UNIVERSITY OF PADOVA

Department of General Psychology

Department of Developmental Psychology and Socialization

Master Degree in Cognitive Neuroscience and Clinical Neuropsychology

Final dissertation

Using sentence context and mouth cues to aid speech comprehension: an electroencephalographic study on Cochlear Implant users

Supervisor:
Prof. Francesca Peressotti

Co-supervisor:
Dr. Simone Gastaldon

Candidate: **Noemi Bonfiglio**
Student ID number: **2002425**

Academic Year: 2021/2022

Contents

Introduction

Speech comprehension involves several interacting processes which allow to transform a series of sounds in meaningful units, to combine such units and extract complex messages. Even if we are not consciously aware of it, these processes are influenced by variables which can facilitate (or hinder) comprehension. First, our brain is not a passive receiver of information, but it uses actively stored knowledge and contextual cues to predict upcoming input (or features of it; Clark, 2013). The possibility to predict allows to process part of the information before encountering it, ultimately leading to more efficient processing. Second, speech perception is often multimodal, and information coming from different sensory modalities are integrated by the brain to enhance processing efficiency. In this context, an important source of information is seeing the mouth articulatory movements of the speaker, which have been shown to aid sound recognition and lexical access (Peelle and Sommers, 2015). While there is extensive literature on the role of each of these variables in comprehension, little is known about possible interactions between them. In fact, prediction is studied mostly with presentation of auditory-only or written sentences, and it is unclear whether the presence of visual information would impact such process. Among the few existing studies examining this issues, Brunelliere et al. (2022) found that neural oscillations were more strongly modulated by sentence context during comprehension of audiovisual sentences compared to auditory-only sentences. However, these results have been interpreted in terms of enhanced attentional engagement in the audiovisual condition rather than interaction between prediction and audiovisual speech processing. While attention likely has a role in predictive processes, possible interactions specifically in the linguistic domain remain to be explored.

Sentence context and visual information are particularly useful in difficult listening situations, such as in noisy environments or when the auditory input is poor (Davis and Johnsrude, 2007; Beauchamp, 2016). A particular population who continuously face adverse listening conditions is represented by deaf people with cochlear implant (CI). CI users receive an auditory input which lacks important speech features and consequently is more difficult to process (Hunter and Pisoni, 2021). Thus, they might compensate by exploiting other sources of information to a greater extent with respect to people with normal hearing. This has been consistently demonstrated in regard to processing of visual speech, which is enhanced in this population (e.g., Rouger et al., 2007). In contrast, data about linguistic prediction in CI users are more scarce and inconsistent, both because of large individual variability in this special population and because of a lack of studies on the topic. Furthermore, among the few studies examining predictive processes in CI users, no studies to date have used electroencephalography (EEG) to explore the neural correlates of prediction in this population.

This manuscript presents the methodology and some preliminary data from a research project which is currently in progress. The aim of this project is to examine possible group differences between CI users and people with normal hearing in specific electrophysiological correlates of prediction, both in the presence and in the absence of visual speech cues. This comparison might be particularly interesting to shed light on the functional changes that occurs in the brain in response to continuous exposure to a degraded speech input. However, because data collection and EEG analyses are still in progress, this manuscript presents only behavioural data. Nevertheless, it also presents a theoretical overview on the topics we investigate in the project, including the EEG measures we will consider in the final analyses.

To examine effects of sentence context and mouth cues and possible interactions between them, we designed a 2x2 experimental paradigm in which each participant is exposed to each possible combination of the two variables. With regard to predictability, sentence frames (i.e. the sentences preceding the last word) can be Low-Constraint (LC, which do not allow prediction of the last word) or High-Constraint (HC, which favours prediction of the last word). Following previous literature (e.g., Gastaldon et al., 2020), analyses on neural oscillations will focus on the silent gap between the sentence frame and the last word (the target), under the rationale that differences between LC and HC sentences should reflect predictive processes occurring during this gap. Also, the N400, a negative event-related potential (ERP) response peaking around 400 ms after stimulus onset, has been shown to reflect word predictability across literature, i.e., less predictable targets elicit a more negative deflection relative to more predictable ones (e.g., Van Petten and Luka, 2012). Since targets in this study are identical in the LC and HC conditions, any difference in such response can be attributed to differences in sentence context, which influences how a word is processed. Regarding visual speech cues, audiovisual stimuli (i.e., videos of a person speaking) are presented with the mouth fully visible (v+ condition) or covered by a grey rectangle which impedes to see lip movements (v- condition). Since sentences were matched across conditions for other possibly relevant variables (e.g., lexical frequency, number of syllables) and features of videos (e.g., voice tone and head position of the speaker) were held constant across conditions, this design allows to isolate how predictability and visual speech modulate electrophysiological activity during speech comprehension. Also, participants have to simply listen and comprehend the sentences, which facilitates the exclusion of other possible processes such as response selection or motor response preparation.

Use of predictive strategies are highly variables across individuals, and literature shows that this variability is even more pronounced within the CI population (e.g., Winn, 2016; Blomquist et al., 2021, Holt et al., 2021; Nagels et al. 2020). In this context, linguistic abilities are important predictors of use of sentence context to comprehend speech, since the ability to predict requires good vocabulary knowledge, efficient lexical access and ability to use grammatical rules. To investigate whether and how these factors are related to the neural modulations associated with prediction, we also administered to participants four tests which provided a general picture of their linguistic abilities in production and comprehension. Since people who were born deaf and were implanted during childhood sometimes show delays and difficulties in language development, it is important to consider general linguistic abilities when interpreting the results. These measures might contribute to explaining which factors are important for prediction also in people with normal hearing, as well as to examining if the relevant factors differ in CI users.

The manuscript is organized as follows. The first chapter discusses what is prediction in language comprehension and its benefits. The second chapter examines the neural correlates of prediction and audiovisual speech integration, including previous studies that associate modulations in alpha and beta oscillations and in N400 ERP responses to predictive processes. The third chapter presents previous studies on prediction and audiovisual speech processing in CI users. Finally, the fourth chapter presents the methods of the research project and the behavioural data we have collected from a sample of participants with CI (N = 11). As this project is still currently in progress, the results discussed here will be limited. Nevertheless, the data presented will constitute a starting point to discuss some important aspects of language comprehension with a CI, and how it could be improved in everyday situations.

Chapter 1. Prediction in language comprehension

1.1 What is prediction in language comprehension?

In the most general sense, prediction is the process by which the comprehender pre-activates linguistic information before the associated input is perceived. Such linguistic information can be a concept or a word, with its orthographical or phonological form, but also other high-level information, as will be discussed in this chapter. Thus, prediction is distinguished by integration of a word into the sentence context, as in the latter case the word has been already perceived. However, it is not straightforward to experimentally isolate one process from the other (Pickering and Gambi, 2018). Classically, psycholinguistic experiments have compared through various measures the processing of words with high and low predictability with respect to the preceding sentence context. Usually, predictability is estimated by asking to an independent group of participants to complete the sentence and taking the proportion of responses associated with each word. This measure corresponds to the Cloze Probability (CP) of that word given the sentence context. Words that are more often reported by responders present a high CP and result more predictable. CP is influenced by the semantic constraint of a sentence. For instance, the CP of the word "cow" is high when completing the sentence "The farmer milks a…", relative to the sentence "The child draws a…". The former sentence is highly constraining, and therefore "cow" is a very likely completion; in contrast, the latter sentence is low constraining and allows for many possible completing words.

Highly predictable words have been shown to cause a facilitation with respect to low predictable words, that is reflected in reaction times, for example in lexical decisions (Schwanenflugel and Shoben, 1985), in reading times (Ehrlich and Rayner, 1981) and event-related potential (ERP) components such as the N400 (Kutas et al., 1984). As pointed out by Pickering and Gambi (2018), these effects might be explained both by integration and prediction, since those words are more coherent with the knowledge activated by the sentence context and thus easier to integrate. However, other paradigms have more strongly demonstrated that it is possible to pre-activate words, by revealing effects of predictability prior to word presentation. For example, in the visual word paradigm, participants are shown a display with different images and simultaneously they hear a sentence while their eye movements are recorded through eye-tracking. By exploiting this technique, Altmann and Kamide (1999) found that in highly constraining contexts participants look at the image corresponding to the target well before its presentation. Other studies have examined neural correlates of predictability observing effects prior to the presentation of the target words. For example, also the amplitude of the N400 in response to the article preceding the target was modulated (Wicha et al., 2004; DeLong et al., 2005; Martin et al., 2018), and neural oscillations during a silent gap between the target and the preceding word were shown to be modulated as a function of the preceding sentential constraint (Rommers et al., 2017; Wang et al., 2018; Terporten et al., 2019; Gastaldon et al., 2020).

In most accounts, prediction is viewed as an "active" process, requiring attention and cognitive effort, that is usually outweighed by its benefits. Indeed, pre-activating the information that will be received makes processing faster and easier. Thus, prediction is commonly distinguished from priming, the phenomenon by which the exposure to a stimulus influences processing of a following stimulus, that is ascribed instead to implicit memory processes such as passive spreading of activation between conceptual or linguistic representations (DeLong et al., 2014). The two processes are defined differently also at the operational level, with very different paradigms used to study them. Priming is usually elicited with the presentation of single words in sequence, in order to isolate the effects of their relations (e.g., orthographic, phonological or semantic); prediction is more often studied during sentence comprehension and does not depend on associations between single words but on sentence context as a whole. The requirement of working memory and attention to maintain and update contextual representation has led to the traditional notion of prediction as an active and conscious process. In contrast, the passive pre-activation associated with priming does not require neither attention or consciousness and, in fact, priming effects occur even when the prime is presented too briefly to reach awareness.

Despite this broad differentiation, it should be noted that the traditional dichotomy between strategic, conscious and effortful processes on one side and automatic, implicit and effortless processes on the other side has now been abandoned by most researchers. Kuperberg and Jaeger (2016) point out that the tendency to maintain some of the assumptions of the early models of prediction and priming, based on such dichotomy, might be not useful in current research. For example, these authors claim that prediction is not necessarily conscious and strategic but might occur implicitly and notice that, on the other hand, even priming is sometimes influenced by attention and strategies (Hutchinson, 2007). Nowadays, a more nuanced approach recognizes a spectrum of predictive processes that might be more or less under attentional control according to the circumstances (DeLong et al., 2014). As a consequence, current research is more focused on the cognitive and neural mechanisms underlying prediction rather than on the issue of its automaticity.

1.1.1 Prediction by production

The model proposed by Pickering and Gambi (2018) is a good example of how recent accounts integrate automatic and controlled mechanisms of prediction. The model makes a distinction between “prediction by association” and “prediction by production”. In the first case, pre-activation of words is due to spreading of activation across associated words, as in priming effects. Critically, this form of prediction is not constrained by linguistic context: spreading of activation is assumed to reach many related words regardless of their plausibility within the sentence. As a consequence, the pre-activated words often do not match what will be perceived, limiting the benefit of this pre-activation for sentence comprehension. In contrast, “prediction by production” is a mechanism that more properly accords with the notion of active prediction discussed above. It involves a covert imitation of the sentence processed up to that point via the comprehender’s production system, which is facilitated by the shared linguistic representations between comprehension and production. After imitation, the intention underlying the sentence is inferred on the basis of both linguistic and extra-linguistic context (e.g., background knowledge and relevant elements in the environment). This process is called by the authors “inverse mapping” because it goes in the opposite direction to what happens during production, when appropriate linguistic representations are derived from the intention that motivates the person to speak. Finally, assuming that the same intention will underlie also what will be said in the continuation of the sentence, the comprehender transfers it in his own production system; they thus reproduce processes involved in production, except for overt articulation, to predict the linguistic representations that will complete the sentence.

The authors use the phrase "linguistic representations" in a very general way, not necessarily encompassing the forms of the word (i.e., their phonology) but also, more often, their high-level representations such as semantics and syntax. Such representations at various levels are predicted following the stages of the production hierarchy: conceptualization, syntactic encoding, lexical selection and phonological activation. These stages are sequential but are not always completed up to phonology, and their completion might vary in different points during the processing of the sentence.

The model by Pickering and Gambi (2018) is not the only proposal of prediction through production. For example, in the PACS (Production-, Association, Combinatorial-, Simulation-based prediction) model, Huettig (2015) includes production among four interacting mechanisms of prediction, along with association (similarly to Pickering and Gambi, 2018), use of combinatorial rules and event simulation. The knowledge of the combinatorial rules of language, such as the ones that regulate syntax, is indeed crucial to identify linguistic constraints and build higher-order meaning when we comprehend a sentence. For this reason, Huettig (2015) proposes that they can also guide the pre-activation of upcoming information. Combinatorial rules are also important when we produce a sentence to build the appropriate linguistic structure, but it is still unclear if and to what extent the same combinatorial mechanisms are shared between comprehension and production. Event simulation is intended as the ability to imagine future events on the basis of present and previous experience; this is often used as a heuristic to predict the likelihood of future events on the basis of the ease with which it is simulated (Tversky and Kahneman, 1973). In the PACS model, a similar heuristic can be used to pre-activate linguistic representations by accessing to imagined perceptual representations of what we are comprehending. Interestingly, also this model entails a large degree of flexibility: because the four mechanisms are independent, it is not necessary to implement all of them in order to predict. Nevertheless, they often overlap and interact with each other when the comprehender engages in prediction.

Most of the evidence for the engagement of production mechanisms in prediction are indirect. For example, Mani and Huettig (2012) found that predictive gaze shifts of 2-years-old children correlated with their production, but not comprehension, vocabulary size. Also, the timing of ERP effects associated with prediction of syntactic or phonological representations of words are coherent with the production hierarchy mentioned above, with activation of gender preceding activation of word form (Ito et al., 2020). Martin et al. (2018) more directly tested whether the availability of the production system is necessary to predict during word-by-word reading. They compared three groups of participants who were engaged in syllable production, tongue-tapping and syllable listening while reading words presented one at a time, and found that only the first group showed a reduced prediction-related N400 effect at the article preceding the target (this effect and other neural correlates of prediction will be better discussed in the next chapter). Critically, the only difference between the syllable production group and the two control groups was the engagement of the language production system; thus, this result supports the model proposed by Pickering and Gambi (2018).

A different approach was used by Gastaldon et al. (2020), who studied prediction during speech comprehension and production planning by directly comparing the neural correlates of the two processes. The same sentences were used in both tasks across the experiment, but in one case participants had to wait and listen to the target word, while in the other case they had to produce it, completing the sentence they heard, by naming a picture. The authors analysed the neural oscillations relatively to the silent gap between the sentence frame and the target (the final word) and found spatiotemporal correlations in prediction-related alpha and beta desynchronization between the two conditions, suggesting that the two processes share at least some neural and cognitive mechanisms.

Interestingly, in a subsequent study the same paradigm was implemented to evaluate the neural correlates of prediction and production in people who stutter (Gastaldon et al., under review). Stuttering is characterized by dysfunctions in neural circuits involved in planning and executing speech; therefore, the study specifically tested the involvement of such processes in prediction during comprehension. While in the control group predictable sentences were associated with pre-target alpha and beta desynchronization (as in the first experiment) and a post-target N400 effect, in the group with stuttering these effects were weaker. Also, in this group anomalous neural patterns were found in right supplementary and associative premotor regions that are crucial for motor planning. In particular, these areas showed an alpha power synchronization that reflects reduced cortical engagement, and that was indeed associated with a reduced N400 effect, indicating a smaller facilitation in processing the target. Even though at the correlational level, these results suggest that this region, primarily involved in speech-motor control, is also actively involved in prediction processes and that its abnormal inhibition might lead to a reduced efficiency in word prediction. Overall, the studies performed so far indicate that production mechanisms might have an important role in prediction, and future research might further elucidate this role by means of new experimental manipulations and studies with clinical populations.

1.1.2 Flexibility of prediction

Most accounts of prediction recognize that it is a graded phenomenon, varying in the degree of attentional control involved, in the strength of predictions and in the type of information that is predicted. In practice, predictive processes are adapted flexibly to the situation and to the structure and content of the sentence that is being comprehended. However, how such flexibility is attained is still object of debate. One question is how prediction can change when new bottom-up information is encountered. For example, Van Petten and Luka (2012) propose that only the word with the highest cloze probability is initially pre-activated, and if that is disconfirmed by the incoming input (e.g., syntactic or grammatical constraints) the second word with the highest cloze probability is considered, then the third if the second is disconfirmed and so on, until the critical word is encountered. In contrast, Kuperberg and Jaeger (2016) propose that several words are pre-activated in parallel with different degrees of belief according to their probabilities, that are re-weighted accordingly when new input change the sentence context. As these authors notice, it can be difficult to experimentally distinguish whether prediction is implemented in a serial or parallel probabilistic manner. Recent experimental research has primarily focused on another issue that is source of debate in the field: which levels of representation can be predicted, and which factors influence if they will be actually predicted.

Preliminarily, it should be considered that the notion that prediction can comprise several levels of representation is not accepted by all authors. For example, Van Petten and Luka (2012) explicitly distinguish between expectation and prediction. Expectation is a more general term that indicates the anticipation of some semantic content, that may or may not be referred to specific words, while prediction involves specific lexical representations, including their physical forms. According to the authors, evidence of prediction should therefore involve non-semantic lexical features of words, such as their phonology. While there is a consistent body of evidence that people predict semantic features of words, evidence of prediction of their phonological form are more mixed and debated. The most important finding in support is the mismatch effect in response to articles, described for the first time by DeLong et al. (2005): the N400 is greater when the article "a"/"an" mismatches with the phonology of the predictable word (e.g., "an" when sentence context suggests to predict "kite"). This seems to indicate that participants pre-activated this word also at the phonological level. The reliability of such result has been object of discussion, also because a multi-lab replication study failed to find robust effects (Nieuwland et al., 2018), but subsequent re-analyses and new studies supported the presence of a consistent effect of phonological mismatch at articles (Urbach et al., 2020; Ito et al., 2020).

Van Petten and Luka (2012) recognize that expectations of semantic features facilitate the processing of subsequent related words but interpret this as an index of a facilitated integration, because the pre-activation of such features remains at the conceptual level and therefore does not constitute prediction. Indeed, there is partial evidence that the pre-activation of conceptual and lexical representations can be dissociated, at least at the level of ERP components (e.g., Thornhill and Van Petten, 2012). Nevertheless, whether pre-activation of semantic features constitutes a different process with respect to prediction, as claimed by Van Petten and Luka (2012), is an issue based mostly on the choice of the terminology. In fact, other researchers propose a much less strict definition of prediction. For example, Kuperberg and Jaeger (2016) claim that context at multiple levels of representations can be used to predict upcoming input at multiple other levels, including semantic, syntactic and phonological, and that high-level information such as meaning can lead to the pre-activation of low-level features such as phonology.

Similarly, the "prediction by production" mechanism proposed by Pickering and Gambi (2018) can stop at different levels of the production hierarchy in different points of the heard (and imitated) sentence. A good example of this variability is the difference between what is typically predicted at the beginning and at the end of a sentence. At first, for example when only the subject has been presented, comprehenders are only able to predict representations associated with early stages of production such as semantics and syntax (for example, that an action expressed by a verb will follow the subject). In contrast, at the end of the sentence, when only one last information is missing, it is more likely that they predict also representations at later stages such as the phonology of the word. This is one of the reasons why in psycholinguistic experiments studying prediction, including the one discussed in this manuscript, the very last word of the sentence is used as target in such a way that a more complete prediction is facilitated. A crucial factor in this example is the semantic constraint of the preceding context: in high-constraining sentences (and, in part, also in low-constraining sentences), constraint is generally low at the beginning and increases as the sentence unfolds, when the representation of sentence meaning becomes richer and almost complete. Indeed, Kuperberg and Jaeger (2016) identify constraint as one of the major determinants of both the strength of the pre-activation and the level of representation that is pre-activated.

Another factor pointed out by the same authors is the speed of the sentence that is being comprehended: slower presentations are associated with greater contextual facilitation, which might reflect a more precise representation of the predicted word. In the framework proposed by Pickering and Gambi (2018), this association is explained by the time required to the comprehender to reproduce the sentence in their production system: when rate of word presentation is faster, early stages of production are more likely to be completed than later stages. In practice, in such situations the comprehender does not have enough time to predict the form of the word (its phonology or orthography) before it is encountered, while they are able to predict its semantics. According to the authors, this often happens in everyday situations, because natural speech has usually a fairly fast rate. In contrast, in psycholinguistic experiments the presentation rate is slower and participants are more likely to reach phonological representations of predicted words (especially of the last word, as mentioned before). The role of time constraints on "prediction by production" also implicates that slower producers are less likely to predict words at later stages of representation or to engage in active prediction at all. This proposal is in line with the study by Gastaldon et al. (under review), whose results suggest that prediction is less efficient in people who stutter. Pickering and Gambi (2018) also point out that predictions can be generated well in advance with respect to the actual perception of the word, and are updated in response to new contextual information in a continuous manner. As a consequence, high and low-level representations might be dissociated and predicted at different stages during sentence processing, in such a way that comprehenders might predict the meaning of a word well before the immediately preceding word, while they might represent its phonology much later. According to the authors, such dissociation is similar to the one observed in production, in which speakers plan semantics in advance with respect to phonology.

Finally, because prediction is a costly process, the amount of available cognitive resources influences if the comprehender will engage in it. Indeed, the models based on "prediction by production" claim that this mechanism is optional and not necessary to sentence comprehension. In particular, in the framework proposed by Pickering and Gambi (2018), both time constraints and lack of resources (e.g., in dual task situations) might impede the use of the production system in prediction and limit it to the association mechanism, that is effortless but less accurate. The authors also note that the resources used to covertly imitate the sentence and predict its completion are likely shared with other aspects of comprehension. This means that prediction by production might, in theory, interfere with comprehension processes if resources are limited. This does not happen because comprehension is usually prioritized. In fact, the goal of the comprehender is an additional factor influencing the likelihood of prediction, as shown by experiments manipulating task instructions (Kuperberg and Jaeger, 2016). In everyday situations, the goal of comprehension is usually to infer the message conveyed by the speaker, and in such contexts both high cognitive load and individual characteristics such as slow processing speed make active prediction less likely (Pickering and Gambi, 2018).

Overall, a comprehensive view of prediction, comprising several types of information that can be predicted, is now directing the research towards the study of the different factors that can influence it. To conceptualize prediction in such broad framework of interacting representational levels and mechanisms can be a fruitful approach to clarify its role in language comprehension, and ultimately better understand the great adaptability that characterizes it in everyday life.

1.2 Benefits of prediction

In the theoretical views discussed so far, prediction is not a necessary process for the construction of meaning, but a valid aid to it. Such an optional process therefore has to bring some benefits in order to be implemented, and indeed evidence of increased processing speed and efficiency in conditions favouring prediction are now robust. The benefits of formulating hypotheses about the upcoming content seem particularly important in certain circumstances, that are common in everyday life. For example, it might reduce much of the ambiguity that is abundant in language use, ultimately reducing working memory load (Huettig, 2015). It might also aid perception and help disambiguate words in noisy environments or when auditory input is not optimal (e.g., Obleser and Kotz, 2009). This is particularly relevant for populations with hearing impairments, such as deaf people with cochlear implant (CI), which are the focus of the study here presented. For these people, the auditory input is poorly encoded and lacks many of the spectral information that characterizes speech. For this reason, perceptual benefits of prediction will be discussed more in depth in the last section of this paragraph, while prediction specifically in CI users will be better discussed in the third chapter.

When discussing benefits and costs of prediction, the underlying assumption is that this process requires cognitive resources. If the predicted representation matches the input, the benefits in terms of processing speed and efficiency outweigh such consumption. If it does not match, the comprehender "wastes" both the cognitive resources used to formulate the prediction and the ones needed to dismiss the predicted representations and build new ones based on the real input. Such waste is generally referred to as the "cost" of prediction. Kuperberg and Jaeger (2016) propose a framework to understand how the benefits of prediction can outweigh its costs. The fundamental assumption of this approach is that cognitive (and metabolic) resources are used in a rational fashion, according to a "utility function" that weigh advantages and disadvantages of a given process. According to the authors, it is possible for a comprehender to maximize the utility of prediction by flexibly modulating its implementation. For instance, they can save resources by pre-activating only the levels of representations that are useful to the goal. More generally, utility of prediction depends on a variety of factors, including the specific task at hand, availability of resources, structure of the input and its speed. Nevertheless, Kuperberg and Jaeger

(2016) underline that the existing evidences of prediction during language processing and conversations suggest that overall its utility is relatively high in everyday situations.

If prediction is routinely implemented as a valid aid to language comprehension, it is natural that sometimes unexpected inputs disconfirm our hypotheses; interestingly, DeLong et al. (2014) suggest that the processes required to correct the error, although costly, might also contribute to redirecting attention and learning. A different approach to prediction, the predictive coding framework, leads this concept to the extreme, by claiming that it is precisely the computation of prediction errors that biases the brain towards making correct inferences through a process of error minimization (Friston, 2010). Unlike the models discussed so far, this view proposes that the formulation and optimization of hypotheses is a general mechanism by which we can interpret the functioning of mind and brain. Thus, when examining prediction, a distinction between two broad approaches can be delineated. On the one side, an important strand of literature focuses on its benefits in those cases in which the hypotheses are confirmed, studying the effects of facilitation on processing the predicted word (see Van Petten and Luka, 2012 for a review). On the other side, the predictive coding framework highlights the evolutionary advantages of prediction, regardless of how much it matches the upcoming input, on the basis of the "Bayesian brain hypothesis", for which perception consists in optimizing the internal representation of the world by continuously comparing it to sensory inputs (Friston, 2010).

1.2.1 How does prediction increase processing speed and efficiency?

Intuitively, what is predicted is partly processed in advance and therefore requires less processing effort after it is actually encountered. One way to view this principle is as domain-independent: in the same way in which predictable words are associated with higher speed, there is ample evidence that such relation is present also for a variety of non-linguistic stimuli and tasks (Smith and Levy, 2013). According to this hypothesis, predictability modulates efficiency of cognitive processing in many different situations, specifically affecting linguistic processes during sentence comprehension. Interestingly, following this assumption, Smith and Levy (2013) propose that words are not the fundamental units involved in prediction, even during language processing. According to their "highly incremental processing" model, we basically process sequences of sub-word fragments that have their own conditional probability (that is, the probability of each sub-fragment is conditioned by the preceding context and phonemes; the authors equate this to predictability), and these probabilities together affect processing speed of the word as a whole. While this view has been proposed on the basis of reading times results, it might be particularly valid for auditory comprehension, in which speech unfolds over time and the listener might start to process a phoneme prior to encountering the subsequent phonemes of the same word. Similarly, they might also predict word continuation according to both semantic context and prior phonemes, pre-activating several possibilities in a graded way (similarly to the proposal by Kuperberg and Jaeger, 2016). Under this hypothesis, predictable words would increase processing speed because they contain predictable fragments, that require quicker processing (Smith and Levy, 2013).

If prediction is implemented in a manner that is incremental (Smith and Levy, 2013) and continuous (Pickering and Gambi, 2018), it is interesting to ask how it can be sufficiently fast to be beneficial. In fact, it is obvious that if the process of pre-activating relevant information was too time-consuming, the related input would be encountered before its completion, making it eventually useless. To clarify this issue, it should be considered what is the contextual knowledge on which predictions are based. Much of the information conveyed by a sentence is not explicit but is inferred from word meanings and common use. Van Berkum (2009) explains this concept saying that, for example, if someone says “some of the guests left” they do not need to add “but not all guests”, because that is implicit in how “some” is commonly used. Context-sensitive inferences like this do not require time, because active representations in working memory work as automatic recall cues for related information in long-term memory (Van Berkum, 2009). Thus, while prediction itself is assumed to be active and costly, it can be implemented in a very fast way because the knowledge used as “context” is not consciously recalled from long-term memory but is readily available.

Another proposed mechanism by which prediction can increase processing speed is constraining low-level input (such as phonological word form) from high-level linguistic representations (such as the semantics of the sentence). According to Christiansen and Chater (2016), this is particularly useful because the existing constraints in perceptual processing and memory (what the authors call the “now-or-never bottleneck”) makes the input be rapidly processed in a “chunk-and-pass” fashion. The input is chunked into more and more abstract representations: sounds are recoded into phonemes, which in turn are recoded into words, then phrases and sentences. Thus, long sequences of information are maintained together at an abstract level of the linguistic hierarchy, leaving space in the perceptual system to keep track of the incoming input. In such a model, constraining upcoming information through prediction is an efficient way to initiate the "Chunk-and-Pass" process as soon as possible.

The "now-or-never" constraint also implies that linguistic processing is incremental, because representations need to be built rapidly on the basis of previous information, while the input is still unfolding. An additional assumption made by the authors is that once a chunk is created it is difficult to re-process that information at lower levels of representation. Indeed, in the cases in which this is required there are significant costs in terms of slowing and difficulty in processing. This is the case, for example, when an incorrect interpretation of an ambiguous input must be revised (the classic "garden path" phenomenon). To avoid such cost, our linguistic system is pressured to choose the correct interpretation as soon as the input is processed, according to a principle of "Right-First-Time". Thus, as prediction constrains and speeds up processing, it might be particularly useful in resolving linguistic ambiguities during processing itself rather than after encountering additional information (Christiansen and Chater, 2016).

1.2.2 Prediction and auditory perception of language

The "garden path" phenomenon mentioned by Christiansen and Chater (2016) involves syntactic and semantic ambiguities, but the linguistic input can be ambiguous also at the sensory level. In these cases, semantic context can influence perception in such a way that comprehenders often remain unaware of the sensory ambiguity, both in the visual domain during reading and in the auditory domain while listening to speech. In the latter domain, a phonetic segment that is intermediate between two phonemes is perceived differently according to sentence context. This phenomenon is called "Ganong effect" (Ganong, 1980) and according to Davis and Johnsrude (2007) it reflects top-down processes that from lexical representations influence phonetic categorization at the perceptual level. An fMRI study by Guediche et al. (2013) investigated this phenomenon manipulating ambiguity of the syllable sound and semantic constraint of the sentence, and found an interaction between these two variables in the neural activation of an area in the left temporal cortex encompassing part of the middle and superior temporal gyri. On the basis of their pattern of results, the authors propose that bottom-up information (the sound being heard) and top-down information (the semantic constraints) are independently weighted and then integrated in superior and middle temporal gyri. However, while the neural and behavioural results of this study support an influence of semantic representations on perceptual processes, they do not tell if this is exerted through predictive processes or other top-down mechanisms. Nevertheless, it is likely that prediction is one possible way in which low-level processing can be constrained from higher-level representations, as proposed by Christiansen and Chater (2016), thus facilitating the correct interpretation of ambiguous sounds.

This constraining process is an important aid to perception also in adverse listening conditions such as in presence of background noise. In this situation, knowledge of semantic content of the sentence has been shown to greatly improve perceived clarity, and listeners who are familiar with sentence content also estimate background noise as being quieter, even though this estimation is based only on basic sound perception (Davis and Johnsrude, 2007). In the same vein, many studies have explored the perception of distorted or degraded speech (such as sine-wave and noise-vocoded speech, NV) to elucidate the mechanisms by which the speech recognition system is robust to great variability in the auditory input. As for speech masked with noise, distorted speech is well intelligible if listeners are informed of the identity of the original sentence or even, in some cases, of the mere fact that what they are hearing is speech. According to Davis and Johnsrude (2007) these results reflect the influence of high-level knowledge and expectations on a basic perceptual process, the “perceptual grouping” of speech. This process allows humans to perceive speech as a single stream of sounds, even though it contains several different acoustic elements that can be noisy or distorted, and to automatically attribute it to a single source (a human voice, in contrast to background noises).

The same authors propose that, in addition to perceptual grouping, a mechanism of perceptual learning allows listeners to comprehend degraded speech. This is suggested by the fact that the number of words that participants can correctly repeat from NV speech rapidly increases after exposure to additional sentences of this type. Such learning generalizes to words that were not previously heard in NV form, suggesting that it consists in “re-tuning” phonetic representations, regardless of the specific words the sounds are embedded in. Also in this case, top-down influences have a crucial role in facilitating both sentence intelligibility and perceptual learning, especially if the same sentence is presented first in clear form and then in NV form. However, for words (as opposed to pseudowords) a previous clear presentation is not a necessary requirement to generate learning effects, suggesting that listeners are able to derive word phonological representations from the distorted input itself. According to Davis and Johnsrude (2007), this is possible due to predictive processes that constrain the interpretation of the distorted speech by pre-activating relevant lexical representations. According to this proposal, the presence of clear speech anticipating the content of the distorted sentence is not as crucial as the presence of semantic constraints that make it possible to predict such content.

An fMRI study by Obleser and Kotz (2009) explored this hypothesis by manipulating level of degradation and predictability (measured as high/low cloze probability) of NV sentences, and found patterns of neural activity that are compatible with the role of prediction proposed by Davis and Johnsrude (2007). Higher intelligibility was found to increase BOLD response in bilateral superior temporal gyrus (STG) and superior temporal sulcus (STS). However this effect was modulated by cloze probability: it was localized in anterior and posterior STG/STS for sentences with low-cloze targets, while it was restricted to middle regions of this area for sentences with high-cloze targets. Also, the left inferior frontal gyrus (IFG) was activated only for sentences with low-cloze targets and only with increasing intelligibility. These results are particularly interesting because they allow to make some hypotheses about the neural mechanisms by which prediction can facilitate perception. In particular, the authors propose that prediction, by cascading from semantic to phonological representations, facilitates the conversion of acoustic input into a phonological percept, reducing and restricting this process to a smaller region of STG/STS. Similarly, they propose that the activation of IFG for unconstraining sentences might indicate an extra processing effort with respect to constraining sentences, as this region is thought to be involved in integration processes that might result unnecessary when semantic and phonological features can be predicted.

An extreme case of influence of top-down processes on auditory perception is phonemic restoration, the effect by which listeners are capable of perceiving a phoneme even if it has been replaced by a non-speech sound within a word. In certain aspects, this is similar to the Ganong effect, as in both cases a particular segment of the input that might potentially cause difficulties in comprehension is quickly “resolved” through top-down processes. Even if this phenomenon has been demonstrated independently from sentence context (e.g., Bhargava et al. 2014), it has been shown to be influenced by it. For example, in a study by Groppe et al. (2010) phonemic restoration was found to be facilitated and biased towards a certain word if the preceding sentence context was informative rather than ambiguous. However, as for the Ganong effect, it is not yet clear which top-down mechanisms mediate effects of context. Prediction might be particularly advantageous to this aim, but it is likely to be one of many possible mechanisms that might be implemented according to availability of resources and characteristics of the situation.

In conclusion, prediction has been shown to be an efficient strategy by which semantic and syntactic knowledge can constrain upcoming input, providing benefits in terms of cognitive resources and processing speed. Studies examining effects of context on speech perception complement this view by shading light on an additional level in which prediction can be beneficial. These studies suggest that high-level processes can influence perceptual processing of language in a top-down fashion, facilitating intelligibility when the acoustic input is ambiguous, noisy, distorted or even incomplete. The cognitive mechanisms underlying these phenomena are still not completely understood, but the study of prediction in suboptimal perceptual contexts, including studies on clinical populations, might be a promising approach to elucidate this issue.

Chapter 2. Neural correlates of prediction and audiovisual speech processing

2.1 Electrophysiological correlates of prediction

Electroencephalography (EEG) is a widely used method in the study of cognitive processes, including language processing. In fact, the high temporal resolution of this technique allows to precisely identify the timing of effects of experimental manipulations, thus providing important insights about the cognitive mechanisms involved (Luck, 2014). EEG measures brain electrical activity through electrodes placed on the scalp, and it detects in particular voltage fluctuations given by post-synaptic potentials of large populations of cortical neurons. The following discussion will focus on two EEG measures that are particularly relevant in the literature about prediction and that we exploit in our study: event-related potentials (ERPs) and power of neural oscillations at certain frequencies, detected through time-frequency analysis.

In regard to prediction in sentence comprehension, the effects on neural activity can be measured either before or after the target, which is the linguistic item (usually a word) that is supposed to be predicted. Both ERPs and variations in neural oscillations can be computer before and after the target, with different interpretations with regard to the underlying cognitive process. In particular, modulations observed after the target can be considered indirect effects of prediction, as they reflect processes that act on the critical word rather than the predictive processes themselves. Nevertheless, post-target measures are informative because the pre-activation of features of the word changes and facilitates its processing when it is encountered. On the contrary, modulations detected before the target are thought to reflect more directly content pre-activation, especially if other potentially relevant processes are held constant across experimental conditions. Importantly, pre-target modulations can be related to post-target measures of facilitation to further support their functional interpretation. Thus, both approaches are valuable and complementary to each other, and can be differently exploited according to the specific research question at hand.

ERPs are measures of activity time-locked to a certain event, such as the onset of a word. This activity is isolated from the noise by averaging together segments of signal corresponding to several trials of the same experimental condition, in such a way that random variations cancel each other and the signal that is constant across trials is maintained. The ERP response is usually composed by many different waves (or components), some of which are constant for certain stimuli and have been defined by researchers according to their polarity and latency. For example, the N100 (or N1) response is a negative peak around 100 ms after stimulus onset and is considered an “early” component, reflecting low-level sensory processing. One ERP component that is particularly relevant for prediction is the N400. It is considered a “late” component, associated especially with semantic processing (though its functional significance is debated, see Kutas and Federmeier, 2011 for a review), and emerges as a negative deflection peaking around 400 ms after the onset of a word, and it is maximally visible in centro-parietal electrodes. Modulations of this component (the so-called N400 effect) have been proposed to reflect either how easily the critical word is integrated into the semantic context (e.g., in Hagoort, 2008) or how easily its lexical representation is accessed (e.g., in Szewczyk and Schriefers, 2018). In particular, easier words elicit a reduced N400 peak respect to control words, while more difficult words elicit a more pronounced (more negative) N400. N400 effects of facilitation can be elicited by a variety of manipulations, including repetition, semantic priming, plausibility and predictability. Therefore, many studies have attempted to distinguish effects of prediction by other possible sources of facilitation.

The EEG signal reflects oscillatory activity which can be also described in terms of its phase, which is the position along the oscillation at a given moment. The ERP technique allows to detect responses which are both time-locked and phase-locked to an event, meaning that they have the same phase at each time-point. In contrast, by averaging many epochs of the signal to compute an ERP, responses with different phases will cancel each other out and the final ERP response will result flat. This limitation is overcome by the time-frequency analysis of the EEG signal, which estimates modulations at different frequencies prior to averaging, thus maintaining also non-phase-locked activity. Frequency quantifies the number of cycles of an oscillation for second (measured in Hertz, Hz), and the EEG signal can be decomposed in several oscillations with different frequencies. Such frequencies have been divided by researchers in five main bands: delta (0.1-4 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz) and gamma (>30 Hz). The amount of energy of a given frequency in the signal (power, the squared amplitude of the oscillation at that frequency) has been associated to the level of synchronization of the cortical neurons that generate the EEG signal: a power increase reflects a synchronization in neural firing, while a power decrease reflects a desynchronization at a given frequency band (Hanslmayr et al., 2012). Critically, power modulations have been associated with a variety of cognitive, motor and perceptual processes (Cohen, 2014). In regard to prediction during language comprehension, some studies analysed power modulations before the target and found a desynchronization in alpha and beta bands in presence of constraining sentence contexts respect to unconstraining contexts (e.g., Gastaldon et al., 2020; Terporten et al., 2019; Wang et al., 2018). These results are particularly relevant for the study discussed in this manuscript, in which we will consider both pre-target power modulations and post-target N400 response.

2.1.1 N400 component and prediction

Since its first discovery by Kutas & Hillyard (1980), the amplitude of the N400 has been considered to reflect influences of semantic context on word processing. In their review, Van Petten and Luka (2012) illustrate some of the early studies that have explored such influences using ERPs. In particular, the authors underline three key features of contextual effects as revealed by the N400. First, they are immediate, as they are observed before the offset of the critical word and sometimes even before the point at which it can be univocally identified. Second, they are incremental, as the facilitation for words presented serially in a sentence increases progressively (the N400 becomes smaller). Third, they are graded, as they scale with the predictability of the word to be processed (the reduction of N400 is correlated with its cloze probability). These features are in line with the idea that linguistic pre-activation underlies effects of semantic context, as they are coherent with the mechanisms of prediction previously delineated. However, a mechanism of rapid integration might as well explain effects of contextual facilitation.

A recent study by Szewczyk and Schriefers (2018) explicitly tested the hypothesis that the reduction of N400 is related to pre-activation of lexical items rather than their congruency (and ease of integration) with the sentence context. While in most studies on prediction pre-activation is correlated with congruency, these authors isolated the two factors by explicitly warning the participants that they would encounter a certain word, which could be either congruent or incongruent with the preceding context. Results showed a strongly reduced N400 in both conditions, suggesting that lexical pre-activation of the word, induced by the preceding warnings, facilitated its processing regardless its ease of integration within the context. In contrast, the classical N400 increase for incongruent respect to congruent words was observed in the control condition without preceding warnings. Such difference is attributed by the authors to a cognitive effort in accessing the lexical representation of incongruent words once they are encountered.

According to Szewczyk and Schriefers (2018), the critical difference between predictable and not predictable words is that only for the former lexical access can be anticipated during the sentence. This is the reason why an N400 effect is often detected in response to the word preceding the target. In this study, in the condition with warnings, the authors found a greater N400 at the verb preceding incongruent targets with respect to the verb preceding congruent targets. This was interpreted as an index of lexical access occurring with slightly different mechanisms in the two conditions. For congruent words, cues for lexical access were given by the preceding semantic context, in such a way that the process required no effort and resulted in a reduced N400. For incongruent words, the verb worked as a syntactic cue to access the word indicated by experimenters, leading to an effortful lexical access associated with an anticipated N400.

According to these authors, a similar mechanism might explain the effect of mismatch at articles or adjectives preceding the target (i.e., a greater N400 when these are not congruent with the expected target). This would reflect an attempt to access an appropriate lexical representation on the basis of both the preceding context and the article or adjective just encountered. In case of mismatch, this process would be effortful because the prediction must be updated according to the new information available. Interestingly, mismatch effects have been found in different languages and for both phonological and gender features, suggesting that these levels of information can be successfully predicted (e.g., DeLong et al. 2005; Wicha et al., 2004; Martin et al., 2018; Ito et al., 2020). In addition, timing of ERP effects can also contribute to exploring the mechanisms of prediction at different levels. For example, Ito et al. (2020) found that the time window of mismatch effect starts earlier (at about 250 ms) for articles mismatching for gender respect to articles mismatching for phonology (at about 450 ms). These results have been interpreted in line with the "prediction by production" mechanism, as in the production hierarchy gender is activated earlier than phonology.

Studies on mismatch effects at adjectives or articles exploit pre-target N400 to explore which level of representation can be predicted. However, post-target N400 can be also used to this aim, for example by examining the response to target words that are related to the predicted word. In one of the first study on the N400, Kutas et al. (1984) found that a word that constitutes an anomalous completion for the sentence but is semantically related to the expected word elicits an N400 that is greater respect to the expected word, but smaller respect to unrelated anomalous completions. This result suggests that prediction is not limited to a single lexical item but involves also higher conceptual levels, in such a way that pre-activated conceptual features can facilitate even words that are incongruent with the sentence context. This does not mean necessarily that post-target N400 is sensitive only to pre-activation of general concepts and not of specific lexical items, as argued by some authors (Van Petten and Luka, 2012; Thornhill and Van Petten, 2012). In fact, similar results have been observed for perceptual features of predicted objects: incongruent words referring to objects with the same shape as the expected word elicited a significantly smaller N400 with respect to completely unrelated objects (Rommers et al., 2013). Because perceptual features are not usually shared among conceptually related objects, their pre-activation has been associated to the prediction of a particular word.

As in the study by Ito et al. (2020), also in Rommers et al. (2013) the timing of the ERP effects provides insights about the levels of representation involved. All the unexpected words elicited an increased negativity respect to expected words very early (at about 150 ms), but the difference between unexpected words with and without the critical shape was observed only in a late time window (starting from 500 ms post-target). According to the authors, the two effects occur at different levels of processing: the first reflects a violation of expectation at the auditory level (i.e., the word "sounds" different than expected), while the second occurs only after the meaning (and the shape) of the unexpected word has been accessed. Because these effects are detected post-target, their timing follows the order by which the word is processed once it is encountered. In contrast, the pre-target effects observed in Ito et al. (2020) were driven by predicted representations, which might be accessed in the reverse order following the production hierarchy.

Overall, the studies discussed here show that the N400 component is an important marker of prediction, and that experimental manipulations are crucial factors in their interpretation.

2.1.2 Pre-target alpha and beta desynchronization in prediction

Pre-target alpha (8-12 Hz) and beta (13-30 Hz) desynchronization in constraining with respect to unconstraining sentences has been observed consistently across modalities, as it has been replicated both during written (Rommers et al., 2017; Wang et al., 2018; Terporten et al., 2019) and spoken language comprehension (Gastaldon et al., 2020; Leòn-Cabrera et al. 2022). In studies employing the written modality, words were presented one at a time for fixed duration, in such a way that it was possible to estimate in which time period the power modulations occurred. All these studies indeed found a desynchronization in a time-window immediately preceding target onset. Similarly, in the studies investigating the auditory modality, the effect was observed in the silent interval between the sentence frame and the target. However, the functional interpretation of these results is much less consistent.

In the predictive framework proposed by Lewis and Bastiaansen, (2015), oscillations in the beta band reflect the maintenance of representations at the sentence level. In this view, beta desynchronization in response to words that are unexpected given the sentence context, either semantically or grammatically, reflects the change and update of the sentence representation in line with the new bottom-up input. Similarly, beta desynchronization observed before the target has been proposed to reflect the same mechanism of representation updating driven by the top-down pre-activation of upcoming information (Gastaldon et al., 2020; Prystauka and Lewis, 2019). The study by Gastaldon et al. (2020) has provided empirical evidence in support of this interpretation of the effect. In fact, this study found a desynchronization in HC sentences respect to LC sentences in both comprehension and production tasks, as well as faster reaction times for HC sentences in the production task. According to the authors, this pattern of results suggests that in the constraining condition the lexical retrieval of the word to produce was carried out before the onset of the related picture and that the observed alpha and beta desynchronization is the electrophysiological correlate of this process. Critically, because of the correlations found between production and comprehension, the authors propose that this effect reflects an anticipated lexical retrieval, and possibly other stages of production planning, also in the comprehension task.

This interpretation is in line with the more general hypothesis that alpha and beta desynchronization reflects the semantic richness of the information that is encoded or retrieved by the underlying neuronal population (Hanslmayr et al., 2012). In fact, highly constraining contexts favour the formation of strong predictions through retrieval of rich and specific representations from long-term memory, which is associated, in this framework, with a lower alpha and beta power respect to contexts that do not allow such retrieval (neural desynchronization; e.g., Gastaldon et al., 2020). The "information by desynchronization" hypothesis has been proposed as a general principle of neural coding of semantic information, thus including not only linguistic material but all types of information carrying a meaning. Indeed, power modulations in this direction have been observed both in different language tasks, such as word memory formation (Griffiths et al., 2016) and word encoding (Meeuwissen et al., 2011), and in non-linguistic prediction in different modalities (e.g., Bauer et al., 2014; Chang et al., 2018).

These and other studies rise the possibility that alpha and beta desynchronization reflect domain-general processes usually associated with prediction, rather than pre-activation of linguistic information itself. For example, in a study by Bidet-Caulet et al. (2012) alpha desynchronization was observed prior to the occurrence of a target triangle presented after a particular sequence of triangles with different orientations, which predicted its appearance. This effect was interpreted as an indication of attentional preparation to detect the input. Rommers et al. (2017) propose that similar attentional mechanisms might be involved during sentence processing in constraining contexts. In this view, the pre-activation of information is associated with enhanced preparation to receive the input, which in turn is reflected by the electrophysiological effects. Notably, attentional and working memory demands usually correlate with semantic richness, therefore the “information by desynchronization” mechanism would be in line both with this proposal and with the alternative hypothesis that the effects reflect pre-target lexical retrieval.

The studies by Terporten et al. (2019) and Wang et al. (2018) both exploited magnetoencephalography (MEG) and estimated the cortical sources of pre-target alpha and beta desynchronization. Notably, both studies refer to the proposal by Jensen and Mazaheri, (2010) that alpha oscillations are involved in the active allocation of resources through inhibition of task-irrelevant regions, and therefore alpha desynchronization reflect the engagement of task-relevant regions. However, there is a great disagreement in which regions generate the observed effects, leading to different interpretations of the results.

Terporten et al. (2019) localized the sources of the effects in bilateral parietal and dorsolateral prefrontal cortex, whose engagement has been associated to working memory load and processing capacity across literature (e.g., Sauseng et al. 2005; Barbey et al., 2013). Importantly, in this study, sentence constraint was manipulated across three levels (low, medium and high constraint) and the power modulation showed a non-monotonic relation with constraint. The strongest desynchronization was elicited in the medium condition, followed by the high and low conditions. According to the authors, this pattern further suggests that the observed pre-target modulations reflect working memory demands rather than linguistic pre-activation per se. In fact, assuming that constraining contexts favours the pre-activation of multiple lexical candidates, working memory demands would vary according to the number of competing items that must be retrieved and maintained. In particular, medium constraining sentences would pre-activate a higher number of candidates with respect to high constraining sentences, in which only few possible completions are possible, leading to higher working memory demands. In contrast, in low constraining sentences working memory is only minimally engaged, as the high number of possible alternatives to pre-activate makes prediction too demanding to be implemented.

Wang et al. (2018) interpreted the alpha and beta desynchronization as reflecting specifically the pre-activation of lexical representations. They found the sources in areas of the language network such as left inferior frontal cortex (LIFC), left posterior temporal cortex (PTC) and Visual Word Form Area (VWFA). In particular, VWFA has been associated to processing of orthographic information across literature, thus increased engagement of this area in constraining contexts might reflect the pre-activation of the predicted word up to its orthographic representation. Also, because PTC is associated with storage and retrieval of lexical representations at multiple levels, the authors suggest that lexical access in this area might support and direct the pre-activation of VWFA. Regarding LIFC, the authors propose that it might have a role in integrating the lexical retrieval process in the temporal cortex with information from the preceding context. This interpretation is also supported by the finding that alpha band activity in temporal and VWFA region was correlated with gamma activity (60-90 Hz) in the left prefrontal cortex only in constraining contexts, both before and after the target. In fact, this pattern suggests the exchange and integration of information between frontal and temporal areas is enhanced during the entire processing of constraining sentences.

The functional interpretation proposed by Wang et al. (2018) is compatible with the proposal by Gastaldon et al. (2020) that alpha and beta desynchronization reflects linguistic pre-activation, and that this constitutes an internal top-down signal to update the current representation of the sentence. In the study by Gastaldon et al. (2020) alpha and beta power modulations were correlated between comprehension and production primarily in areas of the language network, such as the left temporal and inferior frontal cortex. This further suggests that these effects reflect predictive processes which are specific of the language domain. However, this interpretation does not imply that linguistic prediction involves uniquely language areas. In the same study, source localization analysis highlighted also temporo-parietal regions in the right hemisphere, for which the authors propose different explanations. For example, constraining contexts might induce differences in attentional and working memory demands, as proposed by Terporten et al. (2019), and right areas might be engaged in these general processes. Another possibility is that regions of the right hemisphere actively contribute to the maintenance and updating of internal contextual representations. In particular, the temporo-parietal junction (TPJ) has been proposed to have a special role in updating current representations according to expectations of future events (Geng and Vossel, 2013). This hypothesis has been also supported by a meta-analysis that examined neuroimaging studies of prediction in language, action perception and music domain and identified the TPJ as part of a network involved in prediction in all three domains (Siman-Tov et al., 2019; see also Masina et al., 2022, for a review). Critically, this domain-general prediction network shows a bias for the right hemisphere, supporting the idea that the engagement of right areas in Gastaldon et al.

(2020) reflects predictive processes and not only the associated attentional and working memory demands.

In conclusion, whether alpha and beta desynchronization in constraining contexts reflect domain-general or domain-specific processes related to linguistic prediction is the main issue in the interpretation of the effect. However, studies on this topic are still scarce and future research might contribute to the debate by further exploiting source localization techniques on EEG and MEG data.

2.2 Visual enhancement and audiovisual speech integration

Speech processing is aided not only by linguistic information, such as semantic context, but also by cues from other sensory modalities. In particular, visual information provided by the speaker's face, especially their lip movements, has been shown to benefit phoneme recognition, word identification and speech comprehension in multiple tasks (see Peele and Sommers, 2015 for a review). Improvement in performance is also accompanied by shorter latency of auditory ERPs in response to audiovisual with respect to auditory-only stimuli (e.g., Arnal et al., 2009). Additionally, several studies have showed that activity in primary auditory cortex is modulated by audiovisual interactions (e.g., Möttönen et al., 2004). This phenomenon is called "visual enhancement" and seems particularly important when the auditory input is suboptimal, either because of environmental conditions or hearing difficulties (Peele and Sommers, 2015; Beauchamp, 2016). In line with this notion, electrophysiological responses to audiovisual speech in visual cortex is increased when auditory input is noisy with respect to when it is clear, suggesting that visual processing is enhanced especially when acoustic information is not sufficient to extract speech meaning (Beauchamp, 2016).

While it is well established that the combination of auditory and visual information facilitates processing, the mechanisms underlying such facilitation are still not completely understood. Peele and Sommers (2015) propose a multistage model of audiovisual integration, in which visual cues influence auditory perception at early stages and undergo a later stage of integration with auditory information in associative cortical areas. This model is in line with evidence that visual enhancement is subjected to individual differences which cannot be predicted from single measures of auditory perception and lip-reading ability (Grant et al., 1998; Tye-Murray et al., 2007a). In fact, the interaction among multiple integration mechanisms might result in different behavioural outcomes as a function of subjective integration abilities and the perceptual and linguistic context (Peelle and Sommers, 2015). Compatibly with this multistage hypothesis, Karthik et al. (2021) found that the electrophysiological response of the auditory cortex to audiovisual speech differed from that to auditory-only speech in multiple ways, across different time windows, spatial locations and frequency bands. The authors propose that multiple features are embedded in the visual signal, and that these features might influence auditory perception through different neural mechanisms.

A common feature of current models of audiovisual speech integration is that auditory and visual information are combined prior to lexical identification, but lexical information can act in a top-down manner to constrain perceptual processes (Peelle and Sommers, 2015). This is in line with the studies discussed in the previous chapter, showing interactions between predictive processes and speech perception. Peelle and Sommers (2015) illustrate an additional mechanism by which perceptual and lexical processes can interact, which is constraining lexical competition between a target word and its neighbours (i.e., similar words in the lexicon). In competitive models of auditory-only word recognition (e.g., Luce and Pisoni, 1998), phonemes activate both the target and its phonological neighbours, in such a way that words with a greater number of neighbours receive more competition and are more difficult to recognize. Analogous effects have been observed for visual-only speech, with a greater number of visual competitors (i.e., words associated with similar orofacial movements) leading to less accurate word recognition (Mattys et al., 2002; Tye-Murray et al., 2007b). During audiovisual speech processing, both modalities are taken in account to select the appropriate word. This restricts the set of competitors to the words that are both phonological and visual neighbours of the target, leading to easier word recognition (Tye-Murray et al., 2007b; Peele and Sommers, 2015).

2.2.1 Neural mechanisms of audiovisual integration

As previously mentioned, the latency of early auditory ERPs such as the N100 is shorter in response to audiovisual with respect to auditory-only stimuli (e.g., Arnal et al., 2009).

This effect is commonly attributed to the lag between visual speech movements (whose perceptual unit is the viseme) and the auditory input: because visemes have an early onset, they allow comprehenders to predict both when a new phoneme will occur and which type of phoneme it will be (Michon et al., 2020). This is possible because each viseme is associated with a limited number of phonemes, and evidence suggests that such association is activated even if the viseme is not accompanied by any sound. Indeed, early fMRI studies showed that silent lip-reading activates auditory cortices (e.g., Calvert et al., 1997) and more recent electrophysiological studies found that auditory and visual speech induce similar modulations in neural oscillations (e.g., Park et al., 2016). Other studies have supported the existence of specific viseme-phoneme binding in the brain. For example, silent lip-reading of syllables was found to elicit an N400 response as opposed to exposure to the same syllables played backwards. This effect was modulated by viseme salience, in line with the idea that more salient visemes generate stronger cross-modal predictions. In contrast, backward syllables did not elicit any N400 as they were not pronounceable and lacked an auditory counterpart (Michon et al., 2020).

In a study combining MEG and fMRI data, Arnal et al. (2009) investigated the cortical network responsible for the N100 effect and the associated behavioural facilitation. In the MEG experiment, they found that the ERP effect was not modulated by the mismatch between the auditory and visual information, suggesting that phoneme prediction is not driven by integration of the two modalities. Such proposal was supported by the fMRI data, which showed enhanced connectivity between motion-sensitive visual cortex and auditory regions when syllables had higher visual predictability. This suggests that visually predictable syllables facilitate predictions of the associated phoneme through direct communication between auditory and visual regions. In contrast, activity of the Superior Temporal Sulcus (STS) in the fMRI experiment showed an interaction between visual predictability and audiovisual congruency. In line with other studies across literature (e.g., Möttönen et al. 2004; Nath and Beauchamp, 2011; Beauchamp, 2016), this pattern suggests that STS is involved in audiovisual speech perception through a later integration of information from the two modalities, including detection of discrepancies. In line with this proposal, electrophysiological responses related to audiovisual integration have been observed in auditory cortex after 150-200 ms from stimulus onset and later in STS, after 250-600 ms from stimulus onset (Möttönen et al., 2004).

Modulations in functional connectivity constitute a fundamental mechanism by which STS integrate sensory information according to the specific features of the input. In the study by Arnal et al. (2009), functional connections of this region with both motion-sensitive visual area and auditory cortex were enhanced when syllables were less predictable. This suggests that the integrative functions of STS might be more prominent when visemes are less informative and the mechanism of direct cross-modal prediction is reduced. Also, functional connections of STS have been proposed to mediate the weighing of auditory and visual information according to their reliability: connections with associative visual areas increase when auditory input is noisy, and connections with associative auditory areas increase when visual input is blurred (Nath and Beauchamp, 2011). Therefore, STS might drive visual enhancement especially when the auditory input is less informative, in which case visual input can compensate and become more relevant for comprehension.

Overall, these data point to a dual route model in which two distinct neural pathways contribute to the effects of visual enhancement during speech processing (Arnal et al., 2009). One is the direct cortico-cortical pathway from visual to auditory areas, that might be responsible for early (or "online") visual effects on auditory processing (i.e., the early stage of integration in the multistage model). The second pathway influences activity in auditory areas from STS, after this region has integrated information from the two sensory modalities (i.e., the later stage of integration in the multistage model). Therefore, this route provides a feedback signal for the auditory cortex, possibly with different effects depending on the congruency between auditory and visual input. In addition, other regions might also contribute to audiovisual interactions, including parietal and prefrontal cortex, which send feedback signals to auditory cortex similarly to STS (Arnal et al., 2009). Some models of auditory speech perception underline the role of motor regions and articulatory representations of speech, which might contribute to activating phonemic representations. For example, in the study by Michon et al. (2020), when orofacial movements of the comprehenders were experimentally restrained, visual-only syllables did not elicit any N400, suggesting that availability of movement effectors is necessary for cross-modal binding of visemes and phonemes.

Neural oscillations play a fundamental role in modulating both auditory and visual perception. In particular, the phase of the oscillation at a given moment is related to the excitability of the neuronal population that has generated the signal. Consequently, inputs reaching neurons at a certain phase are more likely to elicit a response relative to inputs at other phases. This general mechanism of neuronal functioning is exploited by the brain to efficiently process speech: during perception of connected speech, low frequency (<8 Hz) neural oscillations show phase-locked modulations that track the amplitude envelope of the speech signal, in such a way that they synchronize to the syllabic rhythm (Peelle and Sommers, 2015). Importantly, evidence shows that modulation of oscillatory entrainment to speech is a crucial mechanism underlying visual enhancement effects: the quasi-rhythmic orofacial movements of visual speech increase the precision of entrainment in the auditory cortex when they match auditory information (e.g., Luo et al., 2010; see Peelle and Sommers, 2015 for a review). In particular, visual input might reset the phase of auditory oscillations leading the neurons to an excitatory and more responsive state, a mechanism that has been observed in non-human primates through direct cortical electrophysiological recordings (Lakatos et al., 2007; Kayser et al., 2008).

Arnal et al. (2009) propose that the input to auditory cortex guiding phase resetting might come from both the direct and indirect pathways described in their study. In fact, even if their data did not directly suggest any electrophysiological mechanism, a common mechanism of phase resetting from multiple cortical regions might be a parsimonious explanation combining evidence from multiple approaches. This mechanism might be implemented, to a certain extent, also in presence of single audiovisual syllables, which are largely used in studies on audiovisual integration. In line with this notion, Arnal et al. (2009) interpret their results on single syllables as effects of syllable-dependent phase resetting. Also, similar neural regions involved in audiovisual processing have been found for both syllables and connected rhythmic speech, which is in line with the proposal that effects of cortical entrainment are not limited to phoneme perception, but likely extend to multiple timescales and levels of linguistic processing (Peelle and Sommers, 2015).

In conclusion, studies examining the neural correlates of audiovisual speech integration report results which are coherent with a multistage model, in which visual information can contribute to speech processing with several mechanisms, both at the level of functional connectivity and at the level of cortical oscillatory activity.

Chapter 3. Language and prediction in Cochlear Implant users

3.1 Language perception and audiovisual integration in Cochlear Implant users

Cochlear Implants (CIs) are neural prostheses that are commonly used to treat severe and profound deafness, allowing sound perception even in presence of damages or congenital malfunctioning of the cochlea (Macherey and Carlyon, 2014). These devices process and encode sounds, transmitting auditory information to the brain through the electrical stimulation of the intact auditory nerve from an array of electrodes implanted in the cochlea. In this way, CIs support speech perception in deaf people and, importantly, allow the development of spoken language skills in deaf children, especially if they are implanted before three years of age (Hunter and Pisoni, 2021). However, even if in many cases CIs provide sufficient information to comprehend speech, the auditory input conveyed is qualitatively degraded with respect to the one conveyed by an healthy cochlea. For example, in a normal hearing ear different frequencies of sounds are encoded by stimulating different portions of the cochlea. Such spatial coding has a poorer resolution in case of stimulation of the auditory nerve through CI, which results in a weaker differentiation of the frequencies that characterize vowel identity (Hunter and Pisoni, 2021). Similarly, many other features of speech are processed with poor precision by CIs, which often generates difficulties in speech processing for CI users (for a more in-depth discussion of the acoustic features of the CI input, see Hunter and Pisoni, 2021).

According to time of insurgence, deafness is commonly distinguished between prelingual (or preverbal) and postlingual (or postverbal). Prelingual deafness arises before the acquisition of language (i.e., before about three years of age), while postlingual deafness after it. As a consequence, cochlear implantation in the two cases might lead to different linguistic outcomes, which are due to different neuroplastic changes according to age and subjective experience with language. In fact, effective spoken language development requires exposure to speech during a critical period, in which the brain is particularly prone to experience-induced plasticity (see Werker and Hensch, 2015 for a review). This explains why in case of prelingual deafness linguistic outcomes are largely predicted by age at implantation, as children implanted earlier can receive auditory input when brain plasticity is still at its maximum. In contrast, prelingually deaf people implanted in adulthood unlikely develop normal linguistic functions, and though they perceive speech, they are hardly able to comprehend it (Werker and Hensch, 2015).

In case of postlingual deafness, shorter duration of deafness has been associated with better linguistic outcomes after implantation and with greater neural activity in auditory areas in response to speech (Green et al., 2005). This relation is due to neuroplastic changes that occur in periods of auditory deprivation, which might be more or less reversible when perception is restored through CI. Importantly, neural plasticity involves multiple aspects of speech processing. One of them is audiovisual integration, which is subjected to adaptive changes after cochlear implantation in order to compensate for the impoverished auditory input conveyed by the CI.

3.1.1 Audiovisual speech integration in CI users

Audiovisual speech processing is usually associated to behavioural benefits with respect to auditory-only processing at multiple levels (i.e., in syllable perception, word recognition, and sentence comprehension; Peelle and Sommers, 2015). However, studies examining possible group differences in audiovisual benefits between CI users and normal hearing (NH) controls have obtained a variety of findings, according to both demographic features of the CI group and the type of stimuli used. For this reason, in their review of the issue Stevenson et al. (2017) explicitly distinguished between studies using phonemes and studies using words and sentences. Regarding phoneme perception, when auditory and visual information is congruent, both adults and children with CIs show similar audiovisual benefits relative to their NH peers. In contrast, in presence of a mismatch between the two modalities, CI users typically show reduced audiovisual integration and give more relevance to the visual modality (Stevenson et al., 2017). This is in line with the notion that visual information is more important when the auditory input is less reliable (Beauchamp, 2016), a phenomenon which might be further enhanced when auditory information is both incongruent due to experimental manipulations and degraded as conveyed by the CI (Stevenson et al., 2017).

Interestingly, a different pattern of results has been observed by using word and sentence comprehension tasks. In this case, studies on both postlingually and prelingually deaf adults implanted at various ages have shown consistently greater audiovisual benefits for CI users respect to NH controls (Stevenson et al., 2017). This difference in stimulus type relative to the studies previously mentioned suggests a crucial influence of linguistic information on integration processes, which is in line with the multistage model by Peelle and Sommers, (2015). Interestingly, Strelnikov et al. (2009) propose that CI users' lip-reading abilities might be enhanced when it is possible to implement predictive strategies on the basis of linguistic cues, which are provided by words and sentences but not single phonemes. As linguistic and semantic knowledge can improve perception in several ways, these top-down processes might as well facilitate audiovisual integration, especially in CI users. However, this hypothesis has not yet been explicitly tested, and interactions between prediction and audiovisual speech processing remain poorly understood.

Enhanced processing of visual speech in CI users seems to be the result of long-term neuroplastic processes, rather than a transient compensation for the auditory input they receive. Indeed, CI users have shown a higher audiovisual benefit also with respect to NH controls exposed to simulations of the CI input (Rouger et al., 2007). This suggests that CI users adapt permanently to their input in a different way than people without experience with CI would do. Critically, once auditory and audiovisual abilities have become stable after implantation, CI users often show superadditive audiovisual benefits (i.e., their audiovisual performance is greater than the sum of the performance in the two single modalities), suggesting that during their experience with CI they develop better integration abilities rather than better lip-reading alone (Stevenson et al., 2017; Rouger, 2007).

In fact, evidence suggests that cooperation between auditory and sensory modalities is stronger in CI users. A stronger association between phonemes and visual speech cues (e.g., the visemes) has been proposed to be crucial to match the inputs provided by the CI to the corresponding phonological representations (Anderson et al., 2017). In postlingually deaf CI users, this matching is similar to a perceptual learning process in which the existing phonological representations are "re-tuned" to the new distorted input (Davis and Johnsrude, 2007). In this context, lip-reading might contribute both to maintain such representations during the period of deafness and to re-activate and re-map them after implantation (Anderson et al., 2017). Lip-reading has been shown to benefit also prelingually deaf paediatric CI users, who need to build phonological representation solely on the basis of the CI input. Enhanced processing of visual cues has been proposed to benefit this process as well as the development of general linguistic skills, by aiding the access to the general structure of spoken language (Anderson et al., 2017).

Cross-modal binding of auditory and visual speech information is mediated by a direct cortico-cortical pathway which connects the two sensory areas. In CI users, such functional connections have been shown to be enhanced with respect to NH controls, as suggested by bidirectional cross-modal activations: CI users show greater activity of visual areas in response to auditory speech (e.g., Giraud et al., 2001) and greater activity of auditory areas in response to visual speech (e.g., Rouger et al., 2012). These enhancements might be mediated by the functional features of the Superior Temporal Cortex (STC), including the primary and associative auditory areas. In fact, this region shows a multimodal propensity which allows for cross-modal plasticity shortly after deafness onset and implantation, through the engagement of latent multimodal connections (Anderson et al., 2017).

While neuroimaging data about direct audio-visual functional connections are generally consistent across studies, data about group differences for multimodal regions such as STS are more variable. For example, a study by Song et al. (2015) found reduced recruitment of posterior STS in response to congruent audiovisual speech in CI users with respect to NH controls. This supports the hypothesis that this area might be less important relative to primary auditory and visual areas in CI users (Stevenson et al., 2017). In contrast, a longitudinal study by Rouger et al. (2012) found enhanced response to silent visual speech in posterior STS both a few days and several months after implantation, which suggests the involvement of this region in matching visual and phonological representations of syllables in CI users. Interestingly, in the same study, the response of the Inferior Frontal Cortex (IFC) to visual speech was found to increase with time of experience with CI. This area is prominently involved in language production and speech motor representations (Indefrey, 2011), which might also contribute to audiovisual integration of speech. Thus, the authors interpret the results as indicating a progressive reactivation of the visuo-audio-motor speech network after cochlear implantation, through an increased coupling of phonological and articulatory representations. Importantly, similar results have been observed for STS and surrounding areas of the temporal cortex in another longitudinal study by the same research group. In this case the authors examined audiovisual rather than visual-only speech and found that the response of STS gradually increased with CI experience (Strelnikov et al., 2015). Overall, this pattern of results suggests that different areas and functional connections might require different amount of time and experience with CI to reorganize and become efficient. However, the existing data are still scarce and inconsistent, and more studies are needed to further explore how the language network adapt to the new auditory input after implantation.

3.2 Top-down compensation and prediction in Cochlear Implant users

Language perception and comprehension rely on a combination of bottom-up and top-down processes. Thus, it is not surprising that individual differences in linguistic outcomes after cochlear implantation cannot be entirely attributed to differences in transmission of the auditory input. In this regard, Hunter and Pisoni, (2021) underline the importance of studying the top-down processes involved in language comprehension in CI users, in order to better understand the existing individual differences and improve rehabilitation. Başkent et al. (2016) specifically discussed some of the mechanisms that might work as cognitive compensation in presence of hearing impairments. One is phonemic restoration, which allows to perceive interrupted speech as continuous even though some phonemic information is missing. Bhargava et al. (2014) tested this effect in CI users using a variety of acoustic conditions and found that, at the group level, there was no restoration for CI users in most of the conditions where it was present for NH controls. This result suggests that the features of the auditory input, including the nature and the amount of the speech cues that remain available after CI processing, might limit the capacity to use this cognitive mechanism. However, this might not be true for all CI users. Additional analyses showed that CI users with the highest intelligibility scores at baseline showed the restoration benefit, suggesting that there is a significant variability in the use of this mechanism of compensation (Bhargava et al., 2014). Although the factors underlying such variability are not clear, Başkent et al. (2016) suggest that both cognitive components and individual differences in CI functioning might have a role.

According to the same authors, effective use of the context is another possible mechanism of top-down compensation. Semantic processing of the context can influence perception, including the presence of phonemic restoration and the comprehension of degraded speech for NH listeners. However, for CI users data on use of sentence context are greatly inconsistent across studies. As for phonemic restoration, the low quality of auditory input and the limited amount of speech cues available might compromise the capacity to represent context and generate predictions about upcoming words (Başkent et al., 2016). Also, some authors underline that prelingually deaf children who use CIs often show slower and more effortful speech processing than their NH peers (e.g., Pisoni et al., 2011). This might result in lack of time and cognitive resources sufficient to use semantic cues and predictive processes (Holt et al., 2021). Also in this case, the great variability among CI users makes it difficult to identify trends at the group level. Nevertheless, many studies have explored possible factors influencing the exploitation of semantic context, both at the input quality level and at the cognitive level.

Smiljanic and Sladen (2013) found that neither children with CI nor with NH could use sentence context when speech signal was masked by noise and provided with a less clear communicative style, suggesting that adverse features of the input can impact the use of semantic cues. Importantly, when speech was noisy but provided in a clear, listener-oriented style, children with CI benefited from the context less than NH controls, in line with previous studies showing that CI users are more vulnerable to background noise during speech processing (e.g., Spahr et al., 2007). Thus, smaller benefit from the context can be partly explained by the enhanced difficulty in identifying speech in noise, which is especially relevant in everyday situations (Smiljanic and Sladen, 2013).

Availability of cognitive resources might have a role in determining the efficiency of comprehension in CI users. For example, Smiljanic and Sladen (2013) propose that the perceptual effort required to identify speech in noise might exhaust the resources to incrementally represent the semantic context of the sentence. In the same vein, Başkent et al. (2016) suggest that CI users might be able to use sentence context, but likely with greater effort and expense of cognitive resources respect to NH listeners. This hypothesis is in line with the notion that individual characteristics in cognitive functioning have a crucial role in determining linguistic outcomes, as people with better cognitive capacities would have more resources to use compensatory mechanisms (Hunter and Pisoni, 2021). Several studies have supported this view. For example, Moberly and Reed (2019) found a positive relation between inhibitory control and accurate recognition of meaningful sentences in adult CI users. This result highlights the importance of inhibiting lexical competitors during comprehension, and it is in line with the view that effective lexical competition underlies the ability to form predictions (Winn, 2016; Nagels et al., 2020). In the same vein, Dingemanse and Goedegebure (2019) found that verbal working memory correlated with use of context for sentence recognition in CI users. Importantly, this study found that CI users as a group relied more on contextual constraints than NH controls, both during recognition of single words (i.e. phonotactic constraints determining the possible subsequent phonemes) and sentences (i.e. syntactic and semantic constraints). According to these authors, CI users spontaneously learn to recognize words on the basis of scarce information, through practice with the CI input.

A great source of variability that might partly explain inconsistencies across studies is the age of participants along with the type of deafness. In fact, studies with adults typically have postlingually deaf participants, who have acquired language with a normal auditory input and have later adapted to the CI input after hearing loss. This adaptation might be qualitatively different from the one of prelingually deaf children, who have experienced the CI input from the beginning of life. Such an input might cause difficulties and delays in language acquisition and might also affect the development of cognitive abilities (Hunter and Pisoni, 2021). Furthermore, the factors and processes underlying the relation between the deployment of compensatory strategies and individual linguistic abilities are not yet clear. The study of prediction might shed light on this issue, by exploring a possible mechanism by which sentence context can be exploited by CI users.

3.2.1 Prediction in CI users

The degraded speech input conveyed by the CI causes an increased listening effort which might contribute to difficulties in the language domain. According to Winn (2016), such increase in effort and expense of resources might be related to models of lexical competition (e.g., Luce and Pisoni, 1998). When the auditory input is unreliable, the lexical representation corresponding to the correct word is less strongly activated, while incorrect representations such as its phonological neighbours are improperly activated. These processes can result in a greater expense of resources to maintain active several lexical options (Winn, 2016). In contrast, predictive processes restrain the possible lexical candidates even before encountering the input, thus reducing the expense of resources required for lexical competition. For this reason, Winn (2016) measured predictive processes in CI users and NH controls through an online measure of arousal and cognitive effort, which is pupil dilatation. In particular, a reduction in pupil dilatation response has been associated across literature to a reduction in cognitive demands (see Laeng et al., 2012 for a review). In the study by Winn (2016), such reduction was interpreted as an indication of predictive processes during processing of high with respect to low constraining sentences. Indeed, a difference between the two constraining conditions was found for both groups of participants, but with different timing: pupillary reduction appeared before the offset of the sentence for NH listeners, and after the offset for CI users. A possible explanation is that sentence context is not used predictively by CI users, but rather to facilitate a post-hoc restoration of sentences, as highly constraining contexts allow to infer words that have been misperceived during online processing (Winn, 2016).

Different explanations might underlie lack of prediction for CI users. For example, slower speech processing might hinder the ability of CI users to rapidly exploit context to predict in an effective manner (Pisoni et al., 2011; Holt et al., 2021). In contrast, a period of silence after a sentence might leave time enough to use context to revise its interpretation (Winn, 2016). An alternative hypothesis is that CI users might modulate their use of predictive processes in order to spare cognitive resources. In fact, adolescents with CI have been shown to use a wait-and-see strategy for word recognition (McMurray et al., 2017). Blomquist et al. (2021) propose that a similar strategy might be enacted at the sentence level to avoid the costs of revising incorrect predictions, as errors are more likely to occur when the input is less clear. To test the presence of predictive strategies in children with and without CI, these authors measured the timing of lexical access to the target word and competition from its phonological neighbours through eye-tracking in a visual world paradigm. In particular, constraining contexts were expected to elicit faster looks to the target and reduced looks to the phonological competitor respect to neutral contexts. This pattern was indeed observed even before the onset of the target for both CI users and NH children, suggesting that they were both able to predict. However, comparisons at the group level revealed a smaller effect of prediction in children with CI, who showed a smaller proportion of looks to the target and a greater proportion of looks to the competitor with respect to controls in the constraining condition. While this decreased efficiency might be attributed to difficulties and delays in speech processing, Blomquist et al. (2021) highlight that it might also be an adaptive strategy, in order to predict while at the same time reducing the commitment to the predictions and the revision costs for inaccurate predictions.

The same authors discuss also other variables that might influence prediction in children with CI. In particular, they found that a measure of individual vocabulary partly explained the variability in predictive processes among CI users, suggesting that better linguistic knowledge facilitates prediction in this group. However, this variable was not the only factor underlying differences at the group level, as there were significant differences in predictive abilities also between children with CI and NH children with similar vocabulary. Another relevant factor is the quality of phonological representations built on the basis of the CI input, which have been shown to be less structured and distinct from each other (e.g., Bouton et al., 2012). According to Blomquist et al. (2021), this might make access to lexical information more difficult, with cascading effects on sentence-level predictive abilities. Similarly, experience with the CI input during language development has been associated with a less developed semantic structure (Kennett et al., 2013), which might reduce the possibility to activate semantically associated words and use this information to predict (Blomquist et al., 2021).

It is also important to consider that all these factors are highly variable across individuals. Many children with CI are able to acquire normal language skills and build efficient linguistic representations, especially in the last few years, as quality of devices continues to improve and age of implantation has generally decreased (Holt et al., 2021). This can explain why in a similar visual world paradigm as in Blomquist et al. (2021), Holt et al. (2021) did not find any group differences between NH children and CI users, who showed similar timing of lexical access and speech processing in both constraining conditions. This suggests that it is possible for children who have experienced CI input early in life to develop well-structured phonological, lexical and semantic representations, which in turn allow to develop predictive abilities similar to their NH peers.

Difficulties in lexical competition and subsequent lexical access, which Nagels et al. (2020) call "lexical uncertainty", might be compensated by exploiting semantic context to predictively restrain the possible candidates even if the auditory input is unclear. This would be in line with the finding that postlingually deaf CI users rely more than NH peers on contextual information (Dingemanse and Goedegebure, 2019), and would point to lexical uncertainty as a possible variable influencing the strategic use of predictive processes. Nagels et al. (2020) tested this hypothesis in postlingually deaf adults with CI, exploiting a visual world paradigm in which both phonological and semantic competitors of the target were presented to participants. Predictive strategies were supposed to change the pattern of lexical competition not only by reducing the activation of phonological competitors, but also by increasing the activation of semantic competitors, as they were predictable completions for the sentence as well as the target. This pattern was indeed found for both the CI and the NH groups, but the differences between the two constraining conditions were smaller and occurred later for CI users, in line with the findings by Winn, (2016) and Blomquist et al. (2021).

Importantly, Nagels et al. (2020) found that participants with higher lexical uncertainty showed more pronounced differences between constraining conditions, which is in line with the hypothesis that reliance on context might work as a compensatory strategy for these participants. Higher lexical uncertainty was also associated with longer time course of lexical competition, suggesting that the delay found at the group level was primarily driven by this subgroup of CI users. Such delay might be explained by more difficult lexical competition, which might not be entirely compensated by the use of semantic context. Alternatively, it might suggest that these participants wait to receive more information to later revise their interpretation of the sentence, as suggested by Winn, (2016). Similarly, it might be possible that delaying and reducing commitment to predictions constitute an adaptive strategy to avoid the costs of errors, which are more frequent when uncertainty is higher, as suggested by Blomquist et al. (2021).

In conclusion, the studies discussed here show that several factors might underlie individual variability among CI users in prediction, including an adaptive modulation of the strategies routinely used in sentence comprehension. In fact, utility of prediction depends on a variety of factors which might vary both across individuals (e.g., linguistic abilities, processing speed) and across situations (Kuperberg and Jaeger, 2016). This makes the study of the causes underlying particular patterns of predictive behaviour very complex, especially among an heterogeneous clinical population such as CI users. Nevertheless, this line of research is still at its beginning, and studies in this direction will benefit from new approaches and experimental manipulations, such as in the research project here presented.

Chapter 4. Studying prediction and audiovisual speech processing in CI users: a research project

4.1 Introduction

Informative sentence context and visibility of mouth movements are important aid to speech comprehension in everyday life, but mechanisms of interaction between them are poorly understood, even for NH listeners. Some authors argue that such interactions might be particularly relevant for CI users, as they show greater visual enhancement effects than NH listeners for words and sentences, but not for single phonemes (Strelnikov et al., 2009). This suggests that lip reading might benefit from implementation of predictive strategies, and this combination might further facilitate speech perception. To our knowledge, the study here outlined is the first that examines the interaction between these two factors in people with CI. In particular, the aim of our project is to investigate the neural correlates of prediction, and how they could be affected by the presence of mouth cues, in both CI users and NH listeners. A within-group comparison across experimental conditions will contribute to elucidating the interactions between predictive and audiovisual integration processes. The comparison between the two groups, on the other hand, will shed light on the functional neuroplastic changes that occur after years of experience with the CI input.

Great variability within the CI population has hindered the emergence of a clear trend indicating reduced, similar or increased use of sentence context with respect to NH listeners. Linguistic abilities likely explain part of this variability, as more proficient comprehenders have wider vocabulary knowledge, access to this knowledge faster and use grammatical rules more efficiently. All these abilities are crucial to formulate predictions which are both accurate and fast enough to result useful. For this reason, we chose four tasks to measure general linguistic abilities both in comprehension and in production. We measure speed and efficiency of lexical access in production with a verbal fluency task (semantic and phonological). In comprehension, we test lexical access with a lexical decision task, which also provides a measure of the width of vocabulary knowledge. Then, we test grammatical abilities in production and in comprehension. For the first aspect, participants have to generate sentences which are evaluated by semantic appropriateness, syntactic complexity and grammatical correctness; all these aspects of sentence generation might indeed be relevant for predictive abilities. For the second aspect, we use a Sentence Picture Matching (SPM) task, which tests comprehension of sentences with different levels of grammatical complexity. In particular, the stimuli of the SPM test the recognition of the grammatical function of prepositions and pronouns, especially when complements are implicit, and of the passive verbal mode, which have been shown to be particularly difficult to learn for children with CI (Artesini, 2019).

For the EEG experiment, we implemented a within-subject design in which each participant is exposed to each condition for constraint (high-constraint HC vs low-constraint LC) and visibility of mouth movements (presence v+ vs absence v-) in sentence frames. The paradigm consists in an audiovisual sentence comprehension task, in which participants are presented with the sentence frame (i.e., the video of the speaker uttering a sentence without the final word), followed by a silent pause of 800 ms, and then the target stimulus completing the sentence (i.e., the video of the speaker uttering the final word). Time-frequency analysis will focus on the silent interval between the sentence frame and the target, where prediction is argued to take place, and ERP analysis will examine the response to the target in the range of the N400 component (300-500 ms). Following the literature, the main effects we expect to observe is alpha and beta desynchronization and reduced N400 component in the HC condition with respect to the LC condition. Interactions between constraint and mouth cues are still largely unknown; however, two hypotheses might be formulated. On one side, the higher informativity provided by the integration of visual and auditory speech information might allow for more accurate predictions, which might be reflected in stronger effects (i.e., power decreases and N400) in v+ relative to v-. On the other side, less informative signals might induce increased predictive effort to compensate the lack of information, which would be reflected in weaker effects in v+ relative to v-. In addition, the qualitatively different nature of the two information streams (visual and auditory) may be reflected in spatial and temporal differences of power modulations, rather than in intensity. With regard to the between-group comparison, we expect CI participants who have good linguistic abilities to show more pronounced prediction-related effects with respect to NH participants, as they might use sentence context adaptively to aid their speech comprehension. Nevertheless, we expect to find large individual differences within the CI group due to different linguistic abilities, strategies adopted in everyday life, age and experience with CI use (i.e., different timing of insurgence of deafness).

Data collection and EEG analyses of this research project are currently in progress. This chapter will present the methods we implement in the project and the behavioural results relative to the participants for the first period of data collection. Even though these data are limited, they provide a preliminary picture of the characteristics of the sample we expect to recruit for the totality of the study. Also, these preliminary results on the linguistic tests and some considerations risen during this period of data collection will be the starting point to discuss some relevant aspects of the complex interaction among sound perception, language comprehension and CI use.

4.2 Materials and methods

4.2.1 Participants

So far, 11 participants with CI (4 females; mean age = 19.3, sd = 10.9) were recruited through direct contact from the Otorhinolaryngology Unit at the hospital of Padova. Participation was voluntary and a monetary refund was offered. All participants had an age between 12 and 23 except for one who was 49 years-old. All of them were native speakers of Italian and right-handed, except for one who was left-handed (handedness evaluated by means of an Italian translation of the Edinburgh Handedness Questionnaire, Oldfield, 1971; mean laterality index = 72.63, sd = 47.82). None of them reported a history of neurological or psychiatric disorders. All participants (or their parents, if underage) signed an informed consent to participate in the experiment. The study was approved by the Ethical Committee for the Psychological Research of the University of Padova (prot. 4188).

All participants were experienced CI users (mean years of experience with CI = 12.7, sd = 5.9) but they were implanted at different ages according to time of insurgence of deafness. 7 participants were deaf since birth and implanted within 3 years of age. Among them, 4 had bilateral CIs, 2 had monolateral CI and 1 had monolateral CI with a hearing aid on the controlateral ear (configuration indicated as "bimodal" in Table 1). 2 participants had progressive deafness since birth, used hearing aids from early childhood with sufficient compensation and were implanted later during childhood after acquisition of language. Both of them had monolateral CI with a hearing aid on the controlateral ear. 2 participants were postlingually deaf and lost hearing function after full acquisition of language. Both of them had duration of deafness shorter than 3 months and had bilateral CIs.

All participants had good functioning of the CI, as indicated by a Pure Tone Average (PTA) below 28 decibel (db) in every case (mean = 22.09, sd = 3.64). The PTA is an audiological measure which consists in the mean of hearing thresholds at 500, 1000, 2000 and 4000 Hz. It is a general measure of hearing functioning and it is considered to be normal within 25 db. For the purpose of our study, we considered the lower PTA between the two ears, since our aim was to ensure a good perception of stimuli irrespectively of laterality. In our sample, only 2 participants had a PTA higher than 25 db, with the highest value being 27.5 db. Table 1 summarizes relevant characteristics of the participants.

Age (years)	Laterality of CI	PTA in db (best performing ear)	Insurgence of deafness	Age at first CI activation	Duration of CI use (years)
14	Bimodal	22.5	Prelingual, not progressive	3	11
23	Monolateral	21.25	Prelingual, not progressive	2	21
14	Bilateral	21	Prelingual, not progressive	2	12
21	Monolateral	20	Prelingual, not progressive	2	19
12	Bimodal	27.5	Prelingual, progressive	8	4
12	Bimodal	27	Prelingual, progressive	5	7
12	Bilateral	17.5	Prelingual, not progressive	1	11
23	Bilateral	17.5	Postlingual, not progressive	17	6
12	Bilateral	23.75	Prelingual, not progressive	2	11
49	Bilateral	26.25	Postlingual, not progressive	29	20
20	Bilateral	18.75	Prelingual, not progressive	2	18

Table 1. Characteristics of participants for this part of the study

The total sample of the study will be heterogeneous both for age and aetiology of deafness. In particular, we plan to recruit 40 participants with an age between 12 and 50, and we will consider both prelingually and postlingually deaf people, examining possible differences between them.

4.2.2 Stimuli

We selected 196 sentence frames and 98 target nouns from the stimuli used in the study by Gastaldon et al. (2020). In that study, two sentences frames were constructed for each of 128 target nouns, for a total of 256 sentence frames. Among the two possible frames, one favoured the prediction of the target noun with a high probability (high constraint, HC), and one did not allow to predict a particular word but was constructed in such a way that the target word was still a plausible completion (low constraint, LC). Constraint was modelled as cloze probability (CP) of the target word given the frame, assessed with an online sentence completion questionnaire. However, in the study by Gastaldon et al. (2020), participants were all adults, unlike in the study here presented, which also includes participant from 12 to 17 years of age. Therefore, some of the sentences that were devised in that study might not be as effective for younger participants. In other words, the sentential context that allowed to predict a given word in adults may not be as constraining for younger participants, given that the contexts we encounter and the language we use vary with age. For this reason, we administered the same online questionnaire used in Gastaldon et al. (2020) to assess CP to an independent set of 10 responders whose age was more comparable to the age of the participants we expected to recruit more (mean age = 14.2, sd = 2.5). On the basis of the results of the questionnaire, we selected the stimuli which best suited the age of our sample (i.e., we excluded those that did not elicit the completion we expected), resulting in 196 sentence frames, half HC and half LC, and 98 corresponding target nouns.

A female native speaker was recorded while pronouncing the frames and the targets in a quiet room by means of a videocamera. Frames and targets were recorded separately. The speaker was instructed to keep the reading pace and the voice tone as steady as possible and to avoid any movement of the head or the eyes, which pointed towards the videocamera. Audiovisual recordings were then appropriately centred with respect to the centre of the screen and trimmed at the end and at the beginning using Kdenlive, a software for video editing. Trimming was performed in such a way that mouth movements associated with speech onset and offset were not trimmed out at the beginning and at the end of each recording. By means of the same software, from each sentence frame it was created a corresponding version with a grey rectangle superimposed on the video, in such a way that mouth and lip movements were not directly visible. To subsequently determine which frames would be shown to each participant with mouth fully visible (v+ condition) or not visible (v- condition), target words and their associated sentence frame were divided into two lists, A and B. Each list contained 49 target words and the associated 98 sentence frames (49 HC and 49 LC). The two lists were matched both between conditions and within conditions for the following parameters: lexical frequency (log-scaled; obtained from COLFIS, Bertinetto et al., 2005), number of phonemes and syllables (obtained from PhonIta 1.10, Goslin et al., 2014),. Each list contained a comparable level of constraint, as difference in CP was not significant between lists but it was significant between HC and LC sentences, both in the whole set and within each list. The association between lists and v+/- was counterbalanced across participants (i.e., participant 1 was presented with List A in v+ and List B in v-, participant 2 with the opposite and so on). In this way, each sentence frame is presented equally in the v+ and the v- conditions across the entire experiment.

Finally, in regard to the stimuli used in the comprehension tests, for the lexical decision task we used 60 words and 30 pseudowords from LexITA (Amenta et al., 2021). For the Sentence Picture Matching task (SPM) we selected a subgroup of stimuli from the work by Artesini (2019), and in particular: 8 active transitive sentences, 8 active dative sentences with explicit subject, 8 active dative sentences with implicit subject, 8 passive sentences with complement expressed and 8 passive sentences without complement expressed. For the subsequent analysis, these sentences were divided in 3 types: active (transitive), clitic (active dative with explicit or implicit subject, which tested grammatical function of pronouns and prepositions) and passive (with or without complements expressed). Table 2 summarizes the types of sentence stimuli for the SPM task.

Grammatical structure	Type	Example
Active transitive	Active	La nonna sta guardando il gatto *The grandmother is watching the cat*
Active dative with explicit subject	Clitic	Lo dà la mamma *The mother gives it*
Active dative with implicit subject	Clitic	Lo dà al bambino *(He/she) gives it to the child*
Passive with complement expressed	Passive	Il cane viene guardato dal bambino *The dog is watched by the child*
Passive without complement expressed	Passive	Il bambino viene baciato *The child is kissed*

Table 2. Types of stimuli used in Sentence Picture Matching (SPM) task.

4.2.3 Procedure

The linguistic tests were performed before the EEG session and lasted about 25 minutes in total. For the production part, semantic fluency task, phonemic fluency task and sentence generation task were administered according to the guidelines presented in "Batteria per la Valutazione Neuropsicologica per l'adolescenza" (BVN 12-18, Gugliotta, 2009). In the fluency task, participants were asked to utter as many words as possible during one minute of time. For the semantic fluency, the categories were "animals" and "fruit", while for the phonemic fluency the phonemes were /f/ and /l/. Responses were recorded through the software Audacity from a computer placed in front of the participants, and an experimenter ended the recording when the 1-minute trial ended. This resulted in four recordings for each participant, one for each trial (two for semantic fluency, two for phonemic fluency). In the sentence generation task, participants were asked to produce a meaningful sentence containing a pair of words pronounced by the experimenter, without the conjunction "e" (Italian for "and") and with no time pressure. 5 pairs of words were selected from BVN ("filo/bottone"; "fuoco/legna"; "casa/luce";

“sciopero/salario”; “vita/libertà) and an additional pair (“matita/carta”) was used as an example before the test. Responses were recorded in a single recording through Audacity.

The comprehension tasks were administered through the same computer, the lexical decision task through the Psychopy software and the SPM through the OpenSesame software. Participants were asked to read the instructions and to ask for clarification if needed. For the lexical decision task, they were presented with words and pseudowords in written form and had to press two different keys on the keyboard to indicate whether they thought the string was an Italian word or not. Each string was preceded by a fixation cross (500 ms) and remained on the screen until the response. Reaction times (RTs) and accuracy were recorded. For the SPM task participants were shown a written sentence at the centre of the screen for 2 seconds. Subsequently, four pictures appeared on four sections of the screen (upper left, upper right, bottom left and bottom right), which remained until a response was provided. Participants had to select the picture corresponding to the sentence by using the mouse, and RTs and accuracy were recorded.

For the EEG session, participants were seated in a soundproof room in front of a computer. Before starting the experiment, they were instructed to watch the videos and listen carefully to the sentences, while minimizing eye movements and muscular activity of the face and the limbs. At the beginning of each trial, an empty square appeared for 1 second followed by the video of a sentence frame which could be either HC or LC, and either with mouth visible or not visible (v+ or v- conditions respectively). Then, there was a silent interval of 800 ms and the video of the corresponding target was presented after the silent interval (Figure 1). To ensure that participants paid attention to the sentences, in line with preceding literature (e.g., Gastaldon et al., 2020),42 trials (20%) were followed by a question about the preceding sentence, which was presented in written form after the target. Participants had to respond yes or no by pressing a key on the keyboard.

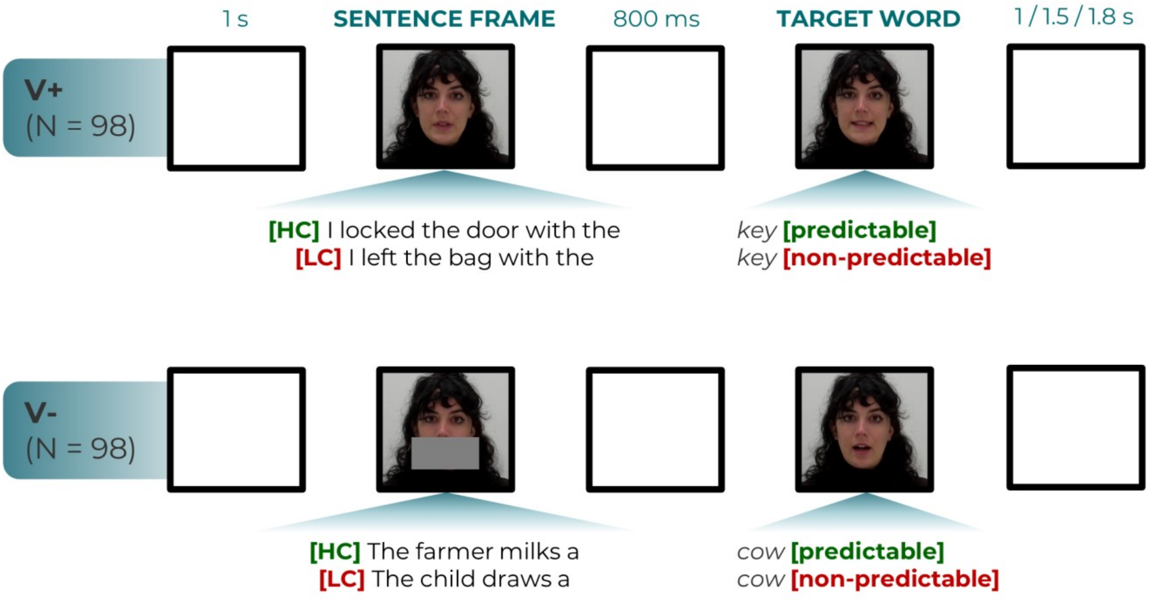


Figure 1. Trial structure in the v+ (top) and v- (bottom) conditions.

For each participant, list A or B was presented in v+ condition and the other list in v- condition, and the lists associated to v+ or v- were counterbalanced across participants. The order of presentation of each trial was pseudorandomized for each participant through Mix (van Casteren and Davis, 2006). In particular, two trials with the same target were separated by at least seven trials, and no more than three consecutive trials belonged to the same condition of constraint or mouth visibility. The inter-trial interval varied from trial to trial (1, 1.5 and 1.8 seconds). After every 28 trials participants could take a short break, which ended when they pressed a key on the keyboard. Before starting the experimental session, 8 practice trials (not included in the analysis) were presented to participants in order to familiarize them with the type of stimuli used. In this phase participants were also asked to adjust the volume of the speakers according to their preference, in order to ensure that the task was performed in optimal perceptual conditions. The EEG session lasted about 35 minutes in total.

4.2.4 EEG data acquisition, preprocessing and analysis

Electroencephalogram was recorded with a system of 64 active Ag/AgCl electrodes (Brain Products), placed according to the 10–20 convention. Reference was placed in the position of the Cz electrode. Of the 64 electrodes, 3 were used to record blinks and saccades (external canthi and below the left eye). For each participant, we took note of the electrodes which were placed near the external part of the CI (magnet and processor) and we avoided to apply the electroconductive gel in such locations, to ensure that the device would not get in touch with it. As a consequence, these electrodes generated a noisy or flat signal and will be later excluded during preprocessing. This resulted in a slightly different configuration of remaining electrodes for each participant. This procedure will be followed for all the participants with CI throughout data collection. Subsequently, for the total sample of the study, pre-processing and analyses will be performed using the MATLAB toolbox Brainstorm (Tadel et al., 2019). In particular, we will perform time-frequency decomposition and ERP analysis to statistically compare the experimental conditions and the CI and NH groups.

4.2.5 Behavioural performance analysis

For the verbal fluency tasks, for each trial we measured the number of correct words produced and the “fulcrum point”, which is the mean of the latencies of responses subsequent to the first, calculated from the beginning of the task (Sandoval et al., 2010). This value represents the time point at which 50% of responses have been produced and reflects the relative distribution of responses across the one-minute period. Thus, fulcrum points near 30 s indicate that the capacity to retrieve and access words is approximately steady across the 1-minute trial, while shorter fulcrum points indicate that participants exhaust their responses quickly and have difficulties in retrieving responses in the second half of the trial. To compute both these measures, we excluded 1) words that had been already said during the trial, 2) variations of the word said immediately before (e.g., cervo/cerbiatto, figlio/figlia) or 3) words composed from immediately preceding words (e.g., pesce/pesce palla). Then, for each participant we averaged the number of correct words and the fulcrum points relative to the trials “animals” and “fruits”, and the trials “f” and “l”, to obtain separate measures of semantic and phonological fluency, respectively. For the sentence generation task, we attributed a score to each sentence produced according to the guidelines of the BVN (Gugliotta, 2009), then we summed such scores to obtain a single value for each participant. In particular, to each sentence was attributed a score ranging from 0 to 3 according to semantic appropriateness and syntactic and grammatical correctness.

For both the comprehension tasks, we considered accuracy scores and the RTs relative to the correct trials and we performed the analysis through the software R. For each task, results were analysed by means of generalized linear mixed-effect models using the lme4 package (Bates et al., 2015), with random intercept for participant and stimulus, and a fixed effect of stimulus type (words vs nonwords, for lexical decision; active vs passive and clitic, for SPM). We used binomial distribution (link function “logit”) for accuracy scores, and gamma distribution (link function “log”) for RTs (Lo & Andrews, 2015). We used the lmerTest package (Kuznetsova et al., 2017) to estimate p values of model predictors. Because in SPM task there were three types of sentences, we also performed post-hoc contrasts by using estimated marginal means (with the emmeans package; Lent, 2021) in order to test for differences between all three types, both in RTs and accuracy.

To evaluate if duration of CI experience was associated with the measures of linguistic abilities, we correlated linguistic scores with the years of experience normalized by the age of each participant (years divided by age). Linguistic scores were also correlated with age and PTA scores. We note that the sample here presented is very small, and a clearer picture will emerge when a bigger sample will be tested. These analyses will be especially insightful when considering prelingually deaf participants, as literature suggests that earlier implantation is associated with better linguistic outcomes (Hunter and Pisoni, 2021). Furthermore, we will consider time of insurgence of deafness as another variable which might modulate both linguistic abilities and the electrophysiological results. As suggested by the studies presented in chapter 3, there might be important differences between prelingually and postlingually deaf people, as life-time auditory experience and neuroplastic changes are different in the two populations. Importantly, we will also consider associations between specific linguistic abilities and the patterns of neural activity associated with prediction. Such associations might contribute to elucidating which factors and mechanisms underlie the ability to use prediction in CI users and NH listeners.

4.3 Behavioural results

4.3.1 Production tasks

Table 3 shows the results of production tasks for each participant. Mean of total correct responses was 15.6 (sd = 3.6) in the semantic fluency task and 9.4 (sd = 3.4) in the phonological fluency task. Fulcrum points were generally slightly shorter than 30 (the point indicating a steady rhythm of production), with a mean of 19.8 (sd = 3.1) for semantic fluency and 24.9 (sd = 5.8) for phonological fluency. This suggests that, in general, participants produced more correct responses when retrieving and producing words belonging to semantic categories than phonological categories (figure 2), which is in line with previous literature (e.g., Schmidt et al., 2017). However, they were more constant in producing responses across the trial in the phonological relative to the semantic task (figure 3). With regard to sentence generation, participants generally obtained good scores, with a mean of 9.8 out of 15 (sd = 3.2). Future statistical analyses on a larger sample will be needed to confirm these trends, also through the comparison with NH peers.

Participant	Total words in semantic fluency	Fulcrum point in semantic fluency	Total words in phonological fluency	Fulcrum point in phonological fluency	Sentence generation score
1	13	16.02	12.5	20.51	8
2	22.5	23.23	8	24.86	13
3	13	14.44	13.5	31.31	11
4	10.5	24.89	3.5	26.18	2
5	14.5	18.27	12.5	23.38	9
6	13	17.29	5.5	25.03	11
7	13	21.33	5.5	26.02	8
8	18.5	21.23	11	24.07	11
9	17	20.05	10	13.07	9
10	17.5	19.29	12.5	23.86	14
11	19	21.78	9	36.28	12

Table 3. Individual measures for the production tasks.

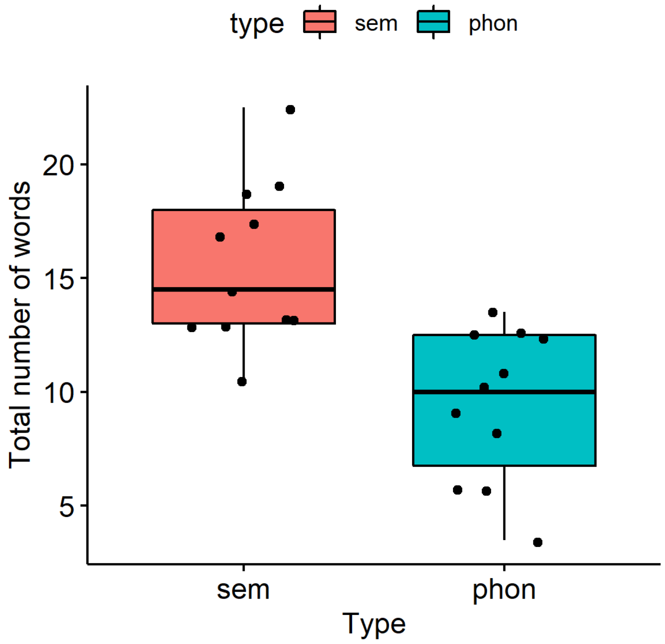


Figure 2. Number of correct responses in the semantic and phonological fluency tasks.

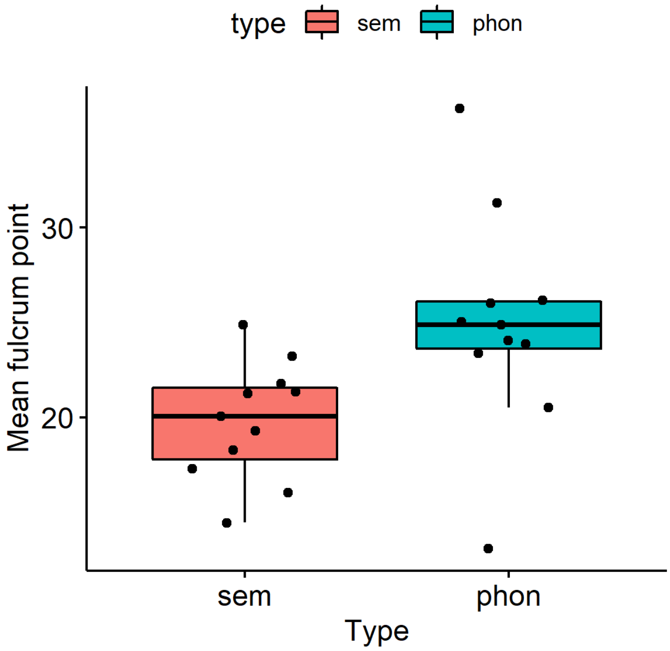


Figure 3. Fulcrum points in the semantic and phonological fluency tasks.

4.3.2 Comprehension tasks

In the lexical decision task, we found an effect of lexical status in both accuracy and RTs: words were more accurate and faster to recognize than nonwords, which is line with previous literature (e.g., Forster and Chamber, 1973). Parameters of the models for accuracy and RTs are reported in tables 4 and 5, respectively. Figure 4 shows RTs for each participant. With the only exception of participant 4, all participants showed fast RTs, with larger variability across the group for nonwords relative to words. RTs and accuracy at the group level are shown in figure 5 and 6, respectively. Also in accuracy, nonwords show much greater variability than words, for which most participants were highly accurate.

Parameter	Estimate	Standard error	z value	p value
Nonwords (intercept)	2.4490	0.5982	4.094	< 0.001
Words vs Nonwords	1.6294	0.4083	3.991	< 0.001

Table 4. Parameters of the generalized linear mixed-effects model for accuracy in lexical decision.

Parameter	Estimate	Standard error	t value	p value
Nonwords (intercept)	0.8007	0.1826	4.384	< 0.001
Words vs Nonwords	-0.5715	0.0673	-8.492	< 0.001

Table 5. Parameters of the generalized linear mixed-effects model for RTs in lexical decision.

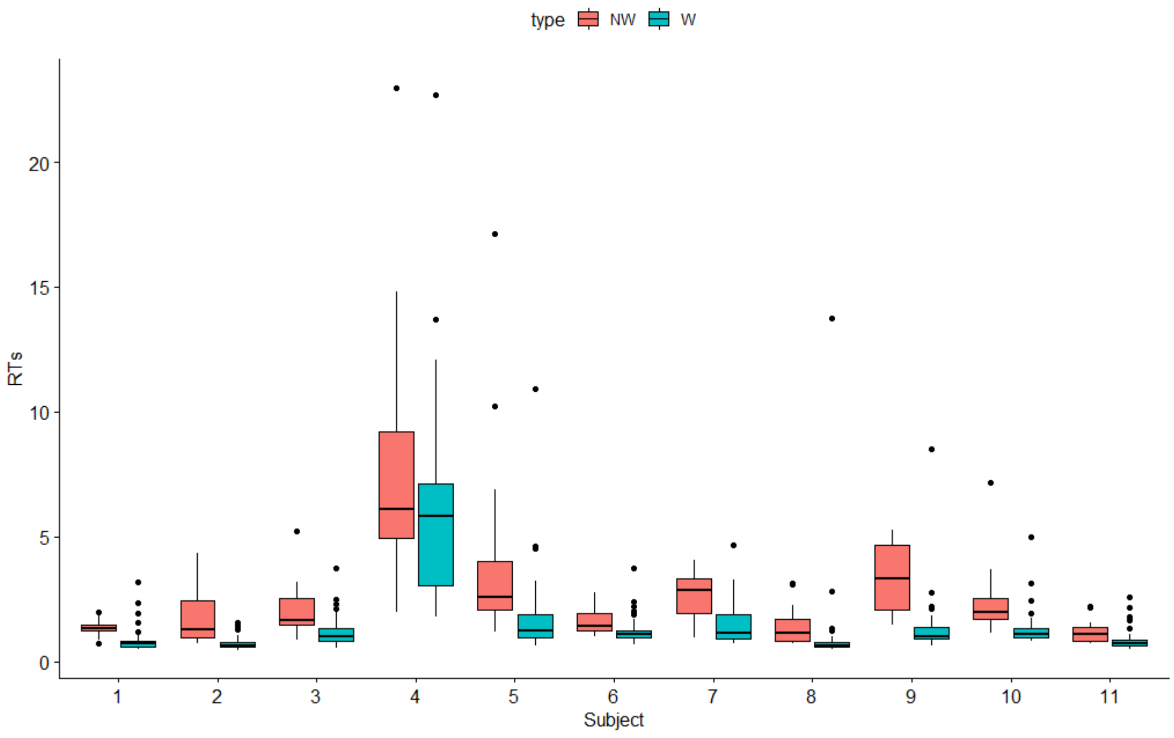


Figure 4. Individual RTs in the lexical decision task (values on the y-axis provided in seconds).

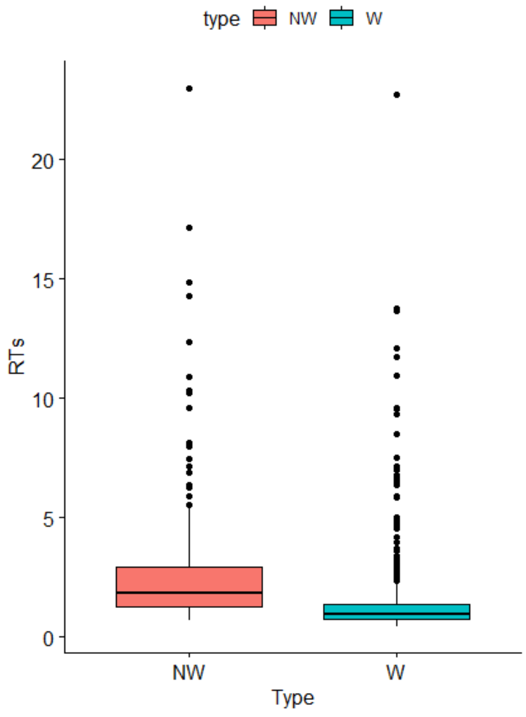


Figure 5. RTs in the lexical decision task (in seconds).

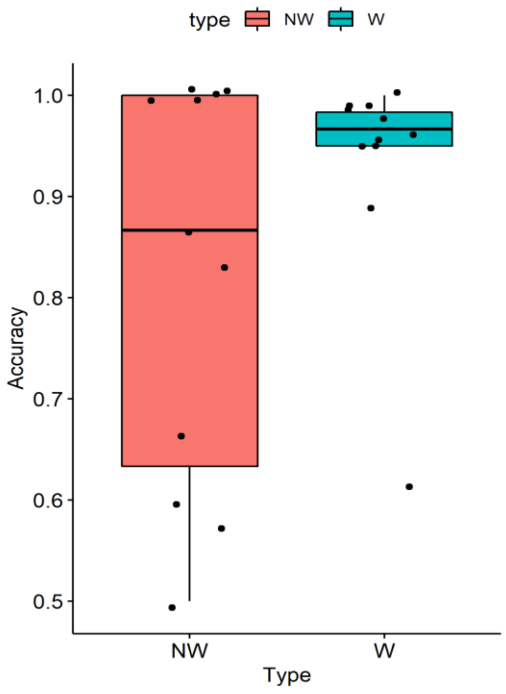


Figure 6. Accuracy in the lexical decision task.

In the SPM task, we found different effects in RTs and accuracy. In regard to RTs, participants were significantly slower in responding to clitic relative to both active and passive sentences, which did not differ to each other (parameters of the model and of the estimated marginal means are summarized in tables 7 and 8, respectively). Figures 7 and 8 show RTs at the individual and group levels, respectively. In regard to accuracy, participants were more accurate in recognizing active sentences relative to both clitic and passive sentences (figure 9 and table 9). However, in the post-hoc comparison with estimated marginal means, only the difference between active and passive sentences resulted significant (table 10). This might be due to the variability observed for responses to clitic sentences, as shown in figure 9.

Parameter	Estimate	Standard error	t value	p value
Active (intercept)	1.5900	0.1821	8.731	< 0.001
Clitic vs active	0.2745	0.1074	2.556	0.010
Passive vs active	0.0177	0.1085	0.164	0.870

Table 7. Parameters of the generalized linear mixed-effects model for RTs in SPM.

Contrast	Estimate	Standard error	z ratio	p value
Active vs clitic	- 0.2745	0.1074	-2.556	0.0286
Active vs passive	- 0.0178	0.1085	-0.164	0.9853
Clitic vs passive	0.2568	0.0889	2.887	0.0109

Table 8. Post-hoc constrasts of estimated marginal means for RTs in SPM.

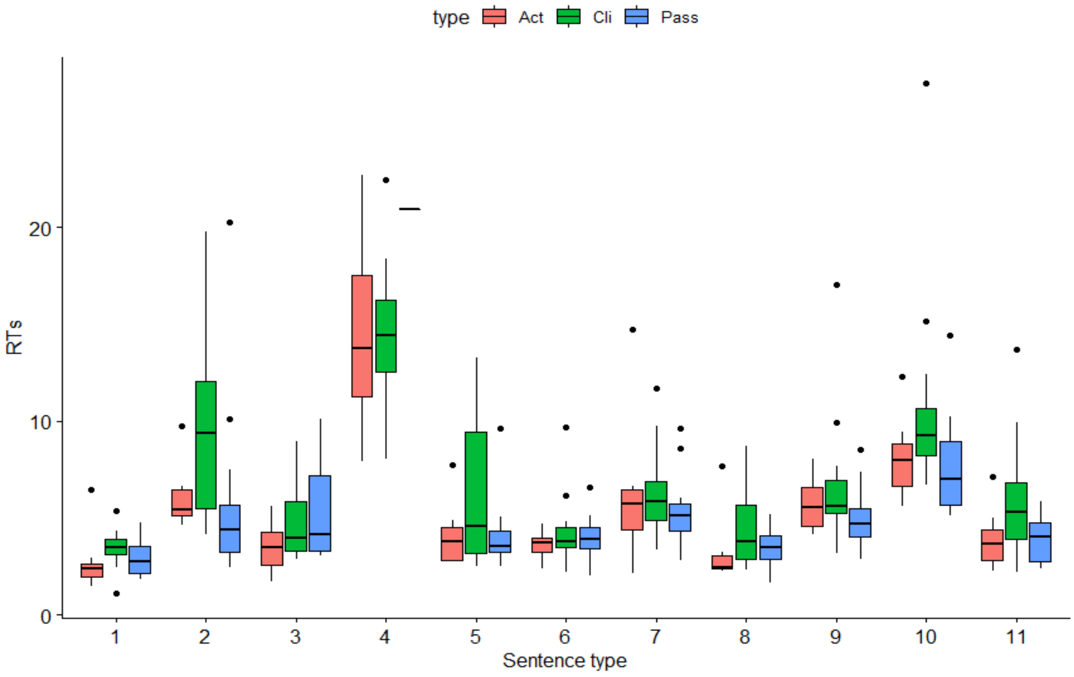


Figure 7. Individual RTs in the SPM task (values in the y-axis provided in seconds).

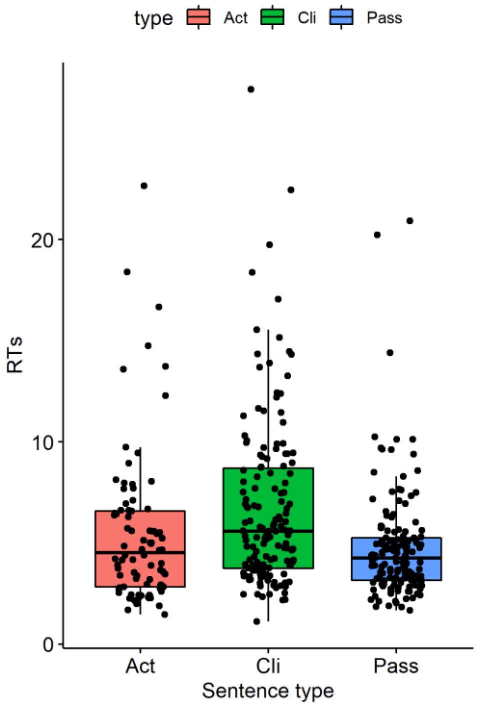


Figure 8. RTs in the SPM task (in seconds).

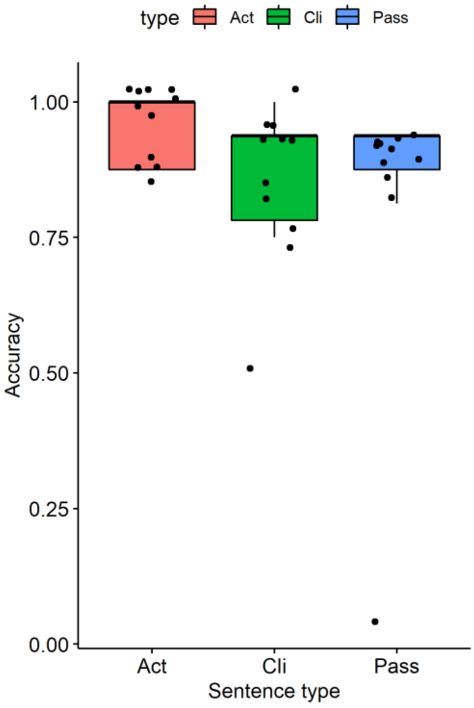


Figure 9. Accuracy in the SPM task.

Parameter	Estimate	Standard error	z value	p value
Active (intercept)	4.2315	0.8539	4.955	< 0.001
Clitic vs active	-1.7696	0.8225	-2.152	0.031
Passive vs active	-1.9335	0.8231	-2.349	0.019

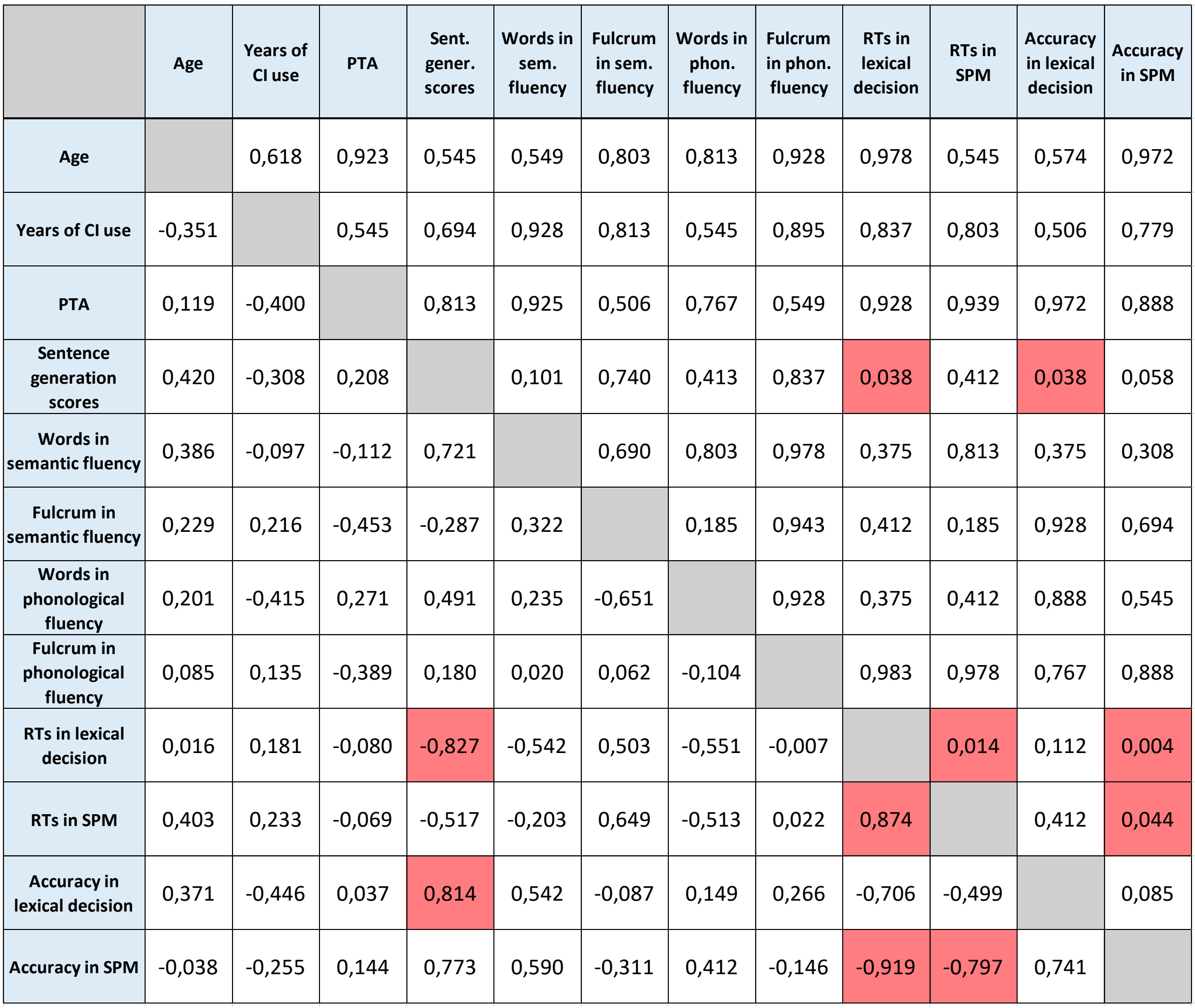
Table 9. Parameters of the generalized linear mixed-effects model for accuracy in SPM.

Contrast	Estimate	Standard error	z ratio	p value
Active vs clitic	1.770	0.822	2.152	0.0798
Active vs passive	1.933	0.823	2.349	0.0494
Clitic vs passive	0.164	0.538	0.305	0.9502

Table 10. Post-hoc constrasts of estimated marginal means for accuracy in SPM.

4.3.3 Correlations

Figure 10 shows the correlation matrix for our analyses. We did not find any significant correlation between measures from the linguistic tests and age, years of CI use normalized by age and PTA. However, we found some correlations between different measures of linguistic abilities. Sentence generation scores were correlated with both RTs and accuracy in lexical decision (R = -0.827, $p_{\mathrm{FDR}} = 0.0038$ for RTs; R = 0.814, $p_{\mathrm{FDR}} = 0.0038$ for accuracy). This suggests that participants who were faster and more accurate in recognizing words and nonwords were also better at producing sentences (figures 11 and 12). Other significant correlations were found between the two comprehension tasks. In particular, both RTs and accuracy in the SPM task were correlated with RTs in lexical decision (R = 0.874, $p_{\mathrm{FDR}} = 0.014$ for RTs; R = -0.919, $p_{\mathrm{FDR}} = 0.004$ for accuracy), in such a way that participants who were faster in recognizing words and nonwords were both faster and more accurate in comprehending sentences (figures 13 and 14). Within the SPM task, RTs were negatively correlated with accuracy (R = -0.797, $p_{\mathrm{FDR}} = 0.044$), meaning that participants who were faster were also more accurate in responding to sentences (figure 15).



	Age	Years of CI use	PTA	Sent. gener. scores	Words in sem. fluency	Fulcrum in sem. fluency	Words in phon. fluency	Fulcrum in phon. fluency	RTs in lexical decision	RTs in SPM	Accuracy in lexical decision	Accuracy in SPM
Age		0,618	0,923	0,545	0,549	0,803	0,813	0,928	0,978	0,545	0,574	0,972
Years of CI use	-0,351		0,545	0,694	0,928	0,813	0,545	0,895	0,837	0,803	0,506	0,779
PTA	0,119	-0,400		0,813	0,925	0,506	0,767	0,549	0,928	0,939	0,972	0,888
Sentence generation scores	0,420	-0,308	0,208		0,101	0,740	0,413	0,837	0,038	0,412	0,038	0,058
Words in semantic fluency	0,386	-0,097	-0,112	0,721		0,690	0,803	0,978	0,375	0,813	0,375	0,308
Fulcrum in semantic fluency	0,229	0,216	-0,453	-0,287	0,322		0,185	0,943	0,412	0,185	0,928	0,694
Words in phonological fluency	0,201	-0,415	0,271	0,491	0,235	-0,651		0,928	0,375	0,412	0,888	0,545
Fulcrum in phonological fluency	0,085	0,135	-0,389	0,180	0,020	0,062	-0,104		0,983	0,978	0,767	0,888
RTs in lexical decision	0,016	0,181	-0,080	-0,827	-0,542	0,503	-0,551	-0,007		0,014	0,112	0,004
RTs in SPM	0,403	0,233	-0,069	-0,517	-0,203	0,649	-0,513	0,022	0,874		0,412	0,044
Accuracy in lexical decision	0,371	-0,446	0,037	0,814	0,542	-0,087	0,149	0,266	-0,706	-0,499		0,085
Accuracy in SPM	-0,038	-0,255	0,144	0,773	0,590	-0,311	0,412	-0,146	-0,919	-0,797	0,741	

Figure 10. Correlation matrix. The upper half reports p values corrected for multiple comparisons, the lower half reports R values. Significant correlations ($p_{\mathrm{FDR}} < 0.05$) are coloured in red.

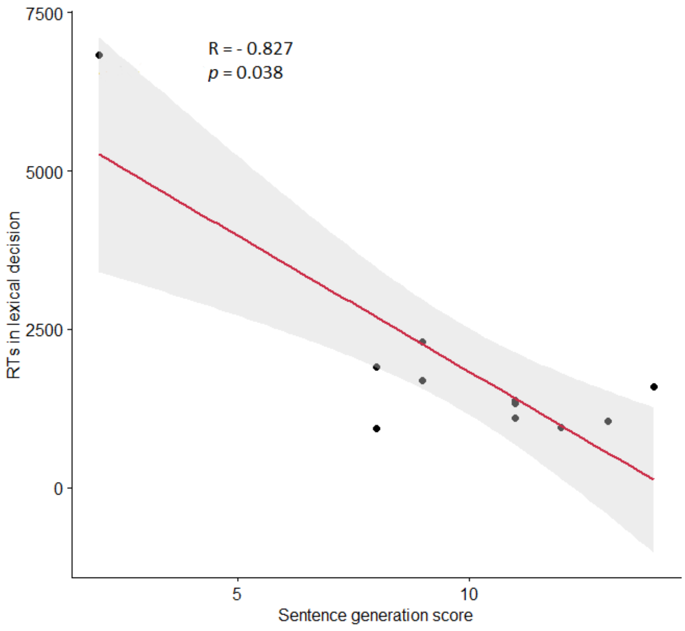


Figure 11. Scatter plot showing the correlation between sentence generation scores and RTs in lexical decision task.

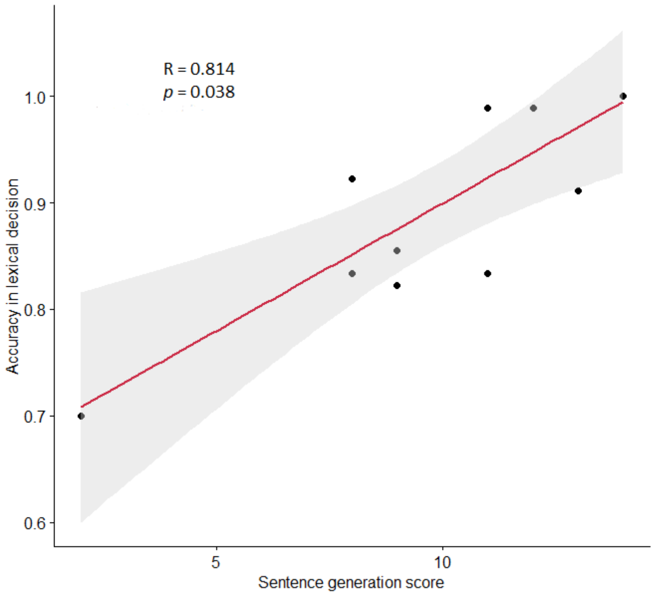


Figure 12. Scatter plot showing the correlation between sentence generation scores and accuracy in lexical decision task.

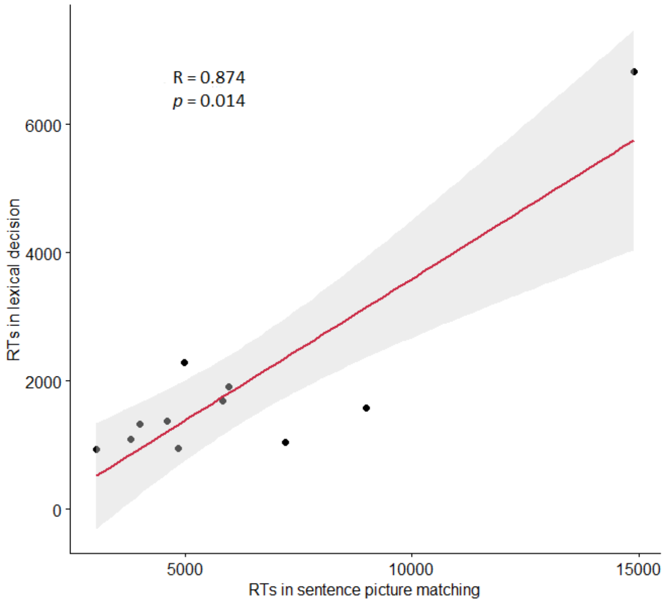


Figure 13. Scatter plot showing the correlation between RTs in SPM task and RTs lexical decision task.

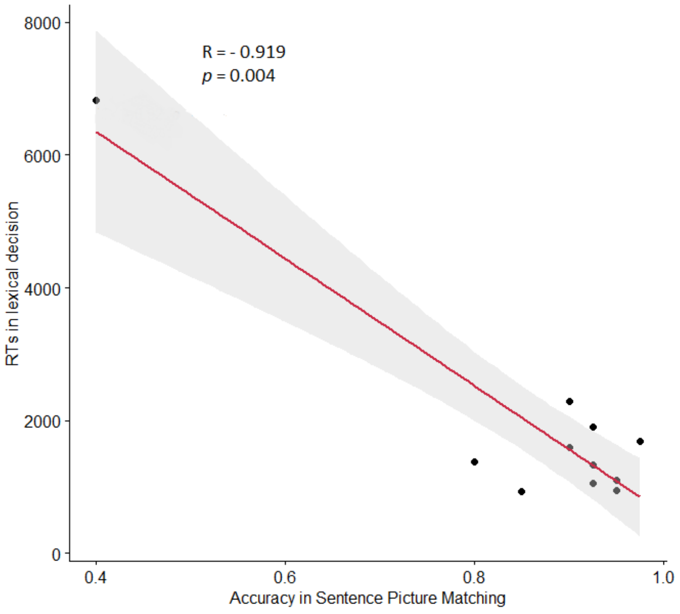


Figure 14. Scatter plot showing the correlation between accuracy in SPM task and RTs in lexical decision task.

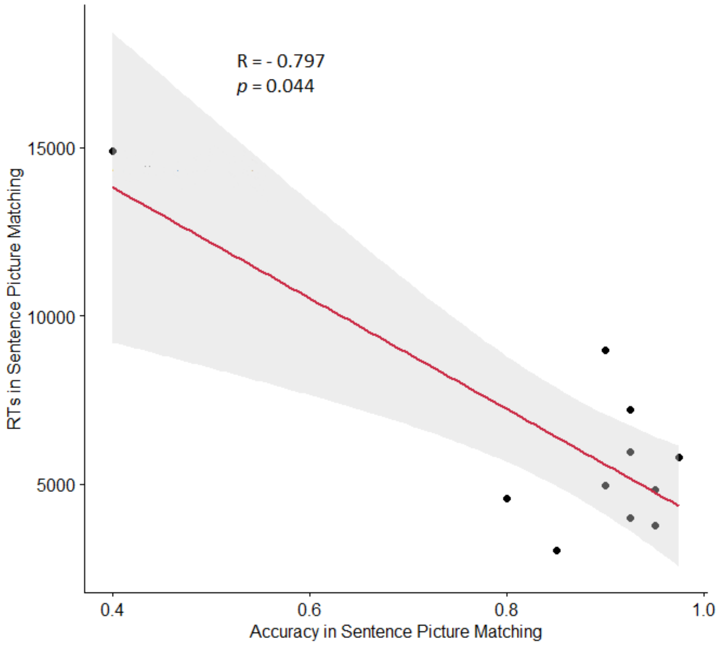


Figure 15. Scatter plot showing the correlation between RTs and accuracy in SPM task.

4.4 Discussion

We performed preliminary analyses on the behavioural data we have collected from 11 participants with CI. The tasks that we employed measured general linguistic abilities in comprehension and production domains, including lexical access, width of vocabulary knowledge and grammatical abilities. One main finding is that none of the measures from the tests was correlated with age of participant, years of experience with CI or PTA. On the one side, this might be due to the fact that the sample here presented is small, while a larger sample with greater variability might highlight correlations with these variables. On the other side, the null result might suggest that such characteristics do not influence the linguistic abilities tested in these tasks. Indeed, we expressly chose tasks that are simple enough to not tax excessively younger with respect to older participants; this might explain the lack of correlation between age and our linguistic measures. Also, irrespective of age, all participants had several years of experience with CI. After implantation, it is generally observed an improvement in speech perception and then a plateau once CI users have sufficiently adapted to the CI input (e.g., Strelnikov et al., 2009). According to the developmental age of the individual, a similar pattern might be present also for linguistic abilities, which would be in line with our finding that years of CI use do not correlate with linguistic measures. Regarding the PTA, we suggest that a lack of correlation indicates that a good functioning of the device does not necessarily lead to good language development in prelingually deaf CI users, who are the majority of our sample. Perceiving and categorizing speech sounds are only the first steps for the complex processes involved in speech comprehension. Thus, spoken language development largely depends on how the brain develops and adapt to external stimuli over time, well beyond the mere capacity to perceive isolated sounds. We suggest that audiological measures might not be actually informative of how an individual has adapted to the CI input, especially of an acoustically highly complex input such as speech: these measures neglect the most important aspect of CI use, which are real-life linguistic outcomes.

We found correlations between production and comprehension tasks, suggesting that they partly share common underlying abilities. In particular, sentence generation scores were correlated with both accuracy and RTs of lexical decision: participants who were faster and more accurate in discriminating words from nonwords were also better in generating meaningful sentences. Indeed, a sufficiently wide vocabulary knowledge and efficient access to stored lexical representations are a important requisites for efficient language production. Sentence generation in particular requires to retrieve, maintain and combine several linguistic representations to convey a more complex meaning. The correlation we found might suggest that comprehension and production partly share the same linguistic representations and access mechanisms, which is line with recent accounts unifying the study of the two domains (e.g., McQueen and Meyer, 2019; Pickering and Gambi, 2018). However, this interpretation will need further confirmation with analyses on a larger sample. In fact, some cautionary notes on these correlations are needed. First, some of the correlations are primarily driven by the presence of a single participant, who obtained low scores and was particularly slow in almost all tasks. Greater variability in the final sample might either confirm or disconfirm these trends. Second, if the hypothesis of shared mechanisms is correct, we might expect to find additional correlations between production and comprehension tests, such as correlations between fluency tasks and lexical decision, and between sentence generation and the Sentence Picture Matching task.

We administered a sentence generation task because we were interested in evaluating how participants were able to use syntactic and grammatical abilities to form new sentences. However, it should be considered that this task might not be sensitive to fine differences in grammatical competencies, as participants often formed short and simple sentences which could be evaluated with a score ranging only from 0 to 3. More sensitive measures with more nuanced scoring might be more appropriate to probe subtler lexical and grammatical skills of linguistic production. The Sentence Picture Matching task is more appropriate to evaluate comprehension of specific syntactic and grammatical structures. In particular, clitic sentences such as “lo dà la mamma” (“the mother gives it”) and “lo dà alla mamma” (“he/she gives it to the mother”, with implicit subject) are particularly difficult to distinguish in Italian without robust grammatical competencies. The crucial difference is between the article “la” (“the”) and the preposition “alla” (“to the”), which is a very small difference at the orthographic (and auditory) level, but that completely changes the meaning of the sentence. Children with CI learn articles and prepositions through a degraded speech input, even before studying them in the written modality, which might cause difficulties in distinguishing them and learning their function. Such difficulties have been shown to be transferred to the written modality, even after language acquisition. In fact, Artesini (2019) found that prelingually deaf CI users were significantly less accurate than NH controls in the SPM task with clitic sentences, both with implicit and explicit subjects, while such difference was not present for postlingually deaf CI users, who learnt grammar with a normal auditory input before hearing lost.

In our sample, we found a significant difference for clitic relative to active and passive sentences only in RTs; in accuracy, in the contrasts of estimated marginal means, only the difference between active and passive sentences resulted significant. This suggests that while clitic sentences require more time to be associated with their meaning, they do not necessarily lead to errors. In fact, as shown in figure 9, clitic sentences showed greater variability respect to both passive and active sentences, suggesting that some participants had difficulties in comprehending them, while others did not. This is coherent with the results reported by Artesini (2019), who also observed variability in accuracy for clitic sentences within the (prelingual) CI group. The comparison with the control group of NH controls will clarify the direction of these results. First, it will show whether slower processing of clitic sentences is typical of CI users or it is observed also in people with normal hearing. Second, it will help to clarify whether the lower accuracy for passive sentences with respect to active sentences is due to specific difficulties of prelingually deaf CI users with this syntactic structure (as reported anecdotally by some of our participants).

As for the sentence generation task, also for the SPM task we found correlations with results of lexical decision. In particular, we found that both RTs and accuracy in SPM were correlated with RTs in lexical decision, in such a way that participants who were faster in lexical recognition also performed better and faster in sentence comprehension. This is coherent with the idea that the benefits of a wider vocabulary knowledge and an efficient lexical access and can cascade to benefit also higher-level processing of sentences, such as syntactic and semantic processing. Furthermore, because both lexical decision and SPM tasks employed the written modality, individual reading abilities might partly explain the association between RTs in the two tasks. The relevance of reading abilities for our comprehension tests might raise the question of whether these measures can be actually related to individual abilities in spoken language comprehension, which is the domain we test in the EEG experiment. We suggest that this is indeed the case.

In a recent article reviewing evidence from multiple studies, Huettig and Pickering (2019) propose that reading skills contribute to develop better predictive abilities which transfer to comprehension of spoken language. According to the authors, this is due not only to the secondary benefits of reading, such as increased vocabulary knowledge and verbal working memory, but also to its direct effects on linguistic representations shared with the spoken domain. For example, the presence of evident word separations in written language creates a strong bias towards word units, which in turn leads to more precise word representations (Huettig and Pickering, 2019). This mechanism might be particularly useful for prelingually deaf CI users, who have been shown to display a reduced efficiency in exploiting prosodic cues, which are important to identify boundaries (e.g., Holt et al., 2016). Some authors also propose that learning to read leads to more fine-grained phonological representations (Pattamadilok et al., 2010) and to faster retrieval of such representations (Araújo et al., 2018). This might benefit especially CI users, who often show difficulties in fine phonological distinctions (e.g., Bouton et al., 2012). Because all these benefits in turn facilitate the ability to form predictions, we expect that participants who show better lexical recognition and reading abilities will also show more pronounced prediction-related neural activity in our EEG experiment. We also suggest that reading might be proposed as a valid aid for children and adolescents who use CIs. On the one hand, developing good reading abilities and reading habits might benefit both language acquisition and everyday speech comprehension. On the other hand, employing the written modality might be a useful adjunct in school education. For example, one of our participants reported that it is much easier for him to learn new words at school when the teacher writes them on the blackboard during the explanation. However, teachers are often not aware of how much simple actions like this can make a difference for students with CIs; cognitive psychologists might have a role in raising awareness on these topics.

Behavioural measures have often been used to study the mechanisms of access to linguistic knowledge in the lexicon. In this regard, we replicated robust effects documented in literature, both in lexical decision and fluency tasks. First, we found the effect of lexical status by which words are easier and faster to recognize in lexical decision. Several models have been proposed to explain such effect. For example, early models of visual word recognition conceptualized the process underlying lexical access as a search in the lexicon (e.g., Forster and Chamber, 1973); later computational models instead viewed such process as a propagation of activation from simple visual features, such as shapes of letters, to more abstract lexical representations (e.g., Coltheart et al., 2001). In any case, the process is faster for words than nonwords, as it stops when a corresponding lexical representation is found or sufficiently activated. It is also more accurate for words, as activation of orthographical neighbours of nonwords reduces the accuracy in rejecting them; this is likely the case in our experiment, in which nonwords were similar to existing words (Amenta et al., 2021). Second, we found a difference between semantic and fluency tasks, with the former eliciting a higher number of responses than the latter. Even though we did not test the effect statistically, this trend is in line with previous findings suggesting that semantic fluency is easier because of the network organization of semantic knowledge (Schmidt et al., 2017). Such organization facilitates the retrieval of semantically associated words belonging to the target category in the semantic fluency task; in contrast, phonological fluency is based on systematic search of words' initial phonemes, and requires to suppress the automatic retrieval of semantically associated words.

In the final analyses of this study, it will be interesting to compare these two effects in CI users and NH controls. Such comparison might shed light on the organization of linguistic knowledge in CI users, especially those who have experienced the CI input early in life. In this regard, Kenett et al. (2013) analysed results of a semantic fluency task extracting correlation-based semantic networks and found that networks of children with CI were under-developed with respect to the ones of their NH peers. As mentioned in chapter 3, this might hinder the ability to efficiently retrieve associated semantic information during language comprehension, and to use such information to predict upcoming input. The preliminary results here presented suggest that this is not the case for our sample, but this general trend will need further confirmation. Furthermore, a comparison with the control group will help to clarify the functional interpretation of the fulcrum point, which seems to be higher for phonological with respect to semantic fluency. For example, the lack of semantic associations between words in the phonological fluency task might lead to a more uniform distribution of responses across the trial; if that is the case, we should expect to observe a similar pattern also in NH participants.

We administered to participants the behavioural tests here presented to obtain various measures of their linguistic abilities. However, we are aware that these measures are approximate and do not fully represent individual abilities to produce and comprehend speech in real-life situations. For example, in everyday life the context of the discourse allows to infer much of the information that might be missed at first, either by predicting or later revising one’s own interpretation. Context might also help to compensate for possible difficulties with specific grammatical structures. Nevertheless, our primary aim is to explore if and how these measures are related with the results of the EEG experiment. As discussed in chapter 3, different explanations have been proposed for the observed group differences between CI users and NH peers in predictive behaviour, including slower processing speed and more difficult lexical competition. The studies here reviewed suggest that individual differences cannot be overlooked when studying language in this clinical population. In the same vein, we believe that measuring different aspects of linguistic competence will contribute to interpreting the EEG data and to exploring the sources of individual differences both among CI users and NH participants.

Conclusions

Prediction is an adaptive strategy to aid language comprehension in everyday life. It is implemented flexibly according to the situations, and might interact with other cognitive mechanisms involved in comprehension. In particular, the research project presented in this manuscript aims at investigating prediction during auditory comprehension of language and its possible interactions with processing of mouth articulatory cues, which also facilitate speech comprehension at multiple levels. We are interested in examining the electrophysiological correlates of prediction in a special population which faces adverse listening conditions on a daily basis, cochlear implant users, as these mechanisms of facilitation might be particularly relevant for them. We are also interested in investigating possible sources of individual differences in prediction, by means of linguistic tests which provide us with a general picture of the linguistic abilities of each participant.

With this study, we want to point out that it is not sufficient to perceive sounds in order to comprehend spoken language. The quality of the auditory input makes the difference when it comes to complex auditory stimuli such as speech. However, the difficulties faced by people who use cochlear implants are often not acknowledged. In the medical field, clinical evaluations of implantation's outcomes are mainly based on audiological measures such as detection of pure tones and on "intelligibility measures" which evaluate the capacity of the individual to repeat single words. While these tests ensure that the person correctly perceive the words, they neglect whether they are able to comprehend and use them in everyday life, which is not automatic. The dialogue with the participants and their parents during the first period of data collection gave us important insights on the their subjective experience with language development and everyday language comprehension.

An aspect that emerged as particularly important is the experience with speech therapy during childhood. While some participants reported that it was particularly useful to learn to distinguish and pronounce similar phonemes (e.g., "f" from "v"), others did not have very positive experiences (one even refused to continue the treatment). Obviously, speech therapy must be adjusted to individual capacities and needs, and its efficacy ultimately influences linguistic outcomes, such as vocabulary knowledge and production abilities.

One common experience among the families of prelingually deaf children is the suggestion from speech therapists to not use sign language, but expose the child to oral language only. This is indeed a debate topic in the field. According to the experience of our participants, most speech therapists and physicians claim that learning sign language would be pointless as these children are able to hear thanks to the device. Some even claim that exposure to sign language would be detrimental, as it would inhibit the development of oral language. However, evidence suggest that is not the case. Researchers report that children with CI who learn both sign and oral languages have better linguistic outcomes than children who are exposed to oral language only (e.g., Newman, 2019). This is because sign languages have many fundamental properties in common with oral languages (e.g., form-to-meaning mapping, syntactic structures, a proper grammar) and the benefits of learning such general features transfer also to oral language development. In a similar way, benefits deriving from reading transfer to the spoken language domain, as discussed in the previous chapter. We suggest that a greater consideration of these factors might significantly benefit language development of CI users.

Our participants also reported various experiences with school institutions. In some cases, schools took measures to facilitate students with CIs, such as soundproofing the classroom to reduce background noise or make the teachers use special microphones connected to the implant. In other cases, parents reported to have found a lack of collaboration from some teachers, when they asked to take simple actions such as writing new words on the blackboard during explanation, or let their child be seated at first desks in the classroom. This is due to a lack of knowledge of the real experience of CI users. The common idea, in those cases, is that these children do not need additional facilitation as the device allows them to hear. The experiences reported by our participants and their parents clearly indicate that a raising in awareness among teachers and other relevant figures in children’s life would really make a difference.

Finally, our participants described some of the strategies they use to aid their comprehension of speech in everyday life. First, they confirmed to us that they often do not understand clearly some words, but they infer them from the context most of the times. One participant also told us that he often misses the first word of sentences if he is not paying attention to the speaker. Thus, he suggests to the people close to him to make a gesture or make eye contact when they have the intention to speak; in this way, he can direct his attention and comprehend much better. Second, they generally pay great attention to mouth movements, especially if the words they hear are ambiguous in some way (e.g., if the speaker does not mark the difference between similar phonemes). Surprisingly, in the last few years some of them have learnt to use some visual articulatory cues even when the speaker wears a face mask. For example, the movement of the cheeks, visible immediately below the eyes, indicates to them that the person is starting to speak; particular movements of the chin direct their interpretation of ambiguous sounds towards certain phonemes rather than others.

To conclude, use of sentence context and other compensatory strategies are largely overlooked in clinical practice, in which the functioning of the device in terms of sound perception is prioritized. We hope that this and future studies in the same direction will provide a better understanding of these aspects and will ultimately contribute to improving the experience of CI users. For example, speech therapy during childhood might benefit from use of predictive contexts to facilitate comprehension and acquisition of new words. Trainings on predictive abilities and lip-reading abilities might also be useful to enhance comprehension abilities in everyday life. Furthermore, raising awareness on the cognitive aspects underlying speech comprehension with a CI and on the factors that facilitate it might improve the approach of familiars, teachers and other relevant figures to this topic. Finally, from a theoretical perspective, the comparison among NH listeners, prelingually deaf CI users and postlingually deaf CI users might provide important insights on how language experience influences the mechanisms of spoken language comprehension.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264

Amenta, S., Badan, L., & Brysbaert, M. (2021). LexITA: A quick and reliable assessment tool for Italian L2 receptive vocabulary size. *Applied Linguistics*, *42*(2), 292-314.

Anderson, C. A., Lazard, D. S., & Hartley, D. E. (2017). Plasticity in bilateral superior temporal cortex: Effects of deafness and cochlear implantation on auditory and visual speech processing. *Hearing research*, *343*, 138-149.

Araújo, S., Fernandes, T., & Huettig, F. (2019). Learning to read facilitates the retrieval of phonological representations in rapid automatized naming: Evidence from unschooled illiterate, ex-illiterate, and schooled literate adults. *Developmental Science*, *22*(4), e12783.

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A.-L. (2009). Dual neural routing of visual facilitation in speech processing. The Journal of Neuroscience, 29, 13445–13453.

Artesini, L. (2019). *ERP correlates of semantic and syntactic processing in cochlear implant users* (Doctoral dissertation, University of Trento).

Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *cortex*, *49*(5), 1195-1205.

Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., ... & Gaudrain, E. (2016). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: How and to what degree can it be achieved?. *Trends in Hearing*, *20*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bauer, M., Stenner, M. P., Friston, K. J., & Dolan, R. J. (2014). Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. *Journal of Neuroscience, 34*(48), 16117e16125.

Beauchamp, M. S. (2016). Audiovisual speech integration: Neural substrates and behavior. In *Neurobiology of language* (pp. 515-526). Academic Press.

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A. M. (2005). Colfis (corpus e lessico di frequenza dell'italiano scritto). *Available on http://www. istc. cnr. it/material/database*, 67-73.

Bhargava, P., Gaudrain, E., & Başkent, D. (2014). Top–down restoration of speech in cochlear-implant users. *Hearing research*, *309*, 113-123.

Bidet-Caulet, A., Barbe, P. G., Roux, S., Viswanath, H., Barthélémy, C., Bruneau, N., ... & Bonnet-Brilhault, F. (2012). Dynamics of anticipatory mechanisms during predictive context processing. *European Journal of Neuroscience*, *36*(7), 2996-3004.

Blomquist, C., Newman, R. S., Huang, Y. T., & Edwards, J. (2021). Children with cochlear implants use semantic prediction to facilitate spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *64*(5), 1636-1649.

Bouton, S., Serniclaes, W., Bertoncini, J., & Colé, P. (2012). Perception of speech features by French-speaking children with cochlear implants.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., ... & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *science*, *276*(5312), 593-596.

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2018). Beta oscillatory power modulation reflects the predictability of pitch change.*Cortex, 106,* 248e260.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39,* e62.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, *108*(1), 204.

Cohen, M. X. (2014). Analyzing Neural Time Series Data: Theory and Practice. The MIT Press.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing research*, *229*(1-2), 132-147.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and linguistics compass*, *8*(12), 631-645.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117–1121.

Dingemanse, J. G., & Goedegebure, A. (2019). The important role of contextual information in speech perception in cochlear implant users and its consequences in speech tests. *Trends in hearing*, *23*.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior, 20*(6), 641-655.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, *12*(6), 627-635.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127-138.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of experimental psychology: Human perception and performance*, *6*(1), 110.

Gastaldon, S., Arcara, G., Navarrete, E., & Peressotti, F. (2020). Commonalities in alpha and beta neural desynchronizations during prediction in language comprehension and production. *Cortex*, *133*, 328-345.

Geng, J. J., & Vossel, S. (2013). Re-evaluating the role of TPJ in attentional control: Contextual updating? Neuroscience and Biobehavioral Reviews, 37(10), 2608–2620.

Giraud, A. L., Price, C. J., Graham, J. M., Truy, E., & Frackowiak, R. S. (2001). Cross-modal plasticity underpins language recovery after cochlear implantation. *Neuron*, *30*(3), 657-664.

Goslin, J., Galluzzi, C., & Romani, C. (2014). PhonItalia: a phonological lexicon for Italian. Behavior Research Methods, 46(3), 872–886.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. Journal of the Acoustical Society of America, 103, 2677e2690.

Green, K. M., Julyan, P. J., Hastings, D. L., & Ramsden, R. T. (2005). Auditory cortical activation and speech perception in cochlear implant users: effects of implant experience and duration of deafness. *Hearing research*, *205*(1-2), 184-192.

Griffiths, B., Mazaheri, A., Debener, S., & Hanslmayr, S. (2016). Brain oscillations track the formation of episodic memories in the real world. *Neuroimage, 143,* 256e266.

Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., & Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain research*, *1361*, 54-66.

Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. *Journal of cognitive neuroscience*, *25*(5), 706-718.

Gugliotta, M. (2009). *BVN 12-18: batteria di valutazione neuropsicologia per l'adolescenza*. Edizioni Erickson.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1055-1069.

Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in Human Neuroscience, 6*.

Holt, R., Bruggeman, L., & Demuth, K. (2021). Children with hearing loss can predict during sentence processing. *Cognition*, *212*, 104684.

Holt, C. M., Demuth, K., & Yuen, I. (2016). The use of prosodic cues in sentence processing by prelingually deaf users of cochlear implants. *Ear and Hearing*, *37*(4), e256-e262.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626,* 118–135.

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in cognitive sciences*, *23*(6), 464-475.

Hunter, C. R., & Pisoni, D. B. (2021). Some Neuromyths and Challenging Questions about Cochlear Implants. *The Handbook of Speech Perception*, 540-569.

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 645–662.

Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. Frontiers in Psychology, 2, 255.

Ito, A., Gambi, C., Pickering, M. J., Fuellenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia, 136*, 107291.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in human neuroscience*, *4*, 186.

Karthik, G., Plass, J., Beltz, A. M., Liu, Z., Grabowecky, M., Suzuki, S., ... & Brang, D. (2021). Visual speech differentially modulates beta, theta, and high gamma bands in auditory cortex. *European Journal of Neuroscience*, *54*(9), 7301-7317.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, *18*(7), 1560-1574.

Kenett, Y. N., Wechsler-Kashi, D., Kenett, D. Y., Schwartz, R. G., Ben-Jacob, E., & Faust, M. (2013). Semantic organization in children with cochlear implants: Computational analysis of verbal fluency. *Frontiers in Psychology*, *4*, 543.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology*, *62*, 621.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Kutas, M., Lindamood, T.E., Hillyard, S.A., (1984). Word expectancy and event-related brain potentials during sentence processing. In *Kornblum, S., Requin, J. (Eds.), Preparatory States and Processes.* Erlbaum, Hillsdale, NJ, pp. 217–237.

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious?. *Perspectives on psychological science*, *7*(1), 18-27.

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.

Lenth, R.V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.0.

León-Cabrera, P., Piai, V., Morís, J., & Rodríguez-Fornells, A. (2022). Alpha power decreases associated with prediction in written and spoken sentence comprehension. Neuropsychologia, 173, 108286.

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex, 68,* 155– 168.

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. Frontiers in Psychology, 6, 1171.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1.

Luck, S. J. (2014). An introduction to the event-related potential technique. A Bradford Book.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS biology*, *8*(8), e1000445.

Macherey, O., & Carlyon, R. P. (2014). Cochlear implants. Current Biology, 24(18), R878–R884.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake- But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843–847.

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports, 8*(1).

Masina, F., Pezzetta, R., Lago, S., Mantini, D., Scarpazza, C., & Arcara, G. (2022). Disconnection from prediction: a systematic review on the role of right temporoparietal junction in aberrant predictive processing. Neuroscience & Biobehavioral Reviews, 104713.

Mattys, S. L., Bernstein, L. E., & Auer, E. T. (2002). Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception & Psychophysics*, *64*(4), 667-679.

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition*, *169*, 147-164.

McQueen, J., & Meyer, A. S. (2019). Key Issues and Future Directions: Toward a Comprehensive Cognitive Architecture for Language Use. In P. Hagoort (Ed.), Human Language: From Genes and Brains to Behavior (pp. 85–96). The MIT Press.

Meeuwissen, E. B., Takashima, A., Fernandez, G., & Jensen, O. (2011). Evidence for human fronto-central gamma activity during long-term memory encoding of word sequences. *Plos One, 6*(6), Article e21356.

Michon, M., Boncompte, G., & López, V. (2020). Electrophysiological dynamics of visual speech processing and the role of orofacial effectors for cross-modal predictions. *Frontiers in Human Neuroscience*, *14*, 538619

Moberly, A. C., & Reed, J. (2019). Making sense of sentences: top-down processing of speech by adult cochlear implant users. *Journal of Speech, Language, and Hearing Research*, *62*(8), 2895-2905.

Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neuroscience letters*, *363*(2), 112-115.

Nagels, L., Bastiaanse, R., Başkent, D., & Wagner, A. (2020). Individual differences in lexical access among cochlear implant users. *Journal of Speech, Language, and Hearing Research*, *63*(1), 286-304.

Nath, A. R., & Beauchamp, M. S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *Journal of Neuroscience*, *31*(5), 1704-1714.

Newman, A. J. (2019). Language Development in Deaf Children. *The Oxford Handbook of Neurolinguistics*, 339.

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia, 142,* 107427.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife, 7.*

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral cortex*, *20*(3), 633-640.

Oldfield, N. (1971). Edinburgh handedness questionnaire. *Neuropsychologia*, *9*, 97-113.

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *Elife*, *5*, e14521.

Pattamadilok, C., Knierim, I. N., Duncan, K. J. K., & Devlin, J. T. (2010). How does learning to read affect speech perception?. *Journal of Neuroscience*, *30*(25), 8435-8444.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, *68*, 169-181.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002–1044.

Pisoni, D., Kronenberger, W., Roman, A., & Geers, A. (2011). Article 7: Measures of digit span and verbal rehearsal speed in deaf children following more than 10 years of cochlear implantation. *Ear and hearing*, *32*(1), 60s.

Prystauka, Y., & Lewis, A. G. (2019). The power of neural oscillations to inform sentence comprehension: A linguistic perspective. Language and Linguistics Compass, 13(9).

Rommers, J., Dickson, D. S., Norton, J. J. S., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience, 32*(5), 576–589.

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object Shape and Orientation Do Not Routinely Influence Performance During Language Processing. Psychological Science, 24(11), 2218–2225.

Rouger, J., Lagleyre, S., Démonet, J. F., Fraysse, B., Deguine, O., & Barone, P. (2012). Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Human brain mapping*, *33*(8), 1929-1940.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences*, *104*(17), 7295-7300.

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, *13*(2), 231-252.

Sauseng, P., Klimesch, W., Schabus, M., & Doppelmayr, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International journal of Psychophysiology*, *57*(2), 97-103.

Schmidt, C. S., Schumacher, L. V., Römer, P., Leonhart, R., Beume, L., Martin, M., ... & Kaller, C. P. (2017). Are semantic and phonological fluency based on the same or distinct sets of cognitive processes? Insights from factor analyses in healthy adults and stroke patients. *Neuropsychologia*, *99*, 148-155.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24*(2), 232-252.

Siman-Tov, T., Granot, R. Y., Shany, O., Singer, N., Hendler, T., & Gordon, C. R. (2019). Is there a prediction network? Meta-analytic evidence for a corticalsubcortical network likely subserving prediction. *Neuroscience & Biobehavioral Reviews, 105*, 262–275.

Smiljanic, R., & Sladen, D. (2013). Acoustic and semantic enhancements for children with cochlear implants.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319.

Song, J. J., Lee, H. J., Kang, H., Lee, D. S., Chang, S. O., & Oh, S. H. (2015). Effects of congruent and incongruent visual cues on speech perception and brain activity in cochlear implant users. *Brain Structure and Function*, *220*(2), 1109-1125

Spahr, A. J., Dorman, M. F., & Loiselle, L. H. (2007). Performance of patients using different cochlear implant systems: effects of input dynamic range. *Ear and Hearing*, *28*(2), 260-275.

Stevenson, R., Sheffield, S. W., Butera, I. M., Gifford, R. H., & Wallace, M. (2017). Multisensory integration in cochlear implant recipients. *Ear and hearing*, *38*(5), 521.

Strelnikov, K., Rouger, J., Barone, P., & Deguine, O. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian journal of psychology*, *50*(5), 437-444.

Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Démonet, J. F., Déguine, O., & Barone, P. (2015). Increased audiovisual integration in cochlear-implanted deaf patients: independent components analysis of longitudinal positron emission tomography data. *European Journal of Neuroscience*, *41*(5), 677-685.

Szewczyk, J. M., & Schriefers, H. (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. *Language, Cognition and Neuroscience, 33*(6), 665–686.

Tadel, F., Bock, E., Niso, G., Mosher, J. C., Cousineau, M., Pantazis, D., ... & Baillet, S. (2019). MEG/EEG group analysis with brainstorm. *Frontiers in neuroscience*, 76.

Terporten, R., Schoffelen, J.-M., Dai, B., Hagoort, P., & Kösem, A. (2019). The Relation between Alpha/Beta Oscillations and the Encoding of Sentence induced Contextual Information. *Scientific Reports, 9*(1), 20255.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*(3), 382–392.

Tversky, A., Kahneman, D., 1973. Availability: a heuristic for judging frequency and probability. *Cogn. Psychol. 5*(2), 677–695.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007a). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. Ear and Hearing, 28, 656e668.

Tye-Murray, N., Sommers, M., & Spehar, B. (2007b). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification*, *11*(4), 233-241.

Urbach, T. P., DeLong, K. A., Chan, W.-H., & Kutas, M. (2020). An exploratory data analysis of word form prediction during word-by-word reading. *Proceedings of the National Academy of Sciences, 117*(34), 201922028.

Van Berkum, J. J. (2009). The neuropragmatics of'simple utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory* (pp. 276-316). Palgrave Macmillan.

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior research methods*, *38*(4), 584-589.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190

Wang, L., Hagoort, P., & Jensen, O. (2018). Language Prediction Is Reflected by Coupling between Frontal Gamma and Posterior Alpha Oscillations. *Journal of Cognitive Neuroscience, 30*(3), 432–447

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: new directions.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288.

Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, *20*.

Acknowledgements

I want to thank some people who helped and supported me during the preparation of this thesis. This section will be written in Italian.

Voglio ringraziare la mia relatrice, la professoressa Francesca Peressotti, per avermi proposto di collaborare a questo progetto, permettendomi di crescere professionalmente e come persona. Grazie per la disponibilità e l'aiuto nel costruire il mio percorso.

Ringrazio Simone Gastaldon, non solo per l'aiuto fondamentale nella scrittura di questa tesi, ma soprattutto per avermi insegnato tantissimo, dalla preparazione dell'esperimento alla raccolta e analisi dei dati, con grande pazienza e impegno. E' stato un piacere condividere questo lavoro con te. E' stato un percorso non sempre facile, ma magnifico. Grazie di tutto.

Voglio ringraziare tutti i partecipanti e le loro famiglie che ho incontrato durante questo progetto per la loro gentilezza e disponibilità a condividere le loro storie, che hanno arricchito questa esperienza al livello umano.

Ringrazio la mia famiglia per avermi supportato nel corso dei miei studi, anche quando le mie insicurezze facevano sembrare tutto difficile. E' grazie al vostro sostegno che sono potuta arrivare fino a qui.

Ringrazio le mie amiche e i miei amici di Messina. Per quanto io possa stare lontano, voi siete un punto fisso a cui tornare sempre.

Ringrazio le mie amiche e i miei amici a Padova, perché grazie a loro mi sono sentita accolta e a casa. Avete reso stupendo il periodo in cui ho vissuto qui.

Ringrazio la mia amica Federica, per aver condiviso le ansie (ma anche le gioie) dei nostri percorsi universitari, e per essere una grande amica per la vita. So che ci saremo sempre l'una per l'altra.

Infine, voglio ringraziare il mio fidanzato, Salvo, per l'immenso supporto che mi dà ogni giorno. Tu mi dai forza, e coraggio verso il futuro. Ti ringrazio infinitamente, perché sono fortunata ad avere la possibilità di viverlo con te.

Novembre 2022