Alexandre José Duarte Amador Oliveira

AUDITORY ATTENTION TO SPEECH DURING CONTINUOUS LINGUISTIC TASKS: AN EEG STUDY

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências Humanas e Sociais

2023

Alexandre José Duarte Amador Oliveira

AUDITORY ATTENTION TO SPEECH DURING CONTINUOUS LINGUISTIC TASKS: AN EEG STUDY

Mestrado em Neurociências Cognitivas e Neuropsicologia

Trabalho efetuado sob a orientação de:

Professor e Investigador Doutorado João Correia

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências Humanas e Sociais

2023

AUDITORY ATTENTION TO SPEECH DURING CONTINUOUS LINGUISTIC TASKS: AN EEG STUDY

Declaração da autoria de trabalho

Declaro ser autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Assinatura:

____

(Alexandre José Duarte Amador Oliveira)

“Da minha aldeia vejo quanto da terra se pode ver no Universo...
Por isso a minha aldeia é tão grande como outra terra qualquer
Porque eu sou do tamanho do que vejo
E não do tamanho da minha altura...”

Alberto Caeiro

Agradecimentos

Depois de um ano repleto de mudança, de novas experiências e de muito trabalho, esforço, dedicação e superação, chegou finalmente o momento de dar por concluída mais uma etapa importante do meu percurso académico e pessoal. Porque este trabalho não é somente meu, mas também de todos aqueles que me acompanharam de perto ao longo deste processo, deixo aqui o meu especial agradecimento.

Em primeiro lugar, gostaria de agradecer ao Prof. João Correia por ter aceitado supervisionar a minha tese de mestrado e acompanhado progressivamente o meu trabalho. Por partilhar comigo a sua experiência e conhecimentos. Pelas aprendizagens proporcionadas, pela exigência e excelência que procura diariamente no seu trabalho, mas também pela confiança e suporte que me transmitiu. Por ter sido incansável e ter estado sempre disponível para responder às minhas questões, dúvidas e preocupações.

A todo o grupo de Neurociências Cognitivas da Universidade do Algarve, agradecer pela oportunidade de partilhar convosco o projeto, pelos debates e discussão de ideias, permitindo clarificar determinados tópicos para o trabalho final. Agradecer também pelo espaço e equipamento disponibilizado para a realização da recolha de dados.

Aos meus colegas de mestrado, não só pela partilha das suas experiências e projetos, mas também das suas inquietações e ansiedades. Ao Daniel e Guilherme pelas sessões de trabalho em grupo e pela companhia e motivação.

Um especial agradecimento à Filipa, por ter acompanhado incessantemente a minha evolução e o meu trabalho durante esta etapa da minha vida. Pela compreensão e motivação essenciais e incondicionais. Pelo reconhecimento do esforço e do trabalho investidos. Aos meus pais e irmão por suportarem desde sempre o meu percurso académico, e apoiarem as minhas decisões ao longo desta etapa. Aos meus amigos, pelos momentos de lazer e pelos risos. A todos, muito obrigado.

Resumo em Português

A aquisição da linguagem é um fenómeno inato, universal e que representa uma das tarefas mais importantes do desenvolvimento humano. O córtex auditivo, por sua vez, desempenha um papel fundamental para a aquisição e desenvolvimento da linguagem durante a infância, papel este que se estende até à idade adulta, onde se constitui como elemento central na rede de processamento linguístico. Vários estudos têm demonstrado que, durante tarefas contínuas de perceção de linguagem, as oscilações neuronais geradas no córtex auditivo, têm a capacidade de captar e sincronizar-se com estímulos auditivos externos, fenómeno este denominado "*Cortical Entrainment"*. Este fenómeno está intimamente relacionado com o processamento contínuo da fala, ao ser responsável por facilitar a sua perceção e compreensão, especialmente durante condições mais ecológicas, que compreendam o processamento de estímulos contínuos. Dado que o processamento da linguagem, no quotidiano, ocorre de forma contínua, estudos de Eletroencefalograma (EEG) mais recentes, têm vindo a utilizar medidas de *Cortical Entrainment* para complementar estudos prévios, cujo foco recaía sobre a análise de estímulos linguísticos discretos (e.g. sílabas, palavras ou frases), através do estudo de potenciais evocados ou ERPs. Igualmente importante, a utilização de uma análise *lag-based*, caracterizada pela aplicação de diferentes períodos de assincronia entre a atividade cortical registada no EEG e o sinal da fala, permite identificar intervalos de tempo onde a sincronização entre os sinais é mais expressiva. Para além de estar envolvido de forma evidente no processamento contínuo da fala, ao nível da perceção e compreensão do discurso, é possível que o *Cortical Entrainment* possa estar também relacionado com outro tipo de tarefas linguísticas (e.g. produção de fala, leitura, memória ou pensamento). Este envolvimento pode ocorrer mesmo na ausência de um estímulo auditivo externo explícito. Dado que, medir *Cortical Entrainment* através do EEG é impossível durante determinadas tarefas linguísticas, devido quer à inviabilidade do sinal ou à ausência de um estímulo auditivo externo (e.g. produção de fala e leitura em voz baixa, respetivamente), é do nosso interesse estudar processos semelhantes ao *Cortical Entrainment* durante tarefas contínuas, para além da perceção de discurso. Desta forma, no presente estudo, testamos o potencial que um paradigma de "*dual-task interference*" com estímulos auditivos concorrentes (bipes) tem, de manipular e aceder indiretamente ao processamento auditivo durante tarefas linguísticas contínuas. Neste estudo de EEG, 18 participantes portugueses, com idades compreendidas entre os 19 e os 34 anos e sem quaisquer dificuldades auditivas documentadas, realizaram uma sessão experimental composta por quatro tarefas linguísticas contínuas (i.e., perceção de discurso; leitura silenciosa, leitura em voz alta e visualização de um vídeo sem som). Cada tarefa teve a duração de 5 minutos e foi realizada duas vezes, de forma aleatória, contabilizando um total de 1 hora de sessão. As tarefas de perceção de discurso e de leitura em voz alta incluíam duas versões distintas: com e sem a apresentação de estímulos auditivos concorrentes (i.e., bipes com frequências de 400 e 800 Hz e duração de 50 ms). A apresentação simultânea e aleatória dos bipes permitiu implementar o conceito de *dual-task interference*, de modo a inferir a participação do córtex auditivo noutros contextos linguísticos, mesmo na ausência de estimulação auditiva explícita (e.g. durante a leitura silenciosa). Adicionalmente, os bipes foram apresentados aos pares, separados por um intervalo de 1s, com o intuito de, não só aumentar a carga atencional necessária para a realização das tarefas, mas também de criar um efeito de expectativa que pudesse ser relacionado com o tipo de tarefa realizada. Métricas relativas aos ERPs em resposta aos bipes, conjuntamente com medidas de *Cortical Entrainment* foram obtidas, de forma a estudar as potenciais relações entre as condições. Como hipótese, propusemos a existência de uma relação entre os ERPs e o *Cortical Entrainment*. De acordo com a literatura acerca dos paradigmas de *dual-task interference*, seria altamente expectável que ocorresse uma redução da amplitude dos ERPs em resposta a estímulos auditivos concorrentes, dada a elevada quantidade de recursos atencionais requisitados durante tarefas de perceção da fala. Relativamente aos resultados, os estímulos auditivos concorrentes (i.e., bipes) geraram ERPs relativamente substanciais. Foi encontrada uma diferença significativa na amplitude do componente P200 (relativo ao primeiro bipe) entre as condições de perceção de discurso e de leitura em voz alta, sendo maior na segunda. Durante a condição de leitura silenciosa, ainda que os estímulos auditivos tenham gerados ERPs mensuráveis, não foram encontradas quaisquer diferenças comparativamente à tarefa de visualização de um filme. Verificou-se ainda a tendência à atenuação do processamento do segundo bipe em relação ao primeiro (nomeadamente do componente P200 relativamente ao N100), refletindo um processo característico de adaptação neuronal. Durante a condição de perceção do discurso, não se verificou adaptação para o componente N100 do primeiro bipe. Quando comparados os valores de adaptação neuronal entre condições, foi encontrada uma diferença significativa entre a condição de perceção do discurso e leitura em voz alta, relativamente ao componente P200. No que concerne aos resultados obtidos acerca do *Cortical Entrainment*, foram obtidas correlações significativas entre o envelope estimado e o envelope real da fala, para as duas condições de perceção do discurso. Ainda que não seja significativa, verificou-se uma diminuição do *Cortical Entrainment* para a condição em que são apresentados simultaneamente bipes, sugerindo um efeito negativo da estimulação auditiva concorrente. Os resultados da análise *lag-based* mostraram que os picos de sincronização entre o envelope da fala e a atividade cortical ocorrem aproximadamente aos 160 ms, para ambas as condições. A posterior análise da relação entre a amplitude dos componentes dos ERPs e o *Cortical Entrainment* revelou uma correlação negativa entre este e o componente P200 (relativo ao primeiro bipe). Com exceção do componente N100 do primeiro bipe, os restantes mostraram a mesma tendência, apesar de não significativa. O *Cortical Entrainment*, também medido durante a leitura em voz alta, apesar de apresentar um valor significativo de sincronização, não foi considerado viável devido aos artefactos contidos no sinal, gerados pelos músculos envolvidos na produção de fala. Para contornar esta limitação, o modelo de *entrainment* gerado durante a condição de perceção da fala foi generalizado para a condição de produção. Os resultados desta generalização mostraram uma capacidade preditiva significativa do modelo. Os nossos resultados foram capazes de replicar o fenómeno de *Cortical Entrainment* evidenciado na literatura durante tarefas de perceção da fala. Este aspeto tem o potencial de direcionar novos estudos no nosso laboratório acerca do processamento da linguagem durante condições contínuas e ecológicas. Relativamente à produção de fala, encontramos resultados bastante interessantes. Ao utilizar uma estratégia de generalização, conseguimos encontrar um *entrainment* significativo com o envelope da fala, contornando assim as dificuldades inerentes ao estudo da produção da linguagem, (i.e., artefactos relacionados com movimentos dos músculos articulatórios da fala). Este resultado sugere a possibilidade do envolvimento de processos cognitivos semelhantes durante as diferentes tarefas. Ainda que os estímulos auditivos concorrentes tenham sido capazes de gerar ERPs significativos, não foram observadas diferenças entre a condição de visualização de vídeo. Isto sugere que, ou o envolvimento do córtex auditivo durante a leitura silenciosa não é tão evidente quanto o esperado, ou a condição de visualização do vídeo requere níveis semelhantes de processamento auditivo. Estes resultados abrem horizontes a futuras investigações para melhor estudar e compreender o papel da rede de processamento auditivo na linguagem, bem como ao desenvolvimento de novos paradigmas de reabilitação envolvendo sistemas de neurofeedback.

Palavras-chave: Entrainment Cortical; Perceção de fala; Produção de fala; Paradigma *dual-task interference*; Potenciais evocados auditivos; Eletroencefalograma

Abstract

From development to everyday life, the auditory cortex proves to be central for language. It is suggested that during continuous speech perception, the auditory cortex tracks the natural rhythm of incoming speech, a process referred to as cortical entrainment. Cortical entrainment facilitates speech perception and comprehension, especially under ecological conditions of continuous stimuli processing. Beyond the known involvement of the auditory cortex in speech perception and comprehension, cortical entrainment may also be relevant for other linguistic tasks, including speech production, reading, memory or thinking, even in the absence of a physical external speech stimulus. Interested in measuring entertainment-like processes beyond speech perception, we tested the potential of a dual-task interference paradigm using auditory stimuli, to indirectly probe the auditory cortex during other linguistic tasks where entrainment measurement is impossible (i.e., speech production and reading). In our EEG experiment, electrical activity from eighteen normal-hearing adults was recorded during the performance of four different continuous linguistic tasks (i.e., speech perception, covert-reading, overt-reading, and watching a silent movie). Measures of cortical entrainment and ERPs to auditory probe tones were obtained to study potential relationships between conditions. Our results showed i) reduced ERPs during high auditory load tasks (i.e., speech perception and production); ii) a negative correlation between ERPs and cortical entrainment in speech perception; iii) successful generalization of a speech perception entrainment model to predict overt-speech production. Our results shed lights on further investigations for better studying and understanding the role of the auditory network for language.

Keywords: Cortical Entrainment; Speech perception; Speech production; Dual-task interference paradigm; Auditory-evoked potentials; Electroencephalogram

Acronym list

ERP: Event-Related Potentials

EEG: Electroencephalogram

IC: Independent Components

ICA: Independent Component Analyses

N100: Negative deflection in EEG signal occurring approximately 100 ms after stimulus presentation.

P200: Positive deflection in EEG signal occurring approximately 200 ms after stimulus presentation.

SD: Standard-Deviation

SNR: Signal to Noise Ratio

Index

Figures Index

1. Introduction

Language acquisition is an innate and universal phenomenon which constitutes a major developmental task during infancy, despite the many cross-cultural differences (Benasich & Tallal, 2002). For most people, language acquisition is driven by acoustically transmitted signals, for which the auditory cortex (and the superior temporal lobe) plays a crucial role in decoding linguistic content (Friederici et al., 2011). Sound signals need to be segmented into low-level units of signal information (i.e., pitch and frequency), and then synthesized into high order information, such as phonemes, words, and sentences, or prosody, carrying emotional intentions (Giraud & Poeppel, 2012). Beyond speech perception, acoustic processing capabilities, together with the underlying brain substrates involved, are central for speech production (Benasich & Tallal, 2002; Miller, 2011) and reading (Molinaro et al., 2016; Žarić et al., 2014).

Even newborns seem to be sensitive to acoustic-phonological regularities in the auditory and speech input, allowing the perception, processing, and spontaneous learning of language (Friederici et al., 2011). In fact, infants preferentially attend to the sounds of language, recognize, and discriminate specific patterns from the speech input (Benasich & Tallal, 2002; DeCasper & Spence, 1986; Eimas & Miller, 1992; Plunkett, 1997). During the first year of life and before they master phonological processing abilities, children are capable of categorizing speech sounds independently of their parent's native language, capturing relations between adjacent syllables, detecting syllable repetitions, taking advantage of prosodic cues for detecting possible words in the continuous auditory input, as well as processing and categorizing auditory stimuli occurring within tens of milliseconds (Benasich & Tallal, 2002; Dehaene-Lambertz & Dehaene, 1994; Gervain et al., 2008; Plunkett, 1997; Teinonen et al., 2009; Thiessen et al., 2005; Tsao et al., 2004). Importantly, during speech comprehension, listeners need to segment a continuous stream of acoustic input into smaller units, analyze the meaningful sounds within it, and ultimately comprehend its intended message. For this purpose, each "acoustic signal must be simultaneously parsed and integrated into meaningful linguistic units" (Palana et al., 2022). In the production domain, the auditory cortex is involved in monitoring the sounds we produce, by comparing the expected sounds (based on phonological representations) with the actual output (Hickok et al., 2011). Monitoring self-produced speech helps us adjust and correct speech output to ensure accuracy, communication coherence (Hickok et al., 2011) and regulate loudness (i.e., Lombard effect) (Brumm & Zollinger, 2011). This capability is believed to require a sensorimotor system, which depends on the functional connectivity between auditory and (pre)-motor cortices and may involve neural circuits capable to translate motor programs to possible auditory consequences (Hickok et al., 2011), as well as, from speech sounds to motor programs (Correia et al., 2015). The latter may provide a neural framework for learning to speak by imitation (Philippsen et al., 2014).

Learning to read and write is also suggested to involve the auditory cortical regions (superior temporal lobe). A prominent theory suggests that, during this process, humans need to establish an association between phonemes (i.e., speech sounds) and graphemes (i.e., letters) (Amitay et al., 2002; Blomert, 2011; Bradley & Bryant, 1978). Consequently, phonological awareness, phonological recoding and recalling (in working memory and from long-term memories, respectively) represent fundamental basic linguistic skills in this context (Ligges & Lehmann, 2022). Evidence from previous research shows a link between poor auditory processing/phonological awareness and reading difficulties in children and adults (Ahissar et al., 2000; Amitay et al., 2002; Sharma et al., 2006). In turn, some authors postulate that alterations in the auditory processing and acoustic perception are one of the causes of dyslexia. (Elshafaei et al., 2022; Ünalgan, 2022; Žarić et al., 2014)

Importantly, under natural conditions, speech processing, speech production and reading are continuous tasks. However, more often than not, brain imaging studies including EEG and MEG, focused at researching the neural mechanisms of these tasks, employ experimental paradigms based on discrete stimuli, such as isolated phonemes, syllables, words, or sentences (Osterhout et al., 1997; Salmelin, 2007). In order to study these processes in real time, EEG researchers have measured event-based neural responses (i.e., Event-related potentials, ERP´s) elicited during language processing (Palana et al., 2022; Zhou et al., 2016). Analyzing and interpreting ERP components in the context of speech perception has provided valuable insights about the timing and strength of neural responses to discrete speech sounds (Palana et al., 2022). For example, the N100 response has been associated with the primary processing of the acoustic-phonetic features of the auditory stimuli and is considered an automatic brain response;

the P200 and P300 are related to a language-specific phonetic–phonological analysis; and the N400 and P600 are involved in the lexical-semantic and syntactic processing and integration (Hernández et al., 2022; Salmelin, 2007).

Research focusing on the involvement of the auditory cortex during language-related tasks (i.e., speech production or reading) has also benefited from evoked responses recorded from concurrent auditory stimulation (Kok, 1997). More specifically, auditory tones/beeps have been used to indirectly modulate the auditory processing system during the performance of such tasks, based on the idea of a dual-task interference (Kober & Neuper, 2012). Daliri & Max (2015) investigated pre-speech sensory modulation during speech production in stuttering and non-stuttering individuals by recording auditory evoked potentials in response to probe tones presented immediately prior to speech onset, in a delayed-response speaking condition. The results showed a statistically significant modulation of auditory processing (reduced N1 amplitude) for the non-stuttering group that was not observed in the stuttering group, which was argued to reflect differences in auditory cortex availability during speech preparation. The notion of dual-task interference (Kok, 1997), using auditory probes, offers the possibility to investigate the participation of the auditory cortex in linguistic contexts, even in the absence of explicit auditory stimuli. If it is possible to modulate auditory processing during speech production planning using concurrent external auditory stimuli, it may also be possible to probe neural responses to auditory cues (i.e., tones) competing for the same brain resources during other linguistic tasks, where the auditory cortex plays a known role (i.e., speech comprehension, reading, memory encoding, memory retrieval or reasoning).

Despite the considerable amount of scientific knowledge produced by studies comprising ERPs, these often fail to incorporate the continuous nature of language. Most linguistic tasks we perform in our everyday life are continuous. Language processing (i.e., perception and production), or reading, in natural (i.e., ecological) conditions, require a system capable of coping with their continuous nature. We rarely hear or produce isolated syllables, words or sentences. Instead, these linguistic units are part of a bigger discourse. Because speech consists of a continuous stream of sounds, composed by quasi-rhythmic phase amplitude waves (Alexandrou et al., 2020), when analyzing isolated speech sounds, there is a loss of information relative to the neural activity necessary for integrating continuous acoustic inputs into meaningful speech (Peelle & Davis, 2012). In this way, measures of cortical entrainment have drawn researchers' attention, to examine how the brain follows segments of continuous speech as they are processed in real-time.

Our brain is daily challenged with an enormous amount of sensory information. Given that only a small part of this information is relevant to interpret and act in a specific context, there is a need to expand or suppress receptivity to sensory input, depending on its relevance (Zoefel & VanRullen, 2015). Neuronal oscillations, which correspond to cyclic variations of neuronal excitability, constitute a fundamental tool for the aforementioned purpose, as they play a preponderant role in optimizing the selection of sensory input (Giraud & Poeppel, 2012; Schroeder & Lakatos, 2009). Although neuronal oscillations occur passively, they can be manipulated by the brain (Zoefel & VanRullen, 2015). Furthermore, neuronal oscillations in the auditory cortex are able to entrain (i.e., phase-lock) with external rhythmic stimuli (Lakatos et al., 2005) and this phase-locked responses are enhanced when congruent information is present (Lakatos et al., 2007). The alignment of oscillatory activity with external cues is related to increased sensitivity for relevant information (i.e., timing, location, and stimulus features), thereby contributing to the optimization of sensory processing (Calderone et al., 2014; Peelle et al., 2013). This process of synchronization is called Cortical Entrainment, and, for it to take place, the external stimulus needs to have certain characteristics, such as rhythmic properties, stable frequency and to alternate between periods of high and low content or relevance. Interestingly, the sound of human speech fulfills these criteria (Larsson, 2017).

Cortical entrainment has been increasingly studied during recent years (Alexandrou et al., 2020). In their review, Calderone et al. (2014) highlighted several studies focusing on better understanding the relationship between entrainment for specific oscillation types and behavioral performance across different sensory modalities. The results suggested that dysfunctional entrainment may underlie a wide range of processing difficulties across psychiatric and language-related conditions. Also, decreased access to temporal predictability, language development difficulties, and selective-attention deficits, are pointed out as possible consequences of impaired entrainment (Calderone et al., 2014). Additionally, Molinaro et al. (2016), found reduced speech-brain synchrony in dyslexic readers versus normal readers while listening to spoken sentences, highlighting the strong potential of impaired entrainment to cause severe consequences in both phonological and reading skills. They also found that low levels of synchrony remain preserved through the development from childhood to adulthood, suggesting that the evaluation of cortical entrainment could be used as a diagnostic tool for early detection of children at risk of dyslexia (Molinaro et al., 2016). Additionally, evidence that speech rhythm distortions impair speech comprehension (Ghitza & Greenberg, 2009) and that the degree of neuro-acoustic entrainment modulates its intelligibility (E. Ahissar et al., 2001; Peelle et al., 2013; Vanthornhout et al., 2018), has highlighted the role of cortical entrainment for speech perception and comprehension. Furthermore, it was found a significant effect of aging in the linguistic processing of speech, as reduced cortical entrainment and increased neural response latency of speech processing were found in older subjects (Gillis et al., 2023). These structural/functional alterations across the adult lifespan occur simultaneously with age-related decline in cognitive functioning.

In summary, the role of the auditory cortex in language perception and production is undoubtable. Studies based on cortical entrainment have been complementing those based on independent, single, and short speech events, such as isolated syllables, words, and sentences, and are able to provide the means to better study the continuous nature of language-related cortical processes. Because the brain networks required for speech processing are central for language overall, it is possible that training our brain’s ability to activate the phonemes and words in continuous speech, transfers to our ability to read better and speak better. Therefore, the mentioned EEG metrics could be used to guide subjects on their brain’s auditory processing performance using neurofeedback.

Critically, cortical entrainment depends on external stimuli (e.g., speech signals) since it analyzes the level of synchrony between two signals: the external signal and the cortical neuroimaging signal. However, in several important linguistic tasks, such as speech production or reading, there is no external signal to enable a classical cortical entrainment analysis. It is reasonable to presume that these linguistic tasks convey a similar brain process to cortical entrainment, even when no physical external stimulus is involved. In such cases, an imagined or endogenous speech signal is central to the linguistic task. For example, speaking involves auditory planning and prediction of its auditory consequences, as well as, monitoring of speech output in the case of overt speech; covert reading requires the activation of imagined speech sounds. These endogenous speech signals are created by the subject in absence of an external auditory signal and are considered of relevance to the task. Other tasks that involve endogenous speech signals may include memory retrieval, thinking and reasoning. But how can cortical entrainment be computed in these tasks, in absence of an external speech signal? Is it possible to assess the involvement of the auditory cortex during these tasks, where classical brain-speech entrainment seems impossible to measure?

In the current EEG study, we investigate the brain's speech processing system in healthy individuals, by assessing the activity of the auditory cortex during continuous linguistic tasks. More specifically, we were interested in assessing auditory cortex activity during tasks where classical entrainment seems impossible to measure, due to the absence of an external auditory signal. To this end, we implemented an experimental session consisting of four continuous conditions: a) speech perception, b) Overt-reading, c) Covert-reading and d) watching a silent movie. In order to study the potential relationship between these conditions, we used two different EEG analysis approaches: i) cortical entrainment and ii) event-related potentials to auditory probes (i.e., tones) in a dual-task setting. Apart from the non-linguistic condition (i.e., silent movie) and the covert-reading task, the remaining tasks had two versions: with and without the presentation of auditory tones (400 and 800 Hz tones of 50 ms duration), making a total of six different conditions. Each condition had a five-minute duration and was presented twice in a random order. Importantly, tones were presented in pairs, separated by 1 second, in order to stimulate more attentional resources and allowing to distinguish between the passive perception of the first tone and the active prediction of the second tone. This coupling is thought to generate a reduction in ERPs amplitude, a phenomenon referred to as Neural Adaptation (Kudela et al., 2018).

Cortical entrainment was measured during speech perception and Overt-reading conditions. ERPs were computed in the conditions for which tones were presented. We expected to find a relationship between the ERP analysis and the cortical entrainment analysis during the speech perception conditions (with and without tones). According to the dual-task interference theory (Kok, 1997), we strongly expect that more auditory resources employed in listening the audiobook (i.e., more cortical entrainment) would reduce the ERPs response to the concurrent auditory probes (i.e., tones). If this relation is indeed confirmed, the next aim of our analysis is to investigate the suitability of auditory probes in dual-tasks comprising covert and overt reading.

2. Methodology

2.1. Participants

The sample consisted of 18 Portuguese-speaking adults (11 females, 55,55%) aged between 19 and 34 years old (Mean ± Standard Deviation (SD) = 22,39 ± 3,45) and 12 to 17 years of school education (Mean ± SD = 13,78 ± 1,70). All participants reported normal hearing abilities and no history of psychiatric, neurological or language-related disorders. The experimental procedures were performed in accordance with the approved guidelines and the Declaration of Helsinki and Oviedo Convention. Informed consent was obtained from each participant before conducting the experiment.

2.2. Stimuli

In order to guarantee that participants were focusing their attention on the task, a general comprehension assessment composed of three questions was presented immediately after each task. All participants were considered committed to the task, since their comprehension performance was flawless.

During the listening conditions, the spoken stimuli (i.e., audiobooks) consisted of ten five-minute audio segments taken from different episodes of a podcast in the native language of the participants (i.e., Portuguese), called – "*Portugueses no Mundo*". These comprised excerpts of interviews with Portuguese people living around the world, stereo recorded with a sampling rate of 48 KHz. The relatively long audiobooks were selected to simulate continuous speech processing and comprehension, as it happens in ecological environments. Throughout the reading conditions, participants were presented with the written version of the previously mentioned audio segments. Thus, written stimuli consisted of ten texts (i.e., dialogues), with an average reading time of approximately five minutes.

The listening and reading tasks included two versions: with and without auditory probes (i.e., tones with 400 and 800 Hz, 50 ms duration and 10 ms ramping onset and offset). Tone presentation was used to implement a dual-task interference setting, allowing us to infer the participation of the auditory cortex in linguistic contexts, even in absence of explicit auditory stimuli (e.g., silent reading). Importantly, tones were presented in pairs, separated by 1 second interval. The use of tone pairs intended to stimulate more attentional resources, as well as allowing to distinguish between the passive perception of the first tone and the active prediction of the second tone, thus increasing the neuronal resource competition imposed by the dual-task interference. Coupling tones not only enables more attentional processing, but also creates an effect of expectation. Participants were informed about the possibility of hearing "beeps" (i.e., tones) that were non-relevant for the task, and therefore, should be ignored. In order to have a control sample for auditory ERP´s, tones were also presented during a five-minute silent movie, whose content implies no sound.

2.3. Experimental Procedure

The experiment was conducted in an electrically-shielded, sound-attenuated room. The EEG recording preparation started by placing a 10/20 system cap (centered through the midpoint between the inion and nasion bones in the sagittal plane, and the midpoint between the ears in the coronal plane). Conductive gel was applied using a plastic syringe. The 32 electrodes were then placed in the predefined areas of the right and left hemisphere, in a Biosemi EEG system (*BioSemi Amsterdam, Netherlands*). This included: Fp1/Fp2, F3/F4, F7/F8, FT9/FT10, FC1/FC2, FC5/FC6, C3/C4, T7/T8, CP1/CP2, CP5/ CP6, TP9/TP10, P3/P4, P7/P8, O1/O2, Fz, Cz, Pz and Oz. Participants were comfortably positioned in the chair and the quality of electrode connectivity was verified (i.e., impedance kept below 20 kΩ).

The experiment session consisted of a set of four continuous conditions (*Silent-movie condition, Speech Perception condition, Overt-Reading condition,* and *Covert-reading condition*). Each condition had a five-minute duration and was performed twice, in a random order. After the performance of each condition, we asked participants three content questions to guarantee they paid attention to the task stimuli. It is important to note that, for each participant, the audiobooks and texts were never repeated, and that the written version of the audiobook was never presented if the participant had already listened to the corresponding auditory version.

In the *Silent-movie condition*, participants had to watch a silent movie while being randomly stimulated with coupled tones. After pressing the spacebar button, a fixation cross was displayed for approximately 1s before the video started. This task was designed to provide a control condition for the auditory ERP data, as the tones were the only auditory stimuli concurring for the auditory cortex resources.

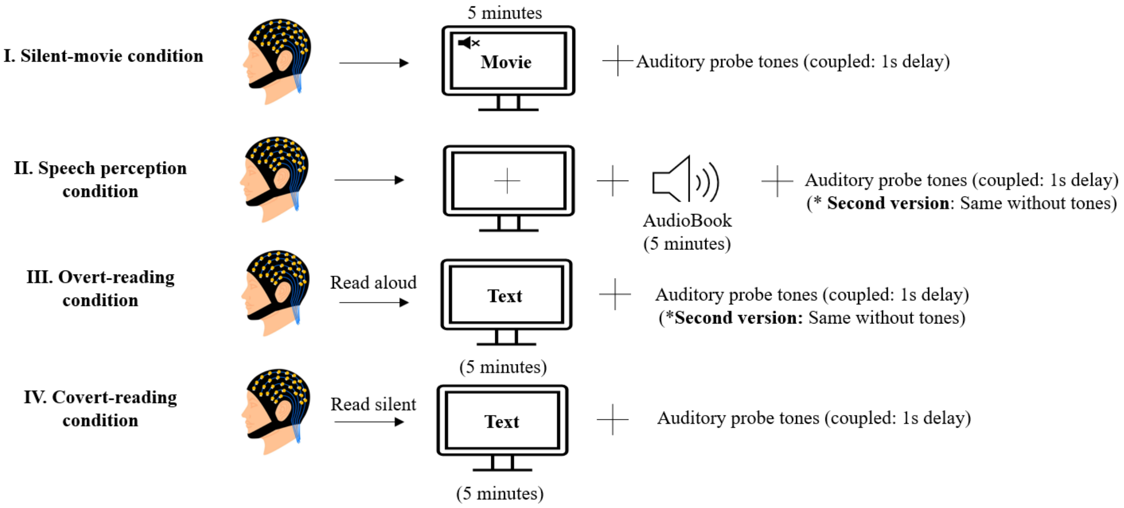
The *Speech Perception condition* comprised two tasks. In the first one, participants listened to a 5-minute audiobook and, in the second one, the presentation of a different audiobook was combined simultaneously with random coupled tones. In both conditions, written instructions were presented on the screen, and participants were asked to press the spacebar button when ready to start the task, time at which the audio started and the task ran without interruptions for 5 minutes. This approach was designed to study speech perception and comprehension in a continuous manner. The duration of the stimulus allowed the measurement of cortical entrainment, a measure of neural tracking between cortical activity and the external auditory signal. On the other hand, the second task is compelling because tones will be concurring for the same auditory resources as the linguistic stimuli. This enables studying potential changes of entrainment or in ERP responses, hence establishing a relationship between them (i.e., between cortical entrainment and concurrent tone processing).

The *Overt-Reading condition* also had two different tasks. In the first one, participants had to read aloud a written version of the audiobooks and, in the second one, the same happened combined with the random presentation of coupled tones. Whenever a participant finished reading a page, the next one appeared after pressing the right-arrow button. The participant's voice was recorded by a Shure SM58 microphone (*Shure inc., Niles, Illinois, USA)* placed on the table. This reading condition allowed studying the participation of the auditory cortex in speech production, which is suggested to require a sensorimotor network linking auditory and pre-motor brain regions, both for planning speech and monitoring ongoing auditory consequences of speaking.

Finally, in the *Covert-Reading condition*, participants silently read the written version of the audiobooks, as in the overt-reading condition, while being simultaneously stimulated with randomly presented coupled tones. In this condition, we did not implement a task without tone presentation, because such would not have any auditory stimuli to assess the participation of the auditory cortex.

Figure 1

Experimental session design



Note: Summary of the 1-hour experimental session design with all conditions presented to participants (N = 18). In silent-movie condition participants watch a silent movie while being randomly presented with coupled tones. In Speech perception condition, participants listened to a 5-minute audiobook (with and without tones). In Overt-reading condition, participants read aloud a written version of the audiobooks (with and without tones). Finally, in Covert-reading condition, participants silently read a written version of the audiobooks while simultaneously stimulated with coupled tones.

2.4. Data recording and pre-processing

Continuous EEG recordings were performed using the BioSemi ActiveTwo 32-channel Ag/AgCl electrode system. The signal amplification was performed by the BioSemi ActiveTwo amplifier, with a sampling rate of 264 Hz during the entire period of the experiment and recorded by the ActiveView600 – Hires software (*BioSemi Amsterdam, Netherlands*). For all pre-processing stages and statistical analyses of the epoched EEG data, EEGLab (Delorme & Makeig, 2004) and MathLab (version R2019b) custom-made scripts were used. EEG data processing started with re-referencing the EEG channels to the Fz electrode, followed by band-pass filtering (0,5-40 Hz). Removal of signal artifacts was performed by correcting stereotypical artifacts related to eye movements, eye-blinks, muscle contractions and/or noisy electrodes, using independent component analysis (ICA, INFOMAX-runica algorithm as implemented in EEG Lab

(Delorme & Makeig, 2004). Participant-specific independent components (ICs) were categorized as neural activity or non-neural artifacts by visual inspection of their scalp topography, spectral peak(s) at typical EEG frequencies, range of voltage, and consistency of their response time-course using the continuous 5-minute blocks of acquired data. ICs representing non-neural artifacts were removed. Fourier Transform was used to decompose the signals from the time domain to the frequency domain allowing to compute a spectrogram, which served to assess data quality.

2.5. Data analysis and Statistics

2.5.1. ERP analysis

For each trial, cleaned epochs of 2000 ms duration including a 500 ms pre-stimulus baseline, were referenced to baseline. Individual-subject ERPs obtained from the average of all trials were used further to calculate grand-average ERPs and group-statistics. Furthermore, time-windows of interest were used to assess mean ERP responses per condition (i.e., 4 windows-of-interest: first N100 = [120-140 ms]; first P200 = [210-230 ms]). For second N100 and P200 components, same time-windows were utilized. These periods of interest were chosen according to existing literature, namely evidence from N100 and P200 peaks following stimulus presentation. Parametric paired and independent sample t-tests were used to obtain values for statistical deviation from baseline (i.e., zero) and statistical comparisons between conditions, respectively (‘*’ and gray lines used to depict p-values below 0.05) (as depicted in figure 2).

Neural adaptation, which reflects a decrease over time in sensory neuron sensitivity to process constant or repetitive stimuli, was assessed by determining differences in amplitude between the ERP components (N100 and P200) of the two tones. This procedure involved subtracting the mean N100 and P200 amplitudes of the second tone from the respective components of the first tone, as shown in figure 3. Positive values represent a reduction in amplitude (i.e., high neural adaptation) between tones, whereas negative values represent the opposite. Similarly, paired sample t-tests were conducted to investigate whether values of neural adaptation differ from baseline (i.e., zero meaning no adaptation) and independent sample t-tests for comparisons between conditions (‘*’ and gray lines used to depict p-values below 0.05).

2.5.2. Entrainment analysis

To obtain a general cortical entrainment metric across conditions, learning a multivariate decoding model was required. Multiple EEG channels (i.e., 32 channels) and multiple time-asynchrony versions of the EEG signal were used to find a link to the speech envelope signals. First, the relation between the multi-channel EEG signal and the envelope of the speech signal was modeled and cross-validated. This process encompassed two phases: i) a training phase and ii) a testing phase. Given that each condition had two trials, data was splitted in half (i.e., one trial used for training and one trial used for testing). In the training phase, an algorithm/model was applied to learn and recognize patterns/associations between EEG signals and speech envelopes. The model's internal parameters were adjusted iteratively to minimize prediction errors and improve its ability to make accurate predictions. Next, during the testing phase, aiming for cross-validating the created entrainment model, we assessed its ability to generalize and predict previously unseen data (i.e., second trial dataset), generating a predicted speech envelope. To assess model performance, correlation between predicted and real speech envelopes was calculated using Pearson correlation coefficient (r). Parametric paired sample t-test were used to calculate individual-subject correlations statistical deviation from baseline (i.e., zero) (see figures 4A, 6A and 7A)

Also, it is possible to determine time-asynchronies where entrainment reaches maximum peaks by repeating the same procedure using individual lags rather than the total number of employed lags ("lag" refers to each time-asynchronous variant of the EEG signal in respect to the speech signal). This procedure allows to better understand the underlying neural mechanisms involved in continuous speech processing (i.e., speech perception and prediction). A lag-based analysis was conducted over multiple asynchrony lags (from -1000 to +1000 ms) of the speech envelope in respect to the EEG signal. For negative lags, the speech envelopes were shifted backwards in respect to the EEG signals, reflecting the EEG capacity to predict speech signals, while positive lags indicate that the speech envelopes were shifted forward in respect to the EEG signals, reflecting the EEG capacity to predict past speech signals. At neutral lags (lag = 0) no asynchrony was imposed between both signals, reflecting the prediction of instantaneous speech envelopes. Parametric paired sample t-tests were used to assess which lags evidenced statistical deviation from baseline (i.e., zero) (see figures 4B, 6B and 7B).

2.5.3. Probe Tones ERPs influence on Entrainment

Statistical analysis aiming to explore the possible relation between auditory probe tones and entrainment was conducted by calculating Pearson correlation coefficient (r) to assess the strength and direction of the linear association between these two variables (‘*’ used to depict p-values below 0.05). Additionally, a linear regression analysis was performed allowing us to construct a predictive model for better understanding how changes in ERPs might be associated with variations in entrainment or vice-versa. (see figure 5)

3. Results

3.1. Auditory ERPs

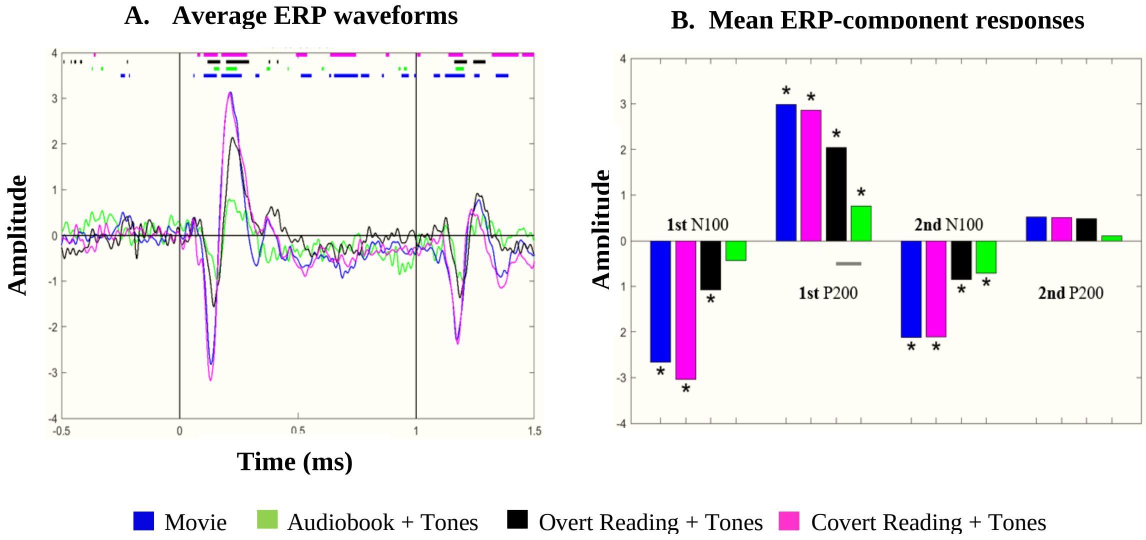
The average auditory ERP waveforms (for the central Cz-channel) elicited by the auditory probes across conditions (i.e., Movie, Audiobook, Overt Reading, Covert Reading) are shown in figure 2A. For all conditions, both tones produced a prominent negative peak around 100 ms after stimulus onset (N100 ERP-component) followed by a positive peak around 200 ms (P200 ERP-component) after stimulus onset. A mean ERP response analysis over time-windows of interest, centered on each peak (i.e., first-N100, first-P200, second-N100 and second-P200), was conducted. For the first tone, statistically significant N100 (Movie: X = -2.66 µV; df = 17; p = 1.70e^{-6}; Overt-reading: X = -1.07 µV; df = 17; p = 2.64e^{-5}; Covert-reading: X = -3.03 µV; df = 17; p = 8.06e^{-7}) and P200 (Movie: X = 2.99 µV; df = 17; p = 1.72e^{-4}; Audiobook: X = 0.77 µV; df =17; p = 1.90e^{-3}; Overt-reading: X = 2.05 µV; df = 17; p = 2.18e^{-5}; Covert-reading: X = 2.86 µV; df = 17; p = 9.50e^{-5}) responses were obtained across all conditions, except for the N100 component, produced while listening to an audiobook (X = -0.44 µV; df = 17; p = 0.0832). In respect to the second tone, a significant N100 response was obtained for all conditions (Movie: X = -2.12 µV; df = 17; p = 2.08e^{-7}; Audiobook: X = -0.71 µV; df = 17; p = 0.0123; Overt-reading: X = -0.85 µV; df = 17; p = 0.00360 ; Covert-reading: X = -2.11

μV; df = 17; p = $3.72e^{-7}$) but not for the P200 response (Movie: X = 0.52 μV; df = 17; p = 0.0517; Audiobook: X= 0.11 μV; df = 17; p = 0.643; Overt-reading: X = 0.48 μV; df= 17; p= 0.0587; Covert-reading: X= 0.51 μV; df = 17; p = 0.139)(see figure 2B).

In order to identify significant differences in ERP responses between conditions, independent sample t-tests were computed. Comparisons followed a specific criterion, as we matched values obtained during conditions where an external sound signal was absent (i.e., Movie vs. Covert-reading) and conditions where a physical external speech stimulus was present (i.e., Audiobook vs. Overt-reading). No significant differences were found between Movie and Covert-reading conditions (1st and 2nd N100: p = 0.219 and 0.973, respectively; 1st and 2nd P200: p = 0.681 and 0.972, respectively). Audiobook and Overt-reading conditions, when compared, revealed a significant difference relative to the P200 ERP-component in the first tone (p = $8.31e^{-4}$). Despite this, the remaining components did not differ statistically (1st and 2nd N100: p= 0.075 and 0.723, respectively; 2nd P200: p=0.292) (see Figure 2B).

Figure 2

Event-Related Potentials from auditory probe tones



A. Average ERP waveforms

B. Mean ERP-component responses

Note: A) Average auditory ERP waveforms of random coupled auditory probes presented during task performance across conditions. The horizontal-colored bars represent time-points where statistically significant results were found ($p < 0.05$, uncorrected); B) Summary of mean ERP-component responses elicited by random coupled auditory probes. The chart bars were obtained after calculating the mean amplitude of the signal in each condition during a specific time interval. All responses marked with a * differed significantly from zero ($p < 0.05$); Grey horizontal lines represent statistically significant differences found between conditions relatively to specific ERP components.

3.2. Analysis of neural adaptation

Differences in amplitude (i.e., neural adaptation[1]) between ERP components relative to the two sequential tones were obtained by subtracting the mean N100 and P200 amplitudes of the second tone from the corresponding values of the first tone. The results shown in figure 3 demonstrate a general tendency for significant neural adaptation for both N100 (Movie: X = 0.54 μV; df = 17; p = 0.0128; Overt-reading: X = 0.22 μV; df = 17; p = 0.444; Covert-reading: X = 0.92 μV; df = 17; p = 0.0165) and P200 (Movie: X = 2.48 μV; df = 17; p = 4.90e^{-4}; Audiobook: X= 0.66 μV; df = 17; p = 0.0160; Overt-reading: X = 1.56 μV; df = 17; p = 3.08e^{-4}; Covert-reading: X= 2.35 μV; df =17; p = 1.42e^{-5}) ERP-components across conditions. A deviant result was obtained for the N00 ERP-component in the Audiobook condition, where adaptation between tones was not significant (X= -0.28 μV; df = 17; p = 0.401). In fact, an increase in N100 response between tones was found, although not statistically significant. Moreover, in general, we observed larger neural adaptation for the P200 ERP-component comparatively to the N100 component.

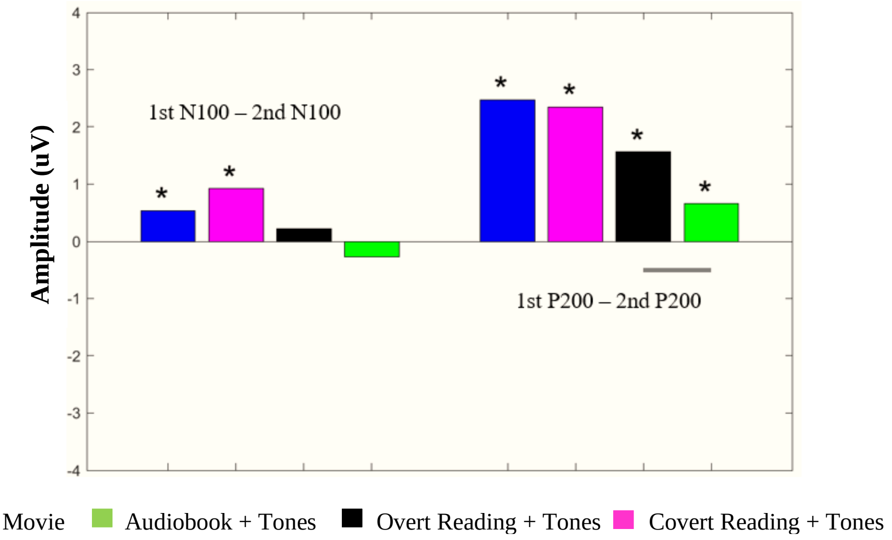
Comparisons between neural adaptation values across conditions were also performed using independent sample t-tests. No significant differences were found in N100 (df = 17; p = 0.334) and P200 (df = 17; p = 0.782) adaptation values when comparing movie and covert-reading conditions. Additionally, in the audiobook versus overt-reading comparison, a marginal significant difference was found in respect to the P200 ERP-component response (df = 17; p = 0.0429 but not for N100 (df = 17; p = 0.349).

---

[1] Throughout this thesis, we use the term ‘Neural Adaptation’ to refer to the reduction in amplitude of ERPs in response to the presentation sequential auditory tones (Kudela et al., 2018). However, we do recognize that, due to the large variety of cognitive processes involved in this phenomenon (eg. auditory attention and expectancy), this use can be controversial.

Figure 3

Neural Adaptation (second – first tone ERP components)



Note: Differences in amplitude (i.e., neural adaptation) between the ERP components (N100 and P200) of the two tones, across conditions. Positive values represent a reduction in amplitude (i.e., neural adaptation) between tones, whereas negative values represent an increase in amplitude (opposite of neural adaptation). (*) Represent significant values of adaptation ($p < 0.05$) and gray horizontal lines represent significant differences between conditions ($p < 0.05$).

3.3. Cortical Entrainment Analysis

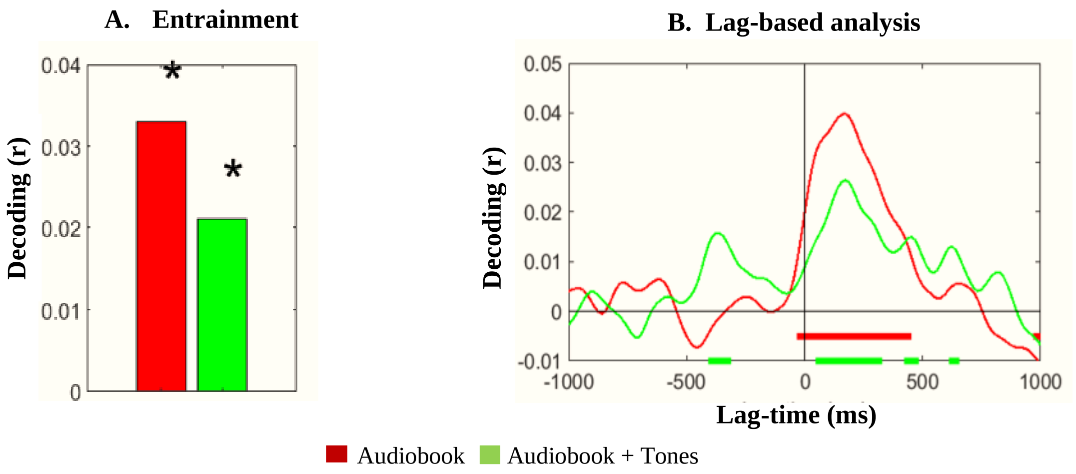
The main analysis of this study focused on verifying the relationship between classical cortical entrainment and an alternative method based on dual-task interference (using auditory probing tones). Cortical entrainment was measured in both speech perception conditions (i.e., audiobooks with and without tones) using a “training-testing” cross-validation approach. This crucial methodological approach was used to assess the robustness and generalizability of the multivariate model, particularly when investigating neural entrainment phenomena. It involved dividing the dataset into multiple subsets (i.e., a training set and a testing set) to validate the model's performance. During the training phase, a learning algorithm/model was used in order to recognize EEG patterns in the data that help to predict the speech stimuli (in this case, to decode the envelope of the speech signal). The model (Ridge Regression) was adjusted iteratively to minimize prediction errors and improve its ability to make accurate predictions. Next, during the testing phase, the purpose was to assess the model's ability to be generalized to unseen data (i.e., how well the model can be generalized to make predictions on new datasets).

Since each condition encompassed two presentation blocks, we were able to divide the recorded data in two parts, serving as training and testing of one another. Results show significant cortical entrainment for both listening conditions (Audiobook: $r = 0.033$; p = 2.0155e^{-5} and Audiobook + tones: $r = 0.021$; p = 0.0023) (see figure 4A). Additionally, although cortical entrainment values differ between conditions (i.e., lower value when tones are presented simultaneously), it is not statistically significant (p=0.174). Nevertheless, the use of probing tones was found to be related to a decrease in cortical entrainment, meaning that the model's ability to predict cortical activity was reduced during this condition.

To help understanding the neural process involved in entrainment, a lag-based analysis was also conducted by isolating asynchrony delays (i.e., lags) between the speech envelope and the EEG signal (from -1000 msec to 1000 msec, spaced by 16 ms units). Figure 4B shows the group average decoding performance across lags. Results show significant higher levels of entrainment around a 160 ms lag, for both conditions ($p < 0.05$, uncorrected) (i.e., when the speech signal precede the EEG signal by 160 ms).

Figure 4

Cortical Entrainment during Audiobook conditions



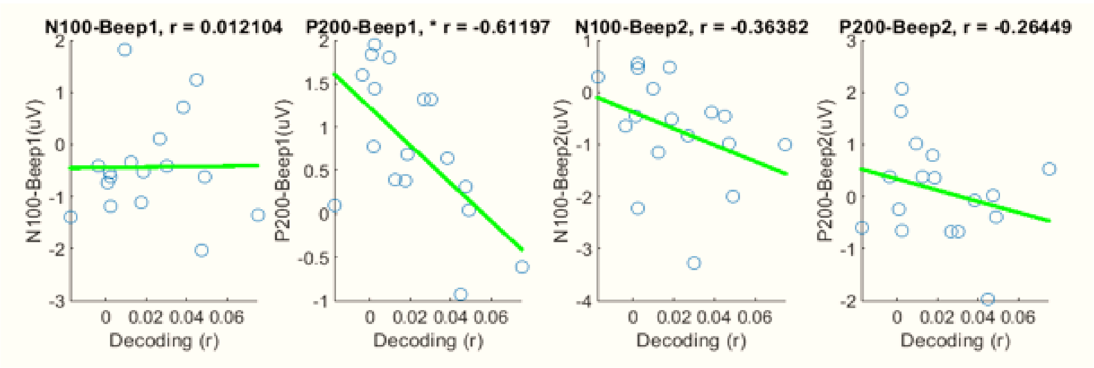
Note: A) Correlation values between the real speech envelope (audiobook) and the estimated envelope created by a multivariate computational learning model. Significant cortical entrainment values, given the cortical activity measured in EEG, represent the model´s ability to predict speech envelope. * Represents significant correlation values ($p < 0.05$). B) Results from a lag-based analysis. Asynchrony between the speech envelope signals and the EEG signal was used to estimate the time-course of envelope prediction. Negative lags indicate that the audio signal is preceding the EEG signal and positive lags indicate that the brain activity follows the auditory signal. Zero lag is when the auditory and EEG signals are synchronous. Colored horizontal bars represent delay time intervals where correlation between signals was statistically significant ($p < 0.05$)

3.4. Correlation between dual-task-interference and cortical entrainment

In order to examine a possible tone effect during entrainment, we calculated the correlation between the amplitude of the summarized ERP components and the level of cortical entrainment to the speech envelope (see figure 5). Results show a significant negative correlation between the first-P200 ERP-component and cortical entrainment (*r* = -0.611; p =0.009). Apart from the first-N100 ERP-component (*r* = 0.012; p = 0.963), the remaining component correlations, although non-significant (second-N100: *r* = -0.364; p = 0.151 and second-P200: *r* = -0.264; p =0.305) were found to follow the same negative correlation tendency.

Figure 5

Correlation between ERP components and Cortical Entrainment



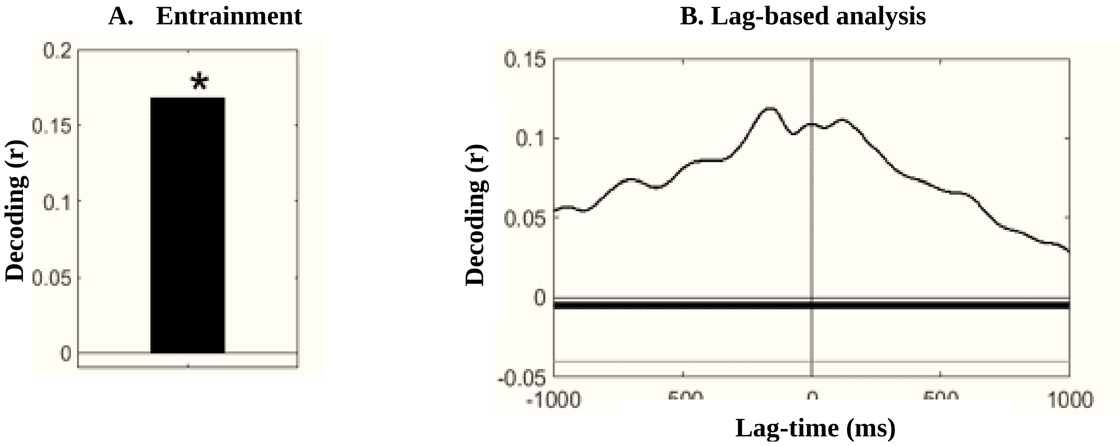
Note: Correlation values between auditory ERP components and Cortical entrainment. (*) Represents significant correlation values ($p < 0.05$). Results evidence that higher ERP-component amplitudes are associated with lower cortical entrainment.

3.5. Entrainment to self-generated speech during speech production

Cortical entrainment was also measured during the Overt-reading condition by analyzing levels of synchrony between cortical activity and participants own voice envelope (recorded through a microphone). Results showed the presence of significant cortical entrainment (*r* = 0.167; p = $1.3732e^{-5}$) (see figure 6A). A lag-based analysis was also performed, which indicated significant decoding (i.e., correlation between predicted and actual speech envelope) across all lag-time intervals ($p < 0.05$, uncorrected), with a peak centered around the 0 ms time-lag (i.e., complete synchrony between the EEG and the speech signals). Based on the conductance of speech muscle activity, and consequence contamination of the EEG signal, we cannot exclude speech muscle artifacts influence in signal integrity. In fact, based on previous results, decoding of speech muscle artifacts were highly expected, and a reason for the unsuitability of speech entrainment methods during overt-speech production. This expected limitation motivated an alternative analysis of entrainment for overt-speech production, based on the generalization of the entrainment model obtained in the speech perception condition to a speech production condition, as described below.

Figure 6

Cortical Entrainment during Overt Reading condition



Note: A) Correlation values between the real speech envelope and the estimated envelope created by a multivariate computational learning model. Significant cortical entrainment values represent the model ability, given the cortical activity measured in EEG, to predict the self-generated speech envelope. * indicates significant correlation values ($p < 0.05$, uncorrected). B) Speech envelope decoding results derived from the lag-based analysis. Black horizontal bars represent time-lags at which correlation between the real and the estimated speech envelopes was statistically significant ($p < 0.05$, uncorrected). Note that in this figure, cortical entrainment may be heavily driven by speech muscle artifacts (non-neural), since speech output and speech muscle contamination are typically correlated.

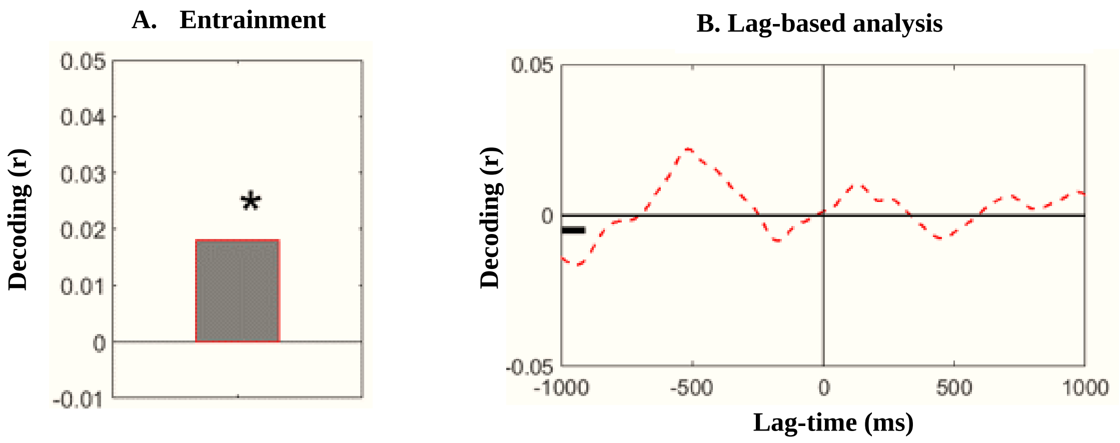
3.6. Entrainment generalization: from speech perception to speech production

To train and test a model of entrainment, usually we split data recorded from the same condition. In this case, we tested the possibility to generalize a model of entrainment for independent conditions. More specifically, the model obtained from a speech perception task (i.e., listening an audiobook) was generalized to a speech production task (i.e., overt-reading). Consequently, the model of entrainment constructed from the speech perception condition (training phase), was generalized to the overt-reading condition (testing phase). The results of the entrainment-generalization analysis are shown in figure

7. Statistical analysis revealed a significant correlation value (r = 0.018; df =16; p = 0.0442). Although marginally significant, this result proves that the neural processes involved in tracking incoming speech during the speech perception condition are similar to those used when speaking aloud. Despite significant entrainment-generalization, the lag-based analysis using the same generalization strategy yielded less clear results, and failed to identify specific lags where entrainment was maximum.

Figure 7

Cross-Correlation analysis



Note: A) Results of a cross-decoding analysis, involving the generalization of the entrainment model obtained from the speech perception condition to the overt-reading condition. * indicate significant correlation values ($p < 0.05$, uncorrected). B) Results from the lag-based analysis. Black horizontal bar represents time-lags for which the correlation between the real and the estimated speech envelopes was statistically significant ($p < 0.05$, uncorrected).

4. Discussion

In this EEG study, we investigated the brain-speech processing system by assessing cortical entrainment to ongoing speech during continuous linguistic tasks. Importantly, we addressed methodological challenges to study cortical entrainment in a number of situations where the use of speech input is unattainable (e.g., covert-reading) or possibly confounded by non-neural artifactual EEG sources (e.g., articulatory gestures in speech production). Hence, we tested an alternative cortical entrainment method, based on a dual-task interference paradigm using auditory probing tones. Cortical activity was recorded during four conditions (i.e., speech perception, overt-reading, covert-reading and watching a silent movie). We used the speech perception condition to validate that both methods evaluated (i.e., the classical cortical entrainment method and the dual-task interference method) show congruent results. Next, we compared the dual-task interference method across conditions to broaden our knowledge on the potential of this method for language research in continuous tasks, thus with additional ecological validity that is often lost in classical ERP approaches (i.e., syllable, word, or sentence processing). Finally, we investigated the possibility that an entrainment model created during a speech perception task, could be generalized for a speech production task. Our main results are i) reduced ERPs during high auditory load tasks (i.e., speech perception and production); ii) a negative correlation between ERPs and cortical entrainment during speech perception; iii) successful generalization of a speech perception entrainment model to predict overt-speech production.

The brain's sensory systems are able to track rhythmic environmental cues for optimizing their processing outcomes (Calderone et al., 2014; Peelle et al., 2013). The same principle applies in the context of speech signals because of their quasi-rhythmic properties (Alexandrou et al., 2020). Consequently, cortical entrainment to speech represents the alignment of neuronal oscillatory activity with the speech envelope (Barajas et al., 2021; Giraud & Poeppel, 2012), and it is a fundamental neural property of speech perception and comprehension (Peelle et al., 2013). In our study, we found significant cortical entrainment values for speech perception conditions, indicating that participants brain was actively engaged in listening and understanding the audiobook stories. These results are consistent with the literature, as this synchronization phenomenon has been replicated several times (E. Ahissar et al., 2001; Aiken & Picton, 2008; Howard & Poeppel, 2010; Luo & Poeppel, 2007; Molinaro et al., 2016; Peelle et al., 2013). Since it was the first time that the method of cortical entrainment was applied in our lab, this replication is important, opening new and relevant research directions to future research at the University of Algarve, under ecological conditions (i.e., continuous speech processing).

Entrainment requires learning a multivariate decoding model, where multiple EEG channels (i.e., 32 channels) and multiple time-asynchrony versions of the EEG signals, are used to find a link to the speech's envelope signal. Each time-asynchrony version of the EEG signal is referred to as a ‘lag’. Repeating the same entrainment analysis using individual lags, instead of all used lags combined, offers the possibility to identify which time-asynchronies are most relevant for entrainment. In other words, it adds temporal resolution to the entrainment analysis. Lag-based results showed that maximum cortical entrainment to auditory speech signal, during speech perception, took place, on average, about 160 ms after speech onset. This positive lag-interval indicates that the brain is processing speech sounds that were presented 160 ms before, continuously. This finding is consistent with several studies, positing that, for listeners, synchrony between brain activity and the speech envelope occurs normally at positive lags (Brodbeck et al., 2018; Broderick et al., 2019; Brumberg et al., 2016; Pérez et al., 2017).

Many studies have highlighted the role of attention in entrainment modulation (E. Ahissar et al., 2001; O’Sullivan et al., 2015). When listeners attend to a relevant speech stream in detriment of competing speech streams, auditory neurons preferentially track changes in the speech-sound envelope of the attended stream (Wikman et al., 2021). Although tones differ from continuous speech, they do require auditory attentional resources that may be necessary for entrainment. Importantly, when auditory probe tones were introduced during speech perception (i.e., dual-task interference), we found a decrease in cortical entrainment values comparatively to the same task in absence of tones. This result suggests that concurrent auditory tones may affect speech perception, as it decreases the brains’ ability to process incoming speech. This tone effect was strongly expected according to the dual-task interference theory (Kok, 1997). The fact that participants' auditory resources were being recruited while listening to the audiobook caused less responsiveness and sensibility to process irrelevant concurrent auditory stimuli (i.e., probe tones). Interestingly, we also found a significant negative correlation between the amplitude of the P200 ERP-component (first tone) and cortical entrainment (i.e., the higher P200 response, the lower cortical entrainment). Although not significant, the second tone components showed the same tendency of a negative correlation. Taken together, this suggests the presence of a relation between the speech envelope and the ERP responses. Whether entrainment is influenced by the presence of concurrent auditory stimulation, or the processing of these stimuli is influenced by entrainment is not possible to infer from our data. Based on the dual-task interference theory, we expected a bidirectional relationship between cortical entrainment and the ERP responses to the tones. However, it is possible that the tones influence entrainment, but not the opposite. This possibility would explain the fact that we did not find significant ERP differences between the silent movie and the covert-reading conditions. During covert-reading, the possible internal (i.e., endogenous) cortical entrainment necessary for reading was not significantly modulating the ERP responses to the concurrent auditory probes, comparatively to a condition for which no entrainment is expected (i.e., the control silent movie condition).

The next step in our analyses focused on comparing the dual-task interference method across conditions, to broaden our knowledge on the potential of this method for language research in continuous tasks. In order to do that, we focused on the average ERP waveforms obtained from auditory probe tones. As expected, auditory probe tones produced substantially reliable ERPs across conditions, except for the P200 ERP-component in the second tone. This might be due to poor signal quality or channel/recording issues. Also, during overt-reading and speech perception conditions, it was found a reduced N100 and P200 ERP-component amplitudes for the first tone. Throughout these conditions, the recruitment of a considerable amount of auditory attentional resources may have led to reduced sensitivity for processing task-irrelevant concurrent stimulation competing for the same auditory resources. Also, these two conditions were the only ones that involved an auditory outcome (i.e., auditory speech envelope and participants' own voice). Our results also show a significant difference in the mean amplitude of P200 ERP-component peak (first tone) between speech perception and overt-reading conditions (higher for overt-reading). This suggests that listening to an audiobook requires more attentional resources than processing our own voice sounds. This might be justified by the fact that during overt reading, speech is a self-generated sound that can be predicted by the brain, resulting in a sensory attenuation effect (Mifsud & Whitford, 2017)

Neural adaptation is a phenomenon in which the sensitivity of sensory neurons decreases over time in response to a constant or repetitive stimulus (Kudela et al., 2018). This adaptive process allows our sensory systems to filter out unchanging or irrelevant information, focusing on detecting new or important sensory inputs. We measured neural adaptation to auditory probed tones by assessing differences in ERP responses of coupled probe tones. After subtracting the mean N100 and P200 ERP-component amplitudes of the second tone from the respective components of the first tone, we found a general tendency for significant neural adaptation across conditions. This means that an effect of expectancy was produced. Also, neural adaptation for the P200 ERP-component was higher comparatively to N100. This might represent our brains' ability to optimize sensory and attentional resources for processing relevant or new, rather than irrelevant or expected, stimuli. The speech perception condition was the only where adaptation was not straightforward, as an increase in N100 ERP-component response was found, despite statistically insignificant. Interestingly, the lowest ERP responses were also found in this condition. We can argue that the reduced sensory sensibility to process concurrent auditory stimuli suggested in this condition may be related to a redirection of the attentional focus for the second tone, to ensure the presentation of the first one.

EEG studies have undoubtedly contributed to uncover and better understand the underlying auditory processing networks and neural substrates involved during speech perception and comprehension. However, the same does not apply in the context of language production (Lovenia et al., 2019). Speech production is a complex task relying on the orchestration of respiratory, phonatory and articulatory muscles (Tremblay et al., 2016). Consequently, experiments using classical cortical entrainment to study the auditory processing involved in self-generated speech (i.e., language and word production), suffer from low signal-to-noise ratios (SNR) caused by non-brain activities recorded in EEG (i.e., speech artifacts) (Wei et al., 2019). SNR quantifies the strength of desired task-related electrical brain signals (i.e., EEG signal) relative to unwanted spontaneous electrical activity and other sources of interference (i.e., noise). In speech production, electrical activity generated by muscle movements (e.g., mouth, jaw, tongue, etc.) will inevitably contaminate the neural signal. Despite the development and improvement of sophisticated methods to surpass this limitation over recent years (Lovenia et al., 2019), the presence of these artifacts overlapping recorded data during speech production, constitutes a challenge distinguishing whether components truly represent brain activity (Wei et al., 2019).

Pérez et al. (2022) found that participants' brains entrain to the self-generated speech signal at a maximum peak immediately prior to vocalization (i.e., in negative lags around -10 to -20 ms). They also posit the validity of this finding, since there is evidence that auditory modulation occurs before self-generated sounds, based on internal sensorimotor models of speech production (Hickok et al., 2011). Similarly to Pérez et al. (2022), we also found significant cortical entrainment during the overt-reading condition, suggesting that participants listened and processed the self-generated speech signal. Although this could reflect cognitive processes related to auditory planning and prediction of speech output, we assumed that entrainment measures for speech production are contaminated by speech muscle artifacts, which are difficult to disentangle. Beyond this, three additional pieces of evidence supported our decision to not consider this cortical entrainment valuable. First, cortical entrainment values found for speech production in this experiment, were three times larger than those obtained during speech perception (i.e., while listening an audiobook), which is unlikely. Second, results from a lag-based analysis showed significant entrainment across all lag-intervals, with a maximum peak around 0 ms. Finally, prediction of speech output hypothesized from sensorimotor speech planning models seems to occur in earlier negative lags, as reported from an intracranial-EEG study (Brumberg et al., 2016, see figure 8). Taken together, these results suggest that, in this study, cortical entrainment found during overt speech production reflects signal artifacts caused from muscle movements occurring simultaneously with voicing.

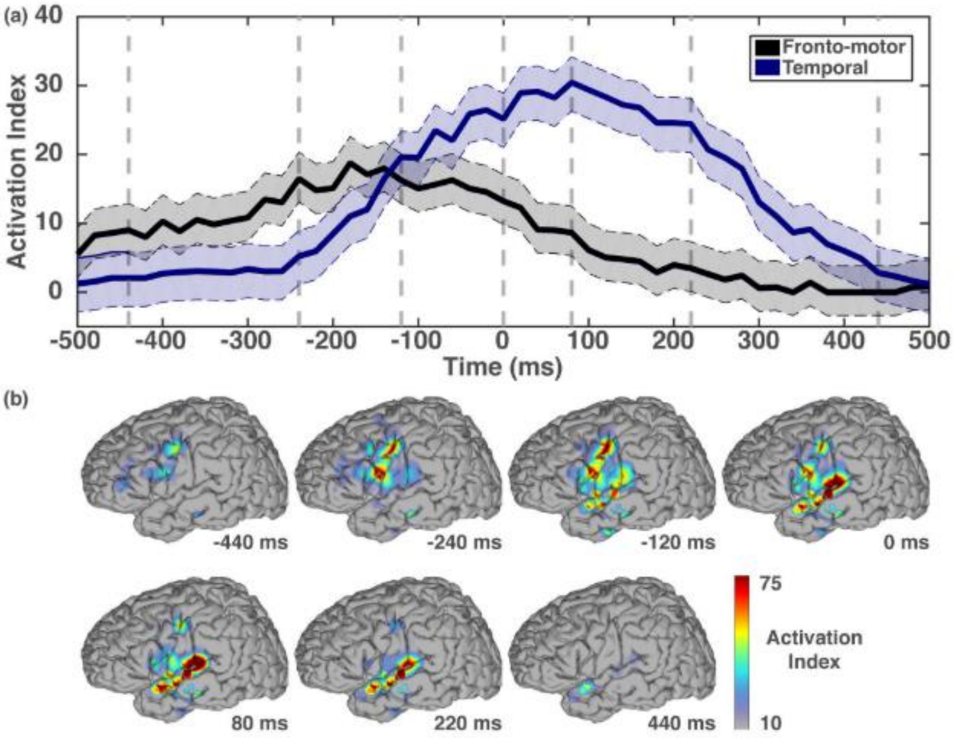
In order to avoid the evident limitations of classical cortical entrainment in assessing auditory processing during speech production, we tested the possibility to generalize an existent entrainment model, created during speech perception (i.e., listening to an audiobook), to a speech production task. For this purpose, a cross-decoding analysis was conducted, reducing possible articulatory speech artifacts in the entrainment metric. Importantly, we found that the speech perception entrainment model was able to predict the self-generated speech signal. This suggests that the neural processes involved in speech perception have similarities with those involved in planning/monitoring the auditory consequences of speech production (i.e., speech output). Furthermore, this novel result opens the possibility to use the generalization entrainment method for studying and understanding the role of the auditory network during speech production, which has insofar required intracranial EEG recordings (Brumberg et al., 2016).

Altogether, our results suggest that further research is needed in order to investigate and clarify the relationship between cortical entrainment and ERPs in response to concurrent auditory probing tones. Although we were not able to modulate the ERP responses to auditory probes during covert-reading conditions, results obtained from speech perception conditions reinforce the possibility to extend the use of a dual-task interference paradigm to study similar processes to cortical entrainment, during continuous linguistic tasks. Also, if ERPs and cortical entrainment are in fact related, we could use auditory probes to study cortical entrainment.

Neurofeedback, also known as EEG biofeedback, relies on the possibility to recondition and retrain brainwave patterns (Hammond, 2011). This technique allows individuals to increase awareness of their neuronal activity, providing them with the means to influence, and gradually change it (through real-time feedback). Promising results on the benefits of neurofeedback for several clinical conditions, such as attention deficit hyperactivity disorder (ADHD), learning disabilities, stroke, head injury, epilepsy, cognitive dysfunction, depression, anxiety, obsessive-compulsive disorder (OCD) and autism, among others, have been reviewed recently (Hammond, 2011). In the context of speech-related disorders, the ability to use ERP components as proxies for cortical entrainment, could complement the training and rehabilitation of linguistic abilities in language-impaired individuals. Furthermore, alternative strategies to cortical entrainment, such as ERPs could be used to facilitate the diagnosis of language related disorders. Besides, because the brain networks required for speech perception are central for language overall, it is likely that training our brain’s ability to activate phonemes and words in continuous speech transfers to our ability to read better and speak better**.** Our results suggest an overlap between the neural networks involved in both speech perception and production. Whether individuals who receive neurofeedback training in one of these domains will, in fact, experience transfer effects to the other remains unknown.

Figure 8

Illustration taken from (Brumberg et al., 2016). Synchrony between frontal and temporal cortical activity with self-generated speech output using ECoG.



Note: Illustration taken from Brumberg et al. (2016). a) Variations of synchrony between frontal (shown in black) and temporal (shown in blue) cortical regions and self-produced speech during a speech production task. Dashed horizontal gray lines identify time-points of interest. B) Spatiotemporal topography imaging illustrating indices of activation from cortical areas of interest during specific time-points. Color values represent activation indices. Overall, this figure is interesting, as it illustrates the preferential involvement of frontal/motor regions prior to speech output (i.e., programming speech motor sequences) and temporal /auditory regions more involved simultaneously and after speech output (i.e., monitoring auditory consequences). Additionally, temporal regions involvement starts even earlier, approximately at -120 ms, suggesting that auditory cortex is predicting the self-generated speech signal.

5. Conclusion

Our literature review highlighted the importance of the auditory cortex for language processing. Results from previous studies led us to hypothesize a potential relationship between ERPs in response to concurrent auditory stimulation and cortical entrainment, enabling novel ‘entrainment-like’ analyses in the absence of external auditory stimulation. Importantly, we replicate, for the first time at University of Algarve, the measure of cortical entrainment during speech perception, with potential for exciting new research directions in our lab, that explore language under continuous and ecological conditions. Furthermore, in respect to speech production, we obtained interesting results. Despite the common limitations in the field of speech production related to confounding articulatory-muscle artifacts present in the EEG signal, we were able to show entrainment to self-produced speech using a generalization strategy of cross-validation. Thus, the entrainment model created during speech perception (i.e., Audiobook condition) was successfully used to predict the speech envelope produced in the overt-reading condition, suggesting similar cognitive processes in both linguistic tasks. In conditions without external auditory stimulation (i.e., covert-reading), we obtained reliable ERP activations from the concurrent auditory tones. Nevertheless, we did not observe differences to the silent movie condition, which should involve minimal auditory processing. This aspect suggests that either the auditory involvement in covert reading is not as pronounced as expected, or the movie condition evokes similar levels of auditory processing.

Overall, our results shed light on further investigations for better studying and understanding the role of the auditory network for language, in particular under continuous ecological conditions.

6. References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367–13372. https://doi.org/10.1073/pnas.201400998

Ahissar, M., Protopapas, A., Reid, M., & Merzenich, M. M. (2000). Auditory processing parallels reading abilities in adults. *Proceedings of the National Academy of Sciences*, *97*(12), 6832–6837. https://doi.org/10.1073/pnas.97.12.6832

Aiken, S. J., & Picton, T. W. (2008). Human Cortical Responses to the Speech Envelope. *Ear & Hearing*, *29*(2), 139–157. https://doi.org/10.1097/AUD.0b013e31816453dc

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2020). Cortical entrainment: what we can learn from studying naturalistic speech perception. *Language, Cognition and Neuroscience*, *35*(6), 681–693. https://doi.org/10.1080/23273798.2018.1518534

Amitay, S., Ahissar, M., & Nelken, I. (2002). Auditory Processing Deficits in Reading Disabled Adults. *Journal of the Association for Research in Otolaryngology*, *3*(3), 302–320. https://doi.org/10.1007/s101620010093

Barajas, M., Guevara, R., & Gervain, J. (2021). The origins and development of speech envelope tracking during the first months of life-NC-ND license The origins and development of speech envelope tracking during the first months of life. *Developmental Coginitive Neuroscience, Elsevier, 48,* 100915. https://doi.org/10.1016/j.dcn.2021.100915ï

Benasich, A. A., & Tallal, P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. *Behavioural Brain Research*, *136*(1), 31–49. https://doi.org/10.1016/S0166-4328(02)00098-0

Blomert, L. (2011). The neural signature of orthographic–phonological binding in successful and failing reading development. *NeuroImage*, *57*(3), 695–703. https://doi.org/10.1016/j.neuroimage.2010.11.003

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature, 271*(5647), 746–747. https://doi.org/10.1038/271746a0

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. *Current Biology*, *28*(24), 3976-3983.e5. https://doi.org/10.1016/j.cub.2018.10.042

Broderick, M. P., Anderson, A. J., & Lalor, E. C. (2019). Semantic Context Enhances the Early Auditory Encoding of Natural Speech. *The Journal of Neuroscience*, *39*(38), 7564–7575. https://doi.org/10.1523/JNEUROSCI.0584-19.2019

Brumberg, J. S., Krusienski, D. J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A. L., & Schalk, G. (2016). Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. *PLoS ONE*, *11*(11). https://doi.org/10.1371/journal.pone.0166872

Brumm, H., & Zollinger, S. A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour*, *148*(11), 1173–1198. https://doi.org/10.2307/41445240

Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, *18*(6), 300–309. https://doi.org/10.1016/j.tics.2014.02.005

Correia, J. M., Jansma, B. M. B., & Bonte, M. (2015). Decoding Articulatory Features from fMRI Responses in Dorsal Speech Regions. *The Journal of Neuroscience*, *35*(45), 15015–15025. https://doi.org/10.1523/JNEUROSCI.0977-15.2015

Daliri, A., & Max, L. (2015). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language*, *143*, 59–68. https://doi.org/10.1016/j.bandl.2015.03.002

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, *370*(6487), 292–295. https://doi.org/10.1038/370292a0

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Eimas, P. D., & Miller, J. L. (1992). Organization in the Perception of Speech by Young Infants. *Psychological Science*, *3*(6), 340–345. https://doi.org/10.1111/j.1467-9280.1992.tb00043.x

Elshafaei, R. A., Kozou, H., Maghraby, R. M., Hamouda, N. H., Adel, R., & Aly, E. (2022). Impact of central auditory processing rehabilitation on literacy and phonological awareness skills in dyslexic children with central auditory processing disorder: a quasi-experimental interventional study. *Senses Sci (Educ Sci Tech)*, *9*(2), 1561–1569. www.sensesandsciences.comArticle

Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to Natural Grammar Learning: Preliminary Evidence from 4-Month-Old Infants. *PLoS ONE*, *6*(3), e17920. https://doi.org/10.1371/journal.pone.0017920

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, *105*(37), 14222–14227. https://doi.org/10.1073/pnas.0806530105

Ghitza, O., & Greenberg, S. (2009). On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica*, *66*(1–2), 113–126. https://doi.org/10.1159/000208934

Gillis, M., Kries, J., Vandermosten, M., & Francart, T. (2023). Neural tracking of linguistic and acoustic speech representations decreases with advancing age. *NeuroImage*, *267*, 119841. https://doi.org/10.1016/j.neuroimage.2022.119841

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Hammond, D. C. (2011). What is Neurofeedback: An Update. *Journal of Neurotherapy*, *15*(4), 305–336. https://doi.org/10.1080/10874208.2011.623090

Hernández, D., Puupponen, A., & Jantunen, T. (2022). The Contribution of Event-Related Potentials to the Understanding of Sign Language Processing and Production in the Brain: Experimental Evidence and Future Directions. *Frontiers in Communication*, *7*(750256). https://doi.org/10.3389/fcomm.2022.750256

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, *69*(3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019

Howard, M. F., & Poeppel, D. (2010). Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. *Journal of Neurophysiology*, *104*(5), 2500–2511. https://doi.org/10.1152/jn.00251.2010

Kober, S. E., & Neuper, C. (2012). Using auditory event-related EEG potentials to assess presence in virtual reality. *International Journal of Human-Computer Studies*, *70*(9), 577–587. https://doi.org/10.1016/j.ijhcs.2012.03.004

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources̊: a review and synthesis. *Biological Psychology*, *45*(1–3), 19–56. https://doi.org/10.1016/S0301-0511(96)05221-0

Kudela, P., Boatman-Reich, D., Beeman, D., & Anderson, W. S. (2018). Modeling Neural Adaptation in Auditory Cortex. *Frontiers in Neural Circuits*, *12*. https://doi.org/10.3389/fncir.2018.00072

Lakatos, P., Chen, C.-M., O’Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron*, *53*(2), 279–292. https://doi.org/10.1016/j.neuron.2006.12.011

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology*, *94*(3), 1904–1911. https://doi.org/10.1152/jn.00263.2005

Larsson, R. (2017). *Neural Entrainment to Speech Analyzed with EEG* [Bachelor Thesis, Linköping University]. http://www.ep.liu.se/.

Ligges, C., & Lehmann, T. (2022). Multiple Case Studies in German Children with Dyslexia: Characterization of Phonological, Auditory, Visual, and Cerebellar Processing on the Group and Individual Levels. *Brain Sciences*, *12*(10), 1292. https://doi.org/10.3390/brainsci12101292

Lovenia, H., Tanaka, H., Sakti, S., Purwarianti, A., & Nakamura, S. (2019). Speech Artifact Removal from Eeg Recordings of Spoken Word Production with Tensor Decomposition. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, *2019-May*, 1115–1119. https://doi.org/10.1109/ICASSP.2019.8682414

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Mifsud, N. G., & Whitford, T. J. (2017). Sensory attenuation of self-initiated sounds maps onto habitual associations between motor action and sound. *Neuropsychologia*, *103*, 38–43. https://doi.org/10.1016/j.neuropsychologia.2017.07.019

Miller, C. A. (2011). Auditory Processing Theories of Language Disorders: Past, Present, and Future. *Language, Speech, and Hearing Services in Schools*, *42*(3), 309–319. https://doi.org/10.1044/0161-1461(2011/10-0040)

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*, *37*(8), 2767–2783. https://doi.org/10.1002/hbm.23206

Osterhout, L., McLaughlin, J., & Bersick, M. (1997). Event-related brain potentials and human language. *Trends in Cognitive Sciences*, *1*(6), 203–209. https://doi.org/10.1016/S1364-6613(97)01073-5

O’Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. https://doi.org/10.1093/cercor/bht355

Palana, J., Schwartz, S., & Tager-Flusberg, H. (2022). Evaluating the use of cortical entrainment to measure atypical speech processing: A systematic review. *Neuroscience & Biobehavioral Reviews*, *133*(104506), 104506. https://doi.org/10.1016/j.neubiorev.2021.12.029

Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*(320). https://doi.org/10.3389/fpsyg.2012.00320

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Pérez, A., Carreiras, M., & Duñabeitia, J. A. (2017). Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Scientific Reports*, *7*(1), 4190. https://doi.org/10.1038/s41598-017-04464-4

Pérez, A., Davis, M. H., Ince, R. A. A., Zhang, H., Fu, Z., Lamarca, M., Lambon Ralph, M. A., & Monahan, P. J. (2022). Timing of brain entrainment to the speech envelope during speaking, listening and self-listening. *Cognition*, *224*, 105051. https://doi.org/10.1016/j.cognition.2022.105051

Philippsen, A. K., Reinhart, R. F., & Wrede, B. (2014). Learning how to speak: Imitation-based refinement of syllable production in an articulatory-acoustic model. *4th International Conference on Development and Learning and on Epigenetic Robotics*, 195–200. https://doi.org/10.1109/DEVLRN.2014.6982981

Plunkett, K. (1997). Theories of early language acquisition. *Trends in Cognitive Sciences*, *1*(4), 146–153. https://doi.org/10.1016/S1364-6613(97)01039-5

Salmelin, R. (2007). Clinical neurophysiology of language: The MEG approach. *Clinical Neurophysiology*, *118*(2), 237–254. https://doi.org/10.1016/j.clinph.2006.07.316

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/j.tins.2008.09.012

SHARMA, M., PURDY, S., NEWALL, P., WHELDALL, K., BEAMAN, R., & DILLON, H. (2006). Electrophysiological and behavioral evidence of auditory processing deficits in children with reading disorder☆. *Clinical Neurophysiology*, *117*(5), 1130–1144. https://doi.org/10.1016/j.clinph.2006.02.001

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21. https://doi.org/10.1186/1471-2202-10-21

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Tremblay, P., Deschamps, I., & Gracco, V. L. (2016). Neurobiology of Speech Production. In *Neurobiology of Language* (pp. 741–750). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00059-6

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study. *Child Development, 75*(4), 1067–1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x

Ünalgan, S. A. (2022). Online evaluations of reading in dyslexia diagnosis: auditory mismatch negativity (MMN) event related potential (ERP). In *Anatomy: International Journal of Experimental & Clinical Anatomy* (Vol. 16).

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. *Journal of the Association for Research in Otolaryngology, 19*(2), 181–191. https://doi.org/10.1007/s10162-018-0654-z

Wei, Q., Zhu, S., Wang, Y., Gao, X., Guo, H., & Wu, X. (2019). Maximum Signal Fraction Analysis for Enhancing Signal-to-Noise Ratio of EEG Signals in SSVEP-Based BCIs. *IEEE Access, 7,* 85452–85461. https://doi.org/10.1109/ACCESS.2019.2925078

Wikman, P., Sahari, E., Salmela, V., Leminen, A., Leminen, M., Laine, M., & Alho, K. (2021). Breaking down the cocktail party: Attentional modulation of cerebral audiovisual speech processing. *NeuroImage, 224,* 117365. https://doi.org/10.1016/j.neuroimage.2020.117365

Žarić, G., Fraga González, G., Tijms, J., van der Molen, M. W., Blomert, L., & Bonte, M. (2014). Reduced Neural Integration of Letters and Speech Sounds in Dyslexic Children Scales with Individual Differences in Reading Fluency. *PLoS ONE, 9*(10), e110337. https://doi.org/10.1371/journal.pone.0110337

Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of Frequency Domain Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00274

Zoefel, B., & VanRullen, R. (2015). The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound. *Frontiers in Human Neuroscience, 9*(DEC651). https://doi.org/10.3389/fnhum.2015.00651