeroing those below this limit. The problem with this solution is that correlated variables are treated independently and could all fall below this limit and be eliminated from the model or be above it and stays. It would be desirable to group these characteristics into one, more relevant to the model, which would be maintained in the model. Thus, the number of variables (characteristics) is reduced and the solution interpretability becomes better since only the characteristics that explain the answer are maintained. The cost function for LASSO is similar to that of Ridge regression only with the difference of the Norm used in the second term of the equation which is the L1 here as can be seen in Eq. C.6.

$$Cost(\hat{\omega}) = RSS(\hat{\omega}) + \lambda \|\hat{\omega}\|_1 \tag{C.6}$$

In this case, the parameter  $\lambda$  regulates the fit of the model and its sparsity.

If  $\lambda = 0$ ,  $Cost = RSS(\hat{\omega})$ , which is the solution given by least squares regression characterized by complex models with low trend and high variance. If  $\lambda = \infty$ ,  $\hat{\omega} = 0$ , where no coefficient is selected. If  $0 < \lambda < \infty$ , we have the LASSO solution where for each value of  $\lambda$  a number of different characteristics are selected for the model while the others are eliminated (zero value for their coefficients  $\hat{\omega}$ ). Once again, the parameter  $\lambda$  can be selected using cross validation techniques. More details can be found at [Murphy \(2012\)](#).

In the case of correlated characteristics, LASSO tends to arbitrarily select those to be eliminated. Thus, some of the better variables in explaining the model may be excluded. Ridge regression has a better solution in model prediction but lacks the sparsity of LASSO. Thus, the Elastic Net method arises to merge the best of these two techniques.

## C.2 Elastic Net

The Elastic Net (EN) regression technique is a penalized least squares regression technique that combines the regularization of the cost function (avoiding overfitting problems) with a feature selection, reducing the model complexity and improving its interpretability (see more in [Zou and Hastie, 2005](#))).

The resulting coefficients of the EN regression can be found by the solution of Equation C.7 where  $\mathbf{y}$  is the response vector,  $\mathbf{X}$  is the predictors matrix and  $\omega$  is the model coefficient vector with  $\hat{\omega}$  being its estimator. The tuning parameters  $\lambda$  and  $\alpha$  controls the regularization level of the model and the sparsity, respectively. The terms  $\|\cdot\|_1$  and  $\|\cdot\|_2$  represents the  $L_1$  and  $L_2$  Norms, respectively. The implementation of this method was performed through the package `glmnet` in R software ([Friedman et al., 2010b](#), [Simon et al., 2011](#)) and the best parameters for the regression model were defined through a k-fold cross-validation algorithm provided by the same package and by testing the regression for different values of alpha.

$$\hat{\omega} = \min_{\omega} \{ \|\mathbf{y} - \mathbf{X}\omega\|_2^2 + \lambda [\alpha \|\omega\|_1 + (1 - \alpha) \|\omega\|_2^2] \} \quad (\text{C.7})$$

The next section describe other three regression techniques tested in this work.

### C.3 Principal Component Regression

The idea behind the principal component regression (PCR) is to perform a principal component analysis (PCA) on the input matrix and then use only the first  $k$  principal components to do the regression in this new space. Thus, PCR reduces the dimensionality and eliminates the correlated variables.

The first step is to perform the PCA and define  $\mathbf{T} \in \mathbb{R}^{n \times k}$  tal que,

$$\mathbf{X} = \mathbf{TP}^T \quad (\text{C.8})$$

where  $\mathbf{P} \in \mathbb{R}^{p \times k}$  is the matrix of the principal components of the input matrix  $\mathbf{X}$ ,  $\mathbf{T} \in \mathbb{R}^{n \times k}$  is the matrix with the  $k$  linear combinations (scores) of  $\mathbf{X}$  where the  $t$  vectors that make up this matrix are unrelated. In possession of the matrix  $\mathbf{T}$  it is possible to obtain the regression coefficients  $\omega$  by doing,

$$\mathbf{Y} = \mathbf{T}\omega + \mathbf{F} \quad (\text{C.9})$$

where  $\mathbf{F}$  is the regression error matrix.

One problem with PCR is that there is no selection of variables relevant to the explanation of the answer, since each principal component (PC) is a linear combination of all variables. Another problem is that the output variables  $y$  are not used in the definition of the PCs so it is not guaranteed that the direction vector found will be the best predictor of the response. To solve this last problem, one option is partial least squares regression.

### C.4 Partial Least Squares Regression

The partial least squares (PLS) regression technique takes into account when decomposing the structure of the data sets of the input variables  $x$  and the responses or output variables  $y$ . The matrices of outputs  $\mathbf{Y}$  and inputs  $\mathbf{X}$  are decomposed into vectors that function as their principal components. A  $\vec{u}$  vector that corresponds to the direction of greatest variation in  $\mathbf{Y}$  is found, and based on it, it is computed the direction vector  $\vec{t}$  in  $\mathbf{X}$  that best explain  $\vec{u}$ . This technique reduces

the dimensionality of the problem while solving the issue of variable collinearity by maximizing the covariance between  $\mathbf{Y}$  and  $\mathbf{X}$ . Be,

$$\begin{aligned}\mathbf{X} &= \mathbf{TP}^T + \mathbf{E}, & \mathbf{X} \in \mathbb{R}^{n \times p} \\ \mathbf{Y} &= \mathbf{UQ}^T + \mathbf{F}, & \mathbf{Y} \in \mathbb{R}^{n \times q}\end{aligned}\tag{C.10}$$

where  $\mathbf{U}$  and  $\mathbf{T} \in \mathbb{R}^{n \times k}$  are the matrices with the  $k$  linear combinations (scores) of  $\mathbf{Y}$  e  $\mathbf{X}$ .  $\mathbf{P} \in \mathbb{R}^{p \times k}$  and  $\mathbf{Q} \in \mathbb{R}^{q \times k}$  are the matrices with the PCs and  $\mathbf{E} \in \mathbb{R}^{n \times p}$  and  $\mathbf{F} \in \mathbb{R}^{n \times q}$  are the matrices of errors. The idea of PLS is that the decomposition of the matrices  $\mathbf{Y}$  and  $\mathbf{X}$  be performed by taking information from each other into account. Once computed the matrices  $\mathbf{T}$ ,  $\mathbf{U}$ ,  $\mathbf{P}$  and  $\mathbf{Q}$ , the regression model relating  $\mathbf{Y}$  and  $\mathbf{X}$  is obtained by calculating,

$$\mathbf{U} = \mathbf{T}\boldsymbol{\omega}\tag{C.11}$$

e substituindo na Eq.C.10 obtendo-se,

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} = \mathbf{T}\boldsymbol{\omega}\mathbf{Q}^T + \mathbf{F} = \mathbf{XP}\boldsymbol{\omega}\mathbf{Q}^T + \mathbf{F}\tag{C.12}$$

Both PCR and PLS do not perform variable selection. In addition, PLS is inconsistent when the number of variables ( $p$ ) is large relative to the number of observations ( $n$ ). To solve these problems, sparse versions of these two algorithms were created. As in PCR the definition of  $\mathbf{X}$ , PCs does not take into account output variables, only the sparse version of PLS was studied.

## C.5 Sparse Partial Least Squares

Sparse PLS (SPLS) includes the sparse feature that was missing from PLS and performs a feature selection by means of a term  $\lambda_1 \|c\|_1$  and additionally a regularization by means of a term  $\lambda_2 \|c\|_2^2$  just like Elastic Net, aiming to improve its performance when  $p > n$ . More details can be found at [Chun and Keleş \(2010\)](#). In this study the authors compare SPLS with some other techniques including EN. They conclude that SPLS outperforms EN in selecting relevant variables when  $n < p$  in case of high noise-to-signal ratio but conclude that EN performs well for the right size of the regularization parameter  $\lambda_2$ . They attribute this superiority to the fact that



each main component used can be composed of more than  $n$  variables and there is no elimination of important ones by limiting the number of variables imposed by the sparsity of the method.

Preliminary tests performed with EEG data from a passive auditory task acquisition showed that SPLS resulted in fewer coefficients than EN. In some conditions and subjects, no coefficients were found.

## C.6 Assembly of the EN input matrix and response vectors

Using EN, only the samples of the AEP that are really relevant to explain the responses (physical and psychophysical) to the stimuli are selected, improving the results interpretability and gaining computational efficiency. The regression coefficients ( $\omega$ ) computes the degree to which the neural measurement (AELR) predicts the behavior (slope of psychometric curve - categorization) or the physical characteristic of the stimuli (VOT or formant frequencies). For each one of the four acquisition conditions (VOT-act, VOT-pass, Form-act and Form-pass) four regression cases were performed: physical left, physical right, psychophysical left and psychophysical right. Depending on the condition, the response vectors and the predictors (input) matrices for the regression of each case was changed as will be explained below.

For the input (predictor) matrices assembling, the general procedure is described as follows. For each subject, after filtering, there is a certain amount of ALRs for each stimulus since many trials were performed for each one in each task. A certain number of trials were grouped for each stimulus in each task, and the stimuli were averaged to obtain 7 averages per stimulus per subject and per task. Here, the stimuli shifted in  $180^\circ$  (see section 4.4) were considered as a separate stimuli, so instead of 5 stimuli, it was considered 10 different stimuli. To help the understanding of this grouping, consider the following example. For a given subject, in the active stage with the VOT continuum, 1000 stimuli were applied. From these, 200 are the  $stim1$  where 100 are normal and 100 had a lag of  $180^\circ$ . Each one of them, at each trial, generated an AELR. During the cleaning of the data, suppose that 30 trials were removed from each case leading to 70 AELRs for the normal  $stim1$  responses and 70 for the  $180^\circ$  lagged responses. Each case was considered as a separate stimulus as, for example,  $stim1$  and  $-stim1$ . In order to obtain 7 groups of averages, the 70 trials of  $stim1$  were averaged in groups of 10 trials and the same was performed for the  $-stim1$  trials providing 14 groups of averages for this  $stim1$ .

This was done to reduce the noise in the AEP that would be passed to the EN. In addition, a certain amount of observations are required for the proper functioning of EN so that one cannot average with all trials per stimulus, as this would reduce each individual's AEP to just 5 instead

of the  $7 \times 5 \times 2$  used. There is a trade-off here between signal quality used in the input matrix and the amount of observations. The quantity 7 was found by trial and using the Akaike information criterion (AIC) to find the number of groups that resulted in the best fit for EN. After grouped and averaged, data for each subject was normalized for the two tasks of each acquisition day.

Thus, the predictors matrix was assembled by arranging the matrix lines so that the groups with the mean AELRs for stim1 of all subjects were aligned, followed by the average AELRs of stim2 and so on until stim5. The final matrix had  $7 \times 2 \times 5 \times 9$  lines (number\_of\_groups x  $180^\circ$ \_lagged\_stimuli\_separate x number\_of\_stimuli x number\_of\_subjects\_temporal\_data) and as many columns as the amount of samples of each trial. It was used 0.35 seconds of each trial (already baseline corrected) so that, with a sampling rate of 5kHz, there were 1750 columns. This matrix was used for the physical regressions. In the k-fold cross-validation it was considered 45 folds ( $5 \times 9$  - number\_of\_stimuli x number\_of\_subjects) grouping at each one the 14 groups of averages of the same stimuli ( $7 \times 2$ ) for the same subject.

For the psychophysical, two variations of the input matrix were used:

1. The same input matrix used for the physical regressions.
2. Three reduced input matrices obtained with relations between stim1, stim2, stim4 and stim5 groups of averages ( $7 \times 2 \times 1 \times 9$ ):
  - relation 1:  $|mean(stim1, stim2) - mean(stim4, stim5)|$
  - relation 2:  $|mean(stim1, stim5) - mean(stim2, stim4)|$
  - relation 3:  $mean(|stim1 - stim2|, |stim4 - stim5|)$

The first assembly was tested but will not be further discussed due to the difficulty in interpret the results according to a variation of the AELR with  $\beta$  and relate it to the hypothesis raised at the introduction of this work. Those last three assemblings were tested in an attempt to obtain the relation between separate stimuli with the categorical effect (through  $\beta$ ) so that it is possible to compare the results with the hypothesis and analyze if is true or not, i.e., if the neuralcorrelates of the categorical perception for stim2 will be similar to those of stim1 and the same for stim4 and stim5.

For relation 1, if its value increases (a distance measurement), this means that stim2 categorical perception is becoming more similar to that of stim1 and the same for stim4 relative to stim5 proving the hypothesis. For relation 2, considering that  $\beta$  is more related to a better or worse discrimination of stimuli; so stim1 will be more similar to stim5 as they are clearer (less ambiguous)

and the stim2 more similar to stim4 as they are more ambiguous. So, if this distance measure decreases, this means that stim2 and stim4 are better categorized and their neural correlates of categorical perception would be more similar to those of stim1 and stim5 proving the hypothesis. Relation 3 has an idea similar to that of the relation 2. If stim2 is similar to stim1 and stim4 to stim5, so the mean will tend to zero, otherwise it will tend to a larger number showing that no categorization occurs for the ambiguous stimuli. Signal (negative or positive) is not important here as this is a measure of distance, so the modulus is used.

For the physical cases the EN response vectors were constructed using physical characteristics of the stimuli, i.e., VOT values or the difference  $F2 - F1$  of the formant frequencies obtained for the 5 stimuli of each subject. It was a single vector where each VOT (or formant differences) of each subject for each stimuli, was repeated as many times as the amount of groups for that stimulus. So it was a  $7 \times 2 \times 5 \times 9$  vector, with the same size of the number of lines of the input matrix. In these two cases, the goal was to see which components of AELR relate to the physical characteristics of VOT and formant frequencies that make up the stimuli.

For the psychophysical cases the response vectors were constructed with the psychometric curve slope values ( $\beta$ ) obtained for each subject in the VOT continuum (used in VOT conditions regressions) and the Formant continuum (used in formant conditions regressions). For the second and third input matrices variations, this vector was constructed with repetitions of the  $\beta$  value as many times as the number of groups of averages of one stimuli for each subject. So it was a  $7 \times 2 \times 9$  vector (number\_of\_groups x  $180^\circ$ \_lagged\_stimuli\_separate x number\_of\_subjects). For the first input matrix, where all stimuli are used, the response vector was similar to the previous one but now repeated 5 times, each one for a stimuli. As the  $\beta$  is related to the subject and not to the stimuli, this vector is really 5 repetitions of the previous one ( $7 \times 2 \times 9 \times 5$ ). Here, the goal was to see which components of AEP relate to the psychophysical characteristic of the categorical perception of the stimuli.

Considering the first assembly for the psychophysical input matrix, each regression in those four cases (physical left and right, psychophysical left and right) were performed 4 times being each for one acquisition condition (VOT-act, VOT-pass, Form-act or Form-pass) totalizing 16 regressions. For the second assembly of the psychophysical input matrix, with the three relations, the total amount of regressions was 32. Those set of regressions were performed twice being one for the temporal electrodes data and another for the frontal electrodes.

A test of this method was performed randomizing the order of the lines of the input matrices to prove that inside noise the EN would not find valid coefficients that can explain the response. In our tests this proved to be true, showing that our results once again, are related with the response vector used in the regression. Correlations between the response vector and the one

predicted by the coefficients were above 75% for the normal regressions and below 30% in the randomized regressions. The randomization tests did not work well for the second and third matrices variations for the psychophysical responses resulting in high correlations (around 60%) in some cases. This occurred due to the small number of subjects of this work so that even randomizing the lines, the probability of the  $\beta$  value from a given subject be aligned with the group average of this same subject was high. In these two matrices variations the number of lines used was  $\frac{1}{5}$  of the complete matrix (the first variation) increasing very much the probability cited before.

For the physical response vector regressions where the complete input matrix was used, the correlations of the randomized cases were absent or below 30% against the above 83% of the normal cases. Figure 143 shows this result for the VOT-act condition in the temporal cortex. With this physical results it was possible to conclude that the EN is an effective technique to find the relevant coefficients from the input that explain the responses and that they will not be found inside noise. Furthermore, this result was found even considering in the case of this work where the number of dimensions is very superior to the number of observations.

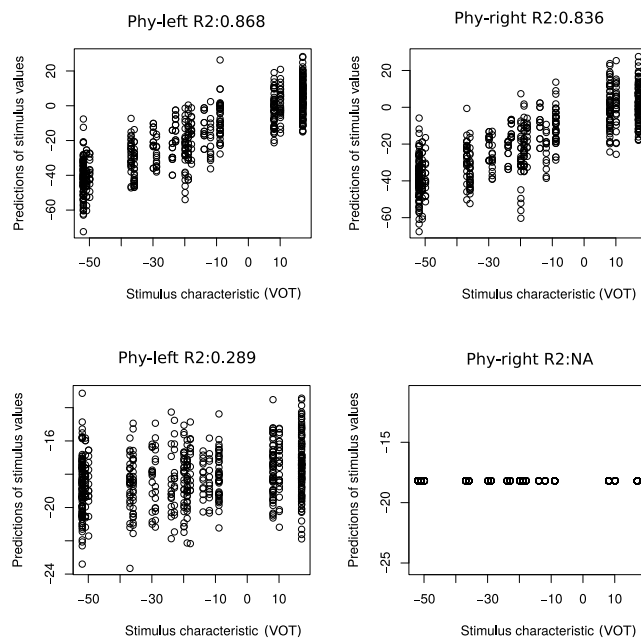


Figure 143 – Correlation of the regression predictions with the physical response for the normal (top) and randomized (bottom) input matrices of the VOT-active condition in the temporal cortex. Responses are separated by side (left or right) and the Pearson's correlation coefficient is showed as the variable R2.

In all regressions performed here it is worth to point out that the amount of characteristics are very high as it is represented by each sample composing the AELR mean used. For the regressions a window of 0.35 seconds of the trials was used. With a sample rate of 5kHz the

amount of characteristics is given by  $0.35 \times 5000 = 1750$ . Considering the assemblings of the input matrices stated before, for the first assembly the amount of observations (lines) is 630 (for the temporal data considering 9 subjects) and for the second this amount decreases to 126. Also, input matrices data and  $\beta$  values were normalized before regression to match the orders of magnitudes of those values.

An additional set of regressions were performed initially before the idea to use  $\beta$  for the response vectors of the psychophysical regressions. In the place of  $\beta$ , was used the sequency of percentages related to the psychometric curve: 0%, 5%, 50%, 95% and 100%. Also, instead of using the VOT or formant differencies from each subject to compose the physical response vectors, the mean of these values for all subjects were used. This resulted in coefficients that modeled very well both physical and psychophysical responses but, specifically for the psychophysical, the interpretation of these coefficients according with the response vector of percentages was not clear and the coefficients found for the phychoophysical regressions were very similar to those found for the physical ones. This occurred possibly due to the ascending sequence observed for both response vectors. So, was necessary to replace this psychophysical response vector by something more related to the auditory categorization neuralcorrelates, which are searched in the AELR data.

In the next section this result is showed as a test for the EN regression on a specific acquisition condition. With the coefficients found by EN for each stimuli and condition, the prediction of the response vectors were computed and projected on the physical and psychophysical axis, only for stim2 and stim4. It was possible to prove the hypothesis of this work in this test as can be seen in the measurement of the distancies between the stim2 and stim4 in the physical and psychophysical projections.

## C.7 Testing Elastic Net Responses

According with the predictors matrix and response verctor assembling, described in the previous section, each subject have 14 groups of averages for each stimulus. So, there was 14 predictions of those groups. The values of predictions for the same stimulus of the same subject were averaged resulting in a single value. So, for each subject, there was 5 projections, one for each stimulus. Those averages were plotted in Figure 144 for all subjects and also the density distributions for the stim2 and stim4. This results belongs to the condition VOT-act.

To quantify how far the distribution of output projections for the psychophysical response to stimuli 2 and 4 is more distant from each other than the distribution of output projections of those same stimuli for the physical response (according to the hypothesis test described in chapter 1),

a paired t-test was performed.

For the analysis of this work, we tested if the distance between the means of stimuli 2 and 4 in the psychophysical axis is greater than that in the physical axis, confirming the categorical perception of these stimuli by the tested individuals in a paired one sided t-test. So the following hypothesis was tested:

$$\begin{cases} H0 : (\mu_4 - \mu_2)_{psy} = (\mu_4 - \mu_2)_{phy} \\ H1 : (\mu_4 - \mu_2)_{psy} > (\mu_4 - \mu_2)_{phy} \end{cases} \quad (C.13)$$

This test was performed separately for the right and left hemispheres. As observed for each individual and also for the average of subjects, the null hypothesis was rejected with a significance level of  $\alpha = 5\%$  in all cases ( $p\text{-value} \ll \alpha$ ), that is, there is not enough evidence to accept the null hypothesis and with a confidence level of 95% it can be stated that, on average, the difference between the projections of stimuli 4 and 2 on the psychophysical axis is greater than that on the physical axis. Figure 144 illustrates this result for the average projections for all individuals tested in the left and right hemisphere at the temporal cortex for the VOT-active condition.

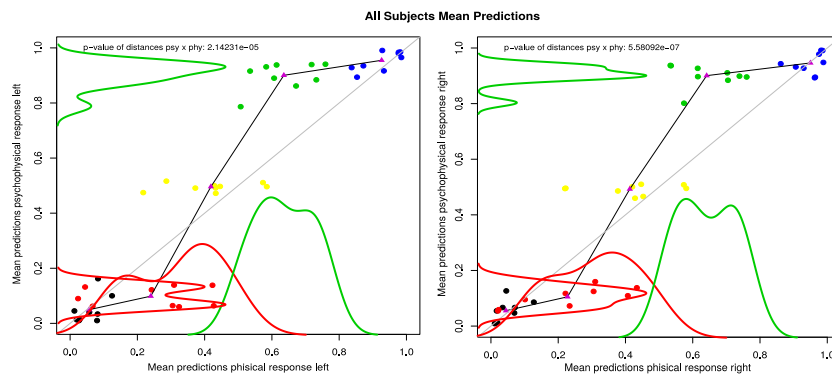


Figure 144 – Average projections of stimuli 4 and 2 on the physical and psychophysical axis of all subjects for the VOT-active condition.

The coefficients for the physical left and right responses and psychophysical left and right responses of VOT-act condition (for temporal cortex data) are represented in Figure 145. A moving average filter (in red) with a width of 100 samples and a gain of 15 was adjusted to the coefficients in order to facilitate the visualization and interpretation of their mean value.

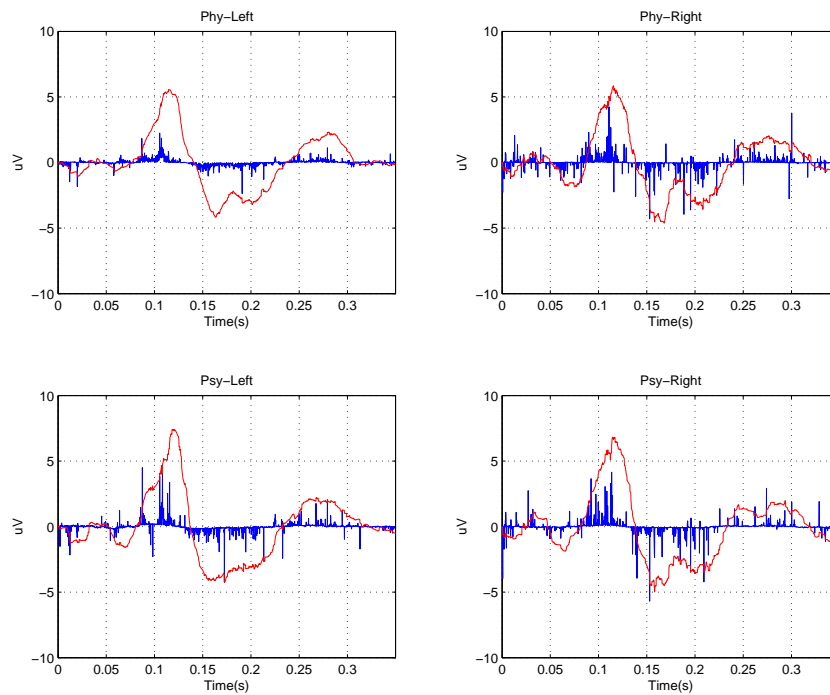


Figure 145 – Elastic Net coefficients for the VOT-active condition for physical left and right and psychophysical left and right responses.

As can be observed in the moving average signal of the Figure 145, the waves of the physical and psychophysical responses are very similar in both hemispheres. We believe that this occurred possibly due to the ascending sequence observed for both response physical and psychophysical response vectors. Also, as stated before, the interpretability of the results for the psychophysical data, which were so important, became difficult with the use of percentages, so this approach will not be used.

# APPENDIX D

## Discrete wavelet transform

The **DWT** is a linear transformation as in the discrete Fourier transform (DFT), used to obtain a time-frequency representation of time domain signals. The number of wavelet coefficients obtained after the transform are equal the number of samples of the original signal in the time domain. As in the DFT, the number of samples have to be a power of two. Figure 146 illustrate a comparison between DFT and **DWT**.

The wavelet basis functions are orthogonal. This is important for the use of wavelet transformed signals in statistical inference, as in multivariate ANOVA (MANOVA) which does not work with correlated variables. The orthogonality warrants that observations in wavelet domain are nearly uncorrelated (McKay et al. (2013) apud Angelini and Vidakovic (2003)).

Wavelets are localized in time. In a wavelet transformed signal, many coefficients are very small or have zero magnitude so that this signal can be represented by fewer wavelet coefficients than the original amount of samples of the signal if those coefficients are eliminated. This enables the use of the wavelet transform to perform compression of signals (McKay et al. (2013) apud Unser and Aldroubi (1996)).

In the **DWT** the transformed coefficients are arranged in a blocked structure orgnized by frequency bands. This structure is obtained through a recursive filtering and downsampling process (see figure 146) used to perform the transform that works as folows: As explained by McKay et al. (2013) apud Cohen and Kovacevic (1996), “The first block of coefficients, referred to as the “approximation” coefficients, comprise a low-pass filtered, downsampled representation of the original signal produced by convolving the signal with a low-pass filter (based on the approximation or “father” wavelet) and downsampling the filtered signal by a factor of two”. The



remaining blocks have the “detail” coefficients which comprises the high frequency information. They are obtained by convolving the signal now with a high-pass filter based on the “mother” wavelet. Coefficients in all blocks are arranged in order of increasing time. The amount of coefficients in each detail block increases with the frequency band of the level.

To obtain a multilevel wavelet decomposition the signal is first decomposed into detail and approximation coefficients, than the approximation ones are recursively divided into new approximation and detail coefficients composing the higher levels.

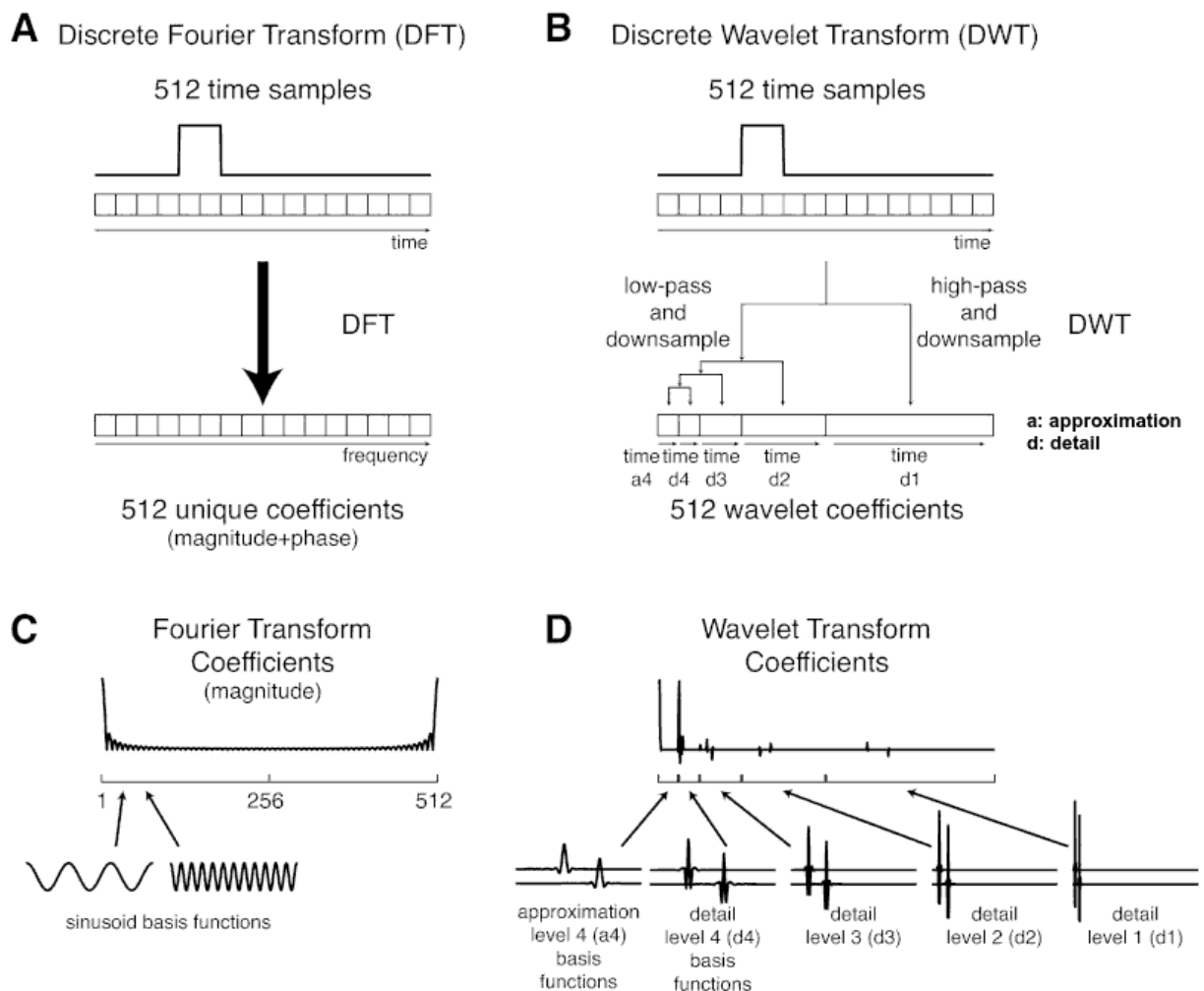


Figure 146 – Comparison between the DWT and the discrete Fourier transform (DFT). The transforms are applied in a square wave with 512 samples as example. (A) DFT transforms the 512 samples of the time domain signal into 512 coefficients in frequency domain. (B) DWT transforms the 512 samples of the time domain signal into 512 wavelet coefficients (‘a’: approximation coefficients; ‘d’: detail coefficients) obtained through a recursive filtering and downsampling process. (C) DFT sinusoidal basis function is infinite in time while (D) wavelet basis functions are limited and localized in time. CREDITS: McKay et al. (2013) (adapted)

Thus, in short, the DWT works by filtering the signal by a “filter bank” with different time-

frequency features. Each filter is a wavelet ('mother' wavelet) described by its coefficient which is a function of scale and shift parameters as can be seen in figure 147 (left). The scale parameter determines the contraction/dilatation of the wavelet which defines the frequency 'detected' (filtered out) by the wavelet. The smaller the dilation of the wavelet, the smaller the number of samples of the signal over time filtered by this wavelet and, thus, the greater the number of coefficients necessary to cover the same signal. The shift parameter determines the position of the wavelet in time. The signal is decomposed at each level generating detail and approximation coefficients. Going from the highest to the lowest frequency level, each level has half of the coefficients from the previous one. An equation for the wavelet transform is given by (Kaushik et al., 2014),

$$X(e, f) = \sum_{-\infty}^{\infty} x(m) \phi_{e, f}(m) \quad (\text{D.1})$$

where  $\phi(m)$  is a window of finite length,  $f$  is the window translation parameter and  $e$  is a contraction parameter.

The scalogram (figure 147 (right)) is a graphical time-frequency representation of the wavelet coefficients organized by decomposition level. The amount of divisions represents the amount of coefficients of each level. The total amount of divisions is equal the amount of samples in the original signal. Tones of colors can be used to represent the value of each coefficient in the scalogram.

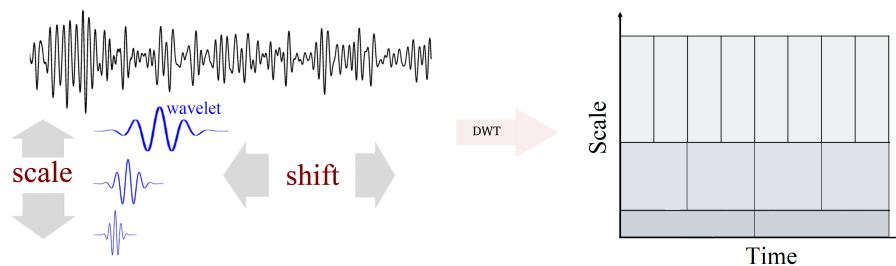


Figure 147 – Scale and shift of mother wavelets filtering a signal (left). A DWT scalogram (right).

1

In Figure 148 is represented the DWT applied in a chirp signal. The time domain signal is represented at the bottom of the figure where it is possible to see that it is a signal which frequency increase with time. The signal is decomposed in 4 discrete wavelet levels. The detail

<sup>1</sup> CREDITS: <http://www.seismology.harvard.edu/research/wavePicking.html> (adapted).

coefficients are described by the letter “W” and the approximation coefficient by the letter “V” according with the nomenclature of the “wavelets” package from the software R (Aldrich, 2013).

Wavelets from the highest levels (as in W1 and W2), have large scale and are short in time so that their coefficients will have greater values in the region aligned with the high frequency part of the time domain signal. On the other hand, wavelets from the lower levels (as in W4 and V4) have short scale and are large in time so that the coefficients at those levels will have greater values at the region aligned with the low frequency part of the chirp signal. Summing up all the wavelets weighted by their coefficient values, it is obtained the original signal again. Here, a linear filtering (zero phase) can be performed by zeroing the coefficients of the frequency band (level) to be filtered out so that the reconstructed signal will not present those frequencies.

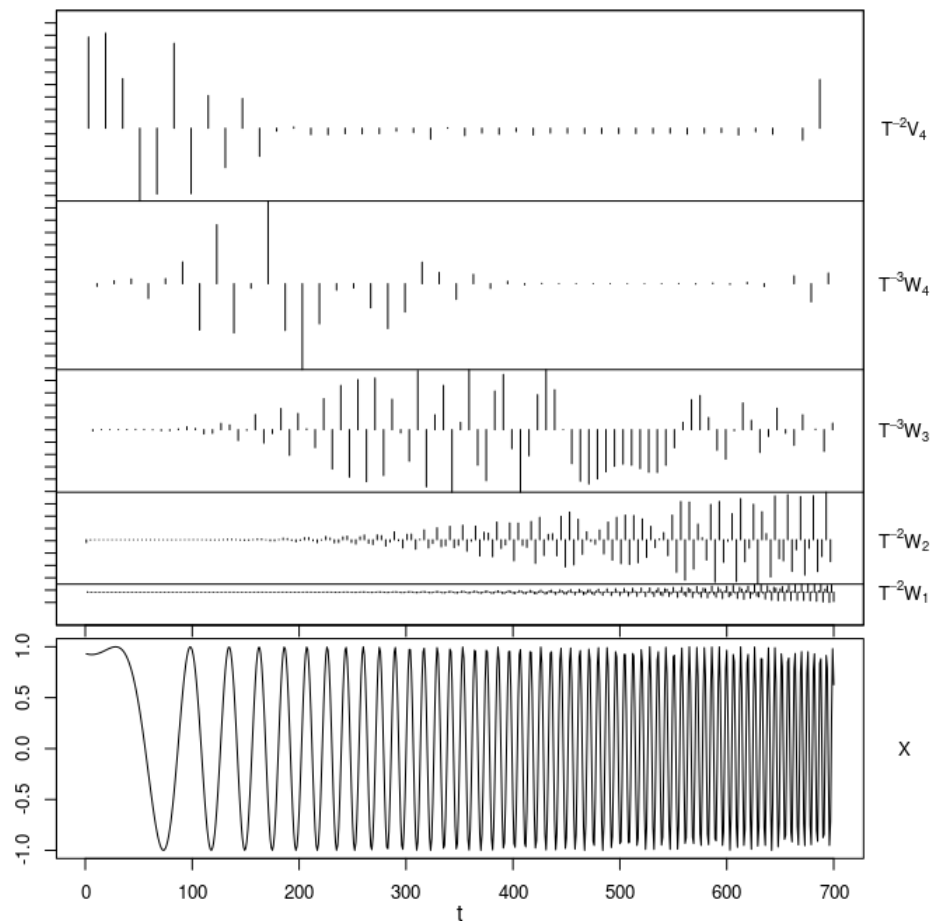


Figure 148 – DWT of a chirp signal (signal with variable frequency over time) decomposed in 4 wavelet levels.

More details about the wavelet transform can be consulted at (Mallat, 1999).

# APPENDIX E

## Complete anova and ranova tables

This Appendix presents the ANOVA and RANOVA tables and the graphic representation of the complete mixed model obtained for the discrepancy computed for the coefficients of the ROIs 2 to 8. These results for the ROI-1 was already presented in Section 8.

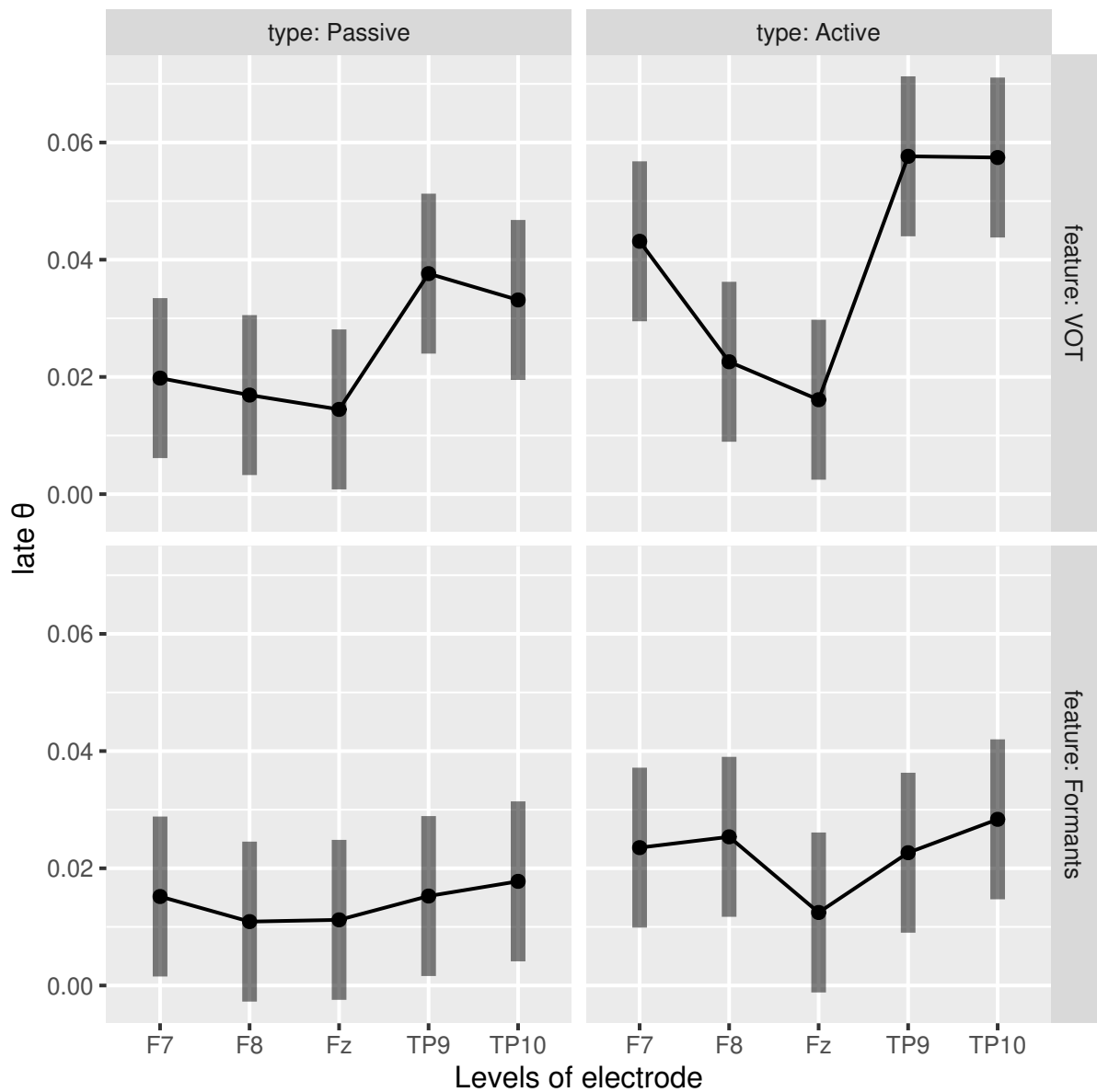


Figure 149 – Representation of the complete mixed-effects model for the discrepancy computed for the ROI-2 including the fixed factors: feature, type, electrodes and regression.

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.0101943	0.0101943	1	190	20.9435	8.527e-06
type	0.0075335	0.0075335	1	190	15.4771	0.000117
electrode	0.0141154	0.0035288	4	190	7.2498	1.873e-05
feature:type	0.0005973	0.0005973	1	190	1.2272	0.269356
feature:electrode	0.0060548	0.0015137	4	190	3.1098	0.016538
type:electrode	0.0017793	0.0004448	4	190	0.9139	0.456949
feature:type:electrode	0.0011898	0.0002975	4	190	0.6111	0.655120

ANOVA-like table for random-effects: Single term deletions

Model:

```
r2.late.theta ~ feature + type + electrode + (1 | subject) +
  feature:electrode
```

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	13	479.12	-932.23			
(1   subject)	12	476.08	-928.17	6.0657	1	0.01378

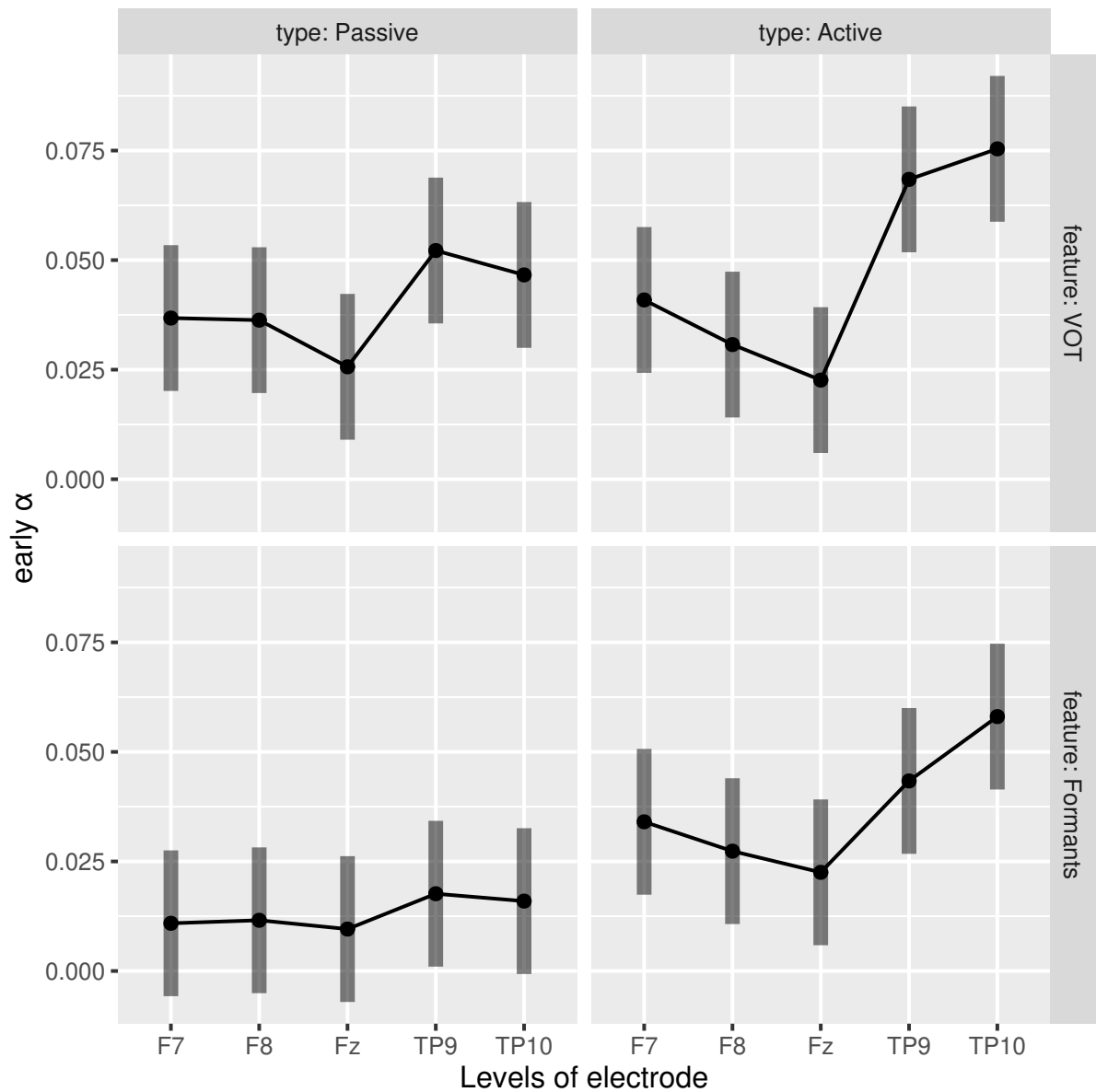


Figure 150 – Representation of the complete mixed-effects model for the discrepancy computed for the ROI-3 including the fixed factors: feature, type, electrodes and regression.

## Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.0187561	0.0187561	1	190	31.7840	6.158e-08
type	0.0141267	0.0141267	1	190	23.9390	2.115e-06
electrode	0.0270946	0.0067736	4	190	11.4786	2.290e-08
feature:type	0.0034495	0.0034495	1	190	5.8455	0.01656
feature:electrode	0.0031977	0.0007994	4	190	1.3547	0.25131
type:electrode	0.0071391	0.0017848	4	190	3.0245	0.01899
feature:type:electrode	0.0002393	0.0000598	4	190	0.1014	0.98189

## ANOVA-like table for random-effects: Single term deletions

Model:

```
r3.early.alpha ~ feature + type + electrode + (1 | subject) +
  feature:type + type:electrode
```

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	14	450.58	-873.17			
(1   subject)	13	433.10	-840.21	34.959	1	3.367e-09

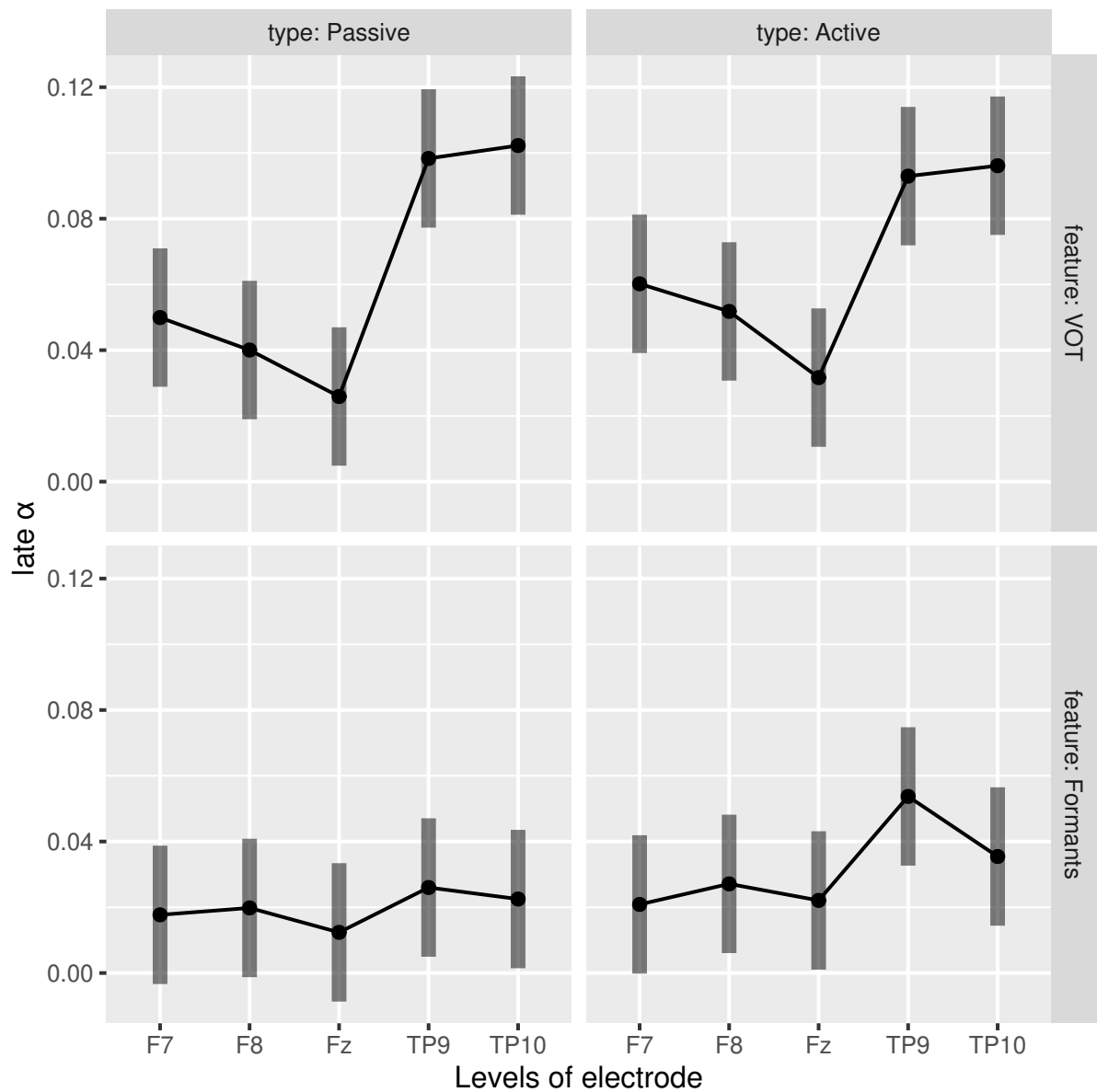


Figure 151 – Representation of the complete mixed-effects model for the discrepancy computed for the ROI-4 including the fixed factors: feature, type, electrodes and regression.

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.084303	0.084303	1	190	91.4917	< 2.2e-16
type	0.003267	0.003267	1	190	3.5460	0.06122
electrode	0.067439	0.016860	4	190	18.2976	9.863e-13
feature:type	0.001095	0.001095	1	190	1.1880	0.27711
feature:electrode	0.025213	0.006303	4	190	6.8409	3.643e-05
type:electrode	0.000383	0.000096	4	190	0.1038	0.98106
feature:type:electrode	0.003154	0.000789	4	190	0.8558	0.49162



ANOVA-like table for random-effects: Single term deletions

Model:

r4.late.alpha ~ feature + electrode + (1 | subject) + feature:electrode

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	12	410.92	-797.84			
(1   subject)	11	391.54	-761.09	38.746	1	4.827e-10

<none>

12 410.92 -797.84

(1 | subject)

11 391.54 -761.09 38.746 1 4.827e-10

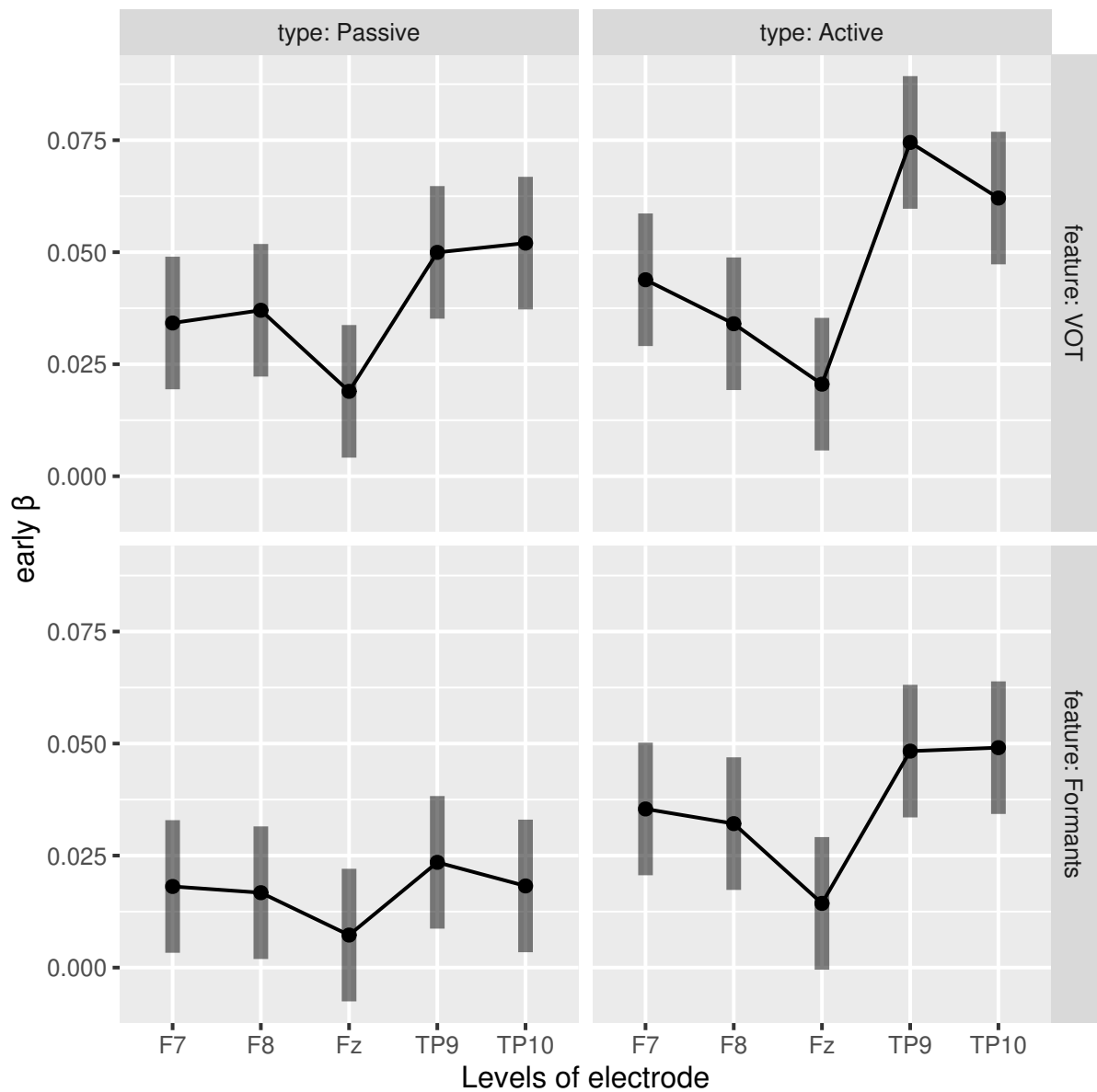


Figure 152 – Representation of the complete mixed-effects model for the discrepancy computed for the ROI-5 including the fixed factors: feature, type, electrodes and regression.

## Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.014814	0.0148144	1	190	29.9197	1.407e-07
type	0.010521	0.0105208	1	190	21.2482	7.391e-06
electrode	0.031800	0.0079501	4	190	16.0563	2.420e-11
feature:type	0.001523	0.0015235	1	190	3.0769	0.08103
feature:electrode	0.002728	0.0006821	4	190	1.3776	0.24324
type:electrode	0.003411	0.0008528	4	190	1.7224	0.14667
feature:type:electrode	0.000843	0.0002108	4	190	0.4258	0.78988

## ANOVA-like table for random-effects: Single term deletions

Model:

r5.early.beta ~ feature + type + electrode + (1 | subject)

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	9	483.98	-949.96			
(1   subject)	8	471.51	-927.01	24.944	1	5.903e-07

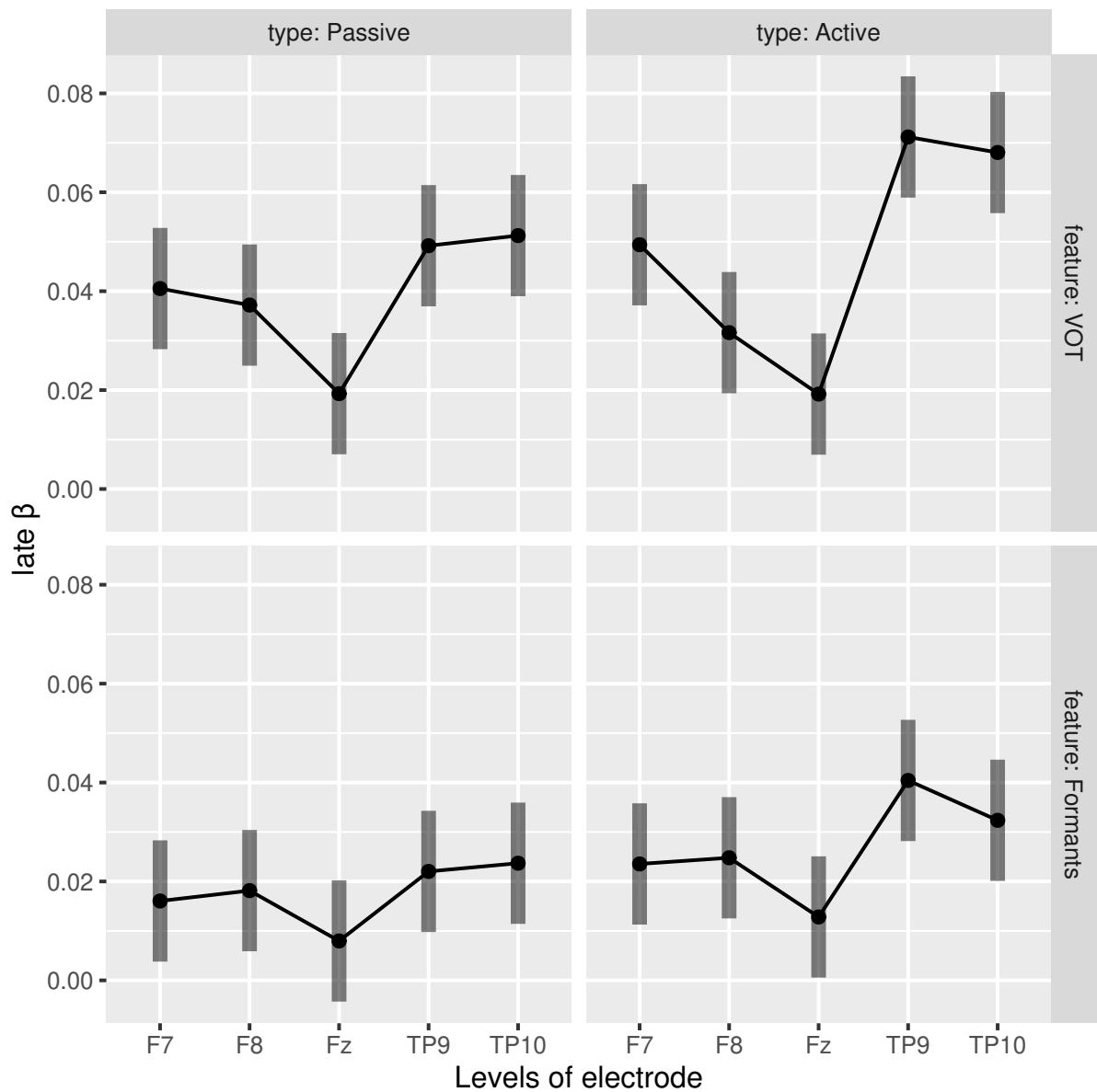


Figure 153 – Representation of the complete mixed-effects model for the discrepancy computed for the ROI-6 including the fixed factors: feature, type, electrodes and regression.

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.0254369	0.0254369	1	190	79.1949	4.443e-16
type	0.0042628	0.0042628	1	190	13.2719	0.0003475
electrode	0.0279623	0.0069906	4	190	21.7644	8.447e-15
feature:type	0.0000091	0.0000091	1	190	0.0283	0.8665036
feature:electrode	0.0044554	0.0011138	4	190	3.4678	0.0092410
type:electrode	0.0028072	0.0007018	4	190	2.1850	0.0722164
feature:type:electrode	0.0006902	0.0001726	4	190	0.5372	0.7085340

ANOVA-like table for random-effects: Single term deletions

Model:

r6.late.beta ~ feature + type + electrode + (1 | subject) + feature:electrode

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	13	515.74	-1005.49			
(1   subject)	12	498.88	-973.75	33.735	1	6.315e-09

<none>

13 515.74 -1005.49

(1 | subject)

12 498.88 -973.75 33.735 1 6.315e-09

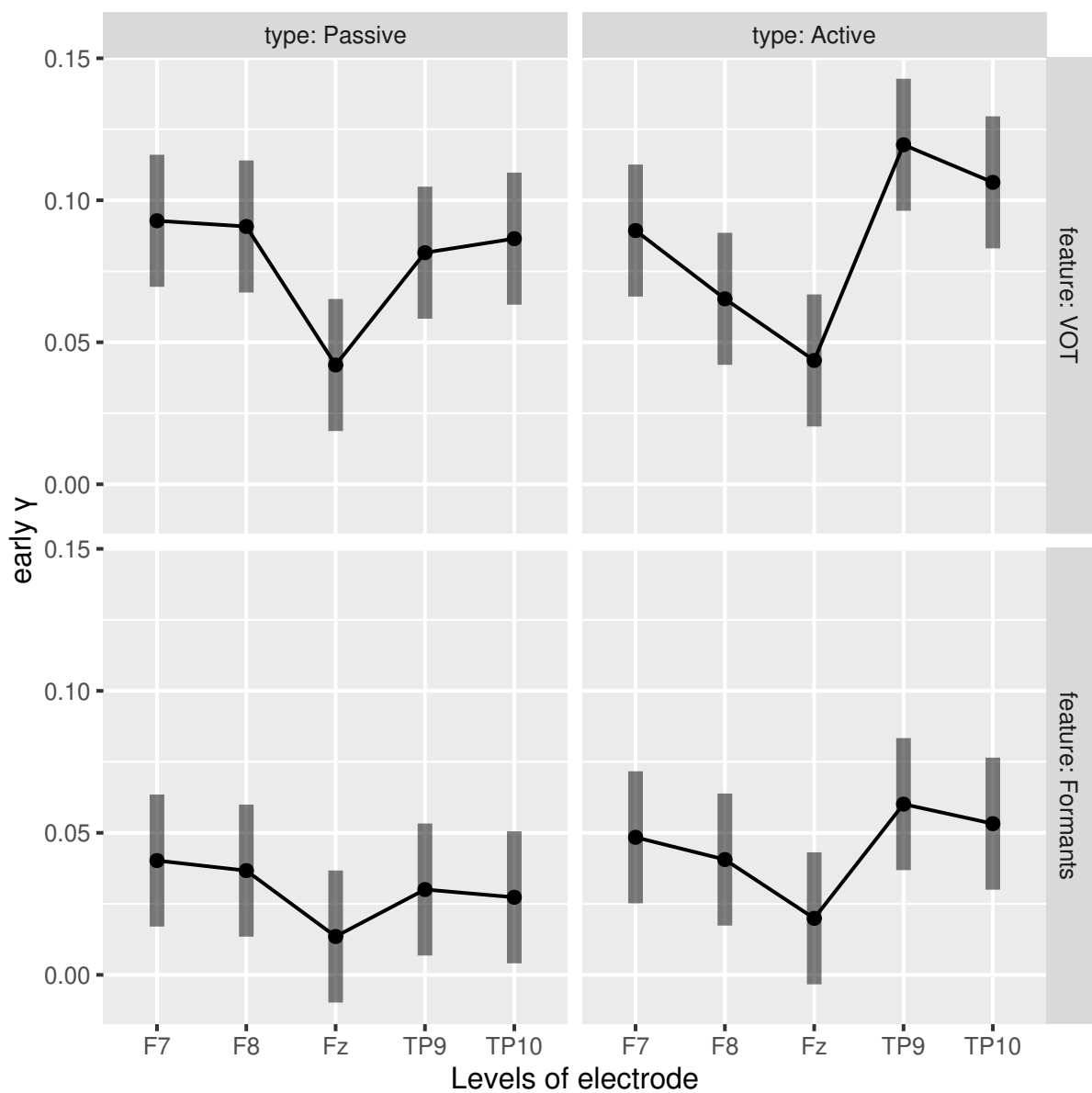


Figure 154 – Representation of the complete mixed-effects model for the discrepancy computed for the ROI-7 including the fixed factors: feature, type, electrodes and regression.

## Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.110190	0.110190	1	190	93.6597	< 2.2e-16
type	0.006062	0.006062	1	190	5.1525	0.02434
electrode	0.053186	0.013297	4	190	11.3019	3.009e-08
feature:type	0.001063	0.001063	1	190	0.9037	0.34301
feature:electrode	0.006872	0.001718	4	190	1.4603	0.21591
type:electrode	0.013951	0.003488	4	190	2.9646	0.02091
feature:type:electrode	0.002020	0.000505	4	190	0.4292	0.78746

## ANOVA-like table for random-effects: Single term deletions

Model:

```
r7.early.gamma ~ feature + type + electrode + (1 | subject) +
  type:electrode
```

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	13	381.62	-737.24			
(1   subject)	12	365.90	-707.79	31.451	1	2.046e-08

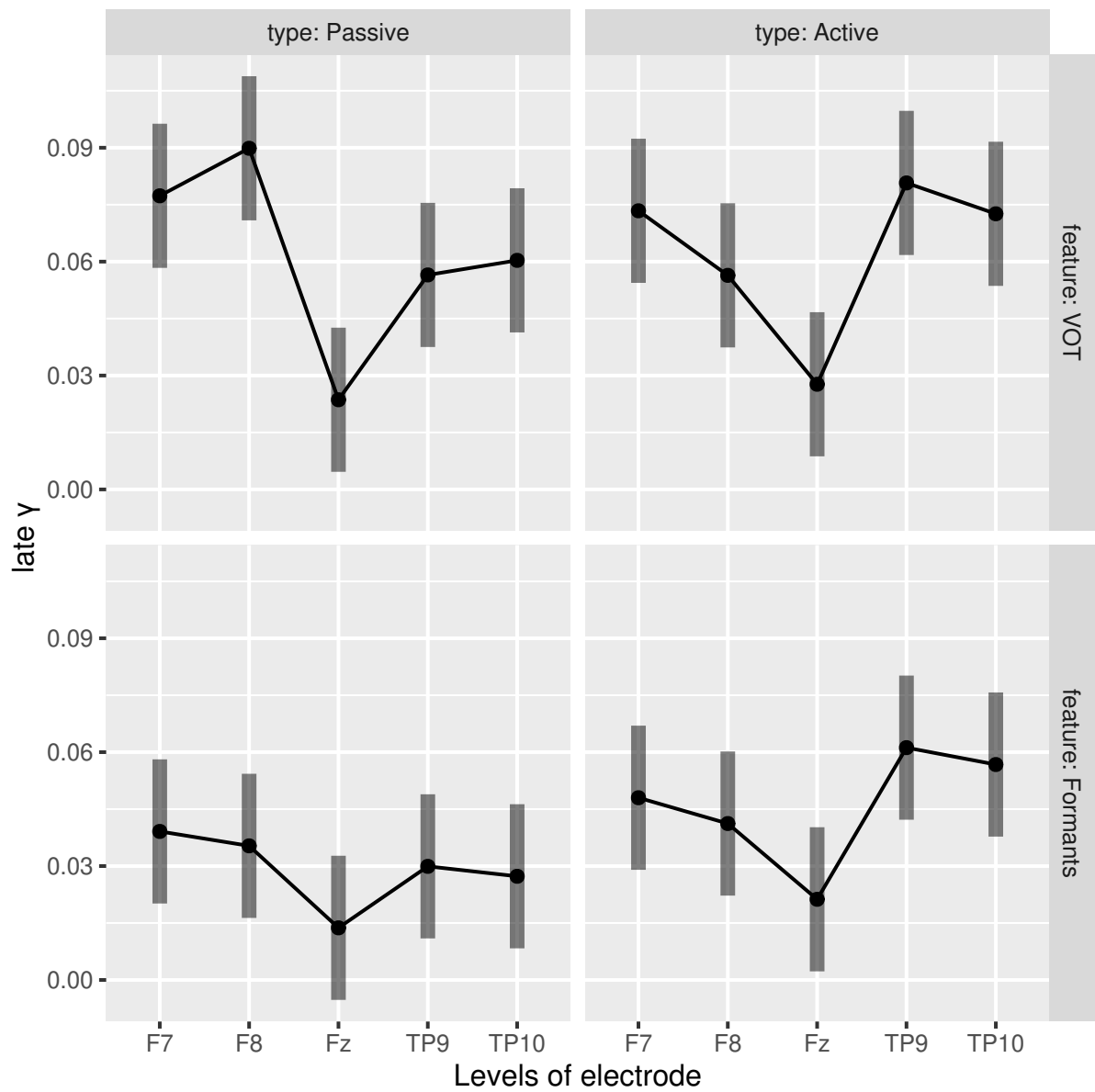


Figure 155 – Representation of the complete mixed-effects model for the discrepancy computed for the ROI-8 including the fixed factors: feature, type, electrodes and regression.

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
feature	0.032918	0.032918	1	190	37.5972	4.931e-09
type	0.004080	0.004080	1	190	4.6597	0.03213
electrode	0.043885	0.010971	4	190	12.5307	4.562e-09
feature:type	0.003503	0.003503	1	190	4.0009	0.04690
feature:electrode	0.004723	0.001181	4	190	1.3487	0.25348
type:electrode	0.011702	0.002925	4	190	3.3412	0.01136
feature:type:electrode	0.002186	0.000547	4	190	0.6242	0.64579

ANOVA-like table for random-effects: Single term deletions

Model:

```
r8.late.gamma ~ feature + type + electrode + (1 | subject) +  
  feature:type + type:electrode
```

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	14	411.22	-794.44			
(1   subject)	13	403.56	-781.12	15.32	1	9.074e-05

# APPENDIX F

## Assumptions

In this section, it is presented the test of the ANOVA assumptions and explained why slight violations to those assumptions are not a problem for our results given the robustness of the linear mixed-effects models and, specifically, to the *lmer* function we used to generate such models in this work.

The use of ANOVA is classically restricted to some assumptions such as:

- Uncorrelated fixed effect predictors;
- Homocedasticity of random effects and of residuals;
- Gaussian distribution for random effects and residuals.

However, linear mixed-effects models are quite robust to violations of those assumptions to some level. [Schielzeth et al. \(2020\)](#) tested this robustness evaluating effects of skewed, bimodal and heterocedastic random effect and residuals as well as correlated fixed effect predictors. The authors showed that violations of distributional assumptions of residuals and random effect variances is small and even skewed and heterocedastic data resulted in little overall bias, particularly for fixed effect estimates. Estimates for random effect components that violated distributional assumptions became less precise but remained unbiased and this occurred only for parameters of the model that violated the assumption. This pattern was also observed for strongly correlated fixed effects (i.e. only the correlated parameters being affected).



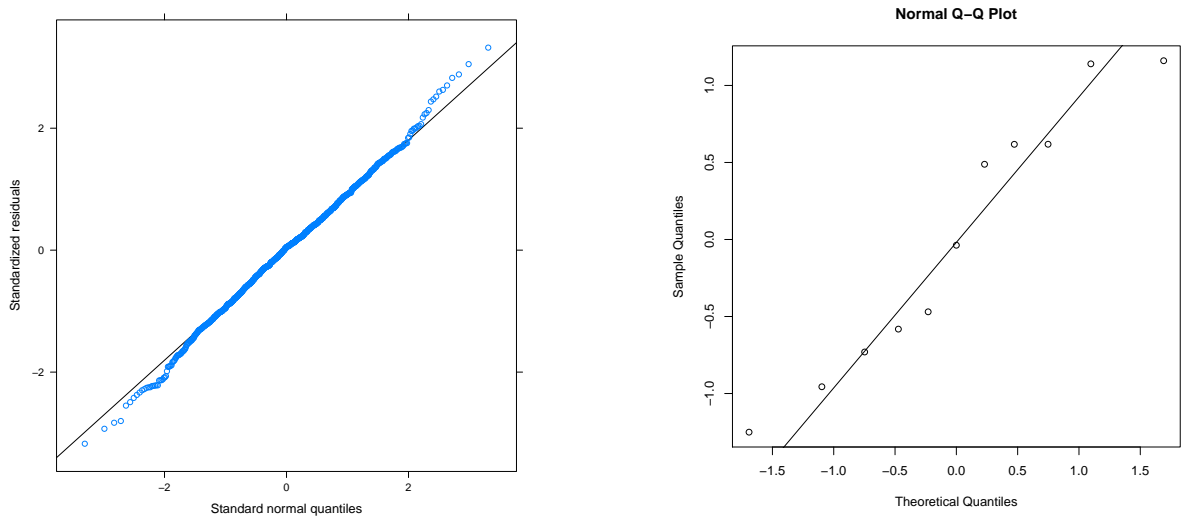
Specifically for the case of non-gaussian distributions, a common solution is perform nonlinear transformations in the response variable to improve the fit to normal distributions. The problem with this approach is that the reduce the interpretability of the model because, now, the data is not in its original scale. Since the violations to the assumption of normality of distributions lead to small prediction errors in linear mixed-effects models, it is worth to analyze if this transformation is really necessary. It is a trade-off between interpretability and conformance to model assumptions.

In this work, to generate the linear mixed-effects models, we used the *lmer* function from the *lme4* package for the R software. This function applies the maximum likelihood or restricted maximum likelihood fitting of the linear mixed-effects models using a penalized least squares method (Bates et al., 2014). In Molenberghs and Verbeke (2000) and Bartlett (2014) a mathematical development is made showing why parameters of linear mixed-effects models estimated by maximum likelihood or restricted maximum likelihood made these models robust even when random effects or residual errors are not normally distributed or even if their variances are not constant. Then, p-values and confidence intervals for the fixed effects will be valid.

Following it is showed the results of the Shapiro-Wilk test of normality and plots of the distribution of residuals and of the residuals vs. fitted values (to analyze homocedasticity) for the dependent variables of this work: N1, P2, N1-P2, T1, T2 and importance of ROIs 1 to 8. If the p-value of the Shapiro-Wilk test is greater than 0.05 than there is no evidence to reject the hypothesis that the sample tested was drawn from a normally distributed population.

We consider that our groups (participants) predictors are uncorrelated because we performed a randomized experiment and the participants were different. The Levene test for homocedasticity was not performed here because it does not perform well for big datasets as is our case. It can be seen in the following results that not all dependent variables met the assumption of normality and we rely in the robustness of the model to address those cases. Besides, observing the ANOVA results in the Section 6 and in the Appendix E, it can be seen that the majority of the factors or factor interactions that presented significant effects have very small p-values. This indicates that even if some correction was performed, the increase in the p-value will make no difference and we will still observe effects in those cases.

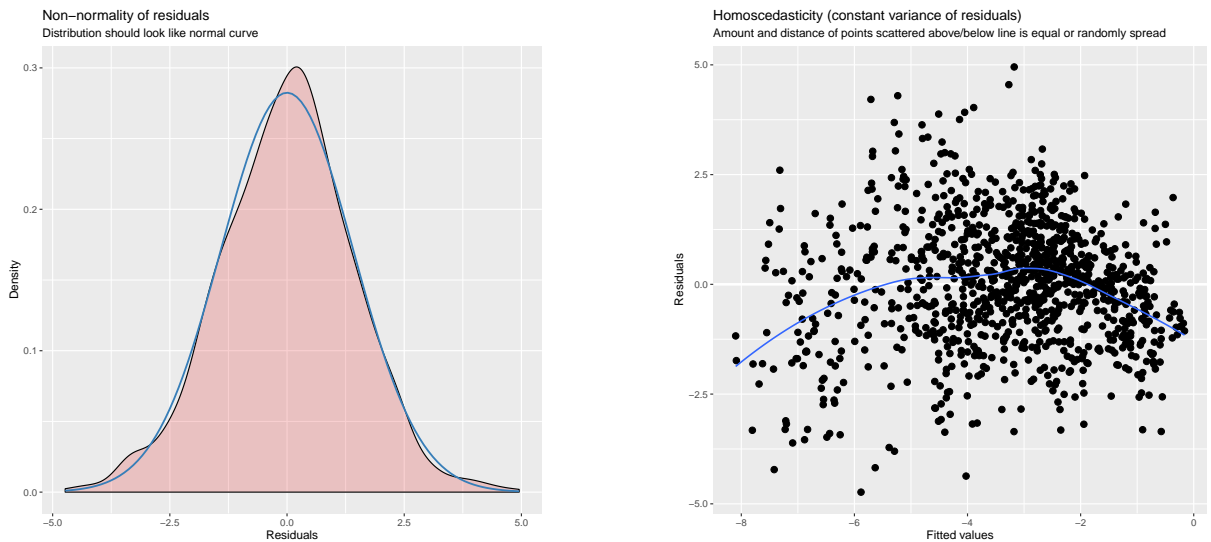
**N1**



(a) Representation of the quantiles of the normal distribution by the standardized values of the residuals of the model for N1.

(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 156 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for N1 and a theoretical normal distribution (blue).

(b) Representation of the fitted values of N1 by the residuals of the model (black dots) and the regression line (blue).

Figure 157 – Graphical representation for normality and homoscedasticity verification.

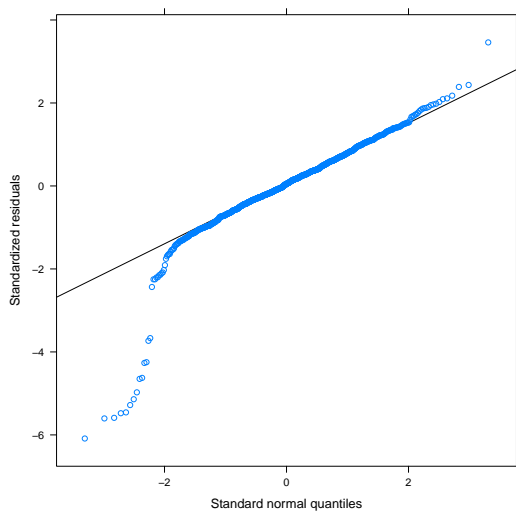
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.99789, p-value = 0.1996
```

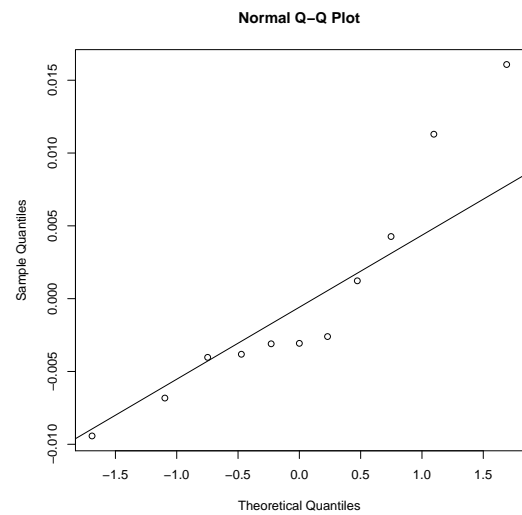
The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.92805, p-value = 0.3915
```

**T1**



(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for T1.



(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 158 – Graphical representation for normality verification.

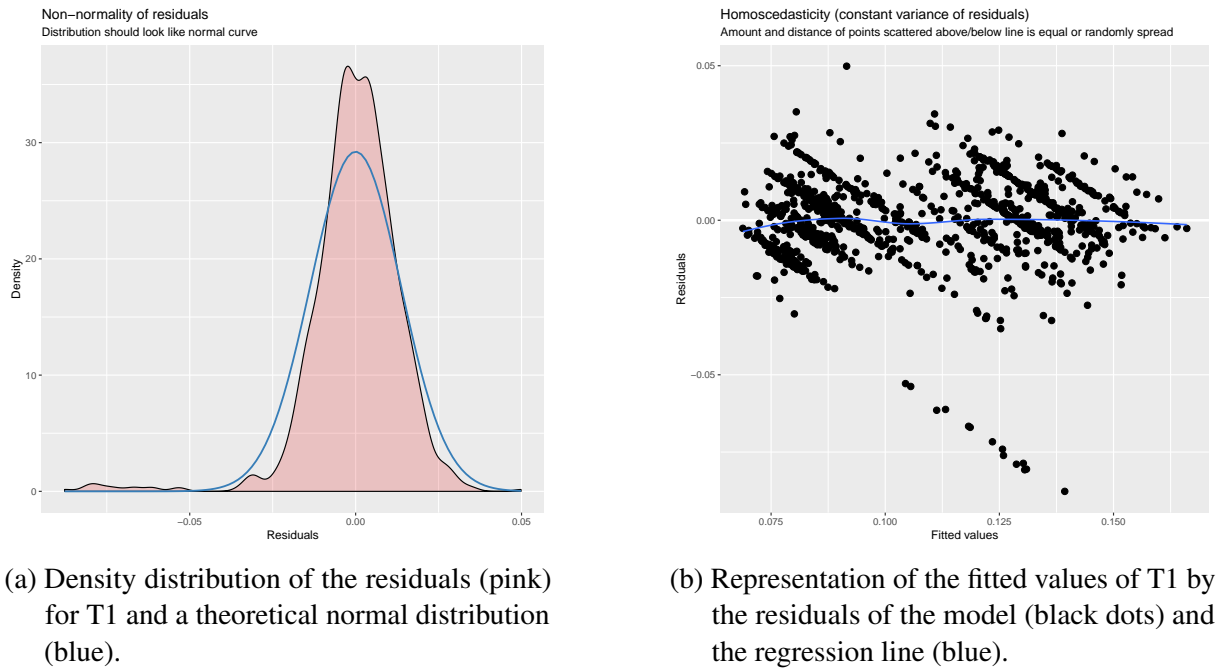


Figure 159 – Graphical representation for normality and homoscedasticity verification.

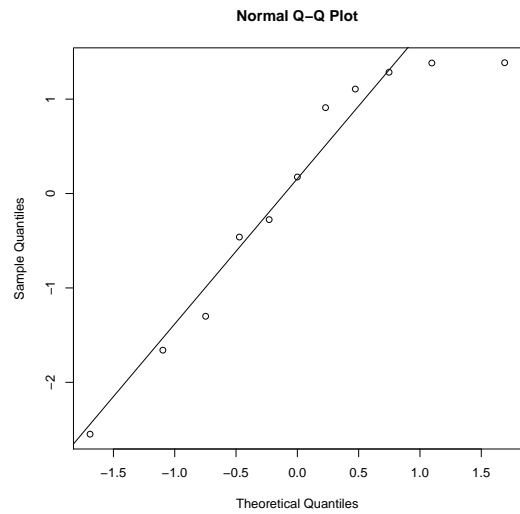
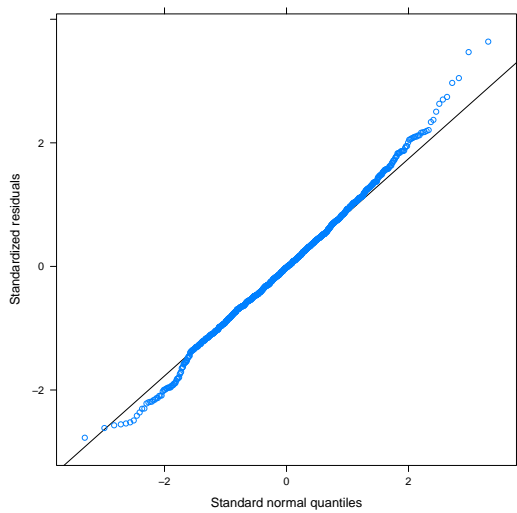
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.87705, p-value < 2.2e-16
```

The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.87892, p-value = 0.1008
```

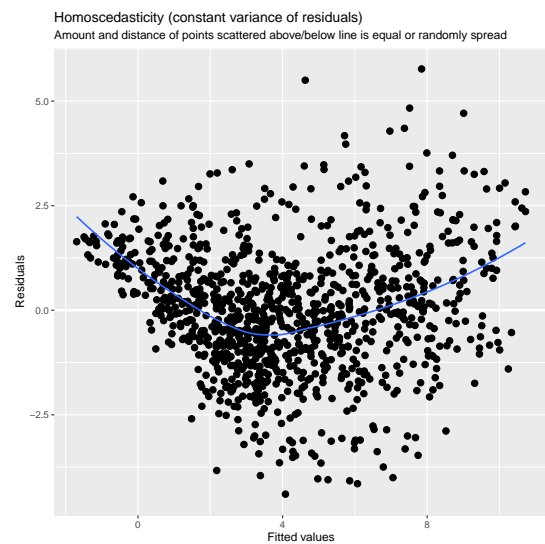
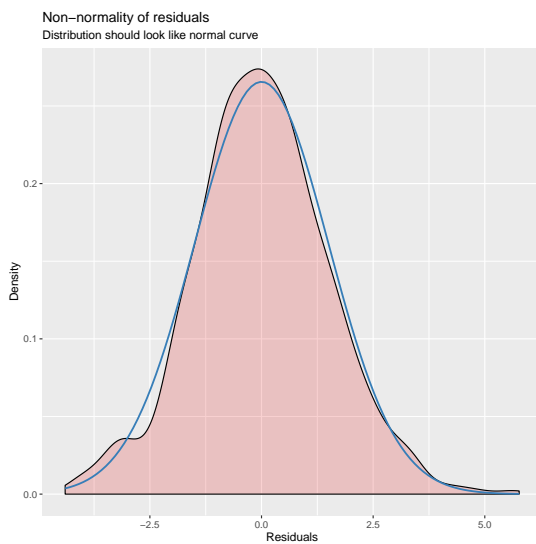
**P2**



(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for P2.

(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 160 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for P2 and a theoretical normal distribution (blue).

(b) Representation of the fitted values of P2 by the residuals of the model (black dots) and the regression line (blue).

Figure 161 – Graphical representation for normality and homoscedasticity verification.

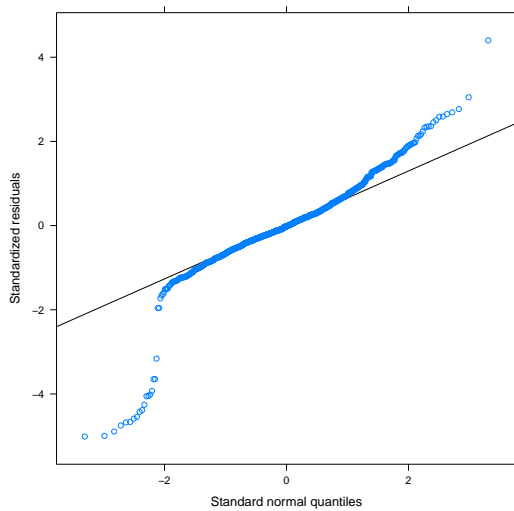
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.99613, p-value = 0.009463
```

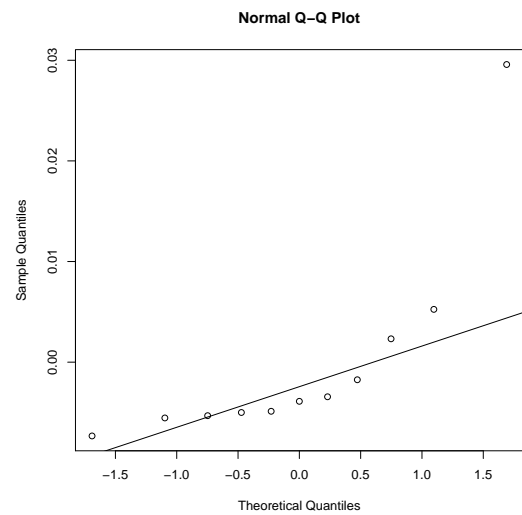
The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.89592, p-value = 0.1647
```

## T2

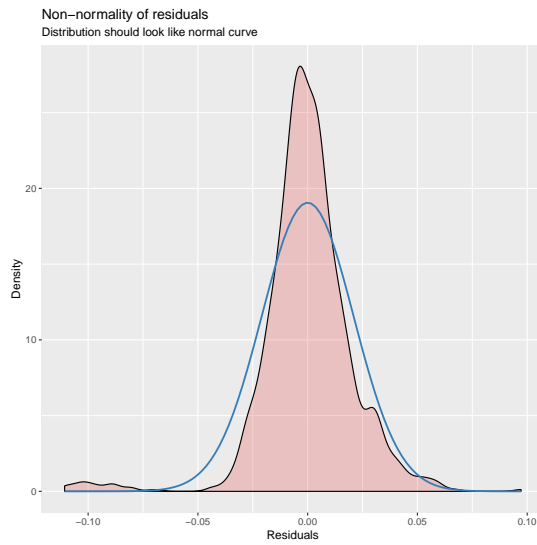


(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for T2.

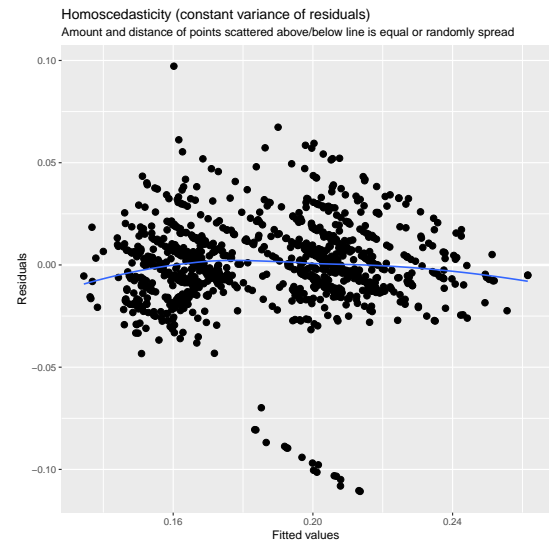


(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 162 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for T2 and a theoretical normal distribution (blue).



(b) Representation of the fitted values of T2 by the residuals of the model (black dots) and the regression line (blue).

Figure 163 – Graphical representation for normality and homoscedasticity verification.

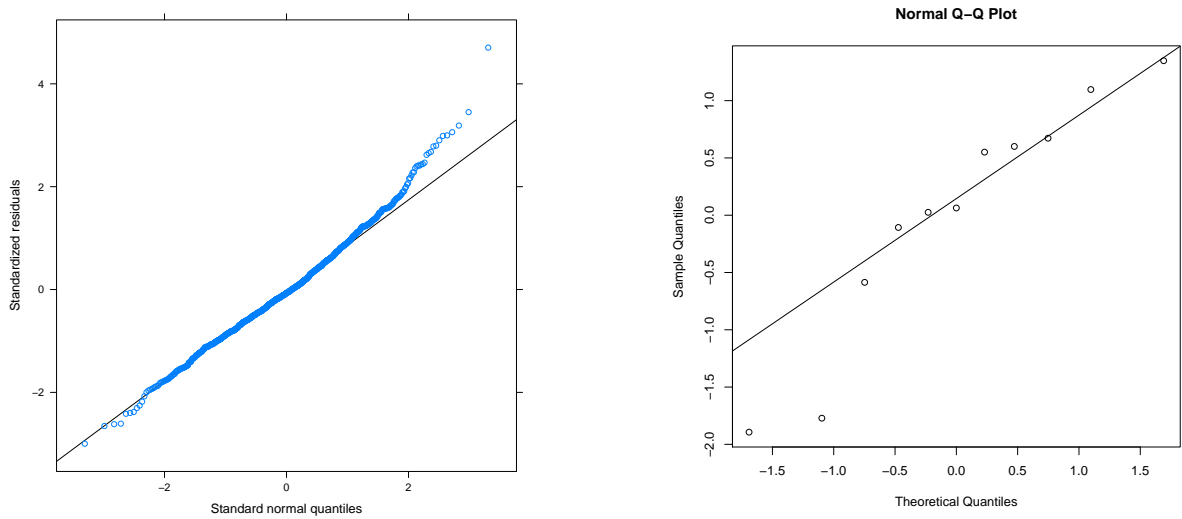
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.88687, p-value < 2.2e-16
```

The Shapiro-Wilk normality test for the random variable is:

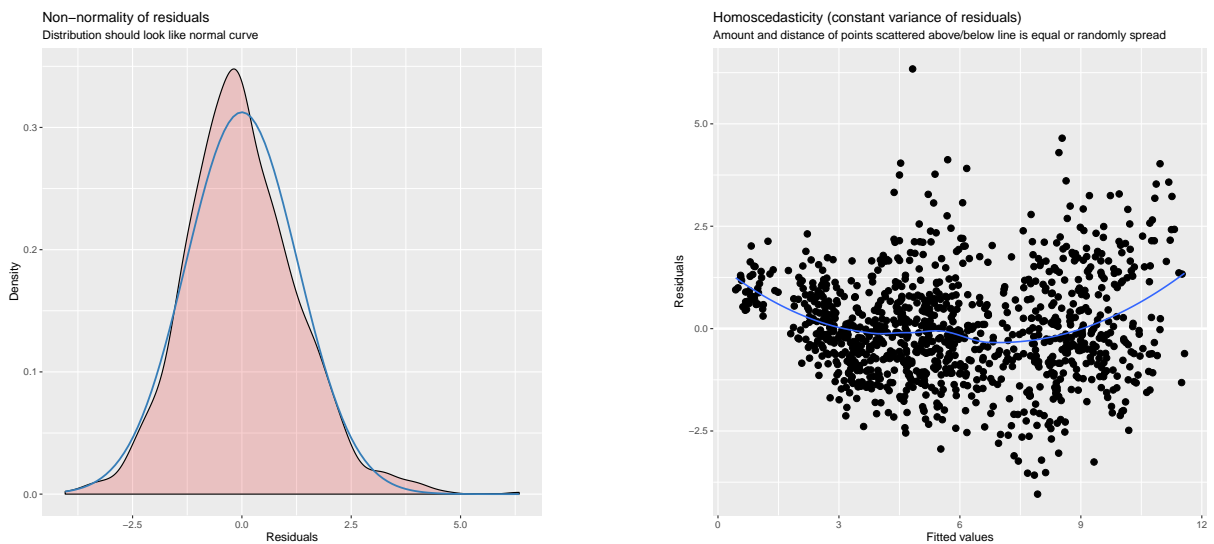
```
data: r
W = 0.64068, p-value = 8.216e-05
```

**P2-N1**



- (a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for P2-N1.
- (b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 164 – Graphical representation for normality verification.



- (a) Density distribution of the residuals (pink) for P2-N1 and a theoretical normal distribution (blue).
- (b) Representation of the fitted values of P2-N1 by the residuals of the model (black dots) and the regression line (blue).

Figure 165 – Graphical representation for normality and homoscedasticity verification.

The Shapiro-Wilk normality test for the residuals of the model is:

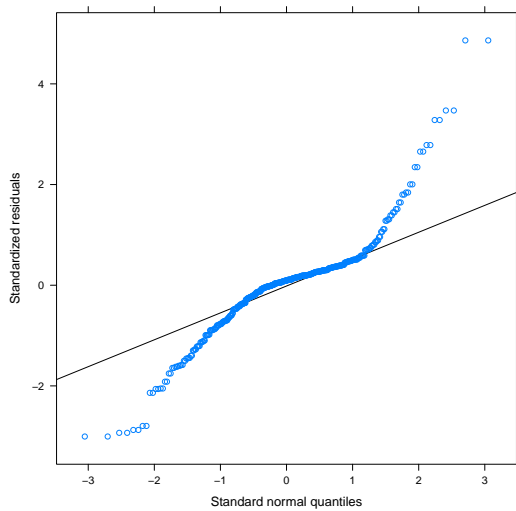


```
data: resid(fm)
W = 0.98874, p-value = 2.904e-07
```

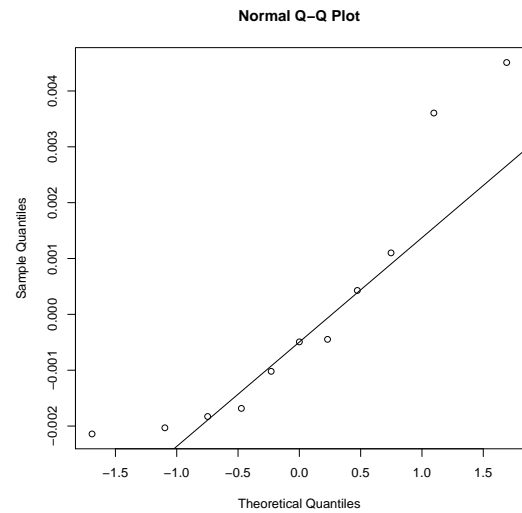
The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.90628, p-value = 0.2202
```

### ROI-1

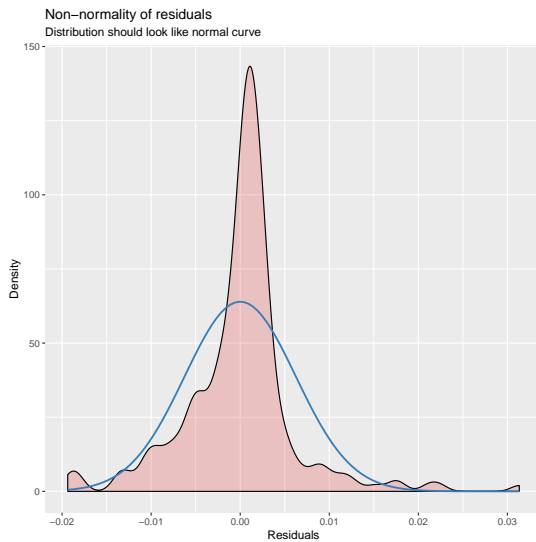


(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for ROI-1.

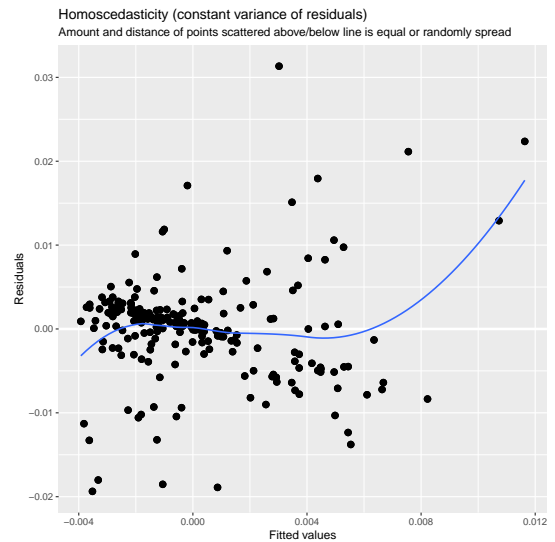


(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 166 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for ROI-1 and a theoretical normal distribution (blue).



(b) Representation of the fitted values of ROI-1 by the residuals of the model (black dots) and the regression line (blue).

Figure 167 – Graphical representation for normality and homoscedasticity verification.

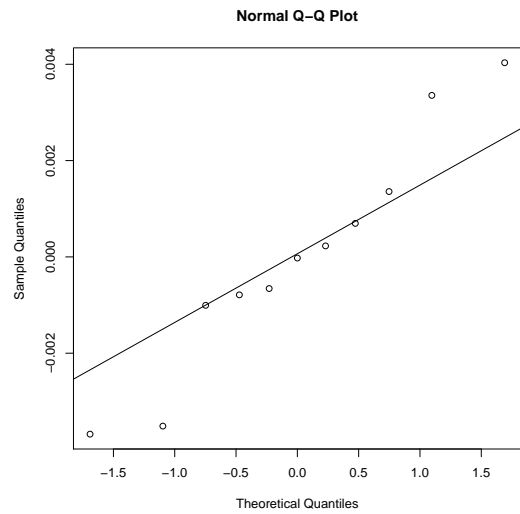
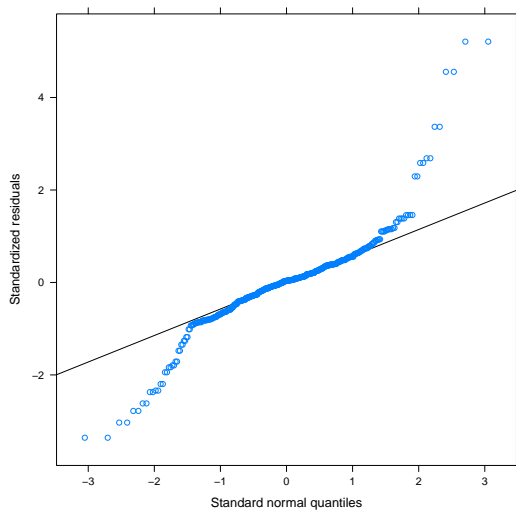
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.89151, p-value < 2.2e-16
```

The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.85183, p-value = 0.04509
```

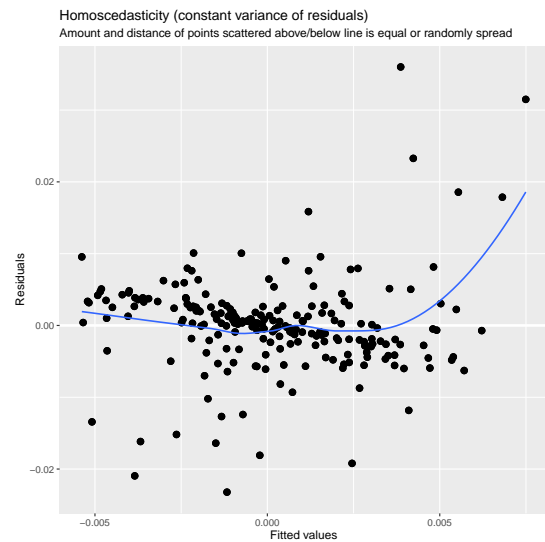
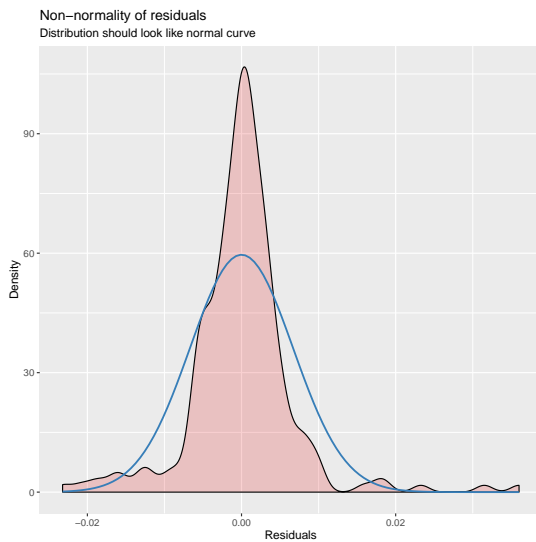
ROI-2



(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for ROI-2.

(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 168 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for ROI-2 and a theoretical normal distribution (blue).

(b) Representation of the fitted values of ROI-2 by the residuals of the model (black dots) and the regression line (blue).

Figure 169 – Graphical representation for normality and homocedasticity verification.

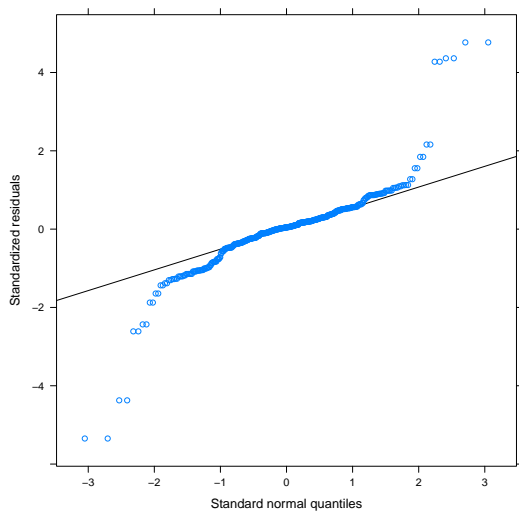
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.86751, p-value < 2.2e-16
```

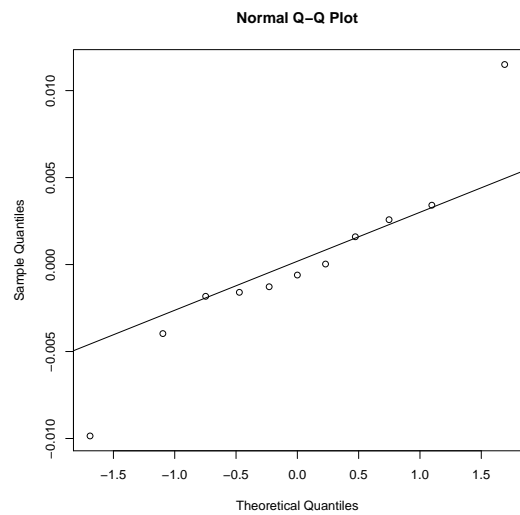
The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.94784, p-value = 0.6165
```

### ROI-3



(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for ROI-3.



(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 170 – Graphical representation for normality verification.

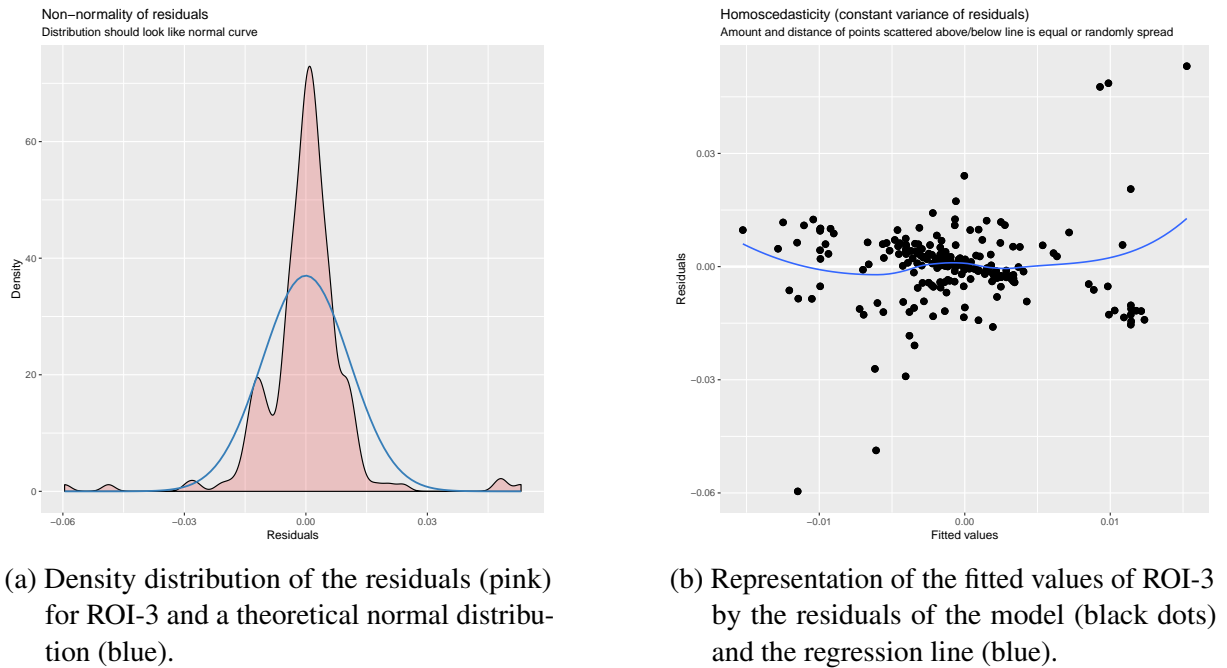


Figure 171 – Graphical representation for normality and homoscedasticity verification.

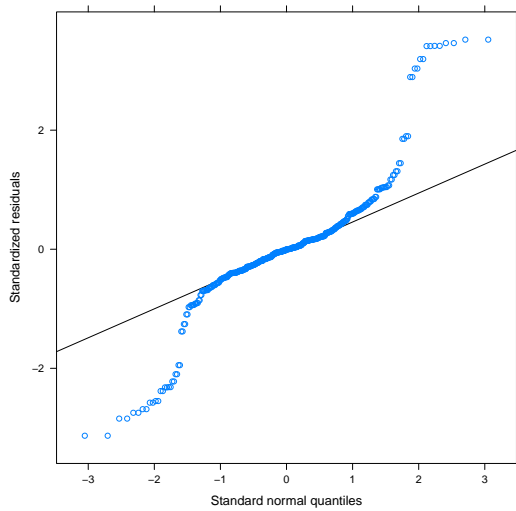
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.81682, p-value < 2.2e-16
```

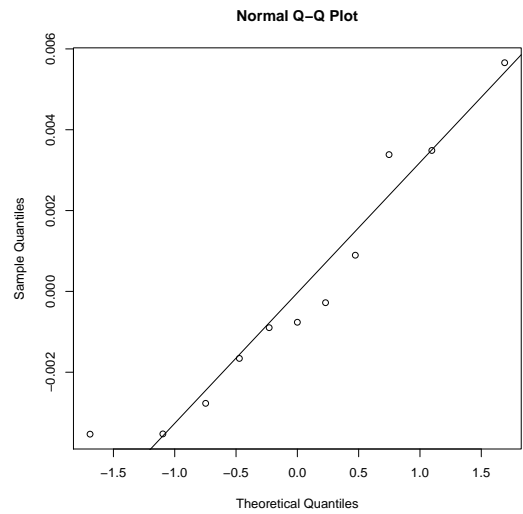
The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.92905, p-value = 0.4013
```

ROI-4

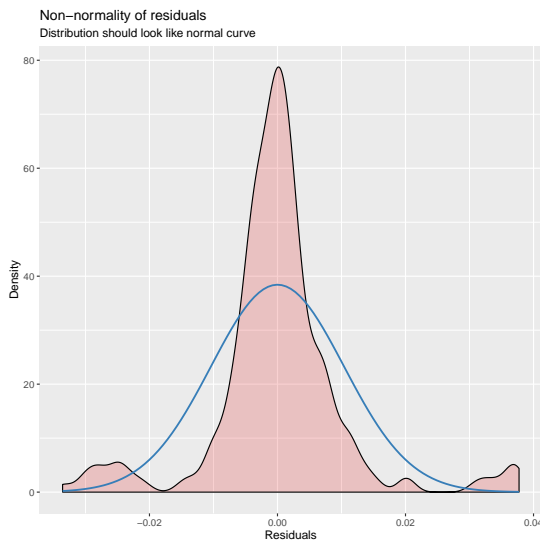


(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for ROI-4.

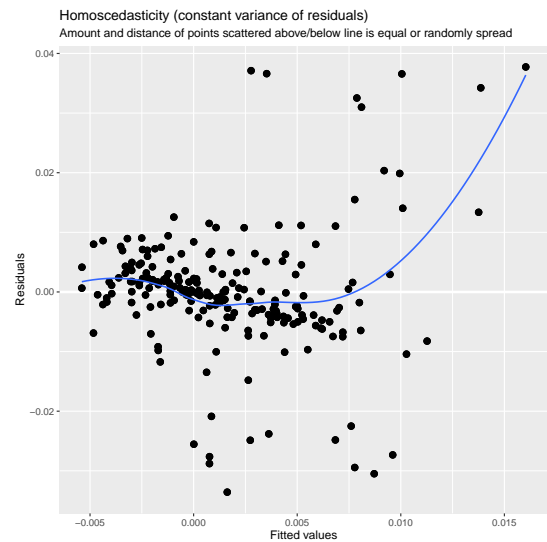


(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 172 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for ROI-4 and a theoretical normal distribution (blue).



(b) Representation of the fitted values of ROI-4 by the residuals of the model (black dots) and the regression line (blue).

Figure 173 – Graphical representation for normality and homoscedasticity verification.

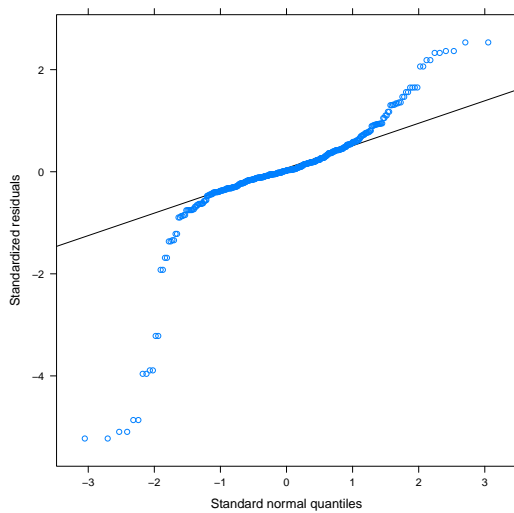
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.85786, p-value < 2.2e-16
```

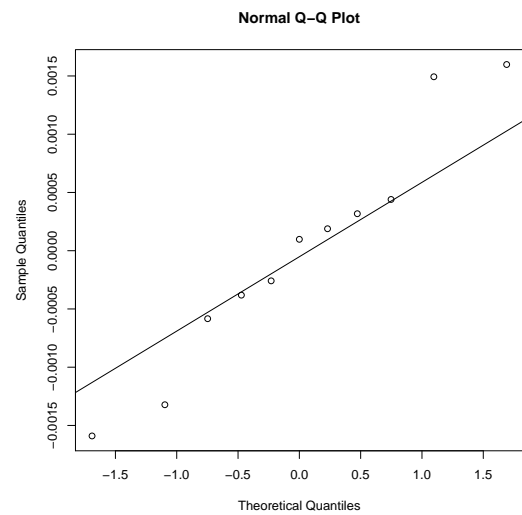
The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.92208, p-value = 0.3363
```

### ROI-5

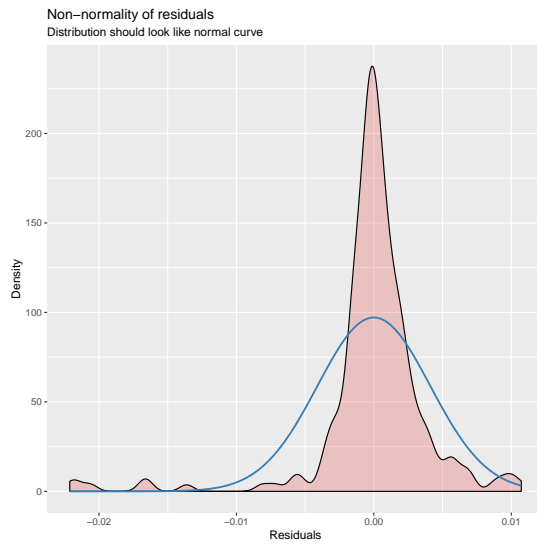


(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for ROI-5.

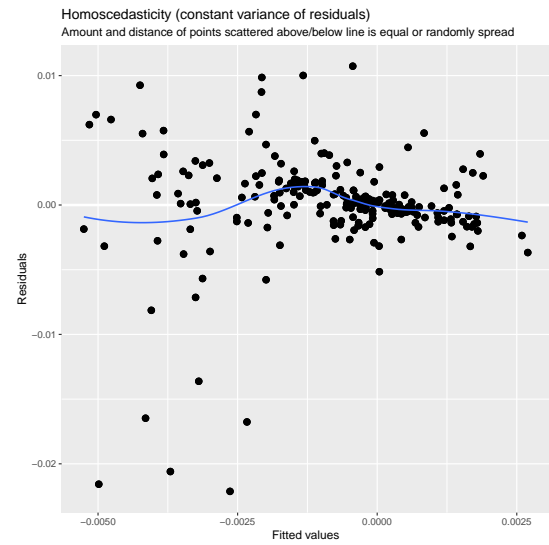


(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 174 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for ROI-5 and a theoretical normal distribution (blue).



(b) Representation of the fitted values of ROI-5 by the residuals of the model (black dots) and the regression line (blue).

Figure 175 – Graphical representation for normality and homoscedasticity verification.

The Shapiro-Wilk normality test for the residuals of the model is:

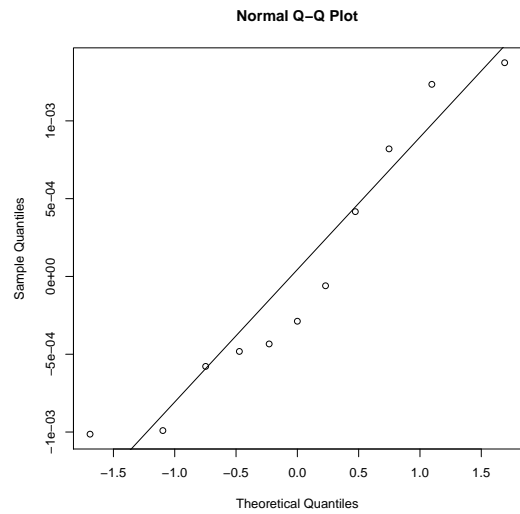
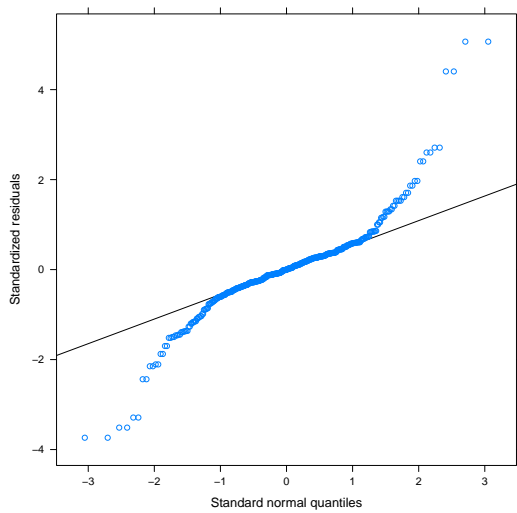
```
data: resid(fm)
W = 0.74553, p-value < 2.2e-16
```

The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.95408, p-value = 0.6963
```



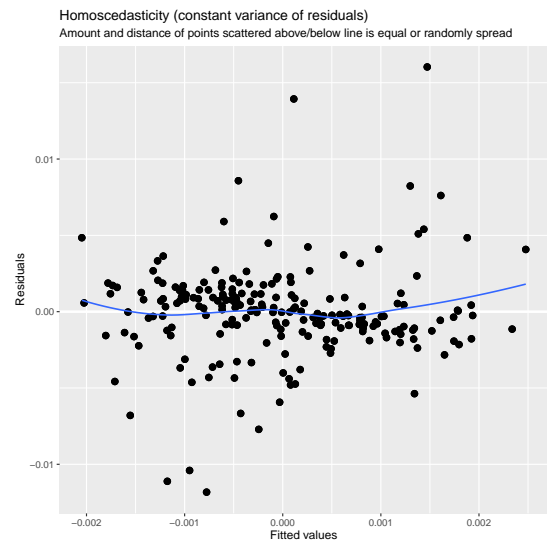
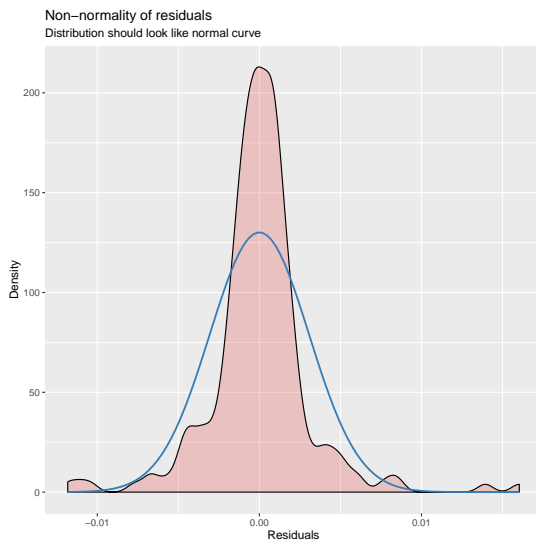
ROI-6



(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for ROI-6.

(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 176 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for ROI-6 and a theoretical normal distribution (blue).

(b) Representation of the fitted values of ROI-6 by the residuals of the model (black dots) and the regression line (blue).

Figure 177 – Graphical representation for normality and homoscedasticity verification.

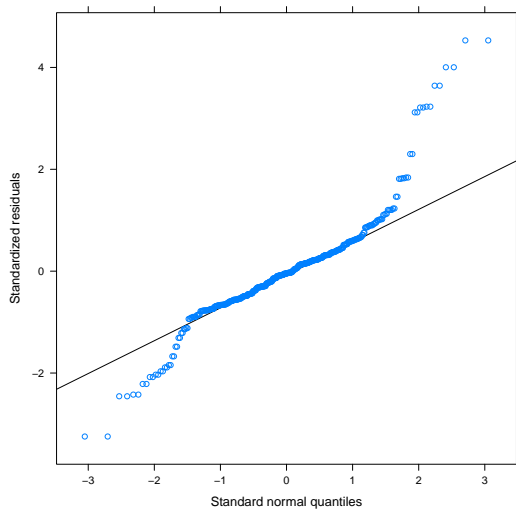
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.87913, p-value < 2.2e-16
```

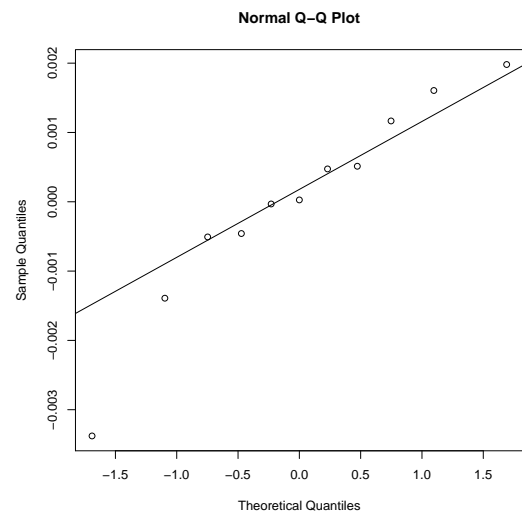
The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.91165, p-value = 0.2551
```

### ROI-7

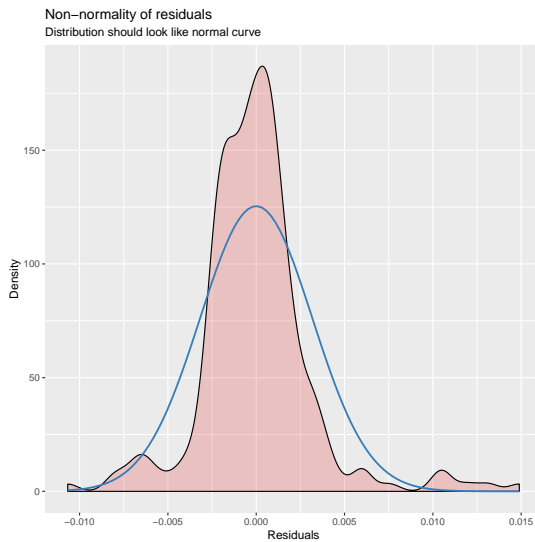


(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for ROI-7.

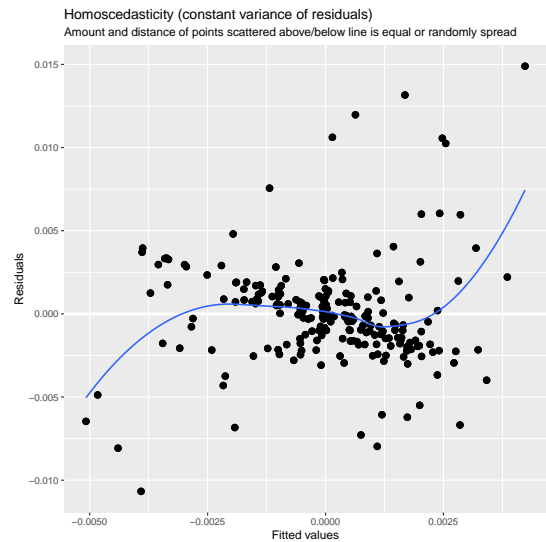


(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 178 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for ROI-7 and a theoretical normal distribution (blue).



(b) Representation of the fitted values of ROI-7 by the residuals of the model (black dots) and the regression line (blue).

Figure 179 – Graphical representation for normality and homoscedasticity verification.

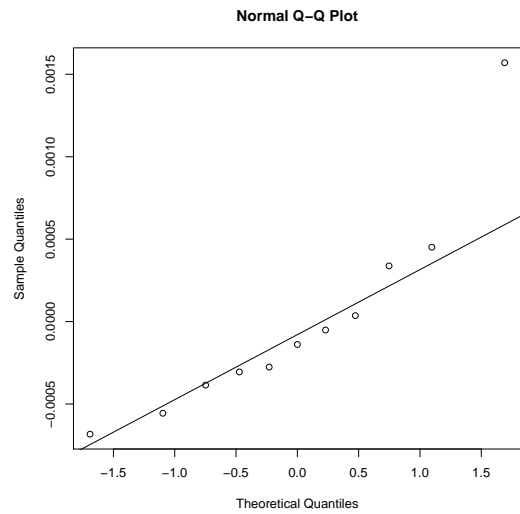
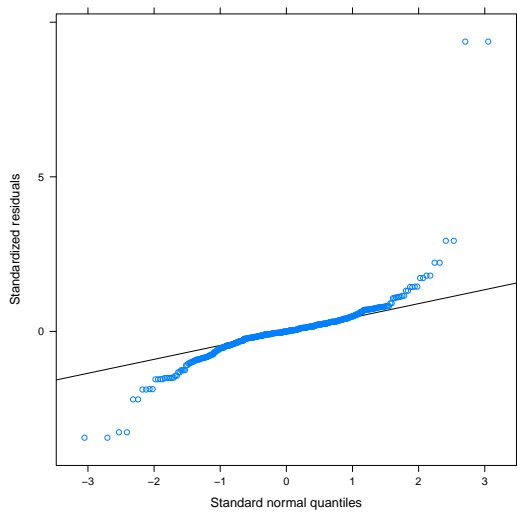
The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.88468, p-value < 2.2e-16
```

The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.93283, p-value = 0.4402
```

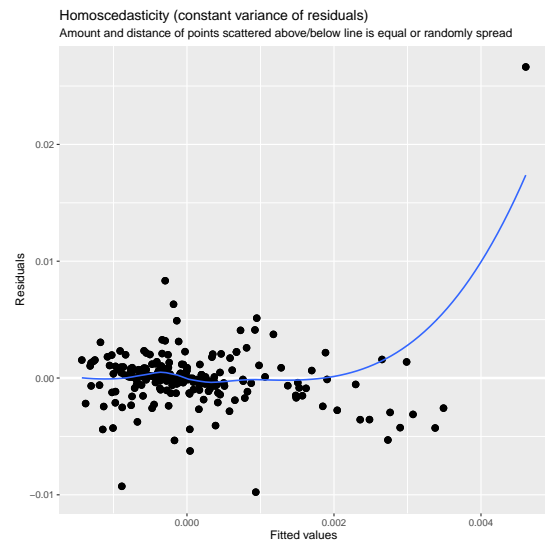
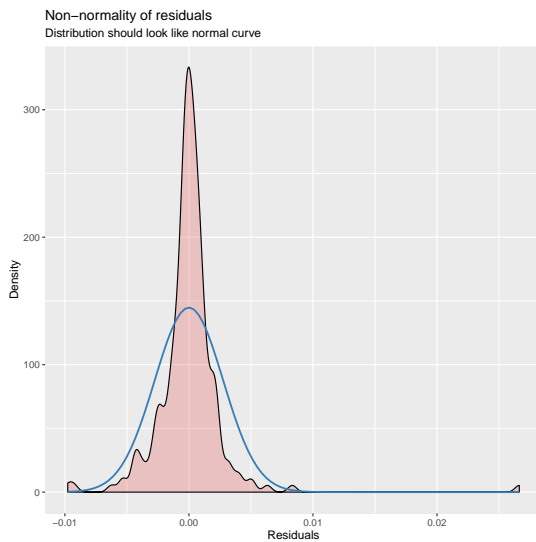
**ROI-8**



(a) Representation of the quantiles of the normal distribution by the quantiles of the standardized values of the residuals of the model for ROI-8.

(b) Representation of the quantiles of a theoretical normal distribution by the quantiles of the distribution of the random variable (subjects).

Figure 180 – Graphical representation for normality verification.



(a) Density distribution of the residuals (pink) for ROI-8 and a theoretical normal distribution (blue).

(b) Representation of the fitted values of ROI-8 by the residuals of the model (black dots) and the regression line (blue).

Figure 181 – Graphical representation for normality and homoscedasticity verification.

The Shapiro-Wilk normality test for the residuals of the model is:

```
data: resid(fm)
W = 0.70056, p-value < 2.2e-16
```

The Shapiro-Wilk normality test for the random variable is:

```
data: r
W = 0.85283, p-value = 0.04646
```

# APPENDIX G

## The need for averaging across trials

This appendix shows the importance of averaging trials in studies that use EEG as acquisition technique and how this, in general, leads to the HDLSS problems.

One of the central reasons why regularization or dimension reduction techniques, like [RoLDSIS](#), are useful for analyzing [EEG](#) data is the fact that the responses for each stimulus must be averaged across trials. Indeed, the number of observations obtained from such averaging is much smaller than the dimension of the feature space. We could speculate that, instead of using the grand average, one could consider each trial as an observation or even averaging ERPs across groups of trials, such that the number of observations would be greater than the dimension of the feature space. For instance, in our [EEG](#) experiment, if we compute the average of every four trials, we will end up with 150 to 250 observations (according to the participant), beyond the number of 128 wavelet features. In this case, the systems resulting from Eqs. [7.25](#) and [7.26](#) would be overdetermined and could be solved with traditional least squares regression. However, the [SNR](#) of the resulting ERPs decreases when the number trials per observation decreases and this could yield unreliable results.

In order to assess this issue, we computed the least squares linear regression varying the number of trials per observation. We did it for all participants. For a number of trials per observation greater than one, the trials were assigned at random to each observation. The results for the root mean square ([RMS](#)) regression errors for the  $\Phi$  and the  $\Psi$  axes across the population are summarized in [Figure 182](#). The  $y$  attributes of our stimuli vary between  $-52$  ms and  $+16$  ms in the  $\Phi$  case (stimuli 1 and 200 in the VOT continuum) and between 0 and 1 in the  $\Psi$  case. From the [Figure](#), one can see that the [RMS](#) regression errors are relatively high when there is one trial per observation, with the population mean being 19.1 ms in the  $\Phi$  case and 0.35 in the  $\Psi$  case.

When the number of trials per observation increases, the **RMS** decreases almost linearly towards zero, a value that is theoretically attained when the number of observations is less than 128.

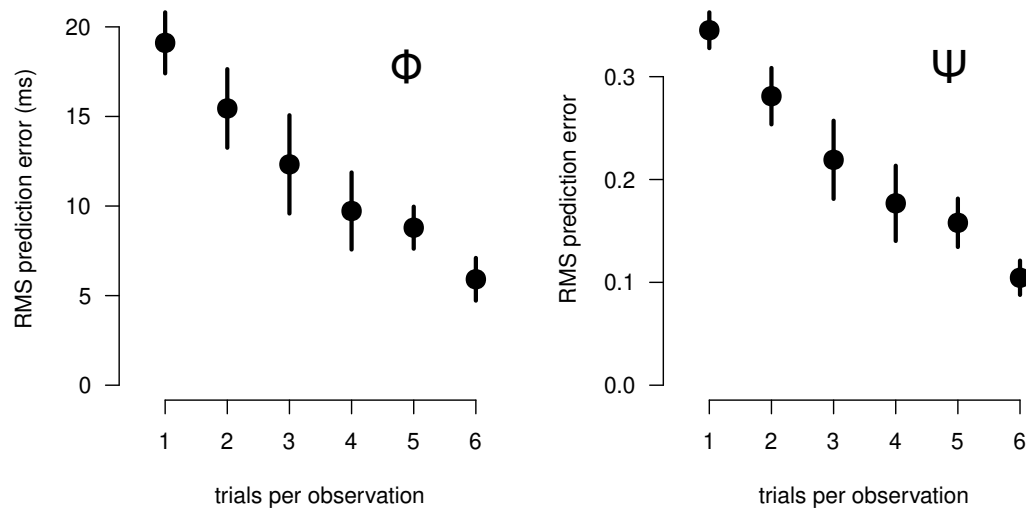


Figure 182 – Prediction error of linear regression for overdetermined cases. Traditional least squares regression applied to the linear model relating ERP feature vectors and either physical (left) or psychophysical (right) attributes. The RMS prediction error is shown in the vertical axis. The number of trials per observation, varying from 1 to 6, is shown in the horizontal axis. Dots and vertical bars represent, respectively, the means and standard deviations obtained for the 11 participants.

**ERP** observations are scarce because hundreds of trials are necessary to obtain a single observation. A typical trial in the experiments carried out has an **SNR** of approximately -15 dB, which is excessively low for any useful analysis. In order to attain an acceptable **SNR** level, we averaged the 200 trials (a little less when corrupted trials were discarded) raising the **SNR** by  $10\log_{10} 200 = 23$  dB, resulting in a signal whose **SNR** is around 8 dB. The recording of 1000 trials (5 stimuli, 200 trials per stimulus) took 35 minutes, which were part of a longer session in which other measurements were made. Thus, in order to avoid the influence of fatigue, the number of trials per stimulus is limited.

An **SNR** of 8 dB is barely sufficient for the analysis of **ERP** signals, hence the need for grand-averaging. The use of single trials or of a small amount of trials per observation to solve an overdetermined linear system by least squares regression yields large mean squared errors, as shown in Fig. 182.

Regularization, such as **LASSO**, ridge regression or **SPLS** can be used when the number of trials per observation results in fewer observations than the number of features. However, those techniques usually need the definition of parameters that require the use of some kind of **CV** which, in turn, requires trials to be spared for validation (see Fig. 109).

In order to maximize the SNR of each ERP observation, we need to average as many trials as possible. In this case, projection techniques are a more suitable solution than regularization. RoLDSIS adapts the PCA regression (principal component regression (PCR)) described in Tibshirani (1996) to the specificities of ERP analysis.



# APPENDIX H

## Psychometric curves

This appendix contain the psychometric curves for all the 11 subjects who participated in the experiments of this work. For each subject one psychometric curve was obtained for each continuum: **VOT** and Formants continuum.

Following it is shown the 11 psychometric curves for the **VOT** continuum. The five stimuli selected for the passive and active experiments for each participant and continuum are indicated in the psychometric curves. Their VOT values are indicated in the abscissa axis.

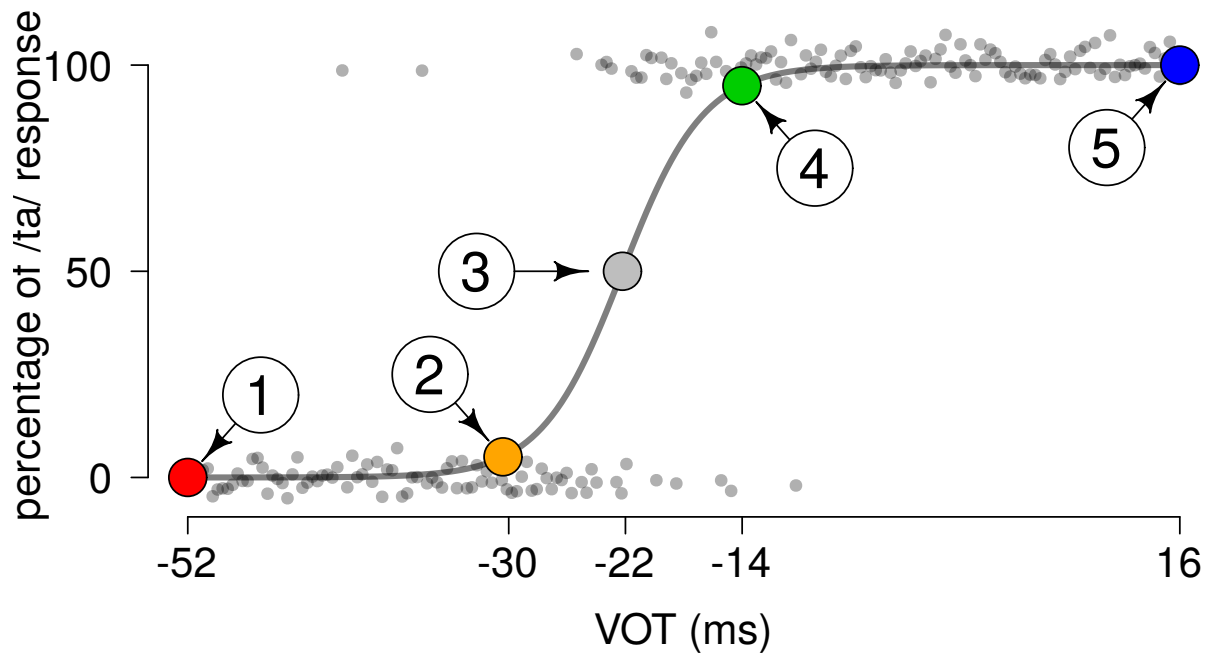
**Participant 1**

Figure 183 – Psychometric curve for the participant 1 for the VOT continuum.

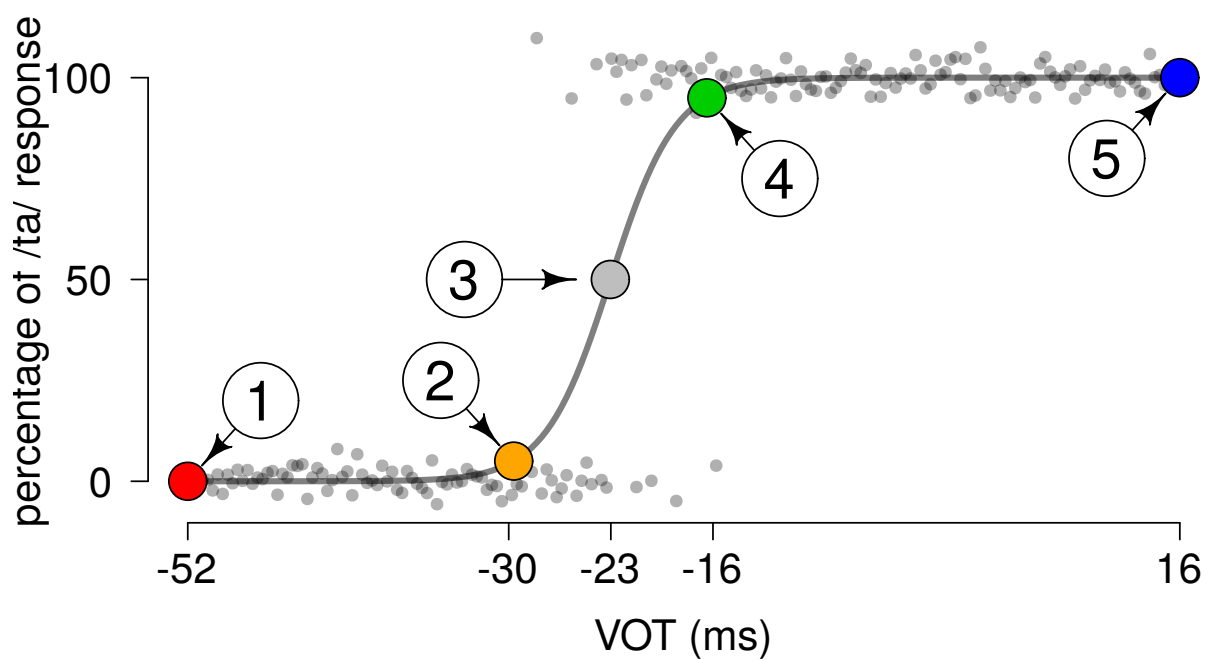
**Participant 2**

Figure 184 – Psychometric curve for the participant 2 for the VOT continuum.

**Participant 3**

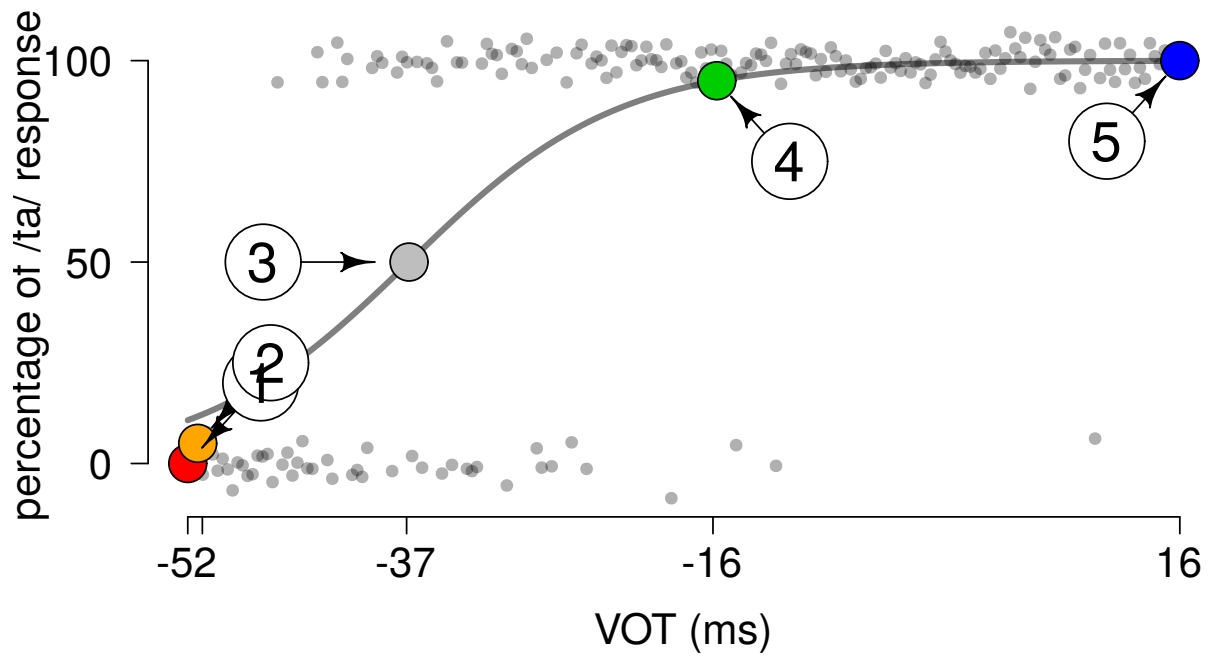


Figure 185 – Psychometric curve for the participant 3 for the VOT continuum.

**Participant 4**

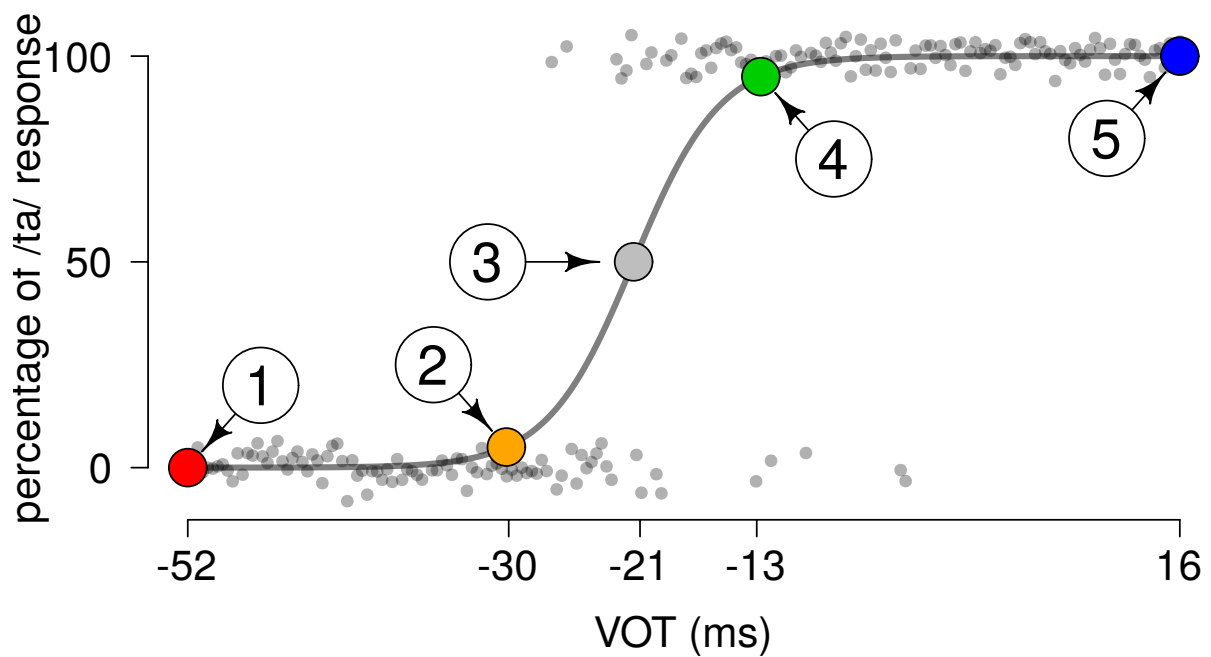


Figure 186 – Psychometric curve for the participant 4 for the VOT continuum.

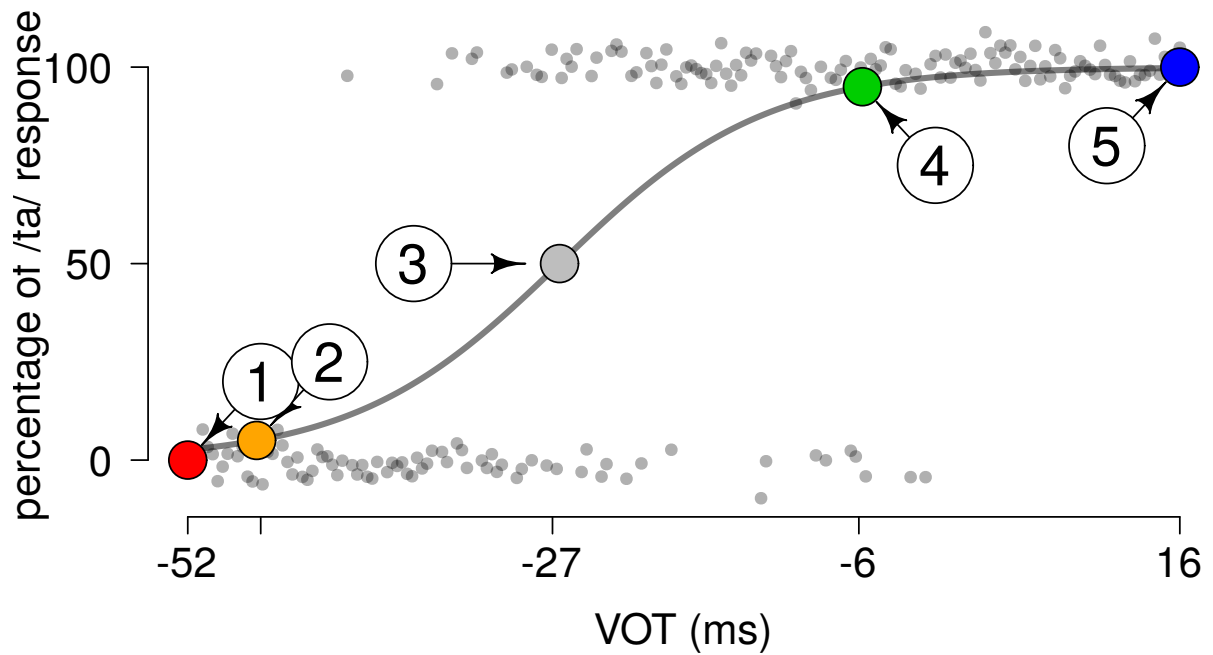
**Participant 5**

Figure 187 – Psychometric curve for the participant 5 for the VOT continuum.

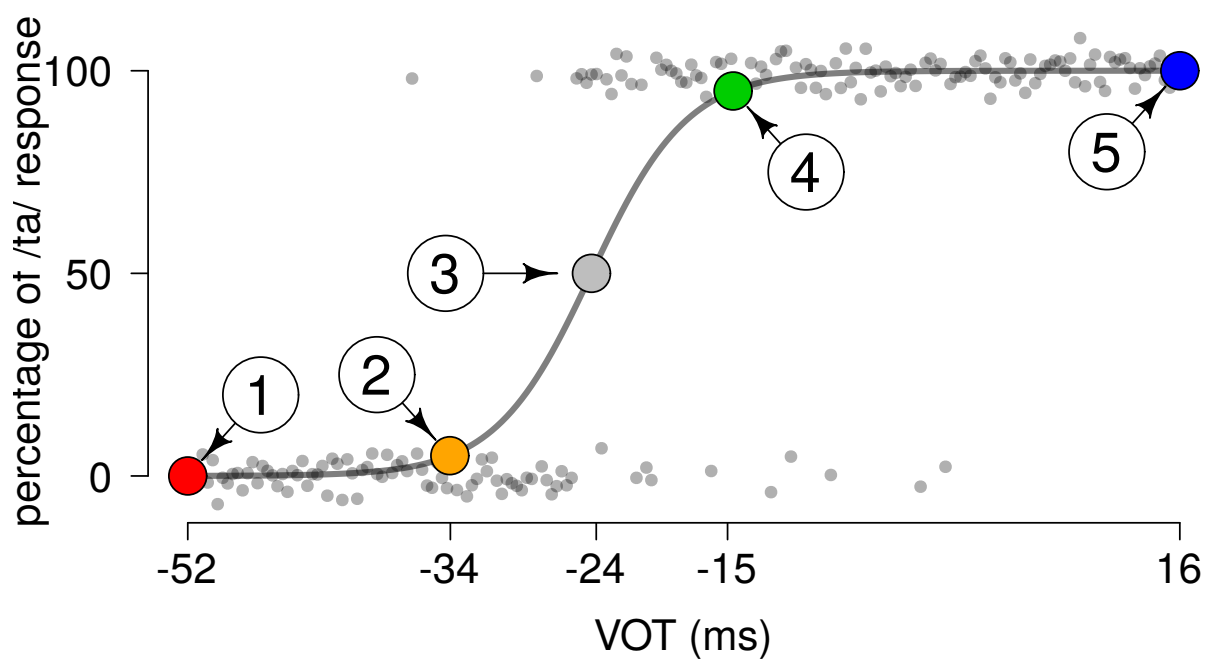
**Participant 6**

Figure 188 – Psychometric curve for the participant 6 for the VOT continuum.

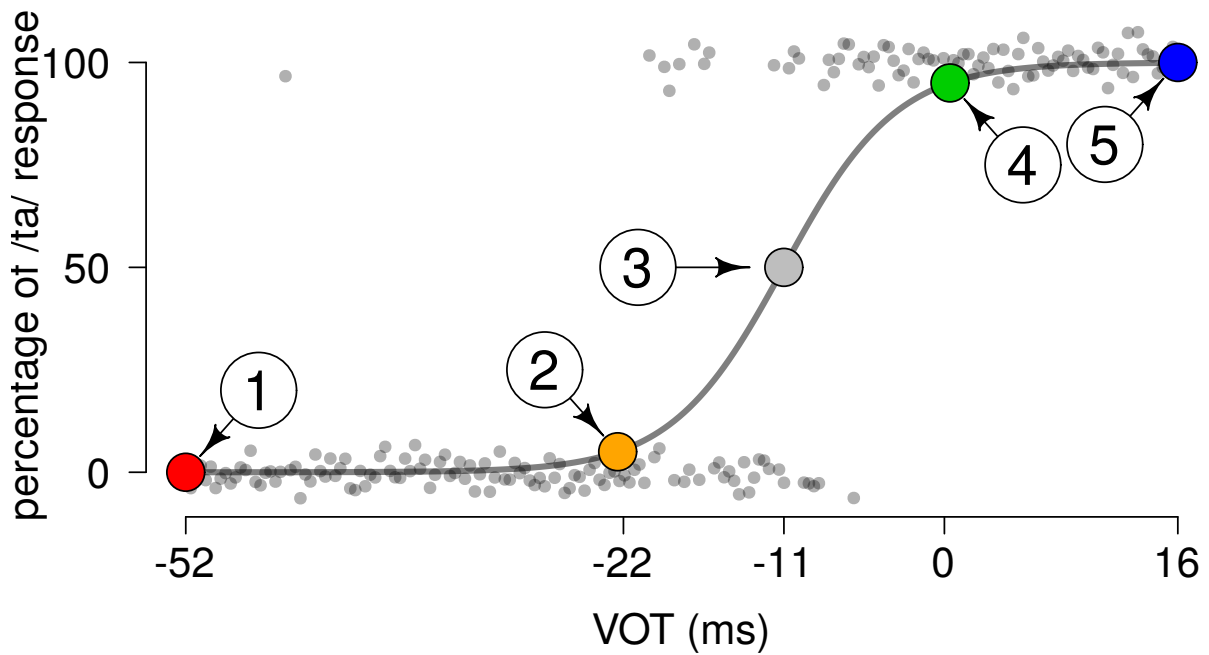
**Participant 7**

Figure 189 – Psychometric curve for the participant 7 for the VOT continuum.

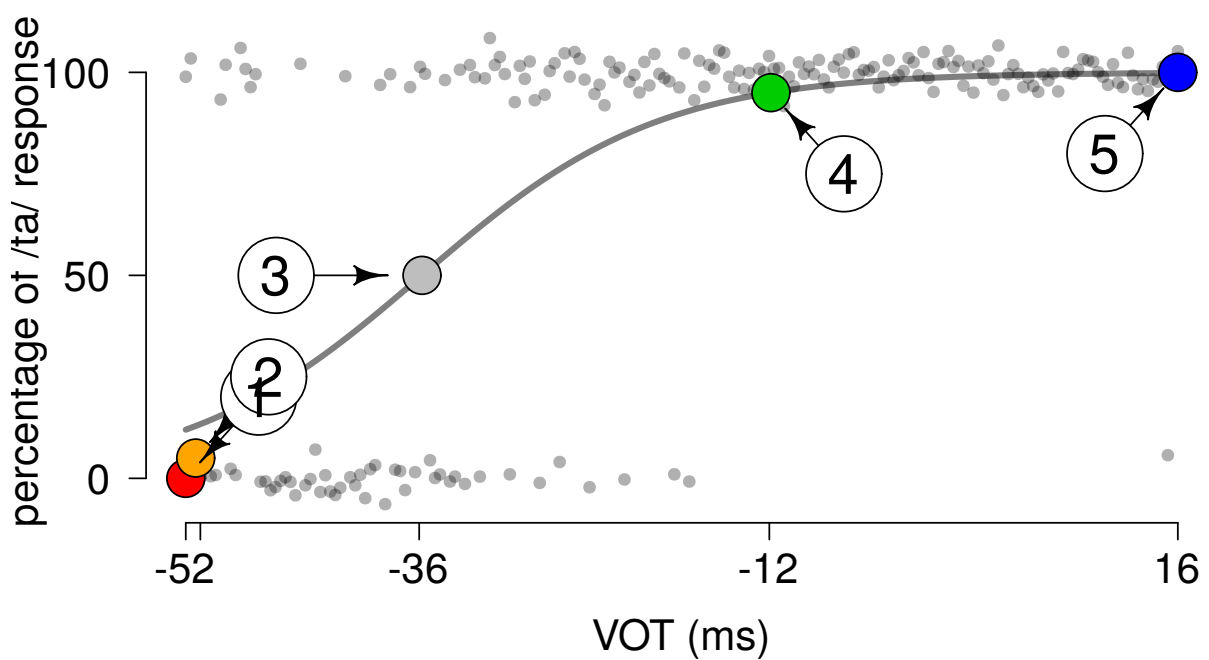
**Participant 8**

Figure 190 – Psychometric curve for the participant 8 for the VOT continuum.

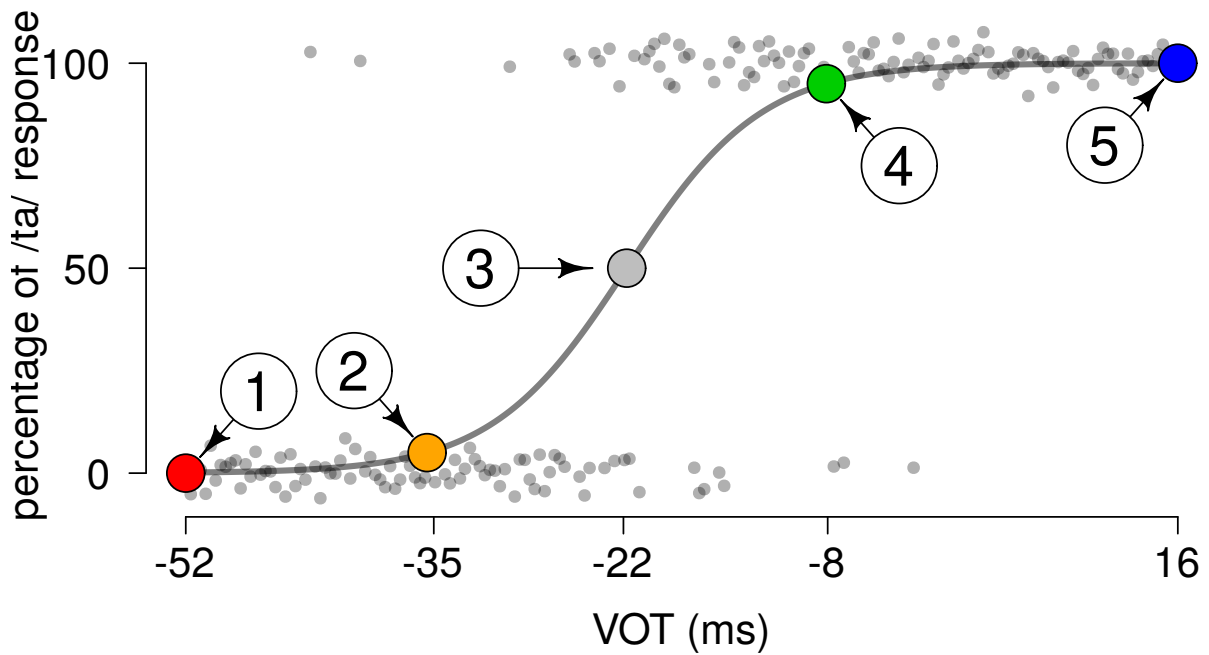
**Participant 9**

Figure 191 – Psychometric curve for the participant 9 for the VOT continuum.

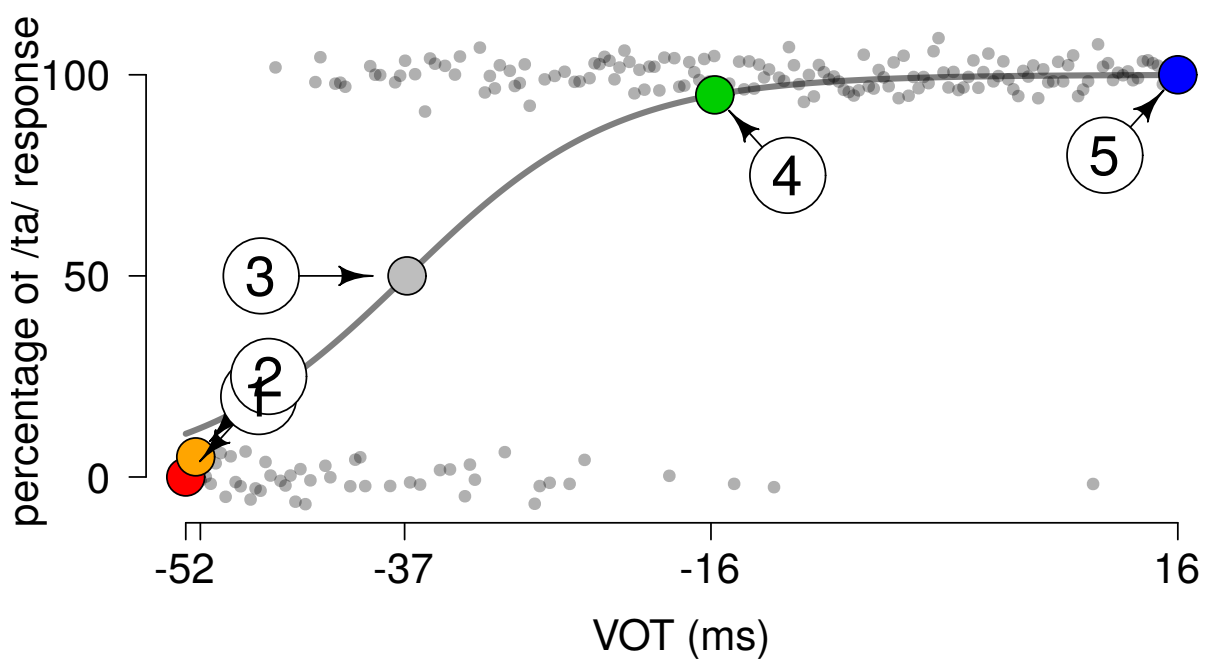
**Participant 10**

Figure 192 – Psychometric curve for the participant 10 for the VOT continuum.

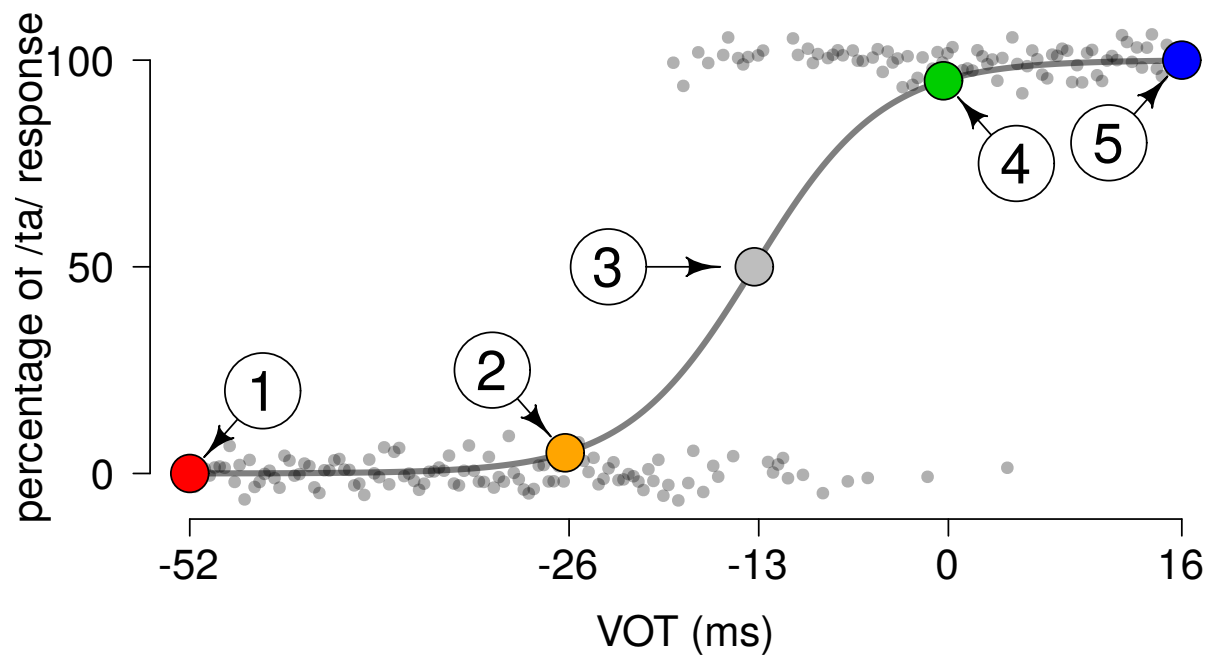
**Participant 11**

Figure 193 – Psychometric curve for the participant 11 for the VOT continuum.

Following it is shown the 11 psychometric curves for the Formants continuum. The five stimuli selected for the passive and active experiments for each participant and continuum are indicated in the psychometric curves. The difference between the second and first formant frequency values are indicated in the abscissa axis.

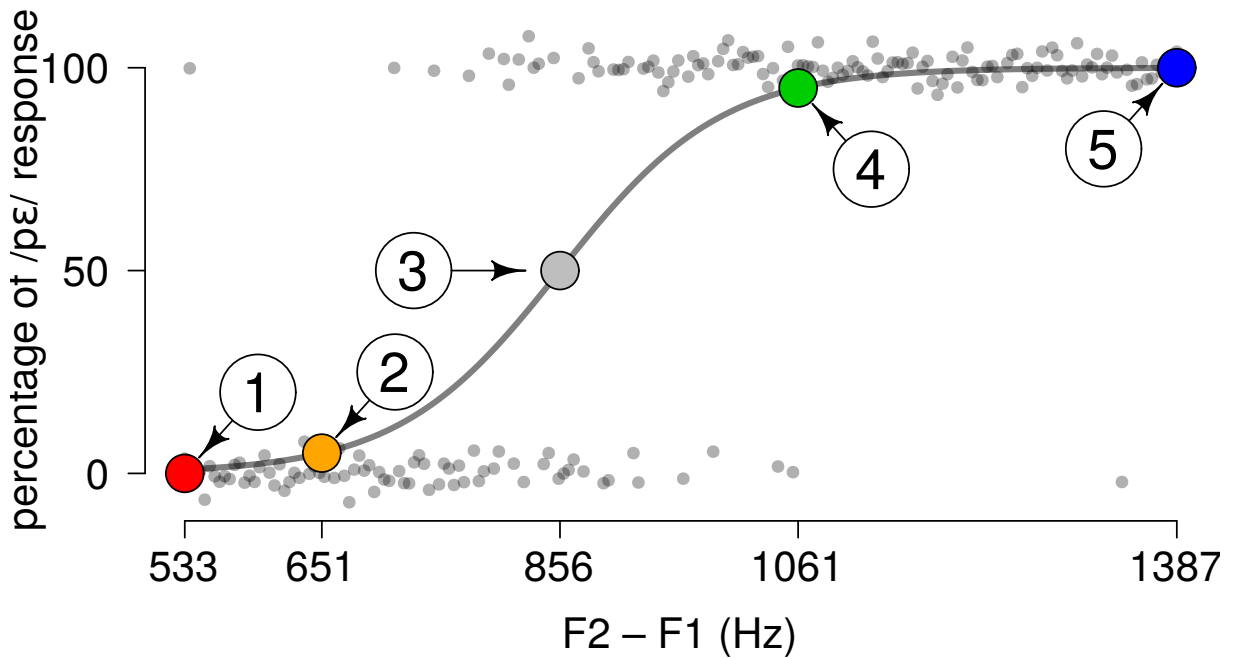
**Participant 1**

Figure 194 – Psychometric curve for the participant 1 for the Formant continuum.

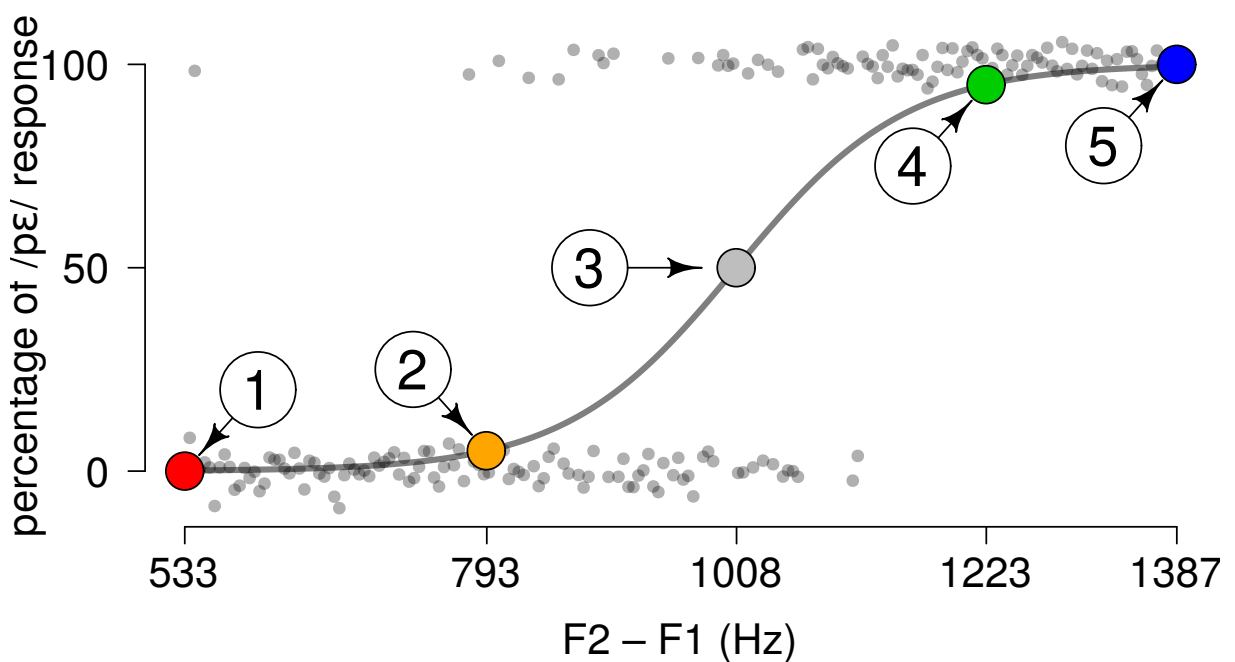
**Participant 2**

Figure 195 – Psychometric curve for the participant 2 for the Formant continuum.



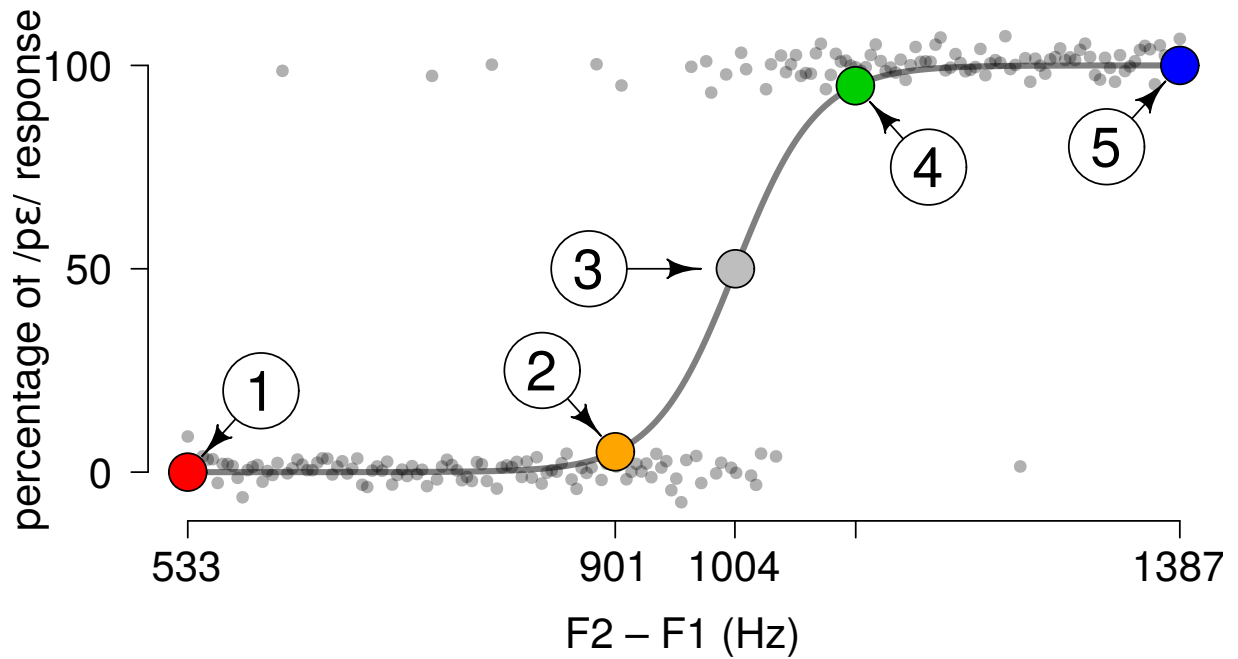
**Participant 3**

Figure 196 – Psychometric curve for the participant 3 for the Formant continuum.

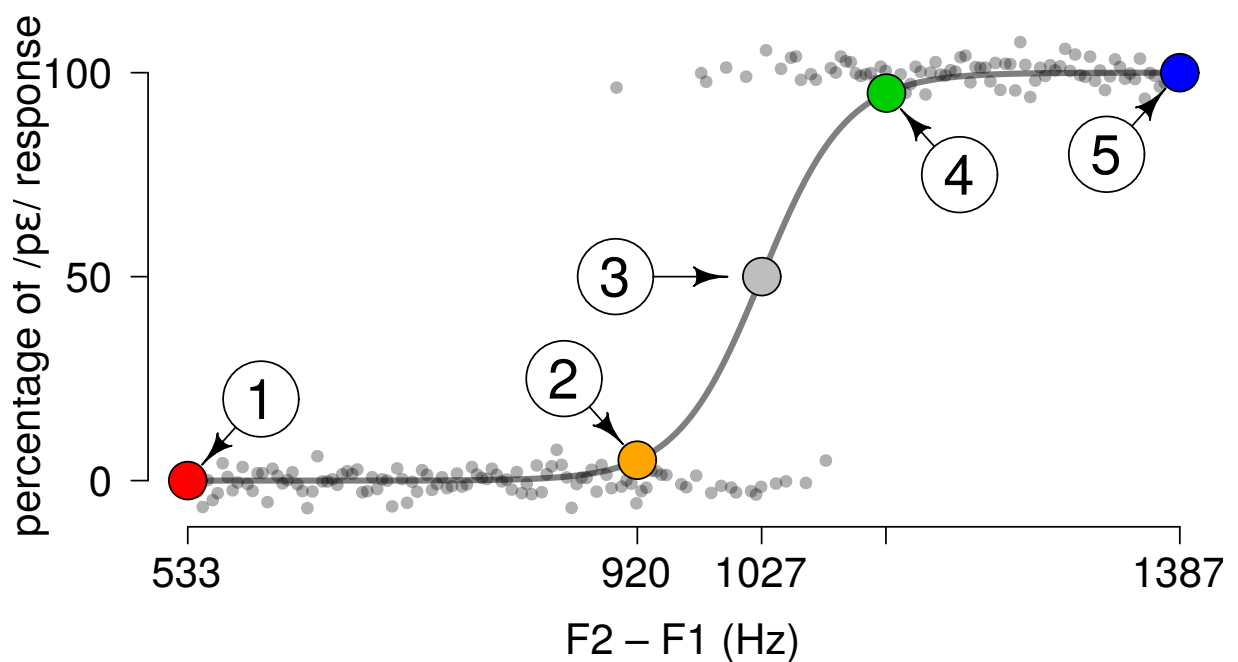
**Participant 4**

Figure 197 – Psychometric curve for the participant 4 for the Formant continuum.

**Participant 5**

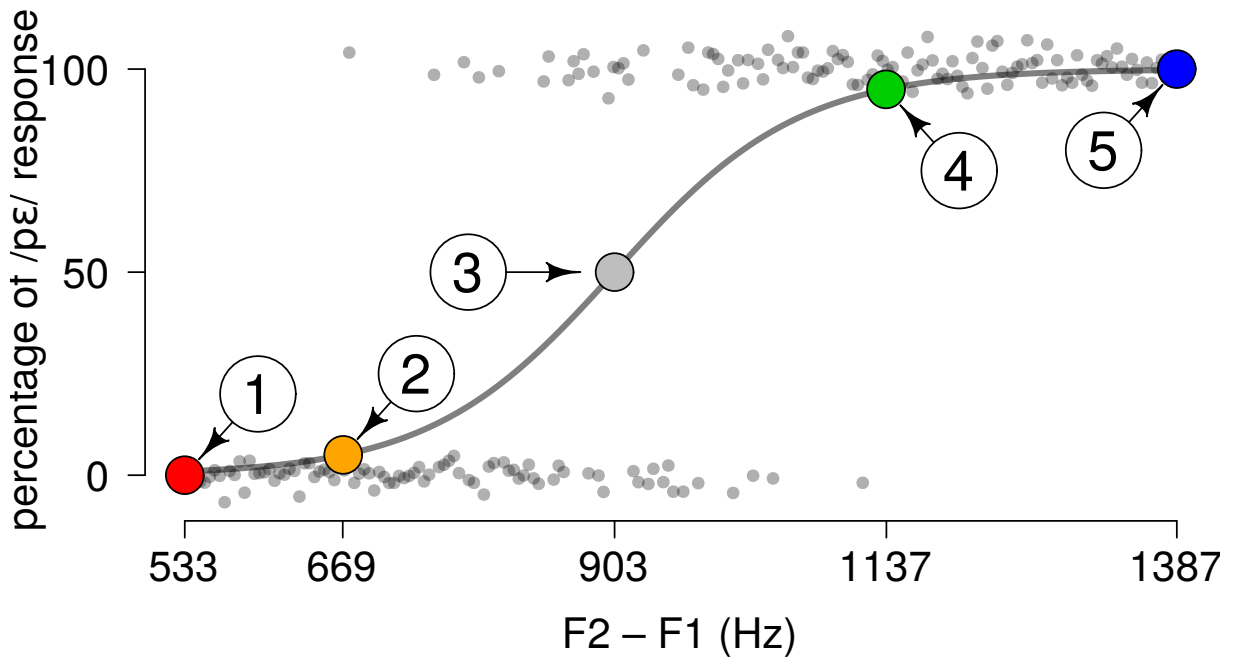


Figure 198 – Psychometric curve for the participant 5 for the Formant continuum.

**Participant 6**

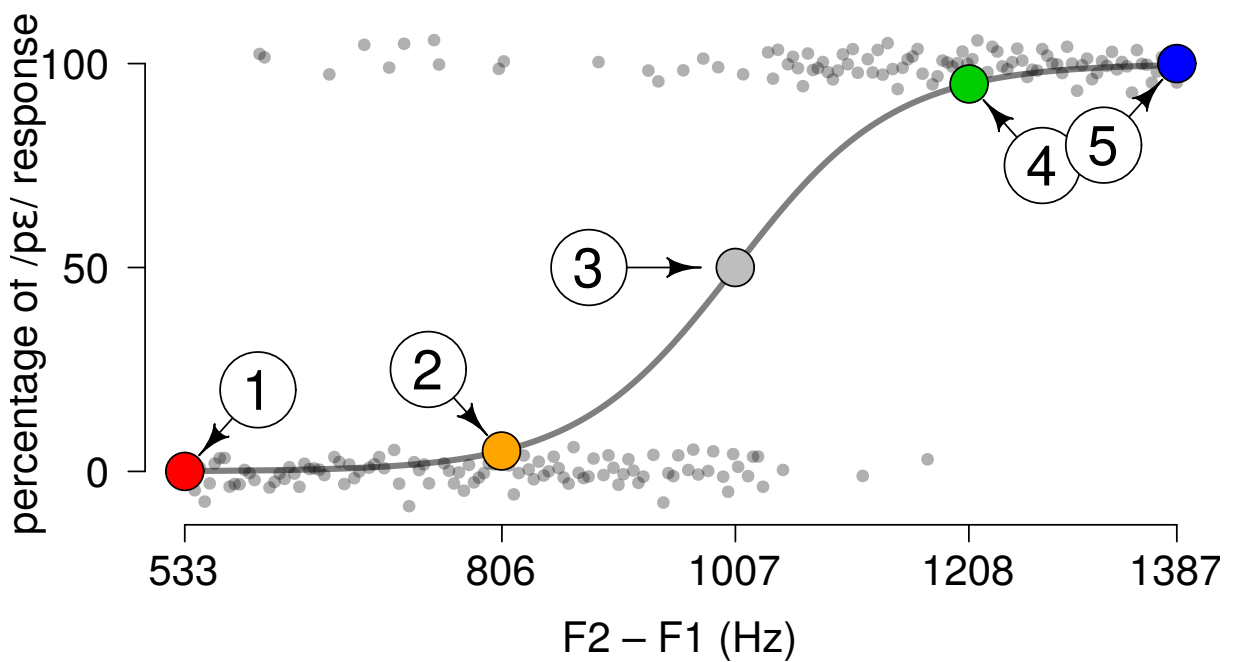


Figure 199 – Psychometric curve for the participant 6 for the Formant continuum.

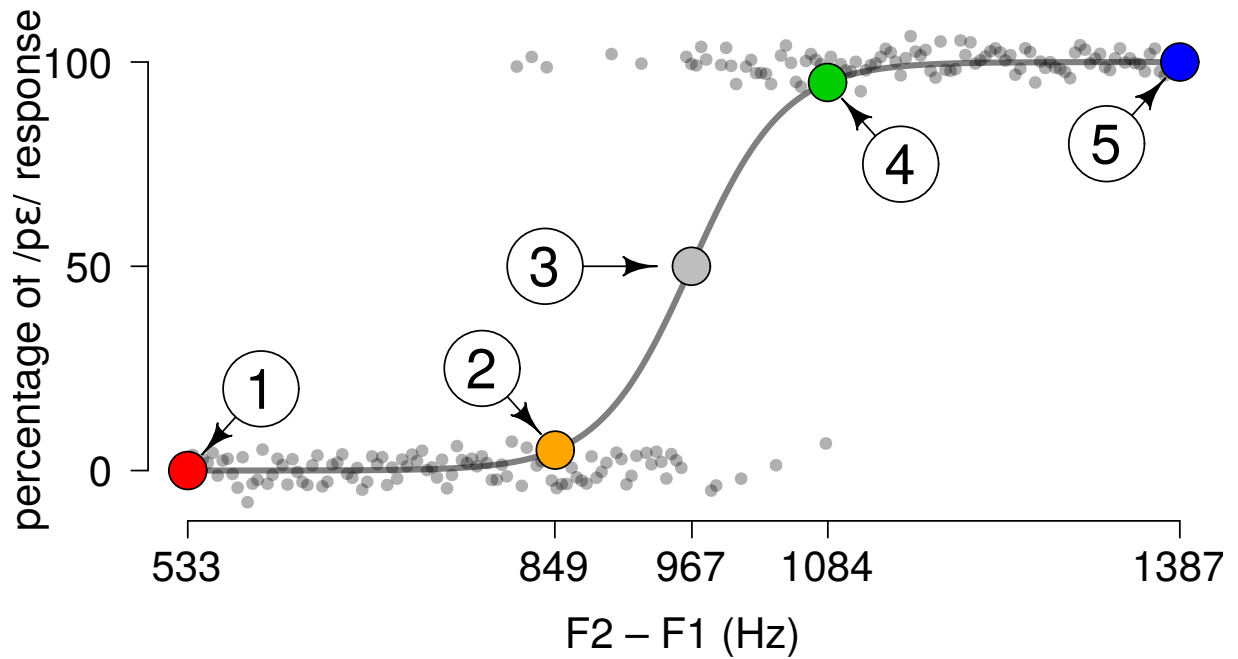
**Participant 7**

Figure 200 – Psychometric curve for the participant 7 for the Formant continuum.

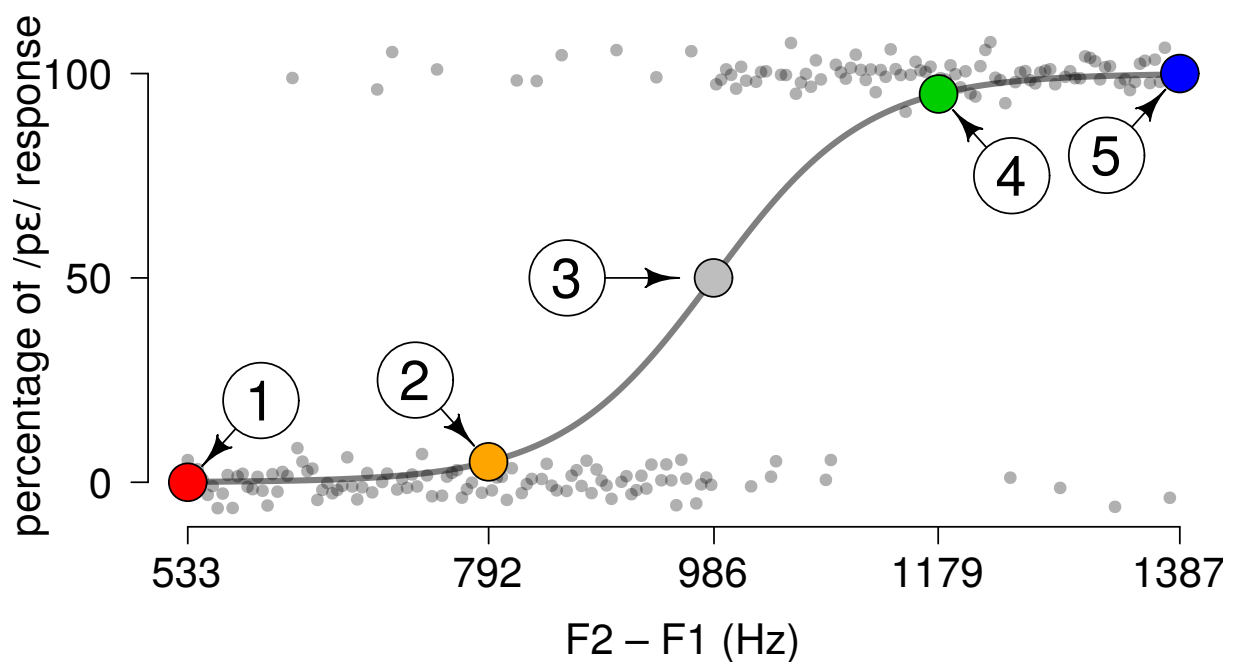
**Participant 8**

Figure 201 – Psychometric curve for the participant 8 for the Formant continuum.

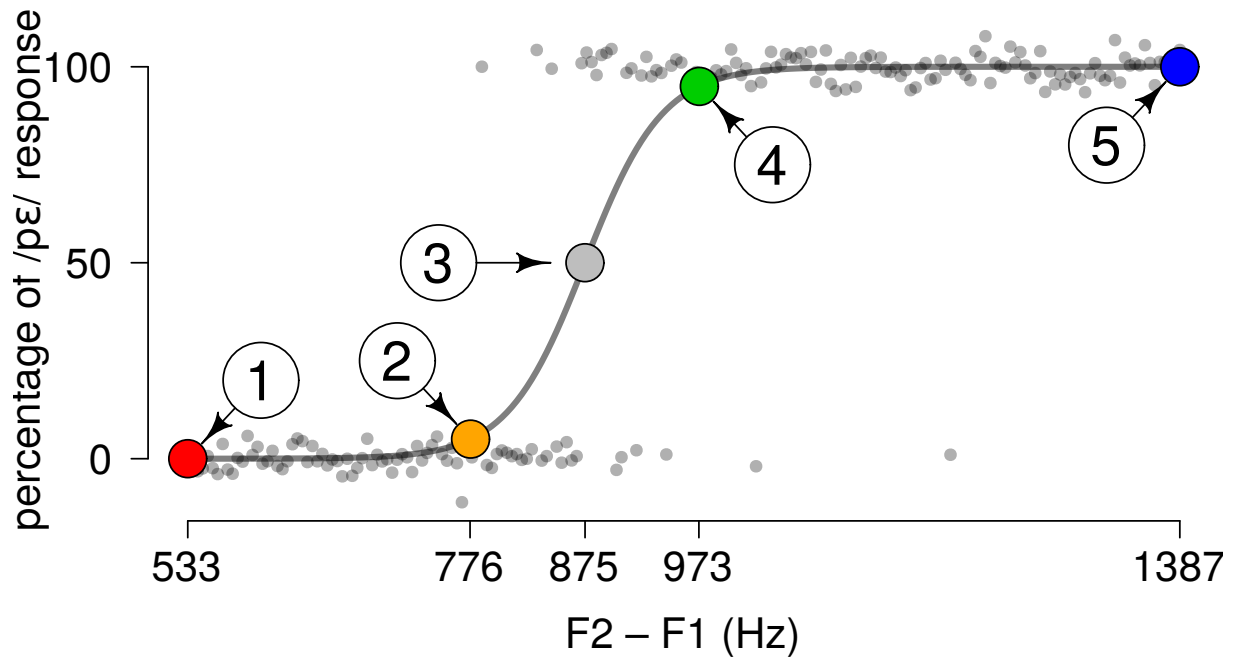
**Participant 9**

Figure 202 – Psychometric curve for the participant 9 for the Formant continuum.

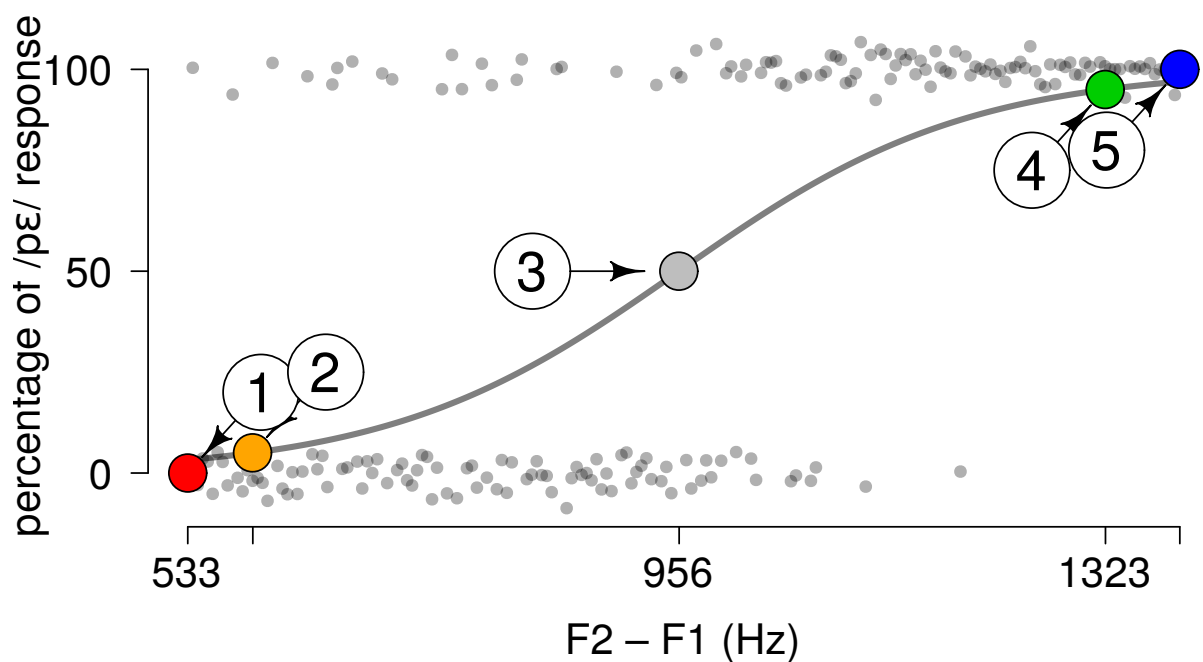
**Participant 10**

Figure 203 – Psychometric curve for the participant 10 for the Formant continuum.

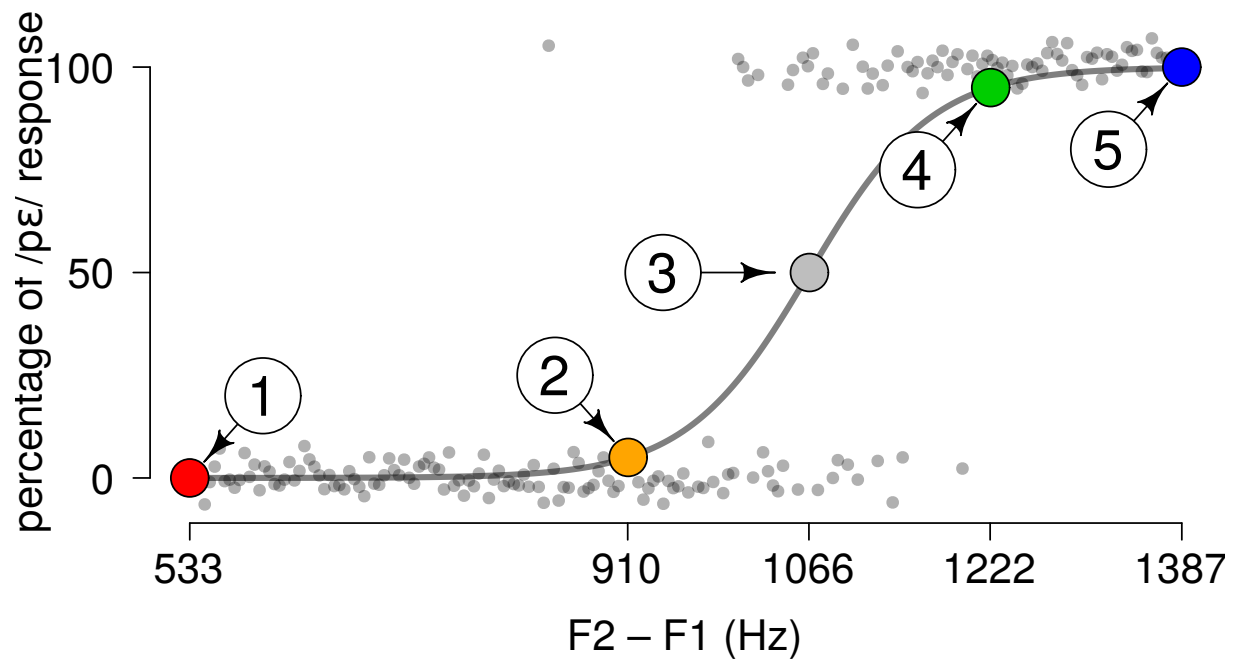
**Participant 11**

Figure 204 – Psychometric curve for the participant 11 for the Formant continuum.

# APPENDIX I

## Scripts

Following are described the scripts used in this work. They are available at the repositories <https://github.com/Adrielle-Santana/ThesisScripts> and <https://github.com/RoLDSIS/code>.

- **read\_Intan\_RHD2000\_file2.m**: function provided by the RHD2000 developer to help in the extraction and organization of the EEG data from the EEG .rhd file. A small alteration was included in this script to keep it from applying its own notch filter.
- **identification.py**: using the library `expyriment`, it implements the data acquisition experiment to survey the psychometric curve.
- **plot-curve.r**: uses the `glmrob` library to fit the psychometric curve with outlier treatment using the data obtained with `identification.py`. Generates .dat file with fit information including slope of psychometric curve and index of stimuli present at positions 0%, 5%, 50%, 95% and 100% for the `passive.py` and `active.py`.
- **passive.py**: controls the random presentation of the stimuli and the jitter in the passive task (using the 5 stimuli selected by `plot-curve.r`) and generates a .pickle extension file with information of which stimuli were used and the order of presentation.
- **active.py**: controls the random presentation of the stimuli and the jitter in the active task (using the 5 stimuli selected by `plot-curve.r`) and captures the answers given through the keyboard generating a .pickle file with information of the presented stimuli, the order of presentation, key pressed on the keyboard in response to each stimulus and the response time.

- **idx.py**: reads the .pickle extension file generated in passive.py and active.py and extracts from each the vector containing the stimulus presentation order in each case generating a .mat extension file for each one.
- **notch60.m**: function that applies the 60Hz notch filter to the signal.
- **Create\_mats\_ativ.m**: reads the .rhd extension file with the EEG signals recorded in the active task, reads the generated idx.py file corresponding to that task, deletes trials with higher than average amplitudes (some those artifacts were also detected visually), and organizes some structures saved with the extension .mat. Those structures were then used in all the processing performed at the R software. One structure contains, for each EEG channel, all the trials for each given stimulus, taking 5000 samples for each trial (including 150ms prestimulus). Each trial is 60Hz notch filtered. Another set of matrices are generated only for the signal of the  $Cz - Tp9$  and  $Cz - Tp10$  electrodes (left and right temporal respectively). Finally, a last set of matrices similar to the latter is generated, with the difference that the signals are baseline corrected and therefore no prestimulus samples are considered in 5000. This script was applied and the matrices generated for each subject and each of the two continua.
- **Create\_mats\_pass.m**: similar to the previous script, now applied to the EEG recorded in the passive task and using its idx.py generated file for that task.
- **resp\_time.m**: compute and plot the average response time of all subjects in the active task using the information recorded in the .pickle file at active.py.
- **compute-dwt-coefs.r**: perform the resampling of the time-domain signal, the baseline correction, compute the average of the corrected signal in the time-domain and compute the discrete wavelet transform organizing the transformed epochs in structures to be used by other scripts.
- **eeg-to-dwt.r**: organize the call of the *compute-dwt-coefs.r* across the experimental conditions.
- **paths.r**: Definition of paths for saving figures and results.
- **parameters.r**: Definition of DWT parameters and other common values used in several scripts.
- **n1-p2-bootstrap.r**: perform the bootstrap analysis for the N1 and P2 peaks to identify the more significant ones across the data of the subjects, stimuli, electrode, feature and type of task. A general figure is generated illustrating all the results.
- **n1-p2-analysis.r**: perform the contrasts analysis for the N1, P2, T1, T2 and N1-P2 values obtained after the bootstrap analysis.

- **geom-lib.r**: Geometry supporting function.
- **dwt-lib.r**: [DWT](#) supporting functions.
- **scalogram.r**: Plot the [DWT](#) scalogram in black and white or blue-white-red color scale.
- **remove-spikes.r**: Remove spikes from signals, using [DWT](#).
- **chisq-to-normal.r**: Convert Chi-square to normal distribution.
- **cross-validation.r**: Basic cross-validation function.
- **compare-methods.r**: Compare regression methods tested to validate the [RoLDSIS](#).
- **run-cv.r**: Main script for running the cross-validation.
- **contrasts.r**: Utility functions for generating contrasts for multiple comparison analysis through mixed-effects models.
- **energy-analysis.r**: Analyses of regression directions obtained from [RoLDSIS](#) using mixed-effects models and contrast analysis.
- **experiments.r**: contain information about features and type of tasks used in other scripts.
- **roldsis.r**: execute the [RoLDSIS](#) technique.
- **roldsis-bootstrap.r**: basic functions for running bootstrap with [RoLDSIS](#).
- **roldsis-coefficients.r**: organize the execution of the [RoLDSIS](#) for subjects, electrodes, regression type, feature and type of task. It also perform the bootstrap [RoLDSIS](#) generating the data for the bootstrap analysis of the technique.
- **pipeline.r**: call the scripts in the correct order to obtain the [RoLDSIS](#) coefficients and their contrast analysis.
- **psy-phy-analysis.r**: generate scalograms with histograms for the regression coefficients obtained for each experimental condition, organized by electrode.
- **figure-cv-scalogram.r**: scalograms with histograms for the regression methods of regressed coefficients.
- **figure-cv-errors.r**: cross-validation errors for [RoLDSIS](#), [LASSO](#), Ridge Regression, and [SPLS](#).
- **figures-mean-erp.r**: Average ERPs for the five stimuli for each subject.
- **figures-roldsis-results.r**: [RoLDSIS](#) result represented in scalogram for each subject and also [RoLDSIS](#) projected responses for each subject. This script also compute and generate the figures with the angles vs. slope analysis.



- 
- **figures-id-exp.r**: Plot results of phonemic identification experiment for each subject (psychometric curve) and obtain the slope of the psychometric curve for each subject and feature used for the angles vs. slope analysis.
  - **figure-trials-observation.r**: **RMS** prediction error for all subjects using different number of averaged points.
  - **figures-roldsis-boot.r**: Generate figure of the **RoLDSIS** bootstrap analysis showing the separation between physical and psychophysical points through **LDA**.
  - **directional.r**: Support functions used in *figures-roldsis-boot.r*.