Effects of Trait Anxiety on Threatening Speech Processing

Implications for Models of Emotional Language and Anxiety

Simon Busch-Moreno

Faculty of Brain Sciences

Division of Psychology and Language Sciences

Dissertation submitted for the degree of Doctor of Philosophy

University College London

November, 2020

Declaration

I, Simon Ernesto Busch Moreno confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Simon Busch-Moreno

Abstract

Speech can convey emotional meaning through different channels, two are regarded as particularly relevant in models of emotional language: prosody and semantics. These have been widely studied in terms of their production and processing aspects, but sometimes overlooking individual differences of listeners. The present thesis examines whether greater intrinsic levels of anxiety can affect threatening speech processing. Trait anxiety is the predisposition to increased cognitions such as worry (over-thinking of the future), and emotions such as angst (felling of discomfort and tension), and can be reflected by an overactive behavioural inhibition system. As a result, according to emotional language and anxiety models, emotional prosody/semantics and anxiety might have overlapping neural areas/routes and processing phases. Thus, threatening semantics or prosody could have differential effects on trait anxiety depending on the nature of this overlap. This problem is approached by using behavioural and electroencephalographic (EEG) measures. Three dichotic listening experiments demonstrate that, at the behavioural level, trait anxiety does not modulate lateralisation when stimuli convey threatening prosody, threatening semantics or both. However, these and another non-dichotic experiment indicate that greater anxiety induces substantially slower responses. An EEG experiment shows that this phenomenon has very clear neural signature at late processing phases (~600ms). Exploratory source localisation analyses indicate involvement of areas predicted by the models, including portions of limbic, temporal and prefrontal cortex. The proposed explanation is that threatening speech can induce anxious people to over-engage with stimuli, and this disrupts late-phase processes associated with orientation/deliberation, as proposed by anxiety models. This process is independent of information type until later phase occurring after speech comprehension (e.g. response preparation/execution). Given this, a new model of threatening language processing is proposed, which extends models of emotional language processing by incorporating an orientation/deliberation phase from anxiety models.

Impact Statement

The present thesis portrays the core of a long-run research project aimed at understanding cognition and emotion by studying the relationship between language and anxiety. In particular, the effects of trait anxiety on threatening speech processing can reveal the intermingling of cognitive and emotional processes via precise measurements of speech, anxiety and comprehension. This relationship can have a direct impact on theoretical models on emotional language and anxiety. By researching them together, it is not only possible to test their derived hypotheses, but also allowing to extend these models. Furthermore, beyond the basic science research carried on by this project, it can develop into many practical applications. For instance, understanding how threatening speech affects trait anxiety behavioural responses can provide crucial information for anxiety assessment through speech. Moreover, understanding the temporal neural signature of anxious responses to threatening speech can help to understand what the particular brain responses of anxiety are. In this way, present research can provide a solid base for developing assessment and/or treatment of anxiety- and speech-related issues. Therefore, the main impact of this thesis is relevant to both experimental and theoretical domains; which can provide useful evidence for developing future applied science approaches. Indirectly, areas such as artificial emotional speech development or applied statistical modelling could receive relevant input from the present project. In short, the present project is a relevant contribution to theory, experimentation, methodological and practical domains.

Acknowledgements

This work would not be possible without the aid of my family. I want to specially thank my parents and sister, who supported me back in Chile and from there during my time in London; coping with the difficulty of being apart and helping me go through. From there, as well, thanks to all my friends for all their care and encouragement. And here, thanks to my family in London, who have received me son candidly and helped me so much. I cannot thank enough to my partner for her support, as she has endured all my downs and raises throughout this project with constant care. Also, thanks to my friends her in London, for their support and encouragement. Thanks, as well, to my supervisors, especially to my principal supervisor for constantly helping me with every detail about my work. And, finally, thanks to all professors, lecturers, family friends and colleagues who lent me hand in the way from bachelor to doctor.

Funding

Contents

Equations

Figures

Tables

Chapter 1
Introduction
A Link between Language and Anxiety

The relationship between emotion and cognition has been considered elusive, and this might result from understanding them as dissociable, instead of understanding them as part of a single broader process or at least tightly intertwined, where processes considered as cognitive like planning or thinking support and/or depend on processes considered emotional such as responses to feelings or appetitive/aversive responses (Damasio, 2012; Pessoa, 2018; Pessoa et al., 2019). Language could be understood as a possible window into cognitive-emotional processing as it can instantiate both communication and thought. Language has been observed to rely on both cognitive and emotional processing; where, for instance, emotional processing networks in the brain can support the interpretation of abstract meanings or emotional states of listeners can affect meaning interpretation in general (Kousta et al., 2011; Pauligk et al., 2019; Pinheiro et al., 2013; Van Berkum et al., 2013; Verhees et al., 2015; Vigliocco et al., 2014). Therefore, if emotion and cognition are not clearly separable and language is rooted in them, then the basic processing of language should be affected by anything that affects cognitive-emotional processing. Thus, individual differences such as cognitive-emotional disorders or traits, should induce effects on language processing. The present thesis is the result of a research project intended to tap into this interaction between language and individual differences. To narrow down observations to a more precise object of study, the main focus is placed on emotional speech and anxiety. Emotional speech, because language can convey complex emotional meaning through different channels simultaneously, which can elicit cognitive-emotional responses with clear temporal and anatomical patterns (Kotz and Paulmann, 2011). Anxiety, because it grasps cognitive-emotional aspects in a very straightforward manner, with clear neurophysiological mechanisms of action (Robinson et al., 2019). More precisely, the present project focuses on how threatening speech processing, conveyed prosodically or semantically, is affected by trait anxiety.

This narrowing down is important, as stimuli conveying different emotions can elicit different responses. For instance, clear double dissociations have been observed for

the processing of fear and anxiety, processed by partially overlapping but distinct neurophysiological mechanisms (Gray and McNaughton, 2000; Robinson, 2019); even more so for positive affective states such as happiness/pleasure (Kringelbach and Berridge, 2009). This fact is sometimes overlooked by research on language or acoustics, which usually test emotional variation across emotions rather than variations within a single emotional expression (e.g. threat) and its effects on intrinsic affect (e.g. trait anxiety). For instance, it has been observed that syntax processing can be impaired by induced negative mood (Van Berkum et el., 2013; Verhees et al., 2015) or by disorders such as bipolar disorder or schizophrenia (Lee et al., 2016); which emphasises the relevance of studying specific effects of threatening speech variation of information channels (e.g. semantics, prosody) on anxiety. On the other hand, research on anxiety tends to assume that all stimuli provide a similar basis of stimulation, independent of their modality. This overlooks the fact that emotional language can be conveyed through different information types (e.g. prosody, semantics), which are not processes in the same way (Kotz and Paulmann, 2011) and may be differently affected by anxiety (e.g. Pell et al., 2015). Therefore, the present approach can bring a more nuanced description and explanation of specific effects of trait anxiety on processing of threatening speech; from which less coarse generalizations could be generated.

A crucial link between these domains is related to general mechanism of attention control. It is generally understood that early attention (up to 200ms) to emotional stimuli involves right lateralised ventral portions of prefrontal cortex, while more integrative and evaluative processes would involve more posterior left lateral cortex (Corbetta et al., 2008; Vuilleumier and Driver, 2007). Greater right prefrontal activity has been associated with anxiety and behavioural inhibition (Gable et al., 2017). Furthermore, right and left lateralised activity have been associated with the arousal and apprehension components of anxiety respectively, where the first is associated with over-attention to threat and the latter to over-engagement with threat (Spielberg et al., 2013). This matches models of attentional biases in anxiety, where earlier phases are related with detection and recognition of stimuli, while later phases are related with evaluation and orientation (Bar-Haim et al., 2007). Models of emotional language, such as the multistep model of emotional language processing (Kotz and Paulmann, 2011), already reveal this match-up between lateralisation and processing phases, but they can also be indicative of possible differences regarding stimulus type. Although speech processing is generally bilateral

(Hickok and Poeppel, 2007), prosody processing is privileged by the right hemisphere while semantic processing is privileged by the left (Poeppel et al., 2008; Zatorre et al., 2002). In the case of emotional language, this can be highly dependent on processing phase, involving distinct limbic and neocortical structures at earlier or later processing windows (Kotz and Paulmann, 2011). This is very important, as prosodic and semantic threat could be affected by anxiety in different ways, depending on both lateralisation patterns and processing phase.

The following chapters will present a theoretical background and supporting evidence used to construct a model of anxious processing of threatening language. Chapter 2 will introduce a literature review of current theory and models focusing on the time-course and anatomical networks of emotional speech and anxiety processing. Chapter 3 will introduce the statistical methods for subsequent chapters, detailing modelling decisions and model structures. Chapter 4 focuses on assessing experimental stimuli and will provide evidence for prosodic and semantic characteristics of threatening speech production. Chapter 5 details a study containing two dichotic listening experiments aimed at understanding possible lateralisation and phasic differences of threatening speech processing given trait anxiety levels. Chapter 6 will provide an in-principle replication of the second experiment presented in the previous chapter, focusing on a non-dichotic task to observe whether similar effects are observed in a different task. Chapter 7 details an EEG experiment, using the event-related potential technique, providing evidence for late phase effects of anxiety on speech processing. Chapter 8 discusses the previously presented results, revising the proposed operative model, and proposes a model for the processing of threatening speech in anxiety.

Chapter 2
Literature Review
Threatening Voices and the Anxious

2.1 Threatening Speech

Before exploring the characteristics of threatening speech, it is important to clarify the features of general speech processing. The current prevalent neuroscientific model of speech processing is the dual stream model, which in terms of language perception/processing involves two main brain networks termed the ventral and dorsal streams (Kemmerer, 2015). While the dorsal stream, involving areas such as inferior frontal gyrus (IFG) and parietal-temporal cortex (with a left hemisphere bias), is mainly involved in speech production and articulation; the ventral stream, involving areas such as bilateral inferior, middle and superior temporal cortex, mainly participates in speech comprehension (Hickok and Poeppel, 2007). As the present project focuses on speech comprehension/processing, features of the ventral stream are particularly relevant. The bilateral involvement of areas in this network has to do with differences in the information properties of the processed stimulus. The key difference is processing rate, where tonal information, processed at a slower frequency rate, would be processed mainly by the right hemisphere (RH), but the left hemisphere (LH) would privilege fast rate segmental information (Poeppel et al., 2008; Zatorre and Gandour, 2007). This implies that speech information that depends upon spectral variation (e.g. pitch or voice quality increases/decreases), would be mainly processed by RH structures, such as right temporal cortex (Belin et al., 2004; Ethofer et al., 2011). Otherwise, faster rate information, such as phonological variation, would be prioritised by areas (e.g. temporal cortex) at the LH (Scott and Johnsrude, 2003; Poeppel, 2014). Note that

As words (i.e. lexical items) are composed of phonemes (i.e. segmental features of speech), semantic information conveyed through words is dependent on segmental features of speech (Kemmerer, 2015). Hence, according to the presented speech processing model, there should be an increased involvement of LH networks during semantic comprehension, as LH language networks tend to privilege fast rate information processing. Otherwise spectral elements of emotional prosody, processed at a slower rate, should be prioritised by RH. Crucially, the dorsality and laterality differences for

acoustic processing not only vary with respect to information type, but also depend on processing time (in particular for emotional sounds), with bilateral support from other brain structures in different time windows (Früholz et al., 2016a). Therefore, speech comprehension can show both laterality and temporal differences depending on whether meaning is conveyed through different information types, such as prosody or semantics.

Prosody is understood as the variation and combination of suprasegmental features such as duration, rhythm, intensity or pitch (Sidtis and Van Lancker Sidtis, 2003); combination and variation which can be intended for the communication of emotion (Banse and Scherer, 1996; Lausen and Hammerschmidt, 2020; Sidtis and Van Lancker Sidtis, 2003). Thus, the comprehension of prosody might depend upon contrasts between and presence/absence of several features, which have been adjusted by the speaker to convey a specific emotion (e.g. anger). This 'tuning up' of sorts can imply the increase of certain features and the decrease of others which together will imbue the stimulus (i.e. sentences) with emotional meaning. For instance, it has been observed that emotional meanings as different as joy and rage are similar in having higher pitch, and mainly differ in being less intense and having less energy at high frequency ranges (Hammerschmidt and Jürgens, 2007). For the case of threatening speech, this would imply that voice quality measures focusing on frequency range differences, such as Hammarberg index which measures differences between 0-2000hz and 2000-5000hz ranges (Banse and Scherer, 1996), should be lower for raging or threatening sentences; while average or median fundamental frequency (pitch) should be higher. Whether increases in median pitch and decreases in Hammarberg index (voice quality) are sufficient to explain threatening/aggressive voices is a relevant empirical question.

Semantic meaning can be understood, linguistically, as the meaning derived from lexical items or explicit sentence context (Kemmerer, 2015). Hence, if a stimulus has no meaningful variations in prosody, it could convey emotionality through its content, as the slow rate phonemic articulations can combine to produce a lexical item (or similar), but not prosodic meaning (at a slower rate). Hence, if the focus is on threatening semantics, threat could be characterised solely by lexical content. Thus, standardised norm of words' meanings, such as the Affective Norms for English Words (ANEW) (Warriner et al., 2013), could provide a good mechanism for quantifying the threat level of lexical items (semantics). For instance, as indicated by ANEW's arousal and valence scores, sentences containing lexical items higher in arousal and lower in valence could be understood as

angrier and more threatening. Consistent with this, it has been observed that words conveying pain (Borelli et al., 2018) or threat (Ho et al., 2015) show high arousal and low valence values. Although sentences not containing emotional lexical items can convey emotional meaning (Lai et al., 2015), if emotionally loaded items are present, these will normally drive the emotional meaning of the sentence (e.g. Paulmann and Kotz, 2008; Yan and Sommer, 2018). Given this, sentences with threatening semantics characterised as high in arousal and low in valence (via their lexical items) could provide a complimentary empirical question to that asking whether threatening prosody can be mainly characterised as high in median pitch and low in Hammarberg index (voice quality).

Although arousal and valence at the lexical level are not necessarily the only features composing semantic threat, or pitch and voice quality are not necessarily the only features composing prosodic threat, they can be sufficient to quantitively characterise stimuli that are able to be felt and comprehended as threatening. This should provide two comparable bidimensional maps of threatening semantics (arousal and valence dimensions) and prosody (pitch and voice quality dimensions). Also, this approach allows stimuli to be concretely categorised, which facilitates making inferences about emotional/cognitive responses (e.g. rumination/worry) elicited by these stimuli. Regarding prosody, some electroencephalography (EEG) studies show mid-phase event-related potentials (ERPs) (~300ms) and theta (4-12Hz) band activity modulated by variations of acoustic measures of angry prosody such as intensity (Chen et al., 2012). Previous research (Chen et al., 2011) indicates that late phase processing (over 400ms) is highly relevant for assessment processes (e.g. integration), when evaluating angry prosody characterised by mean pitch and intensity. Using similar stimuli, more recent research indicates earlier phase theta (100-600ms), but also later phase beta (12-30Hz) modulation by angry prosody (Chen et al., 2014). Similarly, late activity (400-600ms) as a left-lateralised late positive potential (LPC), has been recently observed for angry versus happy vocalisations (also characterised by pitch and intensity), but was interpreted as enhanced attention and evaluation (Burra et al., 2018). Regarding semantics, some EEG studies directly addressed lexical meaning as characterised by arousal and valence (Citron, 2012). A commonly observed pattern is the modulation of LCP by valence and arousal variations (Citron et al., 2013; Delaney-Busch et al., 2016; Imbir et al., 2017; Zhang et al., 2014). Although these studies did not characterise

arousal/valence variation in terms of specific emotions (e.g. anger), they do indicate that lexical meaning properties, such as valence and arousal norms, can modulate responses in a similar way as prosody-related acoustic properties. In the present thesis it is proposed that the homologation of semantic (arousal and valence) and prosodic (pitch, voice quality/roughness) features can be important for comparing emotional (threatening) stimuli and understanding their effects on listeners, as it has been proposed and observed that aggressive or angry voices can be associated with increase in arousal and pitch, and decreases in valence and voice quality (Patel et al., 2011).

Regarding the anatomical aspect of angry prosody some functional magnetic resonance imaging (fMRI) studies indicate the role of various structures, such as the superior temporal cortex (STC), middle frontal gyrus (MFG), basal ganglia (BG), amygdala, IFC and operculum, when participants listen angry vocalisations, syllables or words (Ceravolo et al., 2016; Früholz et al., 2015; Grandjean et al., 2005; Mothes-Lasch et al., 2016). This matches recent studies using functional near infrared spectroscopy (fNIRS) and EEG, testing angry, happy and neutral vocalisations, which show consistency between ERP waves and P2 and LPC and activity in areas such as medial prefrontal cortex (mPFC), IFC and STC (Steber et al., 2020). Consistent with recent magnetoencephalography (MEG) research (Styliadis et al., 2018), previous literature on emotional words (for a review see: Citron, 2012) indicate similar patterns of mPFC, amygdala and additional cingulate cortex involvement; which have been observed in some fMRI studies as well (e.g. Wittfoth et al., 2009). However, most of these studies either focus on language congruency issues or on attentional effects of prosody, and they use different and sometimes not detailed measures for anger in prosody and semantics. Thus, these approaches can give precise information about the speech-related aspects of angry prosody, but not much about differences between semantics and prosody in terms of emotional processing. In view of this, current models of emotional language processing are based mainly on studies assessing emotional recognition or categorisation (for reviews see: Kotz and Paulmann, 2011; Liebenthal et al., 2016). Therefore, the models will tell about the involvement of brain areas and processing phases in regard to responses to a myriad of emotions and how they are recognised and told apart, which makes difficult to assess aspects of emotional communication/elicitation. Even so, the multistep model emotional language (Kotz and Paulmann, 2011) clearly specifies processing phases by combining EEG and fMRI evidence in terms of clearly defined

sources of information, such as written and language, and speech's prosody and semantics.

The model proposes three crucial stages: 1) Early phase processing (∼100ms), comprising sensory analysis, and involving bilateral primary and secondary auditory cortices (STC, BA41, and related structures). 2) Mid phase processing (∼200-300ms), comprising recognition and incipient meaning interpretation, and involving structures such as amygdala and hippocampus. 3) Late phase processing (∼400ms and over), comprising integration and evaluation, and involving bilateral inferior frontal cortex and orbital cortex. More recent models of emotional audition (Früholz et al., 2016a), emphasising the roles of music and prosody, have pointed out a more comprising and defined network involving: mPFC, inferior frontal cortex (IFC), insula, ofbitofrontal cortex (OFC), amygdala, hippocampus, STC, BG and cerebellum. However, phasic processing and specific interaction within the network at such possible phases are not specified. Each area within this network is very broadly defined, which makes difficult to disentangle which type of connectivity tightens them together and what neurochemical mechanisms come into play. Another important gap to be filled is a clear definition of stimuli, such as finding semantic/prosody features which are sufficient to characterise emotional elicitation, such as threat increasing as a function of arousal/valence and pitch/roughness. This focus on more specific emotions and emotional/cognitive elicitation can be crucial to understand how specific emotional/cognitive states or traits, such as anxiety, can affect speech or language processing. This relationship between trait emotional/cognitive features (i.e. anxiety) can help to better identify the involvement of more specific neuroanatomical networks but also the processing time-course of emotional language.

2.2 Trait Anxiety

Anxiety is a process characterised by the tight integration between emotion and cognition. As an adaptive and defensive reaction, common to many animal species, anxiety is understood as a motivation or response to potential or signalled threat (Blanchard et al., 2008; Gray and McNaughton, 2000; Robinson et al., 2019). When threat is not directly actualised, ambiguous or indirect, the response is anxiety as characterised by an emotional response of pervasive unpleasantness and accompanied by a cognitive response of focused thought on threat (McNaughton and Zangrossi, 2008). This could also

be understood as a simultaneous feeling of concern, angst or apprehension, and persistent assessment and re-assessment of risk. Crucially, current models of anxiety (Calhoon and Tye, 2015; McNaughton, 2011; Robinson et al., 2019) emphasise the difference between anxiety and fear. As opposed to anxiety, fear is a defensive response to direct threat or ongoing aggression; which will imply both anatomical and neurotransmission differences (LeDoux and Pine, 2016; McNaughton, 2011; Robinson et al., 2019). This can be understood as a continuum, where serotonin (5HT) and noradrenalin modulate responses to shorter defensive distance (i.e. close proximity threat) to larger defensive distance (i.e. potential threat) through widespread routes (McNaughton and Corr, 2004; McNaughton, 2011). For instance, networks comprising areas such as periaqueductal gray, medial hypothalamus and portions of amygdala mediate short defensive distance and panic responses. When defensive distance increases, areas such as amygdala, hippocampus, cingulate and dorsal and ventral prefrontal cortex mediate anxiety responses.

Current neuroimaging-based models of fear emphasise the role of additional areas in a phasic way (Fullana et al., 2015). Consistent with the aforementioned, physiochemically-based models, areas such as hypothalamus are strongly involved in earlier phases, but accompanied by anterior insula and somatosensory cortex. Later phases involve mPFC and anterior cingulate cortex (ACC). Nevertheless, these could be better explained by anxiety rather than fear, as the continuous exposition to threat in neuroimaging studies could activate anxious responses as anticipation to threat. These have been proposed to be associated with increased levels of anxious arousal as a result of continued anxiety (McNaughton and Grey, 2000). All these features have been taken into account in current models of anxiety (Calhoon and Tye, 2015; Robinson el al., 2019), where the coordinated activity of the basolateral amygdala (BLA), the bed nucleus of the stria terminalis (BNST), the ventral hippocampus (HC) and the mPFC, signalled by phasic theta band (4-12Hz) activity will be the core of anxiety processing. However, these regions are supported by dorsolateral prefrontal cortex (dlPFC), which exercises excitatory control over mPFC (in its paralimbic portion) together with the insula, while mPFC (from its infralimbic or ventral portion) exercises inhibitory control over BLA which communicates with BNST.

Phasic processing of responses to threat has been previously explored by multiphasic models of anxiety, derived from a metanalysis and review of behavioural

evidence (Bar-Haim et al., 2007). The model considers four phases: 1) Pre-attentive evaluation of threat, that is the process before awareness that allows to quickly identify threatening stimuli. 2) Resource allocation, namely aware evaluation of threat to stop or continue ongoing activity. 3) Guided evaluation of threat, which involves threat assessment respect to context and as compared with previous experience. 4) Goal engagement, involving the assessment of threat before decisions oriented to pursue or interrupt current goals. In line with other models (Eysenck et al., 2007; Cisler and Koster, 2010), the core feature of these phases responds to attention control. These models, though underspecified in neuroanatomical and physiological terms, relate to emotion and attention regulation models (Corbetta et al., 2008), which propose adrenergic regulation from locus coeruleus of two networks. First, a right-lateralised ventral network, associated with attention disruption/control, composed by temporoparietal junction (TPJ), middle frontal gyrus (MFG), IFG, frontal operculum and insula. Second, a dorsal network, associated with stimuli integration/evaluation, which comprises intraparietal sulcus (IPS), superior parietal lobule (IPL) and frontal eye field (FEF). Similar models propose early (up to ~100ms), mid early (~200ms), and unclearly specified later phases; integrating structures such as orbitofrontal cortex (OFC), amygdala and hippocampus for supporting earlier attention and later emotional processing in interaction with attention networks (Vuilleumier and Driver, 2007).

More recent models have tried to overcome this issue by focusing on ERP literature (Gupta et al., 2019), proposing early (~50-300ms) over-attention biases and later (over 300ms) over-engagement biases. Although there is a great effort in distinguishing phasic effects as signalled by a number of ERP components, the model conflates fear with anxiety, which makes difficult to disentangle what specific biases might induce each proposed stage. Therefore, for understanding the possible phasic processing of anxious responses to threat it is indispensable to consider emotional attention models (e.g. Corbetta et al., 2008) in the context of models of anxiety (e.g. Robinson, 2019). This would imply that anxious biases should respond mainly to potential or indirect threat, or to the continuous exposure to threatening stimuli. Hence, the effects of anxiety should be more evident at mid or late processing phases, when evaluation and deliberation processes have taken place. This derives from the previously discussed notion of anxiety as an overactive behavioural inhibition system (BIS), implying that over-engagement with threat is prevalent, also indicated by anxious

rumination/worry usually associated with repetitive thinking (McEvoy et al., 2010; McLaughlin et al., 2007). According to this, past oriented (rumination) or future oriented (worry) repetitive thoughts are characterised by increased imageability and by a verbal component. Both can be cognitive components of anxiety, but the latter is specifically associated with inner speech. This is particularly relevant, as inner speech has been shown to work in a cyclic manner, through the phonological loop (Buchsbaum and D'Esposito, 2008; Vigliocco and Hartsuiker, 2002). This model is supported by a wide variety of evidence, and indicates that inner speech activates mainly areas in the language dorsal stream, such as left supramarginal gyrus (SMG) IFG and STC (Geva, 2018). Accordingly, there could be a tight relationship between language and anxiety, particularly when verbal repetitive thinking is elicited.

Having said this, it is relevant to point out that people with increased anxiety and BIS, tend to show greater right lateral frontal involvement respect to people with lower anxiety or BIS levels (Gable et al., 2017). Some models show that this should be a pattern associated with anxious arousal, while anxious apprehension (e.g. worry) should show a left-lateral pattern (Heller et al., 1997). However, further results show a bilateral pattern for worry-related anxiety (Nitschke et al., 1999; Spielberg et al., 2013). The regions observed in this way, tend to match with both attention and language networks, which is consistent with a broader notion of anxious arousal positing a cyclic pattern of physiological hyperarousal as increased by sustained BIS (McNaughton and Gray, 2000). This process has also been proposed to be associated with theta band modulation and right frontal activity (Calhoon and Tye, 2015; McNaughton et al., 2013; Neo et al., 2011). This neatly connects cognitive patterns of sustained anxiety (BIS), characterised by verbal repetitive thinking, with the emotional patterns characterised by amygdala-hippocampal activity and right-frontal regulation. This could also suggest that anxious responses to language stimuli might not be equivalent to other types of stimuli. For instance, verbal threat could be especially impaired by intrinsic predispositions to verbal repetitive thinking, as this would engage similar networks, possibly inducing saturation or interference effects.

For instance, some behavioural studies have shown that socially anxious people will show slower reaction times if enough time is allowed for them to over-engage with angry prosody (Peschard et al., 2016; Tseng et al., 2017). This slow-down has been observed for higher BIS as well, together with fMRI measures indicating a directly

proportional relationship between BIS level and prefrontal cortex activation (Sander et al., 2005). However, different predispositions to higher anxiety (e.g. social anxiety versus trait anxiety) may reflect different patterns when more fine-grained measures are considered (Schulz et al., 2013). Indeed, EEG studies on the semantic domain, researching the effects of abusive words on social anxiety, indicate early EPN and later N400 modulation by social anxiety (Wabnitz et al., 2015); and studies exploring the effects of vocalisations and speech prosody on trait anxiety have shown earlier P2 and later LPC modulation by anxiety (Pell et al., 2015). Earlier studies have found a relationship between mismatch negativity (MMN) ERPs, elicited by deviant angry syllables, physiological hyperarousal and state anxiety (Schirmer and Escoffier, 2010). These studies, however, demonstrate that when stimuli are short and decisions need to be taken quickly, anxiety modulates early pre-attentive or attentive responses, where later processing phases are affected only because of this earlier processing. Even so, several questions are still opened, such as whether the use of semi-naturalistic longer duration stimuli (e.g. sentences) and tasks which allow delayed responses would provide greater opportunity for over-engagement with threatening stimuli thus showing different electrophysiological/anatomical signatures. Another important question is whether behavioural and electrophysiological responses would be equivalent if anxious over-engagement with threat is present. That is, trait anxiety associated with increased BIS might induce both early and late effects on speech processing, which would require re-assessing or extending current models of emotional language processing.

2.3 Operative Model

An important issue with models of emotional language or emotional auditory processing is that they attempt to model responses to all types of emotional language through a single model. However, as portrayed on the first two subsections above, even within the sphere of defensive emotions, the supporting neurological networks and phases of interaction differ between fear and anxiety (McNaughton, 2011). Then, it could be expected that emotions within the appetitive range differ even more from defensive emotions. Similarly, models of anxiety tend to portray responses to threat as if all stimuli are equivalent, overlooking their physical and biological properties (e.g. the acoustics of threatening language). Again, as reviewed above, the features of different informational

properties of language, such as prosody and semantics, can differ in important ways, leading to different processing modes (Kotz and Paulmann, 2011).

Given all this, the present operative model attempts to be a portrayal of anxiety effects (e.g. predisposition to worry or overactive BIS) on threatening speech and a very specific one at that. In other words, there is a departure from emotional type categorisation tasks as a form of assessing emotional processing in general towards a focus on single emotion/cognition stimulation (i.e. threat) and single emotion/cognition elicitation (i.e. anxiety). It is implicit in the present model that signalling direct threat through language is extremely difficult, even more so in tasks of continued expositions to meaningful threatening stimuli, as no threat is actualised to condition the stimulus (e.g. a painful unconditioned stimulus). This is the reason fear conditioning tasks can tap into anxiety processes only if long lapses are placed between unconditioned and conditioned stimulus (Robinson et al., 2019). Semi-naturalistic sentences (or longer stimuli) together with delayed responses to these stimuli could be effective in eliciting anxious behavioural and/or electrophysiological responses. Moreover, when stimuli are speech utterances, their acoustic features might not be irrelevant. The different phasic lateralisation patterns of speech processing might overlap with anxiety processing, being affected in different ways depending on prosody or semantics. Bearing this in mind, the present operative model, specified on Figure 2.1, attempts to integrate emotional language and anxiety processing models. Although the operative model resulting from this integration predicts the involvement of specific anatomical networks, the present focus will be on the time-course of the processing of threatening speech given trait anxiety. Relevantly, this extends the model by including a new fourth phase, where stimuli are re-appraised and/or rehearsed to orient responses (deliberation phase).

Sensorial Phase Pre-attentive Around 100ms	Recognition Phase Aware Around 200ms	Evaluation Phase Integrative Around 300-500ms	Deliberation Phase Cogitant Around 600ms +
Possible ERP components: P1, N1, MMN	Possible ERP components: P2, N2, EPN	Possible ERP components: P300, N400	Possible ERP components: LCP, P600
Possible Network: bilateral auditory sensory cortex, superior temporal cortex (STC), basolateral amygdala (BLA)	Possible Network: extended amygdala, central amygdala, bed nucleus of stria terminalis (BNST), right medial prefrontal cortex (mPFC)	Possible Network: septum, ventral hippocampus (vHC), dorsolateral prefrontal cortex (dlPFC), temporal pole, basal ganglia (BG), insula	Possible Network: middle temporal cortex (MTC), superior temporal cortex (STC), basal ganglia (BG), IFG, hippocampus, amygdala

Figure 2.1. Operative model of phasic processing of threatening speech and anxiety.

The four phases of the operative model can be understood as follows: 1) An auditory language stimulus elicits processing in the ventral stream while simultaneously activating re-orienting attention networks modulated by anxiety; this corresponds to sensory processing and happens pre-attentively. 2) The listener becomes aware of the stimulus and recognises it as possibly threatening, enacting the initial steps of anxious responses if threat is not direct or imminent. 3) The stimulus is assessed and its potential threat is integrated with context and past experience, cognitive-emotional anxious processing is enacted. 4) In the absence of time-pressure the stimulus is re-assessed, overactive BIS over-engages with the stimulus and verbal repetitive thinking is elicited (excessive cogitation). This implies that in early and early-mid phases ($\sim$0 to 300ms) over-attention to threat could take place, and anxious people would re-orient attention to threatening stimuli, initialising BIS if no flee-freeze-fight (FFF) response is required. From mid-late to late phases (from $\sim$300-400ms), as threatening speech is evaluated in terms of previously processed prosody and semantics, the BIS response starts. At late phases BIS will be sustained if the stimulus has generated over-engagement with threat, which will involve either rehearsing the stimulus through the phonological loop, or using inner speech to think over past-oriented comparisons or future consequences regarding the stimulus (i.e. rumination/worry). In the experimental context, this could imply comparing the stimulus with concurrent activity and/or pondering upcoming responses.

At the behavioural level, dichotic listening studies have observed no reaction time slow-down and LH accuracy dominance in an emotional categorisation task (Leshem, 2018). Contrary to this, dichotic listening studies normally show a decreased right ear advantage (i.e. LH poorer performance) (Gadea et al., 2011), and slower reaction times (Peschard et al., 2017). Under the current model, this discrepancy is not hard to explain, as the emphasis on categorisation tasks, using single words of varied emotional meaning, would imply a semantic emphasis (i.e. LH privilege at early stages), as opposed to emotional prosody inducing greater RH involvement (Grimshaw et al., 2003; 2009). In addition, no exposure to threatening stimuli and no sufficient delay in responses would not induce a reaction slow-down. In short, reaction time slow-down and hemispheric differences are not absolute processes, but depend on processing phase and type of tasks and type of stimuli. Therefore, different tasks need to be directly compared; in such a way that differences between fast (during stimulus) and delayed (after stimulus) responses can be tested. Furthermore, electrophysiological evidence comparing different moments

of processing, including onset-aligned versus response-aligned ERPs are important veins of exploration to elucidate predictions based on the presented operative model.

The process described by this model in its current state, as noted by literature from both language and anxiety, could be cyclic. For instance, the observed elicitation of theta band activity by angry prosody (Chen et al., 2012; 2014) could correspond to the classic theta activity observed in anxious responses to potential threat, possibly signalling activity in the BLA-vHC-mPFC 5HT route (Calhoon and Tye, 2015). Notably, these language experiments have observed this theta activity to take place over 400ms and similar studies have observed LPC modulations from this time on (Chen et al., 2011; Burra et al., 2018; Steber et al., 2020). Important to emphasise, these studies use sentences or tasks with long inter stimulus intervals. When either semantic or prosody stimuli are short, and responses are not delayed, experiments tend to show that social or trait anxiety induces increases in amplitude on early ERPs (i.e. P1), but not on late-phase ERPs (i.e. P600) (Pell et al., 2015; Wabnitz et al., 2015). So, with enough time to deliberate (over-engage with threat), anxious people should present greater late phase amplitudes. Stated differently, anxiety could disrupt these via over-engagement with threat, inducing excessive re-appraisal suggested to be signalled by LPC-like ERPs (Hajcak et al., 2010).

Therefore, it is possible to experimentally test the relationship between behavioural and EEG measurements, which could provide evidence of the effects of anxiety through the whole processing cycle, from stimulus input to behavioural output. In the present thesis, the focus is on possible laterality differences across the processing time-course (i.e. the four proposed phases). The present project will attempt to observe these differences via behavioural (i.e. dichotic listening) and ERP measurements. These measurements will allow to test differences in the time-course of threatening speech processing and how it is affected by trait anxiety. Through this theoretically motivated experimental approach, some relevant questions can be answered. First, whether threatening language can be sufficiently characterised by specific semantic and prosodic features oriented to emotional/cognitive elicitation. Second, whether both early over-attention and later over-engagement depend on trait anxiety (i.e.. BIS) increases. Third, whether these effects, if present, have different laterality patterns given semantic and prosodic threat. The upcoming chapters will provide a methodological framework, norming procedures and behavioural and ERP evidence aimed at scrutinising the previously presented operative model.

Chapter 3
Statistical Methods
Bayesian Hierarchical Models for Experimental Analysis

3.1 Bayesian Approach

Before assessing the operative model presented in the previous chapter through experiment, it is important to clarify how this evidence will be assessed in statistical terms. Choosing a statistical approach for scientific inference is not necessarily a trivial matter, as data can follow different distributions and experimental design will render data organised in different levels, and standard/predefined tests will not necessarily adequate to these particularities (Kruschke, 2015). However, most if not all the previously presented research does not explicitly model their statistical analyses. As common in life sciences, this implies the use of a 'predefined test' approach, where researchers pick a model from a conventional list to input their data and perform null hypothesis significance testing (NHST), overlooking the hidden assumptions of the model and the fact that not all hypotheses have a null counterpart (McElreath, 2020). The aim of the present project is to explicitly and transparently model all assumptions from the scientific operative model into statistical models. The aim of this chapter is to provide a very brief overview of the Bayesian approach to probabilities and to explain how this will be applied to the design of hierarchical models. These models are designed to deal with specific inferential and predictive problems derived from the hypotheses at hand, namely with the inferential and predictive outcome posited by the current operative model. This involves the effects of different variables from experiments designed to test laterality, information type (semantic, prosody) and anxiety effects on threatening speech processing. These include dichotic listening and EEG experiments plus psychometric measures of anxiety, using stimuli with clearly defined acoustic and lexical threatening properties.

Modelling data from such experiments has both practical and philosophical implications which cannot be overlooked, as they have a direct impact on inference, prediction and interpretation. For instance, it has been proposed that current statistical modelling techniques, such as the use of hierarchical models are a good approach for dealing with issues of balance between bias and variance in observation and

experimental settings as opposed to traditional non-hierarchical modelling (Alday, 2018; West et al., 2014). This is relevant, because the advantage of a hierarchical model (aka multilevel, mixed-effects) lies on its ability to un-pool data in terms of multiple levels within the data structure. As opposed to completely pooling the data, thus assuming a single overarching variance, or to completely un-pooling the data, thus assuming several unrelated variances, a hierarchical model treats data as arising from different interdependent levels and thus avoids underfitting or overfitting problems (Gelman, 2006). These levels are not arbitrary, and they are defined by how data was collected, organised or originated. These levels are also placed in the context of variables, which can be included into the model as covariates to account for possible relationships between them. However, improving inference by the implementation of additional covariates requires very well-defined assumptions with a strong theoretical foundation about model's variables and parameters (Gelman, 2006; Gelman et al., 2014; Imbens and Rubin, 1997; McElreath, 2020; Sassenhagen and Alday, 2016; West et al., 2014). According to this, to avoid introducing spurious or nuisance relationships, or to block relationships between variables, variables should be included into a model in terms of their theoretically/hypothetically proposed relationships.

Besides the relevance of how a model is organised (i.e. what levels and variables to include), it is important to consider where the model's structure comes from. In other words, identifying the mathematical structure of a model and determining its relevance. A common problem in frequentist statistics is that it tends to rely on fixed pre-defined (assumed) distributions for data, and on an assumed space of probabilities (i.e. an imaginary population) where parameters depend on (Kruschke, 2015). As widely discussed in the past, this has led to serious issues in the application of statistical tests and their interpretation, in particular for NHST approaches (e.g. Cohen, 1994). The most important ones for present purposes are: 1) NHST depends on point estimates which are compared to an imaginary population distribution, so it is very sensitive to sample size and stopping rules (Kruschke, 2015). 2) NHST models are usually organised as predefined tests, which may lead to wrong assumptions on how and when to use them, including their use for testing hypotheses that do not have null counterparts (McElreath, 2020). These two problems could easily lead to bias and variance becoming detrimental, leading to false positives or negatives; while instead bias and variance should be expressions of the natural uncertainty of any measurement (Kruschke, 2015; McElreath,

2020). Instead of fighting against these phenomena, a Bayesian approach can help to understand this uncertainty from the data in terms of probabilities. This results from two features of Bayesian modelling. 1) The selection of prior probabilities for model's parameters, which allows great modelling flexibility resulting in the selection of adequate distributions for hypotheses and data at hand; and 2) the sampling of a whole posterior probability distributions form prior distributions and observed data in terms of Bayes rule, allowing to directly account for error an uncertainty (Gelman et al., 2014; Kruschke, 2015; Martin, 2018; McElreath, 2020).

With this in mind, it will be possible to avoid some common issues present in the reviewed literature (and the literature in general), which include: 1) Not using subjects and/or stimuli varying intercepts, which might require un-pooled treatment of their variance, especially if physiological or psychometric measures are involved, as individual variation in features like brain anatomy or personality can have a relevant impact on measurements or responses. 2) Non explicit or inappropriate use of statistical distributions for observed data, which can be detrimental to inference or prediction, reducing the reliability of the analysis. 3) Unjustified integration of variables into models (mainly linear models), inducing spurious effects or obscuring real effects (e.g. including both measures of trait and state anxiety by slopes in a model, which will obscure effects as trait anxiety can have a direct effect on state anxiety). 4) Not using or not justifying the use of varying slopes, in particular for variables with binary or categorical distributions, where variance may not be equivalent across conditions. Explicitly addressing these issues, the following subsection will explain how analyses were statistically modelled by using a Bayesian approach. The organisation of models will be explained in terms of the operative model detailed in Chapter 2 (section 2.3) and in terms of the experimental measurements at hand (Chapters 4, Chapter 5, and Chapter 7); which include accuracy, reaction times (in ms), amplitude (in μV), and experimental variables such as ear presentation, sentences' type, and trait anxiety level.

3.2 Statistical Modelling

An important aspect of a Bayesian approach is the use of conditional probabilities in terms of prior information. Bayes' rule is based on deriving the probability of a parameter that is not directly observed (A) given the probability of an observed parameter (B), which can be expressed as the classic Bayes' rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.2.1}$$

Where $P(A|B)$ is the posterior probability, $P(B|A)$ is the likelihood, $P(A)$ is the prior probability, and $P(B)$ is the marginal likelihood. A fundamental way for adjusting the inferential capabilities of a statistical model in an appropriate way for the analysed data is by using a hierarchical structure (Gelman et al., 2014; Kruschke, 2015). This simply implies new parameters are added in a chain of dependencies respect to A and B, such as:

$$
\begin{aligned}
P(A, c|B) &= \frac{P(B|A, c)P(A, c)}{P(B)} \\
&= \frac{P(B|A)\,P(B|c)P(c)}{P(B)}
\end{aligned}
\tag{3.2.2}
$$

Where c is a parameter associated to a feature, group or similar that is associated to the observed parameter making the parameter A dependent on c. For instance, in the case of accuracy observed data, involving correct (hit = 1) or incorrect responses (false alarm/miss = 0), that is Bernoulli trials, the hierarchical structure can become exceedingly complex.

Considering Chapter 2's operative model (Figure 2.1), different information types (prosody or semantic) could have different effects on attention. Behavioural experiments presented in Chapters 5, 6 and 7 were explicitly designed to address the question of whether anxiety has different effects on behavioural measures given ear of presentation of threatening stimuli, which can be sentences with prosody-only (Prosody), semantic-only (Semantic) or both types of threat (Congruent). In statistical terms, this question is about the probability of answering correctly to stimuli given ear (left or right), sentence's type (Congruent, Prosody, Semantic), anxiety level (scores, treated as continuous), subjects (44 participants) and stimuli (240 sentences). Any of these variables could be parametrised in a hierarchical structure by assigning distributions (hyperpriors) to the parameters of their distributions (priors). For instance, a simplified linear model of this sort could involve a varying intercept for subjects and a distribution for observed data. A varying intercept, sometimes referred (misleadingly) as a random intercept, is simply the portion of the linear model that is not associated to a variable (intercept) but with

hyperpriors for its mean and standard deviation/precision if a Normal distribution is used. This allows the model to vary across levels, in this case each subject. Such a model can be expressed in the following way:

$$\mu_{subject} \sim Normal(\mu = 0, \sigma = 1) \tag{3.2.3}$$

$$\sigma_{subject} \sim HalfNormal(\sigma = 1)$$

$$\alpha_{subject\ j(1\ldots44)} \sim Normal(\mu = \mu_{subject}, \sigma = \sigma_{subject})$$

$$\bar{y}_i \sim Bernoulli(p = sigmoid(subject_j))$$

Where *Normal* is the Gaussian or *Normal* distribution and *HalfNormal* is a special case of a Normal distribution, but folded at the zero mean (i.e. restricted to positive values), *Bernoulli* is the Bernoulli distribution, *sigmoid* is the transformation of the value *subject* for the *Bernoulli* mean p to be expressed as a sigmoid function instead of a linear function, and y is the observed distribution. Normal priors are used as they represent weakly informative priors, a good balance between non-informative and informative priors in the absence of good prior knowledge about the model's distributions. They can also be understood, in machine learning terms, as providing L2 regularisation (regularising priors) (see: Gelman et al., 2014; McElreath, 2020).

$$Normal = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{i-\mu}{\sigma}\right)} \tag{3.2.4}$$

$$HalfNormal = \sqrt{\frac{2}{\pi\sigma^2} \exp\left(\frac{i^2}{2\sigma^2}\right)} \tag{3.2.5}$$

$$Bernoulli = p^i(1-p)^{1-i} \tag{3.2.6}$$

$$sigmoid(p) = \frac{1}{1 + e^{-p}} \tag{3.2.7}$$

It can be seen that building up this simplified model in terms of Bayes rule (3.2.2) is already complex if $P(p, subject|y) = P(y|p)P(y|subject)P(subject)/P(y)$, to the extent of analytic intractability if a more complete model is applied:

$$\alpha_0 \sim Normal(\mu = 0, \sigma = 1)$$

$$\mu_{sub} \sim Normal(\mu = 0, \sigma = 1) \tag{3.2.8}$$

$$\sigma_{sub} \sim HalfNormal(\sigma = 1)$$

$$\alpha_{subj(1\ldots44)} \sim Normal(\mu = \mu_1, \sigma = \sigma_1)$$

$$\mu_{sen} \sim Normal(\mu = 0, \sigma = 1)$$

$$\sigma_{sen\,k(1\ldots240)} \sim HalfNormal(\sigma = 1)$$

$$\alpha_{2sen} \sim Normal(\mu = \mu_2, \sigma = \sigma_2)$$

$$\beta_1 \sim Normal(\mu = 0, \sigma = 1)$$

$$\vdots$$

$$\beta_7 \sim Normal(\mu = 0, \sigma = 1)$$

$$p = \alpha_0 + \alpha_{1j} + \alpha_{2k} + \beta_1 Ear_{codes} + \beta_2 Type_{1,2codes} +$$
$$\beta_3 Worry + \beta_4 WorryEar_{codes} + \beta_5 EarType1,2_{codes} +$$
$$\beta_6 WorryType_{1,2codes} + \beta_7 WorryEarType_{1,2codes}$$

$$\bar{y}_i \sim Bernoulli(p = sigmoid(p))$$

Where α_0 is the main "fixed" (non-varying) intercept, the other two α parameters are the varying intercepts for subject (44) and sentence (240 stimuli), and all β parameters are the non-varying slopes for categorical variables coded as Ear (left = 1, right = 0), Type1 (Prosody = 1, Semantic = 0, Congruent = 0), Type2 (Prosody = 0, Semantic = 1, Congruent = 0), Worry (continuous), and their interactions. (All variables and exact numbers come from the first experiment presented in Chapter 5). This model permits testing of whether log-odds increase as a function of anxiety given type or ear parameters. Although there is valid criticism against using a multiplicative interaction (McElreath, 2020), the present decision seeks a trade-off between predictive accuracy and reducing parameter space in order to improve tractability. Even so, the possibility of using varying slopes across indexed variables remains a good alternative to the present model and could potentially improve the present approach; but this remains an issue of statistical analysis comparison, which is outside the scope of the present thesis.

The base structure of this model can remain constant for other types of data by replacing changing distributions for more appropriate ones for the new data. For instance, in the case of reaction time data would require a continuous distribution. Given the possibility of outliers, a robust regression (Gelman et al., 2014; Kruschke, 2015) could be used. Here, the observed distribution is modelled by using a Student's t-distribution, which has long tails able to capture outlying datapoints.

$$StudentT = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\lambda}{\pi\nu}\right)^{\frac{1}{2}} \left[1 + \frac{\lambda(i-\mu)^2}{\nu}\right]^{-\frac{\nu+1}{2}} \tag{3.2.9}$$

Where, Γ is the gamma function ($\Gamma(n) = (n-1)!$), ν are the degrees of freedom (normality parameter), and λ is a scale parameter which can also be expressed as σ (standard deviation). This might hinder the predictive capacity of the model, as reaction time data do not have negative values (McElreath, 2020). This trade-off, though sacrificing predictive capacity, could imply better convergence and inferential capacity (Gelman et al., 2014). Thus, the full reaction time model (Figure 5.1) can be expressed as:

$$\alpha_0 \sim Normal(\mu = 0, \sigma = 1)$$
$$\mu_{sub} \sim Normal(\mu = 0, \sigma = 1) \tag{3.2.10}$$
$$\sigma_{sub} \sim HalfNormal(\sigma = 1)$$
$$\alpha_{subj(1\ldots44)} \sim Normal(\mu = \mu_1, \sigma = \sigma_1)$$
$$\mu_{sen} \sim Normal(\mu = 0, \sigma = 1)$$
$$\sigma_{sen} \sim HalfNormal(\sigma = 1)$$
$$\alpha_{sen\,k(1\ldots240)} \sim Normal(\mu = \mu_2, \sigma = \sigma_2)$$
$$\beta_1 \sim Normal(\mu = 0, \sigma = 1)$$
$$\vdots$$
$$\beta_7 \sim Normal(\mu = 0, \sigma = 1)$$
$$\mu_{expected} = \alpha_0 + \alpha_{subj} + \alpha_{senk} + \beta_1 Ear_{codes} +$$
$$\beta_2 Type_{1,2codes} + \qquad \beta_3 Worry + \beta_4 WorryEar_{codes} +$$
$$\beta_5 EarType1,2_{codes} + \qquad \beta_6 WorryType_{1,2codes} +$$
$$\beta_7 WorryEarType_{1,2codes}$$
$$\sigma_{error} = HalfNormal(\mu = 0, \sigma = 10)$$
$$\nu_{d.f.} = Exponential\left(\lambda = \frac{1}{29}\right) + 1$$
$$\bar{y}_i \sim StudentT(\mu = \mu_{expected}, \sigma = \sigma_{error}, \nu = \nu_{d.f.})$$

Another advantage of implementing this robust regression model relates to EEG research. When implementing the event-related potential (ERP) technique, it is customary to extract mean amplitudes from pre-defined time-windows (Luck, 2014).

When averaged trial by trial, ERP amplitude measurements can contain extreme outliers, even after rejecting trials over certain amplitude thresholds (Luck, 2014). So, deciding whether these amplitudes represent or are influenced by brain activity or not is not easy. This provides a perfect case for robust regression, and helps to avoid arbitrary data rejection. Taking advantage of this, a more complex model is suitable (variables and exact values come for the EEG experiment presented in Chapter 7, see Figure 7.1). All interactions can be modelled as indexed interactions by using Ear by Type by Channel matrix B to model varying slopes (384 interaction points: 2 ears, 3 types of sentence, 64 electrodes) and a Channel shaped matrix A to model a fixed intercept pooling amplitude at each electrode (64 channels), also including matrices for varying intercepts for subject (S1) and sentence (S2).

$$\tau_{sub} \sim HalfNormal(\sigma = 1)$$
$$\theta_{sub\ j(1\ldots30)} \sim Normal(\mu = 0, \sigma = 1) \qquad (3.2.11)$$
$$\gamma_{sub} = Deterministic(\theta_{sub}\tau_{sub})$$
$$\tau_{sen} \sim HalfNormal(\sigma = 1)$$
$$\theta_{sen\ k(1\ldots162)} \sim Normal(\mu = 0, \sigma = 1)$$
$$\gamma_{sen} = Deterministic(\theta_{sen}\tau_{sen})$$
$$\alpha_{c(1\ldots64)} \sim Normal(\mu = 0, \sigma = 1)$$
$$\tau \sim HalfNormal(\sigma = 1)$$
$$\theta_{n(1\ldots384)} \sim Normal(\mu = 0, \sigma = 1)$$
$$\beta = Deterministic(\theta_n\tau)$$
$$\mu_{expected} = \gamma_{sub} \cdot S1 + \gamma_{sen} \cdot S2 + \alpha_c \cdot A + \beta \cdot BxBIS$$
$$\sigma_{error} = HalfNormal(\mu = 0, \sigma = 10)$$
$$\nu_{d.f.} = Exponential\left(\lambda = \frac{1}{29}\right) + 1$$
$$\bar{y}_i \sim StudentT(\mu = \mu_{expected}, \sigma = \sigma_{error}, \nu = \nu_{d.f.})$$

Note that subject and sentence varying effects are reparametrized as a deterministic function of the product between a half-normal (scale parameter, here denoted as τ) and a Normal distribution (the distribution of the reparametrized parameter, here denoted as θ), and the location parameter is extracted to the linear model itself as a non-varying ('fixed') intercept (i.e. channel intercept: $\alpha_{c(1\ldots64)}$). In this case, scale is the standard

deviation, and the external location acts as the mean (though neither mean nor standard deviation need to be present). This reparameterisation strategy allows for better convergence of intricated parameters (McElreath, 2020). Previously presented models were reparametrized in a more conventional way, by adding individual location parameters in the deterministic function as an added mean (e.g. $\alpha_{sub} = \mu + \sigma\mu_{sub}$, where μ is the location and σ is the sacle).

Amplitude data can be described at the electrode level as well, when informed from the main model about which parameters have sufficiently strong effects. In this case, this model could be pooling across sentence, subject, ear and/or type and use each sampling point (256 epoch + 26 baseline) as a varying intercept and slope, where T is the matrix containing 282 amplitude sampling time-points (from Chapter 7's experiment).

$$\tau_1 \sim HalfNormal(\sigma = 1)$$
$$\theta_{1\ time\ j(1\dots282)} \sim Normal(\mu = 0, \sigma = 1) \qquad (3.2.12)$$
$$\alpha = Deterministic(\theta_1\tau_1)$$
$$\tau_2 \sim HalfNormal(\sigma = 1)$$
$$\theta_{2\ time\ j(1\dots282)} \sim Normal(\mu = 0, \sigma = 1)$$
$$\beta = Deterministic(\theta_2\tau_2)$$
$$\mu_{expected} = \alpha \cdot T + \beta \cdot T$$
$$\sigma_{error} = HalfNormal(\mu = 0, \sigma = 10)$$
$$\nu_{d.f.} = Exponential\left(\lambda = \frac{1}{29}\right) + 1$$
$$\bar{y}_i \sim StudentT(\mu = \mu_{expected}, \sigma = \sigma_{error}, \nu = \nu_{d.f.})$$

This approach is inspired in non-parametric techniques intended to identify statistical differences across the whole epoch (Maris and Oostenveld, 2007). However, the present approach has some advantages. To start with, there is no need for multiple comparison corrections, as no p-values are used, which heavily depend on a space of counterfactual (presumed) possibilities derived from testing decisions (i.e. sample size, testing stop) inflating the false alarm rate (or Type I error). Differently, posterior distributions are unaffected by this issue as they are data-dependent (Kruschke, 2015), and hierarchical models are especially robust to any influence of multiple parameters as they directly

model distributions and each posterior distribution can be directly derived from the model (Gelman et al., 2014).

This conduces to another relevant advantage. For instance, model 3.2.12 directly includes 282 varying slopes, each representing a time-point, each a whole posterior distribution from which a high posterior density interval (HDI) can be derived, thus providing a direct interpretation of how probable (e.g. 90%) a difference between two amplitudes waves is at a given time-point sample. Hence, models are much more informative than pre-designed tests operating with fixed or not explicit parameters, as commonly observed in the discussed literature. Differently, the present approach explicitly portrays modelling decisions and assumptions, such as chosen priors, by mathematically and computationally defining them respect to hypotheses and data. (Importantly, these advantages also apply to all the previously presented models). Nevertheless, this also evidenced some limitations of the present approach. For instance, model 3.2.12 is designed for a single electrode, instead of modelling all 64 of them. This decision responds to practical limitations, such as computational capacity, as with millions of datapoints (e.g. 64 electrodes by 282 time-samples, etc.) and thousands of parameters, model convergence is not necessarily guaranteed just by extended sampling time, as the complexity of the posterior (as parameter space) might not be easy for the sampler (see section 3.3 below). This evidences the need for constructing models as adequately as possible for the explored problem, but also shows some of the limits of the current approach. This also makes the behavioural modelling to seem oversimplified, but this simply responds to the limited range of assumptions about how variables relate in the models (i.e. simple interactions). Nevertheless, it would be possible to explore alternative behavioural models, such as models using ear and/or types variables as slopes, which might serve to support or criticise the present approach.

The final model presented here is aimed to test stimuli's acoustic and lexical features (experiments presented in Chapter 4, see Figure 4.3). These involve the effects of arousal and valence and/or median pitch and Hammarberg index (see Chapter 2) on sentences classification by threat level. To this aim, the following model was constructed:

$$C_{n(1\ldots8)} \sim Normal(\mu = 0, \sigma = 10)$$

$$\beta_1 \sim Normal(\mu = 0, \sigma = 1) \qquad\qquad (3.2.13)$$

$$\beta_2 \sim Normal(\mu = 0, \sigma = 1)$$

$$\lambda_s \sim Normal(\mu = 0, \sigma = 1)$$

$$\tau_s \sim HalfNormal(\sigma = 1)$$

$$\theta_{s\,j(1\ldots22)} \sim Normal(\mu = 0, \sigma = 1)$$

$$\alpha_s = Deterministic(\lambda_s + \theta_s\tau_s)$$

$$\phi = \alpha_s + \beta_1 p + \beta_2 h$$

$$\bar{y}_i \sim OrderedLogistic(\eta = \phi, cutpoints = C_n)$$

Where $C_{n(1\ldots8)}$ are the 9 minus 1 cutpoints of a 0-8 points Likert scale for threat level, $\beta_1\ and\ \beta_2$ are L2 regularising priors for p (median pitch) and h (Hammarberg index), α_s is a reparametrised distribution for a varying intercept on subjects, and ϕ is a simple linear model acting as the predictor parameter η of an ordered logistic log-likelihood across the 8 cutpoints. Note that, if required, valence and arousal measures could be added as additional parameters in the model or simply replace p and h by arousal and valence variables (for details on ordered-logistic regression see: McElreath, 2020).

This model allows to effectively treat trial-by-trial rating data, thusly enabling the model to account for single rating variability, additionally constrained by a varying intercept across subjects. This model can provide information about two relevant things. Firstly, inference about what features are relevant for threat recognition, namely whether increases or decreases on acoustic measures or valence/arousal norms can indicate increases in threatening ratings. Secondly, based on these measures and norms it could be possible to adjust future stimuli to select and record them to be more consistently threatening and to predict future threatening ratings. In this way, the inference chain can link well-defined stimuli (threat production) with the listener's behavioural responses (reaction times and accuracy) and associated neurological activity (EEG measures).

3.3 Model Assessment

Given the complexity of the models, the best current alternative for sampling is using Markov Chain Monte Carlo (MCMC) methods. In particular, the implementation applied here is based on the Python's package PyMC3 (Salvatier et al., 2016), and it uses the no U-turn sampler (NUTS) method (Hoffman and Gelman, 2014), which is a type of

Hamiltonian Monte Carlo (HMC) sampling and it is the default sampling method for continuous parameters in PyMC3 (Martin, 2018). Roughly explained, HMC samples the log-transformed posterior distribution (log-posterior) by simulating the movement of a frictionless particle, where parameters provide a vector and the log-posterior a space or surface for the particle to move through (for more detailed explanations see Martin, 2018; McElreath, 2020). This is relevant, because all convergence tests for a model sampled with HMC/NUTS are based on the fact that as any MCMC method, NUTS is devised to produce a Markov chain, but it does it through leapfrog steps with an initial position and a momentum, which allow the particle to traverse the log-posterior and output a sampled proportional posterior distribution.

This leads to the following convergence tests, which require sampling two chains or more. 1) Energy transition: this is based on the Bayesian fraction of missing information (BFMI), which indicates whether the momentum resampling (momentum is resampled after an iteration) induced energy variation is sufficient (Betancourt, 2016); a good BFMI should be close to one, or at least BFMI > 0.5, for chains to have similar energy transitions. 2) Effective sample size (ESS): indicates whether samples within a chain are autocorrelated, the lower the ESS, the more autocorrelation and the more uncertainty of the estimates (Martin, 2018). Sufficient amounts of EES are usually: ESS > 400, or ESS > 200 if sampling proved too strenuous; autocorrelation plots and trace plots are important check-ups accompanying ESS for assessing autocorrelation. 3) Gelman-Rubin statistic ($\hat{R}$): indicates whether variance within and between chains is large, where large differences indicate non convergence (Gelman et al., 2014); good convergence implies $\hat{R} \cong 1$, or at least $\hat{R} \leq 1.1$. Inference from a model is feasible if and only if all these check-ups are passed. In short, after sampling a model, failing any of these checks will provide meaningful information about what could have caused the problem, such as the requirement for more samples, reparameterisation, observed data transformations, or simply the inadequacy of the approach.

The predictive capacity of a model requires other forms of assessment. As the main goal of the present approach is inference, prediction check-ups will be only briefly discussed. The most straightforward way for testing predictive capacity is a posterior predictive check (PPC). In general terms, a PPC consists in sampling observed data simulations from the posterior; if these simulated data match the actually observed data, the predictive capacity is good (Martin, 2018). In addition, model comparison can test

models containing different parameters of variables. Generally, the recommended options are Watanabe-Akaike or widely appliable information criterion (WAIC) and leave one out cross validation (LOO-CV), they estimate out of sample expectation or predictive fit adjusting for overfit (for more details see: Gelman et al., 2013; Martin, 2018; McElreath, 2020). Both can be used together to test whether models would be efficient for predicting unobserved data, or similar purposes. However, as models are theoretically well defined and constructed a priori, mainly with inferential aims, the assessment of models' predictive capacity will not be a common theme in the present project. Having said this, it is important to keep in mind that this aspect of model assessment could be especially relevant for criticising models in non-inferential terms in view of they capability for applied approaches.

If models fulfil all these convergence requirements (a much more stringent criteria than in conventional statistical approaches), their interpretability will depend upon the scientific model they aim to reproduce. This means that the pre-established relationships between variables, given models' estimates, will evidence effects of one variable over another. In the present case, the operative model presented in Chapter 2 (Figure 2.1) proposes effects of anxiety at different time-windows for both behavioural and EEG measures. The models presented in this chapter provide a way of estimating the effects of trait anxiety, either by worry or BIS levels, given ear presentation (laterality) or stimulus type (information channel). But they also provide a way to examine the relationships between stimuli's properties, namely the effects of acoustic measures and lexical norms on the categorisation of stimuli. This provides a strong statistical basis for linking stimuli's features with stimuli's processing (e.g. through EEG measures) and behavioural outcome (i.e. reaction times and accuracy). The upcoming sections will show experiments where these models are directly applied to the retrieved data, providing results from which sound inferences can be made.

Chapter 4
Stimuli Analysis
Prosodic and Semantic Features of Threatening Speech

4.1 Introduction

This chapter presents the selection, norming and analysis of experimental stimuli: auditorily presented sentences. As pointed out in Chapters 2 and 3, an experimentally based norming procedure will help in ascertaining the threatening features of stimuli, thus providing a better link between stimuli and behavioural/EEG responses. Theoretically speaking, specific acoustic measures and lexical norms should indicate whether the speaker actually conveys threat in the produced sentences (see Chapter 2, section 2.1). Namely, empirical tests should reveal whether the prosodic (pitch and voice quality) and semantic (arousal and valence) features proposed to characterise threat are actually associated with threat comprehension. Therefore, a proper experimental norming procedure is required in addition to comparisons between measurements; so that by comparing norms and measures a proper link between production (stimuli) and comprehension (elicited response) can be established. Together with an appropriate statistical treatment, this approach can help in dealing with some common issues regarding selection and creation of experimental materials.

First, it has been proposed that stimuli norming and comparison is conventionally analysed in statistically inappropriate ways (Sassenhagen and Alday, 2016). Mainly, this responds to the use of NHST for testing the equivalency of stimuli measures (e.g. whether they do not differ in mean fundamental frequency), which cannot be achieved by a frequentist test. Namely, frequentist tests cannot give evidence of absence, mainly due to their dependency on a null hypothesis based on an imaginary population (Kruschke, 2015). Such NHST approach for stimuli comparison is the default in the literature presented in Chapter 2. Second, another important issue present in the literature, is the inappropriate treatment of norming data. Here, data from rating scales (i.e. Likert) are not treated as categorical variables (the appropriate treatment), but they are averaged and analysed as normal distributions instead, thus obscuring relationships between variables (see McElreath, 2020). Otherwise, the present Bayesian approach will be able

to provide a direct comparison between means of acoustic and lexical features. In addition, based on model 3.2.13 (Chapter 3), threat ratings will be given an appropriate statistical treatment. In this way, present experimental norming of stimuli is not simply a check-up on stimuli's singled-out properties, but an inference- and prediction-based approach to defining what acoustic and lexical features are sufficient for characterising threat in an experimental context. In other words, strong evidence will be provided that these features may be causally associated with threat comprehension and categorisation.

Regarding acoustic threat, previous literature indicates that hot anger (rage) needs to be distinguished from cold anger (Banse and Scherer, 1996; Hammerschmidt and Jürgens, 2007), and that rage can be associated with voices higher in pitch and lower in quality (rougher), as reviewed in Chapter 2. Thus, threatening prosody, as similar to hot anger, should imply a higher pitch (median fundamental frequency) and a rougher voice (lower Hammarberg index). These two features are selected based on the following rationale. First, both median pitch and Hammarberg index address slow rate spectral measures in equivalent units (Hz); which can directly relate to laterality and time-course differences (as discussed in Chapter 2, section 2.1). Second, as median pitch is a central tendency measure marking the mid-point frequency in the sentence and Hammarberg index represents a difference between lower and upper frequency ranges, they may not be so sensitive to prosody variations within the sentences, thus reflecting the overall prosodic threat of each sentence more accurately.

Similarly, lexical content could be evaluated from measurements taken from emotional judgements on words; such as measures of valence and arousal from the extended version of the Affective Norms for English Words (ANEW) (Warriner et al., 2013). Here, valence indicates how negative/positive a word is perceived and arousal indicates whether the stimulus is arousing. These norms are developed to indicate affective judgements on words through a self-assessment manikin (SAM), which consists of a visual and numerical Likert scale intended for participants to indicate their feelings/emotions after reading/listening a stimulus (i.e. word) (Warriner et al., 2013). The ANEW norms also include a dominance measure, associated with meekness, but this has been shown to highly correlate with valence (Montefinese et al., 2013), which could create redundancy in the present approach. As reviewed in Chapter 2 (section 2.1), it has been observed that words conveying pain or threat show high arousal and low valence (Borelli et al., 2018; Ho et al., 2015), which gives consistency to the present approach. In

other words, ANEW valence and arousal norms from threatening lexical items within a sentence can give information about the overall threat of each sentence. Although sentential context can contribute or completely determine the meaning of a sentence (Lai et al., 2015), lexical items with clear offensive/aggressive content (i.e. high arousal and low valence) will drive the threatening content of the whole sentence (see Chapter 2, section 2.1).

Given this, median pitch and Hammarberg index are to prosody what arousal and valence are to semantics; and the following experiments will test whether median pitch and Hammarberg index can characterise threatening prosody, and whether valence and arousal can characterise threatening semantics. This leads to the following operative hypothesis: threatening prosody is higher in pitch and rougher (high median pitch, low Hammarberg index), and threatening semantics are arousing and negative (high arousal, low valence). This allows the following predictions to be made: 1) Sentences that contain only threatening prosody (neutral semantics and threatening prosody: Prosody) should not be characterised by arousal and valence, but clearly characterised by median pitch and Hammarberg index. 2) Sentences that contain only threatening semantics (neutral prosody and threatening semantics: Semantic) should not be characterised by pitch and roughness, but clearly characterised by arousal and valence. 3) Therefore, sentences containing lexical items which are high in arousal and low in valence should be easily categorised as semantically threatening, while sentences recorded as threatening (high in median pitch and low in Hammarberg index) should be easily categorised as prosodically threatening. 4) Control sentences (neutral prosody and neutral semantics: Neutral) should be characterised by no increase in median pitch/arousal and no decrease in Hammarberg index/valence.

To test these possible effects, Prosody stimuli were selected in low arousal and higher valence and recorded by asking a London English speaker (untrained in acting: lay speaker) to produce an aggressive voice (as if threatening someone). Semantic stimuli were selected to be high in arousal and low in valence and the speaker was requested to record them in a neutral voice (as in conversation). Neutral stimuli were selected to be around the median in arousal and valence (or at least median arousal and higher valence) and requested to be spoken in neutral voice. Study 1 involves the rating (threat scale) of written stimuli for a Semantic vs. Neutral comparison, and a second rating session for a small sample of Prosody vs. Neutral stimuli. Study 2 involves the rating of fifty-four

acoustic stimuli per type for both sessions in threat and SAM scales. Note that Study 1 ratings were collected as a simple check-up in both written (for Semantic) and acoustic (for Prosody) modalities, so they present uneven sets of stimuli and participants, without detailed demographic information. To ameliorate this, Study 2 implemented a more controlled rating procedure (acoustic for both conditions), with a smaller number of listeners judging sentences (e.g. Hammerschmidt and Jurgens, 2007) and rating all experimental Prosody and Semantic stimuli (from Chapter 7). These decisions seek a balance between appropriateness of the analysis and resources availability.

4.2 Methods

4.2.1 Study 1 Methods

4.2.1.1 Participants

For Study 1, 60 participants (29 females, age over 18) took part on rating written neutral and semantic-threat sentences: first session; and 22 participants (age over 18, no more information available) took part on rating auditorily presented neutral and prosody-threat sentences: second session. Participants were paid at £7.5/hour rate and gave their consent before participating according to 1998 European data protection act.

4.2.1.2 Materials

Four types of sentences were defined: Prosody (neutral-semantics and threatening-prosody), Semantic (threatening-semantics and neutral-prosody), Congruent (threatening-semantics and threatening-prosody), and Neutral (neutral-semantics and neutral-prosody). Semantically threatening sentences were extracted from movie subtitles by matching the subtitles with a list of normed threatening ANEW words (Warriner et al., 2013). Any word over 5 points in the arousal scale, and below 5 points in the valence scale was considered threatening (these scales ranged from 1 to 9 points). Every word with less than 5 arousal points and between 4 and 6 (inclusive) valence points was considered neutral. Words' frequencies were extracted from SUBTLEX-UK (van Heuven et al., 2014), only sentences containing words with Zipf's log frequencies over 3 were included. After this, sentences were recorded in an acoustically isolated chamber using a RODE NT1-A1 microphone by a male English speaker (untrained in acting: lay speaker). The speaker was instructed to speak in what he considered his own angry threatening/angry or neutral voice for recording Prosody/Congruent and

Semantic/Neutral sentences respectively. Sentences were not repeated across type (i.e. each type has a unique set of sentences). Figure 4.1 shows examples of recorded stimuli.
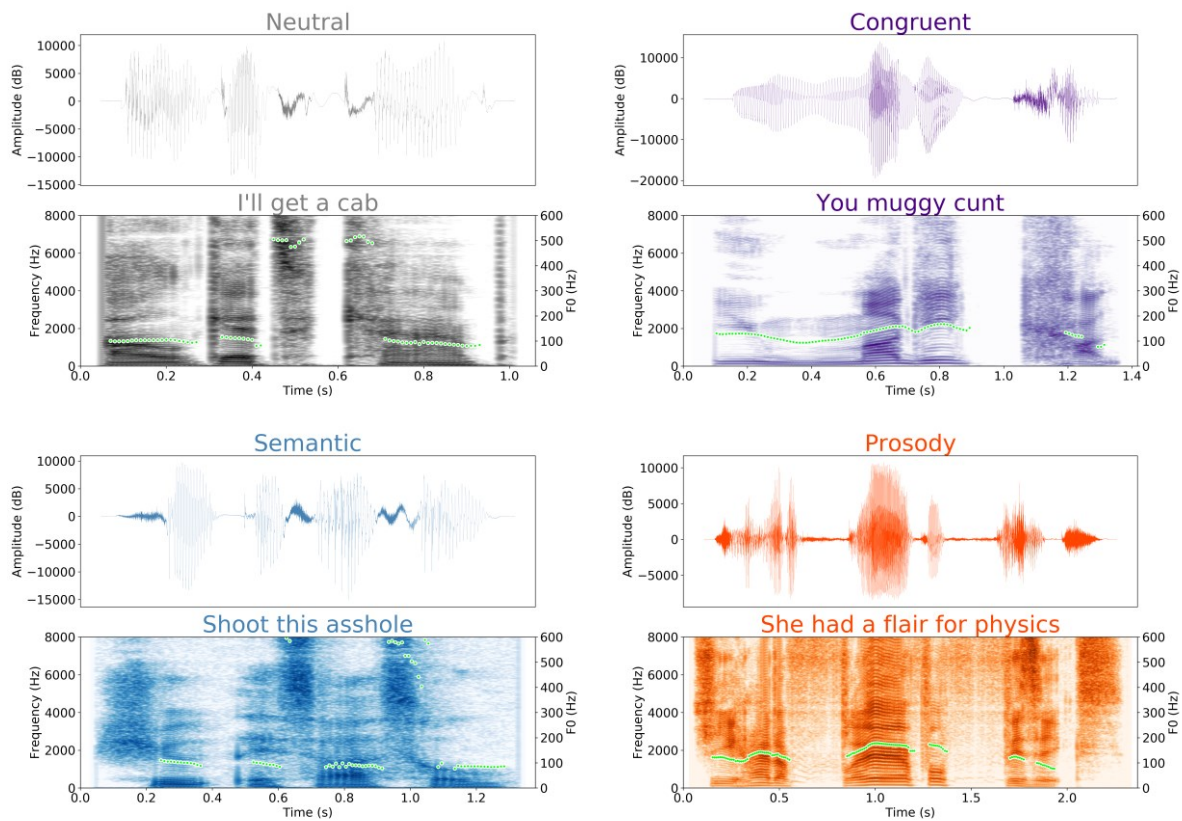
Figure 4.1. Example of four sentences used in Study 1. Top of each image: oscillogram showing amplitude changes. Bottom of each image: spectrogram showing frequency changes. Top left: neutral prosody and neutral semantics (Neutral). Top right: threatening prosody and threatening semantics (Congruent). Bottom left: neutral prosody and threatening semantics (Semantic). Bottom right: threatening prosody and neutral semantics (Prosody). Green dots indicate fundamental frequency (F0) contours.

4.2.1.3 Procedure

In Study 1's rating sessions participants rated how threatening they considered sentences to be in a 0-8 points Likert scale (where 0 is not threatening at all and 8 is very threatening). Participants in the first session rated Semantic sentences (44) together with control Neutral sentences (186) in written form. Participants in the second session rated Prosody sentences (7) together with control Neutral sentences (7) in auditory form. These sessions included sentences selected/recorded to express concern (and respective scale) as well, but they are not included in final analyses as the main focus of present research is only on threat.

4.2.2 Study 2 Methods

4.2.2.1 Participants

For Study 2, 10 University College London (UCL) students (9 Females, mean age = 19.5, all cisgender) rated auditorily presented neutral and semantic-threat sentences, and 10 UCL students (9 Females, mean age = 19.1, all cisgender) rated auditorily presented neutral and prosody-threat sentences. All received coursework credits for their participation and consented to participate informed of GDPR regulations regarding data protection. Participants declared not having auditory, psychiatric or neurological problems, and speaking English as first or native language.

4.2.2.2 Materials

For the second study sentences were re-selected. Any word over 5 points in the arousal scale, and below 5 points in the valence and dominance scales was considered semantically threatening. Every word with less than 5 arousal points and between 4 and 6 valence points was considered semantically neutral. Some sentences remained from the first study's materials and, again, only sentences with Zipf's log frequencies over 3 were included. The same speaker from Study 1 was recruited for a new recording session, and he was instructed to speak in the same manner, but being careful not to de-emphasise threat. Sentences were recorded in the same manner as in Study 1, but minor technical issues were fixed. Figure 4.2 shows examples of recorded stimuli.

4.2.2.3 Procedure

In Study 2's rating session participants rated sentences in the same 0-8 Likert scale for threat, but also judged their arousal and valence levels in SAM scale (as in ANEW norms: Warriner et al., 2013), and categorised sentences as Calm, Neutral, Angry or Enraged. Prosody sentences (54) and Semantic sentences (54) were rated separately. In both Prosody and Semantic rating sessions, Neutral sentences (54) were used as a control. Sentences were presented randomly in both tasks.
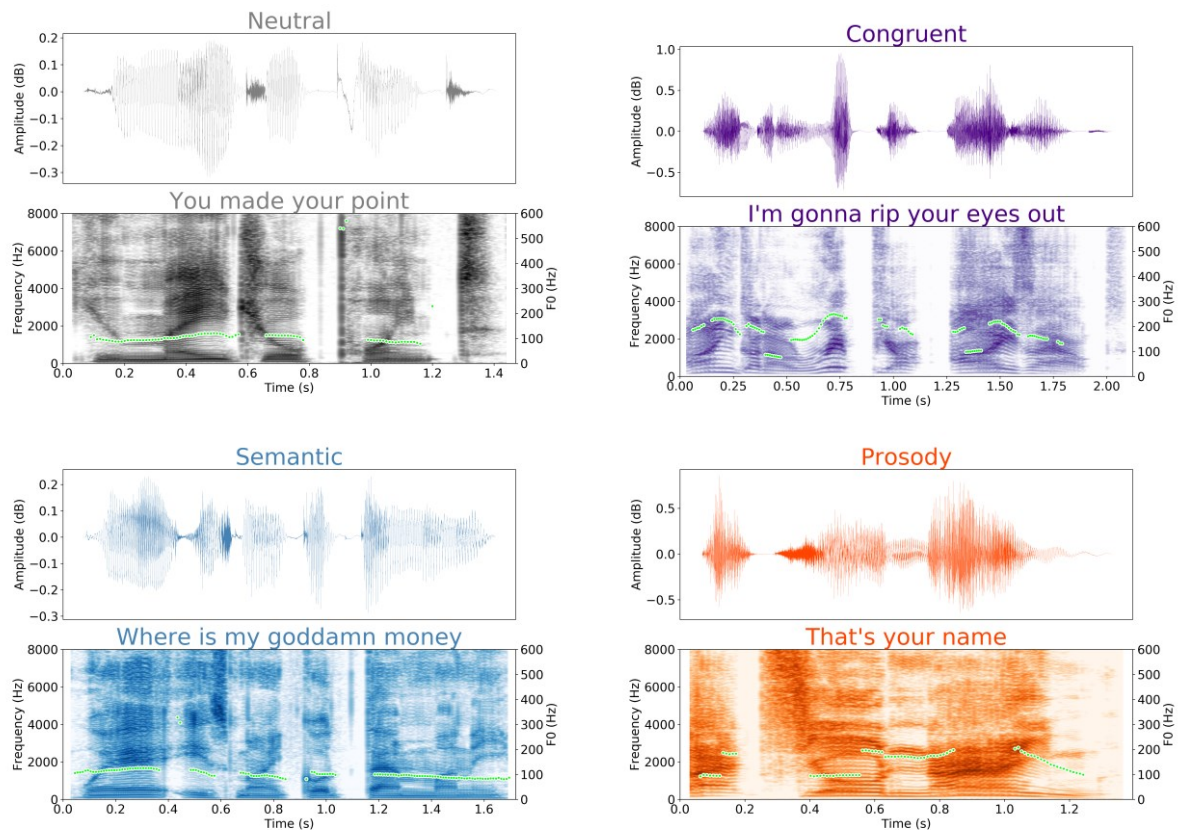
Figure 4.2. Example of four sentences used in Study 2. Top of each image: oscillogram showing amplitude changes. Bottom of each image: spectrogram showing frequency changes. Top left: neutral prosody and neutral semantics (Neutral). Top right: threatening prosody and threatening semantics (Congruent). Bottom left: neutral prosody and threatening semantics (Semantic). Bottom right: threatening prosody and neutral semantics (Prosody). Green dots indicate fundamental frequency (F0) contours.

4.2.3 Analysis

The aim of the present chapter is demonstrating whether acoustic measures and lexical norms can function as markers for threat. Hence, an ordered-logistic regression was implemented to assess the relationship between acoustic/lexical features and threat ratings. The general structure of the model is described on Figure 4.3. Previous analyses on these stimuli were performed using the Bayesian estimation supersede the t-test (BEST) (Kruschke, 2013), but using exponential distributions for standard deviations (Busch-Moreno et al., 2020a; 2020b). These analyses are repeated here as a way of demonstrating that mean differences between stimuli are consistent with threat ratings. All models were sampled with the Hamiltonian Monte Carlo (HMC) no U-turn sampling (NUTS) method as provided by PyMC3 (Salvatier et al., 2016), using 1000 tuning samples and 1000 samples.
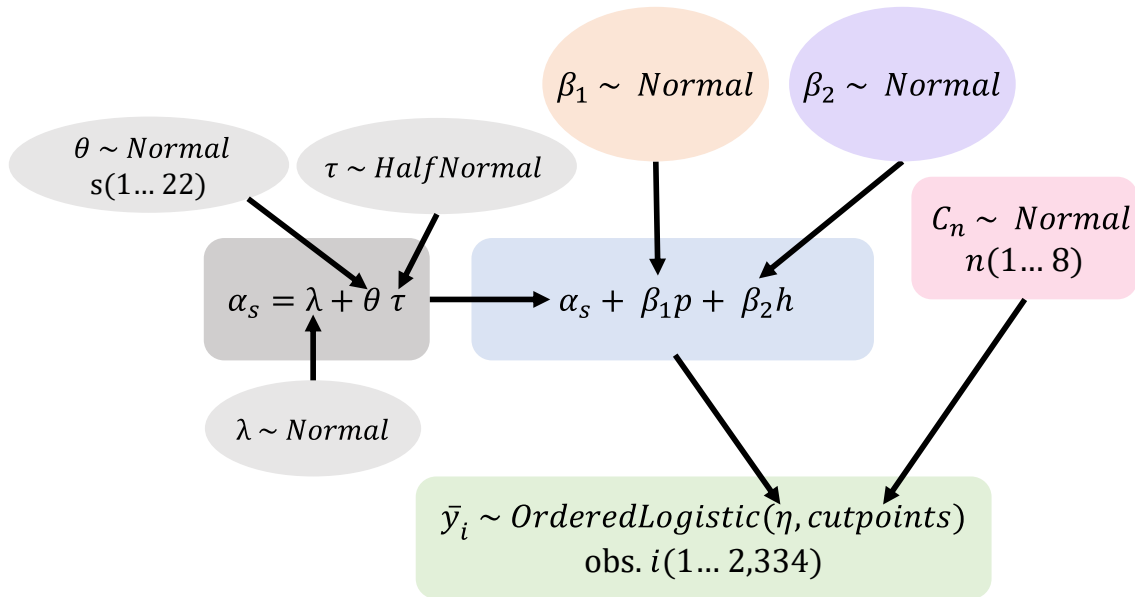
Figure 4.3. Graph representation of hierarchical ordered-logistic model, used for analysing ratings data in terms of acoustic measures or lexical norms. The α_s parameter is the re-parametrised varying intercept across subjects. Parameters denoted as β represent slopes of median pitch and Hammarberg index, denoted as p and h respectively. Parameter C_n indicates the 9 minus 1 cutpoints for the Likert scale. Note that additional β parameters could be added for shimmer and harmonicity without altering the general model's structure. For the analysis of Semantic ratings, the number of participants change and the variables p and h change to v and a, for valence and arousal respectively.

For Study 1 and 2, four sentences' acoustic measures were extracted: Hammarberg index (maximum energy differences between the 0-2000hz and 2000-5000hz ranges), harmonicity (amplitude signal to noise ratio or harmonic to noise ratio), median pitch (fundamental frequency: F0), and shimmer (ratio between absolute amplitude difference of consecutive periods and mean amplitude). Acoustic measures were extracted via Python's Parselmouth interface to Praat (Jadoul et al., 2018). For both studies, lexical norms were ANEW mean Arousal and Valence scores from the lexical items used in the aforementioned sentence selection procedures. For both Studies, the dependent variable consisted of threat scores given by participants. Plots and model comparisons were produced using Arviz (Kumar et al., 2019) and Matplotlib (Hunter, 2007).

4.3 Results

All models showed excellent convergence with $\hat{R} \cong 1$, ESS > 400, and BFMIs > 0.7. All other details, such as stimuli, data, summaries, traceplots, autocorrelation-plots and additional plots of results can be also found at the present chapter's Open Science Framework (OSF) repository (https://osf.io/xrfkq/).

4.3.1 Study 1 Results

Results from Study 1's Semantic model are summarised in Table 4.1, and they indicate a clear modulation of threat ratings by Arousal and Valence. This is evident in Arousal, with a 90% high posterior density interval (HDI) at least 2 SDs away from zero, reliably indicating that the log-odds of a stimulus being rated higher in threat increase by 0.76 on average per Arousal score unit. Similarly, but slightly less reliably (there is a minor HDI overlap with 2SDs), a one-point increase in Valence implies a decrease in mean log-odds of around -0.60.

Table 4.1. Study 1, Semantic ordered-logistic model results

Lexical Norm	Mean	SD	HDI 5%	HDI 95%
Arousal	0.764	0.038	0.705	0.825
Valence	-0.596	0.029	-0.644	-0.550

Prosody results are summarised on Table 4.2, and they go contrary to predictions. They indicate a very small (almost negligible) decrease in Median Pitch; a stronger decreased in Hammarberg Index, but with HDIs completely overlapping zero; which also happens for Shimmer; and an increase in Harmonicity. These results indicate that no consistent inference can be drawn from acoustic measures in this very narrow (14 stimuli) sample. However, a model comparison using LOO-CV indicates that though this model has better in-sample and out-sample predictive capacity (Log score = -2594, error = 41.4, weight = 99%), the model without Harmonicity and Shimmer gives equivalent results for Median Pitch and Hammarberg Index. Interestingly, as seen in Figure 4.4, Median Pitch (pitch) and Hammarberg Index (roughness) are better at explaining higher threatening ratings in the Prosody model, than Arousal and Valence in the Semantic model. Nevertheless, semantic ratings still show a consistent, though less pronounced, increase from the lowest to the highest threatening rating probability when valence decreases and arousal increases.
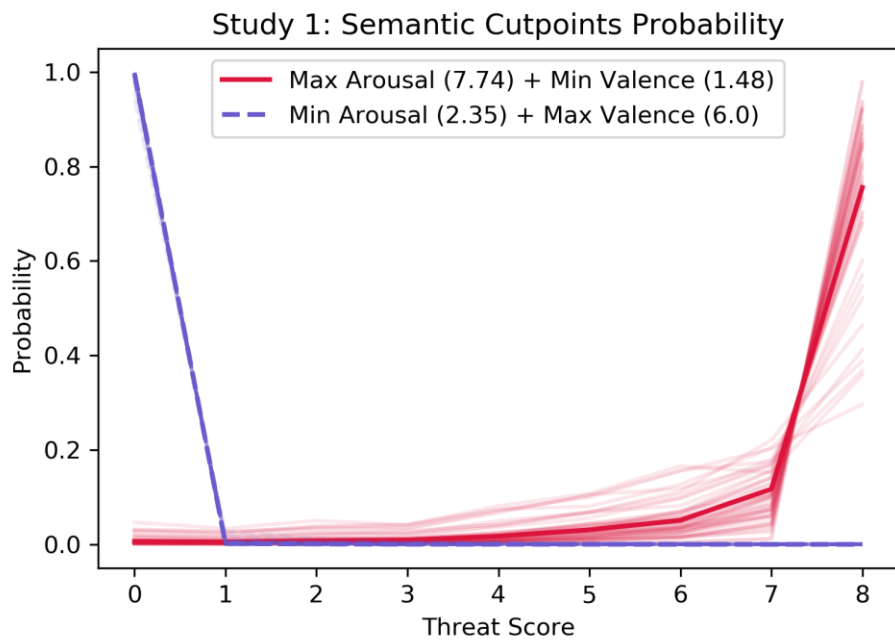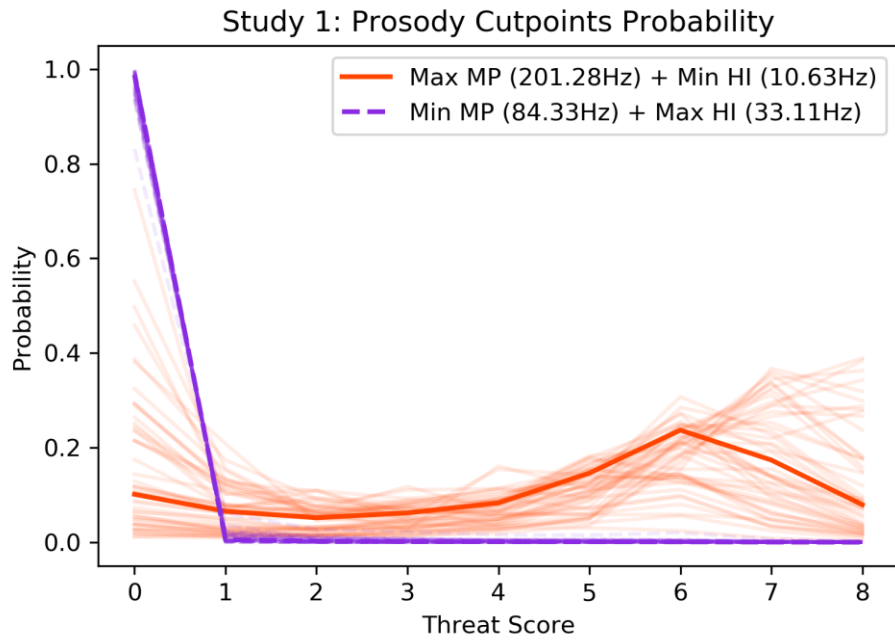
Figure 4.4. Study 1 cutpoints plots. Plots show probability of rating a sentence as not threatening at all (0 points) or very threatening (8 points) given acoustic measures for the Prosody experiment (upper panel) and given lexical norms for the Semantic experiment (lower panel). Prosody orange solid line indicates maximum Median Pitch (MP) plus minimum Hammarberg Index (HI) posterior distributions, violet dashed line indicates minimum Median Pitch (MP) plus maximum Hammarberg Index (HI) posterior distributions. Semantic solid red line indicates maximum Arousal plus minimum valence posterior distributions, blue dashed line indicates minimum Arousal plus maximum Valence posterior distributions. Faded lines are random samples from the posteriors, expressing uncertainty. Note that for both Prosody and Semantic ratings Neutral sentences overfit, most are rated as zero threatening so dashed purple/blue lines drastically drop to almost zero probability for 1 point of threat score.

Table 4.2. Study 1, Prosody ordered-logistic model results

Acoustic Measure	Mean	SD	HDI 5%	HDI 95%
Median Pitch	-0.006	0.004	-0.013	0.001
Hammarberg Index	-0.352	0.033	-0.402	-0.296
Shimmer	-0.931	0.902	-2.427	0.602
Harmonicity	0.356	0.056	0.261	0.442

Notwithstanding these differences, the models have a reasonable predictive capacity. As shown on Figure 4.5, predicted threat scores for Study 2's ratings fall within an acceptable rage of actually collected scores. Better Prosody predictions might indicate that Semantic threat scores changed more consistently by re-selection and/or that Prosody threat scores are given on the basis of multiple acoustic features in combination, from which voice quality and pitch are sufficient but not completely necessary for explaining increases in threat score.

Results from BEST models indicate that mean differences are in coincidence with these predictions. Indeed, BEST models indicate clear similarities across studies. To derive means, 54 stimuli per category were used. All models showed excellent convergence with $\hat{R} \cong 1$, ESS > 1000, and BFMIs > 1. Table 4.3 summarises the estimated means and SDs, which shows that for both studies Arousal means are higher and Valence means are lower for Congruent and Semantic; while Median Pitch means are higher and Hammarberg Index means are lower for Prosody and Congruent. Figure 4.6 shows, differences between means clearly indicate that posterior distributions of matching parameters substantially overlap or are very close to each other; while non-matching parameters' estimates fall wide apart.
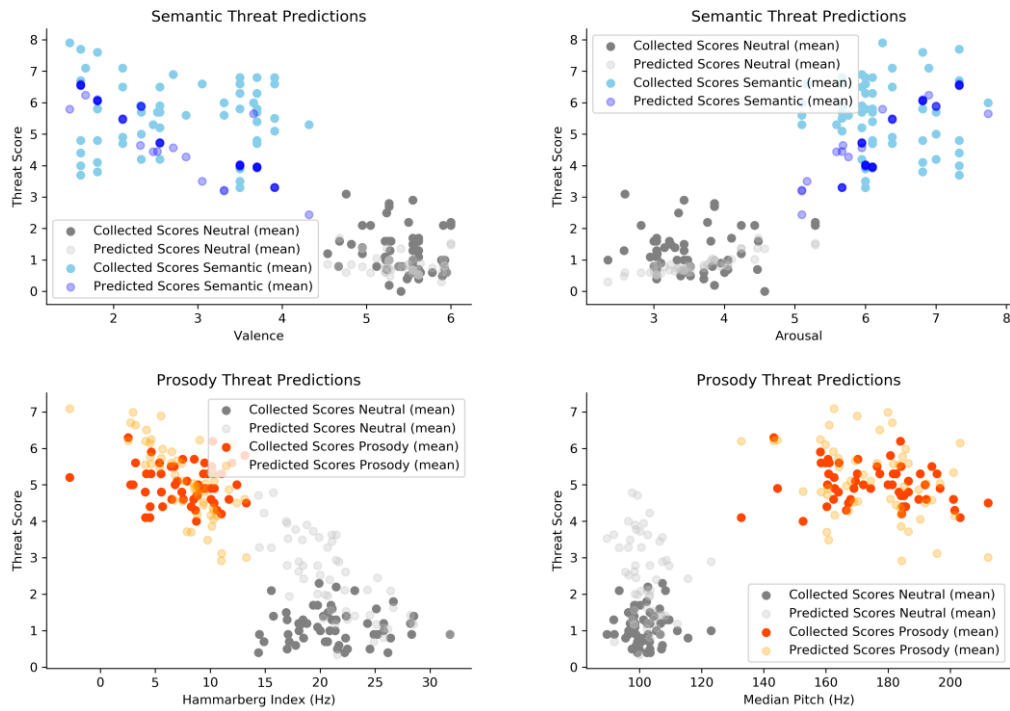
Figure 4.5. Prediction of ratings of Study 2 using ratings from Study 1. Upper left: Semantic threat predictions across Valence. Upper right: Semantic threat predictions across Arousal. Bottom left: Prosody threat predictions across Hammarberg Index (roughness). Bottom right: Prosody predictions across Median Pitch (pitch). Note that Semantic predictions are less accurate for threat (they cluster as solid blue dots) and Prosody predictions are less accurate for neutral.

Table 4.3. Study 1, BEST Mean and SD average posteriors

Study 1	Arousal		Valence		Median Pitch		Hammarberg Index	
Category	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Neutral	3.89	0.74	5.84	0.91	96.48	4.09	33.83	3.57
Congruent	5.88	0.95	3.31	1.25	129.90	9.10	24.15	5.48
Prosody	3.67	0.64	5.93	0.80	133.30	11.60	22.87	4.01
Semantic	6.00	0.93	2.98	1.07	90.06	2.18	31.70	4.62

Study 1: BEST Means Comparisons
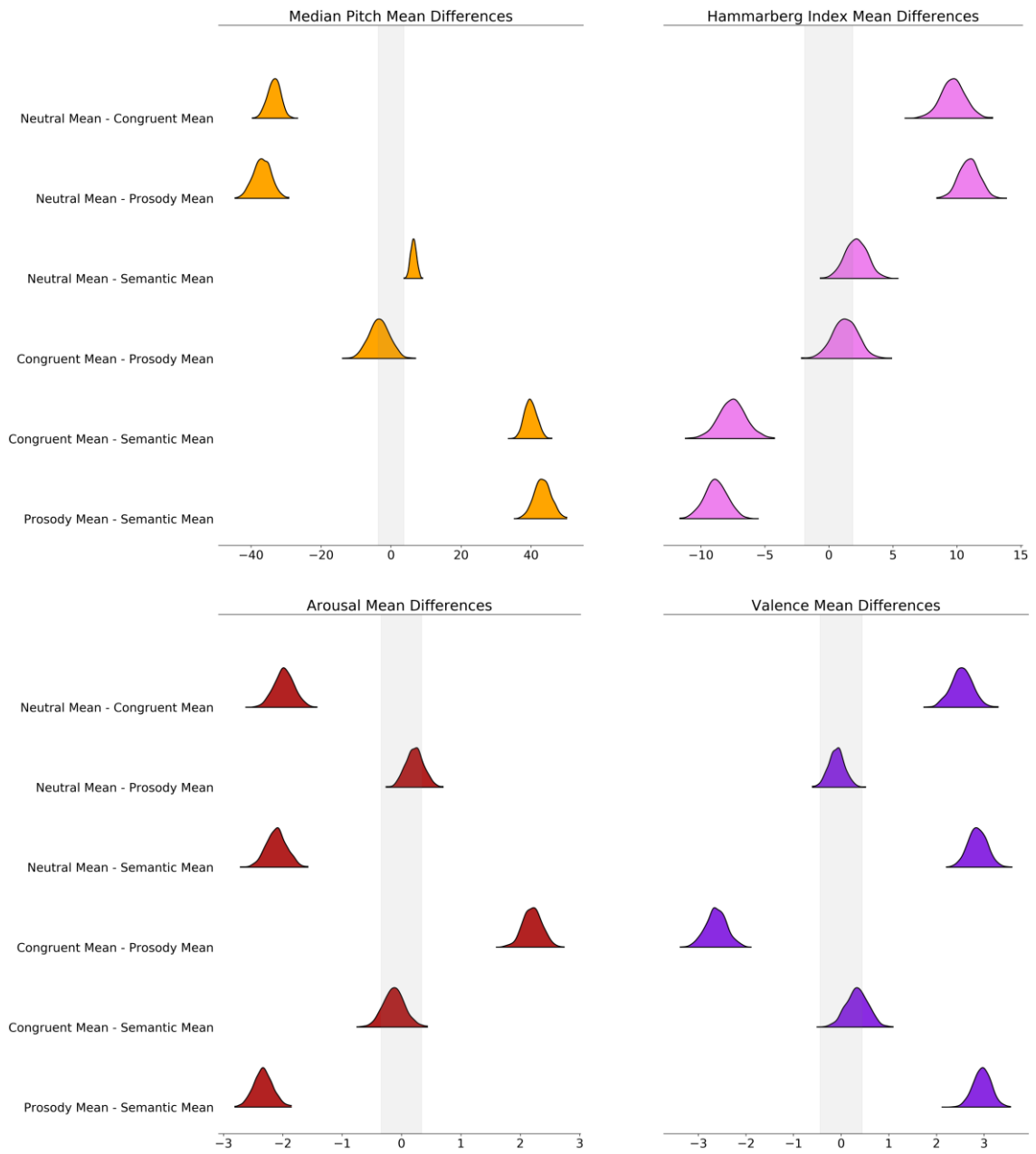
Figure 4.6. BEST model results from Study 1. Top: Median Pitch and Hammarberg Index. Bottom: Arousal and Valence. Ridgeplots represent the HDI distribution (kernel density) from each mean difference's posterior. Grey bands indicate 2SDs regions of practical equivalence (ROPE).

4.3.2 Study 2 Results

For Study 2, all ratings were given to auditory stimuli (54 per category), so both Semantic and Prosody models included acoustic measures and lexical norms. Results of subjects' varying intercepts will not be addressed, as they are not of direct interest. To confirm the contribution of measures of Harmonicity and Shimmer, a model comparison for the Prosody model was performed through the LOO-CV method (see Chapter 3, section 3.3). This indicated that the full model (log score = -1,734, error = 29, weight = 74%) barely outperforms the reduced model (only using pitch and roughness) only in out-sample predictions but not in in-sample predictions, showing a log score difference of only 2 points (log score = -1,736, error = 29, weight = 26%). Although the full model shows more weight, inferential results indicate that estimates of Harmonicity and Shimmer fall across zero or widely overlap the ROPE, thus showing minimal inferential contributions. Thus, only results from the reduced model are presented. As shown in Table 4.4, Arousal and Valence contribute little to explain acoustic threat (HDIs across zero), but Median Pitch and Hammarberg Index show clear increases and decreases respectively. This indicates that as pitch increases and voice quality decreases (more roughness), the ratings become more threatening. Oppositely, Table 4.5 indicates that pitch and roughness contribute little to explaining threatening ratings (HDIs across zero), but Arousal and Valence show a strong increase and decrease respectively.

Table 4.4. Study 2, ordered-logistic Prosody model results

Norm/Measure	Mean	SD	HDI 5%	HDI 95%
Median Pitch	0.033	0.003	0.029	0.037
Hammarberg Index	-0.113	0.013	-0.134	-0.091
Arousal	0.007	0.105	-0.167	0.168
Valence	-0.074	0.158	-0.319	0.199

Table 4.5. Study 2, ordered-logistic Semantic model results

| Norm/Measure	Mean	SD	HDI 5%	HDI 95%
Median Pitch	-0.012	0.010	-0.029	0.004
Hammarberg Index	0.028	0.014	0.002	0.050
Arousal	0.518	0.083	0.385	0.653
Valence	0.356	0.056	0.261	0.442
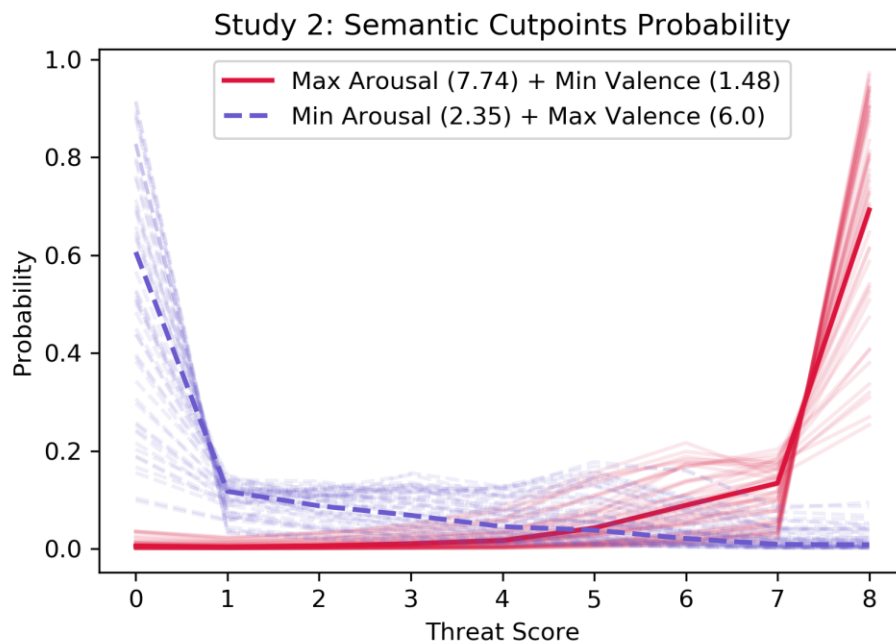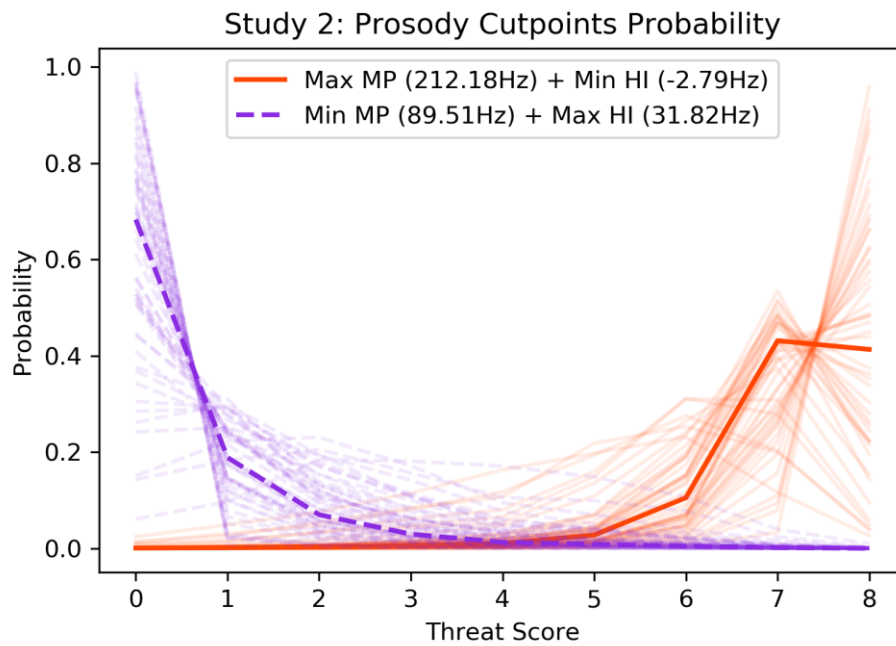
Figure 4.7. Study 2 cutpoints plots. Plots show probability of rating a sentence as not threatening at all (0 points) or very threatening (8 points) given acoustic measures for the Prosody experiment (upper panel) and given lexical norms for the Semantic experiment (lower panel). Prosody orange solid line indicates maximum Median Pitch (MP) plus minimum Hammarberg Index (HI) posterior distributions, violet dashed line indicates minimum Median Pitch (MP) plus maximum Hammarberg Index (HI) posterior distributions. Semantic solid red line indicates maximum Arousal plus minimum valence posterior distributions, blue dashed line indicates minimum Arousal plus maximum Valence posterior distributions. Faded lines are random samples from the posteriors, expressing uncertainty. Note the improvement of estimates respect to Study 1 (Figure 4.4).

Table 4.6. Study 2, BEST Mean and SD average posteriors

Study 2	Arousal		Valence		Median Pitch		Hammarberg Index	
Category	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Neutral	3.55	0.58	5.43	0.36	101.38	3.84	21.09	3.51
Congruent	6.13	0.67	2.88	0.92	167.25	12.65	7.93	3.70
Prosody	3.80	0.52	5.50	0.36	175.85	10.19	7.85	2.81
Semantic	6.26	0.64	2.75	0.85	102.01	3.95	21.20	3.75

Study 2 shows the opposite pattern respect to Study 1, in terms of overall probability across cutpoints. While the Semantic model indicates that lower ratings have low probability, the Prosody model shows a higher probability for lower values and a less pronounced probability increase. This might indicate that Study 1 has more extreme values for Arousal and Semantic, while Study 2 has more extreme values for Median Pitch and Hammarberg Index. Even so, Study 2's mean differences as estimated by BEST are remarkably similar. As summarised in Table 4.6, Semantic and Congruent Arousal means are high but Neutral and Prosody are low; while Prosody and Congruent Median Pitch means are high but Neutral and Prosody are low. Again, the opposite pattern is observed for Valence and Hammarberg Index. Figure 4.8 summarises the differences between means, as estimated by BEST, across all conditions.

In short results indicate that threat ratings increase as Median Pitch and/or Arousal increases and Hammarberg Index and/or Valence decreases. Furthermore, the central tendency (mean) of Median Pitch and Arousal is higher for Prosody and Semantic threat respectively, while the central tendency (mean) of Hammarberg Index and Valence is lower for Prosody and Semantic threat respectively. Study 2's SAM ratings for arousal and valence given to each sentence provide further confirmation of this pattern. These are not ANEW ratings, but ratings given by Study 2's participants to sentences and not words; they can be denominated arousal-sen and valence-sen. Ratings for Prosody show high arousal-sen (m = 5.96, SD = 1.4) and low valence-sen (m = 2.47, SD = 1.49) for Prosody sentences (prosodic threat only), as opposed to Neutral sentences showing low arousal-sen (m = 1.64, SD = 1.66) and valence-sen around the median (m = 4.25, SD = 1.86). Ratings for Semantic indicate high arousal-sen (m = 6.08, SD = 2.10) and low valence-sen (m = 1.65, SD = 1.52) for Semantic sentences (semantic threat only), but low arousal-sen (m = 2.24, SD = 2.01) and valence-sen around the median (m = 4.19, SD = 1.19) for Neutral sentences. Furthermore, 73% of Prosody sentences were classified as Angry, 16% as Enraged, 11% as Neutral and none as Calm. Similarly, 52% of Semantic

sentences were classified as Angry, 27% as Enraged, 17% as Neutral and 4% as Calm. No statistical analyses were performed on SAM ratings or classifications, as these trends are sufficiently clear and consistent with previous results.
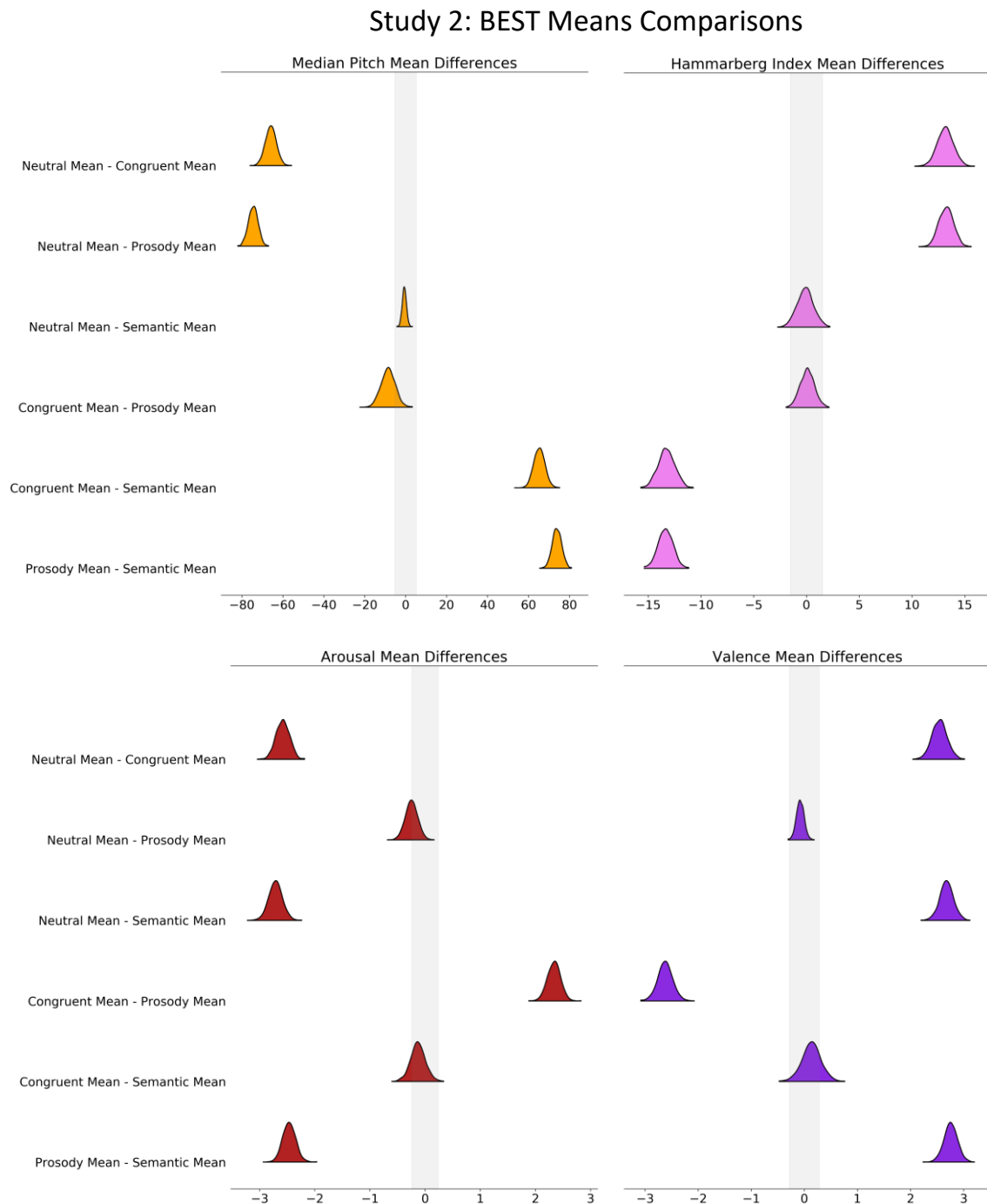
Figure 4.8. BEST model results from Study 2. Top: Median Pitch and Hammarberg Index. Bottom: Arousal and Valence. Ridgeplots represent the HDI distribution (kernel density) from each mean difference's posterior. Grey bands indicate 2SDs regions of practical equivalence (ROPE).

4.3 Discussion

Overall, present results validate stimuli and their selection procedure. This is demonstrated by the predictive capacity of stimuli and ratings from Study 1, by the clear rating pattern from Study 2, and by the replicated BEST results across studies. Results from BEST models indicate that Semantic's high arousal and low valence means homologate Prosody's high median pitch and low Hammarberg index. That is, prosodic threat (Prosody and Congruent) is characterised by higher average median pitch but lower average Hammarberg index, semantic threat (Semantic and Congruent) is characterised by higher average arousal and lower average valence, and Neutral sentences show lower averages in all measures. Ratings, in particular from Study 2, indicate that increases in arousal and median pitch and decreases in valence and Hammarberg index explain and predict threatening scores. In other words, the more arousing and negative the semantics, and the more pitched and rougher the prosody, the more threatening the sentence is perceived. Furthermore, threatening sentences (Prosody and Semantics) from Study 2 show higher arousal and lower valence in a SAM scale, indicating that lexical and spectral features drive sentence-level perception of arousal and valence. In general, these results indicate that the assumption of arousal and valence as homologues of median pitch and Hammarberg index is appropriate, and that stimuli are comprehended as threatening. Hence, these stimuli are potentially good at eliciting emotions/cognitions such as anxiety.

Present results add up to the accumulated evidence indicating that more pitch and roughness are expressions of hot-anger (Banse and Scherer, 1996; Frühholz et al., 2016b; Hammerschmidt and Jürgens, 2007; Juslin and Laukka, 2003). Although most of these previous investigations usually use harmonicity (harmonic to noise ratio) as a measure of noisier less quality speech, present results show that Hammarberg index, a frequency measure of voice quality, can be a better predictor. Indeed, angry prosody's harmonicity has not always shown direct effects on brain activity (Frühholz et al., 2016b). Importantly, noting that most Prosody sentences (89%) are classified by participants as expressing anger or rage, the present approach links this hot-anger/rage production with aggression, namely with the communication of threat. This implies that the speaker of presents studies effectively communicated threatening speech (as requested). This is consistent with research observing that natural expressions of anger have increased pitch and decreased voice quality (Anikin and Lima, 2017). This implies that sentences

with more pitch and rougher (less voice quality) could be more naturally comprehended as threatening.

In terms of semantics, present results support the sufficiency of Valence and Arousal ANEW norms (Montefinese et al., 2013). In other words, a bidimensional ANEW approach, at least for the portrayal of semantic threat. This is supported from previous evidence showing how pain or threat corresponds to increases in negativity and arousal (Borelli et al., 2018; Ho et al., 2015). This indicates that these words convey the potential of inducing harm, which is consistent with prosody measures as conveying potential aggression by increasing voice's pitch but decreasing voice's quality. In addition, present results indicate that lexical items can strongly drive the interpretation of threatening sentences. This is consistent with previous evidence evaluating semantic emotional effects of sentences based on emotional words (e.g. Kotz and Paulmann, 2007; Chen et al., 2011). However, syntax may also drive the perception or comprehension of sentences as threatening. Indeed, it has been shown that sentences lacking lexical items with emotional meaning can be comprehended as emotional (Lai et al., 2015). So, even if present lexical items are sufficient for explaining and predicting perceived threat, addressing possible influences of syntax can prove a relevant further step. Even so, both Prosody and Semantic sentences show higher arousal and lower valence average SAM ratings than Neutral sentences. Also, most Semantic sentences (79%) are classified by participants as angry or enraged. These patterns add up to the evidence provided by present models, indicating clear effects of selected features (pitch, voice quality, arousal and valence) on threatening speech comprehension.

All in all, the present approach is a strong alternative to conventional methods for treating experimental stimuli, usually subject to inappropriate procedures and statistical treatment (Sassenhagen and Alday, 2016). In addition, it is a step forward in researching naturalistic speech, not only by including better statistical modelling and stimuli selection (Alday, 2018), but also by treating stimuli analysis and norming itself through an experimental approach. Even so, there are some crucial limitations that need to be addressed. Firstly, ratings from Study 1 were split between written (for Semantic) and auditory (for Prosody) ratings, also containing later excluded items and deficient recordings, which reduces consistency. In addition, sample sized between Study 1's Semantic and Prosody tasks were dissimilar and insufficient for Prosody. This implies that both inference and predictions from Study 1 are weak, as evidenced by overfit of

Neutral ratings. Nevertheless, these issues were addressed in Study 2, by including improved versions of selection/recording and norming procedures. Despite these limitations, comparisons of means (BEST) showed great consistency across studies. This also evidences that Bayesian approaches are an excellent option for dealing with small data, and also demonstrate that hierarchical models can properly address data inconsistencies if appropriate and meaningful levels are modelled. A limitation of Study 2 is the lack of confidence ratings. However, the fact that participants in Study 2 tend to classify most sentences as angry or enraged in the four items scale (calm, neutral, angry, enraged) strengthens the suggestion that threatening and offensive language resembles hot-anger if understood as emotional expression. Note that the analogous nature between pitch/voice-quality and arousal/valence (Patel et al., 2011), strongly supported by present results, can be extended to an analogy between hot-anger and threatening prosody not only because they match acoustically, but also because they imply more arousing and negative meaning which can be understood as potentially harmful (Borelli et al., 2018; Ho et al., 2015).

In summary, the present study shows strong acoustic and lexical counterparts that underlay prosodic and semantic threat. Namely, Arousal increases and Valence decreases, indicating that semantic threat is comprehended via negatively arousing content. Similarly, Median Pitch increases and Hammarberg Index decreases, suggesting that prosodic threat is comprehended through a rougher higher pitched voice (angry voice). Importantly, these analyses show the relevance of the present procedure for semi-naturalistic stimuli. This extends the scope of evidence supporting theories of threat production and perception. Also, inter-speaker variability needs to be addressed, as different speakers could produce slightly different prosodic cues for conveying threat. Notwithstanding, present sentences can guarantee good threatening stimuli and also evidence core properties of semantic and prosodic threat.

Chapter 5
Behavioural Evidence 1
Is Anxious Repetitive Thinking Influencing Responses?

5.1 Introduction

As discussed in Chapter 2, humans can convey emotion information through different channels, and in the particular case of language the manipulation of tone and/or meaning (i.e. prosody and semantics) are common ways to do so. These different informational features (suprasegmental information associated with prosody, and segmental information associated with semantics) can develop together in a complex language emission such as an emotional sentence, and can convey emotional information simultaneously. Whether intrinsic affect differences between individuals (e.g. variation in trait anxiety) has differentiable effects on prosody and semantics remains a generally unexplored problem in language perception and comprehension research.

The present study, published as a full paper elsewhere (Busch-Moreno et al., 2020a), aims to understand the effect of trait anxiety on these information properties of speech, based on two dichotic listening (DL) experiments. Where DL can be a robust test of functional hemispheric lateralization (Hugdahl, 2011), tapping into features of both speech (language) and anxiety (affect) processing. DL can provide a behavioural test of laterality in such a way that information- and affect-related aspects of processing can be disentangled. Normally, responses to DL tasks that do not involve prosody or emotion indicate a right ear advantage (REA): faster response times and/or higher accuracy for language processing at stimuli presented at the right ear (Hugdahl, 2011). Differently, DL responses to emotional and/or prosodic stimuli show either diminished REAs or a left ear advantage (LEA) (Godfrey and Grimshaw, 2015; Grimshaw et al., 2003).

Few dichotic listening (DL) experiments have researched the effects of anxiety on emotional speech processing (Gadea et al, 2011). They either use speech/prosody as an emotion-eliciting stimulus or use DL mainly as an attentional manipulation technique (e.g. Leshem, 2018; Peschard et al., 2016; Sander et al., 2005). As a result, they are limited in the extent to which they reveal the relationship between dynamic variations in emotion language processing (prosody/semantics). Instead, studies focusing on the dynamic properties of emotional language, whether using DL or not (e.g. measuring

laterality through electrophysiological measures), do not tend to consider individual differences (e.g. Godfrey and Grimshaw, 2015; Grimshaw et al., 2003; Kotz and Paulmann, 2007; Paulmann et al., 2012; Techentin et al., 2009; Wambacq and Jerger, 2004). Therefore, on one side of the picture speech stimuli are typically treated as generic threatening stimuli, so possible differences induced by the informational features of speech that may vary over time are overlooked. On the other side, participants are typically regarded as a homogeneous group, so possible differences induced by anxiety-related processing, that may vary over time and may differ across informational features. Another important thing to consider is that in natural speech, emotional prosody might not be constrained to a single word, as is the case in the experimental manipulations used by most studies cited above. However, semantics is always constrained by sentence's structure and lexical meaning. In other words, while a lexical item needs to be identified within a sentence in order for emotional semantics to be recognized, prosody might be expressed from the beginning of a sentence. This makes difficult to generalize from word level, or highly controlled sentences, to real world emotional utterances.

To address these issues, two web-based DL experiments were designed, using semi-naturalistic sentences in order to ensure dynamic language processing beyond the single word level. Participants were asked to discriminate between neutral and threatening sentences (the latter expressing threat via semantics, prosody or both), in a direct-threat condition: identifying whether a threatening stimulus was presented to the left or right ear, and in an indirect-threat condition: identifying whether a neutral stimulus was presented the left or right ear. Participant's anxiety level was measured by using a psychometric scale. By so doing, the present approach is able take advantage of past studies researching the attentional effects of threatening language on anxiety and of studies researching the dynamics of speech's informational properties within a single study. Also, as both speech processing and anxiety literature seem to converge on theoretical perspectives incorporating multistep models (processing time-course), two experiments were designed to tap into different points in processing for which individual variation in anxiety may affect speech. In particular, experiments were aimed at differentiating responses made at late evaluative stages (delayed response) vs. responses made at earlier attentive stages (online response) as early over-attention to threat (Bar-Haim et al., 2007) might affect earlier prosody/semantic lateralization patterns (Kotz and Paulmann, 2011), and later over-engagement with threat (Bar-Haim et al., 2007) might

affect later emotional language evaluation stages (Kotz and Paulmann, 2011). Thus, Experiment 1 required participants to wait until after sentences' offset to respond (delayed response), and Experiment 2 required participants to respond during sentence presentation (online response).

For Experiment 1 it is hypothesized that anxious over-engagement with threat at mid-late evaluative stages (Bar-Haim et al., 2007) should increase left hemisphere (LH) engagement (Spielberg et al, 2013), disturbing possible LH to right hemisphere (RH) information transferring (Grimshaw et al., 2003; Kotz and Paulmann, 2011). Hence, a left ear advantage (LEA) is predicted; usually observed in DL experiments as an effect of prosody/emotional stimuli (Godfrey and Grimshaw, 2015; Grimshaw et al., 2003), this LEA should decrease as a function of anxiety, especially for semantic threat. This implies slower and less accurate responses for anxious people at their left ear when responding to semantically threatening but prosodically neutral stimuli (here referred as Semantic stimuli). As present sentences are semi-naturalistic, they have varied durations, but are long on average (~2s). This implies that answering after sentence's offset emphasizes late stage processing, understood to start at around 400ms (Kotz and Paulmann, 2011), followed by deliberation (~600ms). This late stage could be sustained for a long period of time, as it is characterized by a cyclic BIS process (McNaughton et al., 2013). Hence, if trait anxiety extends deliberation through excessive worry, then responses locked to sentence's offset should be slower.

For Experiment-2 it is expected that, as responses are forced to be faster (online), prosody should induce the most noticeable effects, as online responses may overlap with early-mid emotional processing stages (Kotz and Paulmann, 2011). Therefore, it is hypothesized that higher anxiety should reduce LH involvement (Spielberg et al., 2013) due to over-attention to threat effects, characteristic of earlier-mid processing stages (Bar-Haim et al., 2007). Hence, predict an enhanced LEA for highly anxious participants is predicted, especially for prosodically threatening but semantically neutral stimuli (here named Prosody stimuli). Thus, faster and more accurate responses for anxious people at their left ear when attending prosodic stimuli. In other words, as participants are required to answer as fast as possible, and prosody is readily identifiable in each sentence, but semantics required the identification of lexical items, processes before quick responses (~100, ~200ms) should take precedence for prosody, while semantics

might be affected by later processes (~400ms) as responses could be naturally slower independent of anxiety.

5.2 Methods

5.2.1 Participants

Participants were recruited using Prolific (prolific.ac). Only participants reporting being right-handed, having English as first language, without hearing and neurological/psychiatric disorders, and using only a desktop or laptop to answer the experiment were recruited. For Experiment 1, after exclusion, due to poor accuracy (below 70%) or not finishing the task properly, 44 participants (mean age = 31.7, 27 females) were retained (26 excluded). For Experiment 2, accuracy rejection threshold was relaxed to 60% (slightly closer to chance); thus 24 participants were excluded and 52 participants (mean age = 31, 24 females) were retained. Participants were remunerated on a £7.5/hour rate. All participants gave their informed consent before participating. It is important to clarify, the web-based nature of the experiment implies that task compliance levels could be low, as there is no direct control over participants meeting requested requirements (e.g. appropriate headphones) or performance (e.g. answering randomly). For this reason, and also to avoid issues related with possible impulsive behaviour or to age-related audition loss, only participants well above the adolescence threshold and amply below critical ages for audition loss were accepted, namely participants between 24 and 40 years old. Participants were informed that all their data is managed by UCL and 1998 data protection act protocols. Note that accuracy exclusion thresholds vary, as the second experiment is more difficult. Also, sample sizes are based on previous literature (e.g. Leshem, 2018; Peschard et al., 2016), and resources availability.

5.2.2 Materials

Four types of sentences were recorded: Prosody (neutral-semantics and threatening-prosody), Semantic (threatening-semantics and neutral-prosody), Congruent (threatening-semantics and threatening-prosody), and Neutral (neutral-semantics and neutral-prosody). Details about sentence selection and semantic and prosodic properties are widely discussed in Chapter 4. As a brief reminder, sentences were recorded in an acoustically isolated chamber using a RODE NT1-A1 microphone by

a male English speaker. The speaker was not a professional actor or voice actor (i.e. untrained speaker). The speaker was instructed to speak in what he considered his own angry threatening/angry or neutral voice for recording Prosody/Congruent and Semantic/Neutral sentences respectively. Sentences were not repeated across type (i.e. each type has a unique set of sentences). Neutral dichotic pairs were also unique across conditions (480 different sentences). Due to a technical problem several sentences were recorded with very low amplitude. Therefore, sentences were normalized and cleaned from noise in Audacity (audacityteam.org).

Next, sentences were paired using Audacity: sentences were paired such as their durations were as similar as possible. Silences between words were extended, never surpassing 40ms, to match sentences' latencies as closely as possible. After this, sentences were allocated to one of the stereo channels (left or right) of the recording; each pair was copied with mirrored channels. A silence (~50ms) was placed at the beginning and at end of each pair. This resulted in a total of 480 pairs where 80 sentences of each type (congruent, semantic, prosody) were each paired with a neutral sentence of the same length twice, so every sentence was presented once at each ear. Sentences' average length is 1720.65ms, and their prosodic features (Chapter 4) are consistent with previous DL experiments (Godfrey and Grimshaw, 2015; Grimshaw et al., 2009).

Table 5.1. Average number of words, duration and reaction time per stimulus type

Type	Words Threat	Words Neutral	Stimulus Duration	Delayed RT	Fast RT
Congruent	4.44 (0.88)	4.9 (0.96)	1744.49 (321)	535.32 (108)	1218.47 (148)
Prosody	4.45 (1.04)	5.01 (0.81)	1853.55 (256)	639.34 (104)	1328.92 (159)
Semantic	4.41 (1.03)	4.4 (1.05)	1554.44 (364)	590.65 (91)	1207.93 (155)

Note. Standard deviation appears in brackets. Duration and reaction times (RT) are expressed in milliseconds (ms).

5.2.3 Procedure

Before starting the experiments, participants answered the Penn State Worry Questionnaire (PSWQ) (Meyer et al., 1990) to assess their worry-level, and the Anxious Arousal sub-scale of the Mood and Anxiety Symptoms Questionnaire (MASQ-AA) (Watson et al., 1995) to assess their arousal level. This follows previous approaches (Nitschke et al., 1999), with the difference that PSWQ scores are used as a continuous predictor instead of splitting participants between high and low anxiety groups. For Experiment 1, PSWQ results indicated a distribution which is varied enough in terms of worry level

(mean = 47.31, median = 48.0, range [33, 67]). For experiment 2, PSWQ scores show similar properties (mean = 45.22, median = 45.0, range = [26,61]). PSWQ measures worry in a scale ranging from 16 to 80 points (median = 48 points), showing a consistent normal distribution in tested samples (mean close to median, as in our samples), and has been shown to have high internal consistence and validity (for details see: Meyer et al., 1990). MASQ-AA scores indicate that participants showed low levels of arousal, as none of them marked above the median. According to previous literature (e.g. Heller et al., 1997; Nitschke et al., 1999; Spielberg et al., 2013), high scores of MASQ-AA would be indicative of trait anxious arousal (hyperarousal), while high scores of PSQW would indicate trait anxious apprehension. Therefore, the present sample does not include participants with high or trait hyperarousal. So, only PSWQ scores were included in the analyses.

After a practice session, participants were randomly assigned to a list containing half of the total number of dichotically paired sentences (threat-neutral pairs) per threatening type (Prosody|Neutral, Semantic|Neutral, Congruent|Neutral), that is 40 pairs per type (120 in total). Sentences' lists were created previous to the experiment using randomly selected sentences from the total pool. Sentences were presented randomly to participants. In one half of the study they were instructed to indicate at which ear they heard the threatening sentence by pressing the right or left arrow keys (direct-threat condition). In the other half of the study they were instructed to respond in the same way, but indicating which ear they heard the neutral sentence in the dichotic pair (indirect-threat condition). This was intended to address attention effects (Aue et al., 2011; Peschard et al., 2016). Starting ear (left or right) and starting condition (direct- or indirect-threat) were counterbalanced. Participants were told to answer, as fast as possible, only when the sentence finished playing and a bulls-eye (target) image appeared on the screen. A 1400ms inter-stimulus-interval (ISI) was used, and the target image stayed on the screen during this period. For Experiment 2, participants were instructed to answer, as fast and as accurately as possible, before the sentence finished playing, and to withhold any response when a stop sign image appeared on the screen after sentences' end.

5.2.4 Analysis

Reaction time (RT) data were recorded in milliseconds, locked to sentence's offset. Accuracy was coded as correct=1 and else=0 (including misses and false alarms).

Participants with hit rates below 70% were excluded, as lower thresholds are too close to chance. This is mainly due to the nature of web-based experiments, where compliance levels cannot be more directly controlled. Thus, using too low exclusion criteria (e.g. ~50% or chance) is not methodologically warranted, as this would make difficult to identify how much variance or bias is added by unidentified non-compliance. Moreover, by setting a higher criterion for inclusion, it is ensured that participants are understanding the sentence content sufficiently for the various proposed stages of processing to occur. Two Bayesian hierarchical models were built for reaction time (RT) and accuracy. Models are described in detail in Chapter 3, but a reminder diagram of the present RT model can be seen on Figure 5.1. For Experiment 2, the same robust regression model was applied without duration, which would be inappropriate as participants answer before offset (incomplete sentence's duration).
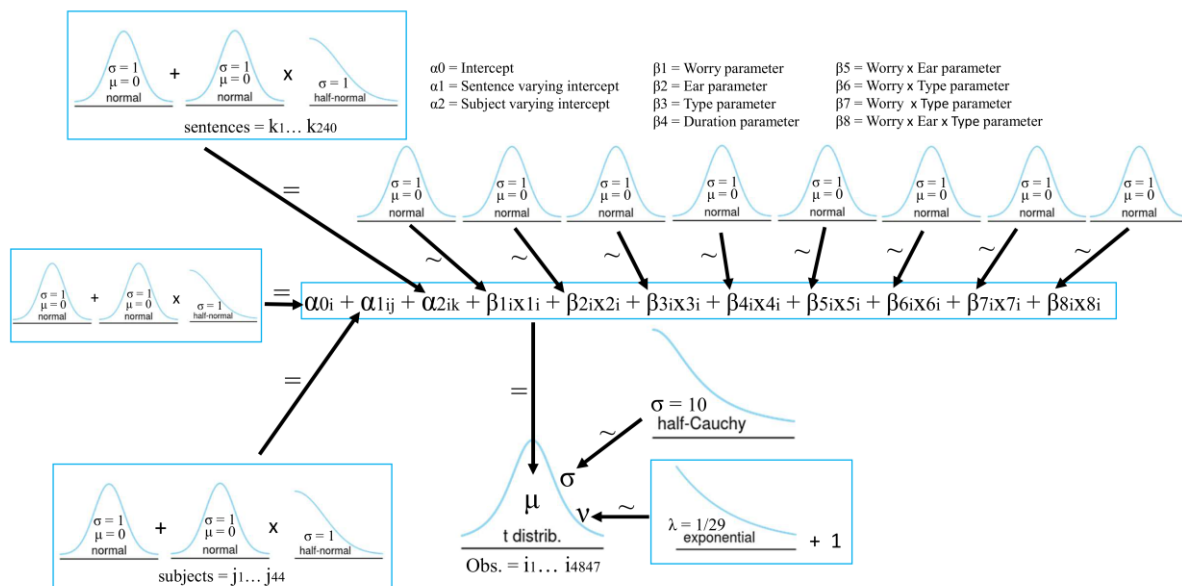
Figure 5.1. Diagram representation of hierarchical robust regression model. Arrows indicate the relationship between a parameter and priors/hyperpriors, where tilde (~) indicates a stochastic relationship and equal (=) indicates a deterministic relationship. Observations (Obs.) represent RTs in milliseconds.

As shown in Chapter 3, accuracy used a similar structure with Bernoulli distribution for observed likelihood. RT models used a robust regression (Kruschke, 2015) in order to account for outliers through a long-tailed Student-t distribution. In this way, RTs that are implausibly fast or implausibly slow do not need to be removed, but can be dealt with statistically. Both accuracy and RT models were sampled using Markov

Chain Monte Carlo (MCMC) No U-turn Sampling (NUTS) as provided by PyMC3 (Salvatier et al., 2016). Two chains of 1000 tuning steps and 1000 samples each were used. Plots and model comparisons were produced using Arviz (Kumar et al., 2019) and Matplotlib (Hunter, 2007). A region of practical equivalence (ROPE) of 2SDs, compared with a 90% high posterior density interval (HDI), was established as main criterion for deciding whether posterior distributions indicated strong or weak effects/differences.

In other words, 90% HDIs which fall completely outside the ROPE are considered to have very high chances of never overlapping with a distribution centred around zero and spanning the ROPE, namely a distribution indicating no change from one condition to another (if categorical) or no change as a variable progress (if continuous). Otherwise, HDIs partially overlapping with ROPEs indicate that chances of distributions being equivalent increases, thus when HDIs completely overlap a ROPE it can be considered that both distributions are indeed equivalent and that the variable had no effect or an extremely negligible effect. The reader is urged to interpret ROPEs as a heuristic of thresholding, but not as an "effect barrier". This implies a basic science interpretation where the interest is in continuous changes in nature rather than an applied science one, where threshold decisions are necessary (see Kruschke, 2018). Note that when HDIs are narrower than ROPEs, this indicates high precision of the estimates.

5.3 Results

5.3.1 Experiment 1: Delayed Response

All models sampled properly ($\hat{R} \cong 1$, ESS > 400, BFMIs > 0.6); energy plots, traceplots and autocorrelation plots also indicate excellent convergence. In addition, HDIs widths indicate high precision of the estimates, as each HDI is narrower than its associated ROPE. Plots and results from these checks, including raw data and full summaries of parameters and conditions, can be found in an Open Science Framework (OSF) repository (https://osf.io/z8pgf/), also associated with present chapter's published version (Busch-Moreno et al., 2020a).

Results for accuracy models are summarized in Table 5.2: Direct-threat, and Table 5.3: Indirect-threat. Tables show summaries for each condition at the lowest and highest worry levels (PSWQ score). As an important reminder: when variables are included in an interaction, their parameters are not free anymore, so main effects (lower-order effects) cannot be understood independently. For the present model, all effects are modulated by

the Worry by Ear by Type interaction. Indeed, when taking worry level into account, slopes show HDIs widely overlapping with zero and/or ROPEs. This makes safe to conclude that evidence supporting accuracy effects is not strong (i.e. weak or negligible effects).

Table 5.2. Experiment 1. Direct-threat task, accuracy logistic regression slopes

			Worry Slopes				
Stimulus Type	Ear	PSWQ Score	Posterior Mean	Posterior SD	HDI %5	HDI 95%	Probability %
Semantic	Left	33	1.00	0.67	-0.08	2.13	73.12
Prosody	Left	33	0.07	0.66	-0.94	1.24	51.75
Semantic	Right	33	0.74	0.56	-0.21	1.65	67.66
Prosody	Right	33	-0.01	0.56	-0.93	0.91	49.64
Semantic	Left	67	2.03	1.37	-0.17	4.32	88.41
Prosody	Left	67	0.14	1.34	-1.91	2.51	53.54
Semantic	Right	67	1.50	1.14	-0.43	3.35	81.74
Prosody	Right	67	-0.03	1.14	-1.89	1.85	49.27

Note. Posterior estimates are expressed in log-odds, the rightmost column contains the mean derived probability.

Table 5.3. Experiment 1. Indirect-threat task, accuracy logistic regression slopes

			Worry Slopes				
Stimulus Type	Ear	PSWQ Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%	Probability%
Semantic	Left	33	-0.43	0.66	-1.49	0.66	39.51
Prosody	Left	33	-0.86	0.62	-1.95	0.12	29.72
Semantic	Right	33	1.06	0.56	0.12	1.92	74.35
Prosody	Right	33	-0.19	0.55	-1.12	0.66	45.29
Semantic	Left	67	-0.87	1.34	-3.02	1.33	29.63
Prosody	Left	67	-1.75	1.26	-3.97	0.25	14.84
Semantic	Right	67	2.16	1.14	0.24	3.89	89.66
Prosody	Right	67	-0.38	1.12	-2.28	1.34	40.53

Note. Posterior estimates are expressed in log-odds, the rightmost column contains the mean derived probability.

Results for reaction time (RT) data in both the direct-threat and indirect-threat tasks indicated strong effects of worry level. Table 5.4 and Table 5.5 summaries show that worry level did not have particularly strong differences between ear or type effects. Slope estimates for worry, however, indicate a strong effect of worry. Increases from lowest worry level (33 PSWQ points) to highest (67 PSWQ points) are almost the same across conditions (~290ms to ~309ms); a negligible difference when considering error. Note

that seemingly slower responses to Semantic can be disregarded due to HDIs overlapping. In short, independent of type or ear, estimates indicate that RT increases around 8ms (± ~0.3ms) per PSWQ score point. At the highest worry level participants answer around 300ms later than at the lowest worry level (see Figure 5.2).

Table 5.4**.** Experiment 1. Direct-threat task, reaction time robust regression estimates

			Worry Slopes			
Stimulus Type	Ear	PSWQ Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%
Semantic	Left	33	285.21	8.95	271.01	299.88
Prosody	Left	33	280.51	9.98	265.42	298.46
Semantic	Right	33	280.06	7.97	267.14	293.42
Prosody	Right	33	266.88	8.86	253.23	282.11
Semantic	Left	67	579.06	18.16	550.23	608.84
Prosody	Left	67	569.52	20.27	538.88	605.96
Semantic	Right	67	568.61	16.18	542.37	595.73
Prosody	Right	67	541.85	17.99	514.14	572.77

Note. Posterior estimates are expressed in milliseconds (ms).

Table 5.5. Experiment 1. Indirect-threat task, reaction time robust regression estimates

			Worry Slopes			
Stimulus Type	Ear	PSWQ Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%
Semantic	Left	33	296.34	9.25	280.50	310.79
Prosody	Left	33	283.14	10.50	264.45	298.82
Semantic	Right	33	288.38	8.02	275.42	301.63
Prosody	Right	33	285.20	9.34	270.94	301.46
Semantic	Left	67	601.66	18.77	569.50	630.99
Prosody	Left	67	574.85	21.31	536.91	606.70
Semantic	Right	67	585.50	16.28	559.18	612.40
Prosody	Right	67	579.05	18.97	550.09	612.06

Note. Posterior estimates are expressed in milliseconds (ms).
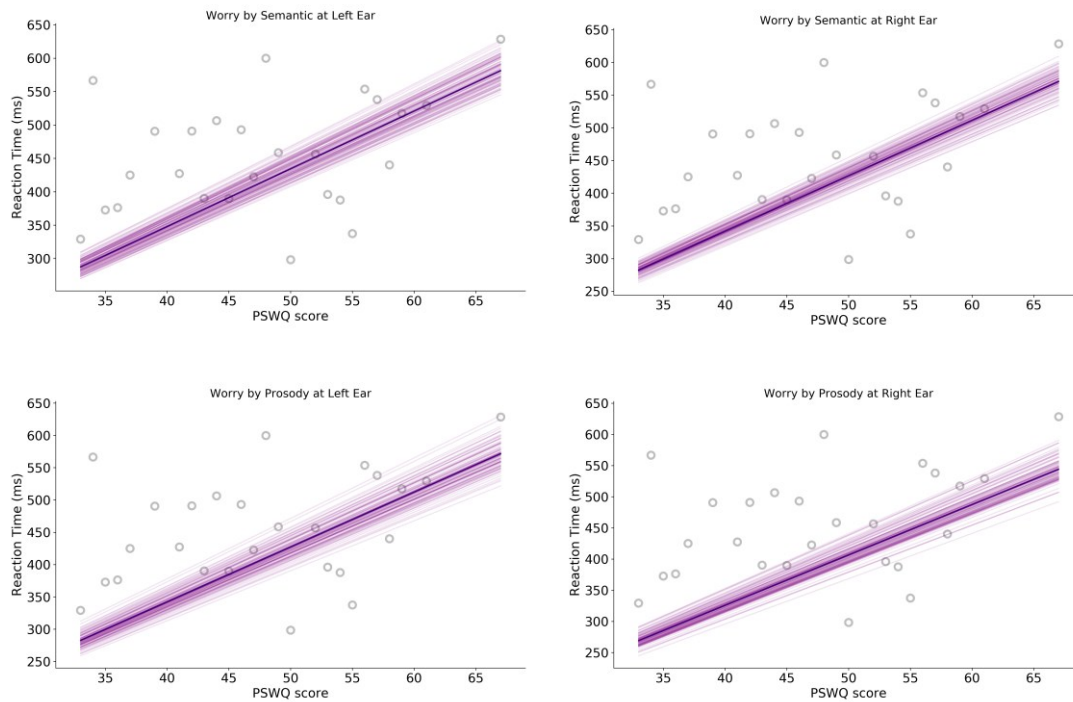
Figure 5.2. Experiment 1 (delayed responses), direct-threat regression lines. Images show posterior distributions across worry levels by Semantic and Prosody at left and right ears for both conditions. Faded lines are samples from the posterior and indicate uncertainty. Grey circles indicate total RT average across worry level (independent of condition). Note that certainty of the estimates is good, and that differences between conditions are small. This indicates that, independent of condition, reaction times substantially increased as worry level increased.

Duration had no relevant impact on results, showing very small increases in accuracy and very small decreases in RTs. Only the indirect-threat RT task shows an HDI outside ROPE indicating a considerable but small effect of around 0.03ms RT decrease per 1ms increase in sentence's duration. So, the maximum duration difference in present stimuli (~1800ms) barely decrease RT by ~50ms from the shortest to the longest sentence. Given the huge variability in sentences' lengths, these effects are negligible.

Overall, results from Experiment 1 (delayed responses) indicate that whether participants answer to threat directly (pressing a button to indicate which ear the threatening sentence was presented to) or indirectly (indicate which ear the neutral sentence was presented to), they are similarly accurate. There are negligible effects of ear and type due to variability in uncertainty and error. Experiment 1's results for RTs also indicate that small effects of Semantic stimuli type, slowing down RTs for higher worry, can also be considered negligible due to overlapping HDIs. RTs are strongly affected by

worry level, where at both direct- and indirect-threat tasks responses increase about eight milliseconds per PSWQ score point.

5.3.2 Experiment 2: Fast Response

Models for Experiment 2 did not include duration in the regression, as the relationship between duration and worry level is not clear, as participants must always answer before the end of each sentence. Again, all effects show good precision and all models sampled properly ($\hat{R} \cong 1$, $\text{ESS} > 400$, $\text{BFMIs} > 0.6$), with energy plots, traceplots, and autocorrelation plots showing excellent convergence (for images see the OSF repository: https://osf.io/z8pgf/).

Accuracy results are summarised in Tables 5.6 and Table 5.7. Note that the direct-threat results seemingly high effect of highest anxiety participants (61 points) of dis-preferring their Right ear for Prosody with a 9.7% probability still slightly overlaps the ROPE, and given this, it cannot be considered good evidence in support for a Prosody right ear disadvantage. Similar conclusions can be drawn for the 12.6% probability for higher worry responses to Prosody at left ear in the indirect-threat task (Table 5.7). All other effects are more clearly near 50% probability, or zero log-odds, with HDIs spanning zero and ROPEs. All other effects are more clearly near 50% probability, or zero log-odds, with HDIs spanning zero and ROPEs.

Table 5.6**.** Experiment 2. Direct-threat task, accuracy logistic regression slopes

			Worry Slopes				
Stimulus Type	Ear	PSWQ Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%	Probability%
Semantic	Left	26	0.35	0.37	-0.24	0.98	58.59
Prosody	Left	26	0.11	0.38	-0.48	0.76	52.63
Semantic	Right	26	-0.62	0.30	-1.07	-0.07	35.09
Prosody	Right	26	-0.95	0.30	-1.49	-0.48	27.87
Semantic	Left	61	0.81	0.88	-0.57	2.29	69.29
Prosody	Left	61	0.25	0.90	-1.12	1.78	56.15
Semantic	Right	61	-1.44	0.71	-2.52	-0.18	19.11
Prosody	Right	61	-2.23	0.71	-3.50	-1.14	9.70

Note. Posterior estimates are expressed in log-odds, the rightmost column contains the mean derived probability.

Table 5.7. Experiment 2. Indirect-threat task, accuracy logistic regression slopes

Stimulus Type	Ear	PSWQ Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%	Probability%
			Worry Slopes				
Semantic	Left	26	-0.47	0.34	-1.01	0.07	38.51
Prosody	Left	26	-0.83	0.34	-1.35	-0.21	30.44
Semantic	Right	26	-0.48	0.27	-0.90	0.00	38.16
Prosody	Right	26	-0.10	0.28	-0.54	0.37	47.54
Semantic	Left	61	-1.10	0.79	-2.38	0.17	25.01
Prosody	Left	61	-1.94	0.81	-3.17	-0.50	12.58
Semantic	Right	61	-1.13	0.64	-2.11	0.00	24.37
Prosody	Right	61	-0.23	0.66	-1.26	0.87	44.24

Note. Posterior estimates are expressed in log-odds, the rightmost column contains the mean derived probability.

Results from RT data are summarised in Table 5.8 and Table 5.9. These indicate a more consistent, but small, effect of Prosody at the highest worry level in the direct-threat task. This effect is independent of ear (but slightly stronger at the left), where increases from lower (26 points) to higher worry (67 points) at the left ear are ~369ms for Semantic but ~418ms for Prosody, and ~368ms for Semantic and ~400ms for Prosody at the right ear. However, estimates still indicate this as weak effects, as HDIs of Prosody at higher worry still overlap with HDIs of Semantic at higher worry. Furthermore, these effects tend to fade in the indirect-threat task. Hence, the strong general effects of worry must be emphasised, which indicate a strong increase of about 400ms, independent of ear or stimulus type, from the lowest to the highest worry level (see Figure 5.3).

Table 5.8. Experiment 2. Direct-threat task, RT robust regression estimates

Stimulus Type	Ear	PSWQ Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%
			Worry Slopes			
Semantic	Left	26	274.43	18.02	244.75	304.10
Prosody	Left	26	310.51	18.47	280.63	340.87
Semantic	Right	26	273.28	17.36	245.23	302.20
Prosody	Right	26	296.75	17.82	267.89	326.86
Semantic	Left	61	643.86	42.27	574.21	713.46
Prosody	Left	61	728.49	43.34	658.39	799.74
Semantic	Right	61	641.16	40.73	575.34	709.01
Prosody	Right	61	696.23	41.81	628.52	766.87

Note. Posterior estimates are expressed in milliseconds (ms).
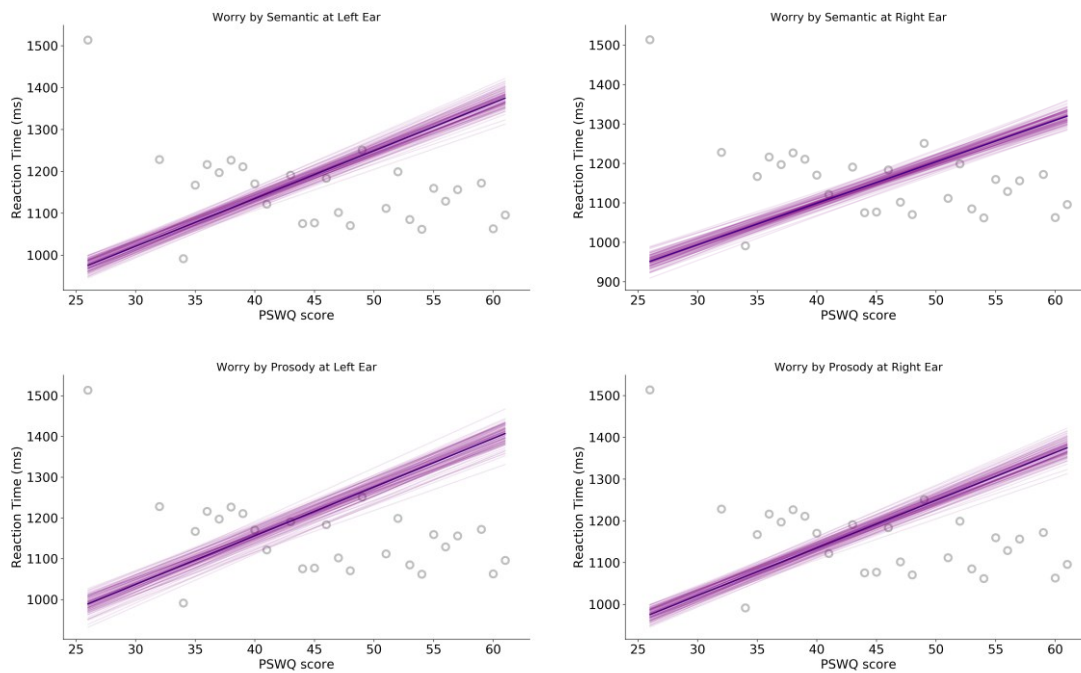
Figure 5.3. Experiment 2 (fast responses), direct threat regression lines. Images show posterior distributions across worry levels by Semantic and Prosody at left and right ears for both conditions. Faded lines are samples from the posterior and indicate uncertainty. Grey circles indicate total RT average across worry level (independent of condition). Note that certainty of the estimates is good, and that differences between conditions are small. This indicates that, independent of condition, reaction times substantially increased as worry level increased.

Table 5.9. Experiment 2. Indirect-threat task, RT robust regression estimates

			Worry Slopes			
Stimulus Type	Ear	PSWQ Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%
Semantic	Left	26	299.96	16.99	269.96	325.15
Prosody	Left	26	308.13	17.27	281.61	337.45
Semantic	Right	26	280.37	16.00	256.25	308.77
Prosody	Right	26	307.72	16.07	280.61	332.93
Semantic	Left	61	703.75	39.87	633.37	762.85
Prosody	Left	61	722.91	40.52	660.71	791.71
Semantic	Right	61	657.79	37.54	601.21	724.42
Prosody	Right	61	721.96	37.71	658.36	781.10

Note. Posterior estimates are expressed in milliseconds (ms).

5.4 Discussion

Results from the delayed response experiment (Experiment 1) indicate small effects of worry on accuracy and strong effects of worry on RT. A noticeable feature is an

increase in accuracy for Semantic in the indirect-threat task, where participants prefer the right ear for semantic independent of worry level (anxious people are slightly more accurate). These effects, however, are very weak and associated with greater uncertainty. The same happens for effects of worry on type. Therefore, the most sound and grounded inference from present results is that worry mainly affects RT in a directly proportional relationship, and does so for any type of threatening stimuli (Prosody, Semantic or Congruent) and independent of ear presentation. These results are generally echoed by the fast response experiment (Experiment 2). To note, higher worry accuracy in the direct-threat task seems to decrease when they answer to Prosody at their right ear, but in the indirect-threat task this happens at the left ear. This reversal is to be expected due to threat-direction, but again these effects are small and somewhat uncertain. A clearer effect of greater reaction times for Prosody as a function of worry level was observed for the direct-threat task of the fast experiment. Although this effect was more reliable than previous Type effects, it was small.

Some effects become more evident when their slopes are taken into account separately, such as a general accuracy increase at left ear for Prosody in the direct-threat task of Experiment 1 (reversed for indirect-threat to a decrease as a function of worry), and an accuracy increase for Prosody at left ear as a function of worry in the direct-threat task of Experiment 2 (also reversed to a decrease for indirect-threat as a function of worry); which can be expected in a dichotic listening task and are partially consistent with present predictions (for plots and addition materials see the OSF repository: https://osf.io/z8pgf/). Nevertheless, for present purposes the most conservative approach is taken. That is, the most straightforward interpretation is upheld; interpretation which takes all effects into account simultaneously due to their involvement into an interaction. This simply indicates that in both experiments higher worriers tend to answer slower to any type of threatening stimuli. Namely, for each point increase of worry score, reaction times increase around eight milliseconds. This implies that this Chapter's hypotheses indicating ear preferences as distinct by type cannot be supported, nor the hypothesis on differing effects of fast responses. However, the main hypothesis indicating a strong effect of worry level (trait anxiety) receives strong support from present results.

In order put the present results in context, it is important to recapitulate important aspects that differentiate the current experiments from previous relevant studies: 1) The

use of worry-level as a continuous variable. Worry is associated with anxious apprehension (Heller et al., 1997), which implies more chances of participants over-engaging with threat. 2) Stimuli were semi-naturalistic sentences, providing stronger contextual effects. In addition, their longer durations can facilitate engagement with their content. 3) Information channels were manipulated to disentangle effects of semantics and prosody from effects of emotional expression (Kotz and Paulmann, 2007). 4) The use of two tasks measuring responses directed to threatening or neutral stimuli (direct vs. indirect threat, e.g. Sander et al, 2005) helps to check whether attention effects could be inducing different response patterns. 5) Two experiments were implemented to verify whether answering after sentences' end or during sentence presentations (delayed vs. fast) can influence laterality patterns by tapping into different moments of a multistep emotional language processing mechanism (Kotz and Paulmann, 2011).

With this in mind, it is important to carefully interpret the lack of laterality (ear) effects in Experiments 1 and 2. Weak ear effects might be explained by the great variability between items and the high duration (also very variable) of sentences. However, the lack of sensitivity of DL when more semi-naturalistic stimuli are provided cannot be discarded as a possible explanation. If DL effects are task dependent (Godfrey and Grimshaw, 2015), increased naturalness on stimuli can bring out a myriad of bilateral processing patterns that might make ear advantages disappear on the long run when prolonged auditory stimuli are listened to. Although there is previous evidence suggesting a right lateralized pattern for prosody vs. semantic evaluation in an EEG experiment (not considering anxiety), using a congruency (not DL) task with sentences as stimuli (Kotz and Paulmann, 2007), further experimentation using a similar paradigm has not observed this pattern (Paulmann and Kotz, 2012). Although this pattern is explained by the strong association between pitch recognition and RH engagement (Kotz and Paulmann, 2007; Zatorre et al., 2002), there are other frequency and spectral features that might be important for recognizing both threatening and neutral sentences (Banse and Scherer, 1996; Hammerschmidt and Jürgens, 2007; Xu et al., 2013; Zatorre et al., 2002). This could imply that distinguishing prosody and semantics might be a continuous process that can have diversified effects even during sentence presentation.

Indeed, by manipulating angry prosody changes at the beginning and end of sentences, an EEG study has observed that when prosody changes from angry to neutral within sentences, processing is more effortful (Chen et al., 2011). This might indicate that

the rich acoustic nature of prosody might be detected quickly but resourcefully. Recent EEG research has observed that anxious people present ERP differences at both early and late processing stages when answering to threatening prosody and non-language vocalizations (Pell et al., 2015). This is consistent with the notion of early over-attention and later over-engagement, and indicates that behavioural responses might change given early or late variations in threat.

Another possible explanation is callosal relay (Atchley et al., 2011; Grimshaw et al., 2003), where increased anxiety would disrupt RH to LH callosal information transferring of threatening prosody. It has been proposed that callosal relay is highly relevant for language informational and emotional processing (Friederici et al., 2007; Kotz and Paulmann, 2011; Steinmann et al., 2017). Hence, interference at one hemisphere (e.g. rumination or worry impacting LH) can have an effect on information transferring to the other. Thus, callosal relay effects could have a relevant impact on how DL tasks are processed, subject to both top-down and bottom-up effects (Westerhausen and Hugdahl, 2008), which is particularly relevant when laterality effects induced by acoustic or lexical properties need to be disentangled from those induced solely by emotional processing (Grimshaw et al., 2003; Leshem, 2018).

Nevertheless, strong effects of worry level (trait anxiety), affecting emotional language processing were observed, which may be due to over-engagement with threat (Bar-Haim et al., 2007; Spielberg et al., 2013). It was proposed that delayed responses facilitate over-engagement with threat due to the long latency between sentence presentation and response. This, together with the high variability in sentences' durations and content might have nullified ear and/or type effects. Furthermore, the present experimental set-up failed to observe any clear Type or ear effect when responses were forced to be fast (during sentence), besides a small effect of worry level on Prosody stimuli (slower RTs). Contrary to predictions, the pattern of Experiment 2 (fast response) is basically the same as in Experiment 1 (delayed response), which gives evidence against early and early-mid emotional language processing effects having a direct behavioural output (at least in the context of semi-naturalistic sentences).

Similarly, previous research using single words, dichotically presented as direct- and indirect-threat (or anger), and measuring anxiety, did not find differences in RT for left or right ears (Sander et al., 2005; Leshem, 2018; Peschard, 2016), but did find differences in attention focus per ear. Present results indicate that RTs differ in neither

of these conditions, which is supported by the remarkably similar posterior distributions for direct- and indirect-threat and the high certainty of these estimates. Recent research (Leshem, 2018) did not find effects of trait anxiety on ear either; present results, going even further, evidence a precise pattern of weak or negligible interactions between ear and worry (trait anxiety). Although the absence of other effects might be induced by stimuli's high variability in length and content, it is also important to emphasize that present analyses are fairly robust.

In addition to observing very weak or negligible effects of ear, results indicated weak effects of Type (i.e. Semantic or Prosody). Variations on Type parameter magnitudes might not necessarily indicate an effect of worry on particular stimulus types, but a trade-off between accuracy and RT for stimuli that are harder to recognize (Robinson et al., 2013). In the delayed response experiment (Experiment 1), Semantic stimuli are easily recognizable by finding the threatening lexical item within a sentence, but this might take longer to achieve, which impact the already slow reactions by participants with higher levels of worry. In the fast response experiment (Experiment 2), as responses are required to be executed as fast as possible before the sentence ends, higher worriers have no time to brood. Thus, possible earlier pre-attentive or attention effects (Bar-Haim et al., 2007) can be still observed as speeding-up the quick categorisation of lexical items as soon as they are identified within a sentence; there is no need to ponder on them while waiting for a sentence's end.

Given this, present results suggest that any type of threatening language, attended either directly or indirectly, strongly affects higher worriers when stimuli are sufficiently long. Therefore, the proposal of adding a fourth stage to a multistep model of emotional language (Kotz and Paulmann, 2011) is partially supported: trait anxiety indeed affects threatening language processing in a way that response times strongly and consistently increase. In other words, the more participants approach a state of trait anxiety, the slower their responses will be. A very plausible explanation for this phenomenon is verbal repetitive thinking, which can be also associated with higher levels of rumination and/or worry as a feature of anxious apprehension (Nitschke et al., 1999; Spielberg et al., 2013), or as a marker of an over-reactive behavioural inhibition system (Corr and McNaughton, 2012). The long duration of present stimuli might have been a decisive factor for inducing a strong effect of worry.

Over-engagement with threat does not need to be in the form of verbal repetitive thinking, but the strong slow-down in responses induced by present threatening speech stimuli suggest that language processing might be specially affected by higher worry, widely associated with verbal repetitive thinking (McEvoy et al., 2010). Previous research has found that non-language simple threatening stimuli (e.g. noise) can indeed induce similar over-engagement effects (slower RTs) in association with BIS but not with trait anxiety (Massar et al., 2011), while other studies indicate that induced anxiety slows down RTs irrespective of stimulus emotional content (Aylward et al., 2017). However, these studies focus on short duration stimuli and compare very short RT differences, in the order of tens of milliseconds. Present findings arise from prolonged exposure to language stimuli and indicate RT increases in the order of hundreds of milliseconds as a function of anxiety. In addition, results indicate that when the task requires to identify the Neutral (not threatening) sentences from the dichotic pair, RT increases are almost equivalent to the task requiring to directly identify the Threatening one of a pair. This indicates that attention effects are not playing a direct role in current responses, neither induce relevant nor sufficiently big indirect effects. It might be that the extended nature of sentences implies that participants have enough time for advancing from early attention to late deliberation phases.

Considering this, the initial assumption that a fast response experiment (Experiment 2) would be enough to identify difference at early processing stages was incorrect, at least given the present stimuli and task. The varied position of threatening lexical items and/or threatening intonation emphasis might cause a general slow-down of responses, as very specific features of sentences need to be identified and participants have time to do so (the whole extent of a sentence). Therefore, without time pressure, attention mechanisms cannot be posited as a plausible explanation for RT increases as a function of anxiety. While evaluation mechanisms could serve as an explanation, the fact that there are not strong effects associated with difficulties categorising of identifying stimuli makes them weak candidates. Differently, long semi-naturalistic speech stimuli might be especially effective in triggering late phase components, such as goal-orientation processing or deliberation. In such case, verbal repetitive thinking, as induced by worry, would be particularly effective for impairing responses to longer and more varied semi-naturalistic speech/language stimuli. Hence the strong association of worry with slower reaction times. Given this, verbal repetitive thinking is a parsimonious

explanation, which could account for patterns such as those of present experiments, and also develops as a promising hypothesis for future experimentation.

With all that being said, a general caveat of the present experimental approach lies on the nature of the experiment itself. Behavioural measures such as DL, though able to portray a very general picture of underlying brain processes, might not be enough. Better spatial and temporal resolution is required to disentangle laterality and early stage effects of threatening language. The latter is particularly relevant, as the time-course of emotional language processing might have crucial differences at much shorter time-scales, as evidenced by previous EEG research (Chen et al., 2011; Kotz and Paulmann, 2007; Paulmann and Kotz, 2012; Pell et al., 2015; Wabnitz et al., 2015; Wambacq and Jerger, 2004). In consequence, present tasks could be replicated by using EEG measures, in particular Experiment 1, where EEG measures such as event-related potentials could provide richer information about processing occurring during sentence listening, before response preparation and response execution. This could also provide lab results as point of comparison with present web-based results. But more importantly, this is crucial for identifying differences in the neural signature of worry and language processing, indispensable for properly understanding time-related models of language and anxiety processing.

In conclusion, present results indicate that extending multistep models of language processing (Schirmer and Kotz, 2006; Kotz and Paulmann, 2011) by including aspects of multistage models of anxiety (Bar-Haim et al., 2007; Corr and McNaughton, 2012) could be a relevant theoretical move. The current multistep model proposes three stages that can be understood as early (perception), mid (recognition), late (evaluation); or as pre-attentive, attentive and evaluative stages. A fourth orientative stage, associated with deliberation, can help to understand aspects of goal-directed processes before response. Late stages which could be particularly impaired by worry components of anxiety, as suggested but not ascertained by present evidence. Complementing this model, however, might be insufficient. Further theoretical development, including quantitative modelling, the inclusion of physiological correlates, and more precise anatomical mappings, might be necessary. Further experimental testing is thus required, in particular by implementing physiological measures such as EEG, and tasks that do not involve DL, using more controlled stimuli and investigating the effects of stimuli below or above the sentence level, such as phrases or narratives.

Chapter 6
Behavioural Evidence 2
Delayed Responses to Non-dichotic Threatening Speech

6.1 Introduction

In the previous chapter, results suggest that in a dichotic listening experiment, the effects of anxiety do not greatly impact ear advantages nor information type, but anxiety induces a big slow-down in reaction times. Slower responses are common finding in previous literature, mainly as a response associated with anxiety and not fear, and is interpreted as anxious people over-engaging with threat (Cisler and Koster, 2010). One of the aims of previous chapters and the present one is to provide a more consistent theoretical background that can better explain the effects of threatening language on anxiety. First, a clear distinction between fear and anxiety responses is required (Gray and McNaughton, 2000), as over-attention to threat might be associated more with fear than anxiety, and if associated with anxiety at all it should be in early processing stages. Second, a distinction between arousing anxious states and apprehensive anxious states (Heller et al., 1997); where apprehension (e.g. worry or rumination) might modulate later over-engagement with threat. Third, integrating phasic models of emotional language (Kotz and Paulmann, 2011) with phasic models of anxiety (Bar-Haim et al., 2007; Cisler and Koster, 2010) to understand how different anxiety-related phases affect speech processing phases. This integration should imply that emotional language models are extended from three phases (perception, recognition, evaluation) to four phases, where the latter occurs after emotional language has been evaluated (~500ms) and is associated with goal-orientation or deliberation. Anxious over-engagement should occur from the evaluation phase, and sustained during the deliberation phase. There, emotional language can be specially affected, as over-engagement can be expressed as repetitive thinking (worry and/or rumination), exhausting language processing resources (e.g. rehearsing the language stimulus) and inducing slower but not less accurate responses.

Consistent with this notion, dichotic listening studies tend observe this phenomenon (e.g. Peschard et al., 2016; Sander et al., 2005), similar to results from Chapter 5. There, participants listened sentences containing only prosodic threat with

neutral semantics (Prosody), only semantic threat with neutral prosody (Semantic), and both combined (Congruent). Anxiety was measured by the Penn State Worry Questionnaire (PSWQ) (Meyer et al., 1990). Results from Chapter 5 studies indicate that reaction times (RTs) to threatening language (semi-naturalistic threatening sentences) increase as apprehensive anxiety (worry) levels increase; but accuracy remains more or less constant, with an uncertain increase for Semantic and uncertain decrease for Prosody (so not allowing strong conclusions about accuracy). From these studies it was concluded that more anxious participants must have recognised threat efficiently but found difficult to disengage from it, as both delayed responses (after sentence) and fast responses (during sentence) induced similar RT delays but did not induced strong accuracy differences. This possible accuracy/RT trade-off has been proposed to occur only when stimuli are emotionally (i.e. negatively) loaded (Robinson et al., 2013). Indeed, recent studies indicate that over-engagement with threatening sounds occurs only when anxious participants have enough time to engage with threat, and is associated with decreased RTs as a function of trait anxiety when participants respond to naturalistic threatening sounds (Wang et al., 2019).

The focus of present chapter is on understanding whether an overactive behavioural inhibition system (BIS) could be associated with increased reaction times, as it was with worry in Chapter 5, even if responses to non-dichotic threatening sentences are required to be fast. This is relevant, because in previous chapter participants showed the same response pattern either in direct-threat (indicate ear of threatening pair) or indirect-threat (indicate ear of neutral pair) conditions. Thus, it is not clear whether accuracy would be also unimpaired in a task where stimuli identification is requested (e.g. responses to presence of threat or type of threat), providing a better accuracy measure (i.e. sensitivity and/or bias towards certain stimuli) in a go/no-go paradigm; which has been associated with anxious response inhibition effects (Neo et al., 2011; Robinson et al., 2013). In addition, this would allow to show whether an anxiety-related slow-down effect persists under less cognitively taxing conditions (i.e. non-dichotic). Importantly, this task should be a fast response (during sentence) as this will help to demonstrate that sentences on their own provide a sufficiently long time for participants to over-engage with threat. In short, the present aim is to understand whether a non-dichotic task requiring fast responses will show a clearer trade-off between accuracy and reaction times; namely, similar increases in reaction time but clearer non-differences in

accuracy as in Chapter 5's Experiment 2. To this aim, two web-based experiments were conducted, having the following features. 1) Experiments consist of non-dichotic sentence-type identification tasks, Experiment 1 (Prosody Experiment): participants answer to acoustic threat; Experiment 2 (Semantic Experiment): participants answer to threatening content. 2) Sentences were recorded by a different speaker to improve generalizability. 3) A full go/no-go paradigm is implemented, where participants answer whether sentences 'sound threatening' in the Prosody Experiment, or whether they 'have threatening content' in the Semantic Experiment, but refrain answers from neutral sentences. 4) A BIS scale from RST-PQ is used instead of the PSWQ, as this measure is better described in relation to physiological measures and has shown to be a reliable (Corr and Cooper, 2016). Given this, the present hypothesis states that anxiety should slow-down responses but should not affect accuracy in both tasks. This implies the following predictions: RTs should increase as a function of BIS scores; but accuracy should not change as a function of BIS scores.

6.2 Methods

6.2.1 Participants

Participants for the Prosody Experiment (n=40; age: mean=34.82, SD=9.99; 39 females) and for the Semantic Experiment (n=49; age: mean=36.06, SD=10.43; 26 females) were recruited using Prolific and completed the study on Gorilla (gorilla.sc). They reported not having hearing, psychological or neurological problems; having English as their first language; and being right-handed. Participants were compensated at a £7.50/hour rate. All the procedure counts with UCL's ethical approval and participant's data was handled according to GDPR protocol as informed to them. Note than sample size tends to be moderately bigger respect to previous studies researching anxiety in the go/no-go paradigm (e.g. Neo et al., 2011; Robinson et al., 2013), which may help to improve possible online-induced non-compliance- or error-related issues.

6.2.2 Materials

Sentences (52 per category) were selected by using the same procedure as in previous chapters (some were replaced and some retained), and recorded on an isolated anechoic chamber by a male native London English speaker using a RODE NT1-A1 microphone. Note that this speaker is different from previous (Chapter 5) and following

(Chapter 7) chapters, and is an amateur actor. Sentences' categories include: Prosody (only prosodic threat) and Semantic (only semantic threat), Congruent (both types of threat), and Neutral (no threat). Table 6.1 summarises sentences' average number of words and duration. Sentences' acoustic measures and lexical norms were analysed following the same procedure described in Chapter 4. However, only Prosody norms were taken for these stimuli; which indicate that's sentences are perceived as threatening and are accurately predicted by previous speakers' norms (from Chapter 5). These analyses' summaries and results can be found in the present chapter's open science framework (OSF) repository (https://osf.io/ptcr9/). Figure 6.1 shows oscillograms and spectrograms of example sentences on each category. Table 6.2 shows Bayesian estimation supersedes the t-test (BEST) estimated means and SDs, and Figure 6.2 shows BEST differences between means. Note the consistency of these results with BEST results from Chapter 4. Note that participants rated previous sentences (Chapter 4, Study 2) as both low in valence and high in arousal in addition to high in threat. This may indicate that either present speaker is closer to a non-trained speaker or that a non-trained or semi-trained speaker may produce threatening prosody in a similar manner.
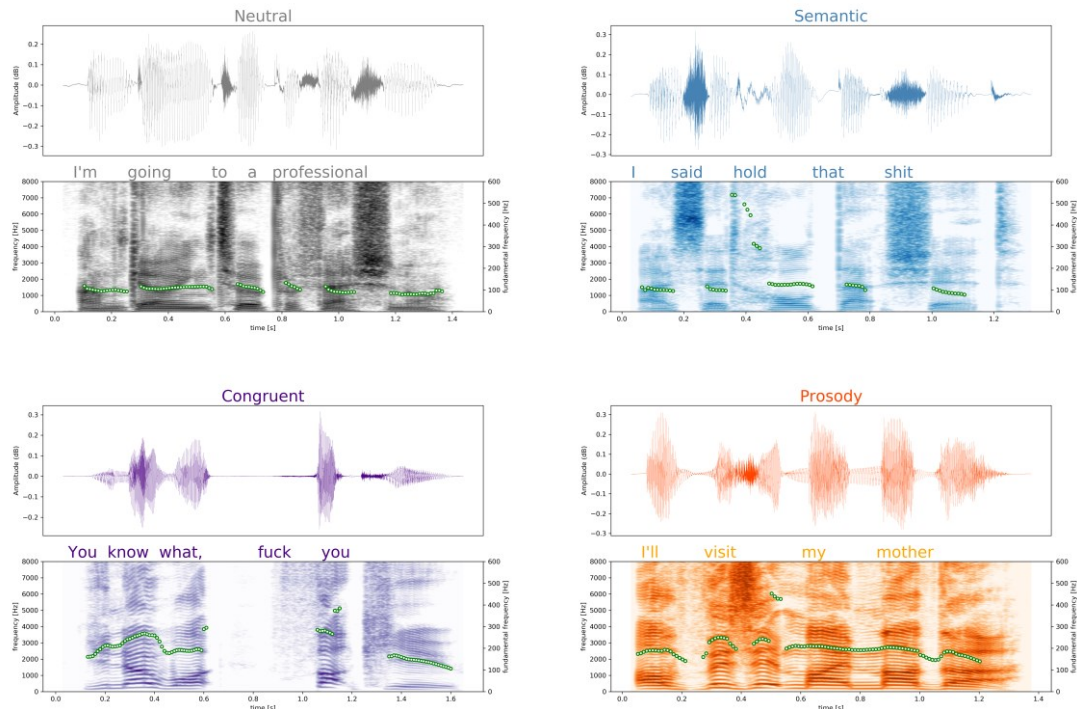
Figure 6.1. Example of four sentences used in this study. Top of each image: oscillogram showing amplitude changes. Bottom of each image: spectrogram showing frequency changes. Top left: neutral prosody and neutral semantics (Neutral). Top right: threatening prosody and threatening semantics (Congruent). Bottom left: neutral prosody and threatening semantics (Semantic). Bottom right: threatening prosody and neutral semantics (Prosody). Green dots: fundamental frequency (F0) contours.

Table 6.1. Sentences average number of words and duration

Type	#Words Mean	#Words SD	Duration Mean	Duration SD
Congruent	5.16	1.03	1519.08	233.67
Neutral	4.61	1.03	1471.37	257.65
Prosody	5.02	0.73	1509.69	172.19
Semantic	5.16	1.01	1518.20	215.96

Figure 6.2. Images show differences between means of lexical norms and acoustic measures. Left to right: Arousal, Valence, Median Pitch, and Hammarberg Index. Ridgeplots show the highest posterior density intervals (HDIs) distributions (kernel density). Grey bands indicate 2SDs regions of practical equivalence (ROPE).

Table 6.2. BEST estimated average posterior distributions for Mean and SD

Measure	Arousal		Valence		Median Pitch		Hammarberg Index	
Category	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Neutral	3.96	0.67	5.94	0.91	109.17	4.74	22.74	4.13
Congruent	5.96	0.89	3.49	1.26	197.89	10.96	19.66	3.50
Prosody	4.04	0.69	6.17	0.88	200.17	11.29	20.38	3.88
Semantic	5.94	0.86	3.44	1.23	108.99	4.57	23.00	3.60

6.2.3 Procedure

Tasks were web-based and were presented in Gorilla (gorilla.sc). In both experiments, after providing consent and confirmation of requirements, participants completed the BIS scale questionnaire and then proceeded to a brief practice task (10 stimuli) before proceeding to the main task. For the Prosody Experiment, participants had to answer to Congruent and Prosody (threatening sound) by pressing the space bar on their keyboards as fast as possible and before the sentence ended but withhold responses to Semantic and Neutral. For the Semantic Experiment, participants had to answer to Congruent and Semantic (threatening content) by pressing the space bar on their keyboards as fast as possible and before the sentence ended but withhold responses to Prosody and Neutral. Stimuli were presented randomly with a 1.5s inter stimulus interval (ISI) between them. From initial 50 subjects performing each experiment, 10 failed to complete the task for Prosody and 1 for Semantic, as they gave zero responses to either target category or answered to the opposite category (i.e. more responses to Semantic than Prosody in the Prosody Experiment).

6.2.4 Analysis

Before analysis, all responses made on ISI before 300ms were coded as hits or false alarms (FAs) depending on correct or incorrect preceding trial respectively, the rest were kept as misses or correct rejections (CR). For accuracy analyses, a probit regression was implemented. This is similar the logistic regression described in Chapter 3 (model 3.2.8), but the logistic/sigmoid function is replaced by a probit/logit function ($-\log(1/p - 1)$, i.e. the inverse of the logistic function). Also, the model used varying slopes over each sentence type (Neutral, Congruent, Prosody, Semantic), input as matrix operation with the parameter (as in model 3.2.11), and varying intercepts for subject and sentence. For RT analyses, a robust hierarchical regression, using the same rationale as

previous research (Chapter 3: model 3.2.10; Chapter 5: Figure 5.1), was used for both experiments. Participants (40 Prosody Experiment, 49 Semantic Experiment) and stimuli (196) were used as varying intercepts, while BIS score (continuous) and stimuli type (categorical) were the interaction terms (non-varying, for consistency with previous approach). As in previous chapters, results were assessed by using ROPEs and HDIs.

6.3 Results

Models converged well (all $\hat{R}s \cong 1, ESS > 200, BFMIs > 0.6$), presented high precision (all widths of HDIs < ROPEs widths) and good certainty of estimates. For detailed summaries and plots see the OSF repository (https://osf.io/ptcr9/). Accuracy results indicate that in both experiments participants were able to discriminate signal and noise efficiently. Note that response corresponds to any type of response, which was input in the model as a Bernoulli trial (0 = no-response, 1 = response), where response can be Hits and False Alarms (FAs) and no-responses are misses and correct rejections (CRs). So, for the Prosody experiment, the computed probability is for Hits (Prosody and Congruent) and False Alarms (Semantic and Neutral). For the Semantic experiment the computed probability is for Hits (Semantic and Congruent) and False Alarms (Prosody and Neutral). As summarised in Table 6.3 and Table 6.4, the probabilities of participants answering to noise (FAs) is very low for both experiments. Differently, the slopes (in log-odds) indicate that response do not differ between each other within signal or noise categories (i.e. log-odds near zero, thus probability near 50%). In the Prosody experiment the Congruent type shows a ~8% probability increase over Prosody; which is negligible due to overlapping HDIs and high uncertainty. BIS level seems to have no effects either, where the Semantic experiment shows that the probability to answering to Semantic or Congruent tends to decrease by ~20% each. However, HDIs widely overlapping zero and each other indicate that these decreases are highly uncertain. Hence, the general conclusion is that participants show high sensitivity to the signal, low bias and this is not dependent on sentence type or BIS level.

Table 6.3. Prosody experiment accuracy estimates

Type	Response	BIS	Int Mean	Slope Mean	Slope SD	HDI 5%	HDI 95%	Reg Prob%
Neutral	Noise (FA)	1	-5.16	-0.01	0.02	-0.04	0.02	0.58
Prosody	Signal (Hit)	1	2.12	0.00	0.01	-0.02	0.02	89.28
Semantic	Noise (FA)	1	-3.27	-0.01	0.01	-0.03	0.01	3.70
Congruent	Signal (Hit)	1	3.38	-0.02	0.01	-0.04	0.00	96.86
Neutral	Noise (FA)	62	-5.16	-0.74	1.09	-2.55	1.04	1.20
Prosody	Signal (Hit)	62	2.12	-0.03	0.83	-1.38	1.29	89.60
Semantic	Noise (FA)	62	-3.27	-0.63	0.87	-1.98	0.87	6.66
Congruent	Signal (Hit)	62	3.41	-1.02	0.83	-2.36	0.30	98.82

Note. Int mean corresponds to the mean of the intercept used to calculate the regression probability (final column). Excepting Reg Prob, all estimates are in log-odds.
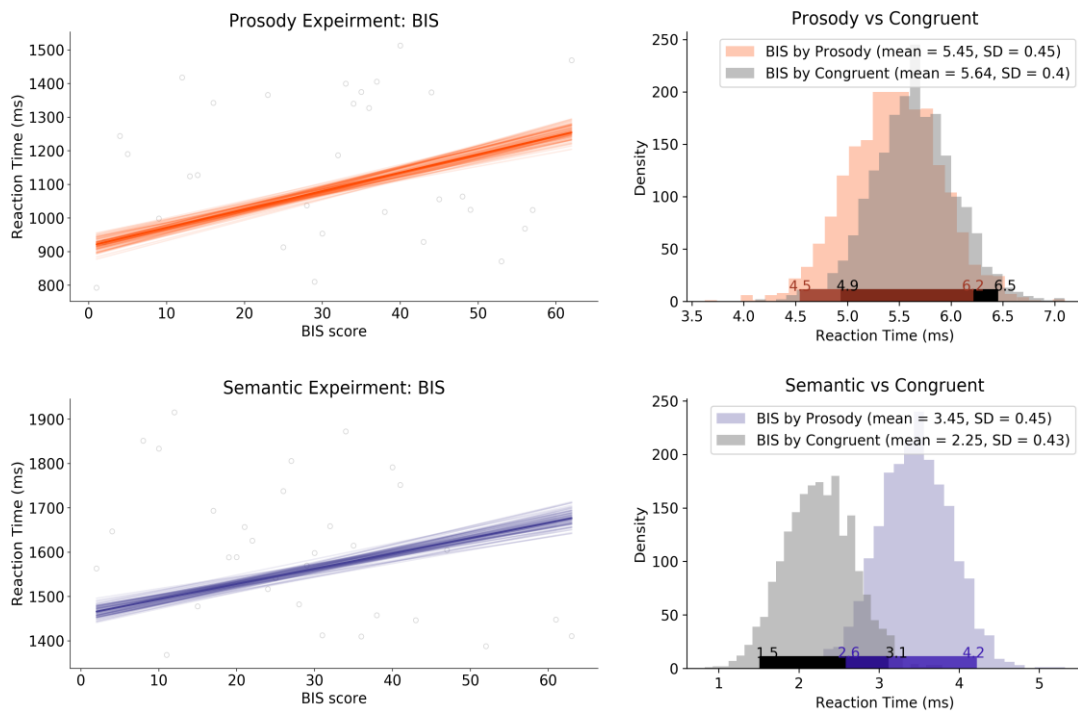
Figure 6.3. Images show posterior distributions, as regression lines, per BIS score point. Upper left: Overall effect of BIS on RT for the Prosody experiment. Bottom left: Overall effect of BIS on RT for the Semantic experiment. Grey circles indicate raw average RT. Faded lines show sample from the posterior and indicate uncertainty. Right panels: histograms of posterior distributions of Prosody vs Congruent (up) and Semantic vs Congruent (bottom); red, blue and black bars represent 90% high posterior densities.

Table 6.4. Semantic experiment accuracy estimates

Type	Response	BIS	Int Mean	Slope Mean	Slope SD	HDI 5%	HDI 95%	Reg Prob%
Neutral	Noise (FA)	2	-5.67	0.02	0.04	-0.05	0.09	0.34
Prosody	Noise (FA)	2	-3.63	0.03	0.03	-0.03	0.08	2.50
Semantic	Signal (Hit)	2	1.62	0.03	0.03	-0.02	0.08	83.00
Congruent	Signal (Hit)	2	1.78	0.04	0.03	-0.02	0.09	85.10
Neutral	Noise (FA)	63	-5.67	0.50	1.33	-1.53	2.73	0.21
Prosody	Noise (FA)	63	-3.63	0.90	1.07	-0.99	2.52	1.06
Semantic	Signal (Hit)	63	1.62	0.98	1.02	-0.67	2.65	65.37
Congruent	Signal (Hit)	63	1.78	1.21	1.03	-0.66	2.78	63.94

Note. Int mean corresponds to the mean of the intercept used to calculate the regression probability (final column). Excepting Reg Prob, all estimates are in log-odds.

Results from RT data analyses indicate that for the Prosody experiment, RTs increase ~334ms from the lowest BIS score (1 point) to the highest BIS score (62 points). For the Semantic experiment, RTs increase ~ 217ms from the lowest BIS score (2 points to the highest BIS score (63 points). For both experiments, sentence type had no effect on RTs; though participants tend to be slower in the Prosody experiment. There is a small speed up for Congruent in the Semantic experiment, but with Congruent posterior HDI overlapping Semantic HDI; which indicates a very small difference (just ~76ms at the highest BIS level). Figure 6.3 summarises these results, from which it can be concluded that RT increases as a function of BIS independent of sentence type, indicating that at the highest BIS level participants can be between 200ms and 300ms slower in answering to threatening speech.

6.4 Discussion

Results indicate that accuracy is not strongly affected by sentence type or BIS. That is, irrespective of anxiety level, participants can distinguish all threatening stimuli with ease; namely, they show high sensitivity and low bias. This is strongly suggested by the low FA rate and high Hit rate that posterior distributions indicate. Results also show that RTs increased directly proportional to BIS score. In other words, as trait anxiety level increases, responses to threatening speech become slower. In the Prosody experiment, where participants answered only to sentences with threatening sound (prosody), RTs increase around 5.5ms to 5.6ms per BIS point; while for the Semantic experiment, where participants answered only to sentences with threatening content (semantics), RTs increased by about 2.2ms to 3.5ms per BIS point. Hence, results from Chapter 5 are

replicated in non-dichotic tasks, with the sole difference of somewhat faster RTs for semantics. Thus, the present hypothesis is strongly supported, as in both experiments participates with higher trait anxiety levels answer similarly accurately but slower than participants with lower trait anxiety levels.

A relevant implication of these experiments is that responses to Semantic are slightly faster or at least similarly fast as responses to Prosody; despite prosodic threat is present across the whole sentence and does not require identifying a specific lexical item. Then, the only strong conclusion which can be taken from present results is that as trait anxiety increases reaction times increase, irrespective of sentence type. This implies that sentences are not sufficiently brief to evidence over-attention to threat (i.e. faster responses to threat given condition). Thus, over-attention to threat is either very hard to observe behaviourally (inconsistent across experiments) or not a feature of trait anxiety; at least not of anxiety understood as an overactive BIS, associated with worry/rumination, and different from fear or anxious arousal (Gray and McNaughton, 2000; Heller et al., 1997).

Further interpretation of these results is straightforward, anxiety inducing slower responses to threatening language could be associated with over-engagement with threatening language. Indeed, repetitive thinking with a verbal component (e.g. rehearsal) has been observed in worry, rumination and post-event processing (McEvoy et al., 2010), where worry is future oriented and rumination is past oriented (post-event processing is recent past oriented). This verbal component, although present together with imagery in rumination, has been more strongly associated with worry (McLaughlin et al., 2007). Also, increases in worry have been associated with brain asymmetries and electrophysiological frequency differences (Nitschke et al., 1999; Spielberg et al., 2013), and increases in BIS have been related with rumination and also with brain asymmetries and electrophysiological frequency differences (Keune et al., 2012; Sander et al., 2005). Hence, whether verbal repetitive thinking is more strongly related with worry or rumination is irrelevant, as long as it contains a verbal component. Although present experiments cannot provide direct evidence for a verbal repetitive thinking component, they reinforce the fact that trait anxiety, understood as an overactive BIS, is associated with a more effortful processing of threatening language. In addition, the notion of an extended multiphasic model of emotional language is strengthened. In Chapter 5, it has been proposed that multistep models of emotional language (Kotz and Paulmann, 2011)

can be extended by adding an additional late processing stage, based on multiphasic models of anxiety (Bar-Haim et al., 2007; Corr and McNaughton, 2012). Given present results, a late deliberation stage (~600ms) extended until or near responses fits well into to a phasic model of emotional language. In particular when long semi-naturalistic sentences are used as stimuli, giving participants enough time to over-engage.

Even so, the Semantic experiment shows a small decrease in accuracy and RT (slightly faster responses) for semantic threat; which may indicate that the accuracy-RT trade-off is slightly weaker for Semantic than for Prosody. Nevertheless, no strong conclusion can be derived from these effects either, as accuracy estimates show great uncertainty (note that in Chapter 5 responses to Semantic were uncertainly more accurate but negligibly slower). Hence, greater RTs from both experiments can be interpreted as an RT-accuracy trade-off for threatening speech, similar to previous research observing a slow-down of RTs for more anxious people but no evidence of accuracy effects (Wang et al., 2019). Although this trade-off has been observed to occur irrespective of threat (Robinson et al., 2013), more recent research on induced anxiety has observed that participants answering to faces in a go/no-go task tend to slow down their responses when under threat of electrical shock (Aylward et al., 2017). This was irrespective of stimuli emotional congruency, and had no effects on accuracy; which was interpreted as a trade-off between accuracy and RT. In other words, participants in the anxiety condition were more cautious with their responses, in particular because of fearful faces capturing their attention, resulting in slowing down responses to achieve better accuracy. This is clearly consistent with present results and the over-engagement interpretation, but could also imply that verbal repetitive thinking is not required for explaining such effects.

However, stimuli length and magnitude increase or RT may explain why verbal repetitive thinking should not be discarded as possible explanation in the present context. Present research uses stimuli of durations ~2s on average and results show evidence of a continuous change from low to high trait anxiety of hundreds of milliseconds. Instead, aforementioned research focuses on stimuli of shorter duration (~250ms), and indicates changes of few dozens of milliseconds between anxious and non-anxious conditions (as opposed to continuous change). In addition, present experiments do not use an external stimulus to induce threat (i.e. shock), which turns the speech stimuli (sentences) themselves as the source of threat. This might indicate that

even when threat is potential and/or weaker (threatening speech instead of an electroshock) the effect is cumulative, inducing longer term disruptions in language processing. Namely, for such a strong continuous delay in less harsh conditions something might be directly disrupting language processing. Although not the only possible mechanism, verbal repetitive thinking is a good candidate, and permits to establish future hypotheses, such as whether verbal disruption tasks would even-out differences between anxious and non-anxious participants as opposed to visual disruption tasks.

Even so, a relevant methodological limitation also needs to be acknowledged. Regarding experimental design, splitting the experiment (Semantic and Prosody) in order to allow a full go/no-go paradigm could reduce the comparability of results. Mainly, because observed data for Semantic and Prosody do not come from the same sample. The fact that previous results are replicated and main predictions were generally precise greatly ameliorates this issue. Future research could directly explore this issue by either increasing the number of non-targets (i.e. Neutral), or by using only congruent stimuli. Another possible limitation is that although verbal repetitive thinking is proposed as a plausible mechanism explaining present results, present experiments do not directly address such mechanism. Nevertheless, verbal repetitive thinking is a plausible proposal and is perfectly testable and falsifiable. As mentioned before, one possibility would be a 'language disruption' experiment, where verbal interference before responses could help in blocking verbal repetitive thinking, thus allowing anxious participants to answer as fast as non-anxious participants; while visual disruption or no-disruption (baseline) conditions would hypothetically show anxiety-related RT increases as in present results.

To conclude, the present in-principle replication shows minor discrepancies with previous experiments and generally replicates previously observed effects. This replication works even with different stimuli, different presentation (non-dichotic), and using a different paradigm (full go/no-go in a split experiment). Therefore, the main interpretation involving an extended multiphasic emotional language processing model and a verbal repetitive thinking component of anxiety interfering at late phases remains a solid candidate for explaining this phenomenon. Even though there is no direct evidence for this yet, and further research needs to be conducted, the present study is a contribution to accumulated evidence in both methodological and experimental terms. These contributions imply that present stimuli are even further validated, as their

features are preserved across speakers, and that statistical models for both RT and accuracy, despite their possible limitations, are inferentially efficient. Withal, more complimentary replications are required for further confirmation of present results, together with electrophysiological measurements which can help in understanding the time-course of neural mechanisms associated with anxiety and emotional language processing.

Chapter 7
EEG Evidence
Late Phase Effects of Anxiety on Threatening Speech Processing

7.1 Introduction

As reviewed on Chapter 2, both anxiety and language processing are characterized by well-defined lateralization patterns. Anxiety has been widely associated to a dual processing pattern: while attention-related arousal and fear-related responses tend to involve greater right hemisphere (RH) processing, evaluation-related apprehension and inhibition-related responses tend to involve greater left hemisphere (LH) processing (Heller et al., 1997; Nitschke et al., 2000; Spielberg et al., 2013). According to current models of speech processing (i.e. dual stream model), speech comprehension recruits both LH and RH (Kemmerer, 2015). While more RH involvement for the slow rate suprasegmental processing of prosody and/or affect/attitude recognition is required, more LH involvement is required for fast rate segmental processing and/or lexical categorization (Belin et al., 2004; Liebenthal et al., 2016; Poeppel et al., 2008; Zatorre et al., 2002). These observations speak not only of general hemispheric activity, but also of very specific activity patterns and brain anatomical structures that, in many cases, are shared by language and anxiety processing. This opens the question of whether these processes simply co-occur, showing superficial similarities, or they actually interact with each other.

The present chapter, also published as a preprint (Busch-Moreno et al., 2020b), seeks to elucidate this issue. Previous research suggests that attention processing, such as that required for fine-grained spatial recognition or recognition of emotional prosody, might be strongly associated to fear arousal, and anxiety responses (Vuilleumier, 2005; Sander et al., 2005). At the same time, worry has been associated with appraisal processes characteristic of environmental evaluation aimed at threat detection, which can be understood as a behavioural inhibition system (BIS) that parses approach or withdrawal responses (McNaughton and Gray, 2000; Corr and McNaughton, 2012). Indeed, neuroimaging evidence indicates that higher levels of arousal are associated to early over-attention to threat, while higher levels of worry are associated to later over-

engagement with threat (Spielberg et al., 2013). Hence, it is plausible that these responses are not only modulated by intrinsic affect (i.e. trait anxiety), but also depend on the type of information being processed (i.e. semantics or prosody) and how this interacts with intrinsic affect by favouring either earlier over-attention or later over-engagement responses. This directly leads to the operative model presented in Chapter 2 (Figure 2.1), which includes three phases of emotional language processing (Kotz and Paulmann, 2011) linked with three overlapping and additional fourth phase as in anxiety processing models (e.g. Bar-Haim et al., 2007). The present operative model was tuned to exclusively address the interaction between threatening speech and trait anxiety. Previous Chapters 5 and 6 provide some behavioural evidence to support the inclusion of a fourth processing phase which is particularly affected by anxiety. However, these experiments do not show laterality differences at the behavioural level. These might be obscured by tasks requesting delayed response or by the length of stimuli. So, whether the time-course of threat processing is affected by anxiety during sentence processing, and whether this involves laterality differences, are questions that need to be addressed through electrophysiology (i.e. EEG).

On the anxiety side, recent EEG evidence indicates that a hyperactive BIS, signalled by higher scores in BIS psychometric scales, presents a right frontal hemispheric pattern (Gable et al., 2017; Neal and Gable, 2017), but not much information is provided about the phasic nature of these asymmetries. Also, some previous EEG studies have observed left or bilateral frontal alpha activity associated with anxious apprehension as measured by worry (Heller et al., 1997; Nitschke et al., 1999), bilateral alpha for rumination-correlated BIS (Keune et al., 2012), and no evidence of hemisphericity patterns of delta or theta waves associated with BIS (De Pascalis et al., 2013). Indeed, recent models of anxiety (stated very generally), propose activation of dorsolateral pre-frontal cortex (dLPFC) inducing medial prefrontal cortex (mPFC) activity, together with hippocampus and insula, that will induce control over amygdala; where structures such as dLPFC show different lateralisation patterns (Robinson et al., 2019). Therefore, evidence suggests that BIS-related lateralization depends on task and stimulus (i.e. environmental conditions). Thus, higher BIS could particularly affect the processing of spoken threatening sentences at LH, especially if they contain semantic information.

Given this, manipulating hemispheric input could reveal possible effects of anxiety on threatening speech processing at different processing phases. In other words, stimuli

that are processed first by LH or RH might be affected differently depending on: 1) their language informational properties (i.e. semantics or prosody) and/or 2) participants' intrinsic lateralization differences when processing threatening stimuli (i.e. anxiety). Considering this, DL stands out as an ideal behavioural test to observe how anxiety-related hemispheric asymmetries might influence emotional language processing hemispheric asymmetries. This relates to the frequently observed phenomenon of a right ear advantage (REA) for non-prosodic language stimuli (Hugdahl, 2011); this implies participants answering faster and/or more accurately to stimuli presented at their right ear when compared to stimuli presented at their left ear. On the other hand, prosody, in particular emotional prosody, has been observed to present either a left ear advantage (LEA) or a diminished REA (Godfrey and Grimshaw, 2015; Grimshaw et al., 2003). Few dichotic listening (DL) experiments have researched the effects of anxiety on emotional speech processing (Gadea et al., 2011). They either use speech/prosody as an emotion-eliciting stimulus, or use DL mainly as an attentional manipulation technique (e.g. Leshem, 2018; Peschard et al., 2016; Sander et al., 2005).

Behaviourally, using two DL experiments, recent research has observed that anxiety does not induce clear ear (laterality) differences and minor sentence type effects (Chapters 5 and 6). It was argued that as sentences were relatively long (near 1800ms on average), experiments allowed ample time for deliberation and thus effects of trait anxiety obscured other possible effects. Given this, it might be that processing differences are not reflected in the behavioural output, but may be measured by electrophysiological activity. The first experiment from the aforementioned task was intended for allowing deliberation, requesting participants to answer only after sentence's end. Participants answered to threatening sentences of three types: 1) containing prosodic threat only, 2) containing semantic threat only, 3) containing both; all dichotically paired with neutral sentences (containing neither type of threat). In a second task, intended to measure possible attention effects, participants answered to the neutral pair instead. The present study aims to replicate this previous experiment by using the same procedure, but with newly recorded sentences and with trait anxiety measured by using the BIS scale from the Reinforcement Sensitivity Theory Personality Questionnaire (RST-PQ) (Corr and Cooper, 2016). In addition, the ERP technique could provide fine-grained temporal information, crucial for testing present operative model, as laterality effects might been obscured behaviourally but could still be observed through EEG activity.

Given this setup and the previously discussed theory and research, higher levels of trait anxiety (higher BIS scores) will induce early over-attention to threat but mid-late over-engagement with threat are expected. The first should be associated with anxiety induced arousal and BIS (McNaughton and Gray, 2000), which has been observed to be right lateralized (Heller et al., 1997; Neo et al., 2011). The second should be associated to a left lateralized or bilateral pattern (Nitschke et al., 1999), which might be due to mid-stage LH thought-induced (e.g. worry or rumination) evaluation and inhibition (Spielberg et al., 2013), but a later RH involvement associated to sustained anxiety-induced arousal. The stages of this processing time-course can be understood thusly: 1) pre-attentive (50-100ms), 2) attentive (150-250ms), 3) evaluative (250-500ms), 4) orientative (500-750ms). Following multistep model predictions: the first stage should show a bilateral pattern for language, as early sensory processing of language might require the use of both hemispheres as prosody and semantics (even if neutral) need to be simultaneously processed. The second stage should be lateralized as a function of stimulus type, modality and emotion expressed. The third stage requires bilateral involvement due to evaluation and integration of different types of stimulus and emotion. The fourth proposed stage here should show an initial bilateral involvement with stronger RH persistent activity. This additional later stage can be understood as a deliberation phase, where participants decide (orientate) their responses to emotional stimuli.

Following previous predictions, the following hypotheses can be specified: 1) Early (~100ms) and Mid-early (~200ms) effects of anxiety on speech processing (Pell et al., 2015), where over-attention to threat will show stronger RH involvement for prosody, facilitating detection. 2) Mid-late stage (~400ms) anxiety effects on LH due to over-engagement with threat (Spielberg et al., 2013); this should particularly affect semantic stimuli by slowing down their processing. 3) A late stage (~600ms) effect of anxiety on goal orientation (Bar-Haim et al., 2007), where both LH and RH (more strongly) should be involved. This could be the results of a worry-arousal loop (McNaughton and Gray, 2000) due to continued exposure to threatening stimuli. In other words, verbal repetitive thinking could take prevalence at this stage due to over-engagement with threat (see Chapters 5 and 6). No specific ERPs are predicted, as present research has no direct precedents in the literature. Hence, it is not clear whether observed ERPs, and their amplitude differences by condition, will be coincidental or not with those observed in previous research.

7.2 Methods

7.2.1 Participants

Participants (mean age = 28.6, age range = [19, 54], 19 females, 17 males) were recruited using Sona Systems' (sona-systems.com) UCL platform. Only participants reporting being right-handed, having English as first language, without hearing problems and with no history of neurological/psychiatric disorders were recruited. Participants were remunerated £7.5/hour rate. All participants gave their informed consent before participating and were informed of their rights and that their data is protected under GDPR protocols. They were debriefed after the experiment. Note that the sample size is moderately big for an EEG experiment, as compared to previous literature (e.g. Paulmann et al., 2012; Pell et al., 2015; Wabnitz et al., 2015), which will serve to guarantee more precision in behavioural measures and a better distribution of BIS measures.

7.2.2 Materials

Stimuli for the present experiment are widely described on Chapter 4 (Study 2). As a reminder, four types of sentences were recorded: Prosody (neutral-semantics and threatening-prosody), Semantic (threatening-semantics and neutral-prosody), Congruent (threatening-semantics and threatening-prosody), and Neutral (neutral-semantics and neutral-prosody). This resulted in 54 recorded sentences per threatening category and respective Neutral pairs (324 in total). As in Chapter 5, each threatening sentence was paired with a neutral sentence of similar length and adjusted by minimally extending silence periods (max 40ms) to match lengths exactly in Audacity (audacityteam.org). Table 7.1 provides a general summary of these stimuli.

Table 7.1. Average number of words, duration and reaction time per stimulus type

Type	Words Threat	Words Neutral	Stimulus Duration	Reaction Time
Congruent	4.63	5.63	1838.63	522.37
Prosody	4.24	4.93	1558.25	601.20
Semantic	4.43	4.89	1543.53	593.40

7.2.3 Procedure

Participants were introduced to the recording room (electrically and sound isolated chamber), signed consent, and sat at 1m distance from a 20" screen used to

display all tasks via PsychoPy2 (Peirce et al., 2019). Next, participants provided their demographic information (age and sex) and completed the BIS and FFFS questionnaires. Participants' head dimensions were measured and EEG caps of according size were placed and centred, conductivity gel was placed and the Biosemi 64 Ag/Cl electrode system (biosemi.com) was connected. Two electrooculogram (EOG) electrodes were connected near left zygomatic bone and near right orbital bone to capture vertical EOG (VEOG) and horizontal EOG (HEOG) activity respectively. Impedance levels were kept below 20Ω and electrodes were checked to be working properly. While EEG recording, participants completed the direct-threat task and indirect-threat tasks as in Chapter 5, but sentences were played twice such that a threatening sentence was presented once at each ear. Participants were requested to swap response hand every other trial. Starting hand, ear and task were counterbalanced.

7.2.4 EEG Data Processing

EEG recordings were pre-processed using Python's MNE package (Gramfort et al., 2014). A completely automated pre-processing pipeline was implemented (based on: Jas et al., 2018). This consisted in the following steps: 1) Importing data, checking events and fixing if misplaced. 2) Set data to average reference (Dien, 1998; Lei and Liao, 2017). 3) Preparing data for independent component analysis (ICA) only: low-pass filtering at 40hz to avoid aliasing artifacts, down sampling to 256hz, high pass filtering at 1hz for better artifact detection, and automatic rejection of noisy segments based on the Autoreject MNE package (Jas et al., 2017; Winkler et al., 2015). 4) Computing ICA components by using python Picard package (Ablin et al., 2018). 5) Removing components that are highly correlated with noise at Fpz channel to remove EOG artifacts from data. 6) Applying average reference to raw data (Dien, 1998; Lei & Liao, 2017) with excluded EOG artifacts and applying first a high pass (0.1hz) and later a low pass (100hz) filter (Luck, 2014; VanRullen, 2011; Widmann et al., 2015). 7) Epoching data from 0-1000ms using a pre-stimulus onset 100ms baseline (10% of epoch), downsamplig to 512hz, and applying automatic detection, repairing and rejection of noisy epochs by using Autoreject's Bayesian optimization (for details on procedure see: Jas et al., 2017). 8) Applying baseline correction (baseline subtraction) (Luck, 2014; Tanner et al., 2015), as integrating the baseline into the model as a regressor (Alday, 2019) could be detrimental for convergence and precision in this particular case. 9) Extracting trial by trial mean amplitudes at 4 a priori

defined time windows (to avoid double-dipping): 50-150ms, 150-250ms, 250-500ms, 500-750ms; these time windows cover the time-windows proposed by the multistep model (Kotz and Paulmann, 2011), plus the proposed fourth time-window. ERP and scalp plots of processed data were produced using the MNE package.

7.2.5 Data Analysis

Models for behavioural data were replications from those used in Chapter 5 (Figure 5.1), all described in Chapter 3. Figure 7.1 shows the model used for EEG data, the diagram is based on Kruschke's (2015) and Martin's (2018) model specification and presentation, and their guidelines on robust regression. The model was sampled using Markov Chain Monte Carlo (MCMC) No U-turn Sampling (NUTS) as provided by PyMC3 (Salvatier et al., 2016). All models were sampled with two chains of 2000 tuning steps and 2000 samples, and initialised using automatic differentiation variational inference (ADVI) as provided by PyMC3. Plots of results were produced using Arviz (Kumar et al., 2019) and Matplotlib (Hunter, 2007). Results were assessed using a region of practical equivalence (ROPE) method (Kruschke, 2015; Martin, 2018), where high posterior density intervals (HDIs) were considered as presenting a considerable difference when far away from ROPEs defined as 1SD to 2SDs around zero.
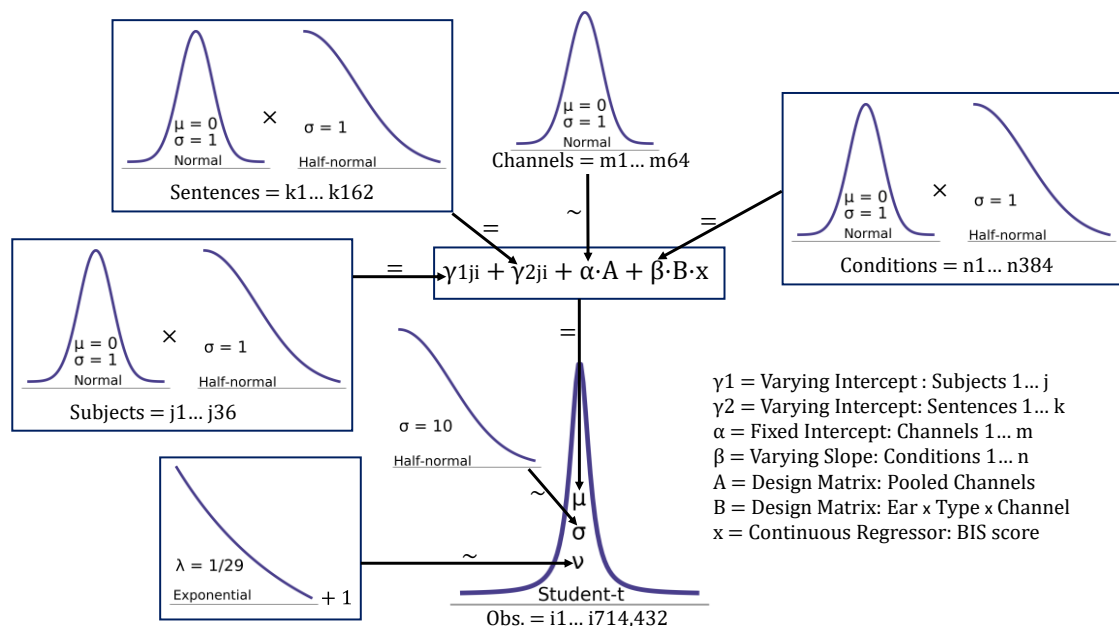
Figure 7.1. Diagram representation of hierarchical robust regression model. Arrows indicate the relationship between a parameter and priors/hyperpriors, where tilde (~) indicates a stochastic relationship and equal (=) indicates a deterministic relationship. Observations in the likelihood distribution are equivalent to mean amplitude at each single trial.

The core idea of the model is to un-pool data from the individual level of subjects and items, the focus can be placed on group level slopes at each single interaction point. Given this, pooling at the electrode level through a normal non-varying prior helped us to define an offset for the other reparametrized priors (e.g. McElreath, 2020), which do not contain individual location parameters. In this way, intercepts from each electrode could be obtained, also improving sampling and convergence in a substantial manner. To be consistent with previous experiments, FFFS ratings were excluded from the analyses, as the present focus is on trait anxiety and not on trait fear. Fear measures (FFFS) were collected as they could be required for comparison in future analyses.

7.3 Results

7.3.1 Behavioural Results

All models sampled well, showing excellent convergence ($\hat{R} \cong 1$, ESS > 200, BFMIs > 0.6); all summaries and plots can be found in present chapter's Open Science Framework (OSF) repository (https://osf.io/n5b6h/), also linked the preprint version of this chapter (Busch-Moreno et al., 2020b). Accuracy results largely replicate results reported in Chapter 5. Table 7.2 and Table 7.3 summarise accuracy results from direct- and indirect-threat conditions. Note that these results consider BIS slopes (in log-odds) across all conditions, and all posterior distributions overlap zero, as clearly indicated by HDIs. For more details on this, please see the OSF repository (https://osf.io/n5b6h/)

Table 7.2. Direct-threat accuracy estimates

Type	Ear	BIS Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%	Probability%
Semantic	Left	1	-0.01	0.02	-0.04	0.03	49.81
Prosody	Left	1	-0.01	0.02	-0.05	0.02	49.66
Semantic	Right	1	0.00	0.02	-0.03	0.03	49.93
Prosody	Right	1	0.01	0.02	-0.02	0.04	50.29
Semantic	Left	55	-0.42	1.08	-2.03	1.45	39.67
Prosody	Left	55	-0.74	1.06	-2.48	0.94	32.20
Semantic	Right	55	-0.15	0.94	-1.61	1.39	46.24
Prosody	Right	55	0.63	0.92	-0.88	2.07	65.19

Note: *All estimates are in log-odds.*

Table 7.3. Indirect-threat accuracy estimates

Type	Ear	BIS Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%	Probability%
Semantic	Left	1	-0.008	0.017	-0.037	0.018	49.80
Prosody	Left	1	0.026	0.017	0.000	0.056	50.66
Semantic	Right	1	-0.006	0.014	-0.030	0.015	49.84
Prosody	Right	1	0.014	0.014	-0.009	0.038	50.36
Semantic	Left	55	-0.444	0.925	-2.059	1.013	39.07
Prosody	Left	55	1.441	0.947	-0.009	3.089	80.86
Semantic	Right	55	-0.344	0.782	-1.647	0.852	41.49
Prosody	Right	55	0.793	0.786	-0.474	2.103	68.85

Note: *All estimates are in log-odds.*

Results from RT data also replicate previous experiments. Table 7.4 and Table 7.5 summarise results from direct- and indirect-threat conditions respectively. Note the big increase in estimates from lowest BIS scores to highest BIS scores, with HDIs well apart from each other. When intercepts are taken into account, the Direct-threat estimates portray an average effective increase of ~175ms from lowest (1 point) to highest (55 points) BIS score by Prosody at the left ear, and ~164ms at the right ear. While for Semantic the increase is ~172ms at the left ear and ~150ms at the right. The apparent left ear advantage for Semantic can be discounted due to HDI overlap. However, overall, RTs were the shortest for Congruent. Indirect-threat results are similar, excepting that in this case slower responses are only for Semantic. For more details on these additional patterns see the OSF repository (https://osf.io/n5b6h/).

Table 7.4. Direct-threat reaction time estimates

| Type	Ear	BIS Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%
Semantic	Left	1	3.20	0.40	2.54	3.84
Prosody	Left	1	3.25	0.39	2.62	3.90
Semantic	Right	1	2.79	0.33	2.25	3.35
Prosody	Right	1	3.04	0.34	2.49	3.60
Semantic	Left	55	176.09	21.76	139.95	211.20
Prosody	Left	55	178.50	21.64	143.94	214.60
Semantic	Right	55	153.21	18.25	123.59	184.04
Prosody	Right	55	167.04	18.87	137.08	198.26

Note: *All estimates are in milliseconds (*ms).

Table 7.5. Indirect-threat reaction time estimates

Type	Ear	BIS Score	Posterior Mean	Posterior SD	HDI 5%	HDI 95%
Semantic	Left	1	4.37	0.41	3.71	5.06
Prosody	Left	1	4.29	0.41	3.64	4.96
Semantic	Right	1	4.23	0.35	3.65	4.79
Prosody	Right	1	3.79	0.35	3.22	4.35
Semantic	Left	55	240.48	22.47	204.31	278.46
Prosody	Left	55	235.85	22.29	200.14	272.72
Semantic	Right	55	232.38	19.09	200.92	263.32
Prosody	Right	55	208.63	19.03	177.11	239.50

Note: *All estimates are in milliseconds (*ms*).*
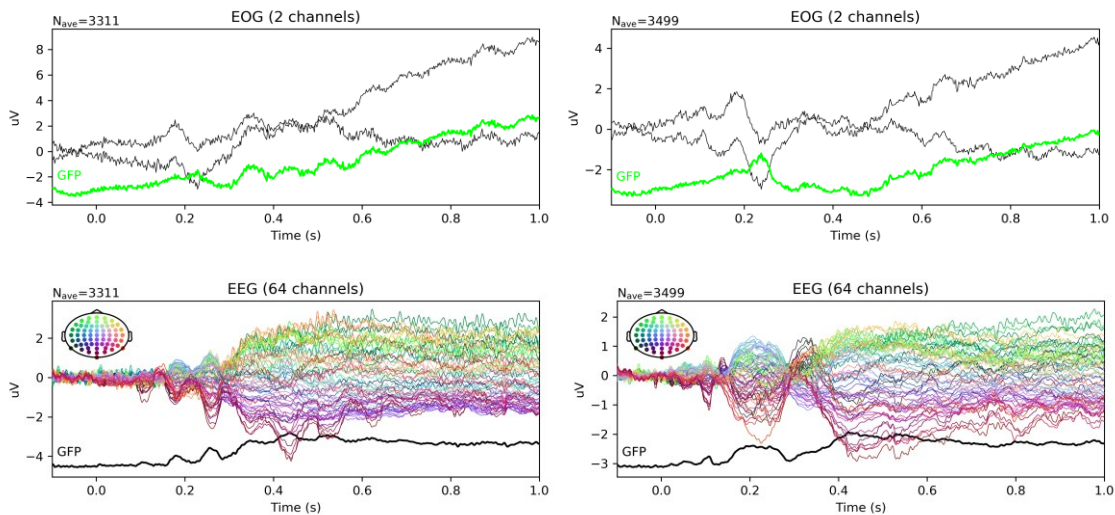
Figure 7.2. Average EOG and EEG activity. Left panels: high BIS. Right panels: low BIS. Up: VEOG and HEOG averages. Bottom: grand averages across all electrodes. Note that polarities of EOG and EEG are aligned and baselines tight around zero amplitude, so drifts are not artefactual (see: Luck, 2014).

7.3.2 EEG Results

All models sampled well, showing excellent convergence ($\hat{R} \cong 1$, ESS > 300, BFMIs > 0.6). Main results indicate that in both the direct- and indirect-threat conditions, the effects of BIS were concentrated on window4 (500-750ms), irrespective of ear or sentence type (Figure 7.9). However, intercepts indicate smaller but clear general effects at all time-windows (Figure 7.8). Hence, for brevity, only results from the direct-threat condition will be presented. All other results' summaries and plots can be found in the

OSF repository (https://osf.io/n5b6h/). Figure 7.2 gives an account of EOG activity and raw average amplitude across electrodes given BIS level (median split for visualisation), and Figure 7.3 shows raw averages of ERPs' topography by BIS score tertiles (Low: [1,15], Mid: [16,28], High: [29,55]). Figure 7.4 shows ERP's topography by Ear (left and right and Figure 7.5 shows ERP's topography by type (Congruent, Prosody, Semantic).
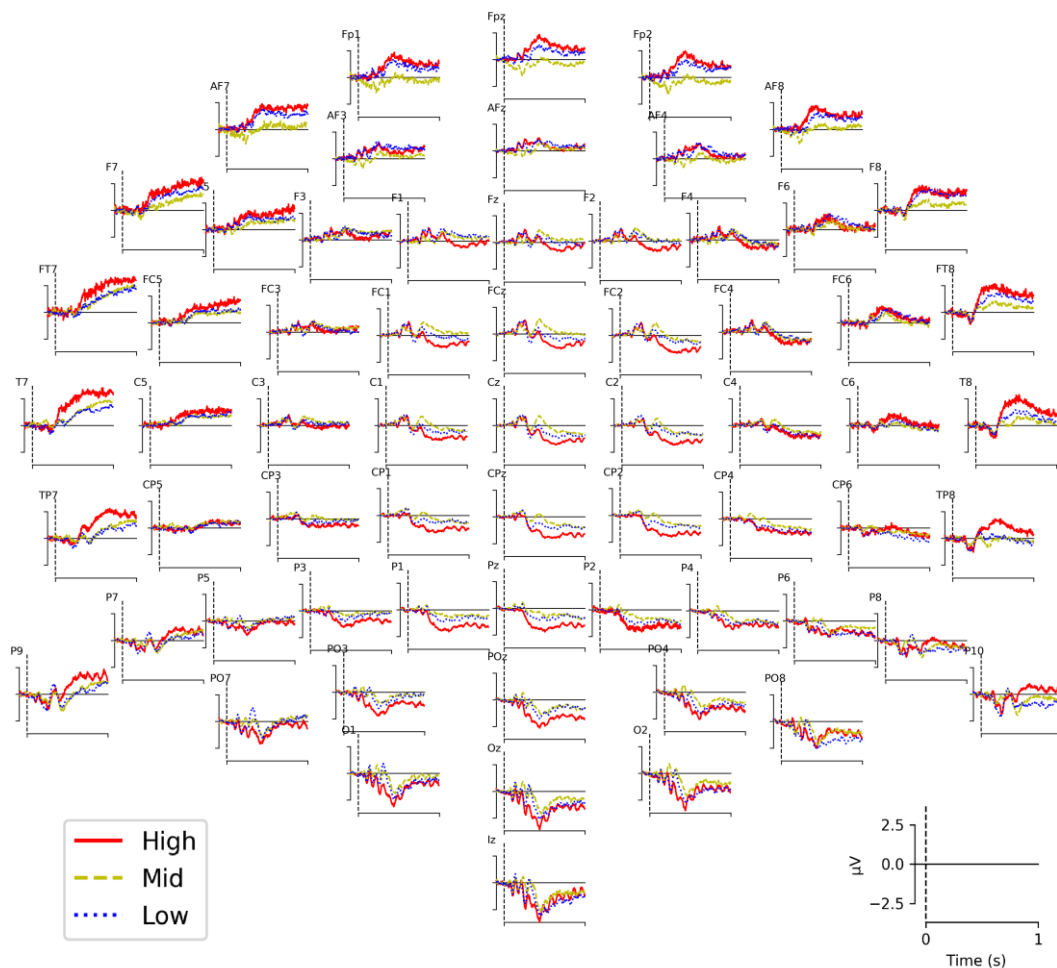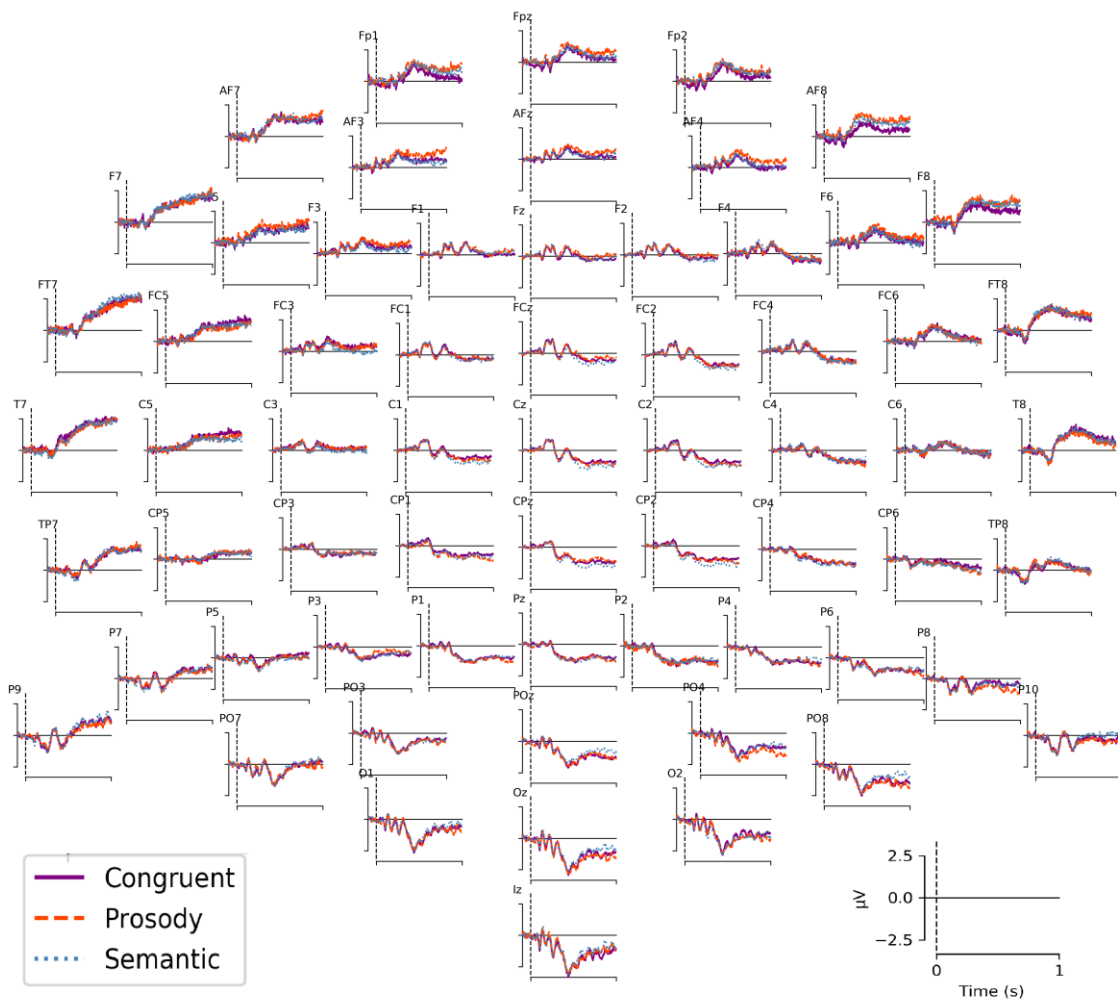

Figure 7.3. Scalp distributions of ERPs by BIS. Red solid line: High BIS score (over 28 points). Yellow dashed line: Mid BIS score (between 16 and 28 points). Blue dotted line: Low BIS (below 16 points). Waves are raw averages.
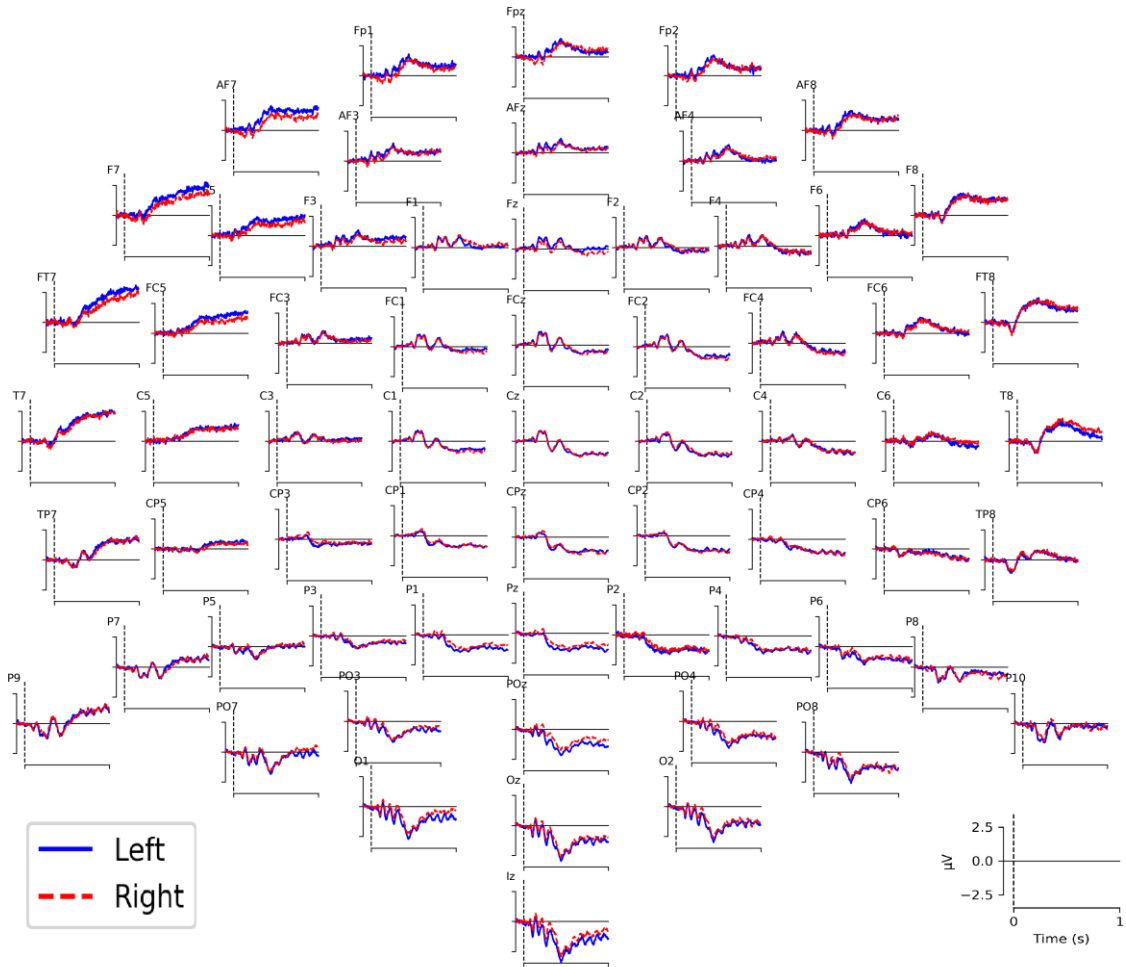
Figure 7.4. Scalp distributions of ERPs by Ear. Blue solid lines: Left ear. Red dashed line: Right ear. Waves are raw averages.

Figure 7.5. Scalp distributions of ERPs by type. Purple solid lines: Congruent. Red dashed line: Prosody. Blue dotted line: Semantic. Waves are raw averages.

Figure 7.6 shows the two electrodes showing maximal increases from the lowest BIS score (1 point) to the highest BIS score (55 points). Plots represent amplitude waves as resulting from a robust regression (Student-t observed distribution) using sampled time-points (downsampled to 256hz) as varying intercepts and as varying slopes (282 samples by BIS score), using reparametrized normal priors. Models, sampled with NUTS (1000 tuning, 1000 samples) passed all convergence criteria as main models. Although intended for display, these analyses further corroborate strong BIS effects at late time-windows, showing a clearer and stronger late positive complex (LPC) at the left hemisphere. Thus, earlier general effects observed in average ERP waves (Figure 7.6 and Figure 7.7) and intercepts (Figure 7.8) are not present when accounting for BIS. This is strongly supported by varying slopes' estimates (Figure 7.9).
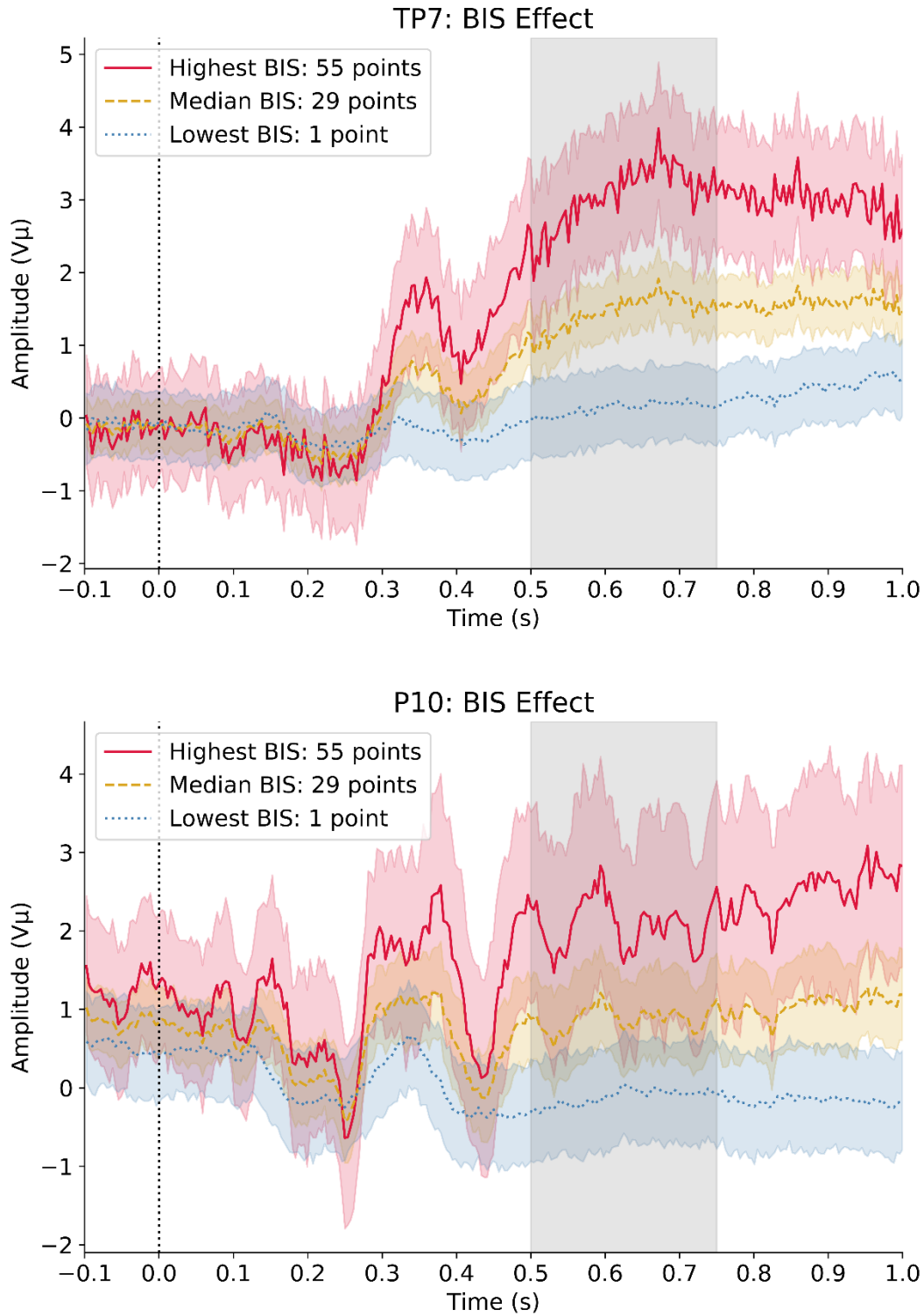
Figure 7.6. Hierarchically regressed amplitudes at TP7 and P10. Images show amplitude waves from a regression across 282 (26 baseline + 256 epoch) samples. Each estimated sample is a whole posterior distribution, shadows indicate 90% HDIs at each sample. Dotted lines in blue: lowest BIS (1 point). Dashed lines in yellow: sample's median BIS (29 points). Solid line in red: highest BIS (55 points). Note that higher levels of BIS tend to be noisier (such as containing more energy/faster frequency), and that effects at right hemisphere (TP7) reach greater positive amplitudes and have fewer overlapping HDIs. Vertical dotted line at zero: sentence onset.
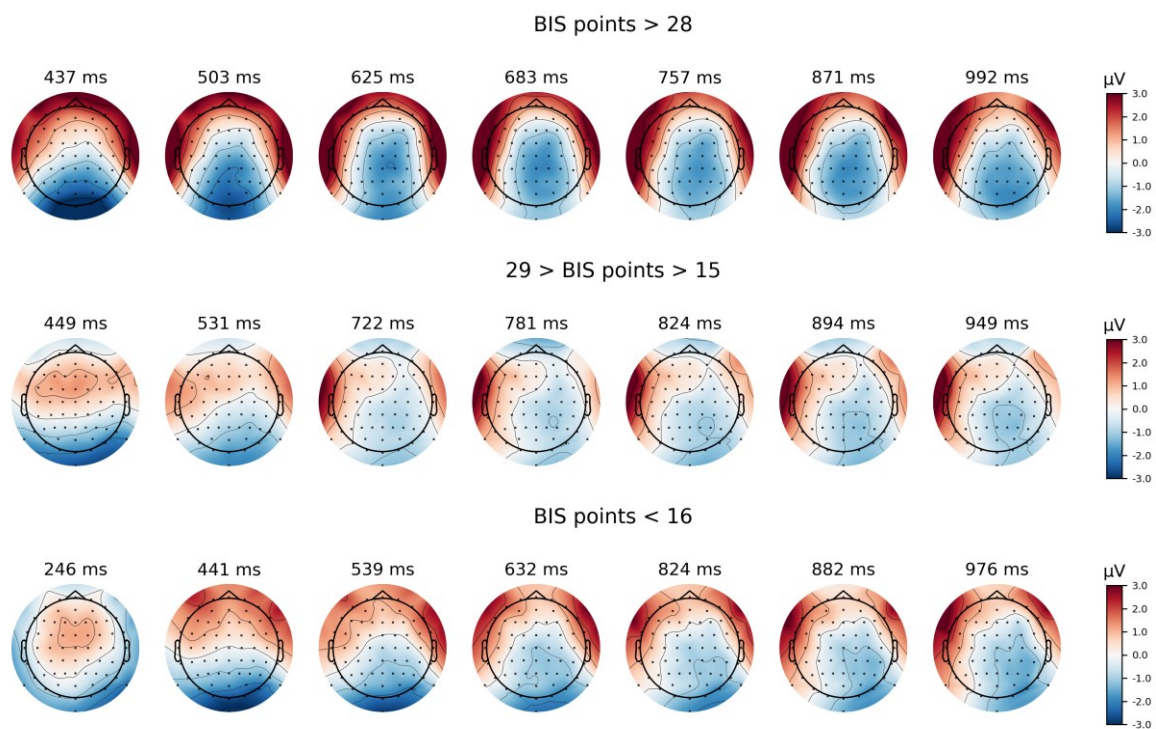
Figure 7.7. Topoplots of average amplitude by BIS tertile. Images show raw mean amplitude across epoch by BIS tertile. Top: upper BIS score tertile. Middle: mid BIS score tertile. Bottom: lower BIS score tertile. Time-points were selected as the local maxim Global Field Power (GFP).

Figure 7.7 shows mean amplitude peaks selected by maximum local global field power (GFP). GFP indicates a spatial standard deviation quantifying electrical activity from all electrodes at specific time-points (Skrandies, 1990), reliably revealing peaks of greater amplitude. These peaks suggest the lack of earlier phases effects by BIS, only indicating sufficiently strong amplitudes from over 200ms for lower BIS levels (below 16 points) and from over 400ms for higher BIS levels (over 15 points). Main analyses indicate that this might be due to very small effects in window1 (50-150ms), with no estimates over $|0.5|\mu V$ for central electrodes (pooled by channel). These values progressively increase to up to around $|1|\mu V$ and $|2|\mu V$ for window2 (100-250ms) and window3 (250-500ms) respectively. Nevertheless, estimates by conditions (BIS, ear and type) indicate negligible increases, with amplitude consistently remaining around zero values. This can be seen in Figure 7.8 and Figure 7.9, which summarise intercepts and slopes respectively (direct-threat task posterior distributions). For brevity, only the condition of BIS by Congruent at left ear is shown, as all conditions evidenced similar patterns. Effects start to be noticed from window3 (250-500ms) and become considerable at window4 (500-750ms), indicating a bilateral temporo-parietal increase

of ~0.05μV per BIS score point; consistent with Figure 7.6, where an increase of around 3μV can be observed from the lowest BIS score (1 point) to the highest BIS score (55 points).

These results suggest general effects of threatening language, but also indicate that BIS highly modulates amplitude effects, in particular at late time-windows. Table 7.6 gives a more thorough summary across all conditions for electrode TP7. It must be emphasised that the same pattern was observed in all electrodes showing strong amplitude increases. Whereas left-side temporo-parietal electrodes increase from near zero values (e.g. T7, TP7, P7, P7), right-side electrodes show increases from negative amplitudes (e.g. TP8, P10) and a less clear LPC. Figure 7.10 shows regression lines for TP7 across all conditions.

Table 7.6. Estimates from TP7 across all conditions for lowest and highest BIS scores

Channel	Type	Ear	BIS score	Mean	SD	HDI 5%	HDI 95%	ESS	$\hat{R}$
TP7	Prosody	left	1	0.05	0.01	0.04	0.06	2273.18	1.00
TP7	Semantic	left	1	0.05	0.01	0.04	0.06	2138.45	1.00
TP7	Congruent	left	1	0.04	0.01	0.03	0.06	2161.74	1.00
TP7	Prosody	right	1	0.05	0.01	0.03	0.06	2140.95	1.00
TP7	Semantic	right	1	0.06	0.01	0.04	0.07	2168.43	1.00
TP7	Congruent	right	1	0.04	0.01	0.03	0.06	2314.09	1.00
TP7	Prosody	left	55	2.77	0.45	2.04	3.51	2273.18	1.00
TP7	Semantic	left	55	2.82	0.46	2.09	3.57	2138.45	1.00
TP7	Congruent	left	55	2.36	0.46	1.64	3.14	2161.74	1.00
TP7	Prosody	right	55	2.67	0.46	1.88	3.39	2140.95	1.00
TP7	Semantic	right	55	3.04	0.47	2.29	3.82	2168.43	1.00
TP7	Congruent	right	55	2.47	0.48	1.65	3.22	2314.09	1.00

Note: All estimates are in microvolts (μV). ESS: effective sample size from centre of distribution (bulk). $\hat{R}$: convergence statistic.
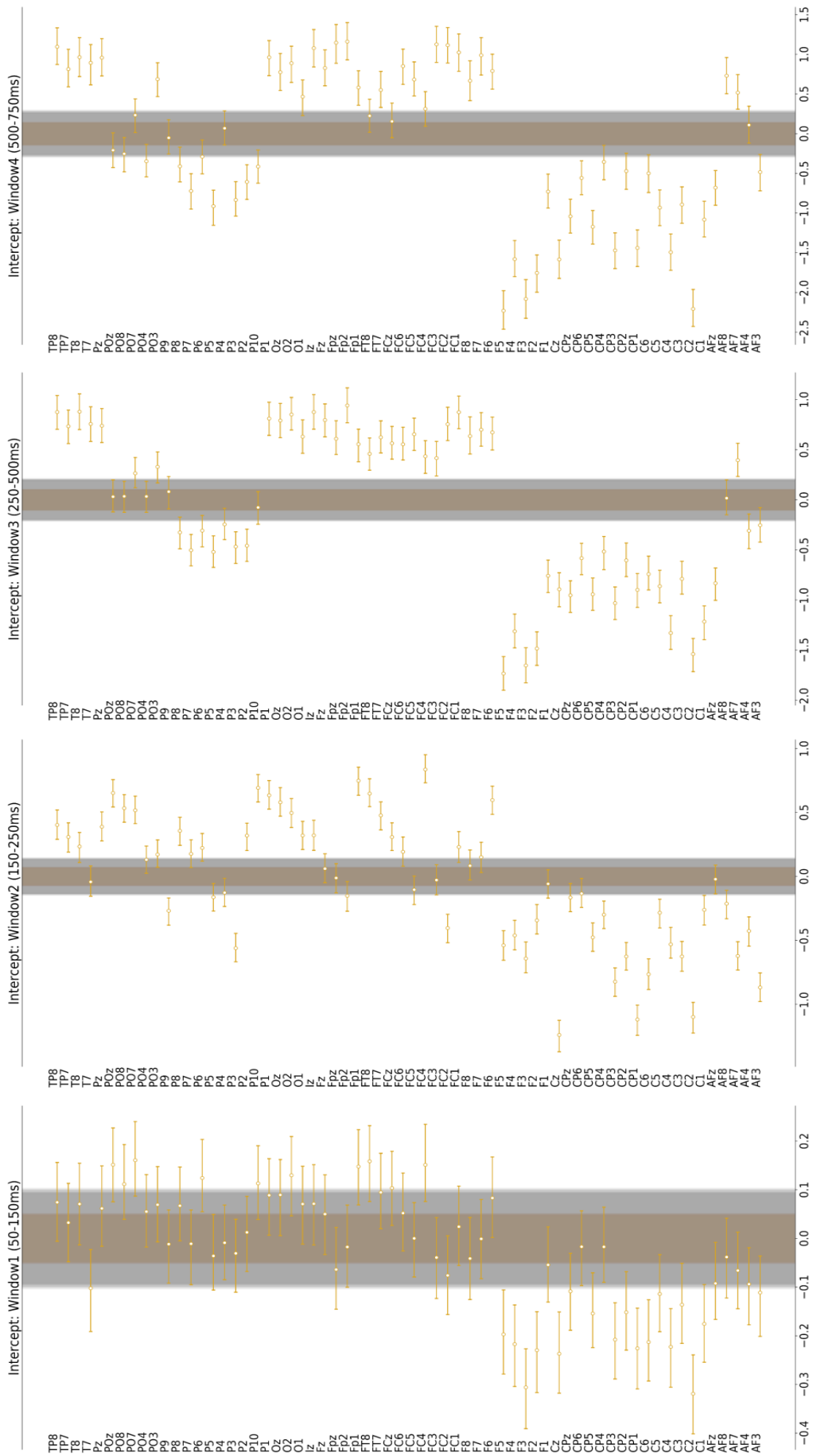
Figure 7.8. Robust regression intercepts. Figure contains forest plots summarising direct-threat results from a robust regression on mean amplitudes. Forest plots show intercepts from all time-windows in the BIS by Congruent by Left Ear condition, which indicate pooled activity (non-varying) across 64 electrodes.
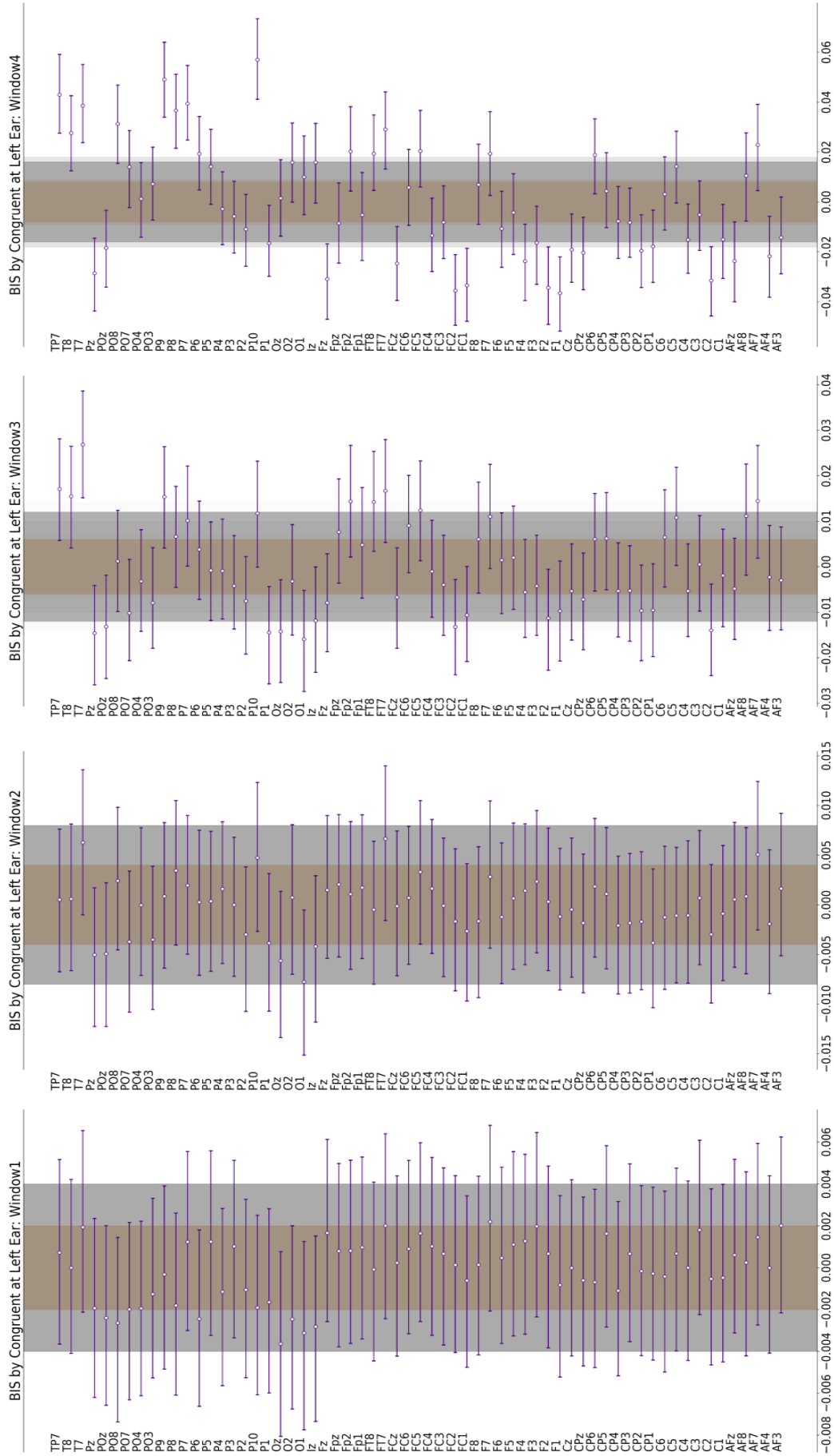
Figure 7.9. Robust regression slopes. Figure contains forest plots summarising direct-threat results from a robust regression on mean amplitudes. Forest plots show varying slopes estimates from four time-windows in the BIS by Congruent by Left Ear condition. Note that effects of BIS are only sufficiently strong at the fourth time window, bilaterally, but particularly clear at left temporal and parietal electrodes such as T7, TP7, P7, P9.
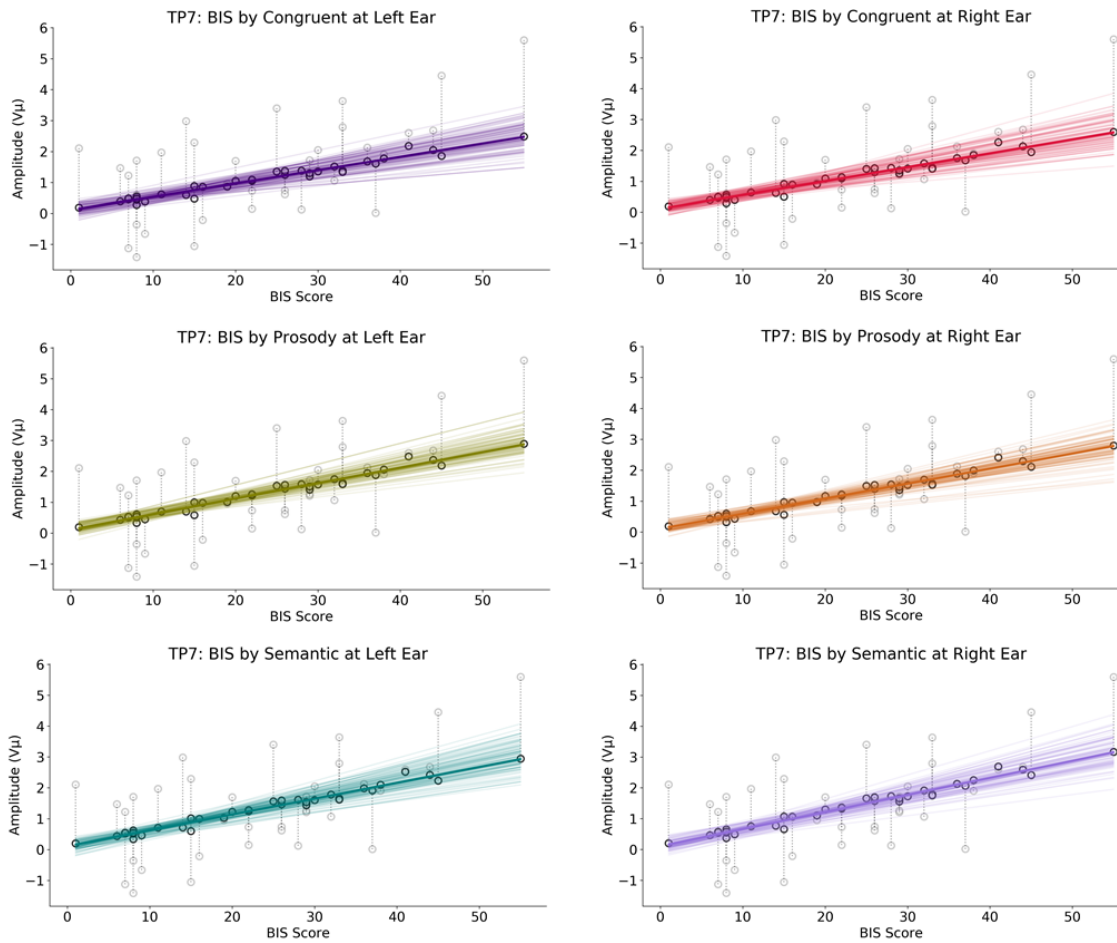
Figure 7.10. Regression lines for TP7 electrode at each condition. Each point in the line is a posterior distribution, faded lines are random samples from the posterior and express uncertainty. Grey circles indicate raw mean amplitude per participant. Black circles indicate estimated mean amplitude per participants. Dotted lines indicate shrinkage.

7.3.3 Exploratory Analyses

Finally, two exploratory analyses were conducted. Firstly, amplitudes were analysed from epochs aligned with response execution. It has been suggested that this method is crucial for observing response preparation P600 ERPs (Sassenhagen et al., 2014; Sassenhagen and Bornkessel-Schlesewsky, 2015). The rationale is assuring that the observed LPC reflects a deliberation phase and not response preparation. Thus, each epoch's offset was locked to each trial's RT (RT event), epoch's onset was placed 1s previous to RT event with a 200ms baseline pre-onset. These analyses should show a P600 similar to the one observed in previous literature, which would corroborate that the observed LPC appearing ~600ms after sentence's onset is a different ERP from the

P600 appearing when epochs are RT locked. Analyses were performed exactly as for the main models, but restricted to as 600-800ms time-window. Both direct- and indirect-threat conditions indicate strong amplitude increases. This confirms a response-aligned P600 which is different from the observed LPC, as average RT (555.21ms) and average sentence duration (1646.30ms) guarantee that responses will generally not overlap or the LPC would be captured by the baseline. Indirect-threat did not show any differences as modulated by stimulus type, ear presentation or BIS score. However, contrary to our expectations, direct-threat indicated small increases for Prosody at the left ear as BIS score increased (~0.025μV per BIS point) at few frontal electrodes, while Semantic at the right ear showed similar but more consistent increases at posterior electrodes (e.g. Oz, O2). Although very small, these effects might be associated with earlier BIS effects, which could be accumulated until near response execution. Figure 7.11 summarises these results for the Semantic/right-ear condition by using electrode Oz as example. See the OSF repository (https://osf.io/n5b6h/) for all summaries and plots
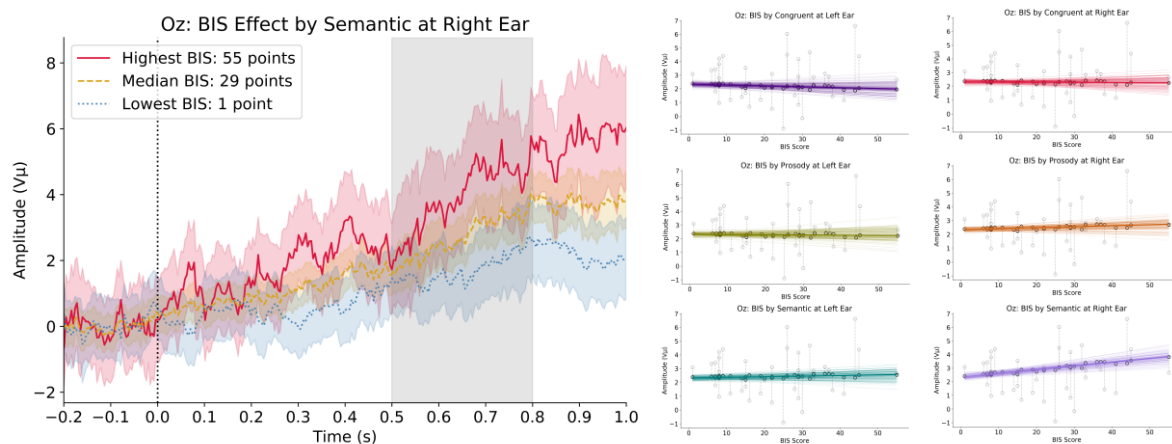
Figure 7.11. Exploratory results example at Oz electrode. Image at the left shows regressed amplitude waves from electrode Oz, shadows indicate HDIs. Vertical grey dotted line: RT-locked onset. Note that there is almost or no overlap between highest and lowest BIS sores at later times in the epoch. Images at the right show regression lines containing posteriors from all conditions, faded lines express uncertainty. Note Semantic at right ear (bottom right), whose uncertainty does not overlap with other conditions at higher BIS scores.

The second exploratory analysis consists of source-localising amplitudes from window4. The aim is to identify potential activity sources for the observed LPC. According to emotional acoustic and language processing (Frühholz et al., 2016a; Kotz and Paulmann, 2011) and anxiety models (Robinson et al., 2019), most probable sources of

activity should be in areas of emotional and language processing. On average across the 500-800ms window, increases of left-temporal activity in association with basal ganglia (emotional language processing) and amygdala-hippocampal activity in association with right prefrontal activity (anxiety processing) are expected. To explore this possibility, source localization over the 500-800ms time window from sentence's onset was performed. For simplicity, estimated activity was averaged across the temporal dimension to test whether activity at given voxel and/or anatomical areas increased as a function of BIS.

Table 7.7. Highest log-activation estimates from voxel by BIS interaction.

Brodmann Area	Activation	MNI x	MNI y	MNI z	Mean	SD	HDI 5%	HDI 95%
Left-BA21	0.487	-50	-50	10	0.020	0.007	0.008	0.031
Left-Caudate (48)	0.199	-10	25	10	0.019	0.007	0.008	0.031
Left-BA22	0.508	-50	-45	15	0.018	0.007	0.007	0.030
Left-Insula (13)	0.218	-20	25	15	0.016	0.007	0.004	0.028
Right-Caudate (48)	0.228	5	25	5	0.016	0.007	0.004	0.028
Left-BA45	0.214	-20	30	15	0.016	0.007	0.003	0.027
Left-PrimAuditory (41)	0.564	-45	-35	10	0.015	0.007	0.003	0.026
Left-BA9	0.207	-20	25	20	0.014	0.007	0.003	0.026
Right-BA9	0.115	10	35	20	0.014	0.007	0.002	0.025
Left-BA10	0.278	-20	35	10	0.013	0.007	0.002	0.025
Left-BA8	0.229	-20	20	25	0.013	0.007	0.001	0.024

Note: Coordinates are in the Montreal Institute of Technology (MNI) system. Activation is the raw mean in log scale.

Epochs from the direct-threat condition were sourced localized with exact low resolution brain electromagnetic tomography (eLORETA) (Pascual-Marqui et al., 2018), using standard values in the Pyhton package MNE (Gramfort et al., 2014), involving 14629 voxels with a minimal distance of sources from skull of 5mm; coordinates in the Montreat Institute of Technolgy atlas (MNI). Due to lack of participants' anatomical images, a Free Surfer average brain was used. The computed sources from the whole epoch, per participant (36) were averaged in the 500-800ms time-window over the time dimension, and extracted as images in NIFTI format. Data and plots from these images were handled using Python packages Nibabel and Nilearn (see: Abraham et al., 2014). For simplicity and to more easily achieve convergence, resulting estimated activation was analysed via Bayesian regression, simply using normal fixed priors and a normal distribution for log-transformed activation. Two analyses were conducted, one

containing an interaction between 4877 voxels extracted from 38 regions of interest (ROIs) and BIS scores; the 38 ROIs were selected as the main areas composing emotional language and anxiety processing (see: Kotz and Paulmann, 2011; Robinson et al., 2019). The second model contains an interaction between the total 116 observed anatomical areas and BIS score, to attest for pooled activity across areas that might have been missed by the first model. Table 7.7 summarises results from effects in voxels showing HDIs over zero, showing peaks at cytoarchitecnotic Brodmann areas (BA). Note that they do not go over 2SD ROPEs, and few go over 1SD ROPEs.
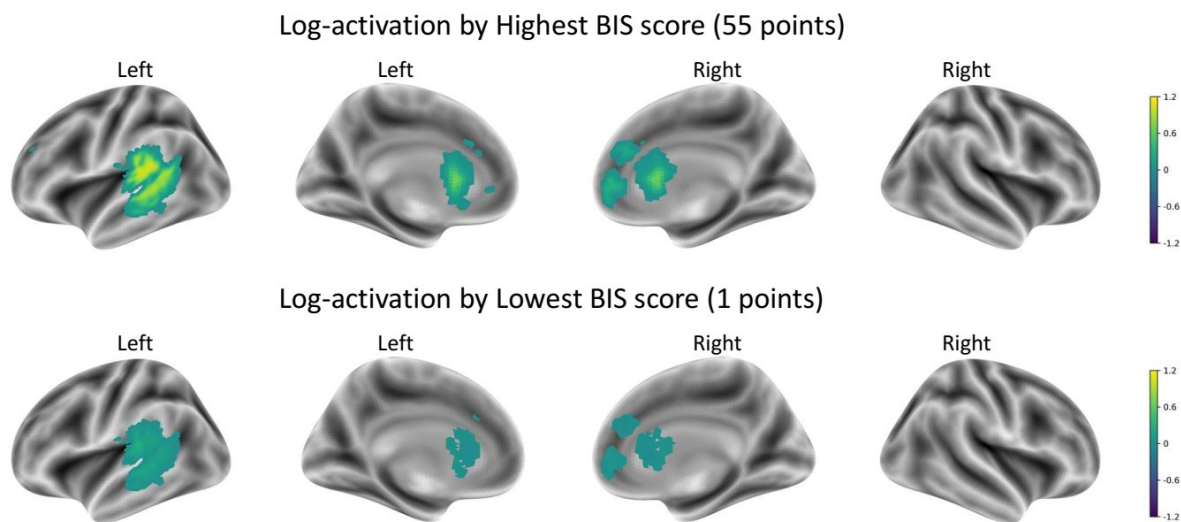
Figure 7.12. Estimated log activation from voxel by BIS interaction. Images show estimated log-activation at the lowest BIS (bottom) and highest BIS (top) levels, plotted over inflated brain surfaces. The strongest increase from lowest to highest BIS happens at left superior temporal cortex (STC) and mid temporal cortex (MTC), followed by caudate nucleus (CN) and later by middle prefrontal cortex (mPFC) at very ventral and interior parts.

Voxel-based results confirm activity in left auditory cortex and temporal language-associated cortex, the greatest increase is 1.08 log-activation a.u. from lowest BIS (0.02 a.u.) to highest BIS (1.1 a.u.) score. In addition, bilateral mPFC, and bilateral caudate nucleus (CN) was observed. This goes in line with predictions of prefrontal, temporal and basal ganglia involvement, but they do not show amygdala or hippocampal activity (See Figure 7.12). As some imprecision is expected from source localization when using average anatomies, some portions of CN might be overlapping with hippocampus and/or

amygdala. The second analysis, summarized in Figure 7.13, shows that when pooling across anatomical region (average activity by area), estimates indicate strong increases of activity (~0.01 log-activation a.u. by BIS point) in areas such as mPFC, CN, hippocampus and amygdala, but also in portions of entorhinal cortex, which connects with temporal cortex. Indeed, some of the greatest increases as a function of BIS occur in the temporal pole, MTC and STC, extending to supramarginal gyrus (SMG). In addition, these analyses also indicate that activity in visual and motor areas tends to decrease as a function of BIS. This is in line with participants withholding responses during LPC activity, as response preparation will come later, signalled by a response-locked P600. Although these results are exploratory and would require further support from MEG and/or fMRI analyses, they are in line with emotional language processing and anxiety operating in tandem as predicted by current models.
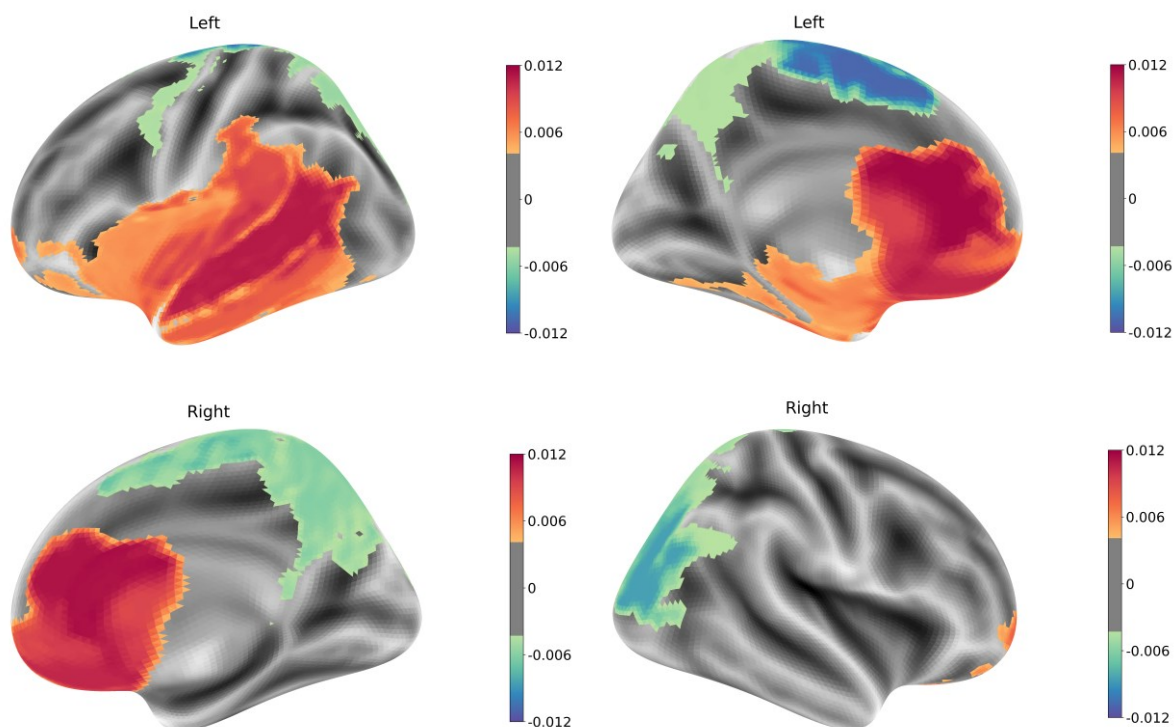
Figure 7.13. Estimated log-activations from BIS by anatomical area interaction. Threshold is placed at the biggest 2SD ROPE from all estimates (0.0042). Note that maximal increases occur at bilateral basal ganglia and left temporal cortex, followed by left amygdala, hippocampus, entorhinal cortex and insula and parts of right prefrontal cortex; while maximal decreases can be seen at left motor cortex, followed by parts of bilateral parietal and right occipital cortex. Scales are in arbitrary units.

7.4 Discussion

Present results indicate that threatening stimuli in both direct- and indirect-threat tasks induce longer RTs as a function of BIS (trait anxiety), but have no effects on accuracy, which only shows differences when selecting slopes by ear or type, indicating conventional dichotic listening effects (see the OSF repository: https://osf.io/n5b6h/). These results replicate the previous experiment very closely (Chapter 5, Experiment 1), consistent with the interpretation of trait anxiety disrupting orientation phases, as accuracy is minorly affected but RTs are strongly affected by trait anxiety levels. EEG results indicate that an LPC increases its amplitude as a function of BIS, this is also consistent with present hypothesis indicating late phase effects of trait anxiety on threatening language processing during orientative (deliberation) phases. However, no earlier effects can be observed as a function of BIS, which goes contrary to the hypotheses predicting the influence of trait anxiety at all time-windows from sentence's onset.

ERP exploratory analyses indicate that when epochs are locked to RT, a P600 signals response preparation. This result rules out the observed LPC as a P600-like ERP, supporting the notion of its relationship with trait anxiety. In addition, the response-locked P600 is modulated by BIS, showing small type by ear effects. Finally, a source localization exploratory analysis indicates that possible sources for the LPC are direct matches with areas involved with emotional language and anxiety processing, as predicted by the current implemented models, but not closely matching the present prediction. Instead of observing stronger RH activity at the ~600ms time-window, main analyses indicate stronger and clearer LH activity, and the exploratory source localization shows that this activity might come from left temporal cortex. Overall, results and exploratory results support the idea of trait anxiety influencing late phase processing (over 500ms), which is associated with response slowdown. Such effects could be explained by verbal repetitive thinking during a pre-response deliberation phase, though present results do not provide direct evidence of this and alternative mechanisms are possible.

Although some EEG studies show lateralization effects associated with emotional semantic and prosody variation (Kotz and Paulmann, 2007), recent research does not show much evidence for these effects (Chen et al., 2011; Paulmann et al., 2012). One proposed explanation is that concurrent multi-channel (prosody and semantic) information obscures laterality effects (Paulmann et al., 2012). Present results are

consistent with this interpretation, in particular when considering that when pooling amplitude by electrode the ERP pattern shows more canonical, though very small, ERP responses to emotional language at earlier time-windows. Most importantly, effects of anxiety at later time-windows in the form of an LPC are not modulated by stimulus type (semantic or prosody). This differs from previous research, which did not include anxiety measures, finding late positive amplitudes in association with prosody/semantic emotional variations and/or congruency effects (Astésano et al., 2004; Chen et al., 2011; Zhao et al., 2015). Thus, it may be the case that the distinct pattern of present LPC signals a more general anxious response to threatening language, similar to previously reported re-appraisal related LPPs (e.g. Hajcak et al., 2010).

However, previous EEG research studying the effects of visually presented (written) abusive words on non-anxious people (Wabnitz et al., 2012) shows much earlier effects, modulating P100 (~100ms) and N400 (~400ms) amplitudes, both ERPs showed greater deflections for threatening words. When a similar procedure included socially anxious subjects (Wabnitz et al., 2015), results indicated a decreased P100 and an unaffected N400 for socially anxious people. Differently, present results indicate very weak evidence for anxiety effects at early or mid-phase time-windows. This aligns better with a study testing the effects of emotional syllables and vocalization, which found that trait anxiety is associated with a lower LPC when participants listened to emotional speech (Pell et al., 2015). This pattern, however, is the opposite of present observations, though both LPCs share very similar latencies and scalp distributions. One possible explanation is that when answering to more controlled stimuli, anxious participants resolve responses at much earlier time-windows, where larger components at an early-mid stage (~100-300ms), such as P2 or EPN, are observed (Pell et al., 2015; Wabnitz et al., 2015).

This could imply that the present LPC requires the deliberation time provided by present stimuli, namely long duration semi-naturalistic sentences. Trait anxiety has been strongly associated with patterns of repetitive thinking (McEvoy et al., 2010; McLaughlin et al., 2007). This has been also proposed as an explanation for lateralisation patterns in higher worriers (Spielberg et al., 2013). It is possible that predispositions to worry and/or rumination, elicited by an overactive BIS, induce verbal repetitive thinking, in particular when stimuli are threatening speech. In line with this interpretation, LPC has been understood as a marker of decision modulation in relation to evaluation through

memory processes (Finningan, 2002; Yang et al., 2019). This LPC was elicited in response to judgements on words, and it was a left-lateral central-parietal positive deflection. This partially aligns with present observations, though some differences might be expected as present research uses a higher density setup (64 electrodes) with average reference. Nonetheless, the role of BIS in present observations must be emphasised. That is, the present LPC would not only be associated with response control, but also with that response control in terms of behavioural inhibition. This would imply that trait anxiety, expressing an overactive BIS, would involve over-engagement with threatening stimuli and induce verbal repetitive thinking. This would result in the involvement of left hemisphere, in particular of structures associated with auditory and language processing, understood as an inner speech phonological loop (Buchsbaum and D'Esposito, 2008; Vigliocco and Hartsuiker, 2002). Evidence from fMRI suggest that inner speech mainly involves areas such as inferior frontal gyrus (IFG), angular gyrus and SMG (Geva, 2018).

Present exploratory source localisation analyses, differently, show weak IFG but strong left primary auditory cortex activity, MTC and STC extending to SMG. This might be associated with source localization capturing shorter time-frames as opposed to fMRI capturing longer processing, which might end up in the recruitment of articulatory areas. Also, this may respond to present inner speech, possibly associated with verbal repetitive thinking, being focused more on imagery than rehearsal, thus deriving towards areas involved in phonological and semantic representation (e.g. BA21, BA41). Also, these analyses indicate decreased motor activity (mainly BA6) and increased mPFC activity; which could imply greater inhibition during verbal repetitive thinking, explaining slower RTs and supporting the notion that present LPC occurs before response control. This is indicated by RT-aligned exploratory analyses, which show a positive deflection starting at around 300ms and sustained until over a second. Similar ERPs have been associated with cortical re-orientation towards execution/inhibition of responses (Sassenhagen et al., 2014; Sassenhagen and Bornkessel-Schlesewsky, 2015); resulting from adrenergic secretion eliciting shifts between ventral and dorsal attention networks; where a right-lateralized ventral network (including mPFC) orients attention to environment and inhibits a goal-oriented dorsal network (Corbetta et al., 2008; Vuilleumier and Driver, 2007). Previous research suggest that BIS-related inhibitory control is linked to right hemisphere (RH) frontal cortex (e.g. mPFC) (Gable et al., 2017), in particular when tasks require refraining responses (Neo et al., 2011; McNaughton et al., 2013).

Nevertheless, when tasks are associated with rumination or worry, a strong RH pattern is not usually observed (De Pascalis et al., 2013; Keune et al., 2012). This might indicate an interplay between right and left hemispheres, where the latter works as an inhibitor of action (either approach or withdrawal) and the former as an inhibitor of ongoing deliberation. In other words, when responses need to be refrained, the ongoing required behaviour is response-withdrawal. Right-handed anxious participants might find difficult to refrain their responses due to left hemisphere interference (left-lateralized movement control). Similarly, if the task induces language-related repetitive thinking, LH might interfere through rumination or worry, thus RH would need to inhibit that ongoing activity to permit responses. As present task did not involve a go-no-go paradigm, but all responses needed to be delayed, the additional inhibition time may have allowed for the interaction between behaviour control and evaluation mechanisms, effectively involving left hemisphere temporal structures, limbic system and prefrontal structures. This might imply that more networks could be involved, in particular those associated with behavioural inhibition as mediated by serotoninergic control through raphe nucleus and amygdala-hippocampal routes (Andrade et al., 2013). And, also those involving language areas and basal ganglia, the latter playing a role in verbal emotional language processing (Paulmann et al., 2009; Pell and Leonard, 2003). To note, exploratory analysis also showed strong activity in basal ganglia, in particular CN, accompanied by smaller degrees of amygdala, hippocampal and entorhinal activity. Hence, there could be an interplay between these systems, where attention control is also mediated by BIS in terms of learning and memory. For instance, by comparing current environment conditions with past experience; consistent with notions of worry and rumination as post-event processing repetitive thinking (McEvoy et al., 2010). Therefore, increased left hemisphere activity elicited by repetitive thinking related processes such as internal dialogue and categorization remains a plausible explanation for present results showing that anxiety has no major effects on accuracy but is associated with increased RTs preceded by an increased LPC during sentence processing.

Under this notion, inhibition processes need to be transferred from right to left hemisphere, which would induce faster recession of right hemisphere activity, longer and stronger latency of left hemisphere activity, and different scalp distributions at each hemisphere. Although this is a parsimonious explanation when theories of callosal relay are taken into account (Friederici et al., 2007; Grimshaw et al., 2003; Steinmann et al.,

2017; Westerhausen and Hugdahl, 2008), it becomes difficult to explain why there are not earlier ear or type differences in association with hemisphericity patterns. One possible explanation is the horse-race model of callosal transferring (Grimshaw et al., 2003), which indicates that ear of presentation would privilege the contralateral hemisphere not as an immediate effect, but in the long run. This would be the effect of the delay caused by callosal transferring. Indeed, main analyses, using epochs locked to sentence onset, do not show amplitude differences of this sort. However, when epochs are locked to RT, these differences (though small) become evident, in particular for Semantic stimuli at right ear and with slight central-left posterior pattern (e.g. Oz, O2). Nonetheless, due to BIS effects, instead of creating an advantage, responses become slower. This would emphasise that deliberation is more effortful for people with higher trait anxiety, which is reflected by an increased LPC.

Overall, differently from previous observations (Pell et al., 2015), main results show an LPC increase as a function of trait anxiety; indicating that anxiety can have different effects when task and stimuli change (i.e. whether they allow deliberation to occur). Thus, present observations indicate that core features of auditory and emotional language processing (Kotz and Paulmann, 2011; Frühholz et al., 2016a) might drastically or unexpectedly vary when taking individual differences into account (i.e. trait anxiety). In other words, BIS effects on threatening language processing may reflect a tight relationship between anxiety related processes (i.e. worry/rumination-induced verbal repetitive thinking) and language processes (i.e. phonological loop). Indeed, in anatomical terms, emotional auditory processing models (Früholz et al., 2016a) are strongly consistent with current models of anxiety processing (Robinson et al., 2019). In the time domain, anxious attention models (Bar-Haim et al., 2007; Cisler and Koster, 2010) and models of emotional language (Kotz and Paulmann, 2011) also show similarities. Taking the similarities between these models and present results into account, it is possible to conclude that effects of trait anxiety evidence the relevance of including a late fourth deliberation phase in a model of threatening language processing. This expanded model would include the following phases: 1) Perception (~100ms), 2) Recognition (~200ms), 3) Evaluation (~400ms), 4) Deliberation (~600ms).

Finally, some relevant limitations need to be addressed. Firstly, an obvious caveat is that there is no direct neutral control. Even though we used an indirect-threat task, this is usually intended to control for attentional differences (e.g. Sander et al., 2005; Peschard

et al., 2016), as participants still listen a threatening sentence at the contralateral ear. Although the experimental requirements of present task design do not allow for having a response to Neutral stimuli without threat interference, future research could include a proper neutral speech control by, for example, using between-groups designs. Secondly, although some patterns of right hemisphere activity associated with anxiety were identified in addition to left hemisphere patterns, it is important to address experimental designs that can directly induce response inhibition to differentiated stimuli. Previous research has indeed proposed that tasks such as go-no-go are better for understanding BIS processes as they directly involve inhibiting responses (e.g. McNaughton et al., 2013; Neo et al., 2011). Thirdly, the inherent caveat of semi-naturalistic sentences, where many lexical items appear later in the sentence than prosody acoustic changes, can be an important limitation for dichotic listening experiments. Amongst other things this could imply why canonical ERP patterns were weak and difficult to find on present results, differently from previous dichotic or non-dichotic research using more controlled stimuli (e.g. Pell et al., 2015; Wambacq and Jerger, 2004). So, additional experiments using controlled stimuli are required as a comparison to and extension of present research.

Future research could include such tasks by, for instance, including responses to stimuli type instead to threat in general. In association with this, spotting effective hemisphericity differences requires the support of more precise anatomical measurements. Although previous fMRI studies (e.g. Sander et al., 2005; Spielberg et al., 2013) partially match with the proposed extended model, they do not include tasks equivalent to present experimental task, so differences can be expected. This implies that fMRI or MEG research, or at least more precise source localization (i.e. using individual anatomies) are required for expanding present results and test the extended model on the anatomical side. Last but not least, the possible influence of verbal repetitive thinking needs to be directly tested, with tasks specifically designed to do so (e.g. Nalborczyk et al., 2017). Thus, experimental design using methods such as language interference could be relevant to address whether anxiety-induced verbal repetitive thinking has particular effects on threatening language.

In conclusion, present experimental results replicate previous behavioural results and provide strong evidence for trait anxiety effects. ERP analyses show that these responses are preceded by a clear positive amplitude deflection in temporal electrodes. This event was interpreted as an LPC, and was associated with trait anxiety affecting

deliberation processes through verbal repetitive thinking. As BIS scores increased, the LPC became more positive, suggesting a disruptive deliberation process, such as induced by over-engagement with threat, during an orientative stage. In other words, higher levels of anxiety due to their association with repetitive thinking (i.e. worry and rumination) induce a more effortful decision-making process after evaluation of threatening stimuli. Although this implies strong support for part of the present hypotheses, several limitations were discussed, and need to be addressed. The development of more sensitive tasks and alternative statistical models and approaches would help to understand earlier processes underpinning the presently proposed deliberation stage. In addition, the observed relationship between anxiety and language is not specific to speech information type. Hence, the effects of anxiety on prosody and semantics remain to be further explored. Overall, this experiment paves the way for future research on the relationship between speech, individual differences and emotional language in terms of information channels, anxiety and threatening language. This outlines an improved basis for reshaping models of emotional language processing by providing clear evidence on how anxiety can affect late processing phases.

Chapter 8
Thesis Discussion
A Model of Threatening Speech and Anxiety

8.1 Evidence for The Operative Model

The present thesis focusses on providing a theoretical background and supporting evidence for the interaction between trait anxiety and threatening speech processing. This has been guided by an operative model proposing that anxiety should induce diverse effects depending on processing phase (see Figure 2.1). However, both behavioural and electrophysiological evidence from previous chapters indicate that noticeable effects are mainly induced at late processing phases. These phases are theoretically defined from a multistep model of emotional language processing (Kotz and Paulmann, 2011). As explained in previous chapters, in particular in Chapter 2, this model proposes three processing stages that can be summarised as perception (~100ms), recognition (~200ms) and evaluation (~400ms). However, this model does not distinguish between different types of emotional stimulation or individual differences. To elucidate possible implications of singular emotional stimulation on individual differences, the present research project focused on testing specific effects of threatening speech meaning on trait anxiety. As stimuli are not presented as unconditioned stimuli (i.e. participants listened to threat continuously), it was theoretically considered that effects of threatening speech should have a clearer impact in anxiety processing rather than fear (McNaughton, 2011; Robinson et al., 2019). Furthermore, according to phasic models of anxiety (Bar-Haim et al., 2007; Cisler and Koster, 2010), it was considered that differentiated effects of anxiety should be evident as early over-attention to threat and later over-engagement with anxiety. Hence, the proposed operative model is an extension of the multistep model of emotional language processing with an additional phase, understood as an orientation or deliberation phase.

Present behavioural and EEG results indicate that earlier (100-300ms) and even mid-late processing stages (~400ms) of language processing are not affected by anxiety, at least not directly. Only later processing stages (~600ms) evidence differences as a function of trait anxiety. Chapter 5 aimed to distinguish the effects of trait anxiety, or anxious apprehension measured as worry level (Heller et al., 1997; Nitschke et al., 1999;

Spielberg et al., 2013), on threatening semantics and prosody in two dichotic listening experiments. Participants listened to threatening sentences paired with neutral ones and had to indicate at which ear they heard the threatening sentence, which could contain prosodic threat, semantic threat or both (acoustically and lexically normed and classified in detail on Chapter 4). In the first experiment, participants answered only after sentences ended and in the second experiment they answered before sentences ended. The rationale of this approach was that responses without time pressure (delayed responses: after sentence end) would be affected by late-phase effects of anxiety (over-engagement with threat), while fast responses (before sentence end) would elicit over-attention to threat effects. Nevertheless, both experiments induced similar effects: response times increased as a function of worry level, independent of sentence type and with no detriment to accuracy. Chapter 6 showed that this reaction time increase also happened in a non-dichotic fast-response experiment (in-principle replication of Chapter 5's fast-response task). This indicates no certain accuracy differences and no major influence of sentence type on responses. Finally, Chapter 7 showed that in delayed-response task (directly replicating the delayed task in Chapter 5), ERPs locked to sentence's onset indicate an amplitude increase as a function of anxiety peaking around 600ms (LPC) and stronger at left temporal electrodes. This was accompanied by slower reaction times as a function of anxiety and by exploratory analyses indicating that reaction-time-locked ERPs are not associated with observed LPCs and a source localization showing strong basal ganglia, temporal and prefrontal activity at the LPC time-window.

With this evidence in mind, presently hypothesised effects of trait anxiety on right-lateralised prosody at earlier processing stages or on left-lateralised semantics at mid-late stages were not observed. These hypotheses were derived from the multistep model (Kotz and Paulmann, 2011) by specifying differences in speech's acoustic properties and how anxiety could affect their lateralisation patterns at aforementioned time-windows. These lateralisation patterns are mainly based on the anatomical patterns of segmental versus suprasegmental information types (Poeppel et al., 2008; Zatorre et al., 2002), and on dorsality/ventrality patterns of the dual stream model of speech processing (Hickok and Peoppel, 2007). When considering similar patterns of laterality/dorsality found in emotional attention models (Corbetta et al., 2008; Vuilleumier and Driver, 2007) and, in particular, anxiety models (Calhoon and Tye,

2015; McNaughton, 2011; Robinson et al., 2019), a possible relationship emerges. The proposed interaction between anxiety/attention and emotional speech processing was not observed at early- or mid-phase processing, but was shown to be constrained at later processing stages. This may be indicative of a dissociation between anxiety and fear and a partial dissociation between anxious arousal and anxious apprehension (Heller et al., 1997; McNaughton and Corr, 2004), as mentioned above. More importantly, this indicates that trait anxiety, understood as an overactive behavioural inhibition system (BIS), is associated with over-engagement with threat, in particular when stimuli (i.e. long semi-naturalistic sentences) and/or task allow for this over-engagement to occur. Even so anxious delayed disengagement from threatening stimuli has been shown to happen after short and long stimuli presentation (Cisler and Koster, 2010), such response may be dependent on several factors such as length of inter stimulus interval, type of stimuli (i.e. modality, information type), or differences between fear and anxiety and between anxious arousal and anxious apprehension; all of which can affect over-engagement with threat.

For instance, EEG studies have observed that worry is associated with negative reappraisal of visual stimuli and parietal late positive potential (LPP) increases (Moser et al., 2014), namely late-phase processing associated with sustained attention (Hajcak et al., 2010). In other words: delayed disengagement, over-engagement, or deliberation. Studies using a cross-modal paradigm, priming words with emotional faces to observe implicit reappraisal have observed early-phase (e.g. N170) but not late-phase (i.e. LPP) effects (Liu et al., 2018), suggesting that explicit engagement (deliberation) or more engagement time is required for later phases to be affected. Indeed, EEG studies focusing on social anxiety have shown that abusive words (short duration stimuli) induce effects on early- or mid-phase ERPs (i.e. P1, N400) which could be better related with over-attention to threat or threat evaluation issues (Wabnitz et al., 2015). Other studies have shown that angry vocalisations (prosody) are associated with early (i.e. P1) but not late (i.e. LPC) amplitude increases associated with anxiety (Pell et al., 2015). FMRI experiments using longer stimuli or tasks requiring sustained attention (e.g. emotional Stroop or dichotic listening) have shown increased activity in phonological loop and attention control-related areas in association with worry/rumination or BIS (Sander et al., 2005; Spielberg et al., 2013). In short, the broad architecture of the present operative model still holds, but the specific effects given anxiety type (i.e. arousal or apprehension)

need to be more precisely specified at each phase, such as was initially proposed for fear-anxiety dissociations and prosody-semantic processing differences.

8.2 Revision of The Operative Model

This leads to a revision of the proposed operative model, which indicates that trait anxiety, in particular given its worry and rumination effects, will be disruptive only at late processing phases where the stimulus has already been evaluated. Hence, the present model needs to either be revised phase-by-phase, indicating effects of different anxiety-types or states at each processing phase, or different more specific models need to be built for each anxious type. Opting for the latter option, a revised model will be presented only for the apprehensive anxious type, namely related with anxiety traits or states associated with rumination and/or worry. This revised model attempts to account for the phenomena observed in previous chapters, by explaining why trait anxiety has such strong effects at late phases, especially when stimuli allow for long periods before responses. The trend followed in the present thesis is to attribute these late phase effects to verbal repetitive thinking, as this could disrupt speech processing by dragging resources towards worry or rumination. Verbal repetitive thinking has been systematically associated with anxiety (McEvoy et al., 2010), and is thought to depend on inner speech; which occurs via the phonological loop, using speech production networks mainly involving left lateral SMG, STC, MTC, and IFG (Geva, 2018). Consistent with recent research observing that electromyographic activity (from speech production-related facial muscles) increases as a function of participants' rumination levels (Nalborczyk et al., 2017). Even though this is aligned with the present interpretation of trait anxiety (i.e. BIS or worry) affecting late phase effects, no strong conclusion regarding repetitive thinking can be drawn yet, as verbal repetitive thinking has not been directly assessed.

Nevertheless, although these late phase effects of trait anxiety have been associated with over-engagement or reappraisal (Hajcak et al., 2010), no proposal has attempted to demonstrate how this over-engagement happens. Furthermore, no focus has been placed on the fact that when studies use negative language of brief duration and emphasising fast or primed responses (e.g. Liu et al, 2018; Pell et al., 2015; Wabnitz et al., 2015), no evidence for or decreases of late phase ERPs (LPC/LPP) are observed in association with anxiety. But, as the present thesis suggests, when semi-naturalistic

longer sentences are used (and in particular when delayed responses are required), both reaction times and LPCs greatly increase as a function of worry and BIS (in addition to exploratory source localisation indicating engagement of both language and anxiety processing areas at LPC phase). Given this, verbal repetitive thinking is a good candidate for explaining late-phase effects of trait anxiety, not only because it is associated with rumination and worry in anxiety (McEvoy et al., 2010; McLaughlin et al., 2007), but also because it engages language processing and production networks (Nalborczyk et al., 2017); which might be required for efficiently processing of complex utterances. In this way, present stimuli, well-defined as prosodically and semantically threatening by very specific acoustic (pitch and voice quality) and lexical (arousal and valence) characteristics, could induce over-engagement on participants with increased trait anxiety. These people would transiently worry or ruminate over the stimuli while deciding how to respond (deliberation) and this would be enacted through verbal repetitive thinking, thus engaging both anxiety-BIS-related and language-related networks. This would start when or after stimuli are evaluated near 300-400ms according to the multistep model of emotional language and be sustained for a longer period (i.e. LPC around 600ms), thus explaining over-engagement and subsequent slower responses.

8.3 A Model of Threatening Speech Processing in Anxiety

According to multiphasic models (Kotz and Paulmann, 2011; Bar-Haim et al., 2007; Cisler and Koster, 2010), the processing of an emotional stimulus should technically start shortly after stimulus onset, when early (pre-aware, around 100ms, e.g. P1) recognition processes begin, which develop into stimulus recognition (awareness, around 200ms, e.g. N2) and evaluation (integration, around 300-400ms, e.g. N400). If tasks do not present strong time constraints, deliberation (from around 600ms, e.g. LPC) can take place and be sustained until responses are prepared/disinhibited (if responses are required). Note that this event might take place in several circumstances, such as if responses are locked to stimulus offset or stimulus is too short, in which case events such as response-related P300-P600 (e.g. Sassenhagen et al., 2014) could be observed. Evidence linking P300 and LPP responses indicates that starting from 300ms, if stimuli appraisal is required an LPP can develop, signalling both emotional and cognitive processing, which is sensitive to individual differences such as anxiety (Hajcak et al.,

2010). Although not necessarily equivalent to LPP, the present LPC might be associated with similar appraisal processes, but more directly linked to speech processing deliberation phases and the behavioural inhibition system.

Figure 8.1 details the revised model, highlighting the aforementioned phases and some possibly involved brain areas, as interpreted from emotional language (Kotz and Paulmann, 2011) and anxiety models (Robinson et al., 2019), but also from possible involvement of the phonological loop. The main implication is that when stimuli have longer durations (i.e. semi-naturalistic utterances), threat is understood as potential and not actual. This can be associated to the complexity of language stimuli, but also to the very nature of language (i.e. language cannot produce vital damage). This requirement is fundamental to any model of anxiety, as anxiety is sourced on behavioural inhibition due to approach-withdrawal assessment and not on the immediate flee/freeze/fight fear response to direct threat (McNaughton, 2011; Robinson et al., 2019). Therefore, during threatening speech processing higher levels of trait anxiety can induce inhibition and delay responses, as long as state/trait arousal is sufficiently low and/or sufficient processing time is allowed. As suggested by present research, some type of disruption occurs rapidly in the scale of hundreds of milliseconds. These burst of activity that delay responses for anxious people could be either induced by sudden brief concern about the upcoming response (worry) or short rehearsal of the past threatening stimulus (rumination).

Either way, verbal repetitive thinking could take hold of the phonological loop (STC, MTG and IG in Figure 8.1) for a short time, while threat is re-assessed and the anxious response cycles (dlPFC, vmPFC, amygdala, hippocampus route in Figure 8.1). Mediating these two networks, entorhinal cortex (EC in Figure 8.1) links the amygdala-hippocampal route with temporal cortex (densely connected with both) and mediates contextual fear learning (Calhoon and Tye, 2015), thus allowing BIS-related environmental assessment in relation to language. While basal ganglia (BG in Figure 8.1) might mediate responses as it is associated with vmPFC anxiety/reward mediation, maybe after dlPFC input (Robinson et al., 2019), by inhibiting amygdala or activating BG (Calhoon and Tye, 2015); BG involvement which has also been observed in emotional language evaluation (Paulmann et al., 2009). Thus, this BG involvement, in particular nucleus accumbens (Calhoon and Tye, 2015; Paulmann et al., 2009), may be linked to upcoming response execution, such as mediating appetitive/approach disinhibition

upon response evaluated as correct/incorrect (i.e. threatening or not threatening). This whole process occurring during an LPC time-window, however, might slightly differ from a long duration ruminative/worry process. Because, even if stimuli are long in the conventional experimental scale, where commonly used stimuli (e.g. faces, words, vocalisations) are very brief (few hundreds of milliseconds), they are short in terms of the normal verbal repetitive thinking process (i.e. minutes or more). Hence, the brief period of verbal repetitive thinking portrayed by the model, might be better characterised as transient brooding.

This model is also consistent with anxiety driving later ear differences (response-aligned), instead of earlier differences; aligned with callosal relay or horse-race models of dichotic listening (Grimshaw et al., 2003; Steinmann et al., 2017), as these models propose that information transferred between hemispheres through corpus callosum could make left or right hemisphere advantages to express at different processing phases depending on features of stimulus or task. This would explain small exploratory effects of semantic threat when ERPs are aligned with response-times. In addition, although not directly addressed through present research, callosal transferring could explain why present behavioural dichotic listening results show consistency with more conventional dichotic listening results when not accounting for anxiety (e.g. Wambacq and Jerger, 2004; Witteman et al., 2014). Hence, increases in anxiety could imply extended processing, facilitating callosal transferring, thus 'evening out' any possible ear advantage and early-phase ERP differences. This is also supported by exploratory results from Chapter 7, showing a strong positivity when aligning epochs with reaction times, which is consistent with models of response preparation ERPs (Sassenhagen et al., 2014). These models directly link with attention control models (e.g. Corbetta et al., 2008), and show that anxiety could be indeed involved in delaying responses by extending deliberation periods. These models also propose right frontal (e.g. vmPFC) involvement, such as in anxiety models (Robinson et al., 2019), which could be involved in the regulation of amygdala and basal nucleus of the stria terminalis (BNST). So, intermediate areas such as BG and EC might mediate this right to left or left to right transferring, supporting processing associated with both speech (Paulmann et al., 2009) and anxiety (Robinson et al., 2019). This completes a very precise anatomical network which is proposed to be involved in late phase processing from 500-600ms up to 1000-1200ms if threatening speech is sustained for long enough.

Evaluation	Deliberation	Preparation
Process: Integration of emotional content. Example ERP: N400	Process: Assessment of stimulus in emotional/cognitive terms. Example ERP: LPC	Process: response preparation/disinhibition. Example ERP: P300-P600
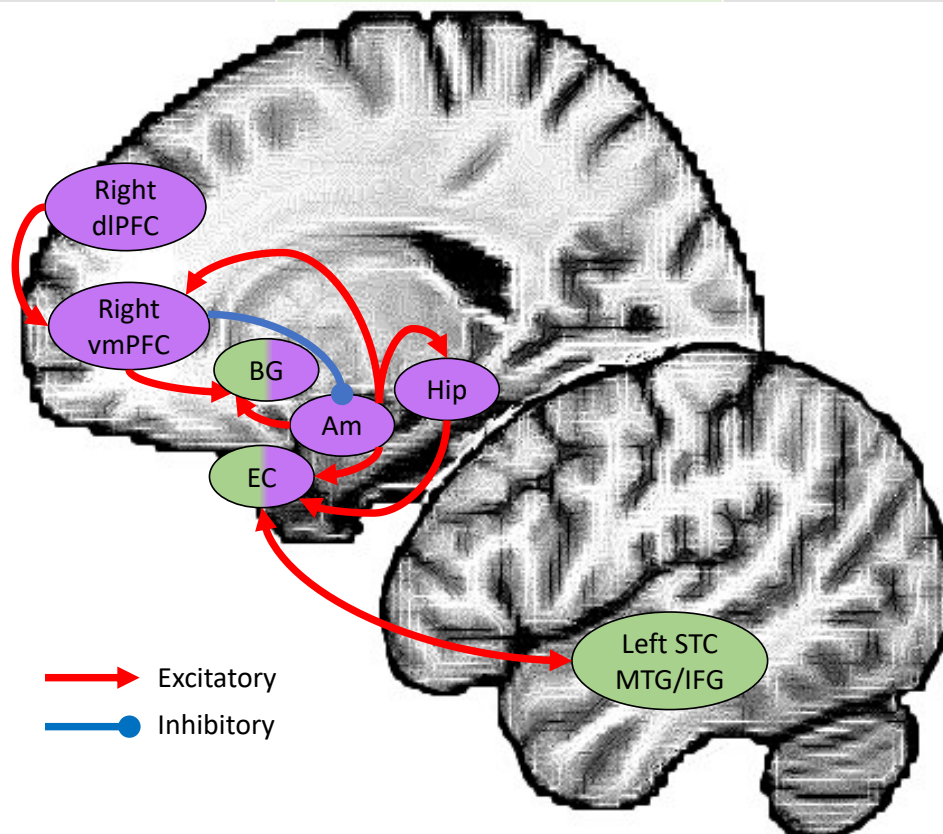
Figure 8.1. A model of threatening speech processing in anxiety. Image above shows the hypothesised pattern of a late deliberation phase, starting from around 600ms. The proposed pattern is understood to be the result of a link between anxiety and language through transient brooding. Note that LPC is preceded by an integration phase (~300-400ms) and followed by a response preparation phase (~300-600ms), this overlap occurs because if time is constrained evaluation is faster (i.e. less or not explicit) or directly precedes response preparation. Otherwise, deliberation can happen or is noticeable until response disinhibition begins (e.g. stimulus offset). Also note that these responses are assumed to be locked to stimulus onset. Below, areas proposed to be involved during deliberation. Purple ellipses indicate areas commonly associated with anxiety. Green ellipses indicate areas commonly associated with emotional language processing. Entorhinal cortex is proposed as an additional structure associating environment evaluation and language, e.g. though environmental scanning and comparison to past experience. dLPFC: dorso-lateral prefrontal cortex. vmPFC: ventro-medial prefrontal cortex. BG: basal ganglia. Am: amygdala. Hip: hippocampus. EC: entorhinal cortex. STC: superior temporal cortex. MTG: middle temporal gyrus. IFG: inferior frontal gyrus.

Having proposed a new processing model there are some limitations that need to be addressed. The most salient one of these is that the present research project lacks experiments providing a direct measure of verbal repetitive thinking (e.g. Nalborczyk et al., 2017). This is a hypothesis in and of itself, which needs to be tested by new research. Experiments testing the effects of disruptive intermediary tasks placed between threatening speech stimuli and response will be required for this. A possible experiment would simply imply occupying speech-related processes with a rhyming task as compared to a visual rotation task and a baseline (no task). Hypothetically, the rhyming task should block the effects of trait anxiety by blocking transient brooding, so reaction times and ERPs in this condition should not change as a function of anxiety. Further behavioural research could also involve induced state anxiety measurements, such as threat of shock or similar manipulations (e.g. Aylward et al., 2017). Other important limitation is the lack of passive listening research, which could involve the analysis of both ERP and brain waves, helping to elucidate effects on sustained brain activity (e.g. McNaughton et al., 2013). For instance. this could allow to determine whether anxiety-related theta-waves are also associated with processes specific to language. This would complement current ERP and source-localisation analyses. Finally, replicating or extending current research in MEG, fMRI or transcranial stimulation settings could reveal whether present speculation on involved anatomical networks are appropriate. All these approaches would put the here presented model to the test and see whether it holds or needs to be amended or replaced.

In addition, there are more practical limitations that need to be addressed. For instance, stimuli norming could have been affected by un-even samples (i.e. more females than males). Further research, with a more diverse sample needs to address this issue, especially by testing the predictive accuracy of present data (coming from un-even samples). This predictive capacity could also ameliorate the issue of present small sample sizes, which though not necessarily uncommon in previous research (e.g. Banse and Scherer, 1997; Hammerschmidt and Jurgens, 2007), may be a relevant limitation. Similarly, sample sizes of present behavioural experiments, although in close proximity to recent literature (e.g. Leshem, 2018; Peschard et al., 2016), could be increased to improve the certainty of accuracy estimates and possible compliance and error issues that may have arose due to the web-based nature of present behavioural experiments. It is also possible that variance and error were induced by the lack of temporal

alignment between prosody and semantics, as lexical items were placed anywhere within sentences but prosody always started at sentence onset. This uneven comparison between prosody and semantics may have obscured both behavioural and EEG measurements, and needs to be addressed by new experimental designs using more controlled stimuli to provide a contrast to present semi-naturalistic stimuli. Finally, the change of scales (used as a proxy of trait anxiety) from PSWQ to BIS could have induced measurement errors. Although this issue might not be so serious, given that estimates across experiments are very similar, more direct and explicit correlation measures across scales could have helped. Also, experiments exploring the effects of induced state anxiety could help to establish a more straightforward causal relationship between anxiety and present RT and ERP results (if they can be replicated in such experiments). This is relevant, as present results do not directly attest for a special effect of threatening language on anxiety, a possible over-engagement with threat effect could be explained by a more general mechanism and not necessarily by verbal repetitive thinking. Hence present interpretation of results is a good explanation, but needs to be addressed more directly by further experimental research.

With these limitations in mind, it is clear that the presently proposed model (Figure 8.1) is not definitely supported yet. This means that the fourth deliberation-related phase, as proposed here, could not be particularly distinct from an evaluation phase. Regarding behavioural effects, the RT-accuracy trade-off could be explained as a general cumulative effect in responses slow down (Robinson et al., 2013) rather than as a consequence of over-engagement with threatening speech. Even so, the effects on reaction times shown in the present experiments indicate particularly big delays of hundreds of milliseconds as a function of anxiety, which may indicate that whatever causes this effect affects anxious people strongly. Although present results do not provide direct evidence for a verbal repetitive thinking mechanism causing this response slow-down, this mechanism may be a reasonable explanation. This implies that the present model is testable and falsifiable, which could be done via the future research directions mentioned in the previous two paragraphs.

Even so, the proposed fourth deliberation phase might be an extended evaluation phase instead of a particularly distinct processing stage, where ERPs such as LPC simply reveal extended evaluation phases (Kotz and Paulmann, 2011). However, the present LCP pattern seems to be strongly associated with anxiety, where preceding ERPs are

very unclear or have very small amplitudes. This may be an artefactual effect in present data, due to misalignment between lexical items in Semantic sentences and emotional prosody in Prosody items; or due to another unidentified problem. This implies that replications and re-analyses are important for testing the feasibility of a fourth phase in emotional language processing. In addition, experiments testing possible differentiated effects of anxiety on semantics and prosody processing could include the effects of anxiolytics on clinical or subclinical populations, such as anti-anxiety drugs (e.g. McNaughton et al., 2013), or the effects of angiogenesis, such as threat of electrical shock (e.g. Robinson et al., 2013). These could also involve the use of more controlled complex stimuli with more precise prosody-semantic alignment, accompanied by clearer predictions of ERPs and/or other measures (e.g. Paulmann et al., 2009). If the present model does not hold to theses tests, alternative models can be implemented. For instance, models where the effects of language on anxiety appear and are sustained on the long term (e.g. rumination induced by sentential context), or models explaining differentiated effects of fear and anxiety on speech processing.

To conclude, this thesis proposes an operative model of threatening speech processing in trait anxiety. Three behavioural and one EEG experiments, together with two exploratory analyses, brought evidence for strong effects of trait anxiety on the processing of semi-naturalistic spoken threatening sentences, or at least on reactions to speech detection. The clear prosodic and semantic characteristics of these stimuli, as evidenced by acoustic measures, lexical norms, and sentences' ratings, provide stimuli with well-defined verbal aggression. This type of threat (potential threat) induces anxious responses characterised by relatively good accuracy but slow reaction times. In neurophysiological terms, event-related potentials (a late positive complex) strongly increase their amplitude as a function of anxiety. This possibly reveals an excitatory neural response that is sustained until behavioural responses are prepared, possibly indicating increased deliberation (over-engagement with threat) while behaviour is inhibited. From this, a model of transient brooding is proposed, where speech and language play a major role. Several hypotheses arise from this model, opening a wide path for future research on the relationship between anxiety and language.

References

Ablin, P., Cardoso, J.-F., & Gramfort, A. (2018). Faster Independent Component Analysis by Preconditioning With Hessian Approximations. *IEEE Transactions on Signal Processing*, *66*(15), 4040–4049. https://doi.org/10.1109/tsp.2018.2844203

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Alday, P. M. (2018). M/EEG analysis of naturalistic stories: a review from speech to language processing. *Language, Cognition and Neuroscience*, *34*(4), 457–473. https://doi.org/10.1080/23273798.2018.1546882

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology*, *56*(12). https://doi.org/10.1111/psyp.13451

Andrade, T. G., Zangrossi, H., & Graeff, F. G. (2013). The median raphe nucleus in anxiety revisited. *Journal of Psychopharmacology*, *27*(12), 1107–1115. https://doi.org/10.1177/0269881113499208

Anikin, A., & Lima, C. F. (2017). Perceptual and Acoustic Differences between Authentic and Acted Nonverbal Emotional Vocalizations. *Quarterly Journal of Experimental Psychology*, *71*(3), 17470218.2016.1. https://doi.org/10.1080/17470218.2016.1270976

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, *18*(2), 172–184. https://doi.org/10.1016/j.cogbrainres.2003.10.002

Atchley, R. A., Grimshaw, G., Schuster, J., & Gibson, L. (2011). Examining lateralized lexical ambiguity processing using dichotic and cross-modal tasks. *Neuropsychologia*, *49*(5), 1044–1051. https://doi.org/10.1016/j.neuropsychologia.2011.01.010

Aue, T., Cuny, C., Sander, D., & Grandjean, D. (2011). Peripheral responses to attended and unattended angry prosody: A dichotic listening paradigm. *Psychophysiology*, *48*(3), 385–392. https://doi.org/10.1111/j.1469-8986.2010.01064.x

Aylward, J., Valton, V., Goer, F., Mkrtchian, A., Lally, N., Peters, S., Limbachya, T., & Robinson, O. J. (2017). The impact of induced anxiety on affective response inhibition. *Royal Society Open Science*, *4*(6), 170084. https://doi.org/10.1098/rsos.170084

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Betancourt, M. (2016). *Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo Michael B*. https://doi.org/arXiv:1604.00695

Blanchard, R. J., Blanchard, D. C., Griebel, G., & Nutt, D. (2008). Introduction to the handbook on fear and anxiety. In R. J. Blanchard, D. C. Blanchard, G. Griebel, & D. Nutt (Eds.), *Handbook of Behavioural Neuroscience. Volume 17. Handbook of Anxiety and Fear* (pp. 3–10). Academic Press. Elsevier.

Borelli, E., Crepaldi, D., Porro, C. A., & Cacciari, C. (2018). The psycholinguistic and affective structure of words conveying pain. *PLOS ONE*, *13*(6), e0199658. https://doi.org/10.1371/journal.pone.0199658

Buchsbaum, B. R., & D'Esposito, M. (2008). The Search for the Phonological Store: From Loop to Convolution. *Journal of Cognitive Neuroscience*, *20*(5), 762–778. https://doi.org/10.1162/jocn.2008.20501

Burra, N., Kerzel, D., Munoz Tord, D., Grandjean, D., & Ceravolo, L. (2018). Early spatial attention deployment toward and away from aggressive voices. *Social Cognitive And Affective Neuroscience*, *14*(1). https://doi.org/10.1093/scan/nsy100

Busch-Moreno, S., Tuomainen, J., & Vinson, D. (2020a). Semantic and prosodic threat processing in trait anxiety: is repetitive thinking influencing responses? *Cognition and Emotion*, 1–21. https://doi.org/10.1080/02699931.2020.1804329

Busch-Moreno, S., Tuomainen, J., & Vinson, D. (2020b). *Trait Anxiety Effects on Late Phase Threatening Speech Processing: Evidence from EEG*. Submitted for publication. Preprint: https://doi.org/10.1101/2020.10.02.323642

Calhoon, G. G., & Tye, K. M. (2015). Resolving the neural circuits of anxiety. *Nature Neuroscience*, *18*(10), 1394–1404. https://doi.org/10.1038/nn.4101

Ceravolo, L., Frühholz, S., & Grandjean, D. (2016). Modulation of Auditory Spatial Attention by Angry Prosody: An fMRI Auditory Dot-Probe Study. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00216

Chen, X., Pan, Z., Wang, P., Zhang, L., & Yuan, J. (2014). EEG oscillations reflect task effects for the change detection in vocal emotion. *Cognitive Neurodynamics*, *9*(3), 351–358. https://doi.org/10.1007/s11571-014-9326-9

Chen, X., Yang, J., Gan, S., & Yang, Y. (2012). The Contribution of Sound Intensity in Vocal Emotion Perception: Behavioral and Electrophysiological Evidence. *PLoS ONE*, *7*(1), e30278. https://doi.org/10.1371/journal.pone.0030278

Chen, X., Zhao, L., Jiang, A., & Yang, Y. (2011). Event-related potential correlates of the expectancy violation effect during emotional prosody processing. *Biological Psychology*, *86*(3), 158–167. https://doi.org/10.1016/j.biopsycho.2010.11.004

Cisler, J. M., & Koster, E. H. W. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, *30*(2), 203–216. https://doi.org/10.1016/j.cpr.2009.11.003

Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*(3), 211–226. https://doi.org/10.1016/j.bandl.2011.12.007

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2013). Effects of valence and arousal on written word recognition: Time course and ERP correlates. *Neuroscience Letters*, *533*(533), 90–95. https://doi.org/10.1016/j.neulet.2012.10.054

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066x.49.12.997

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Corr, P. J., & Cooper, A. J. (2016). The Reinforcement Sensitivity Theory of Personality Questionnaire (RST-PQ): Development and validation. *Psychological Assessment*, *28*(11), 1427–1440. https://doi.org/10.1037/pas0000273

Corr, P. J., & McNaughton, N. (2012). Neuroscience and approach/avoidance personality traits: A two stage (valuation–motivation) approach. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2339–2354. https://doi.org/10.1016/j.neubiorev.2012.09.013

Damasio, A. R. (2012). *Self comes to mind : constructing the conscious brain*. Vintage.

De Pascalis, V., Cozzuto, G., Caprara, G. V., & Alessandri, G. (2013). Relations among EEG-alpha asymmetry, BIS/BAS, and dispositional optimism. *Biological Psychology*, *94*(1), 198–209. https://doi.org/10.1016/j.biopsycho.2013.05.016

Delaney-Busch, N., Wilkie, G., & Kuperberg, G. (2016). Vivid: How valence and arousal influence word processing under different task demands. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(3), 415–432. https://doi.org/10.3758/s13415-016-0402-y

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 34–43. https://doi.org/10.3758/bf03209414

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2011). Emotional Voice Areas: Anatomic Location, Functional Properties, and Structural Connections Revealed by Combined fMRI/DTI. *Cerebral Cortex*, *22*(1), 191–200. https://doi.org/10.1093/cercor/bhr113

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Fang, L., Xu, Y., Prom-on, S., & Yu, A. C. L. (2013). Morpheme-like prosodic functions: Evidence from acoustic analysis and computational modeling. *Journal of Speech Sciences*.

Finnigan, S. (2002). ERP "old/new" effects: memory strength and decisional factor(s). *Neuropsychologia*, *40*(13), 2288–2304. https://doi.org/10.1016/s0028-3932(02)00113-6

Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (2007). Role of the Corpus Callosum in Speech Comprehension: Interfacing Syntax and Prosody. *Neuron*, *53*(1), 135–145. https://doi.org/10.1016/j.neuron.2006.11.020

Frühholz, S., Gschwind, M., & Grandjean, D. (2015). Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. *NeuroImage*, *109*, 27–34. https://doi.org/10.1016/j.neuroimage.2015.01.016

Frühholz, S., Trost, W., & Kotz, S. A. (2016a). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Frühholz, S., van der Zwaag, W., Saenz, M., Belin, P., Schobert, A.-K., Vuilleumier, P., & Grandjean, D. (2016b). Neural decoding of discriminative auditory object features depends on their socio-affective valence. *Social Cognitive and Affective Neuroscience*, *11*(10), 1638–1649. https://doi.org/10.1093/scan/nsw066

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2015). Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21*(4), 500–508. https://doi.org/10.1038/mp.2015.88

Gable, P. A., Neal, L. B., & Threadgill, A. H. (2017). Regulatory behavior and frontal activity: Considering the role of revised-BIS in relative right frontal asymmetry. *Psychophysiology*, *55*(1), e12910. https://doi.org/10.1111/psyp.12910

Gadea, M., Espert, R., Salvador, A., & Martí-Bonmatí, L. (2011). The sad, the angry, and the asymmetrical brain: Dichotic Listening studies of negative affect and depression. *Brain and Cognition*, *76*(2), 294–299. https://doi.org/10.1016/j.bandc.2011.03.003

Gelman, A. (2006). Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics*, *48*(3), 432–435. https://doi.org/10.1198/004017005000000661

Gelman, A., Hwang, J., & Vehtari, A. (2013). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Aki Vehtari, & Rubin, D. B. (2014). *Bayesian data analysis* (Third). Chapman & Hall/Crc.

Geva, S. (2018). Inner Speech and Mental Imagery. In P. Langland-Hassan & A. Vicente (Eds.), *Inner Speech: New Voices* (pp. 1–32). Oxford University Press. Published to Oxford Scholarship Online: October 2018. DOI:10.1093/oso/9780198796640.003.0005.

Godfrey, H. K., & Grimshaw, G. M. (2015). Emotional language is all right: Emotional prosody reduces hemispheric asymmetry for linguistic processing. *Laterality: Asymmetries of Body, Brain and Cognition*, *21*(4–6), 568–584. https://doi.org/10.1080/1357650x.2015.1096940

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146. https://doi.org/10.1038/nn1392

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety : an enquiry into the functions of the septo-hippocampal system, second edition* (Second). Oxford University Press.

Grimshaw, G. M., Kwasny, K. M., Covell, E., & Johnson, R. A. (2003). The dynamic nature of language lateralization: effects of lexical and prosodic factors. *Neuropsychologia*, *41*(8), 1008–1019. https://doi.org/10.1016/s0028-3932(02)00315-9

Grimshaw, G. M., Séguin, J. A., & Godfrey, H. K. (2009). Once more with feeling: The effects of emotional prosody on hemispheric specialisation for linguistic processing. *Journal of Neurolinguistics*, *22*(4), 313–326. https://doi.org/10.1016/j.jneuroling.2008.10.005

Gupta, R. S., Kujawa, A., & Vago, D. R. (2019). The neural chronometry of threat-related attentional bias: Event-related potential (ERP) evidence for early and late stages of selective attentional processing. *International Journal of Psychophysiology*, *146*, 20–42. https://doi.org/10.1016/j.ijpsycho.2019.08.006

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-Related Potentials, Emotion, and Emotion Regulation: An Integrative Review. *Developmental Neuropsychology*, *35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical Correlates of Affective Prosody. *Journal of Voice*, *21*(5), 531–540. https://doi.org/10.1016/j.jvoice.2006.03.002

Heller, W., Nitschke, J. B., Etienne, M. A., & Miller, G. A. (1997). Patterns of regional brain activity differentiate types of anxiety. *Journal of Abnormal Psychology*, *106*(3), 376–385. https://doi.org/10.1037/0021-843x.106.3.376

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Ho, S. M. Y., Mak, C. W. Y., Yeung, D., Duan, W., Tang, S., Yeung, J. C., & Ching, R. (2015). Emotional Valence, Arousal, and Threat Ratings of 160 Chinese Words among Adolescents. *PLOS ONE*, *10*(7), e0132294. https://doi.org/10.1371/journal.pone.0132294

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1351–1381.

Hugdahl, K. (2011). Fifty years of dichotic listening research – Still going and going and…. *Brain and Cognition*, *76*(2), 211–213. https://doi.org/10.1016/j.bandc.2011.03.006

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/mcse.2007.55

Imbens, G. W., & Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, *25*(1), 305–327. https://doi.org/10.1214/aos/1034276631

Imbir, K. K., Spustek, T., Duda, J., Bernatowicz, G., & Żygierewicz, J. (2017). N450 and LPC Event-Related Potential Correlates of an Emotional Stroop Task with Words Differing in Valence and Emotional Origin. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00880

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Jas, M., Larson, E., Engemann, D. A., Leppäkangas, J., Taulu, S., Hämäläinen, M., & Gramfort, A. (2018). A Reproducible MEG/EEG Group Study With the MNE Software: Recommendations, Quality Assessments, and Good Practices. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00530

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kemmerer, D. L. (2015). *Cognitive neuroscience of language*. Psychology Press.

Keune, P. M., Bostanov, V., Kotchoubey, B., & Hautzinger, M. (2012). Mindfulness versus rumination and behavioral inhibition: A perspective from research on frontal brain asymmetry. *Personality and Individual Differences*, *53*(3), 323–328. https://doi.org/10.1016/j.paid.2012.03.034

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*(50), 107–118. https://doi.org/10.1016/j.brainres.2007.03.015

Kotz, S. A., & Paulmann, S. (2011). Emotion, Language, and the Brain. *Language and Linguistics Compass*, *5*(3), 108–125. https://doi.org/10.1111/j.1749-818x.2010.00267.x

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Kringelbach, M. L., & Berridge, K. C. (2009). Towards a functional neuroanatomy of pleasure and happiness. *Trends in Cognitive Sciences*, *13*(11), 479–487. https://doi.org/10.1016/j.tics.2009.08.006

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. https://doi.org/10.1037/a0029146

Kruschke, J. K. (2015). *Doing Bayesian data analysis : a tutorial with R, JAGS, and stan*. Elsevier, Academic Press, Cop.

Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280. https://doi.org/10.1177/2515245918771304

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, *4*(33), 1143. https://doi.org/10.21105/joss.01143

Lai, V. T., Willems, R. M., & Hagoort, P. (2015). Feel between the Lines: Implied Emotion in Sentence Comprehension. *Journal of Cognitive Neuroscience*, *27*(8), 1528–1541. https://doi.org/10.1162/jocn_a_00798

Lausen, A., & Hammerschmidt, K. (2020). Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanities and Social Sciences Communications*, *7*(1). https://doi.org/10.1057/s41599-020-0499-z

LeDoux, J. E., & Pine, D. S. (2016). Using Neuroscience to Help Understand Fear and Anxiety: A Two-System Framework. *American Journal of Psychiatry*, *173*(11), 1083–1093. https://doi.org/10.1176/appi.ajp.2016.16030353

Lee, C. W., Kim, S. H., Shim, M., Ryu, V., Ha, R. Y., Lee, S. J., & Cho, H.-S. (2016). P600 alteration of syntactic language processing in patients with bipolar mania: Comparison to schizophrenic patients and healthy subjects. *Journal of Affective Disorders*, *201*, 101–111. https://doi.org/10.1016/j.jad.2016.05.008

Lei, X., & Liao, K. (2017). Understanding the Influences of EEG Reference: A Large-Scale Brain Network Perspective. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00205

Leshem, R. (2018). Trait Anxiety and Attention: Cognitive Functioning as a Function of Attentional Demands. *Current Psychology*. https://doi.org/10.1007/s12144-018-9884-9

Liebenthal, E., Silbersweig, D. A., & Stern, E. (2016). The Language, Tone and Prosody of Emotions: Neural Substrates and Dynamics of Spoken-Word Emotion Perception. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00506

Liu, B., Wang, Y., & Li, X. (2018). Implicit Emotion Regulation Deficits in Trait Anxiety: An ERP Study. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00382

Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second). Mit Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, O. (2018). *Bayesian analysis with Python : introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ*. Packt Publishing.

Massar, S. A. A., Mol, N. M., Kenemans, J. L., & Baas, J. M. P. (2011). Attentional bias in high- and low-anxious individuals: Evidence for threat-induced effects on engagement and disengagement. *Cognition & Emotion*, *25*(5), 805–817. https://doi.org/10.1080/02699931.2010.515065

McElreath, R. (2020). *Statistical rethinking : a Bayesian course with examples in R and Stan* (Second). Crc Press/Boca Raton.

McEvoy, P. M., Mahoney, A. E. J., & Moulds, M. L. (2010). Are worry, rumination, and post-event processing one and the same? *Journal of Anxiety Disorders*, *24*(5), 509–519. https://doi.org/10.1016/j.janxdis.2010.03.008

McLaughlin, K. A., Borkovec, T. D., & Sibrava, N. J. (2007). The Effects of Worry and Rumination on Affect States and Cognitive Activity. *Behavior Therapy*, *38*(1), 23–38. https://doi.org/10.1016/j.beth.2006.03.003

McNaughton, N. (2011). Fear, anxiety and their disorders: Past, present and future neural theories. *Psychology & Neuroscience*, *4*(2), 173–181. https://doi.org/10.3922/j.psns.2011.2.002

McNaughton, N., & Corr, P. J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neuroscience & Biobehavioral Reviews*, *28*(3), 285–305. https://doi.org/10.1016/j.neubiorev.2004.03.005

McNaughton, N., & Gray, J. A. (2000). Anxiolytic action on the behavioural inhibition system implies multiple types of arousal contribute to anxiety. *Journal of Affective Disorders*, *61*(3), 161–176. https://doi.org/10.1016/s0165-0327(00)00344-x

McNaughton, N., Swart, C., Neo, P., Bates, V., & Glue, P. (2013). Anti-anxiety drugs reduce conflict-specific "theta"—A possible human anxiety-specific biomarker. *Journal of Affective Disorders*, *148*(1), 104–111. https://doi.org/10.1016/j.jad.2012.11.057

McNaughton, N., & Zangrossi, H. (2008). Theoretical approaches to the modeling of anxiety in animals. In R. J. Blanchard, D. C. Blanchard, G. Griebel, & D. Nutt (Eds.), *Handbook of Behavioural Neuroscience. Volume 17. Handbook of Anxiety and Fear* (pp. 3–10). Academic Press. Elsevier.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy*, *28*(6), 487–495. https://doi.org/10.1016/0005-7967(90)90135-6

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The adaptation of the Affective Norms for English Words (ANEW) for Italian. *Behavior Research Methods*, *46*(3), 887–903. https://doi.org/10.3758/s13428-013-0405-3

Moser, J. S., Hartwig, R., Moran, T. P., Jendrusina, A. A., & Kross, E. (2014). Neural markers of positive reappraisal and their associations with trait reappraisal and worry. *Journal of Abnormal Psychology*, *123*(1), 91–105. https://doi.org/10.1037/a0035817

Mothes-Lasch, M., Becker, M. P. I., Miltner, W. H. R., & Straube, T. (2016). Neural basis of processing threatening voices in a crowded auditory world. *Social Cognitive and Affective Neuroscience*, *11*(5), 821–828. https://doi.org/10.1093/scan/nsw022

Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R., Polosan, M., Spinelli, E., Koster, E. H. W., & Lœvenbruck, H. (2017). Orofacial electromyographic correlates of induced verbal rumination. *Biological Psychology*, *127*, 53–63. https://doi.org/10.1016/j.biopsycho.2017.04.013

Neal, L. B., & Gable, P. A. (2017). Regulatory control and impulsivity relate to resting frontal activity. *Social Cognitive and Affective Neuroscience*, *12*(9), 1377–1383. https://doi.org/10.1093/scan/nsx080

Neo, P. S.-H., Thurlow, J. K., & McNaughton, N. (2011). Stopping, goal-conflict, trait anxiety and frontal rhythmic power in the stop-signal task. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(4), 485–493. https://doi.org/10.3758/s13415-011-0046-x

Nitschke, J. B., Heller, W., Palmieri, P. A., & Miller, G. A. (1999). Contrasting patterns of brain activity in anxious apprehension and anxious arousal. *Psychophysiology*, *36*(5), 628–637. https://doi.org/10.1111/1469-8986.3650628

Osvaldo, M. (2018). *Bayesian Analysis with Python. Introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ Martin, Osvaldo. Bayesian Analysis with Python: Introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ* (Second). Packt.

Pascual-Marqui, R. D., Faber, P., Kinoshita, T., Kochi, K., Milz, P., Nishida, K., & Yoshimura, M. (2018). Comparing EEG/MEG neuroimaging methods based on localization error, false positive activity, and false positive connectivity. *BioArxiv*. https://doi.org/10.1101/269753

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, *87*(1), 93–98. https://doi.org/10.1016/j.biopsycho.2011.02.010

Pauligk, S., Kotz, S. A., & Kanske, P. (2019). Differential Impact of Emotion on Semantic Processing of Abstract and Concrete Words: ERP and fMRI Evidence. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-50755-3

Paulmann, S., Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, *50*(7), 1609–1620. https://doi.org/10.1016/j.neuropsychologia.2012.03.014

Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language*, *105*(1), 59–69. https://doi.org/10.1016/j.bandl.2007.11.005

Paulmann, S., Pell, M. D., & Kotz, S. A. (2009). Comparative processing of emotional prosody and semantics following basal ganglia infarcts: ERP evidence of selective impairments for disgust and fear. *Brain Research*, *1295*, 159–169. https://doi.org/10.1016/j.brainres.2009.07.102

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pell, Marc D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 275–288. https://doi.org/10.3758/cabn.3.4.275

Pell, M.D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015).
Preferential decoding of emotion from human non-linguistic vocalizations
versus speech prosody. *Biological Psychology*, *111*, 14–25.
https://doi.org/10.1016/j.biopsycho.2015.08.008

Peschard, V., Gilboa-Schechtman, E., & Philippot, P. (2016). Selective attention to
emotional prosody in social anxiety: a dichotic listening study. *Cognition and
Emotion*, *31*(8), 1749–1756. https://doi.org/10.1080/02699931.2016.1261012

Pessoa, L. (2018). Embracing integration and complexity: placing emotion within a
science of brain and behaviour. *Cognition and Emotion*, *33*(1), 55–60.
https://doi.org/10.1080/02699931.2018.1520079

Pessoa, L., Medina, L., Hof, P. R., & Desfilis, E. (2019). Neural architecture of the
vertebrate brain: implications for the interaction between emotion and
cognition. *Neuroscience & Biobehavioral Reviews*, *107*, 296–312.
https://doi.org/10.1016/j.neubiorev.2019.09.021

Pinheiro, A. P., del Re, E., Nestor, P. G., McCarley, R. W., Gonçalves, Ó. F., & Niznikiewicz,
M. (2012). Interactions between mood and the structure of semantic memory:
event-related potentials evidence. *Social Cognitive and Affective
Neuroscience*, *8*(5), 579–594. https://doi.org/10.1093/scan/nss035

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for
speech and language. *Current Opinion in Neurobiology*, *28*, 142–149.
https://doi.org/10.1016/j.conb.2014.07.005

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the
interface of neurobiology and linguistics. *Philosophical Transactions of the Royal
Society B: Biological Sciences*, *363*(1493), 1071–1086.
https://doi.org/10.1098/rstb.2007.2160

Robinson, O. J., Krimsky, M., & Grillon, C. (2013). The impact of induced anxiety on
response inhibition. *Frontiers in Human Neuroscience*, *7*.
https://doi.org/10.3389/fnhum.2013.00069

Robinson, O. J., Pike, A. C., Cornwell, B., & Grillon, C. (2019). The translational neural
circuitry of anxiety. *Journal of Neurology, Neurosurgery & Psychiatry*, *90*, 1353–
1360. https://doi.org/10.1136/jnnp-2019-321400

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55. https://doi.org/10.7717/peerj-cs.55

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *NeuroImage*, *28*(4), 848–858. https://doi.org/10.1016/j.neuroimage.2005.06.023

Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language*, *162*, 42–45. https://doi.org/10.1016/j.bandl.2016.08.001

Sassenhagen, J., & Bornkessel-Schlesewsky, I. (2015). The P600 as a correlate of ventral attention network reorientation. *Cortex*, *66*, A3–A20. https://doi.org/10.1016/j.cortex.2014.12.019

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010

Schirmer, A., & Escoffier, N. (2010). Emotional MMN: Anxiety and heart rate correlate with the ERP signature for auditory change detection. *Clinical Neurophysiology*, *121*(1), 53–59. https://doi.org/10.1016/j.clinph.2009.09.029

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schulz, C., Mothes-Lasch, M., & Straube, T. (2013). Automatic Neural Processing of Disorder-Related Stimuli in Social Anxiety Disorder: Faces and More. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00282

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*(2), 100–107. https://doi.org/10.1016/s0166-2236(02)00037-1

Sidtis, J. J., & Van Lancker Sidtis, D. (2003). A Neurobehavioral Approach to Dysprosody. *Seminars in Speech and Language*, *24*(2), 093–106. https://doi.org/10.1055/s-2003-38901

Skrandies, W. (1990). Global field power and topographic similarity. *Brain Topography*, *3*(1), 137–141. https://doi.org/10.1007/bf01128870

Spielberg, J. M., De Leon, A. A., Bredemeier, K., Heller, W., Engels, A. S., Warren, S. L., Crocker, L. D., Sutton, B. P., & Miller, G. A. (2013). Anxiety type modulates immediate versus delayed engagement of attention-related brain regions. *Brain and Behavior*, *3*(5), 532–551. https://doi.org/10.1002/brb3.157

Steber, S., König, N., Stephan, F., & Rossi, S. (2020). Uncovering electrophysiological and vascular signatures of implicit emotional prosody. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-62761-x

Steinmann, S., Meier, J., Nolte, G., Engel, A. K., Leicht, G., & Mulert, C. (2017). The Callosal Relay Model of Interhemispheric Communication: New Evidence from Effective Connectivity Analysis. *Brain Topography*, *31*(2), 218–226. https://doi.org/10.1007/s10548-017-0583-x

Styliadis, C., Ioannides, A. A., Bamidis, P. D., & Papadelis, C. (2018). Mapping the Spatiotemporal Evolution of Emotional Processing: An MEG Study Across Arousal and Valence Dimensions. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00322

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *52*(8), 997–1009. https://doi.org/10.1111/psyp.12437

Techentin, C., Voyer, D., & Klein, R. M. (2009). Between- and within-ear congruency and laterality effects in an auditory semantic/emotional prosody conflict task. *Brain and Cognition*, *70*(2), 201–208. https://doi.org/10.1016/j.bandc.2009.02.003

Tseng, H.-H., Huang, Y.-L., Chen, J.-T., Liang, K.-Y., Lin, C.-C., & Chen, S.-H. (2017). Facial and prosodic emotion recognition in social anxiety disorder. *Cognitive Neuropsychiatry*, *22*(4), 331–345. https://doi.org/10.1080/13546805.2017.1330190

Van Berkum, J. J. A., De Goede, D., Van Alphen, P. M., Mulder, E. R., & Kerstholt, J. H. (2013). How robust is the language architecture? The case of mood. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00505

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

VanRullen, R. (2011). Four Common Conceptual Fallacies in Mapping the Time Course of Recognition. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00365

Verhees, M. W. F. T., Chwilla, D. J., Tromp, J., & Vissers, C. T. W. M. (2015). Contributions of emotional state and attention to the processing of syntactic agreement errors: evidence from P600. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00388

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, *128*(3), 442–472. https://doi.org/10.1037/0033-2909.128.3.442

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2013). The Neural Representation of Abstract Words: The Role of Emotion. *Cerebral Cortex*, *24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Vuilleumier, P., & Driver, J. (2007). Modulation of visual processing by attention and emotion: windows on causal interactions between human brain regions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 837–855. https://doi.org/10.1098/rstb.2007.2092

Wabnitz, P., Martens, U., & Neuner, F. (2012). Cortical reactions to verbal abuse. *NeuroReport*, *23*(13), 774–779. https://doi.org/10.1097/wnr.0b013e328356f7a6

Wabnitz, P., Martens, U., & Neuner, F. (2015). Written threat: Electrophysiological evidence for an attention bias to affective words in social anxiety disorder. *Cognition and Emotion*, *30*(3), 516–538. https://doi.org/10.1080/02699931.2015.1019837

Wambacq, I. J., & Jerger, J. F. (2004). Processing of affective prosody and lexical-semantics in spoken utterances as differentiated by event-related potentials. *Cognitive Brain Research*, *20*(3), 427–437. https://doi.org/10.1016/j.cogbrainres.2004.03.015

Wang, Y., Xiao, R., Luo, C., & Yang, L. (2019). Attentional disengagement from negative natural sounds for high-anxious individuals. *Anxiety, Stress, & Coping*, *32*(3), 298–311. https://doi.org/10.1080/10615806.2019.1583539

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

West, S. G., Cham, H., & Liu, Y. (2014). Causal Inference and Generalization in Field Settings: Experimental and Quasi-Experimental Designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 49–80). Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511996481.007.

Westerhausen, R., & Hugdahl, K. (2008). The corpus callosum in dichotic listening studies of hemispheric asymmetry: A review of clinical and experimental evidence. *Neuroscience & Biobehavioral Reviews*, *32*(5), 1044–1054. https://doi.org/10.1016/j.neubiorev.2008.04.005

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Winkler, I., Debener, S., Muller, K.-R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. https://doi.org/10.1109/embc.2015.7319296

Witteman, J., Goerlich-Dobre, K. S., Martens, S., Aleman, A., Van Heuven, V. J., & Schiller, N. O. (2014). The nature of hemispheric specialization for prosody perception. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-014-0255-1

Wittfoth, M., Schroder, C., Schardt, D. M., Dengler, R., Heinze, H.-J., & Kotz, S. A. (2009). On Emotional Conflict: Interference Resolution of Happy and Angry Prosody Reveals Valence-Specific Effects. *Cerebral Cortex*, *20*(2), 383–392. https://doi.org/10.1093/cercor/bhp106

Xu, Y., Lee, A., Wu, W.-L., Liu, X., & Birkholz, P. (2013). Human Vocal Attractiveness as Signaled by Body Size Projection. *PLoS ONE*, *8*(4), e62397. https://doi.org/10.1371/journal.pone.0062397

Yan, M., & Sommer, W. (2018). The effects of emotional significance of foveal words on the parafoveal processing of N + 2 words in reading Chinese sentences. *Reading and Writing*, *32*(5), 1243–1256. https://doi.org/10.1007/s11145-018-9914-x

Yang, H., Laforge, G., Stojanoski, B., Nichols, E. S., McRae, K., & Köhler, S. (2019). Late positive complex in event-related potentials tracks memory signals when they are decision relevant. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-45880-y

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/s1364-6613(00)01816-7

Zatorre, R. J., & Gandour, J. T. (2007). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161

Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., Li, H., & Luo, Y. (2014). Three stages of emotional word processing: an ERP study with rapid serial visual presentation. *Social Cognitive and Affective Neuroscience*, *9*(12), 1897–1903. https://doi.org/10.1093/scan/nst188

Zhao, J., Liang, W.-K., Juan, C.-H., Wang, L., Wang, S., & Zhu, Z. (2015). Dissociated stimulus and response conflict effect in the Stroop task: Evidence from evoked brain potentials and brain oscillations. *Biological Psychology*, *104*, 130–138. https://doi.org/10.1016/j.biopsycho.2014.12.001