Investigations into the neural representation of prosodic, lexical, and syntactic properties of spontaneous, natural speech production using electrocorticography (ECoG)

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Philologischen Fakultät
der Albert-Ludwigs-Universität
Freiburg i. Br.

vorgelegt von
Olga Glanz (geb. Iljina)
aus Dünaburg (Lettland)

SS 2019

Erstgutachter: Prof. Dr. Peter Auer
Zweitgutachter: PD Dr. Tonio Ball
Vorsitzender des Promotionsausschusses der Gemeinsamen Kommission der Philologischen und der Philosophischen Fakultät: Prof. Dr. Dietmar Neutatz
Datum der Disputation: 18.08.2020

Abbreviations[1]

ALCORP: the corpus of spoken Alemannic (data courtesy of Prof. Dr. Guido Seiler, 2013)
ASM: absolute spectral magnitude
BA: Brodmann area
BOLD response: blood-oxygenation-level-dependent response
CAR: common average reference
CVR: consonant-to-vowel ratio (the number of consonants divided by the number of vowels in a spoken word)
DiSynDe: Diachrone Syntax Deutsch (Heilemann 2008)
ECoG: electrocorticography, electrocorticographic
EEG: electroencephalography, electroencephalographic
EMG: electromyography, electromyographic
emoFA: content words carrying a focus accent and stemming from "emotional" clauses
EoA: ease-of-articulation index (Ziegler and Aichert 2015); high values indicate greater ease
ESM: electrocortical stimulation mapping
FA (words): focus accent / content words carrying a focus accent
(f)MRI: (functional) magnetic resonance imaging
FRQ: lemma frequency extracted from the corpus FOLK (2012)
GAT-2: Gesprächsanalytisches Transkriptionssystem 2 (Selting et al. 2009)
IP: intonation phrase
FOLK: Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK 2012)
L1: first language
MAUS: Munich Automatic Segmentation System MAUS (Kipp et al. 1997; Schiel 2015)
NemoFA: content words carrying a focus accent and stemming from "non-emotional" clauses
N400: a negative deflection of the time-locked EEG signal with a peak around 400 ms after stimulus presentation
nFA words: content words not carrying a focus accent
NoS: number of syllables in a spoken word
PCA: principal component analysis
pc(s): principal component(s)
PoS: part or speech
ROI: region of interest
RSM: relative spectral magnitude
STTS: Stuttgart-Tübingen Tag Set conventions for the German language (Schiller, Teufel and Thielen 1995)

---

[1] See table and figure legends and Neuromedical AI Lab (2019, online) for further abbreviations and conventions.

Table of contents

Acknowledgments

First of all, I would like to express my deep gratitude to my supervisors Prof. Dr. Dr. h.c. Peter Auer and Prof. Dr. Tonio Ball. Prof. Dr. Tonio Ball has taught me a lot about gathering, processing, visualizing and interpreting electrophysiological recordings. Prof. Dr. Dr. h.c. Auer has advised me on both linguistic and organization-relevant questions and trained me in empirical and theoretical questions in the regular colloquium at the Department of German Studies throughout this work. I am grateful to my both supervisors for their valuable time and effort and for availability on both long- and sometimes also short-term notice.

I would like to cordially thank Prof. Dr. Stefan Pfänder and Dr. Christine Mertzlufft for their support and advice regarding the collection and evaluation of the linguistic data, and Prof. Dr. Andreas Schulze-Bonhage for giving me an opportunity to work with high-quality ECoG recordings, and Prof. Dr. Guido Seiler for providing me with part-of-speech frequency data from the corpus of spoken Alemannic (ALCORP). Also, I would like to express my gratitude to Prof. Dr. Evelyn Ferstl, Prof. Dr. Lars Konieczny, Prof. Ad Aertsen and JunProf. Dr. Adriana Hanulíková for occasional advice in the course of the here reported work.

Also, I would like to thank Prof. Wolfram Ziegler and Dr. Ingrid Aichert for sharing their software and expertise on estimations of word complexity with our lab, Benedikt Sauerborn for kindly providing us with his software for segmentation of the acoustic signals, and Uli Held for sharing his experiences with computation of the depth of syntactic embeddedness and for supplying me with helpful literature on statistical procedures.

I am thankful to my colleague and friend Johanna Derix, who I have learnt a lot from. I am grateful to her for sharing my interest in communication-relevant ECoG research and for working in this field hand-in-hand with me over years. It has been a great time with her, learning new, hitherto unknown things about how the human brain supports language in conditions of real-world behaviour.

Many colleagues have supported me in this research. I would like to thank all my co-workers for their contributions, which are highly appreciated. Particular thanks go to Bella Diekmann, Marina Hader, Ann-Kathrin Kießner (born Klumpp), Pia Hagen-Wiest, Lisa Dieminger, Dominic Lau, Sheila Albrich, and Judith Bartels for fruitful collaborations.

I am grateful to our technician Andreas Wintermantel (University Medical Center Freiburg) for being available, whenever any hardware and software issues need to be solved. Also, thanks to Lukas Dominique Josef Fiederer for immediate technical support at the lab.

I would further like to acknowledge the funding agencies that have kindly supported me during the here reported research. These are DFG-GRK 1624 “Frequency effects in language” (University of Freiburg), BrainLinks-BrainTools (DFG grant EXC1086, University of Freiburg) and the Baden-Württemberg Stiftung (grant NEU018).

Last but not least, I would like to thank my husband Udo and baby Ella for being there, for being patient, and for giving me an opportunity to continue this work during parental leave. And thanks to the second baby in the belly for behaving recently.

Zusammenfassung

Diese Doktorarbeit berichtet von interdisziplinären Versuchen, die bisher nur wenig erforschten neuronalen Korrelate linguistischer Verarbeitung unter nicht-experimentellen, natürlichen Bedingungen der Sprachproduktion, zu untersuchen. Zu diesem Zweck hat die Autorin der vorliegenden Arbeit, unterstützt von KollegInnen, ein multimodales Korpus (“The Freiburg/First Neurolinguistic Corpus”) erstellt, welches aus synchronisierten Audio-, Video- und Elektrocorticographie (ECoG)-Aufnahmen, wie auch aus linguistischen Annotationen auf mehreren Abstraktionsebenen besteht. Dieses Korpus hat uns einen Zugang zu mehreren linguistischen Aspekten natürlicher Sprache und den damit verbundenen neuronalen Prozessen ermöglicht. Jeder Aspekt ist in einer separaten Studie behandelt worden. Studie 1 widmet sich prosodischer Verarbeitung und konzentriert sich auf die Hirnaktivität, die mit der Produktion des Fokusakzents einhergeht, Studie 2 untersucht die neuronale Repräsentation der Wortkomplexität und Studie 3 behandelt neurolinguistische Fragen zur syntaktischen Verarbeitung während spontaner Produktion einfacher Sätze. Die verwendeten psycholinguistischen Methoden bestehen aus: (i) einem „matching“-Verfahren zur Ziehung kontrollierter Stichproben aus spontansprachlichen Daten (Studie 1), (ii) der Orthogonalisierung linguistischer Parameter mithilfe eines linearen Regressionsmodells, um dem Problem der Kollinearität zwischen einzelnen linguistischen Parametern auszuweichen und (iii) einer Hauptkomponentenanalyse zur Extraktion informativster Anteile des linguistischen Materials. Wir haben uns für neuronale Effekte interessiert, die mit dem Produktionsaufwand sprachlicher Einheiten mit linguistischer Komplexität unterschiedlichen Grades verbunden sind. Ausgehend von Erkenntnissen aus bisheriger Forschung, haben wir neuronale Effekte erwartet, die räumlich lokal stattfinden und in einem breiten Gamma-Frequenzbereich (>35 Hz) auftreten würden, weil dieser Frequenzbereich als ein zuverlässiger und funktional spezifischer Marker des Aufwands bekannt ist. Erwartungsgemäß haben wir anatomisch lokale Effekte beobachten können. In Studie 1 haben sich Signale aus postzentralen Arealen als über prosodische Eigenschaften der Wörter informativ erwiesen. Das Mengenverhältnis zwischen der Anzahl der Konsonanten und der Vokale hat sich in Studie 2 als der informationsreichste Parameter in Bezug auf Wortkomplexität erwiesen, welches in zwei anatomisch fokalen corticalen Bereichen Effekte gezeigt hat: Der frontale Bereich lag an der Grenze des prämotorischen und präfrontalen Kortex zum posterioren Teil des Broca-Areals (Brodmann-Areal 44). Der postzentrale Bereich lag direkt über dem lateralen Sulcus und umschloss Effekte am zentralen Sulcus, im parietalen Operculum und im angrenzenden inferioren parietalen Cortex. Die vorläufigen Ergebnisse der Studie 3 zeigen, dass perizentrale mundmotorische Areale für syntaxbezogene Verarbeitung relevant sein können, welche sich in der ersten Hauptkomponente, die über 30% der Varianz im linguistischen Material erklärt, niederschlägt. Während Studie 3 sowohl zeitlich als auch frequenziell ausgedehnte Effekte in hohen Gamma-Frequenzen gezeigt hat, wie wir sie erwartet haben, konnten wir in Studien

1 und 2 nur zeitlich und frequenziell enge Effekte beobachten, die wenig reproduzierbar im Hinblick auf diese beiden Charakteristiken aufgetreten sind. Dies könnte als ein Hinweis darauf gedeutet werden, dass die in Studien 1 und 2 behandelten linguistischen Phänomene in den untersuchten neuronalen Signalen nur schwach repräsentiert sind. Da das Spektrum der Gamma-Frequenzen als ein Kompositum aus Signalen von mehreren Zelltypen gilt, ist eine alternative Erklärung dieser Befunde möglich: Die zeitlich und frequenziell engen und wenig reproduzierbaren Eigenschaften könnten auf funktional spezifische Verarbeitungsmuster hindeuten, die zwischen linguistischen Parametern und zwischen Probanden variieren. Nach unseren Erkenntnissen gibt es bisher keine publizierten Arbeiten zur neuronalen Verarbeitung unter natürlicher Sprachproduktion, daher ist eine weitere Validierung dieser Beobachtungen und Spekulationen vonnöten. Zusätzlich zu den hier berichteten Erkenntnissen besteht der Beitrag dieser Arbeit zur neurolinguistischen Forschung darin, dass wir Wege gefunden haben, wie man neuronale Effekte bei natürlicher Sprachproduktion untersuchen, Kontrolle über potenziell konfundierende Parameter in solchen Daten erreichen und das Problem der Kollinearität zwischen multiplen linguistischen Aspekten des neurolinguistischen Materials beheben kann.

Abstract

The present thesis reports on interdisciplinary attempts to elucidate the hitherto unexplored neural correlates of linguistic processing in conditions of non-experimental, natural overt speech production. To this end, the author of this thesis, supported by colleagues, built up a multimodal neurolinguistic corpus ("The Freiburg/First Neurolinguistic Corpus"), composed of synchronized audio, video, electrocorticographic (ECoG) materials and linguistic annotations on different levels of linguistic abstraction. This corpus allowed us to study the neural effects related to several aspects of natural language, each of whom was treated in a separate study. Study 1 was dedicated to prosody and addressed the neural activity related to the production of the focus accent, Study 2 was dedicated to questions related to the neural representation of word complexity, and Study 3 investigated the syntactic processing accompanying natural clause production. The psycholinguistic methods we used consisted of (i) application of a matching procedure to select controlled word categories out of the natural language data (Study 1), (ii) orthogonalization of the linguistic parameters with the help of a linear regression model to overcome the problem of collinearity between correlated linguistic parameters, and (iii) the usage of a principal component analysis to extract most informative components of the linguistic material. The neuroscientific approach consisted either of group comparisons, in which neural effects underlying linguistically distinctive groups of words were compared (Study 1) or of correlation of neural activity with individual linguistic parameters (Study 2) and with principal components explaining most of the variances in the linguistic data (Study 3). We were interested in neural effects reflecting differences in the effort related to the production of speech units of different linguistic complexity. We were looking for neural effects which would be spatially focalized and which would be manifested in gamma activity (>35 Hz), since it known as a reliable and functionally specific marker of event-related effort. Based on knowledge from previous research, we were expecting neural effects to be spatially focalized, extended in time and extended over a broad range of gamma frequencies. We were, indeed, able to observe anatomically local effects. In Study 1, activity in postcentral areas proved to convey information about prosodic properties of content words. In Study 2, the proportional relation between the number of consonants and vowels, which was the most informative parameter with regard to the neural representation of word complexity, showed effects in two anatomically focal areas: the frontal one was located at the junction of the premotor cortex, the prefrontal cortex, and the posterior portion of Broca's area (Brodmann area 44). The postcentral one lay directly above the lateral sulcus and comprised effects on the ventral central sulcus, in the parietal operculum and in the adjacent inferior parietal cortex. The preliminary findings of Study 3 indicate that pericentral cortical areas implicated in mouth motor functions show correlations with syntax-relevant information, reflected in the first principal component explaining over 30% of the variances in the linguistic data. While Study 3 yielded temporo-frequentially extended effects in high gamma frequencies, as we had expected, Studies 1

and 2 showed temporally and frequentially narrow effects with little reproducibility in terms of these neural characteristics. This may indicate moderate representation of the phenomena investigated in Studies 1 and 2 in the investigated neural signals. Alternatively, since the spectrum of gamma frequencies is a composite phenomenon relying on multiple cell types, it is also conceivable that these temporo-frequentially narrow effects may point to the functionally specific activation of small local populations of neurons, whose signal properties vary between linguistic parameters and subjects. Since we are not aware of published ECoG works investigating neural effects of linguistic processing during natural speech production, further validation of these observations and speculations is required. Beyond the here summarized findings, the overall contribution of this work to the field of neurolinguistics is that we have developed ways to study the neural effects related to natural language production, to obtain control over potentially confounding parameters in such data, and to surmount the problem of collinearity between multiple linguistic features of the neurolinguistic material.

1 Introduction

1.1 Natural, uninstructed speech production as a focus of this thesis

The work described in the thesis at hand is an attempt to study the neural correlates of linguistic processing during natural, spontaneous expressive language. The author of this thesis, supported by several co-workers, developed a unique and innovative database which contains simultaneous annotations of spontaneous speech recorded on audio and video, together with synchronized electrocorticographic (ECoG) recordings from the left hemisphere of seven epilepsy patients. In addition to those data, we performed extensive annotations of the language material on different levels of linguistic abstraction (syllables, words, and simple clauses) which deal with various linguistic phenomena including phonological, grammatical, and prosodic features. To our knowledge, no other lab worldwide possesses such detailed annotations of ECoG and their associated linguistic data, and it was unclear at the start of this work, how much meaningful information could be extracted from such data. This is one of the main reasons, why we probed not one but several linguistic phenomena. The collected materials, which we have termed "The First/Freiburg Neurolinguistic Corpus" are described in the poster by Diekmann et al. (2016), in a publication by Iljina et al. (2017) and more exhaustive information on how exactly the linguistic data have been gathered and analysed are available online (Neuromedical AI Lab 2019, Diekmann 2019).

Our motivation to study natural language resides in the fact that contemporary neurolinguistic research relies on experiments which are conducted with the help of specific designs that often detach linguistic features from their communicative context and that are conducted in artificially created settings. There is a lot of converging evidence from several scientific disciplines including psychology and neuroscience that natural and experimentally evoked behaviour may differ to a great extent and that the associated neural processes can also differ. For instance, subjects may falsify their behaviour due to the awareness of being studied (Bartlett 1995), or experimental settings may fail to reflect the entire complexity of natural, real-life situations (Gibson 1950). Neuroscientific research in animals has further shown that neural activity during artificial acoustic stimuli is different from activity during complex, natural tones (Aertsen and Johannesma 1981). Such differences are also conceivable between the neural processes related to natural compared to experimentally generated speech in humans. Until we initiated this work, however, research on how the brain processes the here studied linguistic phenomena during real-world communication was missing and no evidence was available for comparisons with previous experimental work.

It was our aim to elucidate the neural correlates of natural language in conditions of non-experimental, real-life speech. We were able to access such unique conditions with the help of extraoperative ECoG recordings in consented neurological patients. The recordings were obtained around the clock from language-relevant cortical areas together with concurrent audio and video materials in the course of pre-neurosurgical diagnostics (1-2 weeks). We performed transcriptions of the patients' continuous speech during this time period, identified linguistic features of interest within them, and inspected the underlying neural activity. ECoG is an attractive method to study language due to its high resilience against movement-related artefacts (Ball et al. 2009) and combination of high spatial and very high temporal resolution (Crone et al. 1998; 2001a,b). These properties allow gaining fine spatiotemporal details of high-quality neural activity accompanying human communication.

In this thesis, we sought to investigate linguistic phenomena specifically in the context of overt speech production. The reason for this is that most neurolinguistic research available to date was conducted using perceptual rather than expressive language paradigms, and the side of speech production remained little explored (Price 2012). In our recent study (Glanz et al. 2018), we compared the temporal evolution of brain activity in the different cortical areas between speech production and perception. There, we found out that the order of activation of the different areas between speech production and perception was not just reverse, which might have been an indication that perceptual processes correspond to speech production processes running backward (Liberman, Shankweiler, and Studdert-Kennedy 1967: 454), but that different areas were activated to a different extent and to different time points. This finding may mean that linguistic processes involved in receptive speech might be different from those related to overt expression.

1.2 Generation of "The Freiburg/First Neurolinguistic Corpus"

1.2.1 A brief description of the corpus

To be able to study, how linguistic properties of speech are represented in neural activity during spontaneous, natural communication, the author of this thesis, supported by colleague neurolinguists, created a multimodal neurolinguistic corpus "The Freiburg/First Neurolinguistic Corpus" consisting of simultaneous video, audio, and neural recordings obtained during overt, spontaneous, uninstructed speech production of epilepsy patients while they were engaged in face-to-face communication with visitors or medical personnel. The data stemmed from subjects who had given their written informed consent for the usage of all recordings obtained in the course of pre-neurosurgical evaluation for retrospective research[2]. The average duration of the time period within which the data were gathered ranged from one to two weeks, so that many hours of recordings from each subject were

---

[2] The Ethics Committee of the University Medical Center Freiburg approved the protocol for subject recruitment.

available for transcription. The total number of investigated subjects was eight. One data set, however, had to be excluded from the here reported neurolinguistic analyses due to the presence of strong inter-ictal activity in the neural recordings. Additional information on the analysed data sets is presented in Tab. 1. The here reported neurolinguistics data sets have been gathered over years at the Neuromedical AI Lab by the author of the thesis at hand with the support of other lab members. The doctoral thesis by Diekmann (2019) and the master thesis by Hader (2016), for instance, report on four of these eight data sets (P1-P4). The author of the present thesis has made the major contribution to the generation of the data base, and other researchers contributed to checking it and made additional project-specific annotations[3]. This thesis reports on the findings that arise from the author's own annotations as well as from re-analyses of the neurolinguistic materials reported in master and bachelor theses by Hader (2016), Dieminger (2017) and Lau (2016), which have been conducted under her supervision. The authors of these previous works have consented to further usage and re-analysis of these data in the present thesis.

The procedure of data acquisition was as follows: we transcribed the subjects' continuous speech production using the GAT-2 linguistics conventions (Selting et al. 2009), extracted simple clauses from the spoken data, and annotated them with regard to numerous syntactic parameters such as parts of speech, depth of syntactic embeddedness in a clause, and sentence constituents. Within the borders of simple clauses, we also identified individual content words and annotated them with regard to the syllable structure, lexical frequency, and several prosodic features (presented below). We then manually identified the analysed simple clauses and the content words in the neural data and marked their starts and ends to be able to align the linguistic analyses with the ECoG recordings. Thereby, a unique, to our knowledge unprecedented corpus containing both linguistic and neural data from spontaneous, real-life conversations could be generated. Detailed information on what linguistic decisions have been made and what theoretical considerations they are based on is available in Methods and in a dedicated online document (Neuromedical AI Lab 2019).

1.2.2 Challenges met during corpus generation

Generation of a spontaneously spoken corpus is associated with a number of challenges.

First, transcription may be difficult, e.g., due to the presence of acoustic background noises, low volume of speech or dialectal features the transcriber is not familiar with. Psycholinguistic research indicates that acoustic information during speech perception can be processed more correctly, when accompanying video materials showing the facial movements of the speaker are available (McGurk and MacDonald 1976). We were in a lucky

---

[3] Diekmann (2019: 13): „Die Gestaltung dieses multimodalen Korpus ist maßgeblich von Olga Glanz entwickelt worden", English translation: "The design of multimodal corpus was predominantly developed by Olga Glanz".

position to have such recording, and they proved helpful not only to process the acoustic information but also to understand the communicative situation in the light of the objects and people present in the hospital ward. Transcription is a fundamental step in the generation of a linguistic corpus. Given its subjective nature, however, it can be prone to errors, and speech can be understood and segmented differently, depending on the transcriber. Since the author of the present thesis is a non-native speaker of German, Bella Diekmann, Pia Hagen-Wiest, among other colleagues, cross-validated the transcriptions and the data were only selected for usage in the corpus, whenever several native speakers trained in linguistics agreed on speech segmentation and prosodic annotation. This double-checking procedure ensured the highest possible quality of the transcriptions, given the available resources.

A second challenge we had to deal with relates to the rendering of the transcribed linguistic material into standard German spelling. This was needed to be able to search the words in external corpora and to apply automated instruments for (morpho-)syntactic annotation. The German language has numerous verbs with detachable verbal particles, and full verbs are often accompanied by adverbs. According to the Duden online dictionary, numerous forms exist, allowing for both separate and agglutinated spelling, such as in "leer trinken"/ "leertrinken" (Engl. "to drink empty"). In the first-case spelling, the PoS analysis would contain an additional adverb, while the spelling in the second case would result in an annotation of a single verb. Clear conventions had to be developed for such and other spelling-related problems in order to keep the annotation consistent throughout the corpus.

Third, a procedure to segment speech needs to be selected, depending on what linguistic demands the corpus is supposed to meet. The corpus Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK 2012), for instance, only allowed us to access words, their lexical frequencies, and the PoS tags. An investigation of the syntactic composition using these data was therefore impossible. An example of a corpus using larger linguistic units is DiSynDe (Diachrone Syntax Deutsch, Heilemann 2008), which has additionally annotated grammatical structures, enabling investigation of the syntactic composition of sentences. Segmentation of spontaneous speech may have consequences for qualitative and quantitative analyses that should not be neglected, such as that pragmatic and prosodic information can get lost when isolating linguistic units from the communicative situation (Auer 2010), and one needs to be aware of this limitation. An important motivation for segmenting the data into simple clauses in our corpus was nevertheless to allow studying linguistic units on several scales of abstraction, ranging from clause-level syntax, to phrases, to lexical and sublexical properties. We additionally retained prosodic information to be able to conduct research on prosody and to control for prosodic variables when studying other questions.

A fourth challenge is related to the correspondence between linguistic analyses. Since our corpus possesses multiple scales of linguistic description, measures needed to be taken to make sure that annotations on the different linguistic levels were mutually coherent. For instance, the German word “auch” can function both as an adverb and as a modal particle, depending on the speech context, and we had to ensure that annotation between PoS and sentence constituent analyses did not differ. Another example of a word group that needed to be carefully controlled for was that of copula verbs, which may sometimes be difficult to distinguish from full verbs. The presence of a predicative construction in sentence constituent analyses was thus dependent on whether a verb was annotated as a copula or a full verb in the PoS analysis. To make sure that these and other necessities for correspondence were adequately met, the author of the present thesis developed a number of Matlab-based tools which allowed for identification and semi-manual correction of potentially problematic annotations. Since our corpus, containing on average around 450 annotated simple clauses and 1000 content words per subject, is comparatively small (Tab. 5), we have been able to control for such fine-grained linguistic differences, although they are usually ignored in large, automatically annotated corpora (which, e.g., usually annotate all instances of “sein” (Engl. “to be”) as an auxiliary and all instances of “auch” (Engl. “too”) as an adverb.

Finally, the management of the corpus was challenging due to its multimodal nature. From time to time, there have been instances of linguistic units which were clearly intelligible in the audio data and which could be annotated with regard to our linguistic features of interest; whenever their annotation was associated with a risk of timing errors, e.g., if the exact onset of a clause was hard to define with high temporal precision using the dedicated software (see Methods), the clause had to be removed from the corpus and all linguistic annotations had to be updated accordingly. Also, whenever post-hoc correction of the transcribed material was necessary according to colleague neurolinguists, multiple linguistic annotations had to be corrected. To control for such modifications without producing errors, the author of this thesis wrote multiple Matlab-based scripts. These can be made available for the reader upon reasonable request.

As this section has illustrated based on some examples, creation of a multimodal neurolinguistic spoken corpus can be challenging. By eluminating some peculiarities here and providing detailed documentation of our linguistic decisions in Neuromedical AI Lab (2019, online), we hope to provide helpful clues for colleagues working on related problems.

1.3 Studied linguistic phenomena and motivation for their selection

1.3.1 Prosody

In the neurolinguistic research reported in this thesis, we set the scope of analyses to prosody (Study 1), word complexity (Study 2), and the syntactic composition of simple clauses (Study 3), which will be presented below in the respective order. Our motivation for Study 1 was that it is controversial, whether and to what extent prosodic processing is supported by the left hemisphere (Friederici 2011; Belyk and Brown 2013). Neurolinguistic research on the complexity of words is sparse, and it was our aim in Study 2 to shed more light on the neural processing underlying this phenomenon. Also, it is an open question, whether the motor cortex, in addition to low-level motor processes, supports syntax (Fogassi and Ferrari 2007), and we were interested to address it in Study 3.

In the study on prosody, we focused our attention on the phenomenon of the “focus accent” (Selting et al. 2009, further abbreviated as “FA”). Whenever we pronounce an utterance, we highlight one of its syllables stronger than the others (Pheby 1975). This happens automatically, without our usually even being aware of doing so, and this happens for certain communication purposes. The placement of the FA, or the main prosodic stress of an utterance, is not arbitrary; it can alter the meaning of what is being said. For instance, if one highlights the subject “I” or the object “cat” in the sentence “I have a cat,” the person will be what the utterance is mainly about in the former case and the animal in the latter. The FA is therefore important for everyday conversations.

1.3.2 Word complexity and lexical frequency as its plausible manifestation

Study 2 deals with the question, in which ways “linguistic complexity” can be operationalized for use in quantitative analyses. Among other phenomena, which will be described below in more detail, we were interested in the lexical frequency of words. In Study 3, dedicated to the syntactic properties of simple clauses, we also used parameters describing the clauses in terms of the frequency of their occurrence. We were interested to study the frequency of linguistic units due to the fact that linguistic frequency is known as measure of a person’s experience with language. Linguistic research indicates that the frequency with which one produces or perceives a language unit affects diachronic developments in language and modulates the strength of the mental representation of linguistic units, or the degree of their entrenchment in a person’s memory (e.g., Ellis 2002; Bybee 2010; Pfänder, Behrens, and Auer 2013). According to Bybee (2010), the latter is a likely explanation for the fact that frequent words are faster processed and pronounced more correctly than the words which occur more seldom (e.g., Oldfield and Wingfield 1965; Rice and Robinson 1975).

Frequency information plays an important role in models of lexical access. Research into the functional role of these effects suggests that word frequency is processed when word forms in the mental lexicon are accessed. Jescheniak and Levelt (1994), for instance, conducted an experiment in which they compared the production times of low- and high-frequency homophones, such as “more” and “moor” and found out that these word categories were associated with the same speed of word production. These authors concluded that it is the word form rather than the lemma that is relevant for the speed of lexical access. Based on this and related evidence, the model of speech production by Levelt, Roelofs and Meyer (1999) accounts for word frequency as an important factor at the stage of accessing word forms. Murray and Foster (2004) observed that lexical access time in a lexical decision experiment had a direct relation to the word’s rank in a frequency-ordered set and postulated the “rank hypothesis” suggesting that words in the mental lexicon are ordered in terms of the frequency of their occurrence.

Neurolinguistic studies have shown that word frequency may modulate non-invasively obtained electrophysiological signals (Rugg 1990; van Petten and Kutas 1990; Strijkers, Costa, and Thierry 2009) and that it is also reflected in inferior frontal blood-oxygenation-level-dependent (BOLD) responses recorded with functional magnetic resonance imaging (fMRI) during speech perception (Prabhakaran et al. 2006; Grande et al. 2011). In this research, the magnitude of the N400 component of the electroencephalographic (EEG) signal, e.g., proved to be larger for low- compared to high-frequency words (Rugg 1990; van Petten and Kutas 1990), and fMRI results showed that high-frequency words in lexical decision tasks were associated with increased BOLD responses in the left anterior and middle posterior temporal cortex, while low-frequency words yielded increased activity in the left inferior frontal gyrus (e.g., Prabhakaran et al. 2006). The usage of frequency-based parameters in our study, in which electrophysiological signals were recorded with the help of invasive electrophysiology from an area involving the inferior frontal cortex, was therefore plausible.

1.3.3 Syntax

Syntax has been proposed to play a particularly important role in language. Our capacity of producing and learning grammar has been singled out in generative linguistic theory as a feature which is only characteristic of the human consciousness and which distinguishes our species from the rest of the living world (Chomsky 1970). Profound interest in this aspect of language, however, was present in linguistics already before the generative turn. Particularly the period up to the mid-1920s witnessed the initial growth of structuralist approaches which allowed for systemic descriptions of syntactic rules and structures in a language, whereby the surface structure of a sentence was meticulously described (Bloomfield 1926). From this time, syntactic annotation and segmentation approaches emerge which have formed a basis for modern tagging procedures, such as those used in the present thesis and in previous modern work our sentence-structure analyses are based upon (e.g., Foth 2006).

Our interest in analysing neural effects related to syntax was motivated by both linguistic and neurobiological considerations. In usage-based linguistics, an assumption has been expressed that syntactic planning is linear and incremental (Auer 2009). One crucial argument supporting this idea, e.g., is that we are able to produce self-corrections as we speak. This would otherwise be difficult if not impossible, once a solid syntactic plan has been accomplished prior to sentence production. A neuroscientific hypothesis with regard to the timing of neural activation can be derived from this assumption: one would expect syntactic planning-related activity not only before but also during clause production. Auer (2005) further assumes that the phenomenon of "projection[4]," or the capacity of the listener and speaker to foreshadow the next linguistic units in the stream of speech, plays a role in syntactic planning. The selection of the next linguistic elements in the stream of speech is restricted by the ones which precede them. Many options are possible in the beginning of a sentence, and their number decreases as the sentence advances. A sentence can begin with nearly any part of speech (PoS). Let us say, the speaker chooses to start a sentence with an article. While a noun and an attributive adjective are possible in German immediately after it, a verb or an adverb cannot stand directly after an article without violating the grammatical rules of this language. The farther the sentence advances, the fewer options are left, the stronger the projectability of the next lexical elements, and the less effort is associated with the completion of a sentence. From these considerations, one can derive a neurobiological hypothesis that the neural activation related to syntactic planning would decline over the course of sentence production.

1.4 General neurolinguistic hypotheses

In all three studies, we were looking for neural indications of effort associated with the production of language units of varying complexity. Based on previous evidence, we assumed that more frequent linguistic units would be easier to process and that they would hence be associated with neural markers reflecting effort. A neural correlate of difficulty was necessary to be able to formulate neurophysiologically adequate hypotheses. Previous neurophysiological research has shown that high-frequency activity in the gamma range (>35 Hz) is a spatially and temporally reliable index of event-related cortical activation and that it takes place in a functionally specific manner (Crone, Korzeniewska, and Franaszczuk 2011). We therefore expected higher levels of activity in the gamma range during the production of more complex and less frequent linguistic units in language-relevant cortical areas. A characteristic feature of gamma-range speech-related cortical activation in our data is that it often presents a homogeneous response extended over multiple frequencies, typically strongest between 70 and 150 Hz, and extended over time (e.g., Derix et al. 2014; Glanz et al. 2018). We were therefore expecting neural effects related to the here studied linguistic

---

1. [4] not to be confused with "projection" in Chomsky's terms, which is an entirely different linguistic phenomenon: cf. Chomsky (2013).

phenomena which would possess these characteristics. Since lower-frequency components of the neural signal are also known to reflect event-related cortical activation (Pfurtscheller 1996), we additionally analysed ECoG activity in the lower frequencies.

1.5 Background for Study 1: prosody

1.5.1 Commonly studied manifestations of prosody

Most of the previous research into the neural correlates of prosodic processing has been conducted using speech perception experiments. Prosody in speech production has largely been studied based on experiments in subjects with lesions in articulation-related areas, and physiological neural activity underlying prosodic processing during spontaneous speech production remained unexplored. It was our aim in Study 1 to elucidate the neural correlates of natural prosody in conditions of non-experimental, real-life speech production. To this end, we performed transcriptions of the patients' continuous speech during this time period, identified prosodic features of interest within them, and inspected the underlying ECoG activity.

Two kinds of prosody – linguistic and emotional – are distinguished in the literature with regard to their conversational function. Linguistic prosody determines the semantic-pragmatic aspects of an utterance. Emotional, or affective, prosody conveys information about the speaker's emotional state or about the utterance's emotional content. In spite of these functional distinctions, both types of prosody rely on a common suprasegmental inventory. Both are realized with the help of vocal pitch, sound volume, the tempo of alternations between vocalization and silence and between accentuated and non-accentuated syllables, as well as combinations thereof (Frick 1985; Féry 1993). The neural underpinnings of these prosody types are currently debated. Early evidence from lesion-based research points to differences in lateralization: the left hemisphere has been found to play an important role in the processing of linguistic prosody, whereas the right hemisphere has been shown to contribute to the processing of emotional prosody (e.g., Brådvik et al. 1991; Van Lancker 1980). Recent hemodynamic research, however, provides evidence that either hemisphere can be activated during the processing of both emotional and linguistic prosody (Belyk and Brown 2013). We were interested to establish whether these two types of prosody show localization differences, or if they express themselves at the same cortical sites.

Context information may be essential to determine whether a unit of speech contains emotional or linguistic prosody. Segmentation of spontaneous, connected speech into linguistically meaningful multi-word sequences was thus needed to differentiate between these prosody types in our linguistic material. An intuitive approach to segment continuous speech, suited for investigation of prosody, is segmentation into "intonation phrases"

(Pierrehumbert 1980), abbreviated below as IPs, also referred to as “intonation units” (Fox 2002). IPs are described in the linguistic literature as major prosodic units with a single, coherent intonation contour. They are identified in the stream of speech based on the above-mentioned suprasegmental characteristics (Selting et al. 2009). Such units normally represent semantically and syntactically meaningful speech excerpts which are about five to eight words long. They often coincide with simple sentences and are capable of conveying individual ideas (Chafe 1994). Chafe (1994: 292) illustrates such segments as follows (modified from his Figure 5.14, square brackets are used to mark borders of IPs): [Yes.] [I can see it.] [Quite right.] [Here’s a pretty kettle of fish.]. Both linguistic and emotional prosody can be manifested within an IP through the phonetic characteristics of one most prominently accentuated syllable. With rare exceptions, as in “nein es heißt nicht baBY[5] sondern BAby” (Selting et al. 2009: 371), the syllable with the main pitch accent (capitalized) mostly corresponds to the syllable carrying a word stress. Such a syllable, referred to as “the nucleus accent” (Pierrehumbert 1980) or “the focus accent[6]” (Selting et al. 2009), is an obligatory element of an IP (Pheby 1975; Uhmann 1991) which highlights the semantically and pragmatically most relevant piece of information within it. Thus, depending on whether the FA lies on “I” or “can” in the above-mentioned IP “I can see it.” (Chafe 1994: 292), different parts of this IP will be interpreted by the listener as the most relevant: in the first case, it is the person who can see “it” that is important (“I” and not, e.g., “he”). The FA on the modal verb “can,” however, would imply that the person is indeed capable of seeing “it” (possibly in spite of the presence of visual drawbacks or against a previously formulated assumption that the person might not be able to see “it”).

1.5.2 The FA as a linguistic phenomenon

The position of the FA plays an important role in natural oral communication. A dense link between the IP constituent carrying the FA and the information structure of an utterance is well-known in linguistic literature: the position of the FA is believed to often convey new information (“rhema”) which is highlighted vocally against the already available, given information (“thema,” Batliner 1989; Prince 1981; von Heusinger 1999). An example of implementation of the FA in this function is in exclamatory sentences, which tend to be of rhematic nature (Altmann 1993). The relatively early placement of the FA in such sentences (Jacobs 1988) as well as the characteristic lengthening of the syllable carrying an FA (Scholz 1987), together with the late amplitude and F0-component peaks in the acoustic signal (Oppenrieder 1988) make it easy for listeners to differentiate exclamatory from non-exclamatory sentences (Altmann 1993). Another noteworthy pragmatic function is that a

---

[5] Syllables carrying an FA will be capitalized in examples from here on.

[6] The “focus accent” (FA) is the most prominently stressed syllable, and the “nucleus accent” corresponds to the last prominent stress within an intonation phrase (IP). The nucleus accent can therefore be a secondary stress in the presence of a primary stress preceding it. Since the position of the primary stress tends toward the right border of an utterance (Chomsky & Halle 1968: 22), however, the FA and the nucleus accent coincide in most IPs. We will stick to the notion of the FA and its respective definition henceforth in the present study.

contrastive FA can be used to mark contrast to another item in the discourse, such as in "yeah i READ the dispatch (but i don't enJOY it)" (modified from Welby, 2003: 54). Such FAs, however, are comparatively rare (Selting et al. 2009) and we shall not go into further detail for this reason.

An FA can highlight an individual lexical element within an utterance, but it can also accentuate multi-word constituents and entire utterances against the discourse's background without inevitably assigning particular pragmatic importance to single words within those units. Prosodic foci can be accordingly classified into "narrow" and "broad," depending on whether the former or the latter is the case (Selting et al. 2009). The FA-highlighted part of an utterance, termed "the focus domain," may correspond to "any constituent whose focus can be marked with a single pitch accent, such as a focused argument, a focused modifier, or a focused argument-predicate combination" (Gussenhoven 1999: 45). In the example ""I can SEE it." (modified from Chafe 1994: 292), the accentuated pragmatically unmarked[7] head of the sentence ("see") carries an FA projecting a broad focus on the entire utterance in its context ("[Yes.] [I can see it.] [Quite right.] [Here's a pretty kettle of fish.]," ibid.). A broad focus can similarly be projected on the entire verb phrase when a single pitch accent lies on the head of a noun phrase governed by the verb, as in "I read the disPATCH" in answer to the question, "How do you keep up with the news?" (modified from Welby, 2003: 55).

Theoretically, an FA can be assigned to any constituent of an utterance at the cost of altering its meaning (Selting et al. 2009). Empirical linguistic research dedicated to the placement of pitch accents, however, suggests that the position of an FA within an utterance and the breadth of the projected focus domain are not arbitrary but that they are determined by the utterance's semantic (Gussenhoven 1999) or syntactic (Uhmann 1991) structure. Gussnhoven's "sentence accent assignment rule" (1999), for instance, postulates that "semantic constituents," corresponding to arguments, predicates and modifiers, with rare exceptions, carry a pitch accent, and the placement of an FA on these elements is hence likely. A rhematic hierarchy model by Uhmann (1991: 209-215) associates the placement of an FA on different syntactic constituents with different breadths of the focus domain: based on Uhmann's experimental observations (ibid.), speakers of German tend to perceive focus domains as broad when the FA is placed on predicatives, objects, and non-agentive subjects, and its occurrence on verbs, agentive subjects, and temporal and causal adverbials is associated with a narrow focus domain. The relation of the FA to the syntactic structure of the sentence is an interesting side question which will be briefly addressed in the discussion of Study 3 dedicated to the neural correlates of syntax.

---

[7] i.e., carrying the standard intonation corresponding to the unmarked prosodic form of the sentence

1.5.3 Main study question and the neurolinguistic hypothesis

A large body of neurolinguistic research has studied the neural infrastructure involved in the processing of general suprasegmental prosodic features of speech such as vocal pitch (Chang et al. 2013; Plante, Creusere, and Sabin 2002; Kreitewolf, Friederici, and von Kriegstein 2014; Sammler et al. 2015). In comparison, only few studies, conducted with fMRI, have specifically addressed the neural correlates of the FA. They have been conducted in conditions of speech perception (Wildgruber et al. 2004; Tong et al. 2005) and during overt production of logatomes (Dogil et al. 2002). We are not aware of a neurolinguistic study addressing the neural correlates of the FA during naturalistic, real-life speech production. One aim was to close this gap by exploring differences between words carrying a focus accent (FA) and words with no focus accent (nFA) which were extracted from continuous, uninstructed language output and matched with regard to a number of linguistic parameters[8] (see Methods).

According to Selting et al. (2009), the FA can be acoustically realized by a vocal pitch movement and/or a higher speech volume and/or a longer duration of the corresponding syllable[9]. On the level of motor execution, one would thus expect a stronger involvement of the articulatory cortex in the production of words carrying an FA. On a cognitive level, the production of FA words could be associated with increased attention of the speaker compared to the production of nFA words (Hirschberg and Pierrehumbert 1986). We tested whether these assumptions would find empirical validation in our data. Previous neurophysiological research has shown that high-frequency activity in the gamma range is a spatially and temporally reliable index of event-related cortical activation which takes place in a functionally-specific manner (Crone, Korzeniewska, and Franaszczuk 2011) and that evoked gamma activity is modulated by the difficulty of the task (Senkowski and Herrmann 2002). We therefore expected higher levels of activity in the gamma range (>35 Hz) during the production of FA words than during the production of nFA words. Since lower-frequency signal components are also known to reflect event-related cortical activation (Pfurtscheller 1996), we also analysed ECoG activity in the lower frequencies.

---

[8] The composite term “linguistic parameters” will be used from here on to refer to numeric parameters describing both the transcribed and linguistically annotated speech material from a linguistic point of view (such as those involving the compositional and statistical properties of the language units) as well as mechanistic phenomena describing the investigated language epochs (such as those related to their duration, loudness, or the accompanying levels of electromyographic activity).

[9] We are aware of the fact that pitch realizations of the FA in German can differ by communication situation and dialect. For instance, Féry (1993) describes six different contours of the main prosodic stress of an utterance, depending on the semantic context, pragmatic interpretation or the style of communication. Also, the realization of the FA in Alemannic is often characterized not by a lower frequency spectrum than the standard German (Prof. Auer, personal communication on August 1, 2019). The fact that heterogeneity in the possible pitch realizations of the FA cannot be excluded does not undermine the notion that increased articulatory effort is likely associated with its acoustic highlighting against the rest of the IP.

1.5.4 Additional study question and the neurolinguistic hypothesis

Besides the main question on the neural representation of the FA, we undertook an exploratory attempt to differentiate FA words by their emotional vs. non-emotional context.

There is evidence from linguistic research that the intensity of stressed syllables is modulated by the degree of emotional excitement. Paeschke, Kienast, and Sendlmeier (1999) found, for instance, that the production of the same noun phrases with disgust, anger, fear or happiness involved higher average F_0 values than in neutral statements. Since more articulatory effort may be needed to produce emotional speech, a neurophysiological difference between these contexts appears plausible. There is a growing body of recent evidence that cortical processing of emotional prosody is mainly associated with a stronger activation of the right hemisphere (Sammler et al. 2015, see also Friederici 2011 and Belyk and Brown 2013 for related reviews). The involvement of the left hemisphere in emotional prosody, however, is debated (cf. Friederici 2011 and Belyk and Brown 2013). We were therefore interested in addressing it with the help of our corpus.

Previous work in (psycho-)linguistics indicates that an objective definition of what is “emotional” speech may be difficult, since a solid theoretical analysis of this concept is lacking (Abelin and Allwood 2000). A variety of studies have shown that listeners are often unanimous in their interpretations of the emotional content of vocal expressions (Davitz and Davitz 1959; Hayashi 1999; Abelin and Allwood 2000). Being aware of this potential difficulty, we therefore employed a measure of inter-rater agreement (Fleiss 1971) to classify the FA-carrying words depending on whether they occurred in “emotional” or neutral, “non-emotional” utterances. The former category will henceforth be referred to as “emotional” FA (emoFA) and the latter as “non-emotional” FA (NemoFA) words. We tested whether differences between these two categories could be identified in our cortical recordings, hoping to observe left-hemispheric signs of emotional prosody and expecting greater activation levels in relation to emotional compared with non-emotional speech.

1.6 Background for Study 2: word complexity

1.6.1 The notion of “linguistic complexity”

In linguistics, the notion of linguistic complexity is of great interest to researchers working in different domains including language typology, history, and contact linguistics (Sinnemäki 2008). Two views are popular with regard to the world’s languages. There are scholars who subscribe to the idea of all languages being equally complex, whereby simplicity on one level of linguistic abstraction is compensated by higher complexity on another level (the so-called “equi-complexity” of languages, Maitz 2014). This claim can be illustrated by the following quotation from Crystal (1987: 6): “All languages have a complex grammar: there may be relative simplicity in one respect (e.g., no word-endings), but there seems always to be relative complexity in another (e.g., word-position).” Some researchers have found support for this claim: Sinnemäki (2008), for instance, reported an inverse relation between head-dependent morphological marking and word order in core argument marking based on a sample of 50 languages, interpreting this finding as evidence for a trade-off between morphological and syntactic strategies in the generation of this linguistic phenomenon. Other linguists who have taken a cross-linguistic perspective on multiple languages, however, have not found comparable evidence: Kortmann and Szmrecsanyi (2009) and Maitz and Németh (2014), e.g., investigated varieties of English and German, respectively, and reported, using a number of linguistic criteria, that no inverse relation between competing linguistic strategies could be observed. For instance, the correlation between the number of analytical (free grammatical forms) and synthetic (bound grammatical forms) was not negative, as one would expect departing from the assumption of equi-complexity, but positive. This is a likely indication that languages may be differently complex and not necessarily oriented toward a general common level (Maitz 2014).

In spite of the wealth of empirical and theoretical debate, the question on the extent to which an influence of linguistic complexity on one’s consciousness is far from being understood. This is partly because an objective definition and precise quantification of complexity on multiple levels of linguistic abstraction still call for elaboration (Kusters 2008; Maitz 2014), which may be difficult due to the wealth of approaches to defining and operationalizing “linguistic complexity” in the literature as well as due to the multitude of linguistic phenomena in which this phenomenon can be manifested (Szmrecsanyi and Kortmann 2012). According to Miestamo (2006a, b), linguistic complexity can be defined in two general ways: the “absolute” and the “relative.” The absolute approach aims at objective, theory-based descriptions which account for the number of structural units and rules involved in a system. From this perspective, the more rules and parts are involved, the more complex a language or language unit is. The relative approach, on the contrary, is not theory- but usage-based. It defines linguistic complexity based on how difficult a particular language or a linguistic unit is for a language user, regardless of the structural properties of the linguistic material.

1.6.2 Motivation for our choice of parameters to investigate word complexity

There is a body of psycho- and neurolinguistic research dedicated to word complexity, which has largely been motivated by clinical interests toward improving the speakers’ verbal capacities after neurological impairments affecting speech (reviewed in Ziegler and Aichert 2015). Previous research has identified such factors as word length, the position of phonemes within a word (particularly the association of stronger impairments in phoneme production at word-initial positions), and the occurrence of consonant clusters as important predictors of articulatory errors, suggesting their contribution to increased complexity of the linguistic material (ibid.) This research, however, has largely been conducted in patients with neurological disorders, and little is known about how the linguistically unimpaired human brain processes words of different complexity. In the present study, we have addressed the modulation of neural activity by word complexity metrics in linguistically unimpaired individuals.

Since the linguistic complexity of words can have multiple facets, previous psycholinguistic research on word complexity has accounted for numerous linguistic parameters simultaneously (Ziegler and Aichert 2015). We have also used several parameters to operationalize the complexity of individual content words so as not to overlook potentially informative aspects of the linguistic material. The measures we have implemented are (i) the number of spoken syllables in a word (NoS), (ii) the consonant-to-vowel ratio (CVR), calculated by dividing the number of consonants by the number of vowels in the spoken word, (iii) the "ease-of-articulation" (EoA) index, calculated according to the model by Ziegler and Aichert (2015), and (iv) the lemma frequency of the words (FRQ). It is intuitive to assume that the more syllables a word has, the more complex it will be from the "absolute" perspective. CVR was chosen because consonants, which tend to involve more articulatory organs and gestures, can be considered more complex than vowels: Shankweiler and Harris (1966), for instance, found that the production of vowels in aphasia patients was less affected compared to the production of consonants, lending support to the idea that vowels are easier to articulate than consonants. Also, consonants are acquired later in life, which is an indication of greater difficulty to the L1-language learner (Kirchner 1998). FRQ was chosen, since frequent words are associated with fewer processing errors and faster reaction times, indicating their greater ease for the language user (Bybee 2010).

The EoA index by Ziegler and Aichert (2015) presents a complex measure encompassing several aspects of articulatory phonology. It arose from these researchers' investigations on what aspects of words are predictive of articulatory errors in patients suffering from apraxia of speech. Aichert and Ziegler (2004) and Ziegler and Aichert (2015) identified the metric pattern, the occurrence of complex constrictions, as well as the number of articulatory gestures involved in word production (i.e., discrete actions of the lips, the tongue, the velum and the glottis) as important error predictors and included them in an EoA model, which they trained and validated in an additional sample of subjects. Thus, a peculiar feature of the EoA index is that it, unlike the rest of the parameters we have analysed, accounts for multiple parameters related to linguistic complexity[10]. Note that high EoA values indicate greater ease and vice versa.

---

[10] Here, it needs to be noted that the work on the latter parameter presented here was not conducted from the scratch but that it builds upon Marina Hader's master thesis (2016), who made first attempts toward investigating the neural infrastructure supporting this phenomenon under co-supervision of the author of the present thesis. Her approach to calculating linguistic complexity in accordance with the model by Ziegler and Aichert (2015) had to be modified in the thesis at hand to accomodate Prof. Ziegler's comments we obtained in personal communication. What modifications needed to be undertaken, will be explained in Methods.

1.6.3 The neurolinguistic hypothesis in the light of previous psycho- and neurolinguistic findings

Like in Study 1, we used gamma activity as a neural index of the processing effort. Since stronger gamma activity is assumed to reflect higher effort (Senkowski and Herrmann 2002), we expected that high CVR, high NoS, low EoA and low FRQ would be associated with high magnitude values in gamma frequencies. If one correlates the aforementioned parameters with gamma activity, one would hence expect to obtain positive correlations with CVR and NoS but negative correlations with EoA and FRQ in language-relevant cortical areas.

1.6.3.1 Functional neuroanatomy

As to the linguistic level at which these effects might be expected, numerous possibilities seemed plausible. According to Levelt (1999) and Levelt and Meyers (2000), word production involves the following consecutive stages: “conceptual preparation,” whereby a lexical concept is selected, “lexical selection,” a stage at which a suitable lemma is chosen, “morpho-phonological encoding, code retrieval, syllabification,” a stage at which incremental syllabification takes place, whereby morphological and phonological forms are selected, “phonetic encoding,” at which the phonetic shape of the intended language output is integrated in its prosodic context, and the final stage is “articulation,” at which the prepared articulatory gestures are motorically executed. From a functional perspective, word complexity can be manifested on each of these levels of linguistic abstraction, since preparatory and executive processes at each level of word production in this model can be difficult to varying degrees. A number of partially overlapping fronto-temporo-parietal areas are known to contribute to these linguistically distinctive processing levels. Concept extraction has been localized to the left pars orbitalis and the middle and superior frontal gyri (Binder et al. 2009), lexical selection to the left posterior middle temporal gyrus and the midsection of the left middle temporal gyrus (de Zubicaray et al. 2006), syllabification to the left pars opercularis, Brodmann area (BA) 44 of Broca's area (Ghosh et al. 2008) and to the left premotor cortex (BA 6, Guenther et al. 2006). Phonetic encoding has been associated with activity in the bilateral superior temporal gyrus (Mesgarani et al. 2014) and articulation with activity in the bilateral sensorimotor cortex (Wise et al. 1999). We were hoping for anatomically selective occurrence of our expected correlations as an indication toward a particular level of linguistic processing.

1.6.3.2 Timing

One attractive property of ECoG is that it is capable of providing temporally resolved data on the scale of milliseconds. Timing-related evidence may be particularly helpful in the context of controversial psycholinguistic literature with regard to the question, at which stage of linguistic processing during word production lexical frequency is manifested. Psycholinguistic research disagrees on whether word frequency in speech production occurs on the lexical or post-lexical level of planning, the latter corresponding to the stage of articulation and articulatory planning and the former taking place prior to these processes (Balota and Chumbley 1985). In a psycholinguistic study, Balota and Chumbley (ibid.) used a delayed naming task, in which the subjects saw a word and pronounced it after a “go” cue that was systematically delayed from stimulus presentation over a range of temporal intervals between 0 and 1400 ms. The aim of inserting temporal delays between word presentation and production was to separate early, lexical processing stages from those reflecting articulatory preparation and articulation. These authors anticipated that, if word frequency only affected the stage of lexical access, its effects would take place only very early after stimulus presentation, and that they would disappear by the stage of more delayed articulatory preparation and articulation. Conversely, if word frequency affected post-lexical processing, its effects must persist also at later stages in spite of long temporal delays between stimulus presentation and articulation. Balota and Chumbley (ibid.) found out that the difference in the naming latency between high-and low-frequency words persisted with an increasing temporal delay of the “go” cue. Based on this, they concluded that “a large component of the word-frequency effect in the production task involves production [i.e., articulation – O. G.] instead of simple lexical access” (ibid.: 95). A later study by Jescheniak and Levelt (1994), however, observed a contrary effect, showing the disappearance of differences in the naming latency between low- and high-frequency words after particularly long delays of the “go” cue. Accordingly, these authors concluded that word frequency effects in their study were manifested only at early, lexical stages of processing. In the study at hand, we were hoping to contribute to resolving this controversy: if FRQ-related effects took take place prior to word production (and likely in articulation-related areas), this would lend support to the findings by Balota and Chumbley (1985). Late effects would provide evidence supporting the conclusions by Jescheniak and Levelt (1994). We also hoped that the timing of correlation effects would provide indications to a particular linguistic level (Indefrey and Levelt 2004), or at least give us cues whether a particular neural effect is an index of feed-forward processes related to planning and early stages of execution, or if it points toward a feedback process, which would likely occur at late stages of processing, such as immediately after word production (Hickok 2012).

Like in Study 1, we were expecting neural effects which would be spatially focalized and extended over a range of gamma frequencies. Since previous research indicates that distinctive linguistic processes switch fast during word production (Indefrey and Levelt 2004), we expected that effects related to linguistic complexity would be of relatively short duration, ca. on the scale of hundreds of milliseconds.

1.7 Background for Study 3: syntactic complexity

1.7.1 Previous neuroanatomical findings on syntax

From a functional neuroscientific perspective, syntax has been predominantly studied using paradigms grounded in speech perception, and research in conditions of speech production is comparatively rare (Caplan 2015). Some researchers assume that linguistic processes, including syntax, which occur during production, have a firm link to perception due to their common representation in the mirror-neuron system (Pulvermüller and Fadiga 2016). This notion, however, still needs empirical validation. Such validation is possible by comparing the locations of syntax-relevant activation loci between speech perception and production. To be able to draw such comparisons, studies addressing the neural activation patterns during expressive speech, such as the one presented here, are necessary. Previous perception-centred research has identified the following areas in the perisylvian region of the language-dominant hemisphere as important contributors to syntactic processing: Broca's area (namely, BA 44 and sometimes also the more frontal BA 45), the angular gyrus (BA 39), which lies mainly in the anterolateral region of parietal lobe near the superior edge of the temporal cortex, the supramarginal gyrus (BA 40), which lies rostral to the angular gyrus, and the superior temporal gyrus (BA 22), occupying the anterior and posterior two-thirds of the superior temporal cortex (e.g., Caplan, Alpert, and Waters 1998, 1999; Kang et al. 1999, summarized in Caplan 2015). Additionally, Caplan (2007) mentioned several other cortical areas as syntax-relevant. Based on previous works, this author also identified the left superior parietal and left anterior inferior temporal cortex as important for syntactic comprehension. Beyond cortical regions, the basal ganglia[11] and the cerebellum proved to contribute to the perception of syntax (Caplan 2015; Friederici and Kotz 2003). We were interested to find out in Study 3, whether the aforementioned cortical areas would show effects related to syntax in conditions of overt speech production. Regarding the areas involved in the production of syntax (Indefrey and Levelt 2004) found that BA 44 and the adjacent Rolandic operculum had shared activity between speech production and perception, but that these areas were more strongly activated during the production of syntactically meaningful linguistic units (noun phrases and simple clauses), compared to the perception of the same units. The activity in the motor cortex during speech production in this earlier study was attributed to syntactically unspecific articulatory processes. It is, however, an open question, if the motor cortex, too, is engaged not only in articulation but also in high-level processes such as syntactic planning (Fogassi and Ferrari 2007).

---

[11] Friederici and Kotz (2003) proposed that inferior freontal and anterior temporal regions contribute to early syntactic processing, and the basal ganglia engage in syntax comprehension at later stages. In the present work, we have not been able to address this difference due to the absence of recordings from the basal ganglia in our sample of subjects.

1.7.2 Motivation for our choice of parameters to investigate syntactic complexity

There is a number of possible ways to assess syntactic complexity. It has been operationalized in (psycho-)linguistic research based on measures ranging from average length of an utterance in morphemes (Klee and Fitzgerald 1985), to the distance between structurally related words in a sentence (Futrell et al. 2015), to processing difficulty experienced by human subjects (Frazier 1985). In neurolinguistic research dedicated to the phenomenon of syntactic complexity, common study paradigms rely on group comparisons between grammatical structures, such as between subject and object relative clauses, since the former are associated with greater processing difficulty than the latter ones (e.g., Braze et al. 2011). It has also been shown that left-branching hierarchical structures require more processing load than right-branching structures (e.g., Cheung and Kemper 1992), and the direction of branching has recently been used in neurolinguistics as a factor contributing to syntactic complexity (Udden et al. 2019). Continuous regressor parameters have also been implemented: Brennan et al. (2012), for instance, operationalized syntactic complexity as the number of syntactic nodes integrating the word into the phrase structure within a sentence, and Udden et al. (2019) accounted for the number of words, letters and syllables, among other factors.

In the present study, we annotated the structure of individual single clauses from spoken German conversations and expressed it in numeric values, which were then correlated with the clause-accompanying frequency-domain components of neural activity. We used several syntactic parameters to describe the linguistic data. Specifically, we investigated the sentence constituent composition, the PoS constellations within the clauses, and the maximum depth of syntactic embeddedness in a clause. The former two parameters were expressed in terms of the frequency of occurrence as a particular constellation: for instance, if a constellation “s vf pv” (subject, finite verb, predicative), as in “Die Katze ist da.” (Engl. “The cat is here.”), occurred 100 times in a subject’s subcorpus, this constellation was assigned a frequency of 100 within the given subject. The same was done with PoS constellations, e.g., if “ART NN VKFIN ADV” (article, normal noun, finite copula verb, adverb, which also describes the clause above) occurred 80 times, this specific constellation was assigned the frequency of 80 in the respective subject. The depth of embeddedness provides the number of syntactic dependency levels within a clause, resulting from a classical tree structure analysis (Foth 2006). The aforementioned parameters were chosen for the following reasons: the depth of syntactic embeddedness is an established measure of syntactic complexity in (psycho-)linguistics (reviewed in Szmrecsanyi 2004) and in neurolinguistic research (e.g., Brennan et al. 2012), and it allows comparisons of our results to previous findings obtained in experimental settings. The frequency of syntactic constellations, on the contrary, presents a novel feature for which we are not aware of any published neurolinguistic report. These parameters may thus allow novel insights into the neural representation of syntactic processing and its relation to the statistical properties of simple clauses. Our motivation to use not one but multiple parameters to describe the syntactic composition of the clauses resides in considerations presented by Grodzinsky and Friederici (2006). In their review encompassing work based on lesion studies, EEG, and magnetoencephalography, these authors proposed a "formal syntax map," which presents an attempt to associate the different subcomponents of syntactic theory with distinctive anatomical areas. In their Figure 1, Grodzinski and Friederici (ibid.) highlighted the frontal operculum (i.e., an area located more medially and inferior to Broca's area in these authors' work) and the anterior superior temporal gyrus as important for building up local phrase structures, Broca's area including both BA 44 and BA 45 as supporting the calculation of syntactic dependencies within a sentence, and the posterior superior temporal gyrus as supporting integration and "possibly involving syntactic and syntax-relevant lexical information" (ibid.: 245). This functional specialization was received controversially for methodological reasons (Caplan 2007). We nevertheless thought that it may be interesting to investigate the possibility that our different syntactic parameters – the number of hierarchy levels reflecting non-linear relations related to embeddedness and the PoS and sentence constituent analyses reflecting more linear processes involved in the production of syntax – might be associated with effects in different cortical areas.

1.7.3 Expectations with regard to ECoG signal properties

Since gamma activity is known as a robust marker of event-related effort (Senkowski and Herrmann 2002; Crone, Korzeniewska, and Franaszczuk 2011), we expected that it would decline over the course of clause production and that this activity pattern would be observable in areas implicated in syntax.

Like in the other two studies reported in this thesis, we were expecting that correlations of gamma activity with the linguistic parameters of our interest would occur in anatomically focalized regions which have been associated with syntactic processing in previous research, and that these correlations would be extended over time and over a range of high gamma frequencies. Since increased activity in the gamma range is known as a marker of increased processing effort (described above), we hypothesized that sentences with more hierarchy levels and with low-frequency constellation frequencies would be associated with greater levels of high-gamma activity. In other words, we expected a positive correlation of high gamma activity with the number of syntactic hierarchy levels and a negative correlation with the frequency of syntactic constellations. Additionally to these parameters, we also investigated a number of other, potentially confounding factors, which will be described in Methods. To the best of our knowledge, the here presented endeavour to investigate syntactic processing in conditions of non-experimental expressive speech is the first of its kind, and it paves a way for research on the neural correlates of syntax during uninstructed overt, spontaneous conversations.

2 Materials and methods

This section will first present the general methods which are relevant to several studies described in this thesis. These will be followed by study-specific methods presented further below.

2.1 Methods applied in all three studies

2.1.1 Subjects

Data from seven native speakers of German (Tab. 1) were analysed. A major criterion for subject selection was that the seizure onset zone, identified in continuous ECoG recordings during pre-neurosurgical evaluation of epilepsy, did not overlay potentially speech-relevant cortical areas. These were defined as areas implicated in speech and mouth motor functions according to the results of electrocortical stimulation mapping (ESM). All subjects gave written informed consent that all audio, video, and neural data obtained in the course of pre-neurosurgical diagnostics would be made available for research, and the Ethics Committee of the University Medical Center Freiburg approved the protocol for subject recruitment. Both ECoG and ESM data were gathered by trained medical personnel at the University Medical Center Freiburg. The locations and numbers of electrodes were defined exclusively by the subjects' clinical needs. The study was performed retrospectively upon completion of diagnostics, and it did not interfere with the pre-neurosurgical procedures. Data from subjects 1-4 were analysed in the study on prosody, data from subjects 1-5 were analysed in the study on word complexity, and data from subjects 4-7 were analysed in the study on syntax.

Table 1: Subject details. Abbreviations: s.: subject, M: male, F: female, R: right, B: bilateral, L: left, R*: right-handed converted from left, FTP: fronto-temporo-parietal, FP: fronto-polar, OP: occipito-parietal, IH: interhemispheric, IP: inferior parietal cortex, PO: parietal operculum, PF: prefrontal cortex, PM: premotor cortex, TC: temporal cortex, TB: temporo-basal cortex.

s.	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**
age during implantation	22	42	29	49	41	54	56
sex	M	M	F	F	F	M	M
handedness	R	R	R	R*	R	R	R
speech lateralization	L	L	L	B	L	L	L
location of the 8x8 electrode grid	L FTP	L FTP	L FTP	L FTP	L FTP	L FTP	L FTP
seizure onset zone	IH, PF, PM	IP, PM, TC, TB	IH, IP, PO, TC, TB	PF, PM	IH, PM	FP	OP

2.1.2 ECoG recordings

ECoG recordings were acquired with platinum/stainless-steel electrodes (4-mm diameter, 10-mm centre-to-centre inter-electrode distance) at the Epilepsy Center, University Medical Center Freiburg. An EEG system with a sampling rate of 1024 Hz was used with a high-pass filter with a cut-off frequency of 0.032 Hz and a low-pass, anti-aliasing filter at 379 Hz. The subjects were additionally monitored with around-the-clock audio and video recordings synchronized with the ECoG signal. The video recordings had a resolution of 640x480 pixels and a sampling rate of 25 Hz. All patients had 8x8 electrode grids in comparable locations of the left fronto-temporo-parietal cortex. Additional electrode strips were placed in other regions of the same hemisphere. Due to the extensive variability in the locations of strip electrodes between subjects, we restricted our analyses to grid electrodes on the lateral convexity.

2.1.3 Assignment of electrodes to anatomical areas

We applied the same hierarchical probabilistic method of anatomical assignment as in our previous studies (e.g., Derix et al. 2012; 2014; Ruescher et al. 2013). This method harvests information from individual structural images of the subjects' cortex obtained when the electrodes were implanted. T1-weighted magnetic resonance images with full-head coverage were acquired with a 1.5-T Vision magnetic resonance imaging (MRI) scanner (Siemens, Erlangen, Germany) using a magnetization-prepared rapid-acquisition gradient-echo sequence at a resolution of 1x1x1 mm. These data were converted into a Matlab-compatible format with MRIcro (Rorden and Brett 2000) and normalized to the standard Montreal Neurological Institute (MNI) brain template using the anatomy toolbox [17] implemented in SPM5 (Friston et al. 1994). We visualized the MRI data with the help of custom-made Matlab-based software and marked each electrode as well as the individual positions of the central and lateral sulci in 3D. The electrodes were subsequently localized to frontal, parietal, and temporal lobes based on their positions relative to the central and lateral sulci in the individual subjects. The entire extent of the central and lateral sulci visible on the lateral surface of the cortex was taken into account for this assignment. Whenever the anatomy toolbox provided no assignment of an electrode to the parietal or temporal cortex after the end of the lateral sulcus, an electrode was defined as either temporal or parietal based on a probabilistic atlas system (Toga et al. 2006). After this, the electrodes were assigned to the most likely anatomical area within the respective lobe based on their MNI coordinates.

2.1.4 Assignment of electrodes to functional areas

We assigned the electrodes to functional areas with the help of ESM using an INOMED NS 60 Stimulator (INOMED, Germany). 50-Hz pulse trains of 10-s durations and with alternatingpolarity square waves of 250 µs were systematically applied to non-overlapping pairs of adjacent electrodes (bipolar stimulation). Monopolar stimulation was then performed to spatially resolve the effects observed in bipolar stimulation against a functionally neutral intracranial reference electrode. The subjects were unaware of the exact timing and cortical region of stimulation until the occurrence of its motor, somatosensory, or speech-related effects. Motor effects were either positive, i.e., corresponding to movements of a particular body part or negative, i.e., reflected in the patient’s inability to move a particular body part. Somatosensory effects were manifested in proprioceptive experiences at a particular body part. Speech-related effects showed as transient impairment in expressive and/or receptive language functions identified in at least one of the six simple tasks: counting, execution of body commands, naming objects, reading, repetition of sentences, and a Token Test (Wellmer et al. 2009). Stimulation intensity was gradually increased up to 15 mA until the induction of a motor effect or up to 18 mA when testing language functions. If no effect could be observed at these highest thresholds, the electrode was annotated as functionally neutral. The observed ESM effects were documented by trained medical personnel at the University Medical Center Freiburg during the ESM procedure in real time, and the correctness of this documentation was checked by other trained members of the medical personnel. We then manually visualized the obtained ESM results using CorelDRAW (version X3) by depicting the effects which could be observed at each cortical site at the highest stimulation threshold (Fig. 1). We used ESM information to define potentially language-relevant cortical areas including those engaged in movements of mouth motor effectors such as the tongue, the lips, the cheeks or those implicated in cognitive language functions. We considered both motor effects of body parts involved in articulation and transient speech production and/or perception effects as potentially important for language. Our definition of these areas relies on the results of bipolar stimulation. The reasons for this were that monopolar stimulation covered smaller portions of the cortex and that monopolar ESM in S3 was conducted only sporadically due to medical considerations.

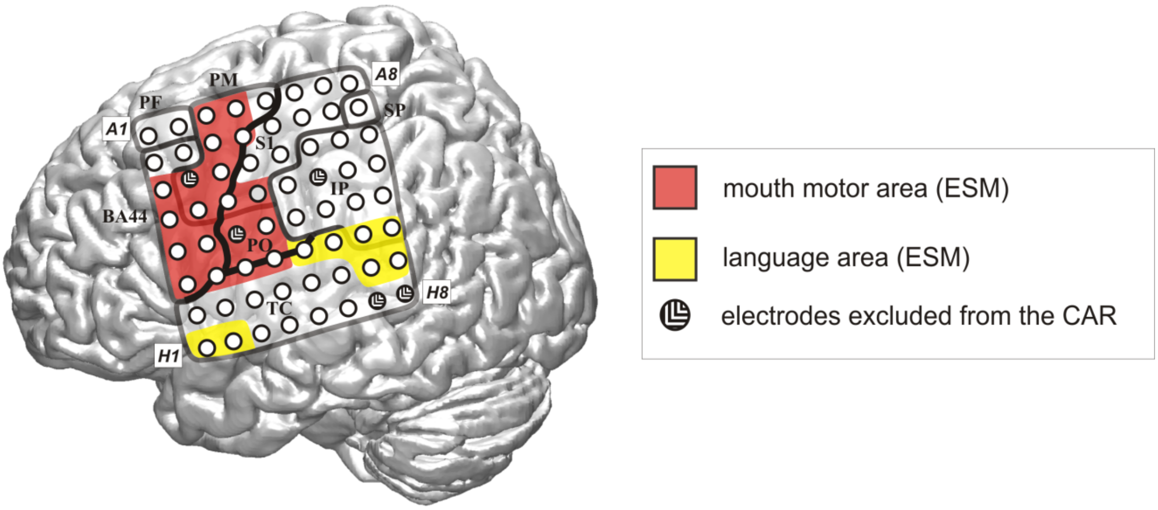


Figure 1: Typical electrocortical stimulation mapping (ESM) responses in relation to individual cortical architecture (S2). Anatomical location of the electrode grid projected onto the standard brain surface from spm5 based on the Montreal Neurological Institute (MNI) coordinates of the individual electrodes. Rows of the electrode grid are labelled by letters A to H, and the columns are labelled by numbers 1 to 8. A1 to H8: labels of the individual electrodes of the grid, included for ease of spatial reference. Conventions for structural anatomy: BA 44: Brodmann area 44, IP: inferior parietal cortex, PO: parietal operculum, TC: temporal cortex, PF: prefrontal cortex, PM: premotor cortex, S1: primary somatosensory cortex, SP: superior parietal cortex. Grey solid lines indicate the borders of these anatomical areas. Black solid lines indicate individual positions of the central and lateral sulci in this subject. Colour coding of the overlays highlights potentially speech-relevant functional areas identified with the help of ESM (see legend and Methods). ESM results for potentially speech-relevant areas are coded by red (articulation relevant) and yellow (cognitive) language functions.

2.1.5 Pre-selection of the spoken material

Upon visual screening of the entire around-the-clock data, we pre-selected recordings in which the subjects were awake, alert, and participating in face-to-face conversations with visitors, hospital ward neighbours, or medical personnel. The numbers of hours of recordings we selected per subject depended on the overall duration of their hospital stay, on how much they spoke on average, and on the intelligibility of the language material. Epochs of acoustically and mechanically undistorted speech production were selected. To avoid erroneous transcriptions, we discarded recordings which took place when the subjects were speaking quietly, in the presence of loud acoustic background noise as well as during long periods of overlapping talk. Speech production in the presence of strong acoustic background noise and when the subjects were eating or drinking were also not analysed.

2.1.6 Continuous transcription of the spoken material

Trained linguists performed transcriptions of the patients' speech using the freeware PRAAT (Boersma 2001) in accordance with the "basic" transcript of the GAT-2 conventions (Selting et al. 2009). This transcription method allows segmenting continuous speech in individual IPs, which are characterized by its cohesive intonation contour and meaning. These conventions allow annotation of both what and how is being said. The latter includes the placement and prosodic prominence of the focus accent (the whole syllable is annotated in capital letters) and of secondary accents, pause durations, suprasegmental (e.g., the IP-final falling intonation is annotated by a semicolon) and also some paralinguistic properties. Examples of IPs are provided in Example 1. This method of transcription allows documenting colloquial, idiolectal, and dialectal properties of language such as elisions and other deviations from the standard German. It is therefore well suited to provide realistic descriptions of natural speech without losing information about such properties. The high precision of annotation, however, comes at a cost of time. In our experience, the average transcription rate was approximately one min. of continuous speech per 45 min. for an expert and around one hour for transcribers with less experience. All transcriptions were checked by at least two other linguists, who cross-validated the borders of the IPs and the linguistic annotations within them. If differences between the transcriber and the cross-validators and/or among the cross-validators with regard to the content and/or the position of the primary accent of the IP were identified, the IP was excluded from neurolinguistic analyses in favour of unambiguous language material. Little spoken material had to be discarded at this step of data acquisition owing to the pre-selection of high-quality language material at a previous stage of data acquisition.

German transcription	**German standard spelling**	**English translation**
das HEIßT-	Das heißt,	That is,
=sie haben jetzt praktisch (.) gestern so ne art LANDkarte [(.) beim gehirn aufgeschrieben;]	sie haben jetzt praktisch gestern so eine Art Landkarte beim Gehirn aufgeschrieben.	they actually wrote down some kind of a brain map yesterday.
aber jetzt können sie dann noch nicht SAgen-	Aber jetzt können sie dann noch nicht sagen:	But they still cannot tell yet,
=ach da verMUten wir was-	Ach, da vermuten wir was.	Oh, we suspect something here.
=da ISCH jetz:-	Da ist jetzt	This is now
DER nerv oder die funktion;	der Nerv oder die Funktion.	this nerve or that function.

Example 1: A typical transcription excerpt. Syllables in capital letters highlight the focus accent. “-” indicates no change in IP-final intonation, “=” indicates a smooth, immediate acoustic transition between adjacent IPs, “(.)” annotates pauses <200 ms, square brackets indicate speech excerpts in which the speaker overlapped with at least one of his/her conversation partners, and “;” is an index of IP-final falling intonation.

2.1.7 Extraction and selection of simple clauses

The next step was extraction of simple clauses which lay within the borders of the acquired IPs. This procedure was part of the acquisition of the Freiburg/First Neurolinguistic Corpus, which we built up over the last years (Diekmann et al. 2016). A simple clause was defined as the main verb (the full verb, the copula verb, or a transitive modal verb in the absence of a full verb (Eisenberg 2004) plus all its arguments. The borders of the simple clauses were set in accordance with the topological field model of German utterances (Musan 2009). It defines discrete syntactic fields which occur in the following order: “Vor-Vorfeld” (VVF) containing lexical elements which have a loose syntactic relation to the main verb and which do not form arguments of the simple clause, “Vorfeld” (VF) containing lexical elements which form sentence constituents such as subjects, objects, or adverbial modifiers and their integral parts, “Linke Klammer” corresponding to the finite verb of the main clause or to a subordinating conjunction of the subordinate clause, “Mittelfeld” containing lexical elements after the finite verb and until the end of the clause or until “Rechte Klammer” (RK) whenever it is present; RK corresponding to the non-finite verb; “Nachfeld” (NF) containing lexical elements which form sentence constituents such as post-positioned objects or adverbial modifiers which can otherwise also occur in VF; “Nach-Nachfeld” containing lexical elements which have a loose syntactic relation to the main verb, do not form arguments of the simple clause, and can otherwise also occur in VVF. All lexical elements between the start of VF and the end of NF were considered as falling within the borders of a simple clause (Musan 2009). Only clauses which were free of hesitation markers, self-corrections, stuttering, and consecutive repetitions of the same word were selected. All clauses which entered our selection contained a finite verb. Since psycholinguistic research has shown that humans perceive pauses ≥200 ms as actual “pauses” in a conversation (Walker and Trimboli 1982), we selected clauses with such pauses for the sake of homogeneity of our linguistic materials. The simple clauses obtained as described above form basic multi-word units of our neurolinguistic corpus which is currently being used by several projects at our lab which are dedicated to the neural architecture supporting linguistic functions.

2.1.8 Rendering the clauses in standard German

Next, the simple clauses from the IPs were rendered in standard German in accordance with the Duden online dictionary. We maintained the proximity to the spoken German as long as a word form was registered in Duden. Some German adverbs, for instance, “heut”/ “heute” (Engl. “today”) have parallel registered variants. Whenever our subjects said “heut” or “heute,” we differentiated between them. Phonological variants for endings of finite verbs in the first-person singular are also admissible in modern grammars, e.g., on Netzwerb (online). We therefore retained the presence or absence of “e” at the end of verbs in this form, e.g., “ich sag”/ “ich sage” (Engl. “I am saying”). German has verbs written together with a detachable verbal particle corresponding to a separate adverbial modifier in English. There are no universal grammatical rules as to which non-finite verbs have to be written together with the counterparts of the English adverb and which not, and parallel forms are often registered in Duden. For instance, while „danebenstehen” (Engl. “to stand next to”) is written together as one word, “da sein” (Engl. “to be here”) is written separately, and two parallel forms are registered in Duden (as of 2017) for “leer trinken”/ “leertrinken” (Engl. “to drink empty”). We always kept to Duden whenever it offered one variant, and we opted for a composite form whenever two variants were listed as admissible. These phonological peculiarities have obvious consequences for linguistic analyses, and they had to be accounted for during the extraction of lemma frequencies (described below).

2.1.9 Selection of content words

For analyses conducted on the level of single words, the words were extracted. Each word in each simple clause was subject to a PoS analysis, conducted according to the Stuttgart-Tübingen-Tag-Set (STTS) conventions for the German language (Schiller, Teufel and Thielen 1995). We modified the tag set in order to differentiate between adverbs and homophone particles, which can only be distinguished from each other in German with the help of the surrounding semantic and phonological context. The PoS annotation was conducted by hand, since automated taggers are not sensitive to such context information. Unlike STTS, we also differentiated between homophone full verbs, copula verbs, and auxiliaries. The correspondence of the tags set in the data to our tag inventory was checked with the help of a custom-made MATLAB-based program to prevent errors due to manual annotation. This program also checked if the number of words in a clause was the same as the number of PoS tags for the respective clause. It also identified impossible or statistically unlikely combinations of linguistic tags in the same clause (e.g., a copula verb and a full verb, two or more non-finite verbs, etc.) and displayed them to the investigator for validation. There is evidence that neural activity can differ between content and function words (Münte et al. 2001; Diaz and McCarthy 2009). We only used content words to have homogeneous samples in the present study. All words of the PoS categories “full verb” (FV), “normal noun” (NN), and “adverb” (ADV) were extracted from the clauses and ordered chronologically within the respective PoS category. The overarching ADV category consisted of normal adverbs (ADV in STTS), adverbial adjectives (ADJD), interrogative or relative adverbs[12] (PWAV), and pronominal adverbs (PAV). The number of occurrences of the same word in the subject’s subcorpus was calculated and used to select unique words per subject. To avoid dominance of repeated words, we discarded all repetitions of the same word in the word list analysed within each subject and only used the word in its first occurrence. For reasons described below, the EoA index (Ziegler and Aichert 2015) can only be calculated for words whose

---

[12] We were able to distinguish interrogative or relative adverbs from the similar PoS category of interrogative or relative adverbial pronouns based on their syntactic role in the clauses. While the former had an adverbial dependence on the full verb, the latter played the syntactic role of objects.

phonological length does not exceed three syllables. We limited the selection of words in the study on word complexity accordingly to meet this criterion.

2.1.10 Lemmatization and extraction of lemma frequencies

Word frequency information was used in the study on word complexity and it was also accounted for as a potentially confounding parameter in the study on prosody. The frequency information was obtained in the following way. We generated the lemma forms of all selected words with the help of the lemmatization tool of the online platform WEBLICHT (Hinrichs, Hinrichs, and Zastrow 2010), designed for automated generation of corpora. We extracted lemma frequency information from the Forschungs- und Lehrkorpus Gesprochenes Deutsch or FOLK (Engl. “Research and Teaching Corpus of Spoken German”), developed by the Institut für Deutsche Sprache (Engl. “Institute for the German Language” (FOLK 2012). It consisted of 45,104 lemmata at the moment of data acquisition. Their lexical frequency was determined by analysing 1,308,786 single words (Dr. Thomas Schmidt, personal communication in 2015). An attractive characteristic of this corpus is that it provides PoS annotations of the word forms to each lemma in accordance with the STTS conventions. The lemma frequencies of the words in the patients’ data were determined by searching for the lemma and the respective frequency in FOLK using a custom-made MATLAB-based program. The lemma frequency for the verbs was determined by accounting for the verb with a detachable verbal particle whenever the verb and its detachable verbal particle were produced together (e.g., in a non-finite verb form “wahrgenommen” (Engl. “perceived”). When the verb and the detachable verbal particle were produced separately (e.g., “Sie *nimmt* das nicht *wahr*.” (Engl. “She is not perceiving it.”), we used the frequency of the verb without a prefix. The reason for this is that verbal particles in such cases obtained an own PoS tag according to the STTS conventions, just as all other words in our corpus did. Also, they are listed in FOLK with own lexical frequencies. The procedure of how word frequencies were extracted was the same as described in the master thesis by Marina Hader (2016), and word frequencies for most subjects were available to the author of this thesis at hand owing to this previous work. Hader (2016) and Dieminger (2017), who have been working on the neural correlated of lexical processing under the supervision of the author of the thesis at hand, annotated the neurolinguistics data describing the individual words together with the author of this thesis. The author of this thesis developed the custom software for data analysis, conducted the statistical tests and visualized the presented quantitative data.

When the lemma from the patient data set was not found in FOLK, its frequency was first assigned to 0, and each lemma with this frequency was screened for correspondence between the lemmatization approaches of WEBLICHT and FOLK. Whenever necessary, corrections of the lemma form and of the associated lemma frequency were undertaken manually. The PoS category ADV accounted with two extra cases: for some adverbs, two parallel lemma forms were listed in FOLK, e.g., “gern” and “gerne” (Engl. “gladly”), each of them with its own frequency. We treated them as variants of one lemma which was assigned the sum of both frequencies. Another peculiarity with regard to some adverbs was that some of our lemmata had no exact matches in the lexical inventory of FOLK due to the fact that FOLK did not allow for the absence of the final “e” in adverbs such as “heut,” which is registered in Duden. Whenever a lemma frequency for an adverb could not be found, our custom-made MATLAB-based software for extraction of lemma frequencies searched for a variant with an “e” at the end of the word and selected the respective frequency.

We took the PoS information into account to avoid erroneous frequency assignments from homographs of a different PoS category. For instance, the lemma “sein” (Engl. “to be”) in FOLK was listed in finite and non-finite auxiliary verb forms but also as an attributive possessive pronoun (Engl. “his”), and we used the sum of the frequencies of the verb forms only, whenever “sein” was used as a verb.

2.1.11 Annotation of the speech-accompanying ECoG data

The next step of data acquisition was identification of the pre-selected simple clauses and content words in the neural data with the help of the software Coherence EEG/PSG System (Deltamed: Paris, France), which was also used to record the data. This software allows for simultaneous visualization of raw ECoG potentials at each of the recorded electrodes together with synchronized video and audio materials. Since automated tagging of the neurolinguistic data using Coherence was not possible, we opted for manual tagging of the speech-accompanying neural data. The tagging of each word was performed using PoS-specific markers for word starts and ends based on aural inspection of the acoustic information in the video recordings frame by frame. We also tagged the starts and ends of the embedding clause boundaries (“cs” and “ce” for the start and end, respectively). Additionally, we tagged the starts and ends of the embedding speech production epoch (“ss” and “se”) in which the words occurred to be able to account for the position of the words and clauses within these language units. Clauses were tagged whenever they lay within the borders of a single IP and whenever the FA of the IP lay within clause boundaries. A “speech production epoch” was defined by the absence of pauses ≥ 200 ms within it for the reasons mentioned above. The tagging procedure is illustrated in Fig. 2.

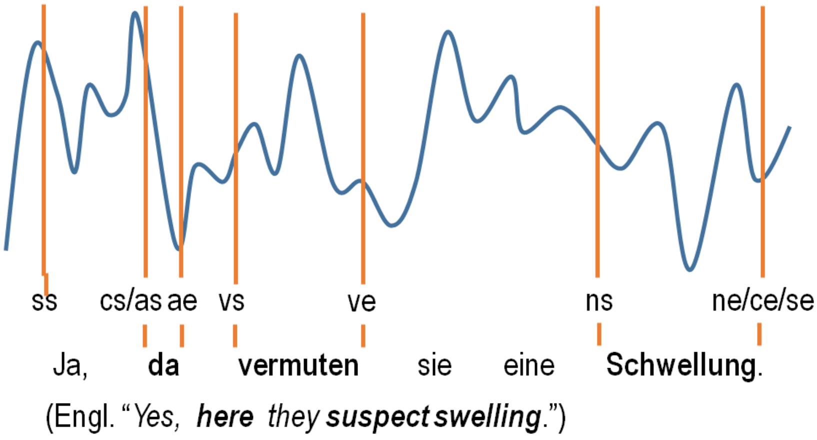


Figure 2: A schematic representation of how the tagging of the selected words in the ECoG data was carried out using the software Coherence. The used tags denote the beginning (ss) and end (se) of the speech production unit and the beginning and end of the selected words (as/ae for adverbs, ns/ne for nouns and vs/ve for verbs).

The annotators were instructed to make sure that the tags for clause boundaries were set as precisely as possible, whenever multiple tags had to be set simultaneously, and validation of each tag by at least two other people was performed to control their temporal precision. Due to the manual nature of our annotations, however, the word start tag could, e.g., slightly precede the tag for speech/clause start tag within the same speech production epoch, and the word/clause end tags could follow the corresponding speech end tag by several ms. Automated post hoc correction using custom-made Matlab-based software was applied to account for such imprecisions. This correction was performed to align word and speech starts to clause starts as well as word and speech ends to clause ends whenever the respective tag combinations occurred within the same time window for automated correction. The time window for automated correction was systematically varied in steps of 5 ms between 5 and 50 ms to the left and to the right from the tag. The software identified missing tags of the respective pair (i.e., word start/end, clause start/end, speech start/end) within the time window of interest. It also produced warnings, whenever impossible or unlikely durations which would likely need tag correction could be identified (e.g., negative or excessively long words, clauses or speech production epochs). To select an optimal time correction threshold for each subject, the outputs of automated correction at each threshold (i.e., the time window for automated correction) were checked by validating the plausibility of the resulting overlaps between the tags based on the content of the transcriptions. It was possible to tell from this comparison, whether a word start/end and a speech start/end in the respective clause were supposed to overlap or not. For instance, if an adverb start had the same timing as "cs" after the automated time correction, we checked if that adverb indeed corresponded to the clause-initial word in the transcription. The best correction thresholds ranged between 10 and 40 ms, depending on the subject. In rare cases, the best thresholds failed to account for 1-2 individual tags per subject. The timing of these tags was then manually adjusted upon automated correction at the selected best threshold.

2.1.12 Spectral analysis of the ECoG data

ECoG data were re-referenced to a common average reference (CAR). All electrodes with artefacts caused by impaired connections to the amplifier and those lying in the seizure onset zone, identified in continuous extraoperative recordings, were excluded from the CAR and removed from subsequent analyses (grey-coloured circles in Fig. 1). The total duration of the analysed time window in Studies 1 and 2 was 2 s before and 3 s after word onset. The analysed frequency range in Studies 1 and 2 was up to 300 Hz. In Study 3, we selected a longer time window of 2 s before and 4 s after clause onset to accommodate the average duration of the clauses, and the analysed frequency range was up to 500 Hz[13]. The data cut out relative to word (Studies 1 and 2) or clause (Study 3) start within the selected time-frequency window will further be referred to as “trials”. We computed word-onset-related, time-resolved spectral magnitude changes for all words/clauses of each subject using a Fast Fourier transformation. 200-ms sliding windows and 20-ms time steps were used for calculation of the absolute spectral magnitudes (ASM). This yielded a frequency resolution of ca. 5 Hz and a time resolution of ca. 200 ms. (We also probed alternative settings with 500-ms sliding windows and 50-ms time steps. They proved consistently worse in statistical analyses, which is likely due to temporal resolution inferior to that resulting from the former settings. For this reason, we will abstain from reporting on them in the following.) We applied a multitaper method (Percival 2000) with 5 Slepian tapers to diminish the cutting artefacts at the edges of the analysed temporal sequences.

2.1.13 Baseline correction

The ASM in each trial was baseline-corrected. In the study on prosody, the baseline was generated by averaging the frequency-resolved ASM in each trial of both FA and nFA conditions over the first 500 ms of the analysed time window within each electrode of each individual subject and by averaging it over trials. The ASM values of each time-frequency bin in each trial were divided by a common baseline activity for both conditions. The same procedure was used for the emoFA/NemoFA contrast, except that the ASM data for this trial selection were extracted from the already available ASM data of the overarching FA category. In the study on word properties, the baseline was generated by averaging the frequency-resolved ASM in the respective trial over the first 200 ms of the analysed time window, corresponding to [-2 -1.8] s relative to the start of each word. The thereby obtained relative spectral magnitudes (RSM) for each word, ordered chronologically within the respective PoS in the given subject, were then concatenated for all three PoS categories. In

---

[13] The selection of the upper time and frequency limit depended on both the upper time and frequency limit within which clear neural effects were visible in the spectra and on the statistical robustness of the neural effects in correction for multiple comparisons: Since the number of time-frequency bins was one of the sources for multiple comparisons, we sought to reduce it in a meaningful way so as not to make the statistics more conservative than necessary.

the study on syntax, the baseline was extracted from each trial over the first 500 ms of the analysed time window within each electrode of each subject and by averaging it over trials. Next, the ASM values for each trial were divided, in a time- and frequency-resolved manner by the resulting baseline. The RSM in all three studies were transformed to a natural logarithmic scale using the log function implemented in MATLAB, trial-averaged and visualized using custom-made MATLAB-based software.

2.1.14 Statistical analyses of the ECoG data

Once the RSM values for each individual condition of Study 1 were calculated, we first assessed the statistical significance of the RSM responses obtained in the respective condition. To this end, we used a two-tailed Wilcoxon sign test implemented in the signtest function in Matlab. We compared the RSM values in all time-frequency bins for all trials at each electrode against a value of "1". Then, we corrected the obtained p-values for multiple comparisons for the number of time bins, frequency bins, and the number of tested grid electrodes ($q<0.05$).

To compare RSM effects between conditions in Study 1 (i.e., FA vs. nFA and emoFA vs. NemoFA), we employed three different variants of time window selection ("long1," "long2," and "short1") in six frequency ranges (0-10, 0-30, 30-45, 60-80, 80-100, and 100-150 Hz). This was done to be able to identify effects which might be specific to a particular time window duration or frequency range. The duration of the time window "long1" corresponded to the mean of the median word duration in all subjects, "long2" was the median word duration in the individual subjects, and "short1" corresponded to 100-ms long time windows within the analysed time period. These variants of time window duration were applied relative to the start of the word within the entire analysed time period of [-2, 2] median word durations, defined as "long1" and "long2". That is, each variant of the long time window was used in the time period from word start to word end, in the time period from minus one word duration to word start, in the time period from word end to plus one median word duration, and in the respective time periods within minus or plus two word durations. The time window selection "short1" was applied in the aforementioned time period of [-2, 2] median word durations of "long1," yielding numerous short time windows within this period. For each of the resulting four time windows "long1," four time windows "long2," and 19 time windows "short1" and for each analysed frequency range, time-frequency-averaged RSM values in the respective condition were calculated, resulting in a single RSM value per time-frequency window at each electrode for each trial.

These RSM values in Study 1 were compared between the word groups of interest using the aforementioned Wilcoxon rank sum test. Although this test does not assume a normal distribution, it requires that the data in the two groups to be compared are equally distributed, while allowing a possible location shift. We conducted a two-tailed Kolmogorov-

Smirnov test, implemented in the kstest2 in Matlab, to test whether the RSM values of each time-frequency window at each tested electrode met this criterion in our contrasts. To account for a possible location shift between the distributions, we subtracted the respective median in each condition prior to the implementation of the Kolmogorov-Smirnov test. We accounted for the results of the Kolmogorov-Smirnov test afterwards (Tab. 4), and no data were rejected prior to application of the Wilcoxon rank sum test. The p-values resulting from the Wilcoxon rank sum test for each time-frequency window underwent Bonferroni correction for multiple comparisons across the number of tested grid electrodes using a number of statistical thresholds ($q<0.05$, 0.01, 0.005, 0.001, etc.). Whenever no significant effect remained upon Bonferroni correction, false discovery rate correction (FDR) was used at the same statistical threshold. If no significant effect was observed for a given time-frequency window, we performed uncorrected testing at a threshold of $<q/50$. Consecutive automated application of these different statistical procedures and thresholds was used to not overlook the possibly weak differences between conditions, which we anticipated upon preliminary visual comparison of the RSM changes between our word groups of interest in the study on prosody.

In addition to conventional statistics, we applied single-trial decoding to test whether neural activity in the FA/nFA and in the emoFA/NemoFA conditions was informative as to what condition of the pair a given excerpt of neural data came from. As in some of our previous studies (e.g., Pistohl et al. 2012; Derix et al. 2012), we applied a regularized linear discriminant analysis. The features of the neural signal used for decoding were RSM values averaged at each individual time point over the respective frequency band. We used the same frequency bands as in the conventional statistics described above. The obtained RSM values from the entire analysed 5-s time interval of the RSM calculation were used together, without averaging over time. The normalized decoding accuracy was obtained by first calculating the decoding accuracy for each condition separately, and then taking the mean decoding accuracy of these conditions. Bonferroni correction of the p-values for multiple comparisons in each frequency band was performed for the number of analysed electrodes in the respective subject.

In the studies on word complexity and on syntax, we also compared the RSM values in all time-frequency bins for all trials at each electrode against a value of “1”. Then, the obtained p values were corrected for multiple comparisons for the number of time bins, frequency bins, and the number of tested grid electrodes at (Wilcoxon sign test, FDR-corrected at $q<0.05$). The small 2x3 grids in the upper right corners of the individual electrode panels in Fig. 10 (Study 2) code for the occurrence of significant effects in six time-frequency intervals. The vertical three columns of this grid indicate the occurrence of significant effects before, during, or after word production. The horizontal two rows code for the frequency range of the effect: the upper row refers to effects in gamma frequencies from 35 Hz and above, and the lower row codes for effects in the lower frequencies. The stars (increases in the spectral magnitude) and the circles (decreases in the spectral magnitude) code for the direction of the effects relative to the baseline period.

2.1.15 Acquisition of linguistic parameters dependent on word-, clause- and speech-epoch duration

The electrode grids in all subjects covered extensive portions of the motor cortex (e.g., Fig. 1), which contributes to low-level processes related to the execution of motor actions (Penfield and Boldrey 1937). All subjects also had electrodes in the superior temporal region, known to engage in domain-general acoustic processing (Steinschneider, Nourski, and Fishman 2013). To disentangle the neural effects of the linguistic processes of our interest from those related to such linguistically-unspecific phenomena, we gathered a set of parameters relevant to the description of words/clauses with regard to their duration and position in the stream of speech. In Study 1, for instance, the acoustic volume of the words, and the electromyographic (EMG) data recorded over the course of pre-neurosurgical diagnostics were additionally analysed.

We collected three duration-related parameters: the duration of the word in ms (ws_we), the duration between word start and speech start in ms (ss_ws), and the duration between word end and speech end in ms (we_se). These were obtained by calculating the temporal distances from ws to the closest ss, and from we to the closest se after the completion of the aforementioned procedures for tag validation and semi-automated correction. The same was done for the starts and ends of the simple clauses, whereby cs_ce, ss_cs, and ce_se were calculated, respectively.

To evaluate the potential interference of EMG activity with our findings, we accounted for the available EMG recordings in the studies on the level of single words. This was not done in the study on syntax due to considerations related to the tight time schedule within which all those time-consuming studies have been conducted. The analysed EMG recordings consisted of one EMG electrode on the left cheek in S3 and S4, two electrodes from the bilateral quadriceps plus two electrodes from the bilateral deltoid muscles in S5, and two bilateral EMG electrodes in the upper chest region of S1. The electrodes in the latter subject lay in the approximate area of the muscle trapezius, which is involved in arm movements. The EMG levels S1 thus likely reflect myographic activity of the upper extremities ipsilateral to the positions of these electrodes. No EMG data were recorded in S2. We analysed the EMG data in the frequency range of 60-200 Hz, since EMG activity in these frequencies was most pronounced. The RSMs of the EMG responses were calculated in the same way as for the neural data and averaged over 60-200 Hz and over the duration of the respective word. The obtained RSM values were trial-averaged and rendered on a natural logarithmic scale.

Another linguistically-unspecific parameter which we sought to control for was the intensity (volume) of the acoustic signals underlying word production. We obtained this information from the .wav data recorded concurrent with the ECoG signals while the patients were speaking. This was done by automated identification of the transcribed words in the acoustic signals and by subsequent manual correction of the resulting annotations. The transcribed linguistic material was pre-processed using custom-made Python-based software (courtesy of Benedikt Sauerborn) to ensure the compatibility of our data with the subsequently applied software to automatically align the transcribed words to the acoustic signal. Word-for-word segmentation of the acoustic signal was then performed using the Munich Automatic Segmentation System MAUS (Kipp et al. 1997; Schiel 2015). The temporal precision of this alignment was checked and improved when necessary by manual inspection of the segmented data using PRAAT. Average (mean) values over the duration of each word for the intensity of the acoustic signals in Db were obtained using this software's designated object infrastructure.

The here described parameters and some other linguistic parameters which are study-specific (described below) were used together with the linguistic parameters to identify the collinearity structure in our linguistic data and accounted for in word-level correlation analyses of linguistic parameters with the neural activity.

2.2 Additional steps of data analysis in Study 1: prosody

2.2.1 Matching words for the FA/nFA contrast

Experimentally unrestricted, spontaneously spoken data have the advantage of granting access to ecologically valid linguistic material. A challenge in dealing with such data, however, is the presence of a priori uncontrollable linguistic parameters which may be different between FA and nFA or between emoFA and NemoFA words. If one does not handle such possibly confounding factors in the acquired data, they may elicit statistical differences between word groups which would have little to do with our research interests. To reduce the likelihood of this happening, we employed a statistical matching procedure using the MS DOS-based freeware Match (Van Casteren and Davis 2007). This program uses two identically-structured data sets which do not necessarily have the same number of entries (in our case words) to generate two new data sets with the same number of entries which will be as similar as possible with regard to a number of pre-specified control parameters. Match probes all possible combinations of entries and updates its solution whenever it is better than the previous one. The program can be stopped manually at any time, and the result of matching can be saved at that time point only. The duration of its operation can be determined by the investigator depending on the schedule of the study, the amount of language material, the number of parameters to be matched, and/or the available computational resources.

The control parameters we used to match the words for each of the two contrasts were the duration of the words in ms, the duration between word start and speech start in ms, the duration between word end and speech end in ms, lemma frequency, the number of phonemes in the spoken word, the position of the word stress in syllables relative to the word start, and the number of repetitions of the same lexeme within the subcorpus of the individual subject. Since content words of different PoS can show differences in neural processing (Luzzatti and Chierchia 2002), we matched FA/nFA and emoFA/NemoFA words for each PoS separately, so as to have equal numbers of matched words from each PoS in our categories of interest. We started by matching the FA and nFA words for each PoS at the maximum number of trials, which in all subjects corresponded to the number of trials of the respective PoS in the FA condition (Appendix 1: Suppl. Tab. 1). We took the same approach when matching emoFA/NemoFA words (Appendix 1: Suppl. Tab. 2). The matching procedure does not automatically produce word groups which are balanced with regard to the desired control parameters, it only elicits the best match at a given time point of the program's operation. A top-up statistical analysis was therefore necessary to determine whether the word groups for the respective PoS (FV, NN, ADV) showed differences between conditions (FA vs. nFA and emoFA vs. NemoFA) with regard to any of the control parameters we were trying to match. We evaluated the success of matching for every individual control parameter using the Wilcoxon rank sum test ($p<0.05$), implemented in the ranksum function of Matlab 2015b. This two-tailed, non-parametric test evaluates the hypothesis that two independent samples come from the same distribution with equal medians. Whenever any of the control parameters showed a significant difference between the two word categories, we reduced the desired number of trials in the matching output by ten in the FA/nFA contrast. Since the amount of available data to match the emoFA and NemoFA words was small, we reduced the number of trials between matching iterations by one. We repeated this procedure until no significant differences were observed with regard to any of the control parameters. In the next step, we merged the three PoS groups, for which the matching procedure had been carried out.

2.2.2 Rating words for the emoFA/NemoFA contrast

We implemented a rating procedure in which three independent native speakers of German were instructed to define whether the word carrying an FA stemmed from a clause that was produced using emotional prosody ("1") or non-emotional prosody ("0"). The emotional content was rated as could be identified based on the semantic, contextual, and phonological properties available to the raters from the continuous transcriptions. The raters were instructed to assign the clauses to the category "emotional" whenever the speaker was, in their opinion, feeling emotional and/or talking about emotionally-relevant topics and to the category "non-emotional" otherwise. All three raters were familiar with the acoustic data and they used transcriptions with available context information as the basis for their judgment. We evaluated the inter-rater agreement based on Fleiss' Kappa (Fleiss 1971), interpreted its quality on a scale by Landis and Koch (1977), and selected words from "emotional" and "non-emotional" clauses, further referred to as the emoFA and NemoFA categories, based on the quality of the inter-rater agreement. Example 2 illustrates both realizations of the focus accent in their embedding IPs at a 100% inter-rater agreement. Since the FA corresponds to the semantic-pragmatic core of an utterance (Selting et al. 2009), the transfer of clause-level emotionality ratings to the level of the individual words carrying an FA appears plausible.

	German transcription	**German standard spelling**	**English translation**
emoFA	ich seh sicher **SCHLIMM** aus;	Ich sehe sicher **schlimm** aus.	I am sure looking **awful**.
	also **ICH** könnt das net bezahlen;	Also, **ich** könnte das nicht bezahlen.	Well, **I** couldn't pay it.
	DAS is mein traum;	**Das** ist mein Traum.	**This** is my dream.
NemoFA	um **FÜNF** uhr ruf ich die nochmal an;	Um **fünf** Uhr rufe ich die nochmals an.	I'll call them again at **five**.
	dAt **WEIß** ich nit;	Das **weiß** ich nicht.	I don't **know**.
	da war de **BRUNnen** da:;	Da war der **Brunnen** da.	There was a **fountain** there.

Example 2: Examples of "emotional" (emoFA) and "non-emotional" (NemoFA) words with a focus accent in the context of their embedding intonation phrases. The respective words in columns 2-4 are in bold font, all other conventions as in Example 1.

In our experience, clauses rated as "emotional" often had a more personal character and they often contained first-person pronouns. In such clauses, the patients, e.g., talked about their appearance, wishes, or other topics related to personal situations and experiences. A topic which often co-occurred with positive emotionality ratings was the disease, the upcoming surgery and its possible outcomes. Sentences rated as "neutral" conveyed less personal information. They, e.g., contained facts about daily routines, such as the time of the day when certain events happened, statements about the location and state of objects within and outside the hospital ward, or corresponded to matrix clauses with relatively little semantic and pragmatic content. Further details on the matching and rating procedures can be found in a related work by Lau (2016), who had contributed to rating and matching the here analyzed data under the supervision of the author of the thesis at hand.

2.2.3 Region-of-interest analysis

Since one of our hypotheses was that stronger RSM responses might occur at articulation-relevant contacts, we took a closer look at the cortical region involved in articulation. Definition of an "articulatory area" based on ESM alone may be inconclusive, since one cannot assume that all electrode contacts implicated in speech and movements of articulatory organs are involved in articulation. Conversely, significant RSM responses during word production may also include electrodes involved in non-articulatory processes, such as monitoring of own language output (Hickok 2012). We therefore identified this articulation-relevant region of interest (ROI) based on both sources of information: an electrode belonging to it had to lie within the area identified by ESM as responsible for speech and/or movements of articulatory organs, and it had to show a significant RSM response during word production in both FA and nFA conditions. We identified such a region based on monopolar stimulation in S1, S2, S4 and based on bipolar stimulation in S3, since monopolar stimulation in the latter subject was conducted only sparsely. For each of the FA and nFA conditions, we then median-averaged the RSM responses at each time and frequency bin over trials at the respective electrode, mean-averaged the obtained values over all electrodes within the ROI, and mean-averaged these values over subjects. The obtained results were visualized together with examples of individual electrodes from each subject. We also performed the same analysis for all electrodes with significant differences between our conditions of interest regardless of their location. Due to the lack of similarity of these effects with regard to their time and frequency between subjects, however, no clear response patterns could be observed in this latter analysis, and its results will not be presented or discussed for this reason.

2.3 Additional steps of data acquisition and analysis in Study 2: word complexity

2.3.1 Calculation of NoS and CVR

Using the phonological information gathered by the author of this thesis together with co-workers who participated in the annotation of our corpus, we generated two additional parameters which are relevant to description of articulatory complexity. We calculated NoS and CVR (i.e., the proportional relation between the number of consonants relative to the number of vowels in a word). The latter was calculated by dividing the number of consonants by the number of vowels in a spoken word. The procedure of how these parameters were extracted was the same as described in the Hader (2016), and NoS and CVR values for most subjects were available to the author of the thesis at hand owing to this previous work[14].

2.3.2 Articulatory complexity index as in Ziegler and Aichert (2015)

We estimated the articulatory complexity of the words using a mathematical model by Ziegler and Aichert (2015). This is a tree-structure model which describes the hierarchical embedding of vocal-tract gestures in single words. It was developed to predict the accuracy of word articulation in patients with apraxia of speech. An earlier master thesis from our lab by Hader (2016) was a pilot study dedicated to this phenomenon. In comparison with this earlier work, however, the calculation formula was slightly modified to accommodate Prof.

---

[14] A crucial methodological difference of the present study from the approach by Hader (2016) is that, unlike in this former work, we correlated the neural data not only with the (updated and extended) raw linguistic parameters but also with the individual mutually-orthogonalized parameters after residualization.

Ziegler’s comments[15]. More detailed information as to how exactly the parameter EoA was generated is available in Appendix 2. Based on the annotations of the consonant cluster structure, syllabic composition, and the prosodic properties of the words, we calculated EoA as described in Appendix 2 with the help of a custom-made MATLAB-based program. What the reader may need to memorize is that a high EoA index reflects greater ease of articulation and a low EoA index is a sign of high articulatory complexity.

2.3.3 Correlations between the linguistic parameters

Each of the linguistic parameters (i.e., FRQ, EoA, CVR, NoS, ss_ws, ws_we, we_se, EMG levels, and the intensity of the acoustic signal) in every subject was concatenated for all PoS of the selected words as for the underlying neural data. Custom-made MATLAB-based software was used to make sure that the arrangement of the linguistic and the neural data was exactly the same. Next, these parameters were correlated with each other. All linguistic parameters were rendered on a natural logarithmic scale, before which 0.001 was added to each value to avoid rendering all zeros to minus infinity (see Solari (1969) for related methodological considerations). We used Pearson's correlation implemented in the built-in MATLAB function corr.m and tested the obtained r values for significance at $p<0.05$. The results of this analysis were used to assess the reproducibility of the correlation structure between parameters in spontaneously spoken language, and they were taken into account when interpreting the results of correlation-based neural analyses.

2.3.4 Correlations of RSM with the linguistic parameters

To identify the neural effects related to the gathered linguistic parameters, we correlated the individual parameters with the neural activity. Since the parameters we were investigating present data distributions which do not fall into discrete natural categories, we chose this procedure over group comparisons (Baayen 2008). Prior to correlation, the RSM values and all linguistic parameters were rendered on a natural logarithmic scale in the same way as done for the linguistic parameters. We will further refer to this procedure as

---

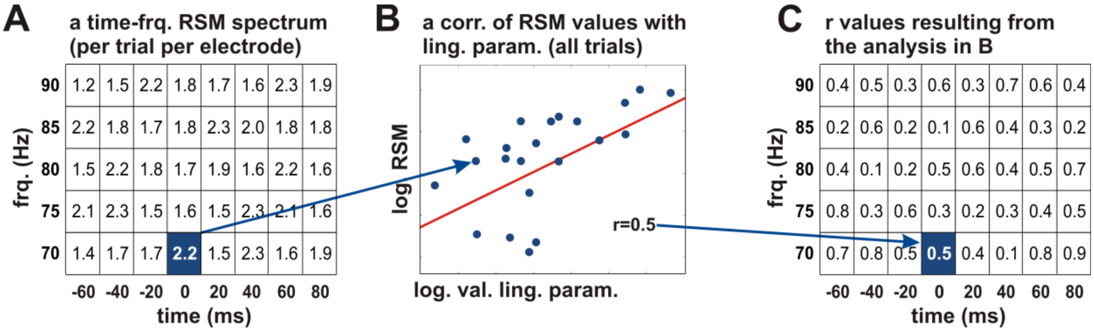
[15] The formula on p. 22 of Marina Hader’s work was changed in agreement with Prof. Ziegler’s feedback from:

$$p_{/St/} = \underbrace{\underbrace{p \times c_{cnstr} \times c_{clust}}_{\text{tongue tip gesture}} \times \underbrace{(p \times c_{glot} \times c_{clust})^{0.5}}_{\text{glottal aperture gesture}}}_{/S/} \times \underbrace{\underbrace{(p \times c_{glot} \times c_{clust})^{0.5}}_{\text{glottal aperture gesture}} \times \underbrace{p \times c_{clust}}_{\text{tongue tip gesture}}}_{/t/}$$

to:

$$p_{/St/} = \underbrace{\underbrace{(p_1 \times c_{cnstr} \times c_{clust})}_{\text{tongue-tip gesture}} \times \underbrace{(p_2 \times c_{glot} \times c_{clust} \times 0.5)}_{\text{glottal aperture}}}_{/S/} \times \underbrace{\underbrace{(p_3 \times c_{glot} \times c_{clust} \times 0.5)}_{\text{glottal aperture}} \times \underbrace{(p_4 \times c_{clust})}_{\text{tongue-tip gesture}}}_{/t/}$$

neurocorrelation (Fig. 3 explains the principle of this analysis). Its results were visualized by projecting the obtained r values on the original time-frequency space of each electrode and coluor-coded for the strength and prefix of the correlation (cf. RSM effects in Fig. 6C and correlation effects in Fig. 8 for S5).



frq. (Hz)	-60	-40	-20	0	20	40	60	80
90	1.2	1.5	2.2	1.8	1.7	1.6	2.3	1.9
85	2.2	1.8	1.7	1.8	2.3	2.0	1.8	1.8
80	1.5	2.2	1.8	1.7	1.9	1.6	2.2	1.6
75	2.1	2.3	1.5	1.6	1.5	2.3	2.1	1.6
70	1.4	1.7	1.7	2.2	1.5	2.3	1.6	1.9
	time (ms)							

frq. (Hz)	-60	-40	-20	0	20	40	60	80
90	0.4	0.5	0.3	0.6	0.3	0.7	0.6	0.4
85	0.2	0.6	0.2	0.1	0.6	0.4	0.3	0.2
80	0.4	0.1	0.2	0.5	0.6	0.4	0.5	0.7
75	0.8	0.3	0.6	0.3	0.2	0.3	0.4	0.5
70	0.7	0.8	0.5	0.5	0.4	0.1	0.8	0.9
	time (ms)							

Figure 3: Correlation of the neural activity with the acquired linguistic parameters. (**A**) The logarithmic relative spectral magnitude (RSM) values at an individual electrode are depicted schematically in every time-frequency bin of the spectrum (the squares of the table) for a frequency range of 70-90 Hz (y axis) and -60 to 80 ms relative to word start (x axis). A Pearson's correlation of the RSM values with a linguistic parameter of interest was calculated across all words of the given subject (**B**). The correlation coefficients were projected on the time-frequency structure of the original spectrum (**C**). This procedure allowed us to identify the time periods and frequencies in which the linguistic parameters showed correlations with the neural activity. The r values were colour-coded and tested for significance (described below). Abbreviations: frq.: frequency, log.: natural logarithm, RSM: relative spectral magnitude, val.: values, ling. param.: linguistic parameters.

We first conducted the neurocorrelation analysis for the frequencies up to 300 Hz over the entire time window of 2 s before and 3 s after word onset. Correlations of RSM values with the linguistic parameters of our interest showed effects which were more local in time and frequency, compared with the duration-related parameters. They did not show clear effect patterns in very high gamma frequencies over 150 Hz and at very early and late time points of the aforementioned time window. To decrease the number of multiple comparisons for subsequent statistical analyses and to focalize the neurocorrelations in the time and frequency space, we repeated the same analysis for all parameters in frequencies up to 150 Hz, which is a commonly used upper limit for time-frequency analyses of speech-related gamma activity in ECoG studies (e.g., Derix et al. 2014; Sinai et al. 2005; Towle et al. 2008). We also shortened the analysed time window to the period from 500 ms before ss to 500 ms after se.

Statistical testing for significance was conducted using Bonferroni correction for multiple comparisons over all time-frequency bins and electrodes at q<0.05. Bonferroni-corrected effects will be reported whenever they could be observed. The results of uncorrected testing at a particular threshold will be reported otherwise. Note that, for the sake of statistical robustness of the results, we considered the uncorrected effects as "significant" if the p values did not exceed 5E-06. All effects with higher p values were treated as not significant. Different parameters showed differently focal effects with regard to their spatial extent when tested at the same significance threshold. The following statistical procedure was hence applied in order to capture the most salient effects related to the given parameter. We tested the neurocorrelation effects for each linguistic parameter using a scale of significance thresholds ranging from 5E-06 (the least conservative) in steps of 5E-06 until no effects related to the given parameter could be observed. The last threshold yielding significant effects for the given parameter will be reported.

We accounted for the occurrence of the possible mutually confounding effects, i.e., the effects of several parameters taking place when testing the neurocorrelations for significance at the same electrode, in the same time range (a maximum temporal distance between effects was 40 ms) and in the same frequency range (a maximum difference in the frequency of the effect was 20 Hz within gamma frequencies over 35 Hz or within the same frequency band in the alpha and beta signal components). We defined “the same” temporo-frequential components as intervals rather than as single time-frequency bins so as not to overlook the overlaps between parameters which would have been evident when testing the parameters using less conservative significance thresholds.

Effects taking place at the same electrode and in the same temporo-frequential components of the neural signal indeed took place in some cases (left column in Fig. 12), and the analysed linguistic parameters were often significantly correlated. Some neural effects describing a particular parameter may therefore have come into being due to such mutual correlations. We were, however, interested in finding out, to what extent parameter-specific effects could be identified. Therefore, we conducted an additional analysis to erase the mutually-orthogonal components in the linguistic data. Using a linear regression lmer function implemented in the lme4 package (Bates et al. 2014) for R, we predicted each linguistic parameter by all parameters with which it was significantly correlated. By doing so, we were able to extract the residuals of the models, which were orthogonal to these parameters. We predicted each parameter of interest only by the parameters with which it was significantly correlated and not by all other parameters in order not to erase too much variance in the linguistic data in the process of residualization[16]. Using all parameters as predictors could have resulted in losing potentially meaningful information. In a next step, we correlated the residuals with the neural activity using the correlation procedure described above.

2.4 Additional steps of data acquisition and analysis in Study 3: syntax

The study on syntax involved generation of several parameters which described the grammatical composition of the simple clauses. We conducted a number of linguistic analyses, which are described in more detail online (Neuromedical AI Lab 2019), dedicated to the linguistic conventions of our corpus. The parameters analysed in the present study

---

[16] Additional details on the residualization procedure can be found in a related work by Dieminger (2017), except for the methodological difference that the original prefix of the residuals in our study was maintained in contrast to this earlier work.

were restricted to several parameters for reasons of time and effort. We focused on three syntactic parameters: “hierarchy levels,” “sentence constituent frequency,” and “PoS frequency”.

2.4.1 Syntactic dependency analysis and extraction of the number of linguistic hierarchy levels in a clause

The parameter “hierarchy levels” was generated by manual annotation of the clauses using the conventions by Foth (2006), which we slightly modified for increased precision (Neuromedical AI Lab 2019, online). Our own innovation is the rendering of this analysis in a written form which can be easily read and interpreted with the help of automated software. Using this approach, one can annotate, for each word, how many words it (directly and indirectly) depends on and how far in the sentence they are located (Example 3).

„Am Mittwoch **hatte** ich die schlimmsten Schmerzen meines Lebens.“

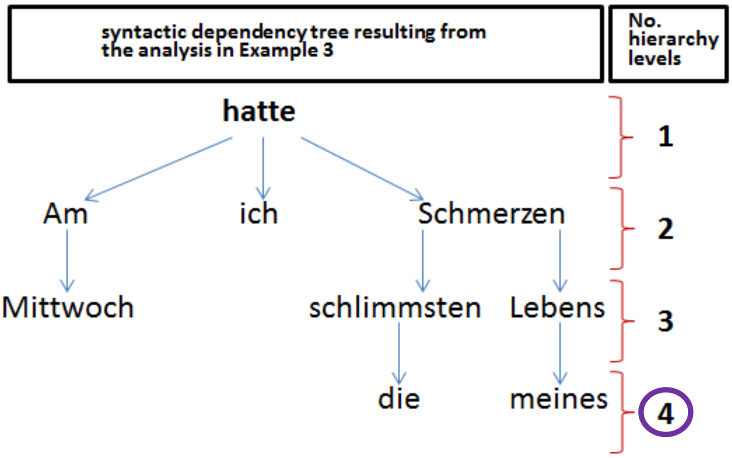
(Engl. “On Wednesday, I had the worst pain in my life.”)

+1pn 0 **-2pp/+1subj/+4obja** 0 0 0 -2det/-1attr/+2objg 0 -1det

Example 3: An example of a simple clause and its syntactic dependency analysis. The full verb “hatte” (Engl. “had”), whose analysis is indicated in bold font, is used to explain the principle of annotation. It has three syntactic constituents which depend on it in this sentence. These are divided by “/” in the annotation. “-2pp” means that one of them is a prepositional phrase whose head (the preposition) corresponds to the second word to the left (“-2”) from the verb (position 0 in each calculation); “+1subj” annotates the subject, which depends on the verb and lies one word to the right of it, and “+4obja” annotates an accusative object whose head (“Schmerzen,” Engl. “pain”) lies four words to the right from “hatte”. Note that “objg” in this annotation system denotes a non-prepositional genitive dependency and not necessarily a genitive object. The same holds for nominal phrases with other casus (“objd” is used for non-prepositional dative and “obja” is used for non-prepositional accusative dependencies).

Once these analyses were completed, they underwent a number of semi-automated checking procedures to make sure that no spelling errors, no omissions, and no multiple dependencies of the same word on several lexical constituents were annotated, as well as that the numeric values of annotation were descending to the left and ascending to the right relative to the position of the analysed word in the sentence (position 0). Obviously, the number of syntactic constituent annotations for each word, separated from the annotations for other words by a space, had to be the same as the number of words in the sentence. We made sure these criteria were fulfilled with the help of exhaustive Matlab scripts, which the author of the present thesis developed for this specific purpose.

After we made sure that the analysis was complete and correct, we rendered this information using an automated Matlab program written by the author of this thesis which rendered this information to calculate, how many levels of linguistic dependency were present in the sentence. This approach can be graphically illustrated as follows (Example 4):

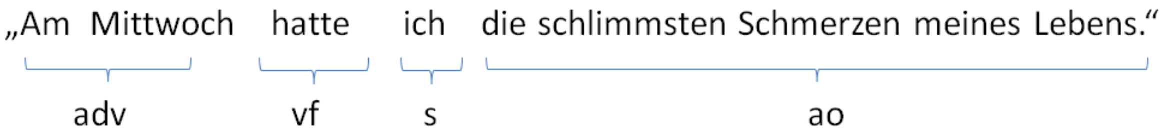


Example 4: Rendering of the analysis in Example 3 into a tree-structure representation and calculation of the number of syntactic hierarchy levels. This step of analysis was conducted using dedicated software and the precision of its outcome was verified by hand. The outcome number of hierarchy levels in the given sentence is highlighted by a purple circle.

Thus, the final products of this analysis were individual numbers (one for each sentence). These were used in consecutive linguistic and neural analyses described below. After all syntactic annotations were completed, we wanted to know, how representative our data were of the spoken German language annotated in other sources. We thus compared the number of the different PoS with those in the corpus of spoken Alemannic (ALCORP), which consists of spoken interviews with the native speakers of Alemannic (2013, courtesy of Prof. Dr. Guido Seiler).

2.4.2 Sentence constituent analysis

For each clause, we also performed a classical sentence constituent analysis as in Eisenberg (2004) and Musan (2009). This analysis involved annotation of the verbal complex which made up the head of the clause and syntactic elements queriable from the head (Example 5).



Example 5: A syntactic constituent analysis of the sentence from Examples 3-4. The corresponding sentence constituent annotations are below the words and phrases highlighted by blue brackets. “adv” is an adverbial modifier, “vf” is the finite verb, which in this example is also the head of the sentence, “s” is the subject, and “ao” is the accusative object.

Once this analysis was completed, it was semi-automatically compared with the syntactic dependency analysis for consistency (One example is that both analyses allow annotation of copula constructions, and annotations of copula constructions had to be present for this clause in both analyses.). Once the checking was complete, we calculated the frequency with which a particular constellation of sentence constituents (in Example 5, it is “adv vf s ao”) occurred in the subcorpus of the individual subject.

2.4.3 PoS frequency

As described above, each word of each sentence was annotated for the PoS according to the STTS conventions (Example 6). After this, this analysis was semi-automatically compared with the other two analyses to ensure that the number of syntactic elements which had to be the same between the analyses was the same (e.g., the number of non-finite verbs between PoS and sentence constituent analyses, which both coded for this information).

„Am	Mittwoch	hatte	ich	die	schlimmsten	Schmerzen	meines	Lebens.“
APPRART	NN	VVFIN	PPER	ART	ADJA	NN	PPOSAT	NN

Example 6: A PoS analysis of the sentence from Examples 3-5. The corresponding PoS annotations are below the words. "APPRART": preposition with a fused article, "NN": normal noun, "VVFIN": finite full verb, "PPER": personal pronoun, "ART": article, "ADJD": attributive adjective, "PPOSAT": possessive pronominal determiner.

Once the checking was complete, the same procedure as with the sentence constituent analyses was applied: the frequency of a particular PoS constellation (in the case of the sentence in Example 6, "APPRART NN VVFIN PPER ART ADJA NN PPOSAT NN," was calculated in the subcorpus of the individual subject.

As additional parameters which may potentially influence/interfere with syntax-relevant processes in our recordings, we accounted for the length of the clause in words, the temporal duration of the clause and the durations from its left and right borders to the start and end of the speech production epoch, respectively, as well as the position of the focus accent within the clause.

2.4.4 PCA

We conducted a principal component analysis to find out which linguistic parameters played an important role in the linguistic data. This was done with a dedicated script in the programming environment R. For each subject, we evaluated the loadings of the individual parameters within the first three major principal components, which together explained over 70% of the variances in our data. Next, we correlated the loadings of the first three principal components with the neural activity of the subjects using the same time-frequency-resolved neurocorrelation approach as in Study 2. The same analysis was done with the individual parameters and the results of correlations with the individual parameters and with the PCA loadings were compared.

3 Results

This section will present the results of the here conduced analyses for the three studies. The results of the study on prosody will be presented, followed by the results of the study on word complexity, followed by the results of the study on syntax.

3.1 Results of Study 1: prosody

3.1.1 Speech data

The data used for the analyses came from recordings which were rich in conversations. The number of content words extracted from these data differed between subjects, and it was not necessarily proportional to the number of hours per subject. Among other factors, this is due to the different degrees of the subjects' involvement in conversations alongside the varying frequency and duration of confounding phenomena which precluded us from analysing particular excerpts of the language material (e.g., high levels of acoustic background noise, speech-accompanying mastication, and overlapping talk). The number of FA words was about half of the number of nFA words in all subjects. This is because a clause could contain only one FA but several content words. Prior to the matching procedure, the three PoS were distributed differently between the FA and nFA conditions. Verbs in the FA condition formed the largest group, nouns were the second largest group in three out of four subjects, and adverbs were the smallest group for three out of four subjects. Verbs were also the largest group in the nFA condition, followed by adverbs, then by nouns. Suppl. Tab. 1 in Appendix 1 provides additional information.

3.1.2 Matching

When taking the maximum possible numbers of words in each PoS, significant differences in control parameters were apparent throughout all tested word groups in every subject (Appendix 1: Suppl. Tab. 1: FA/nFA contrast, Suppl. Tab. 2: emoFA/NemoFA contrast). A matching procedure was therefore applied (see Methods). The number of trials required for successful matching differed between PoS and subjects. The proportion of matched words summated over both conditions of the FA/nFA contrast relative to the initial number or words was ca. 50% in S1, 59% in S2, 49% in S3, and 59% in S4. In the emoFA/NemoFA contrast, it was 31% of all available FA words in S1, 40% in S2, 40% in S3, and 29% in S4. Suppl. Tab. 3 (Appendix 1) provides detailed information on the success of matching FA and nFA words for each subject, PoS, and control parameter and also gives information on what parameters were different between these word groups prior to matching.

3.1.3 Rating

Emotionality is an inner state of the speaker which may be hard to evaluate using objective extrinsic measures (Truong et al. 2012). Like in our previous study confronting a similar problem on the example of semantics (Derix et al. 2014), we asked several independent raters to assess, whether the clauses in which the words carrying an FA were embedded were emotional or neutral. Then, we selected words from the sentences yielding a 100% inter-rater agreement for “emotional” and “non-emotional” assignments to create contrastive samples for further analyses. Our definition of “emoFA” words corresponded to words carrying an FA and stemming from sentences with emotional content, as could be identified based on the semantic, contextual, and phonological properties available to the raters from the continuous transcriptions. The “NemoFA” words were defined as words carrying an FA but stemming from sentences rated as “non-emotional” based on the aforementioned characteristics. The raters reported that, in their experience, a clause could be perceived as “emotional” due to various factors or their combinations, such as either because of the emotionally loaded topic of the conversation (e.g., a pending neurosurgical intervention the speakers were agitated about), or it could be a clause which was neutral with regard to the semantic content but produced with an emotional expression manifested by changes in speech volume, rhythm, or pitch. Inter-rater agreement in the assignment of emotionality proved only “fair” (21-40% in Landis and Koch (1977)) in all subjects (37% in S1, 35% in S2, 37% in S3, and 38% in S4). Due to the relatively low agreement between raters, we selected words for the emoFA and NemoFA categories only if all raters agreed on the assignment of the respective word. We recognize that, in some instances, 100% agreement may have arisen by chance, considering the overall fair quality of inter-rater agreement. We nevertheless expected this conservative method of assignment (100% agreement, cf. 75% agreement threshold in a study by Derix et al. 2014) to obtain word groups which would reflect qualitative differences between the emoFA and NemoFA categories in subsequent analyses of the neural data. On average, 25% of all words were classified as emoFA (S1: 27%, S2: 21%, S3: 27%, S4: 23%) and another 25% as NemoFA (S1: 25%, S2: 26%, S3: 25%, S4: 22%). Around 50% of the FA words had to be discarded at this stage of analysis in order to extract maximally distinctive emoFA and NemoFA categories.

3.1.4 RSM effects for FA, nFA, emoFA, and NemoFA words

The RSM effects underlying the production of the matched FA and nFA words were mainly manifested as increases in gamma activity and/or decreases in alpha/beta activity. These activity patterns are consistent with the typical properties of speech-related (Crone et al. 2001a; 2001b) and general event-related (Pfurtscheller and Da Silva 1999) neural responses. These effects mostly took place along the course of the central and lateral sulci and in Broca’s area and largely occurred at electrodes with mouth motor functions identified using ESM and in their neighbourhood within the pericentral cortex. Most prominent responses started prior to word onset. They were often concentrated already around speech start and ended prior to speech end. Figs. 4A, B show typical response patterns on the example of S4. The only subject in whom RSM responses were observed earlier was S1. RSM increases in this subject ended prior to word start at some of the motor electrodes (e.g., electrode E6 in Fig. 5A). Response patterns underlying FA and nFA words were generally similar in terms of the topography and time-frequency characteristics of the neural responses (cf. Figs. 4 and 5A). The same was true for emoFA and NemoFA words (not shown).

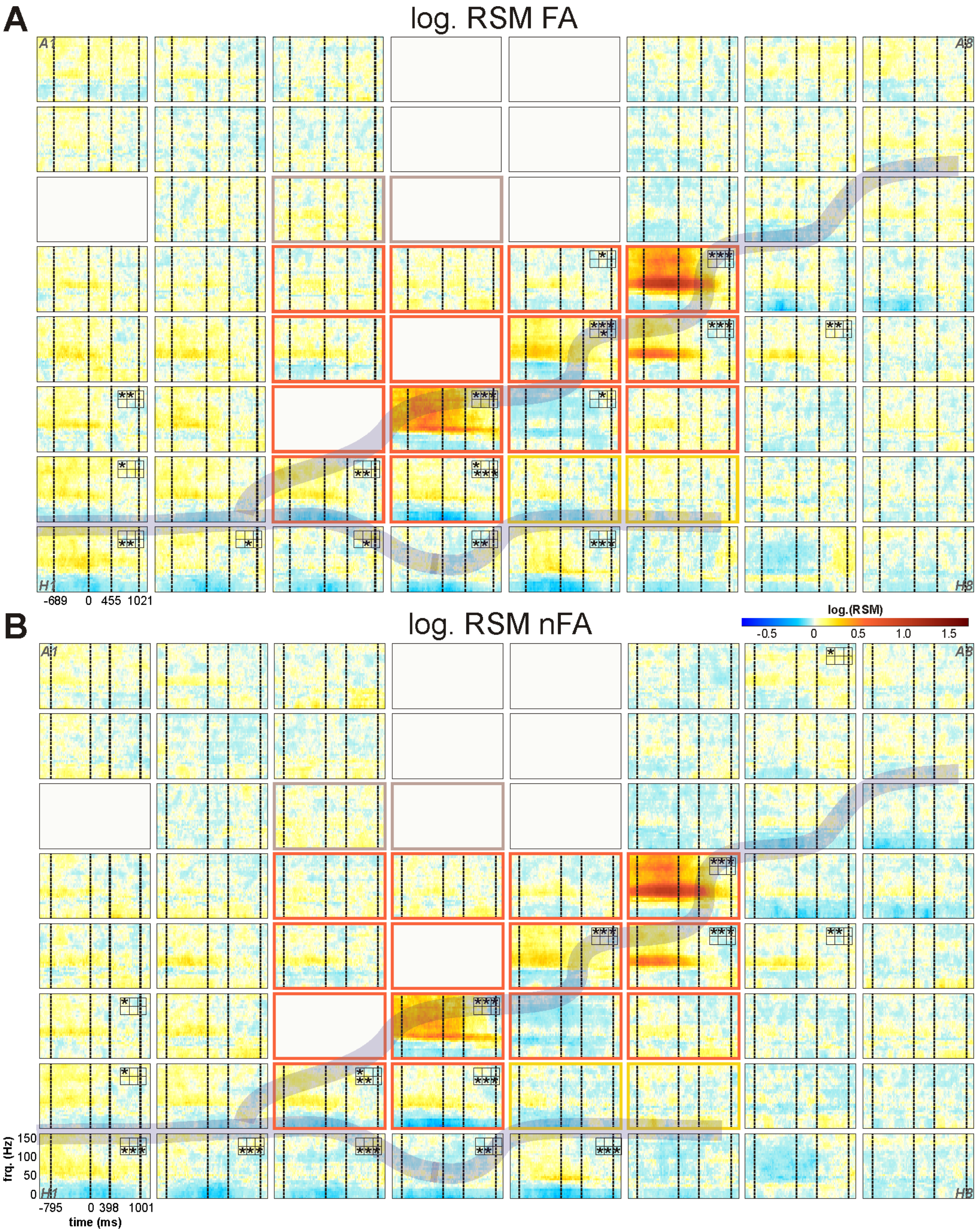

A
log. RSM FA
A1
A8
H1
H8
-689 0 455 1021
B
log. RSM nFA
log.(RSM)
-0.5 0 0.5 1.0 1.5
A1
A8
H1
H8
frq. (Hz)
150
100
50
0
-795 0 398 1001
time (ms)

Figure 4: Typical RSM responses in the FA (A) and nFA (B) conditions in relation to individual cortical architecture (S4). The responses at individual electrodes of the 8x8 grid shown for frequencies up to 150 Hz. Electrodes lying in speech-relevant areas identified using ESM (Fig. 1) are colour-coded; yellow-outlined: electrodes with speech perception and/or production functions, red-outlined: electrodes with mouth motor functions, brown: both. The transparent purple lines indicate the positions of the individual central and lateral sulci. Electrode names in the corners of the outer electrodes are provided for ease of spatial reference. Electrodes in white were excluded from the common average and removed from all analyses. Vertical dashed lines indicate, from left to right, average speech start, word start, average word end, and average speech end in the respective condition in the given subject. The small 2x3 grids with stars in the upper right corners of individual electrodes show the results of statistical testing: their three horizontally aligned squares code for the time of the significant effect (Wilcoxon sign test, FDR-corrected at $q<0.05$, see Methods); left: before, middle: during, right: after the average duration of the word in the respective condition. The two vertically aligned squares code for the frequency range of the significant effect; upper squares: gamma activity between 60 and 150 Hz, lower squares: alpha-beta activity in the range of ca. 10-35 Hz. This information is visualized for each electrode with significant effects. Abbreviations: frq.: frequency, log.: natural logarithm.

To describe differences between conditions in the articulatory cortex, we inspected neural activity at potentially articulation-relevant electrodes, i.e., those implicated movements of articulatory organs during ESM and showing significant RSM responses in both FA and nFA and respectively in both emoFA and NemoFA conditions. Typical RSMs at individual electrodes together with the ROI-averaged responses, the locations of these electrodes on the standard brain surface, and the ESM effects at these cortical sites are visualized on the example of the FA/nFA contrast in panels A, B, and C of Fig. 5, respectively.

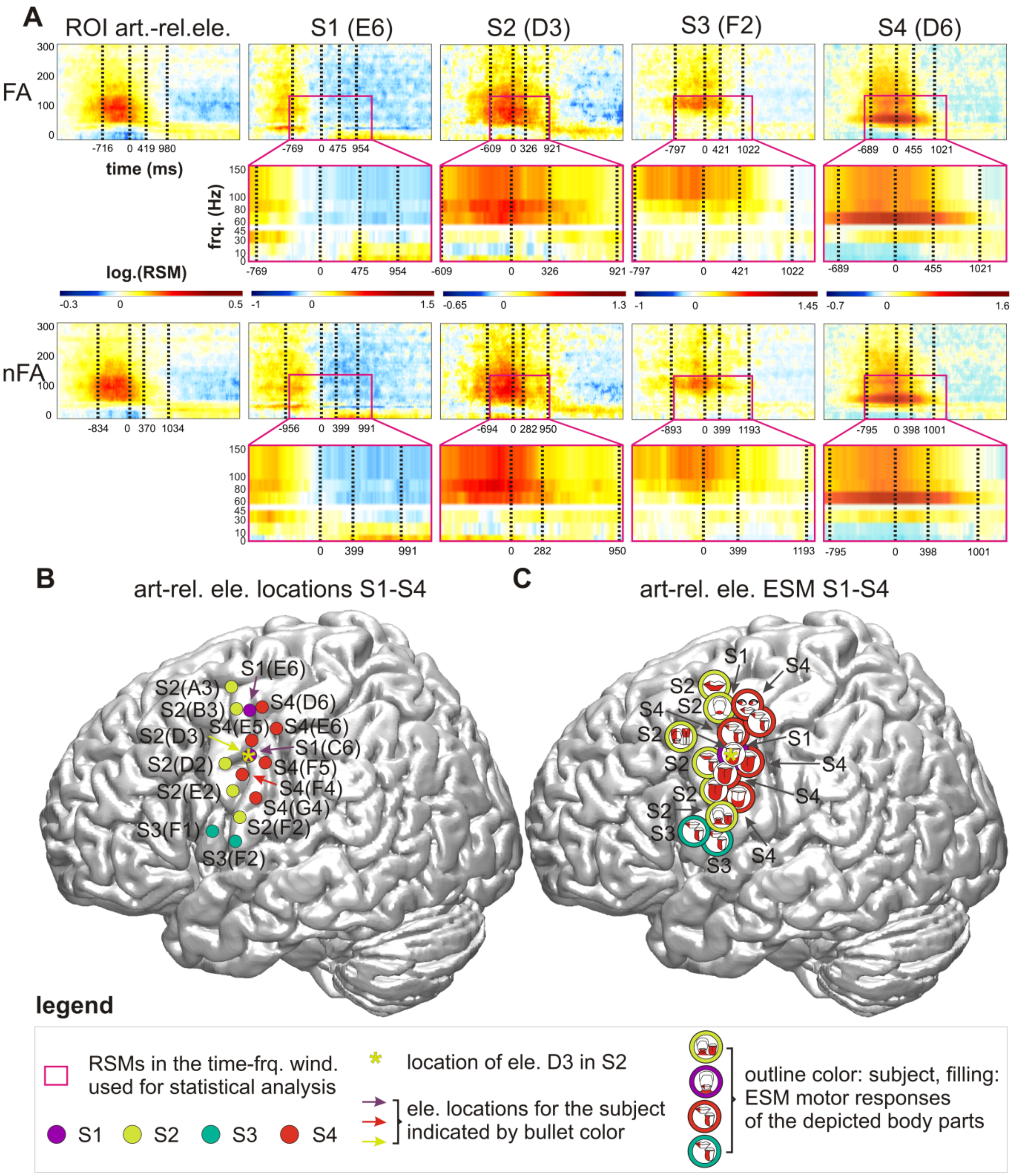

A
ROI art.-rel.ele.
S1 (E6)
S2 (D3)
S3 (F2)
S4 (D6)
FA
300
200
100
0
-716 0 419 980
time (ms)
-769 0 475 954
-609 0 326 921
-797 0 421 1022
-689 0 455 1021
frq. (Hz)
150
100
80
60
45
30
10
0
-769 0 475 954
-609 0 326 921
-797 0 421 1022
-689 0 455 1021
log.(RSM)
-0.3 0 0.5
-1 0 1.5
-0.65 0 1.3
-1 0 1.45
-0.7 0 1.6
nFA
-834 0 370 1034
-956 0 399 991
-694 0 282 950
-893 0 399 1193
-795 0 398 1001
0 399 991
0 282 950
0 399 1193
-795 0 398 1001
B
art-rel. ele. locations S1-S4
S1(E6)
S2(A3)
S2(B3)
S4(D6)
S2(D3)
S4(E5)
S4(E6)
S1(C6)
S2(D2)
S4(F5)
S4(F4)
S2(E2)
S4(G4)
S2(F2)
S3(F1)
S3(F2)
C
art-rel. ele. ESM S1-S4
S1
S4
S2
S2
S4
S2
S1
S2
S4
S4
S2
S2
S3
S3
S4
legend
RSMs in the time-frq. wind. used for statistical analysis
S1
S2
S3
S4
location of ele. D3 in S2
ele. locations for the subject indicated by bullet color
outline color: subject, filling: ESM motor responses of the depicted body parts

Figure 5: RSM changes at potentially articulation-relevant electrodes in relation to structural and functional anatomy in the FA and nFA conditions. A: ROI-averaged RSM responses in both conditions together with examples of individual mouth motor electrodes from each subject included in the articulatory ROI. The electrodes chosen for visualization showed maximum RSM values during word production in gamma frequencies between 60 and 150 Hz in the respective subject and in both conditions. The inset in magenta is magnified for the time period of -2 to 2 median word durations in the given subject and up to 150 Hz, averaged over each of the analysed frequency bands (0-10, 0-30, 30-45, 60-80, 80-100, and 100-150 Hz). Note that frequencies in the range of 45-60 Hz have not been analysed and are excluded in the frequency-averaged presentation (white stripes at each electrode in this frequency range). Art.-rel. ele.: potentially articulation-relevant electrodes (see Methods). **B**: The anatomical locations of the electrodes included in the ROI. The electrodes from each subject (colour-coded, see legend) are projected onto the standard brain surface from spm5 based on their MNI coordinates. **C.**: The ESM responses of the electrodes in the ROI visualized on the standard brain surface. Grey bullets point at the locations of the electrodes belonging to the subject specified by text and outline colour (see legend). Abbreviations in the legend: wind.: window, other abbreviations and conventions as in Figs. 1 and 4.

As can be seen from this figure, the responses at individual electrodes and the ROI-averaged responses (averaged RSM responses for FA and nFA conditions are shown in the two left-most panels of Fig. 5A) looked very similar in both FA and nFA conditions, and they took place at cortical sites which were mostly implicated in movements of the tongue and/or lips. Notably, the electrodes which were identified as articulation-relevant did not cover the entire spatial extent of the ESM-identified cortex with moth motor properties (cf. red overlay in Fig. 1 and red-outlined areas in Fig. 4). With few counterexamples (i.e., potentially articulation-relevant electrode D3 in S2 in conventional statistics and potentially articulation-relevant electrode E6 in S4 in single-trial decoding, see below), significant differences between word groups were found outside this group of electrodes.

3.1.5 Statistical comparisons between conditions

Tab. 2 shows the results of statistical testing for all time windows and frequency components at p<2e-04 (uncorrected). This was the most conservative threshold at which significant (Wilcoxon rank sum test in both contrasts) differences between conditions could be observed. Correction for multiple comparisons yielded no significant results. Most prosody-related effects took place in the postcentral cortex. With the exception of one subject (S2), differences between FA and nFA words in our data mostly occurred prior to the word and the differences between emoFA and NemoFA mostly after the word. This spatial and temporal reproducibility suggests that our findings may be functionally relevant. Fig. 6A illustrates the difference in RSM values between FA and nFA conditions in S4 from Fig. 4, the latter condition subtracted from the former at each individual time-frequency bin. Fig. 6B illustrates how the difference at electrode E6 in S4, which was hard to see by comparing FA and nFA conditions with the naked eye, becomes more prominent after the subtraction.

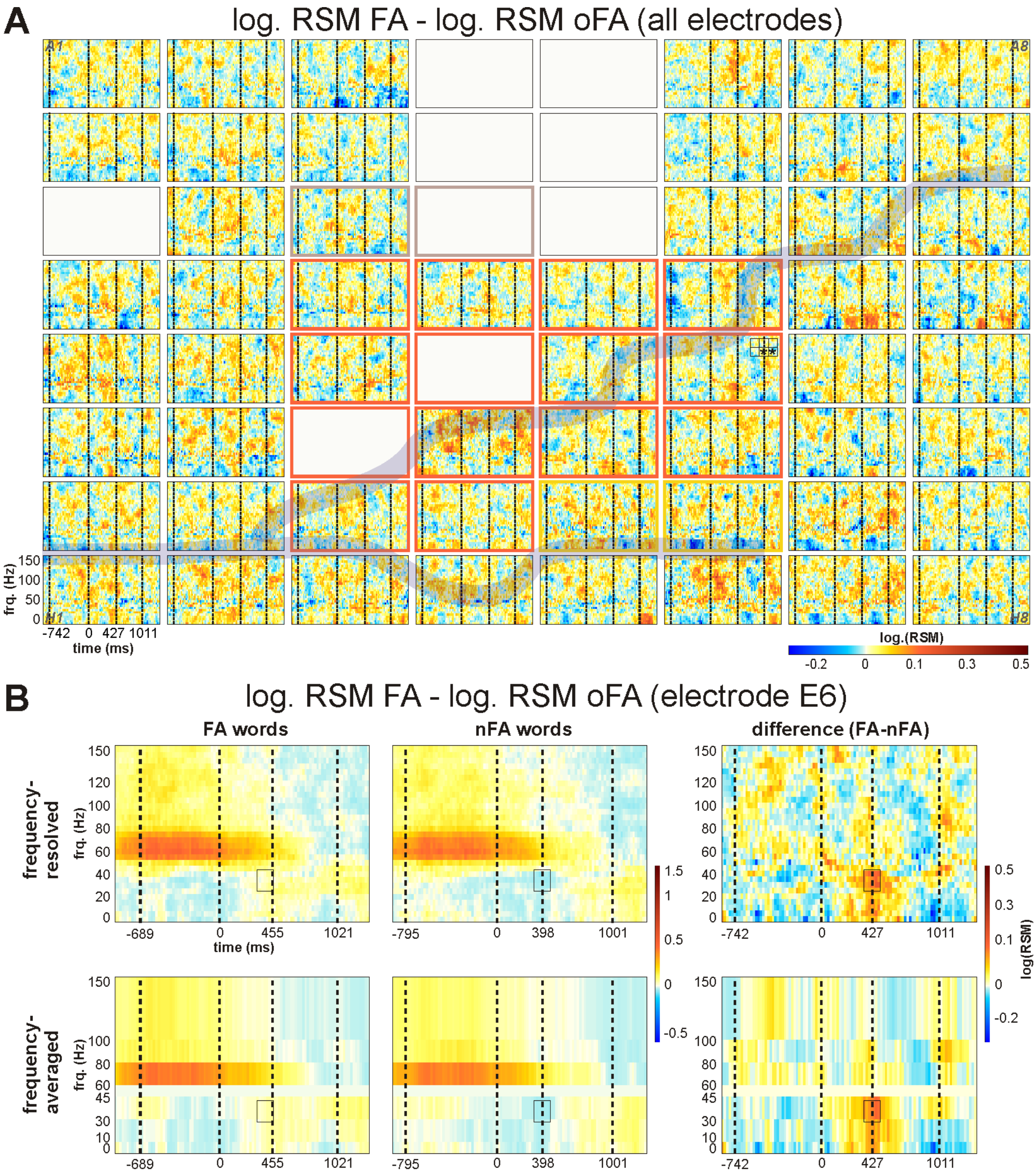


Figure 6: Example of RSM differences between FA and nFA conditions. A: RSM differences between FA and nFA conditions in S4 (cf. Figs. 4A and B for the effects underlying each respective condition) are shown. For this visualization, all RSM values in the nFA condition have been subtracted from all RSM values in the nFA condition at each individual time-frequency bin. The small 2x3 grid with stars in the upper right corner of electrode E6 indicates the presence of significant differences between conditions at this electrode (Wilcoxon rank sum test at $p<2e-04$, uncorrected). **B**: Relative spectral magnitude (RSM) changes at electrode E6 in S4: FA words, nFA words, and the difference between them. Upper row: frequency-resolved presentation, lower row: band-averaged RSM effects averaged over the six ranges of analysed frequencies (0-10, 10-30, 30-45, 60-80, 80-100, and 100-150 Hz). Note that activity between 50 and 60 Hz has not been analysed in the band-averaged data, and it is therefore excluded from presentation. The small black box in each panel highlights the time window within which the reported significant difference was observed. Conventions as in Fig. 4.

An interesting observation from this analysis is that only short, ca. 100-ms time windows yielded significant effects, suggesting that prosody-related processes in the left hemisphere may occur on very short time scales.

Table 2: The results of statistical comparisons between word groups for all analysed time windows and frequency components. The results were obtained using a Wilcoxon rank sum test at p<2e-04 (uncorrected). Abbreviations: s.: subject, time wind.: time window, frq.: frequency, time rel. to w. ons.: exact time of the effect(s) relative to word onset, approx. eff. pos. rel. to w. dur.: approximated effect position relative to average word duration of “long2” (footnote), log. diff. FA-nFA: logarithmic difference between FA and nFA words (rounded to the second decimal), eff. size: size of the effect, RSM in the nFA or NemoFA condition subtracted from RSM in the FA or emoFA condition in per cent rounded to the first integer, bip.: bipolar, sign. ele.: electrode with significant effect, p-val.: p-value, K-S pos.: a positive result of the two-tailed Kolmogorov-Smirnov test, CS: central sulcus, other conventions for structural anatomy as in Fig. 1. The results in this latter column show that the conditions for application of the Wilcoxon rank sum test were nearly always satisfied (except for one comparison in S2, 60-80 Hz, indicated by grey font and an asterisk at the beginning of the respective line in the table. This effect should therefore be interpreted with caution, and it will not be treated as “significant” or visualized in the following.). The anatomical locations of all significant effects are visualized in Fig. 7.

cond.	**s.**	**time wind.**	**frq. (Hz)**	**time rel. to w. ons. (ms)**	**approx. eff. pos. rel. to w. dur.**	**log. diff. FA-nFA**	**eff. size**	**contrast direction**	**sign. ele.**	**p-val.**	**K-S pos.**	**anat. area**	**bip. ESM**
FA/nFA	S1	short 1	10-30	554 to 656	+1 m. dur.	-0.18	120%	nFA > FA	C2	6.23e-05	yes	IP	speech
	S2	short 1	80-100	-62 to 41	- 1 m. dur.	0.21	21%	FA > nFA	D8	1.87e-04	yes	IP	no effect
	S3	short 1	80-100	964 to 1066	+2 m. dur.	-0.27	133%	nFA > FA	G3	4.05e-05	yes	TC	speech
	S4	short 1	30-45	349 to 451	m. dur.	0.13	12%	FA > nFA	E6	4.85e-05	yes	SI	lip & tongue motor
cond.	**s.**	**time wind.**	**frq. (Hz)**	**time rel. to w. ons. (ms)**	**approx. eff. pos. rel. to w. dur.**	**log. diff. emoFA-NemoFA**	**eff. size**	**contrast direction**	**sign. ele.**	**p-val.**	**K-S pos.**	**anat. area**	**bip. ESM**
emoFA/NemoFA	S2	short 1	100-150	923 to 1025	+2 m. dur.	0.31	60%	emoFA > NemoFA	B4	1.40e-04	yes	CS	jaw motor
	S2	short 1	100-150	820 to 1230	+2 m. dur.	0.38	29%	emoFA > NemoFA	B5	1.65e-04	yes	SI	hand sensory
	S2	short 1	80-100	205 to 308	m. dur.	0.45	50%	emoFA > NemoFA	B6	7.58e-05	yes	SI	hand sensory
	*S2	short 1	0-10	103 to 205	m. dur.	0.71	54%	emoFA > NemoFA	E8	1.52e-04	no	IP	thigh motor
	S3	short 1	60-80	-103 to 0	-1 m. dur.	-0.52	154%	NemoFA > emoFA	B3	4.56e-05	yes	IP	thumb motor
	S4	short 1	80-100	-103 to 0	-1 m. dur.	-0.45	146%	NemoFA > emoFA	B3	1.93e-04	yes	PF	no effect

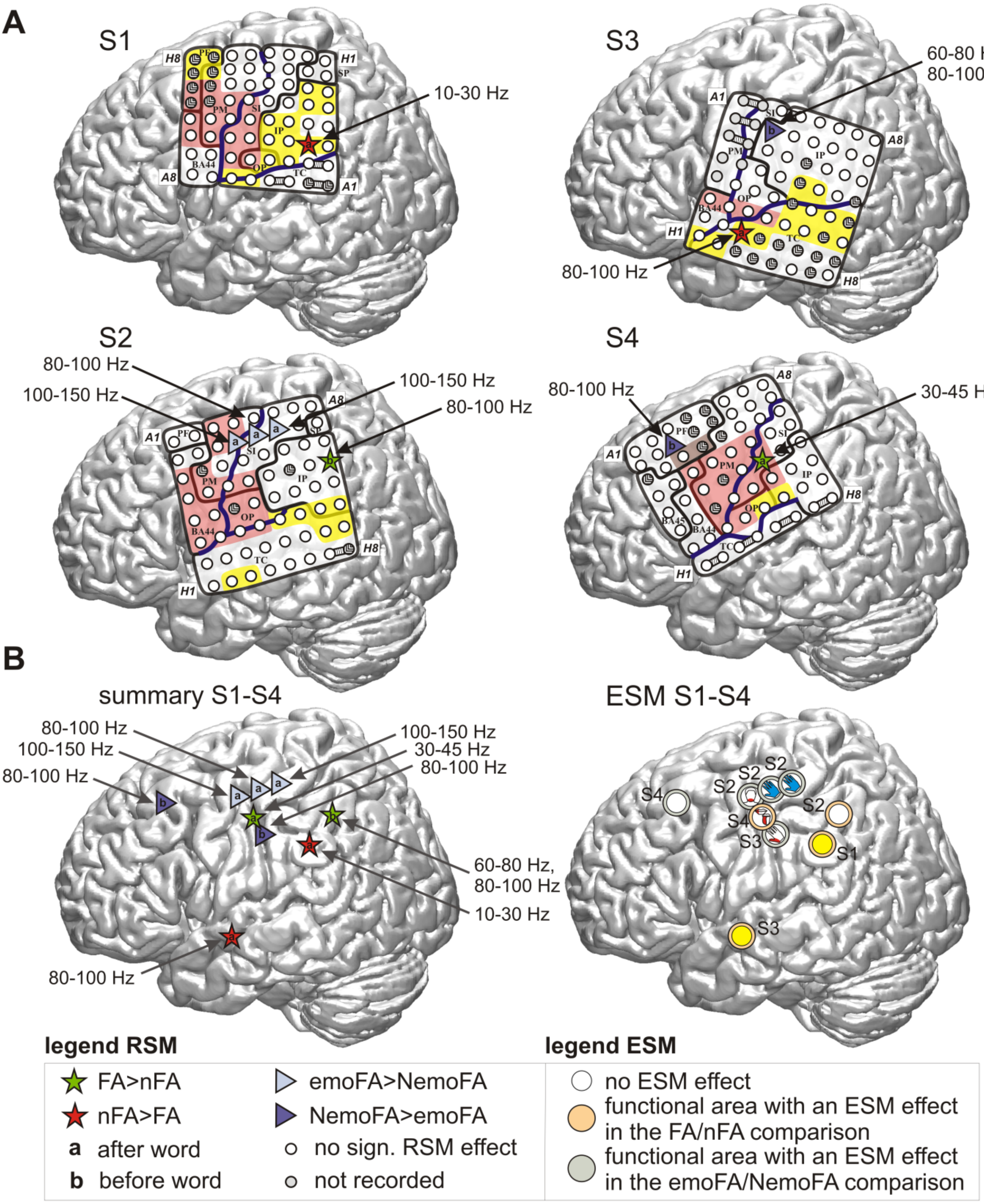


Figure 7: Statistically significant differences between word groups for all analysed time windows and frequency components. The results of the Wilcoxon rank sum test (p<2e-04, uncorrected) reported in Tab. 2 are visualized on the standard brain from spm5 based on the MNI coordinates of the electrodes in relation to the individual structural and functional anatomy of each subject (BA 44, BA 45: Brodmann areas 44 and 45, other conventions for structural anatomy as in Fig. 1).

As can be seen from Fig. 7 and Tab. 2, statistical comparisons between FA and nFA word groups yielded spatially focalized results. They took place in the postcentral cortex (S1, S2, S4) of three subjects and in the superior temporal cortex of one subject (S3, see the summary in the left panel of Fig. 7B). In most subjects (S1, S3, S4), these effects were located either in speech-cognitive or in mouth motor areas identified by ESM (summarized in the right panel of Fig. 7B), suggesting that they may play a role in linguistic processing. In terms of the timing and frequency characteristics, these effects were not systematically reproducible. They were manifested in different frequency bands and occurred at different time points relative to word onset. All effects occurred either before or after the median word duration in the respective subject and contrast. Differences between emoFA and NemoFA words took place only in three out of four subjects (S2, S3, S4). These effects were anatomically less reproducible than those in the FA/nFA contrast: they lay on the central sulcus and in the adjacent primary somatosensory cortex of S2, in the anterior inferior parietal cortex of S3, and in the prefrontal cortex of S4. The assignment of electrodes to functional areas also showed no consistent functional pattern (right panel of Fig. 7B).

3.1.6 Single-trial decoding

As our analysis using conventional statistics reported above was only capable of elucidating differences between word groups in uncorrected testing, we applied machine-learning methods to establish is they would yield better discrimination. This was indeed the case: significant single-trial decoding could be achieved in three subjects for the FA/nFA contrast and in all subjects for the emoFA/NemoFA contrast (Tab. 3, Fig. 8). The normalized DAs ranged between 58% and 61% in the FA/nFA contrast and between 63% and 82% in the emoFA/NemoFA contrast (chance level: 50%). Although these DAs are not very high, they proved significant after Bonferroni correction ($q<0.05$) for the number of tested electrodes within the respective frequency band.

Table 3: The results of single-trial decoding for both FA/nFA and emoFA/NemoFA contrasts. All time points within the time window “long2” in the respective frequency band were analysed together. Significance threshold and method of correction for multiple comparisons: $q<0.05$, Bonferroni; n. DA: decoding accuracy, rounded to the second decimal, other abbreviations as in Tab. 2. The anatomical locations of all significant effects are visualized in Fig. 8.

cond.	**s.**	**frq. (Hz)**	**sign. ele.**	**n. DA**	**p-val.**	**anat. area**	**ESM**
FA/nFA	S1	10-30	F2	0.59	9.15e-05	IP	speech
	S1	10-30	C2	0.58	8.44e-04	IP	speech
	S2	80-100	D3	0.61	4.37e-06	CS	face, lip & thigh motor
	S2	80-100	E8	0.59	3.15e-04	IP	thigh motor
	S3	0-10	A4	0.58	0.001	IP	not stimulated
	S3	0-10	A5	0.59	3.15e-04	IP	no effect
cond.	**s.**	**frq. (Hz)**	**sign. ele.**	**n. DA**	**p-val.**	**anat. area**	**ESM**
emoFA/NemoFA	S1	30-45	A4	0.71	2.49e-04	TC	not stimulated
	S1	30-45	A5	0.71	2.49e-04	LS	speech
	S1	10-30	C3	0.69	6.50e-04	IP	speech
	S1	60-80	D7	0.82	1.39e-07	PM	lip & tongue motor
	S1	0-10	E3	0.69	6.50e-04	IP	index finger motor
	S2	30-45	G4	0.73	9.76e-06	TC	no effect
	S2	80-100	E4	0.68	4.51e-04	OP	tongue motor
	S3	60-80	D3	0.68	4.51e-04	OP	no effect
	S4	0-10	B6	0.64	3.82e-04	PM	hand motor
	S4	60-80	C6	0.63	7.09e-04	PM	eye motor
	S4	30-45	D5	0.64	3.82e-04	PM	eye, eye lid & lip motor
	S4	30-45	G2	0.63	7.09e-04	BA	no effect

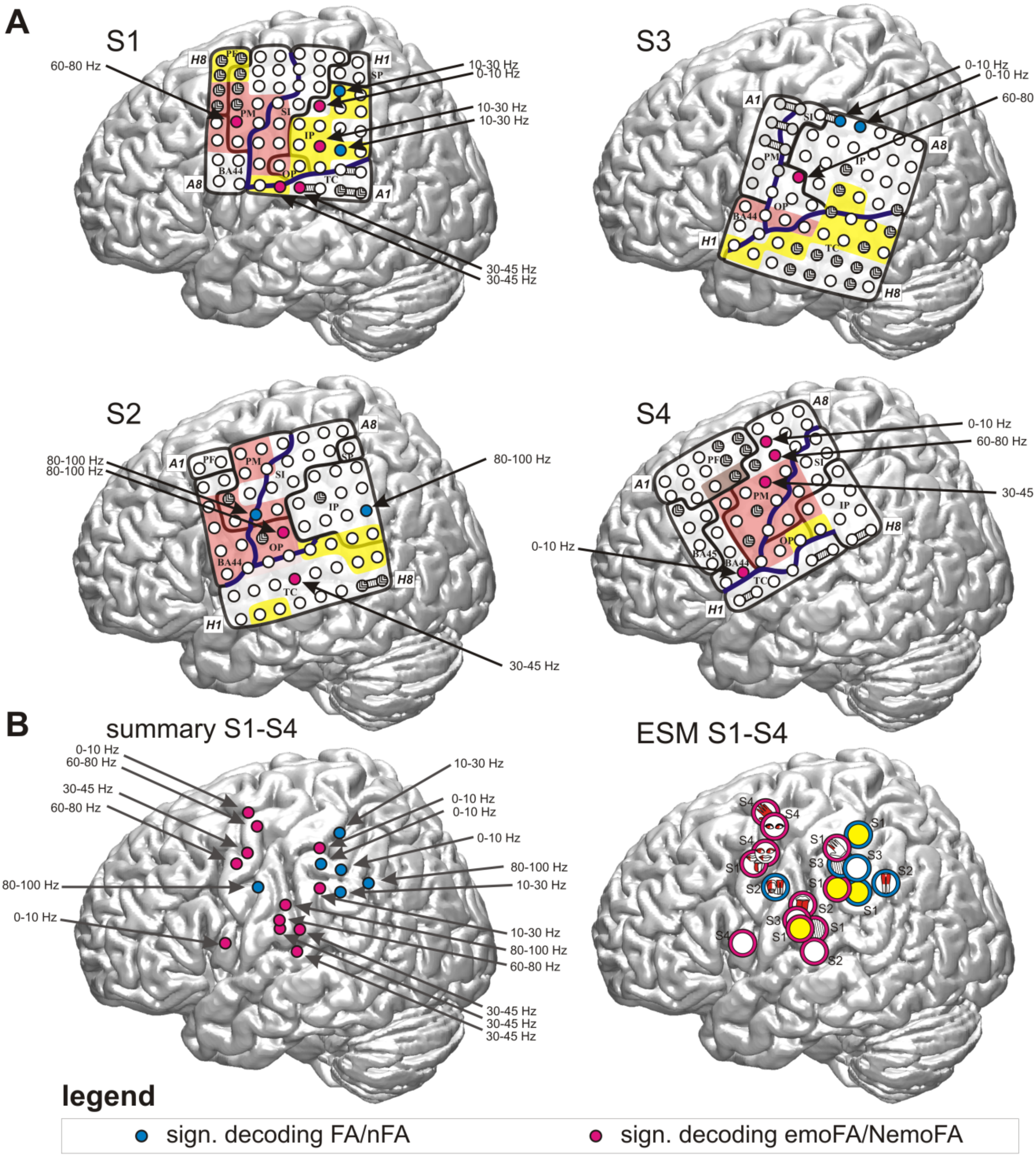


Figure 8: Significant (q<0.05, Bonferroni-corrected) single-trial decoding for both FA/nFA and emoFA/NemoFA categories. Conventions as in Figs. 1 and 7.

Anatomical and functional areas with significant decoding of FA/nFA words were similar to those observed in conventional statistics. The effects lay predominantly in the inferior parietal cortex (left panel of Fig. 8B). In two out of the three subjects in whom significant decoding could be achieved (S1, S2), they occurred in areas implicated in cognitive (S1) or motor (S2) aspects of language (Tab. 3, right panel of Fig. 8B). In S3 and at one electrode of S2, significant decoding results were additionally obtained in leg motor areas identified using ESM. The number of significant effects in the emoFA/NemoFA contrast was larger than in the FA/nFA contrast. Similar to conventional statistics, single-trial decoding yielded effects in anatomical locations which varied between subjects. Postcentral effects could be observed in three subjects (S1, S2, S3), precentral effects in two subjects (S1, S4), and effects in the superior temporal cortex in two subjects (S1, S2). Many significant effects in the emoFA/NemoFA comparison took place in mouth motor and cognitive language-relevant areas identified using ESM. Three subjects (S1, S2, S4) showed significant decoding in such areas. The effects in S1 occurred at contacts implicated in cognitive language functions, and S1, S2 and S4 showed effects in tongue motor areas. Examples of electrodes with hand motor functions were also observed in S1 and S4. The electrode showing a significant effect in S3 was not stimulated. The anatomical locations of these effects overlapped only marginally with those of the effects yielded by conventional statistics (cf. Tab. 2 and Tab. 3, Fig. 7 and Fig. 8). This result most likely resides in methodological differences between the statistical procedures and is partly due to the fact that numerous time points could be used without averaging RSM values over them in the decoding analysis.

3.1.7 Potential influence of acoustic parameters on the outcome of statistical testing

Once we had established that significant differences between conditions of interest could be observed, we were interested to test whether these differences could be explained by acoustic variables intrinsic to our prosodic phenomena, namely, by pitch and sound intensity of the words. Our analysis of word-accompanying parameters from the acoustic data showed that pitch and acoustic intensity values differed between conditions in a subject-specific manner. While FA words in S2 and S4 were associated with higher intensity values than nFA words, S1 and S3 showed no significant intensity differences in this contrast. S2 was the only subject to show higher pitch values in the FA than in the nFA condition. The same comparisons performed for the emoFA/NemoFA contrast revealed significantly higher pitch values underlying the production of emoFA words than NemoFA words in one subject only (S1). As is mentioned in the Introduction, we are aware of the fact that vocal pitch values in the production on the FA can vary not only between but also within word categories as well as between speakers, dialects, and communicative situations. Our comparison of pitch values between word groups does not represent an endeavour to arrive at linguistic generalizations; it merely presents an attempt to find factors which would be helpful to explain the observed differences in the neural data. The overall more pronounced differences in the acoustic parameters for the FA/nFA contrast may be due to the considerably larger number of trials in this comparison. Alternatively, they may reflect the more attenuated acoustic differences in the emoFA/NemoFA contrast.

Table 4: Comparison of word-accompanying pitch and intensity parameters from the auditory signal. A two-tailed Wilcoxon rank sum test was used at p<0.05 to test whether the acoustic parameters ("param.") underlying the analysed words differed between conditions (C1: first word group of the contrast, C2: second word group of the contrast, see column "contr."). "H" refers to the hypothesis of the significance test, H=1 indicates that word groups were significantly different. A grey overlay is used to additionally mark the parameters for which significant differences between conditions could be observed in the respective subject. The direction of contrast information is provided for these word groups in column "log(C1-C2)".

s.		S1			S2			S3		S4		
	ranksum param.	**H**	**p-val.**	**log.(C1-C2)**	**H**	**p-val.**	**log.(C1-C2)**	**H**	**p-val.**	**H**	**p-val.**	**log.(C1-C2)**
FA/nFA	**pitch (Hz)**	0	0.48		0	0.05		0	0.41	1	6.6e-04	0.04
	intensity (dB)	0	0.29		1	0.009	0.03	0	0.48	1	1.2e-04	0.02
emoFA/ NemoFA	**pitch (Hz)**	1	0.03	0.02	0	0.22		0	0.44	0	0.86	
	intensity (dB)	0	0.07		0	0.87		0	0.96	0	0.61	

In the next step, we juxtaposed the presence or absence of significant differences in the aforementioned acoustic parameters between word categories (Tab. 4) with the presence or absence of significant neural differences identified between word categories with conventional statistics (Tab. 2, Fig. 7). In doing so, we accounted for the direction of contrast in both tests. Higher levels of gamma and/or lower levels of alpha/beta activity in the neural signal are known as reliable markers of speech-related cortical activation (Crone 2001a,b), and higher pitch and intensity can be expected to result in stronger articulatory effort and hence in stronger cortical activation. Accordingly, if higher levels of cortical activation are associated with higher values of acoustic parameters in a given contrast, an explanation of the observed neural effects in group comparisons in Tab. 2 and Fig. 7 is likely. Since neither pitch nor intensity of the acoustic signal differed between FA and nFA conditions in S1 and S3 (Tab. 4), an explanation of the observed effects by differences in these parameters does not seem plausible. Another possible reason for the absence of systematic differences in acoustic parameters between FA and nFA conditions in these subjects may reside in the fact that we concentrated all analyses on word-level parameters. Values for acoustic parameters, averaged over the entire duration of the word, may be too coarse to reflect the temporally fine-grained structure of the acoustic signal. Exploration of single syllables was beyond the scope of the present study, and future research will be needed to address this possibility. The higher level of cortical activation (80-100 Hz) in the FA/nFA contrast in the IPC of S2 (Fig. 7) shortly prior to word start may indicate increased preparatory activity related to the higher acoustic intensity of FA compared with nFA words. The higher level of cortical activation (30-45 Hz) in the FA/nFA contrast in the SI of S4 (Fig. 7) after the production of FA compared with nFA words may be due to feedback-related processing of the higher pitch and acoustic intensity values in the former condition (Tab. 4). An explanation of the observed differences between emoFA and NemoFA words by differences in acoustic parameters is conceivable in one (S2) out of the three subjects (S2-S4), who showed significant differences between word groups in this contrast (Tab. 2, Fig. 7). Higher levels of activation (80-150 Hz) observed in S2 at electrodes in SI and on the central sulcus after the production of emoFA words could reflect feedback-related processing of the higher pitch values in this condition.

The single-trial decoding analysis yielded more and statistically more robust effects in group comparisons (Tab. 3, Fig. 8), compared to conventional statistics (Tab. 2, Fig. 7). This analysis, however, does not provide information on the direction of contrast. A more general approach than the one we took above when interpreting the results of conventional statistics will therefore be pursued when interpreting the decoding results. Since significant decoding of the FA/nFA contrast was possible in S1-S3 and considering that significant differences in acoustic intensity values could be observed for this contrast in S2 (Tab. 4), an explanation of significant decoding by higher intensity of the acoustic signal in the FA condition in this subject is conceivable. In the same vein, an explanation of significant decoding by higher vocal pitch values in the emoFA than in the NemoFA condition in S1 (Tab. 4) cannot be ruled out. Note, however, that most of the observed effects could not be explained by these acoustic parameters. Also, the location of reproducible effects in the inferior parietal cortex for the FA/nFA contrast does not agree with the typical locations of pitch-related cortical processing (Belyk and Brown 2013).

3.1.8 Potential influence of body movements on the outcome of statistical testing

The majority of the observed neural effects agreed well with potentially language-relevant areas identified using ESM and in their immediate proximity (Tabs. 2 and 3, Figs. 7 and 8). There were nevertheless also examples of significant differences between word groups outside these areas. Some electrodes with significant differences, e.g., showed extremity movement functions in ESM (FA/nFA: electrodes with leg motor functions with significant decoding in S2, S3; emoFA/NemoFA: electrodes with hand motor functions and significant results in conventional statistics in S2, S3 and electrodes with hand motor functions with significant decoding in S1, S4, cf. Tabs. 2 and 3). A further question was whether there were differences in EMG recordings between words groups, and if our observed neural effects would be accompanied (and thus possibly attributed to) such differences.

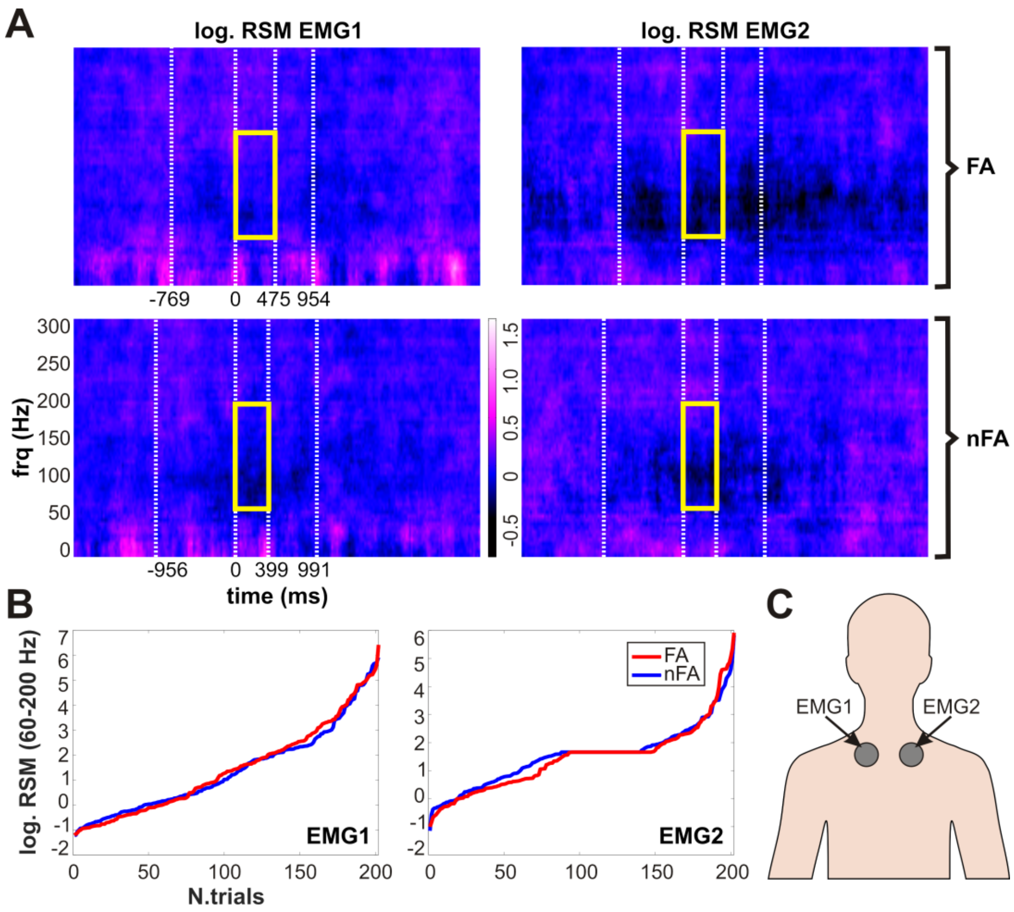


Figure 9: Examples of word-accompanying EMG activity from bilateral upper-chest electrodes of S1. The levels of EMG activity were compared between FA and nFA conditions. **A**: the EMG data in each condition were analysed for the EMG1 and EMG2 electrodes. The yellow box shows the time and frequency range for which the time- and frequency-averaged RSM values were compared. The logarithmic RSM values (see colour map between the two lower panels of A) had been obtained using the same procedure as for the neural data, all conventions as in Fig 2. **B**: a visualization of RSM values at both electrodes in both conditions. **C**: the positions of the EMG electrodes.

Our analysis of the available EMG recordings showed that no significant differences (Wilcoxon rank sum test) between FA/nFA (S1: p=1 in EMG1, p=0.25 in EMG2; S3: p=0.71 in EMG1; S4: p=0.68 in EMG1; Fig. 9B) and emoFA/NemoFA (S1: p=0.53 in EMG1, p=0.46 in EMG2; S3: p=0.64 in EMG1; S4: p=0.2 in EMG1) conditions could be observed during word production. Interestingly, EMG levels in S1 were reduced during the production of both FA and nFA words, indicating that the subject was moving his upper extremities less during speech and word production than prior to it. The levels of EMG activity during word production were comparable between FA and nFA and also between emoFA and NemoFA conditions. Fig. 9A illustrates this on the example of the FA/nFA contrast. EMG in our subjects was not recorded from lower extremities, and EMG reflecting upper-extremity movements was only recorded in S1 (Fig. 9C). We therefore cannot be certain whether or not movements of unrecorded body parts had contributed to the significant differences we observed between conditions outside the language-relevant cortex in the emoFA/NemoFA comparison using conventional statistics in S2 (3/3 electrodes with significant effects had upper-extremity functions in ESM), in S3 and S4 (1/1 electrodes with significant effects had upper-extremity functions in ESM (Tab. 2, Fig. 7). Note, however, that all effects in the respective FA/nFA analysis lay at ESM-identified electrodes implicated in language functions. An explanation of effects in this contrast by extremity movements is thus unlikely. As to the results of the single-trial decoding analysis, 1/2 electrodes with significant effects for the FA/nFA contrast were associated with leg motor functions in ESM. 1/5 and 1/4 electrodes with significant differences in the emoFA/NemoFA conditions in respectively S1 and S4 were localized to upper-extremity motor areas. Accordingly, a contribution of extremity movements to the observed neural differences between our word groups of interest cannot be ruled out, especially for the emoFA/NemoFA contrast. Even if one takes this possibility into account, the majority of our observed effects in the FA/nFA contrast lay in language-relevant areas (Figs. 7 and 8) lending credibility to the assumption that differences between neural activity patterns in this latter contrast reflect prosody-related processing.

3.2 Results of Study 2: word complexity

3.2.1 The selected language material

Words no longer than three spoken syllables were used due to the constraints related to the calculation of EoA (see Methods). We only used unique words without repetitions to avoid domination of the neural effects by repeated occurrences of the same word in the subject's sample. As can be seen from Tab. 5, we were able to obtain relatively large data sets in spite of these constraints. The words were harvested from over a half of the collected simple clauses. The proportional relation between the total numbers of the gathered content words relative their numbers within the respective PoS category was reproducible in all subjects (last column of Tab. 5). More than a half of the nouns could be used for the analysis, while the categories of full verbs and adverbs lost many words in comparison. This is because the latter two word categories showed considerably more repetitions within the lexical samples of the individual subjects. The reproducible proportional relations between and within the analysed PoS categories appear to reflect the subject-unspecific linguistic composition of spontaneously spoken German.

Table 5: The amount of the collected and analysed linguistic material. Abbreviations: ling.: linguistic, ADV: adverbs, FV: full verbs, NN: normal nouns, №. w.: number of all words together regardless of their PoS category, №. cl.: number of clauses; sel./total: number of the respective language units selected for the analysis, divided by the number of such units gathered per subject in total, abs.: in absolute values, %: expressed in per cent rounded to the last integer; mean: the average percentage of the selected language units relative to the total number of these language units, the absolute maximum deviations from this value within our sample of subjects are given in brackets. Other conventions as in Tab. 1.

s. / ling. unit	S1		S2		S3		S4		S5		mean
	proportional relation sel./total										
	%	abs.	%	abs.	%	abs.	%	abs.	%	abs.	%
№. ADV	29%	83/286	26%	65/248	28%	68/240	29%	156/535	32%	135/425	29% (±3%)
№. FV	37%	143/389	38%	115/302	40%	132/327	35%	260/741	34%	173/509	37% (±3%)
№. NN	64%	81/127	60%	60/100	51%	61/120	56%	177/317	52%	84/161	57% (±7%)
№. w.	38%	307/802	37%	240/650	38%	261/687	37%	593/1593	36%	392/1095	37% (±1%)
№. cl.	57%	218/383	57%	170/299	59%	190/325	60%	433/723	56%	282/506	58% (±2%)

3.2.2 Typical RSM changes related to word production

We conducted a spectral magnitude analysis to find out if the gathered content words would be associated with clear and reproducible patterns of neural activity. The RSM spectra, triggered to word onset and averaged over the entire number of content words in each subject, were calculated and tested for significance. The results were very similar as in the study on prosody (not visualized, see Fig. 4 in Study 1 for a typical picture). The most pronounced neural responses occurred in a broad range of gamma frequencies, in which they were manifested as increases in the spectral magnitude. These responses took place at individual electrodes within the pericentral mouth motor cortex, and they were significant not only during but also before and after the onset of word production. The aforementioned effects tended to start around speech start and end between word end and speech end. This suggests their contribution to preparatory and executional processes. Since all words were averaged in this analysis regardless of their linguistic properties, these responses likely reflect general articulatory features related to word and speech production which are linguistically unspecific. More attenuated changes in gamma activity took place at other cortical locations including the posterior part of Broca’s area adjacent to the motor cortex and in the superior temporal and parietal regions. Effects in the lower frequencies could also be observed. These were mainly decreases in the alpha and beta frequencies, which took place prior to and during word production.

3.2.3 Correlation structure in the linguistic data

The evaluation of collinearity between the linguistic parameters revealed a number of statistically significant correlations which were reproducible across subjects (not visualized). These significant correlations will be referred to as "strong" whenever the r values exceeded 0.4 and as "weak" otherwise. Among the linguistic parameters, EoA, for which higher values indicate greater ease of word production, consistently showed strong positive correlations with FRQ. This means that frequent words were modelled as easier to articulate than rare words with the help of Ziegler and Aichert's model (2015). The inverse relation between word frequency and complexity in our data is in line with psycholinguistic findings that high-frequency words are faster recognized and named and that they are more often pronounced correctly in comparison with low-frequency words (Bose, van Lieshout, and Square 2007; Connine et al. 1990; Oldfield and Wingfield 1965). The strong negative correlation of EoA as well as FRQ with NoS also indicates that more frequent words, which were associated with a greater ease of articulation, also tended to have fewer syllables. In a similar vein, the strong negative correlations between both EoA and FRQ with word duration (ws_we) showed that the production of articulatorily complex, low-frequency words lasted longer. As one would expect, NoS was strongly positively correlated with ws_we. The negative correlation between lexical frequency and these measurements of word length is in line with the observation by Zipf (1935: 38) that "the length of a word tends to bear an inverse relationship to its relative frequency." CVR was weakly negatively correlated with EoA, suggesting that consonants made words more difficult to pronounce than vowels did. This agrees with the linguistic and psycholinguistic literature indicating that the articulation of consonants is more demanding than the articulation of vowels (Ziegler and Aichert 2015). A weak negative correlation between CVR and NoS occurred in all subjects. The words which had more syllables thus tended to have fewer consonants in relation to vowels compared with words with fewer syllables. This is not surprising, considering that vowels are the essential building blocks of syllables.

We made an interesting observation with respect to a word's position in the speech production epoch. In S1-S4, the duration from articulation onset to word start (ss_ws) was weakly positively correlated with ws_we. Similarly, the duration from word end to speech end (we_se) in all subjects showed weak to strong negative correlations with ws_we. Longer words were thus farther away from ss and closer to se than shorter words. ss_ws showed reproducible weak negative correlations with FRQ (S1, S3-S5) and a weak positive correlation with NoS (S1-S5). This does not only mirror the aforementioned inverse relation between NoS and FRQ but it also indicates that low-frequency, polysyllabic words tend to have a greater temporal distance to ss. At the same time, the weak negative correlation between we_se and NoS (S2-S4) suggests that the duration we_se is shorter for words with more syllables. Taken together, these correlations provide evidence that longer and less frequent words have a tendency to occur later in the speech production epoch. This agrees with previous findings that high-frequency, shorter words display a tendency to occur earlier in multi-word sequences during speech production, likely because they are easier to access in the mental lexicon than low-frequency, longer words (Keenan and Comrie 1977; Kelly, Bock, and Keil 1986; Navarrete, et al. 2006). See Dieminger (2017) for additional information on collinearity between the aforementioned parameters.

Reproducible correlations were observed for two duration-related parameters with the intensity of the acoustic signal. It was weakly negatively correlated with ss_ws (S2-S4) and weakly to strongly positively correlated with we_se (S2, S4-S5). Accordingly, words which occurred earlier in a speech production epoch had the tendency to be pronounced more loudly. Reproducibility with regard to correlations of the linguistic parameters with EMG activity could not be assessed, since the locations of the EMG channels differed between subjects.

All in all, the reproducible correlations between EoA, NoS, CVR, FRQ and the duration-related parameters were consistent with psycholinguistic literature. They reflected the organizational properties pertinent to human language as a compositional, rule-based system. The presence of multicollinearity described in this subchapter was taken into account in the subsequent analyses aimed to identify parameter-specific effects in the neural signals.

3.2.4 Neurocorrelations of RSM changes with duration-related parameters

Out of all linguistic parameters which we correlated with the RSM values, those related to the position of the word within a speech production epoch elicited most pronounced neurocorrelation patterns. These are illustrated on the example of S5 in Fig. 10. This figure shows neurocorrelation results for the parameters ss_ws and we_se in the respective panels A and B against the individual structural and functional anatomy of the subject. The effects representative of these parameters were spatially focalized to electrodes within the pericentral mouth motor cortex. They occurred at the same electrodes for both ss_ws and we_se, and also at the same electrodes at which the clearest RSM responses underlying the production of content words could be observed (data not shown). The presence of such correspondence lends proof to the suitability of the neurocorrelation approach for identifying the neural signal components modulated by a given parameter. It also suggests that the neural effects observed with regard to the parameters ss_ws and we_se are related to the linguistically-unspecific processes which are likely due to the preparation and execution of articulation. The fact that selective parts of the mouth motor cortex were involved suggests a functional specificity for such processes within the mouth motor cortex.

Like in the RSM analysis, most pronounced effects occurred in the gamma frequencies. Interestingly, the maximum spectral magnitude changes underlying the production of content words and the highest positive correlation values with the duration-related parameters ss_ws and we_se (Fig. 10) took place at the electrodes at which the frequency range of the effect was broadest. The gamma effects in the neurocorrelation analyses differed in the timing of their occurrence between ss_ws and we_se: such effects tended to occur prior to the onset of word production for ss_ws and mostly after word production for we_se. The positive correlations in the gamma frequencies were sometimes accompanied by negative correlations in the alpha and beta frequencies. This is reminiscent of the classical domain-general pattern of the event-related spectral response involving simultaneously increasing gamma and decreasing alpha-beta activity (Pfurtscheller 1996; Pfurtscheller and Da Silva 1999). The correlations of RSM with the parameter we_se, however, yielded more prominent effects in the lower frequencies than the correlations with ws_we did. These likely points to a greater functional role of low-frequency activity in the processes related to termination rather than initiation of motor actions (Demandt et al. 2012) including articulation.

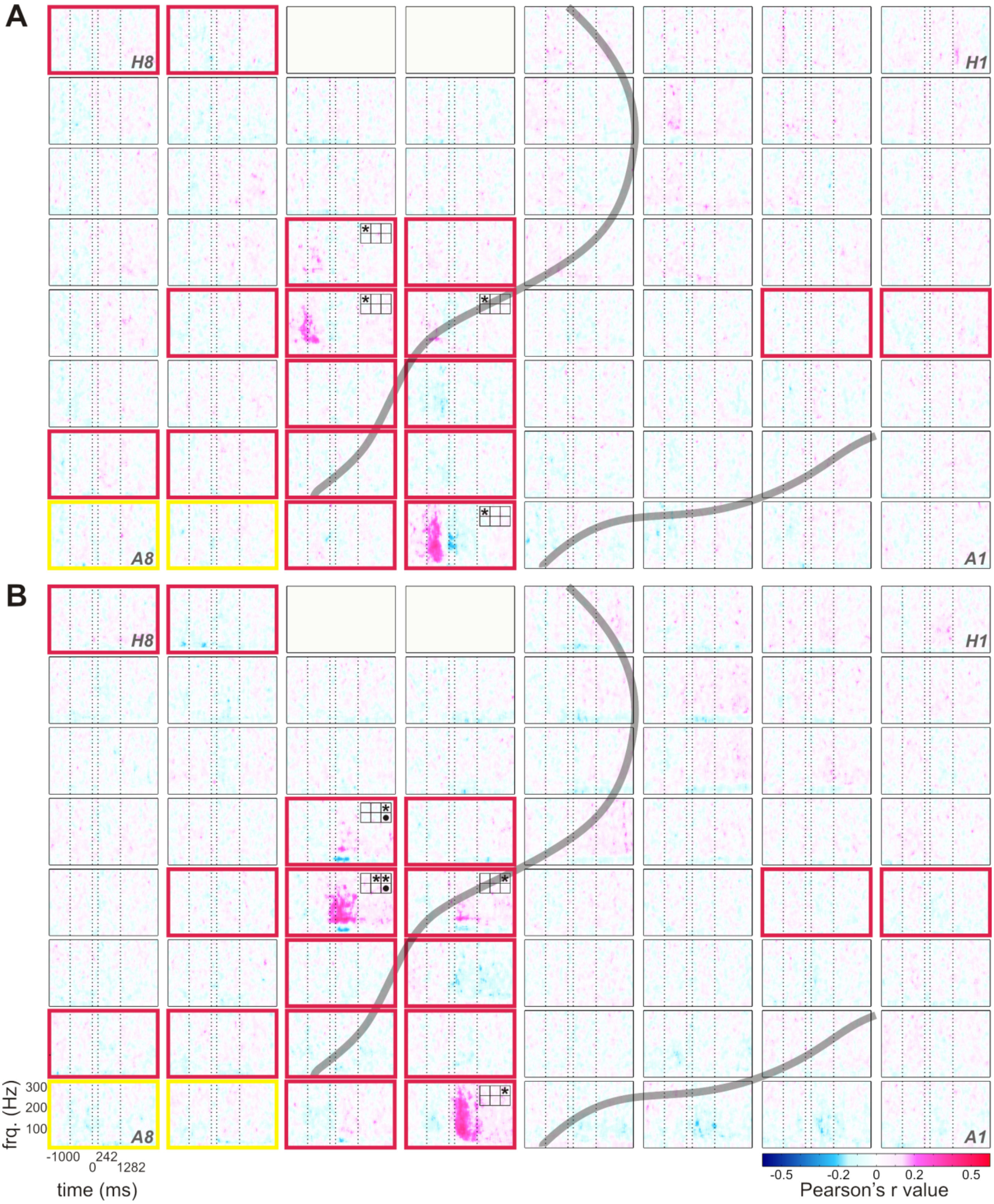


Figure 10: Typical correlations of RSM responses related to the production of content words with the duration from speech start to word start (A) and with the duration from word end to speech end (B), example from S5. The correlation values of RSM with the respective parameters are colour-coded (see legend), all other conventions as in Fig. 4.

In comparison with the parameters ss_ws and we_se, the neurocorrelations with word duration (ws_we) did not show such pronounced effects (data not shown). Although the tags for word and speech starts and ends were set according to the same principles, the neurocorrelation effects related to word duration were more attenuated, and they seldom reached significance in the Bonferroni-corrected testing for multiple comparisons over all time-frequency points and electrodes ($q<0.05$). Generally, ss_ws showed most pronounced positive correlations with RSM values around speech start (before and after it but prior to the start of the word) and the strongest positive correlations with we_se occurred between word end and speech end.

3.2.5 Statistical testing of neurocorrelation results prior to residualization

When applying statistical testing on the correlations of the RSM values with the individual linguistic parameters, we noticed that, while some parameters, such as the duration-relevant ones, yielded very strong effects which survived conservative statistical testing, others yielded significant effects only using less conservative thresholds, if any effects for a given parameter could be observed at all. This is not surprising, since the investigated parameters describe different aspects of the linguistic data which may be differently represented in neural activity. If one selects a single conservative test with a conservative threshold, one might thus overlook effects of some parameters which may be of interest. On the contrary, however, if one tests all data using an unconservative statistical approach, this may compromise the spatial specificity of the neural effects and produce a topographically meaningless picture. Due to these considerations, we applied a sequence of tests which were differently conservative. They ranged from Bonferroni-corrected testing to uncorrected testing at different thresholds up to 5E-06 (these results are shown for the main linguistic parameters of our interest in Tab. 6). Note that even the least conservative test was thus relatively strict and it yielded spatially meaningful, focal neural effects. For each parameter, however, we sought to find out, what effects would survive the most conservative testing possible (i.e., the most conservative test and threshold at which any effects for the given parameter could be observed at all, highlighted in green in Tab. 6). If Bonferroni-corrected testing yielded significant results, the result of this test was reported and visualized in the figures depicting neurocorrelation results (Figs. 11-13), even though uncorrected testing at very conservative thresholds sometimes elicited very similar results (e.g., CVR in S3 in Tab. 6). In total, effects for FRQ could be observed in all 5 analysed subjects, for NoS in 4 subjects, and CVR and EoA yielded significant effects in 3 subjects only. In terms of the numbers of electrodes with significant effects per subject per parameter, spatially specific results could be observed: the numbers of electrodes with significant effects, identified as described above, ranging from 1 to 6 electrodes.

Table 6: An overview of the outcomes when testing correlations between RSM responses and the linguistic parameters before residualization with the help of different statistical procedures. The outcomes of Bonferroni correction for the number of time-frequency bins and electrodes (Bonf.) and of uncorrected testing (uncorr.) at different statistical thresholds (thr.) are summarized for parameters (par.) FRQ (lemma frequency extracted from the linguistic corpus FOLK), EoA (ease-of-articulation index, Ziegler and Aichert 2015), NoS (number of syllables in a spoken word), CVR (consonant-to-vowel-ratio, i.e., the number of consonants in a spoken word divided by the number of vowels). "yes": a test at the respective threshold elicited significant results for at least one electrode, the number of electrodes with significant effects/the total number of tested electrodes are given in brackets. The positive outcomes of the most conservative statistical test (Bonferroni correction at q<0.05 or uncorrected testing at the most conservative threshold per parameter per subject in the absence of Bonferroni-corrected effects) are highlighted by a green background. "no": a test at the respective threshold elicited no significant results. Other conventions as in Tab. 1. The data in this table and in Tab. 8 have been re-analysed and partially reproduced from Lau (2016) with his permission.

s.	**test**	**thr.**	**par.**			
			CVR	**EoA**	**FRQ**	**NoS**
S1	Bonf.	0.05	no	no	yes (1/56)	no
	unc.	5E-06	no	yes (3/56)	yes (4/56)	yes (2/56)
	unc.	1E-06	no	no	yes (1/56)	no
	unc.	5E-07	no	no	yes (1/56)	no
	unc.	1E-07	no	no	no	no
S2	Bonf.	0.05	yes (4/59)	yes (1/59)	no	no
	unc.	5E-06	yes (13/59)	yes (12/59)	yes (6/59)	yes (2/59)
	unc.	1E-06	yes (8/59)	yes (3/59)	no	yes (1/59)
	unc.	5E-07	yes (7/59)	yes (1/59)	no	yes (1/59)
	unc.	1E-07	yes (4/59)	no	no	no
	unc.	5E-08	yes (3/59)	no	no	no
	unc.	1E-08	no	no	no	no
S3	Bonf.	0.05	yes (1/40)	no	no	yes (1/40)
	unc.	5E-06	yes (5/40)	no	yes (6/40)	yes (2/40)
	unc.	1E-06	yes (1/40)	no	yes (3/40)	yes (1/40)
	unc.	5E-07	yes (1/40)	no	no	yes (1/40)
	unc.	1E-07	no	no	no	yes (1/40)
	unc.	5E-08	no	no	no	yes (1/40)
	unc.	1E-08	no	no	no	no
S4	Bonf.	0.05	no	no	no	no
	unc.	5E-06	no	yes (1/55)	yes (16/55)	yes (4/55)
	unc.	1E-06	no	no	yes (4/55)	yes (3/55)
	unc.	5E-07	no	no	yes (1/55)	yes (1/55)
	unc.	1E-07	no	no	no	no
S5	Bonf.	0.05	no	no	no	no
	unc.	5E-06	yes (2/62)	no	yes (1/62)	no
	unc.	1E-06	no	no	no	no

Table 7: The results of testing correlations between RSM responses and the linguistic parameters with the most conservative statistical test and threshold before residualization. Abbreviations: ele.: electrode name, provided for ease of spatial refence (cf. Fig. 11); frq. (Hz): frequency of the significant effect in Hz, frq. r.: frequency range of the significant effect, β: beta (15-30 Hz), Lγ: low gamma (35-45 Hz). Hγ: high gamma (50-150 Hz), the effects in gamma frequencies are in black font (visualized in Fig. 11), the effects in lower frequencies are in grey font (not visualized); corr. pref.: correlation prefix, neg.: negative, pos.: positive; time rel. to w.: time of the effect relative to word production, bef.: before, dur.: during, aft.: after; MNI (x/y/z): coordinates of the electrode with a significant effect; str. area: structural area of the effect; funct. area (mon. ESM)/(bip. ESM): functional area of the effect identified with either monopolar or bipolar ESM; n./s.: electrode was not stimulated, n./.e.: electrode was stimulated but elicited no observable effect, ch.: chin, f.: finger, he.: head, l.: lip, t.: tongue, mot.: a motor response (a movement of the corresponding body part), sens.: a sensory response (a tactile sensation), speech: ESM-identified language-essential cortical site. °: a cortical site outside of the ESM-identified potentially speech-relevant cortex (Fig. 1) which lay in its immediate neighbourhood; ^: a cortical site outside of the ESM-identified potentially speech-relevant cortex (Fig. 1) without ESM-identified potentially speech-relevant cortical sites in its immediate neighbourhood; overl. eff.: a significant effect of at least one other parameter occurred in the same time-frequency range (see Methods for a definition) when using the same statistical test and threshold. EMG: average relative spectral magnitude for electromyographic activity during word production at the subjects' left cheeks, DELL: EMG from the subject's left deltoid muscle, int.: intensity of the acoustic signal, ss_ws: temporal duration from speech start to word start in ms, ws_we: temporal duration from word start to word end in ms, we_se: temporal duration from word end to speech end in ms, abbreviation for the other parameters are the same as in Tab. 6; "eff. surv. in postresid" indicates whether ("yes") or not ("no") the effect survived when repeating the same analysis after residualization (Tab. 9), i.e., by correlating the RSM values with the residuals of the linear model predicting the parameter by all other parameters with which it had significant correlations; "=thr.": effect upon residualization took place at the same electrode and in the same time-frequency range when using the same test and statistical threshold (Fig. 13), "↑thr.": effect upon residualization took place at the same electrode and in the same time-frequency range when using the same test at a more conservative threshold, **: a post-residualization effect took place at the same electrode but in a different frequency range.

s.	par.	test	thr.	ele.	frq. (Hz)	frq. r.	corr. pref.	time rel. to	MNI (x/y/z)	str. area	funct. area (mon. ESM)	funct. area (bip. ESM)	overl. eff.	eff. surv. in postresid.
S1	**EoA**	unc.	5E-06	C1	40-	Lγ	neg.	bef.	-62/-55/27	IP	n./s.	speech	#	no
				D5	45	Lγ	neg.	dur.	-60/-15/44	SI	l. mot.	l. & he. mot.	NoS	no
				D6	75	Hγ	neg.	dur.	-58/-4/44	PM	l. & t. mot.	l. & he. mot.	#	no
	FRQ	Bonf.	0.05	E5	120-	Hγ	neg.	dur.	-54/-16/53	SI	thumb mot.	l. & he. mot.	#	yes (=thr.)
	NoS	unc.	5E-06	A5	45	Lγ	pos.	bef.	-66/-16/8	TC	n./s.	speech	no	no
				D5	45	Lγ	pos.	dur.	-60/-15/44	SI	l. mot.	l. & he. mot.	EoA	no
S2	**CVR**	Bonf.	0.05	B2	140	Hγ	pos.	bef.	-51/10/43	BR	n./e.	n./e.°	#	yes (=thr.)
				B3	135	Hγ	pos.	aft.	-51/-1/53	PM	ch. & t. mot.	ch. mot.	#	yes (=thr.)
				E4	125	Hγ	pos.	bef.	-67/-17/26	OP	n./e.	t. mot.	#	yes (=thr.)
				E5	115	Hγ	pos.	dur.	-66/-30/29	IP	n./e.	hand sens.°	#	yes (=thr.)
	EoA	Bonf.	0.05	C4	50	Hγ	neg.	aft.	-58/-14/47	SI	f. sens.	f. sens.^	#	no
	FRQ	unc.	5E-06	C3	40,	Lγ,	neg.	bef.	-58/-3/44	PM	ch. & l. mot.	hand sens.	no	no
				C7	30	β	neg.	bef.	-53/-44/54	IP	n./e.	thigh mot.^	#	no**
				E1	30	β	neg.	bef.	-63/13/15	BR	n./e.	t. mot.	#	no
				E5	40	Lγ	neg.	bef.	-66/-30/29	IP	n./e.	hand sens.°	no	no
				E7	50	Hγ	neg.	aft.	-63/-47/34	IP	n./e.	thigh mot.°	#	no
				H2	75	Hγ	neg.	dur.	-63/-6/-9	TC	n./s.	speech	#	no
	NoS	unc.	5E-07	B3	130	Hγ	pos.	bef.	-51/-1/53	PM	ch. & t. mot.	ch. mot.	ss_ws	no
S3	**CVR**	Bonf.	0.05	E2	145	Hγ	pos.	aft.	-65/-5/23	CS	n./s.	n./e.	EMG,	no
	FRQ	unc.	1E-06	F1	90	Hγ	pos.	aft.	-64/9/12	BR	n./s.	l. & t. mot.	no	no
	NoS	Bonf.	0.05	D7	75	Hγ	neg.	bef.	-58/-63/18	IP	n./s.	n./e.°	#	no
S4	**EoA**	unc.	5E-06	E8	40	Lγ	pos.	aft.	-56/-36/51	IP	f. sens.	f. sens.	EMG	no
	FRQ	unc.	5E-07	G8	130	Hγ	neg.	bef.	-62/-44/38	IP	n./s.	n./e.	EMG	no
	NoS	unc.	5E-07	E6	65	Hγ	pos.	bef.	-58/-18/46	SI	l. mot. &	l. & t. mot.	int.,	no
S5	**CVR**	unc.	5E-06	D8	25	β	neg.	bef.	-44/21/42	PF	n./s.	aura		yes (↑thr.)
				D8	75	Hγ	pos.	aft.	-44/21/42				DELL	yes (↑thr.)
				F6	120	Hγ	neg.	bef.	-39/-4/63	PM	hand mot.	hand mot.^	#	no
	FRQ	unc.	5E-06	B5	40	Lγ	pos.	dur.	-63/-5/24	SI	t. mot.	t. mot.	no	no

Tab. 7 provides a neuroanatomical description of the effects observed for each parameter using the most conservative test and threshold. It specifies the range of frequencies within which the effects took place, provides the MNI coordinates of the electrodes and gives functional descriptions of these electrodes obtained with the help of ESM. Fig. 11 visualizes the locations of these effects on a standard brain template from spm5. The ESM-based functional assignment of the electrodes with significant effects did not yield a comprehensive picture. Overall, however, the effects observed in relation to the four main linguistic parameters of our interest (i.e., CVR, EoA, FRQ, NoS) took place in areas roughly corresponding to speech and mouth-relevant regions (panel E in Fig. 11). In agreement with our expectation that higher levels of gamma activity would be associated with increased complexity of words, we observed that CVR mostly elicited positive correlations with high-gamma activity, EoA and FRQ yielded effects which were predominantly negative and NoS mostly showed positive effects. Against our expectation, however, the effects did not take place over broad ranges of gamma frequencies and over extended time periods (Tabs. 7, 9), in contrast to the correlation effects observed with the linguistically-unspecific speech-duration-related parameters (Fig. 10). There was little reproducibility with regard to the timing and to the exact frequential components of the spectrum, both within and between the linguistic parameters.

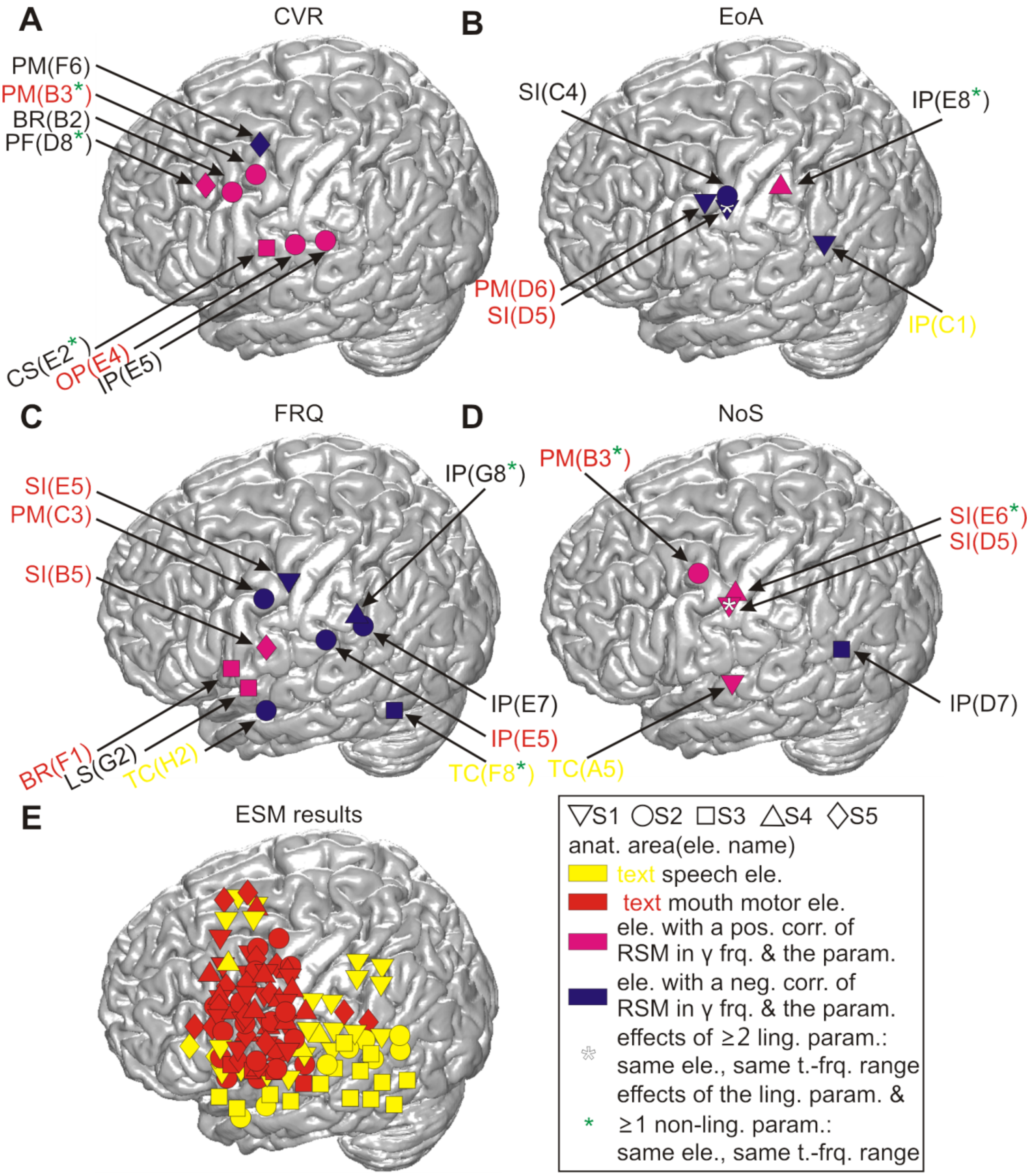


Figure 11: Correlations between RSM responses and the linguistic parameters before residualization in relation to cortical anatomy. All effects in gamma frequencies (in black font in Tab. 7) are visualized on the cortical surface based on their MNI coordinates for **A**: CVR, **B**: EoA, **C**: FRQ, and **D**: NoS. CS: central sulcus, LS: lateral sulcus, other abbreviations for the anatomical areas in which the effects were observed are the same as in Fig. 1, the electrode names in brackets next to the names of anatomical areas are provided for reference to Tab. 7. **E**: The anatomical locations of all potentially speech-relevant electrodes are visualized using the same procedure as in A-D. Abbreviations in the legend: anat. area(ele. name): anatomical area (electrode name), pos./neg. corr.: positive/negative correlation, γ frq.: gamma frequencies (40-150 Hz), param.: parameter, ling.: linguistic, t.-frq. range: time-frequency range; BR: Broca’s area, other abbreviations as in Fig. 1.

3.2.6 Statistical testing of neurocorrelation results after residualization

As has been mentioned above, the linguistic data we were analysing consisted of parameters which were often mutually correlated, e.g., EoA was negatively correlated with NoS in all subjects ($p < 0.05$, uncorrected). As a result, the outcome of the neurocorrelation analysis with an individual linguistic parameter might produce effects which are not specific to the given parameter but which partially reflect the fact that this parameter is correlated with another parameter and that this other parameter plays a role. In an attempt to be able to draw parameter-specific conclusions, we conducted a linear regression analysis and extracted its residuals, representing components of the linguistic data which were mutually orthogonal. Then, we repeated the same neurocorrelation analysis with the residuals and compared the resulting neural effects with those observed prior to residualization.

Three scenarios could be observed as to how the neurocorrelation effects prior to and after residualization behaved in relation to each other (Fig. 12):

(1) there was an effect prior to residualization which did not survive in post-residualization analysis;
(2) there was an effect prior to residualization which survived in post-residualization analysis at the same or at a less conservative statistical threshold;
(3) there was a new effect after residualization which was not there prior to it.

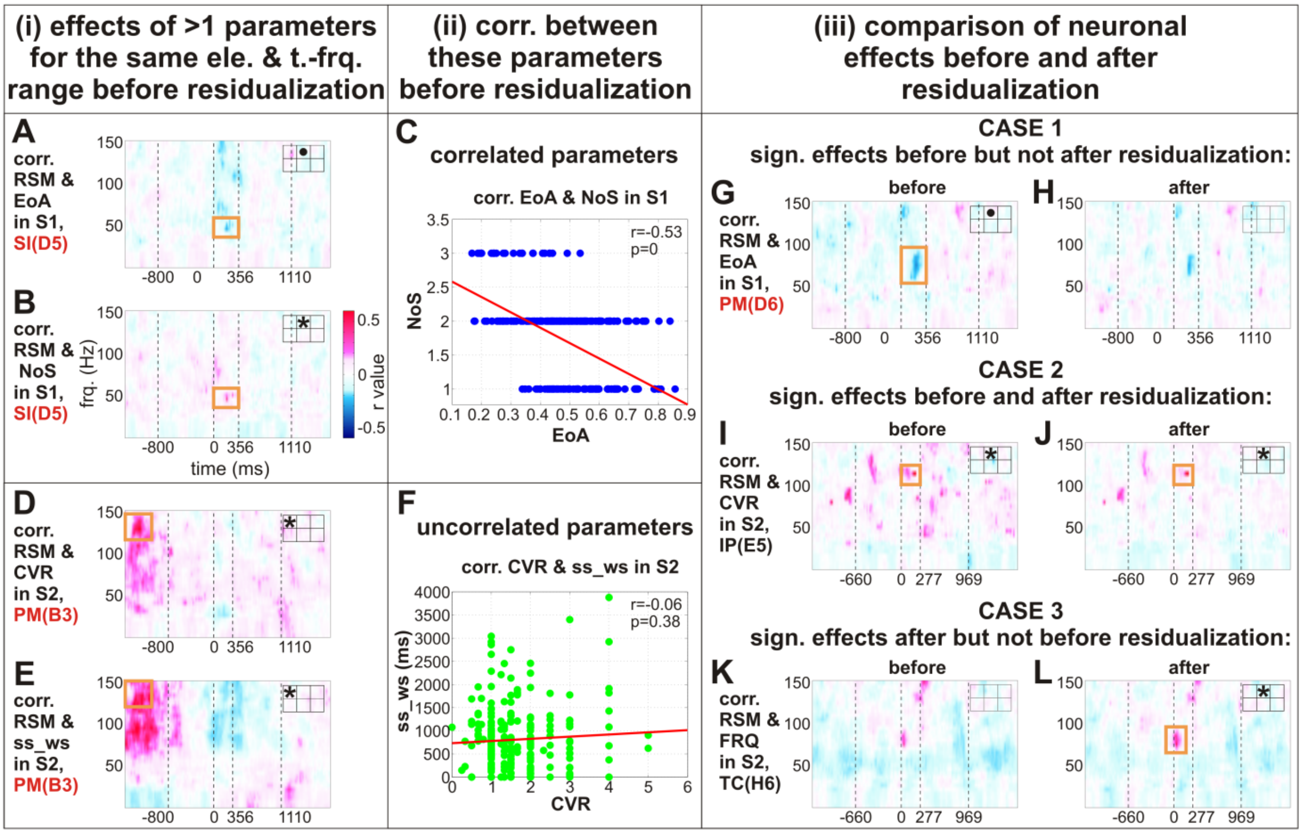


Figure 12: Examples of RSM correlations with word-describing parameters (i) in relation to the collinearity structure between these parameters (ii) and (iii) a summary of the three impacts of removing multicollinearity in the linguistic data on the structure RSM correlations. Abbreviations and conventions: ele. & t.-frq. range: electrode and time-frequency range, >1 more than one, corr.: correlation, sign.: significant, orange frames circumscribe the time-frequency ranges in which the effects were significant at the displayed electrodes in the respective conditions, other abbreviations and conventions as in previous figures. (1) shows examples of significant neural effects which took place at the same mouth motor electrode (red font) and in the same time-frequency range (see Methods) with either different (**A, B**) or the same (**D, E**) correlation prefixes. These parameters either showed a significant correlation with each other (**C**) which could explain the different prefixes when correlating RSM values with these parameters (cf. A, B), or they were not significantly correlated (**F**) and thus could not explain the presence of significant RSM correlations with these parameters in the same time-frequency range at the same electrode (D, E, see Methods). A comparison of correlating RSM values with the linguistic parameters before residualization (**G, I, K**) with the outcomes of correlating the residuals (**H, J, L**) of the linear model predicting the respective parameter by all parameters with which it was significantly correlated in the respective subject resulted in changes post- compared to pre-residualization. These are listed and illustrated in (1) as cases 1-3.

The occurrence of these different options likely has several reasons. We assume that (1) may be due to the fact that mutually-correlated parameters bear meaningful information which cannot be disentangled into individual, parameter-specific components without its loss (Case 1 in Fig. 12), (2) is likely an indication of the robustness of the effect and its relative independence on the presence of collinearity in the linguistic data (Case 2 in Fig. 12), and (3) can reflect the fact that residualization has removed some noise in the linguistic data, making hitherto concealed effects more clearly visible (Case 3 in Fig. 12).

Table 8: An overview of the outcomes when testing correlations between RSM responses and the linguistic parameters after residualization using different statistical procedures. All conventions as in Tab. 6.

s.	**test**	**thr.**	**par.**			
			CVR	**EoA**	**FRQ**	**NoS**
S1	Bonf.	0.05	no	no	yes (1/56)	no
	unc.	5E-06	yes (1/56)	no	yes (2/56)	yes (1/56)
	unc.	1E-06	no	no	yes (1/56)	no
	unc.	5E-07	no	no	yes (1/56)	no
	unc.	1E-07	no	no	no	no
S2	Bonf.	0.05	yes (6/59)	no	no	no
	unc.	5E-06	yes (13/59)	no	yes (3/59)	no
	unc.	1E-06	yes (9/59)	no	no	no
	unc.	5E-07	yes (8/59)	no	no	no
	unc.	1E-07	yes (4/59)	no	no	no
	unc.	5E-08	yes (4/59)	no	no	no
	unc.	1E-08	yes (2/59)	no	no	no
	unc.	5E-09	yes (1/59)	no	no	no
	unc.	1E-09	yes (1/59)	no	no	no
	unc.	5E-10	yes (1/59)	no	no	no
	unc.	1E-10	no	no	no	no
S3	Bonf.	0.05	yes (1/40)	no	no	no
	unc.	5E-06	yes (4/40)	no	no	no
	unc.	1E-06	yes (2/40)	no	no	no
	unc.	5E-07	yes (1/40)	no	no	no
	unc.	1E-07	yes (1/40)	no	no	no
	unc.	5E-08	yes (1/40)	no	no	no
	unc.	1E-08	no	no	no	no
S4	Bonf.	0.05	no	no	no	no
	unc.	5E-06	yes (1/55)	no	yes (8/55)	no
	unc.	1E-06	no	no	yes (1/55)	no
	unc.	5E-07	no	no	no	no
S5	Bonf.	0.05	no	no	no	no
	unc.	5E-06	yes (2/62)	no	yes (2/62)	no
	unc.	1E-06	yes (1/62)	no	yes (1/62)	no
	unc.	5E-07	no	no	no	no

Tab. 8 presents the results of the same statistical testing procedure as applied on the neurocorrelation results obtained with pre-residualized data but after residualization. As can be seen from this table, EoA did not survive statistical testing with our specified tests and thresholds, NoS showed a significant effect in only one subject, FRQ yielded fewer effects but was still significant in three out of five subjects, and CVR showed significant results in all tested individuals. Tab. 9 provides additional information about the locations of the observed neural effects together with their anatomical and functional descriptions.

Table 9: The results of testing correlations between RSM responses and the investigated linguistic parameters with the most conservative statistical test and threshold (after residualization). Abbreviations: α: alpha (5-10 Hz), eff. rel. to preresid: this column indicates whether the effect had been there already in the correlation analysis before residualization ("old," Tab. 7) or if it was taking place at a different electrode and/or a different time-frequency range ("new"), **: an effect before residualization took place at the same electrode but in a different frequency range; mid. f.: middle finger, aura: an ESM-induced feeling preceding an epileptic seizure, other conventions as in Tab. 7. The effects in gamma frequencies (in black font) are visualized for the linguistic parameters CVR and FRQ in Fig. 13.

s.	par.	test	thr.	ele.	frq. (Hz)	frq. r.	corr. pref.	time rel. to w.	MNI (x/y/z)	str. area	funct. area (mon. ESM)	funct. area (bip. ESM)	overl. eff.	eff. rel. to preresid.
S1	**CVR**	unc.	5E-06	H1	120	Hγ	pos.	bef.	-28/-55/71	SP	hand mot.	arm mot.^	no	new
	FRQ	Bonf.	0.05	E5	110	Hγ	neg.	dur.	-54/-16/53	SI	thumb mot.	he. & l. mot.	no	old (=thr.)
	NoS	unc.	5E-06	F3	70	Hγ	pos.	bef.	-49/-36/61	SI	little f. mot.	little & ring f. mot.°	no	new
S2	**CVR**	Bonf.	0.05	A8	150	Hγ	pos.	aft.	-29/-45/71	SI	n./e.	arm sens.-mot.^	no	new
				B2	140	Hγ	pos.	bef.	-51/10/43	PF	n./e.	n./e.	no	old (=thr.)
				B3	135	Hγ	pos.	aft.	-51/-1/53	PM	ch. & t. mot.	ch. mot.	no	old (=thr.)
				E4	125	Hγ	pos.	bef.	-67/-17/26	OP	n./e.	t. mot.	no	old (=thr.)
				E5	115	Hγ	pos.	dur.	-66/-30/29	IP	n./e.	hand sens.°	no	old (=thr.)
				F6	105	Hγ	pos.	aft.	-65/-41/22	TC	n./e.	speech	no	new
	FRQ	unc.	5E-06	C7	100	Hγ	neg.	aft.	-53/-44/54	IP	n./e.	thigh mot.	no	new**
				G3	145	Hγ	pos.	bef.	-68/-15/1	TC	n./s.	n./e.°	no	new
				H6	75	Hγ	pos.	dur.	-68/-44/-2	TC	n./s.	n./e.°	no	new
S3	**CVR**	Bonf.	0.05	D3	145	Hγ	pos.	aft.	-66/-16/28	OP	n./s.	n./e.^	no	new
S4	**CVR**	unc.	5E-06	F7	45	Lγ	neg.	bef.	-61/-32/43	IP	n./e.	n./e.°	EMG, int.	new
	FRQ	unc.	1E-06	B8	5-10	α	pos.	aft.	-30/-22/73	PM	index & mid. f. mot.	hand mot.^	EMG	new
S5	**CVR**	unc.	1E-06	D8	25	β	neg.	bef.	-44/21/42	PF	n./s.	aura°	no	old (↑thr.)
				D8	75	Hγ	pos.	aft.	-44/21/42	PF	n./s.	aura°	no	old (↑thr.)
	FRQ	unc.	1E-06	C6	35	Lγ	neg.	bef.	-62/5/34	PM	t. mot.	t. & l. mot.	no	new

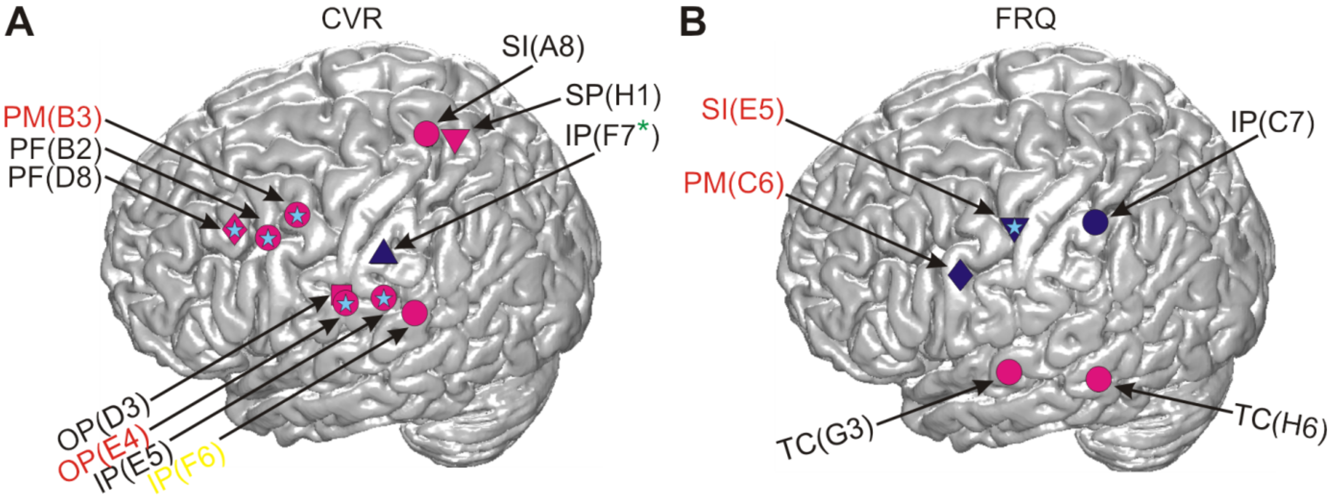


Figure 13: Correlations between RSM responses and the linguistic parameters CVR and FRQ after residualization in relation to cortical anatomy. The light blue stars indicate that the same effect was present in the neural data also prior to residualization and the absence of such a star indicates the contrary. All other conventions as in Fig. 11.

Fig. 13 shows the locations of the results for CVR and FRQ we observed in post-residualization neurocorrelation analyses, described in more detail in Tab. 9. As one can see from this figure, the neurocorrelation effects related to CVR survived residualization and took place at the same time and frequency ranges and at the same electrodes and with the same prefixes in two subjects. Two major anatomical sites at which these effects took place can be distinguished: the border of the prefrontal and premotor cortex and the parietal operculum converging on adjacent other parts of the inferior parietal cortex (Fig. 13A). This anatomical picture is very similar to the one observed prior to residualization (cf. Fig. 11). Several other effects are new and it is hard for us to interpret their location and functional significance due to the relatively small sample of investigated subjects. The effects of FRQ in the fronto-parietal cortex were less conclusive: only one electrode in the primary somatosensory cortex of one subject showed an effect at the same time and frequency range and with the same correlation prefix. The correlation effects related to this parameter prior to residualization mostly occurred in the fronto-parietal cortex with a negative correlation prefix (Fig. 11C). This was also the case with post-residualized data in these anatomical areas (Fig. 13B). In addition to these effects, positive correlations with high-gamma frequencies which had not been visible in the pre-residualized data could be observed in one subject. Due to a relatively small sample of subjects, we shall refrain from functional interpretations of this effect.

3.3 Results of Study 3: syntax

In the study on the syntactic properties of the language material, we evaluated data from four subjects (Tab. 10). The degree to what the data have been analysed, however, differs between the subjects due to pragmatic reasons, and only preliminary results of neural analyses which may require further validation will hence be presented.

Table 10: The numbers of simple clauses (N. cl.) evaluated per subject (s.).

s.	S4	S5	S6	S7	total
N. cl.	195	441	276	242	1154

3.3.1 PoS analysis

For each simple clause, we annotated the PoS according to the STTS conventions yet with slight modifications as described in Methods and online (Neuromedical AI Lab 2019). After this analysis was completed and checked, we wanted to know, how representative our data were of the spoken German language annotated elsewhere. For this reason, we compared the number of the different PoS with those in the corpus of spoken Alemannic (ALCORP, 2013). Due to the aforementioned slight deviations from the STTS in our data, we have evaluated both data sets on a level of abstraction at which these differences would no longer be relevant. This was achieved by grouping the distinctive PoS into overarching categories (Fig. 14).

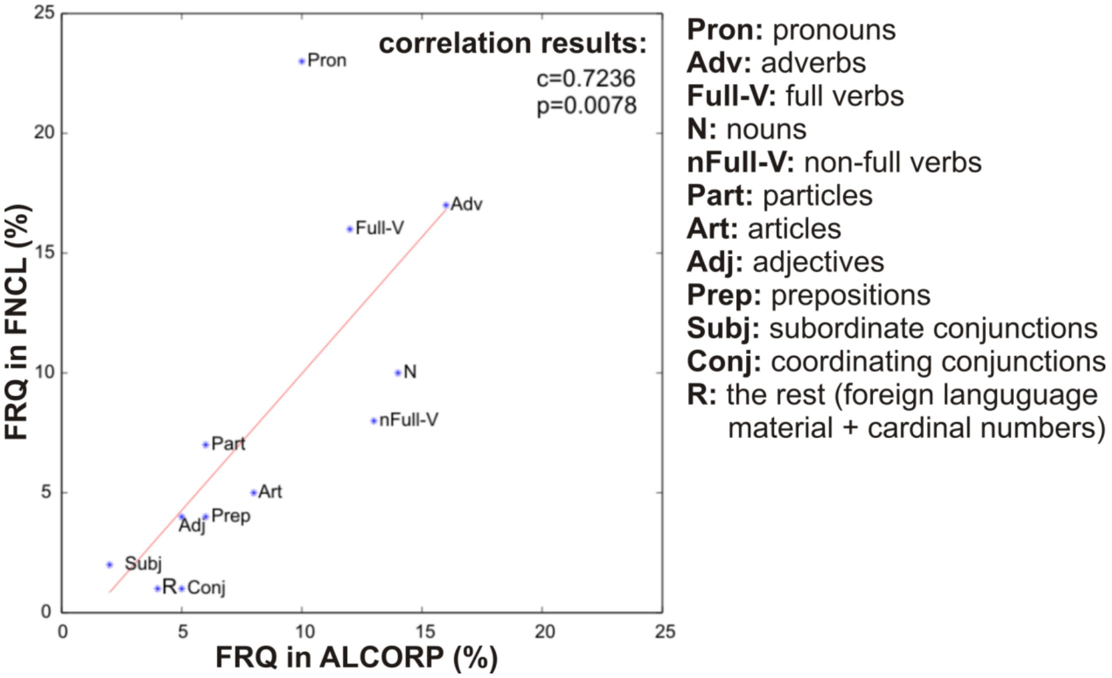


Figure 14: Correlations between the numbers of the different PoS in FNLC and in ALCORP. Correlation results using a Pearson’s correlation are shown.

As can be seen from Fig. 14, our PoS analysis resulted in values close to those in ALCORP. The proportional relations between the different PoS categories were reproducible between the corpora and the numbers of PoS of the same category showed a strong and highly significant positive correlation. Adverbs, verbs, nouns and pronouns represented the vast majority of the words in both corpora. One difference between the corpora which needs to be mentioned, however, is that there were more pronouns in our data than in ALCORP. This may relate to the fact that our subjects were mostly talking to close friends or relatives and that they were thus perhaps more ready to talk about personal topics, such as their clinical situation, while the data in ALCORP were collected from subjects who were talking to an external interviewer. The overall reproducibility of the numbers between both corpora indicates that our data are capable of representing the spoken German language with regard to their PoS composition. We also calculated the frequencies with which the different PoS constellations occurred within our samples. This information was used for further correlation-based analyses described below.

3.3.2 Sentence constituent analysis

After annotation and correction of the sentence constituent analysis (see Methods and Neuromedical AI Lab (2019, online)), we were interested to establish, whether the sentence constituent constellations would show similarities between speakers in terms of the frequency of their occurrence. Since we are not aware of a corpus of spoken German which would contain sentence constituent analyses, we were not able to perform a similar analysis as in the case of the PoS data, unfortunately. We calculated, how many times within the subjects' subcorpora a distinctive PoS constellation occurred and sorted these frequencies in descending order. Tab. 11 shows two most and two least frequent constellations for each subject.

Table 11: Examples of most and least frequent sentence constituent constellations in our data. "s": subject, "frq": the frequency of syntactic constellations in the subcorpus of the individual subject,"vf": finite verb," "pv": predicative, "adv": adverbial, "vnf": non-finite verb," "io": infinitive object, "do": dative object, "po": prepositional object.

s.	frequent	frq (% per s.)	infrequent	frq (% per s.)
S4	s vf pv s vf adv	7.6 5.9	s vf io vnf pv vf s	0.6 0.6
S5	s vf pv s vf adv pv	9.4 8.6	s vf adv do adv s vf adv do	0.7 0.7
S6	s vf pv s vf adv	12.1 5.0	do vf s adv io vf s adv	0.3 0.3
S7	s vf pv s vf adv pv	6.6 5.3	po vf s adv adv vf s io do vnf	0.7 0.7

Indeed, we were able to observe reproducible patterns in terms of most frequent constellations of sentence constituents. In all subjects, the constellation "s vf pv" (subject → finite verb → predicative, e.g., "Er ist da.", Engl. "He is here.") was the most common. It was followed either by a constellation "s vf adv" (subject → finite verb → adverbial, e.g., "Er kommt bald.", Engl. "He is coming soon.") or by "s vf adv pv" (subject → finite verb → adverb → predicative, e.g., "Er ist bald da.", Engl. "He will be here soon."). Infrequent constellations differed in terms of their syntactic composition and, unlike the frequent ones, involved different kinds of objects and sometimes had the subject in a more rightward position.

3.3.3 Syntactic hierarchy analysis

Our analysis of the number of hierarchy levels showed that, unsurprisingly, most words in a clause were dependent on the full verb (ca. 75 %). The rest 35% were dependent on other words within the clause, and the number of hierarchy levels ranged from 1 (no dependent lexical elements) to 7 (6 dependent lexical elements) in our data. A detailed description of these relations is available in the doctoral thesis by Diekmann (2019), so we shall not go into further detail.

3.3.4 Correlation structure between linguistic parameters

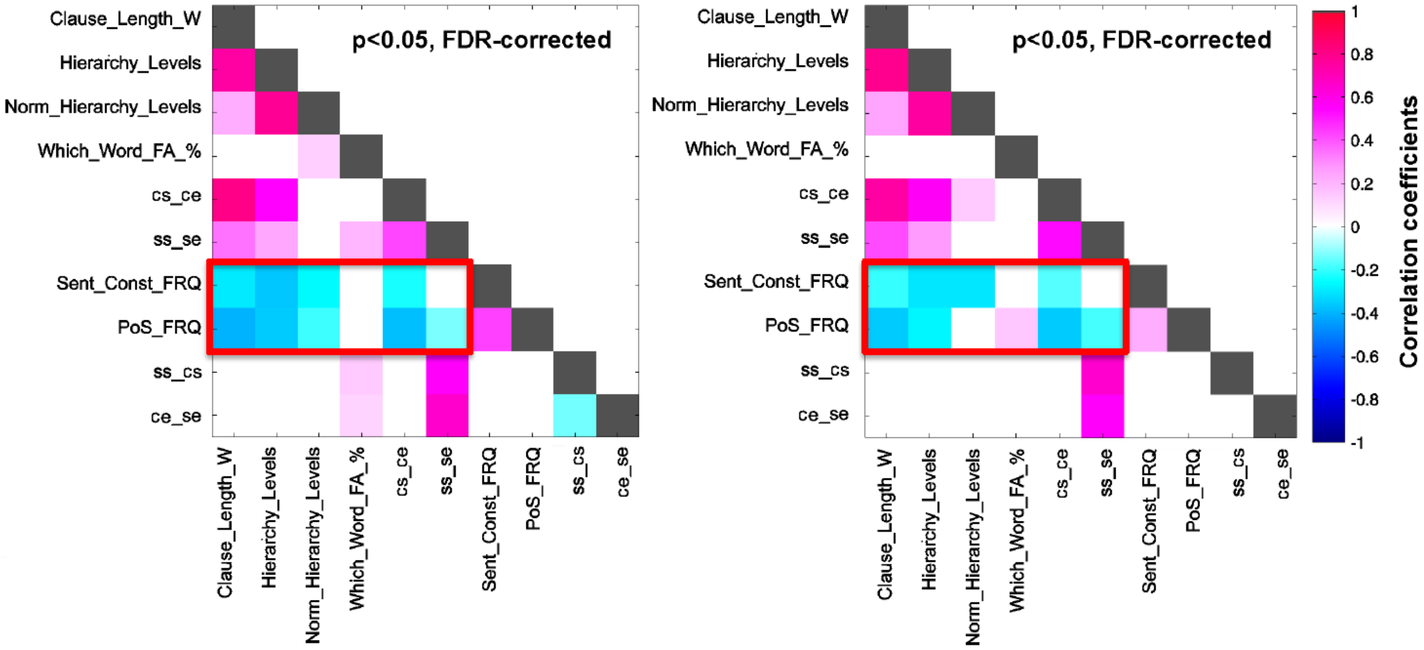


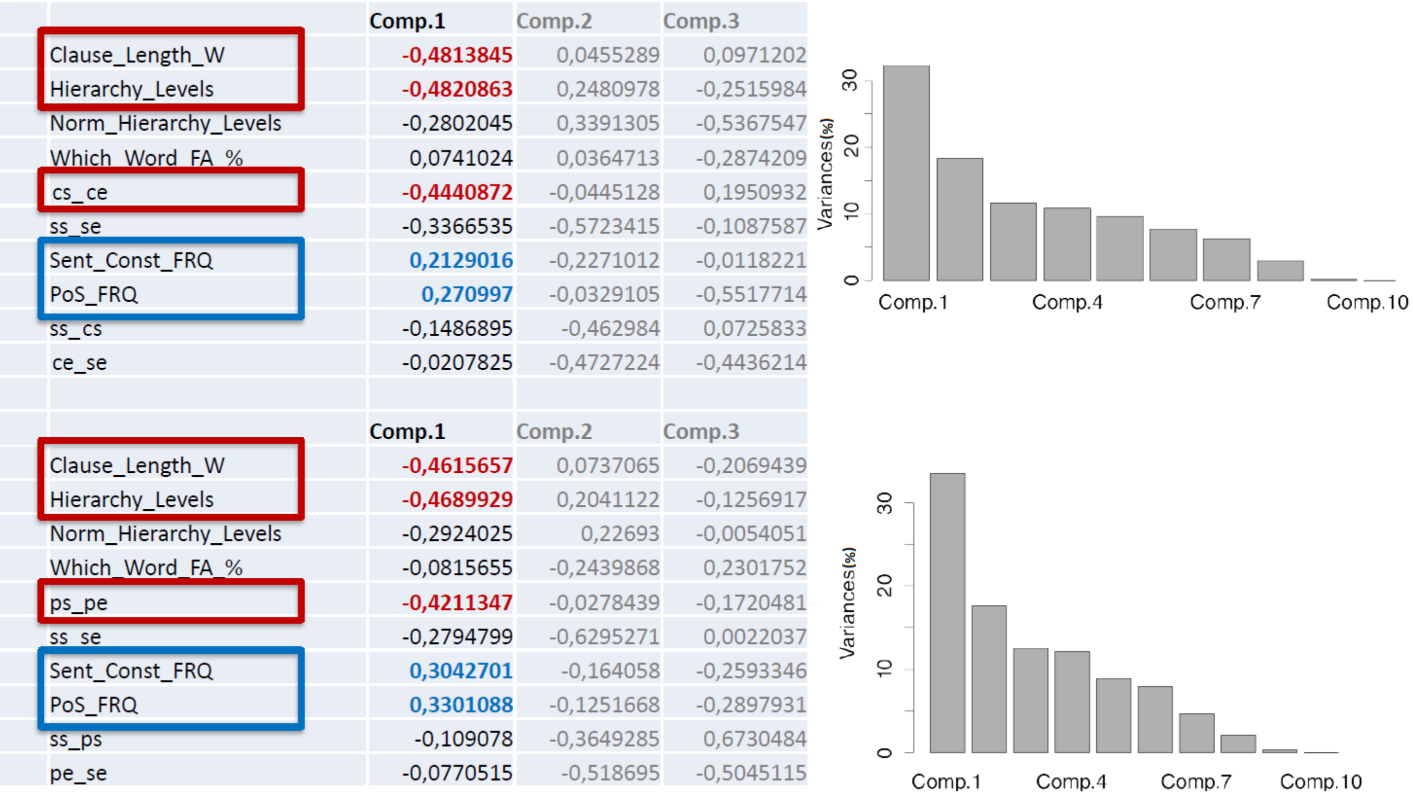
Figure 15: Typical correlation structure between the acquired linguistic parameters shown on the examples of two datasets. The results of a Spearman's correlation are shown, the correlation coefficients are colour-coded (see legend). Only significant correlations ($p<0.05$, FDR-corrected for the number of correlation analyses within the subject) are shown, the rest are in white colour. "Clause_Length_W": clause length measured as the number of words in a clause, "Hierarchy_Levels": the number of syntactic hierarchy levels within the clause, "Norm_Hierarchy_Levels": the number of syntactic hierarchy levels within the clause normalized (divided) by the number of words within a clause, "Which_Word_FA_%": the position of the word carrying the focus accent in relation to the total length of the clause in words, "cs_ce": clause duration from clause start to end in ms, ss_se: the duration of the speech production epoch embedding the clause in ms, "Sent_Const_FRQ": the frequency of sentence constituent constellation within the given subject, expressed in % in relation to the total number of clauses in this subject, "PoS_FRQ": the frequency of the PoS constellations, "ss_cs": the duration from speech start to clause start in ms, "ce_se": the duration from clause end to speech end in ms.

As is exemplified in Fig. 15, we observed reproducible correlation structures in our linguistic data. The number of words was strongly and positively correlated with clause duration in ms and with the number of syntactic hierarchy levels in a clause (this effect persisted also after normalization of this parameter by clause duration). Strong negative correlations were observed between the aforementioned parameters with the frequency of PoS- and sentence constituent constellations. This finding shows that Zipf's law (1935), postulating a negative correlation between word length and frequency, also extends to larger linguistic units such as simple clauses.

An observation consequential for the present work was that, as it was the case with the word-related parameters, the parameters describing our simple clauses were also strongly and mutually correlated. Thus, it was an important question for us, whether one could extract mutually orthogonal parameters which would show significant, meaningful correlations with the neural data and how to deal with the problem of collinearity otherwise.

3.3.5 PCA

All linguistic parameters listed in Fig. 15 were used in this analysis, which was performed using a dedicated function in the programming environment R.



	Comp.1	Comp.2	Comp.3
Clause_Length_W	**-0,4813845**	0,0455289	0,0971202
Hierarchy_Levels	**-0,4820863**	0,2480978	-0,2515984
Norm_Hierarchy_Levels	-0,2802045	0,3391305	-0,5367547
Which_Word_FA_%	0,0741024	0,0364713	-0,2874209
cs_ce	**-0,4440872**	-0,0445128	0,1950932
ss_se	-0,3366535	-0,5723415	-0,1087587
Sent_Const_FRQ	**0,2129016**	-0,2271012	-0,0118221
PoS_FRQ	**0,270997**	-0,0329105	-0,5517714
ss_cs	-0,1486895	-0,462984	0,0725833
ce_se	-0,0207825	-0,4727224	-0,4436214

	Comp.1	Comp.2	Comp.3
Clause_Length_W	**-0,4615657**	0,0737065	-0,2069439
Hierarchy_Levels	**-0,4689929**	0,2041122	-0,1256917
Norm_Hierarchy_Levels	-0,2924025	0,22693	-0,0054051
Which_Word_FA_%	-0,0815655	-0,2439868	0,2301752
ps_pe	**-0,4211347**	-0,0278439	-0,1720481
ss_se	-0,2794799	-0,6295271	0,0022037
Sent_Const_FRQ	**0,3042701**	-0,164058	-0,2593346
PoS_FRQ	**0,3301088**	-0,1251668	-0,2897931
ss_ps	-0,109078	-0,3649285	0,6730484
pe_se	-0,0770515	-0,518695	-0,5045115

Figure 16: Typical three main principal component analysis (PCA) components explaining over 70% of the variances in the data (left) and the extent to what the distinctive components were informative (right). The data are shown on the example of the same subjects whose linguistic data are visualized in Fig. 15. Comp.: components, variances: the percentage of the variances in the data which can be explained by the respective component. Colour coding shows the most dominant parameters in the first principal component with either a negative (red) or a positive (blue) prefix.

The PCA analysis we carried out on the linguistic data yielded a reproducible first principal component (pc) which explained over 35% of the variances in all subjects (Fig. 16). In this component, sentence length and complexity-related parameters (clause duration, clause length, and the number of syntactic hierarchy levels) had loadings which carried a contrary prefix to that of the frequency-related parameters (i.e., the frequency of sentence constituent or PoS constellations). In this sense, this parameter seems to replicate the findings presented in Fig. 15, which indicate an inverse relationship between these parameters. Another fact which needs to be mentioned is that the position of the FA did not contribute to the explanation of the variances in the first pc, suggesting that the FA played a modest role in our linguistic data. The second pc was dominated by the parameter ss_se, and the other pcs varied from subject to subject in terms of the major contributing parameters.

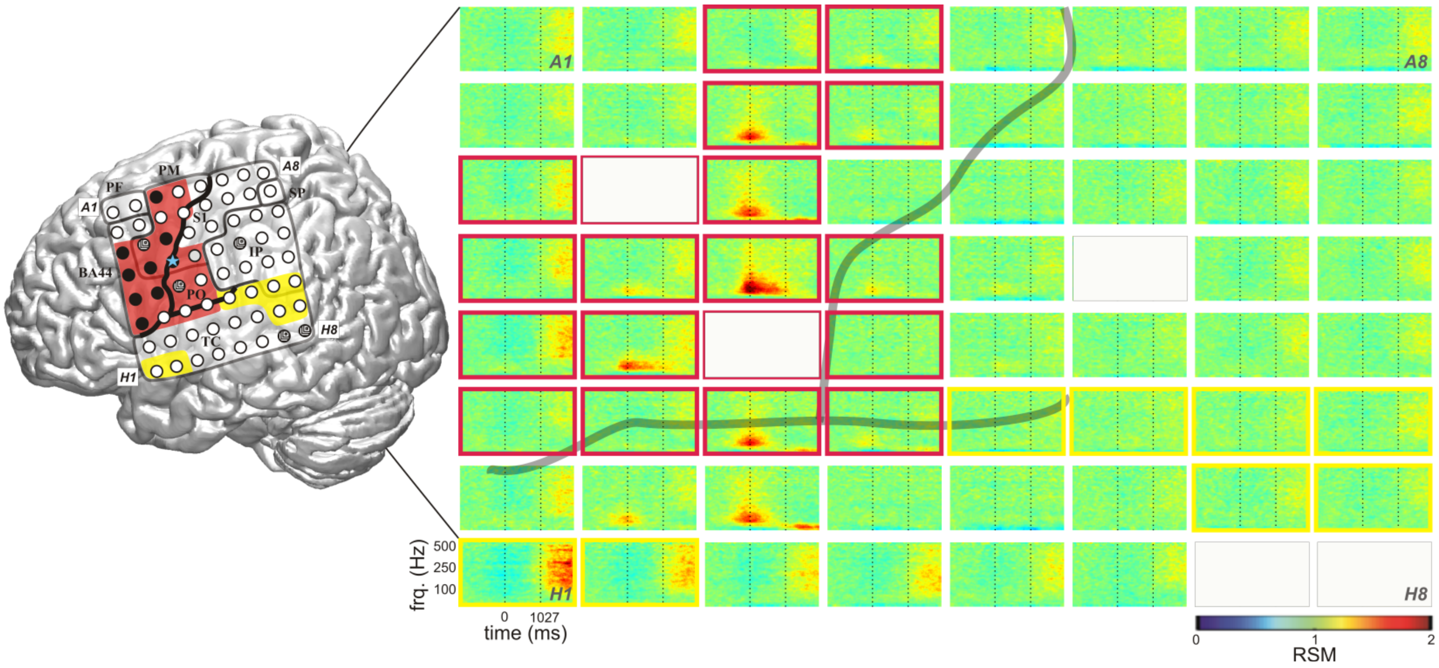


Figure 17: Trial-averaged, time-resolved RSM changes of clause-production-related cortical activity in S2. Left panel: the individual location of the 8×8 electrode grid and the borders of anatomical areas in which the electrodes were located, conventions as in Fig. 1. Right panel: colour-coded RSM changes observed in relation to clause production (see legend). The positions of the individual electrodes correspond to those in the left panel. The vertical dashed lines in the spectrum of each electrode show the average start (0 ms) and end (1027 ms) of the simple clauses in this subject. Other conventions as in Fig. 4.

As is visualized on the example of S2 in Fig. 17, cortical activity changes related to clause production were, similarly to activity we observed in relation to word production, manifested in increased high-gamma RSM changes (colour-coded in red). These mostly took place in mouth motor areas located in the anatomical premotor cortex and in the posterior part of BA 44 or adjacent to them (e.g., on the lateral sulcus and at one electrode in the superior temporal cortex next to it, see Fig. 17), started prior to clause production and ended mostly prior to clause end or around it. We observed no effects which would take place exclusively prior to clause production.

3.3.6 Neurocorrelation results

Like in the study on word complexity, we performed correlations of the RSM values with the individual linguistic parameters prior to and after residualization. The same procedure as in the aforementioned study was applied. As it was the case in Study 2, the neurocorrelation analysis also showed most prominent correlations with duration-related parameters at mouth motor electrodes (not visualized). The results of this analysis with regard to the linguistic parameters were rather inconclusive, i.e., they failed to elicit spatio-frequentially extended effects (with a notable exception in the case of the number of syntactic hierarchy levels in the medial temporal lobe of S7, solely in whom this area had electrode coverage, see Fig. 18). We therefore performed a PCA on the linguistic materials to extract most informative aspects of the acquired linguistic data for correlation of the obtained pcs with the RSM values.

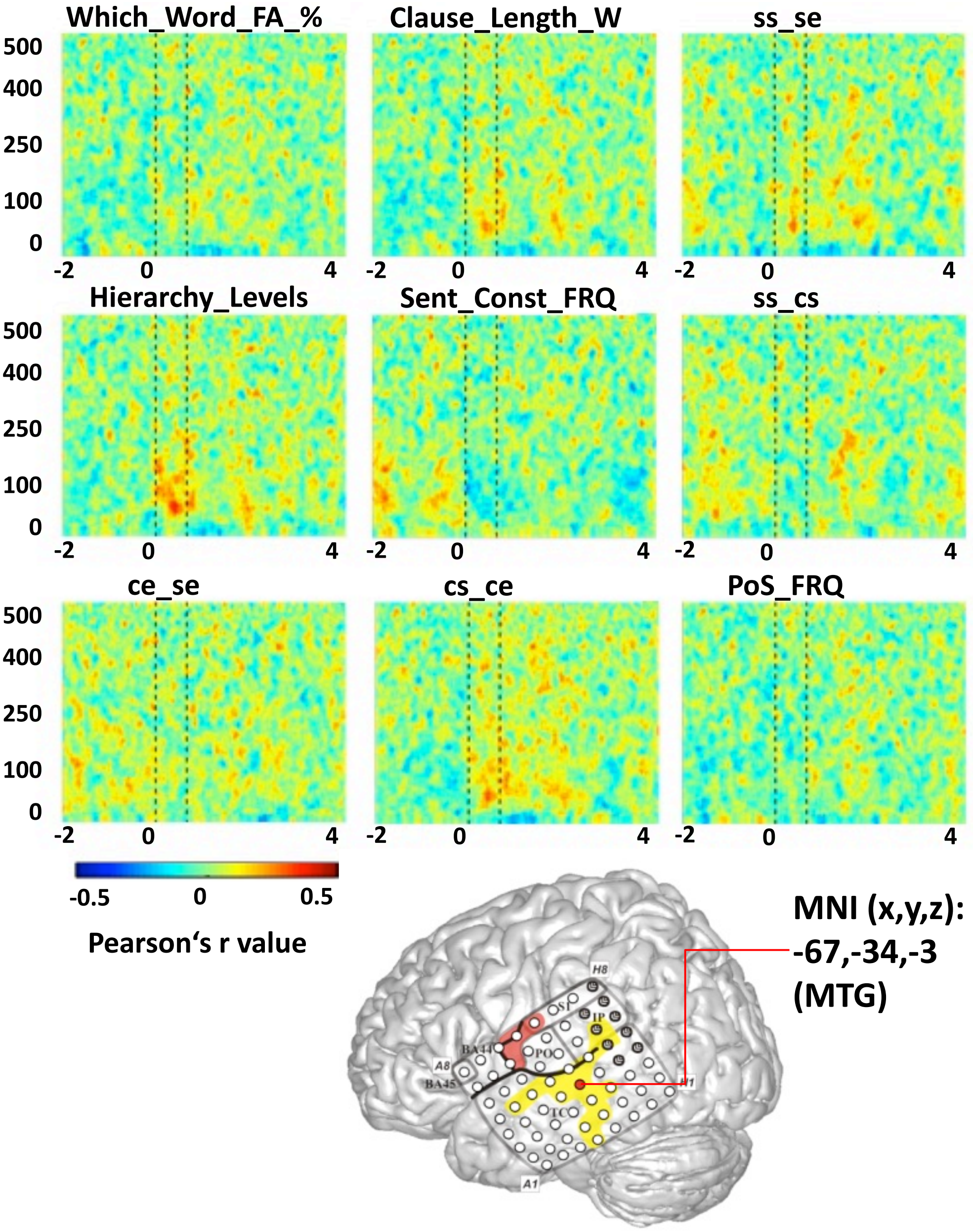


Figure 18: Correlations with the different clause-related parameters in S7 (Electrode E4). y axis: frequency (Hz), x axis: time (ms), “0”: clause start, also indicated by the first dashed line. The second dashed line indicates the average clause end in this subject. Correlation results are colour-coded (see legend). The electrode for which the results are shown is indicated below the neurocorrelation results by a red circle on the standard brain surface (see Methods). The depicted electrode was the only one to show significant correlation results (p<5E-06, uncorrected) with linguistic parameters. It lay in the medial temporal gyrus (MTG) according to the anatomy toolbox of spm5 (v17).

As is shown on a typical example in Fig. 18, the only linguistic parameter which showed meaningful correlations with high-gamma activity was the number of hierarchy levels. This effect was significant at p<5E-06 (uncorrected) and it occurred at one electrode lying in the medial temporal lobe of S7 only. The significant correlation took place over the duration of clause production (left panel of the second row in Fig. 18). When we repeated the same neurocorrelation analysis with the residuals of the models in which we predicted each parameter by all others (Fig. 18), this significant correlation was gone, unfortunately.

3.3.7 Neurocorrelation with the pcs

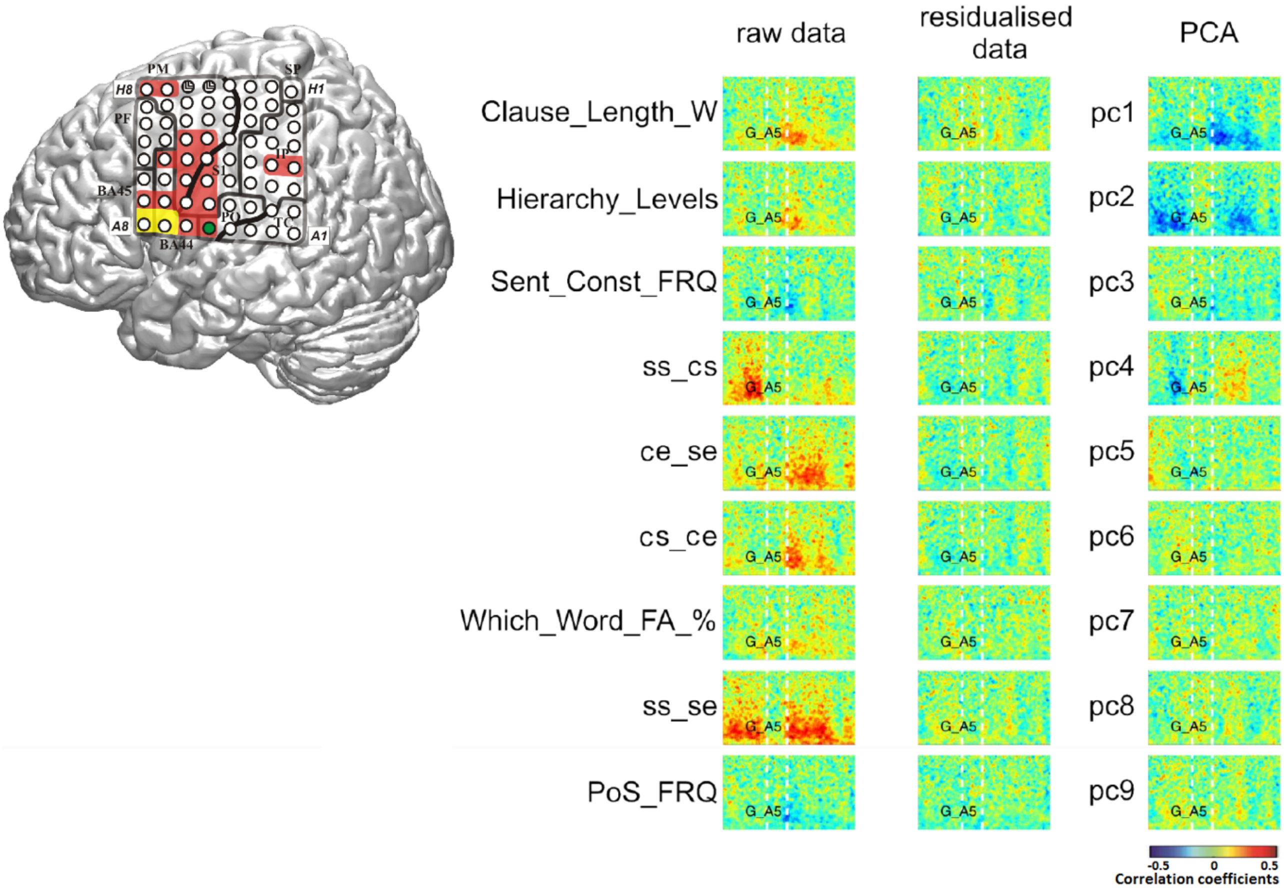


Figure 19: Correlations with the different parameters before residualization (left column), after residualization (middle column), and with the nine principal components (right column) on the example of electrode A5 (marked by a green circle on the brain on the left) in S5. “pc”: principal component, “G_A5”: grid electrode A5, conventions for structural anatomy as in Fig. 1, other conventions as in Fig. 18.

Our correlation analysis prior to residualization showed most effects at electrodes with mouth motor properties (exemplified in the middle row of Fig. 19 using electrode A5 in S5). This analysis showed no meaningful results, with the exception of the effect depicted in Fig. 18. The correlations with the duration-related parameters survived statistical testing ($p<5E-03$, uncorrected) for duration-related parameters only. Like in the word-related analyses, the duration from speech start to clause start showed the most prominent effects prior to clause start. The duration from speech start to speech end showed the most prominent effects before and after the clause, and the duration from clause start to clause end showed most prominent correlations during and after the clause. All meaningful results were unfortunately gone in correlations with the mutually orthogonal parameters (residuals), extracted using a linear regression model predicting each parameter by a set of all other parameters depicted in Fig. 18.

Neurocorrelation analysis with each of the nine pcs at mouth motor electrodes showed most prominent correlations with the first two components. Note that this analysis has, so far, been conducted in S5 only, and the other data sets have not yet been analysed. These negative correlations were significant ($p<5E-03$, uncorrected). If one compares the correlation structure of pc1 (upper panel in the last column of Fig. 19) with the correlations prior to residualization (first column of Fig. 19), one can see that the effect related to pc1 corresponds in time to the neurocorrelation of RSM with such parameters as clause length in words and syntactic hierarchy levels (note that these parameters were among the dominant ones in pc1, see Fig. 16). Thus, it seems likely that the effects of pc1 may reflect combined contributions of several syntax-relevant processes. The correlation with pc2 (second panel from the top in the last column of Fig. 19) is reminiscent of the correlation with the parameter ss_se in the correlations prior to residualization (compare this effect with the one in the upper right panel of Fig. 18). This parameter was the most dominant one in pc2, contributing around 60% to this component in the loadings (Fig. 16). Neurocorrelations with this component thus most likely depict the linguistically-unspecific effects related to speech duration.

4 Discussion

The research described in the thesis at hand is the result of interdisciplinary attempts directed at a better understanding of what signatures of linguistic processing can be observed in the left fronto-temporo-parietal cortex of subjects whose ECoG activity was recorded as they were producing spontaneous, uninstructed oral speech. To be able to undertake such investigations, we built up a neurolinguistic database containing both neural and conversational data, which were transcribed and annotated at multiple levels of linguistic abstraction. The resulting database, which emerged as a result of work of the author of this thesis together with numerous co-workers over years, in entitled "The Freiburg/First Neurolinguistic Corpus" and it contains detailed annotations of spoken German from eight subjects. Data analyses of seven of them are presented in this thesis.

Back in 2012, when this project started, we only had first indications that data obtained in conditions of non-experimental, real-life conversations look meaningful and that they have the potential to unveil linguistically-relevant neural processes, such as related to the identity of the conversation partner (Derix et al. 2012) or to the mnemonic nature of the speech content (Derix et al. 2014). Earlier (unpublished) academic work on linguistic processing (e.g., Kang (2012), who investigated only one ECoG-implanted subject in an attempt to identify word frequency-related processing) was inconclusive. Kang (2012) observed spatially and temporally differential effects of word frequency, depending on whether one- or two-syllable words were studied. Thus, it was unclear back then, how successful our endeavour would be. To be on the safe side, we chose to tackle several levels of linguistic description and to analyse phonological, lexical, and syntactic properties of the available language data, which we addressed in the respective Studies 1-3. The following parts of Discussion will be dedicated to the individual studies. They will be followed by the general Conclusions and Outlook, summarizing the achievements of the present work and its implications for future research.

4.1 Discussion of Study 1: prosody

While the right hemisphere is known to elicit robust effects in tasks related to prosodic processing, it has been debated whether and to what extent prosodic processing is reflected in cortical activity of the left hemisphere (Friederici 2011; Belyk and Brown 2013). Also, the majority of neurolinguistic research has been conducted in conditions of speech perception, while the neural underpinnings of overt expressive speech (Price 2012) and of natural communication (Przyrembel et al. 2012) remain poorly understood. To contribute to a better understanding of these phenomena, we were therefore interested in exploring the neural correlates of prosodic processing in conditions of non-experimental, real-life speech production. We were in the fortunate position of being able to access such conditions by analysing synchronized around-the-clock recordings of audio, video and ECoG data originally obtained for the purpose of pre-neurosurgical diagnostics and then donated for research by consenting neurological patients. Common points of criticism to experimental research on prosody are that it may rely on linguistic stimuli involving intonation patterns which do not necessarily occur in natural language, or that manipulation of prosodic parameters within the same sentence may interfere with the subjects' perception of the associated syntactic structure or of the salience of a given linguistic unit (Wagner and Watson 2010). An attractive property of our data is that they come from experimentally-unconstrained communication which was not subject to artificial modifications and thus present valuable ecological language material.

A particular interest of this study lay on the prosodic phenomenon of the FA, which corresponds to the main prosodic stress of an utterance. Although the FA is an obligatory constituent of each IP of natural spoken language (Pheby 1975; Uhmann 1991), no study so far has addressed the neural correlates of this important linguistic phenomenon in conditions of non-experimental, real-world speech production. We united expertise in linguistics and in human invasive neurophysiology to bridge this gap. We compared words with (FA) and without (nFA) an FA which were matched with regard to a number of linguistic parameters. These included word frequency, the number of phonemes in the spoken word, the position of the word stress, the temporal duration between word start/end and speech start/end, etc. (see Methods). To address the neural underpinnings of emotional compared to non-emotional, linguistic prosody, we did the same for emotional (emoFA) and non-emotional (NemoFA) words with an FA which were assigned to these functional groups based on a rating procedure. Differences between our word categories of interest were not evident from spectral analyses of the underlying neural activity on a coarse time and frequency scale (see Fig. 4 for typical response patterns in the FA and nFA conditions). Statistical analyses using a number of time and frequency components of the neural signals, however, revealed distinctions between words in both contrasts (Tabs. 2 and 3, Figs. 7 and 8). These distinctions were manifested on short time scales in conventional statistics, which speaks in favour of temporally high-resolution methods to study prosodic processing.

4.1.1 Anatomy

Notably, the comparison of FA and nFA words yielded results which were more or less reproducible with regard to both functional and structural anatomy: they lay predominantly in the inferior parietal cortex with cognitive speech- and mouth-movement-relevant areas identified with the help of ESM (Tabs. 2 and 3, Figs. 7 and 8). The observed neural effects could not be explained by systematic differences in vocal pitch and acoustic sound intensity during the production of these words (Tab. 4). The available EMG data suggest that the observed effects in the FA/nFA contrast are unlikely to reside in systematic, language-unspecific differences in myographic activity between FA and nFA words (Fig. 9). Compared with the effects observed in the FA/nFA contrast, those obtained in the emoFA and NemoFA contrast proved anatomically less reproducible, although they also predominated in postcentral regions. Unlike in the FA/nFA comparison, many of the observed effects in the emoFA/NemoFA contrast took place in areas outside the language-relevant cortex, such as in upper- and lower- extremity motor regions. We had no EMG recordings of lower-extremity movements, and EMG data reflecting movements of upper extremities were available for analysis in only one subject (Fig. 9). A contribution of potential systematic differences in myographic activity between emoFA and NemoFA conditions to the observed neural effects therefore cannot be ruled out. Such an explanation appears plausible, since emotions and movements are functionally interconnected (Damasio 1999). Future ECoG studies which will dispose of extremity EMG data are thus needed to validate our findings in this latter contrast. The relation between speech-accompanying gestures and the positioning of the FA may be another interesting topic for future research.

4.1.2 Timing

The neural effects in the FA/nFA and in the emoFA/NemoFA contrasts were found in time periods which differed considerably between subjects and conditions but which nevertheless allowed certain generalizations. With regard to the question whether the effects took place before or after word onset, reproducible patterns could be observed in each condition and in all subjects except for S2 (Tab. 2, Fig. 7): in the FA/nFA contrast, three out of four subjects (S1, S3, S4) showed effects after the average onset of word production (i.e., either during or after the word, cf. Tab. 2), while S2 showed effects prior to it. This general pattern suggests that the effects in the FA/nFA contrast may be more strongly associated with preparation and of speech production rather than with online execution or sensory processing. In the emoFA/NemoFA contrast, two out of three subjects with significant effects (S3, S4) showed effects before and one subject (S2) after word production. The reproducible temporal pattern between S3 and S4 may point to a more prominent role of these effects in perception of own speech rather than in preparation or execution. Our cohort of subjects, however, is small, and these results should be appreciated with caution. Research with a larger cohort is necessary to validate these observations.

4.1.3 Range of spectral frequencies

With the exception of S1 in the FA/nFA contrast, all effects observed in statistical comparisons between word groups took place in gamma frequencies (Wilcoxon rank sum test at $p<2e-04$, uncorrected, Tab. 2, Fig. 7). The exact frequency component in which they were manifested, however, varied between and within subjects and contrasts. This observation did not match our expectation of a broad-banded, frequentially homogeneous effect among subjects. It unclear at present, whether distinctive components of gamma activity support functionally different processes, or if differences in the signal component (and also in the direction of contrast) between conditions may relate to peculiarities of the individual functional neuroanatomy of the subjects. Future research disposing of large neurolinguistic corpora (Iljina et al. 2017) will be essential to answer these interesting questions. A single-trial decoding analysis was able to identify effects not only in the gamma range but also in lower frequencies in most subjects (Tab 3, Fig. 8), suggesting that it may be a helpful for neurolinguistic research and harvest complementary information.

4.1.4 Relation to previous psycho- and neurolinguistic research

Previous research to address the neural correlates of the FA is limited, and it has been conducted with the help of fMRI in experimental conditions of language processing. Wildgruber et al. (2004) studied the effects of linguistic and emotional prosody using fMRI experiments by varying the position of the FA in the acoustic stimuli. In the linguistic prosody condition, the subjects listened to pairs of lexically and grammatically identical German sentences with varying positions of the FA. Based on the position of the FA, they were asked to determine which sentence is better suited to answer a particular question (the linguistic prosody task). The subjects also had to answer the question, “Which of the two sentences sounds more excited?” (the emotional prosody task). Wildgruber et al. (2004) found bilateral effects with rightward asymmetry in both tasks. In the left hemisphere, they found stronger activation of the lateral inferior frontal gyrus related to linguistic prosody and of the orbitofrontal cortex related to emotional prosody. We did not observe such effects in the present study. This may be in part due to the limited spatial coverage of higher-order frontal areas in our sample (Fig. 7) and because of different recording methods and tasks used. Another fMRI study to explore linguistic prosody with the help of the FA was conducted by Tong et al. (2005). These authors also manipulated the position of the FA in the same sentence. Then, they asked the subjects, who were either native speakers of Chinese or of English, to either judge if the FA lay in the same location of the first and in the second acoustic stimuli of the pair or if the sentence was affirmative or interrogative. Rightward asymmetry was observed in both language tasks and subject groups. In the left hemisphere, reproducible effects occurred in the intraparietal sulcus, extending into both inferior and superior parietal cortex. This finding agrees with the effects we observed when comparing FA and nFA words during non-experimental, real-world speech production (Tab. 2 and Fig. 7). Dogil et al. (2002) presented their subjects with pentasyllabic logatomes (e.g., “dadadadada”) in their fMRI experiments on prosody during overt speech production. The subjects were instructed to accentuate logatomes in different ways upon visual presentation. They tasks were to place the main prosodic stress on a particular syllable (a linguistic prosody task), to mimic different emotional states when doing so (an emotional prosody task), or produce the logatomes in a monotonous manner (a baseline task). These authors observed effects relative to the baseline bilaterally in the region of the occipito-basal cortex when comparing both linguistic and emotional prosody tasks with the baseline; they found enhanced activity in the left superior temporal cortex related to the production of the FA and in the right superior temporal cortex related to the production of emotional prosody (their Figure 15). Our comparison of FA and nFA words revealed a difference in gamma activity only in S3 (Fig. 7), although, contrary to the findings by Dogil et al. (2002), a higher level of activation was observed in the nFA condition. The differences between this latter study and our report may reside in the fact that Dogil et al. (2002) used simplified linguistic stimuli and not connected meaningful speech, or also in different recording methods between ours and this latter study.

It is known from the neurolinguistic literature that the inferior parietal cortex contributes to prosodic processing (Belyk and Brown 2013). Our observed differential patterns of neural activity between FA and nFA conditions in the inferior parietal cortex agree with this general notion. They also agree with the report by Tong et al. (2005), who addressed the neural correlates of the FA in an fMRI experiment dedicated to the perception of the FA. The neuroanatomy of such effects may have several functional interpretations. Tong et al. (2005), for instance, proposed that they might reflect domain-general functions of the intraparietal sulcus. The fact that other aforementioned studies have not observed this effect, however, appears to speak against this notion. We rather suggest that the observed reproducible differences between FA and nFA conditions may reside in the speech-relevant mechanisms of attentional prioritization: while FA words are in the speaker's (and respectively also in the listener's) focus of attention, nFA words are in its periphery due to their less prominent role in an utterance (Hirschberg and Pierrehumbert 1986). Attention-related neural effects in previous speech perception research, however, have been localized in areas other than the inferior parietal cortex. They have been reported to occur in the superior temporal cortex (Hugdahl et al. 2003; von Kriegstein et al. 2003) and in portions of the superior parietal cortex (Osaka et al. 2004) during reading tasks. Our observed inferior parietal effects in the FA/nFA contrast may thus reflect a selective attention mechanism which is specific to prosodic processing. Considering that the other aforementioned studies on the neural correlates of the FA did not observe this effect, it will be interesting for future research to determine which factors modulate inferior parietal activity during the production and perception of the FA.

Experimental conditions allow for rigid control of confounding parameters during generation of linguistic stimuli. This is an important motivation for quantitative research on language. To our knowledge, the present study is the first to pursue an alternative approach in which controlled samples from non-experimental language data have been harvested for neurolinguistic research. We implemented a matching algorithm to gain control over potentially confounding linguistic parameters, and at the same time, to get an impression as to which of them may be particularly influential. A rating procedure allowed us to form word groups with different semantic content for the emoFA/NemoFA contrast. By matching each PoS in each subject individually (Suppl. Tabs. 1, 2 in Appendix 1), we were able to obtain word categories with equal numbers of adverbs, verbs and nouns. The parameters which were most difficult to match, i.e., the ones that needed most matching iterations for each

PoS, were word duration and lemma frequency. The duration between speech start and word start and the position of the word stress were the easiest to match (Suppl. Tab. 3 in Appendix 1). This suggests that German spoken words, even within the same PoS category, may vary more strongly with regard to the former parameters, while the latter parameters form a more consistent pattern. Thus, the former parameters are especially important to control for in studies based on natural language.

4.1.5 Plausibility of statistical comparisons using relatively small numbers of trials

Like in previous research (e.g., Davitz and Davitz 1959; Hayashi 1999; Abelin and Allwood 2000), raters in our study were often unanimous in their assignment of “emotional” and “non-emotional” speech. To obtain contrastive emoFA and NemoFA categories, we used a 100% threshold of inter-rater agreement, which was a trade-off between the amount of data and the functional specificity of the analysable word categories, given the overall “fair” quality of inter-rater agreement on the scale by Landis and Koch (1977). Although this methodological decision was associated with a loss in the number of data points for consecutive analyses (ranging from 31 to 68 trials after the implementation of both rating and matching procedures, see Suppl. Tab. 2 in Appendix 1), we do not consider the thereby reduced samples as problematic, or a priori doomed to failure in showing statistically meaningful results. As is illustrated in our recent publication addressing the neural differences between speech and non-speech orofacial behaviours (Kern et al. 2019), comparable numbers of trials as used in the emoFA/NemoFA analysis can, indeed, yield statistically robust results (e.g., ibid.: Fig. 2). Surely, small samples are associated with challenges such as that it can be difficult to estimate the standard error and that additional procedures such as bootstrapping, as implemented in Glanz et al. (2019, Fig. 4), may be needed. Beyond the number of observation points, however, the robustness of a statistical effect is crucially dependent on their distribution: highly significant effects in spite of relatively modest trial numbers can, e.g., occur when comparing samples with narrow distribution of data points around considerably different medians and thus possessing small, non-overlapping standard errors (Sheskin 2007).

While matching and rating procedures allow for control of important potentially confounding parameters, they also inevitably lead to losing data points. The software we used for matching required its output to contain an equal number of words in each word category of the contrast. Prior to matching, the number of FA words was about one third of the number of the nFA words (Suppl. Tab. 1 in Appendix 1), and about a half of the nFA words thus had to be discarded. Approaches which would allow for matching word categories with unequal numbers of trials in the output are desired for future studies. Also the fact that we selected words for the emoFA/NemoFA categories based on rating reduced the sizes of these word groups (Suppl. Tab. 2 in Appendix 1). The raters found it hard to judge, whether the placement of the focus accent was an indication of emotionality during the production of an utterance, and also the inter-rater agreement was low. We therefore selected the words which had a 100% inter-rater agreement to obtain contrastive samples of neurolinguistic data. This is quite rigid compared to, e.g., Derix et al. (2014), where a threshold of 75% was used given a better inter-rater agreement. Determining the influence of the threshold for inter-rater agreement may be an interesting option for future investigations to further disseminate the linguistic and emotional aspects of speech processing in the brain.

4.2 Discussion of Study 2: word complexity

The linguistic complexity of words has largely been studied on the behavioural level and in experimental settings. Only little was, until recently, known about the neural processes underlying this phenomenon in uninstructed, spontaneous conversations. We used ECoG recordings from the fronto-temporo-parietal cortex of five epilepsy patients to investigate, how the linguistic complexity of content words is reflected in cortical activity obtained during real-life speech production. We took an integrative approach involving different measures of word complexity. The investigated parameters were (i) the number of spoken syllables in a word (NoS), (ii) the consonant-to-vowel ratio (CVR), calculated by dividing the number of consonants by the number of vowels in the spoken word, (iii) the "ease-of-articulation" (EoA) index, calculated according to the model by Ziegler and Aichert (2015), and (iv) the lemma frequency of the analysed content words (FRQ).

We performed time- and frequency-resolved correlations of the linguistic parameters of interest with word-accompanying RSM data (Fig. 11). Considering that many parameters turned out to be correlated with each other and also with potentially confounding control variables such as word or speech duration (see Results), we orthogonalized the parameters by extracting residuals with the help of a linear regression model (see Methods). After this, we repeated the correlation analysis with the residuals for each of the linguistic parameters (Tab. 8, Fig. 13). The effects we were able to observe when correlating word-complexity measures with the RSM data occurred, as expected, predominantly in the gamma frequency range (Tabs. 7, 9). Increased activity in gamma frequencies is known as reflecting high effort (Senkowski and Herrmann 2002). High values for the parameters NoS and CVR and low values for the parameters EoA and FRQ are associated with high word production effort. We were thus expecting positive correlations of gamma activity with the former two parameters and negative correlations with the latter two parameters. The correlation effects were, indeed, in agreement with our expectations: prior to residualization, CVR and NoS yielded positive (CVR) or mostly positive (NoS) correlations with gamma activity, and EoA and FRQ yielded predominantly negative correlations (Tab. 7, Fig. 11). This prefix tendency was mostly maintained, whenever significant neurocorrelation effects could be observed after residualization (Tab. 9, Fig. 13).

4.2.1 Anatomy

With regard to the anatomical areas in which the effects related to word complexity could be observed, we were able to witness effects which were not focalized to a single anatomical region but occurred in several areas. This was the case with each word-complexity parameter, both before and after residualization. The effects related to CVR took place in two local clusters, one lying on the border of the three following regions: the premotor cortex, BA 45, and the dorsoventral prefrontal cortex; the other cluster was localized more inferior and posterior, involving electrodes on the central sulcus, in the adjacent parietal operculum and in the neighbouring inferior parietal cortex. Interestingly, this parameter was the only one among the investigated complexity measures which showed effects in these locations, time windows and frequency ranges both before and after residualization (cf. light blue stars in Figs. 11, 13). This may suggest that CVR yielded more reliable effects than the other complexity-related parameters. Prior to residualization, each of the other parameters showed effects in the premotor cortex, in S1, and in the IPC. FRQ and NoS additionally showed effects in the temporal cortex or on the lateral sulcus, and one subject also exhibited an effect in Broca's area in relation to FRQ. These areas are known as language-relevant (Price 2012), and their contribution to word complexity-related processes is anatomically plausible. With the exception of the parameters EoA and NoS, which had a shared effect at the same electrode lying in S1 (marked by a white star in Fig. 11) and also within the same time-frequency range (see Tab. 7 for details), the observed effects took place at different electrodes. The fact that EoA and NoS had a shared effect was not surprising, given that syllable-structure information contributed to the calculation of EoA (Suppl. Tabs. 4, 5 and Suppl. Fig. 1 in Appendix 2). After residualization, most of the observed effects related to these three parameters disappeared, which may be an indication of their weak robustness, or also a reflection of the fact that the removal of collinearity may have left too little meaningful information in the residuals, especially from the parameter EoA, which is composed of multiple articulation-relevant aspects.

4.2.2 Differences between neurocorrelation results between pre- and post-residualization data and the relation to previous work

CVR was the only parameter to show post-residualization effects in all subjects. This likely means that the areas of the pericentral cortex which showed effects related to CVR are sensitive to the distinction between vowels and consonants. The locations of these effects agree astonishingly well with those reported using ECoG by Pei et al. (2011) using single-trial-decoding-based methods (cf. our Figs. 11A, 13A and their Fig. 5). Due to the lack of spatial or temporal reproducibility between subjects and/or between pre- vs. post-residualized effects in the other three complexity-relevant parameters in our study, we shall abstain from functional interpretations of these effects. Since the present evidence is based on a relatively small sample of subjects, which is not uncommon in ECoG studies due to methodological reasons (Diekmann, 2019), further work using larger samples and possibly automated procedures of data gathering will be required. In the research by Ziegler and Aichert (2015), the parameter EoA proved highly informative about errors in apraxia subjects. The present inconclusive results with regard to this parameter do not undermine this valuable previous work: while Ziegler and Aichert's (2015) evidence originated from lesion-based approaches in subjects with apraxia of speech, our data are correlative in their nature and they come from subjects with unimpaired language capacities. The former approach is capable of identifying mechanisms which involve not only individual areas on the cortical surface but likely reflect impairments in conduction of information between areas and also abnormal processes in the deeper cortical layers. Therefore, differences between our findings and those by Ziegler and Aichert (2015) are most likely attributable to methodological reasons.

The fact that lexical frequency, which has proven to modulate neural activity in some studies, did not yield reproducible post-residualization effects in the present work is possibly due to several reasons. First, previous neurolinguistic research dedicated to this phenomenon has largely been conducted in conditions of speech perception and elicitation, such as in lexical decision (Rugg 1990; Prabhakaran et al. 2006) or naming (van Petten and Kutas 1990; Graves et al. 2007) tasks, and it may be that word frequency effects in overt, spontaneous speech production have a distinctive neural infrastructure. Second, it is also conceivable that word frequency effects are not general in the sense that a negative correlation effect is equally visible regardless of the other word-describing parameters but that they are only present when some other parameters are accounted for. Kang (2012), who compared one- vs. two-syllable words in one ECoG-implanted subject, e.g., reported selective involvement of the middle temporal gyrus in the processing of two-syllabic words, and the inferior frontal gyrus proved active only when monosyllabic words were processed. In their EEG study, van Petten and Kutas (1990) observed interactions between word frequency and position of the word in a sentence. They showed that the amplitude of the N400 component of the event-related potential was larger for low-frequency words which occurred early in the presented sentences. These observations tie upon an ongoing discussion in (psycho-)linguistic research, whether or not an effect of word frequency can be seen as an individual, robust phenomenon on its own, whether it is highly context-specific, or, alternatively, if such an effect is epiphenomenal to other linguistic factors. McDonald and Schillcock (2001) showed that contextual distinctiveness, or a corpus-derived measure of word probability in a given context, was a better predictor for lexical decision latencies than word frequency, and that word frequency effects could largely be explained by the presence of syntactic co-occurrence. A follow-up corpus-based investigation by Baayen (2010) addressed the extent to what other factors contributed to word frequency effects. This study, entitled "Demythologizing the word frequency effect," showed that 90% of the variance in word frequencies could be predicted from other lexical properties and concluded that "frequency of occurrence, when understood in the sense of repeated experience, plays only a minor role in lexical processing[17]" (ibid.: 437).

Our evaluation of linguistic parameters is in agreement with Baayen's observation that word frequency information can be correlated with other lexical properties (ibid.) and that it may hence be difficult to isolate a word frequency effect. Lemma frequency in Study 2 was strongly negatively correlated with NoS, weakly negatively correlated with CVR, and strongly negatively correlated with EoA. It also displayed a strong negative correlation with word duration (ws_we). A neurocorrelation analysis prior to residualization showed, as expected, predominantly negative correlations with activity in gamma frequencies (Fig. 11C), but these effects were largely gone when the analysis was performed after mutually-orthogonal parameters were extracted using a linear regression (Fig. 13B). The present findings also agree with those by Diekmann (2019), who did not find statistically robust effects of word frequency with the help of data from our corpus but using different methodological procedures. Note, however, that the absence of evidence (or, in our case, the presence of scarce evidence supporting the neural strength of a lexical frequency effect) should not necessarily be interpreted as evidence of absence (or, in our case, representational weakness): since previous neurolinguistic research showing word frequency effects has observed effects not only within but also beyond the cortical region covered in the FNLC, it is conceivable that recordings from other brain regions beyond our reach would still be informative. More ECoG research involving other brain regions may shed light on this question.

4.2.3 The range of spectral frequencies

Against our expectation of temporo-frequentially extended patterns, the observed correlation effects related to word complexity proved very local in time and frequency (e.g., Fig. 12). At the same time, parameters describing the temporal duration from word start to speech start and also from word end to speech end (ss_ws and we_se) showed correlations with gamma activity in a broad range of frequencies and over more extended periods of time than the investigated linguistic parameters (Fig. 10). This difference may suggest that mechanistic, linguistically-unspecific processes associated with executive functions during speech production are more dominantly represented in the pericentral cortex, compared with the parameters related to word complexity. These findings might also indicate that activity in the gamma range may contain subcomponents tuned to the distinctive linguistic features in temporally and frequentially narrow windows. Considering that the distribution

---

[17] Since spontaneous language is inevitably associated with the problem of collinearity, lexical frequency is likely not the only linguistic property that can largely be explained by other contributors (Baayen 2010). With regard to the psychological plausibility of lexical properties, further research may be of interest which would permit accounting for a larger number of linguistic parameters and the extent to what they are influenced by other factors.

of a frequency spectrum in the gamma frequencies depends on what particular cell types are active (Buzsáki, Anastassiou, and Koch 2012), it is also conceivable that these effects reflect the involvement of particular, localized cell groups which are tuned to individual higher-order processes. Since there is little evidence to support this possibility in current research (Gaona et al. 2011), further work will be needed to address this tentative speculation. Particularly studies with electrodes offering high spatial resolution of ECoG recordings (Wang et al. 2017) may be helpful to this end. The reader is likely wondering whether the temporally and frequentially focalized correlation effects related to word complexity are attributable to the peculiarities of our strict statistical testing procedure, in which we only reported the effects which survived most conservative testing at most conservative thresholds (see Methods), and whether temporally and frequentially extended effects, as we were expecting, would have been observable using less conservative statistics. Unfortunately, this would generally not be the case: the effects related to our word complexity measures were always arranged in temporo-frequentially narrow clusters, as is shown on typical examples in Fig. 12. Therefore, we believe that these properties are rather indications of moderate sensitivity of RSM values to parameters reflecting word complexity, compared to those related to the duration of the word-embedding speech production epochs.

4.2.4 Timing

We also find it difficult to draw conclusions with regard to the timing of the neural effects and its relation to (a) particular stage(s) of linguistic processing. This is especially the case with the parameters EoA and FRQ, which showed effects before, during, and after word production (Tabs. 7, 9). If one assumes that the observed effects are not artefacts of statistical testing (see Methods), the observed disparate time points of activation might indicate involvement of these processes at multiple stages of word production and monitoring of the language output. In line with the assumption of syllabification taking place at relatively early stages of speech (Levelt and Meyers 2000), the effects related to NoS took place either shortly before word onset or during word production. Due to the fact that this parameter lacked reproducibility with regard to the spatial and frequential properties, the related effects, too, need further validation and should be interpreted with caution. As concerning the parameter CVR, the timing of the effects appears interesting: CVR-related effects nearly always occurred either prior to or after but not during word production (with the exception of one electrode in the pre-residualized and one electrode in the post-residualized data, Tabs. 7, 9). Our results do not allow drawing clear distinctions between anatomical areas with regard to the timing of the effects, since the same areas sometimes showed effects both before and after the word (e.g., the dorso-ventral prefrontal cortex converging on the premotor cortex or also the parietal operculum). Network-structure analyses with functional connectivity measures (e.g., as in Kern et al. 2013) may be helpful to understand the relation between the time and region of CVR-related neural effects. Such an investigation was, however, beyond the scope of Study 2 and further work in this direction may be of interest.

To sum up, this ECoG study reports on an innovative undertaking which investigated word-complexity-related effects during non-experimental, real-world speech production by using a combination of linear regression and correlation approaches. We are aware of the fact that these linear methods may overlook non-linear relationships within the data. By removing linear trends, however, we made a step toward reducing mutual dependency on the individual parameters on each another, and this measure was sufficient to obtain parameter-specific neural effects (cf. Figs. 11 and 13). In doing so, we were able to identify CVR as the linguistic parameter yielding most reproducible and robust positive correlations with gamma-range activity in the pericentral cortex. The anatomical location of our observed CVR-related effects agrees with the location of areas which were most informative about the distinction between consonants and vowels in the single-trial-decoding study by Pei et al. (2011). This similarity shows that spontaneous and experimentally elicited speech involve anatomically similar neural resources, at least regarding the production of vowels and consonants. It is also in line with the idea that the phonological composition of speech is a promising way to better understand the functional organization of the language-relevant pericentral cortex (Blakely et al. 2008; Bouchard et al. 2013; Mugler et al. 2014; Ramsey et al. 2018). The lack of reproducibility in a number of characteristics of the neural signal underlying the other investigated word-complexity-related parameters may indicate their moderate representation in the studied neural signals. The fact that parameters related to speech duration, on the contrary, yielded strong, temporo-frequentially reproducible and spatially focalized effects in the articulatory motor cortex using exactly the same method (Fig. 10) speaks for the feasibility of our approach and suggests that linguistically unspecific parameters associated with preparation for articulation and with articulation proper clearly dominate over the here studied linguistically-relevant processes in the investigated portion of the fronto-temporo-parietal region.

4.3 Discussion of Study 3: syntax

In Study 3, we were interested to find out, whether linguistic parameters describing the composition of naturally-produced simple clauses would be reflected in cortical activity, and, if yes, what temporal properties this activity would possess. The question into the temporal aspect on cortical activation was motivated by usage-based linguistic theory. Auer (2005; 2009) assumes that a sentence is planned incrementally and “on-line,” in the act of speaking, and that the cognitive load associated with the planning reduces toward clause end due to the fact that few the already produced language material makes the following linguistic units increasingly predictable. Departing from this assumption, one might expect pre- but also post-onset activity which would decrease in intensity over time. We conducted clause-averaged analyses of the neural data and inspected the neural properties of high gamma activity, which is known as a robust and reliable index of cortical activation (Crone, Korzeniewska, and Franaszczuk 2011). Since syntax-relevant processing is relevant to the production of each sentence, we expected that neural activity related to this phenomenon m be visible in the trial-averaged data.

4.3.1 Anatomy

As is shown in Fig. 17, illustrating typical RSM changes related to clause production, we observed spatially local increases in high gamma activity which, indeed, started prior to clause production and either ended prior to clause end or roughly around it. These effects either occurred in cortical regions with articulatory properties which lay either in the premotor cortex, on the central sulcus, or in the caudal part of BA 44 or in areas belonging to the lateral sulcus and the superior temporal cortex. Our observed effects in BA 44 are in agreement with previous studies on syntactic processing (Caplan 2015), also in conditions of speech production (Indefrey et al. 2004), although the contribution of other processes is equally possible (Price 2012).

4.3.2 Timing

The overall temporal development of neural signals does not lend support to the assumption of a selective feature of the neural signal which would indicate completed syntactic planning prior to articulation. Our results are in agreement with the notion by Auer (2009) in the sense that activation related to clause production took place online, over the course of clause production, and that it diminished over time. Given that most of these electrodes were associated with mouth motor processing, one cannot say with certainty, to what extent the observed effects are due to syntactic processing and whether syntax-relevant activation can be told apart from the syntactically-irrelevant components of speech production. In our opinion, it is conceivable that shared linguistic processes may be reflected in this activity, and that the phenomenon of linguistic projection may take place on several levels of linguistic abstraction but also on the level of motor planning. To be able to single out the syntactically-specific processes, further work comparing linguistic units in which syntactic planning is relevant with units which do not involve syntactic planning (such as in Indefrey et al. 2001 or Indefrey et al. 2004, who addressed this question using fMRI, which possesses only low temporal resolution inherent to this method), will be required. Such work will, however, call for an experiment rather than spontaneous language data, since spontaneous language is, with the exception of – in our experience – rare single-word utterances, syntactically meaningful and connected.

4.3.3 The range of spectral frequencies

Our attempts to find linguistically-specific effects related to the syntactic parameters of interest using the same methods as described in Study 2 mostly did not yield results which would be in agreement with our neurobiological expectations of temporally- and frequentially-extended effects in high gamma frequencies. One notable exception was the number of hierarchy levels in a sentence, which elicited a temporo-frequentially-extended correlation with high-gamma activity in ca. 60-150 Hz over the course of clause production at one electrode in the medial temporal lobe of S7 (Fig. 18). This effect was significant ($p<5E-03$, uncorrected) only before but not after residualization (not visualized). The data obtained from the other subjects did not cover this cortical region, and a comparative analysis was therefore not possible. The finding of effects related to syntactic complexity is in agreement with previous findings by Brennan et al. (2012), who found effects related to syntactic hierarchy in this cortical region in addition to the anterior frontal and anterior temporal areas. Neurolinguistic studies on syntactic complexity which have been conducted with the help of different parameters have also found effects in this cortical region (e.g., Just et al. 1996; Keller et al. 2001). Since the effect we were able to observe could not be reproduced in the other samples due to the lack of comparable electrode coverage and validation in a larger sample is needed, we shall refrain from its detailed discussion in the light of the current psycho- and neurolinguistic literature. It may be interesting to conduct follow-up ECoG investigation of how syntactic hierarchy is represented in other subjects with good electrode coverage on the middle temporal region.

4.3.4 Composition of the pcs

We further undertook an attempt to identify combined aspects of the linguistic data which would, together, be informative of the syntactic features of our interest, and to find out if they would show robust representations in the neural data. To this end, we first conducted a PCA, identified the pcs explaining most of the variances in the data (the first three pcs explained around 70%), and attempted to interpret the pcs from a linguistic point of view. An interesting finding was that pc1 was reproducibly dominated by the parameters “clause length in words,” “the number of hierarchy levels,” “clause duration,” “the frequency of syntactic constellations," and the frequency of PoS constellations," which are linguistically-relevant; pc2 was, on the contrary, always dominated by the linguistically-unspecific parameters describing the duration from speech start to clause start and from clause end to speech end. Other pcs, starting from pc3, proved non-reproducible with regard to the prefix and weight of the individual parameters in the loadings. These findings are illustrated on the example of subjects S4 and S6 in Fig. 16. The position of the FA did not make a substantial contribution to these main pcs. This parameter was a major contributor to pc8 only, and this pc explained little variance in the linguistic data (not visualized). This observation suggests that the position of the FA is largely independent on the here investigated syntactic properties of the language material but presents an individual linguistic dimension. Note that this observation does not undermine previous works identifying links between the placement of the FA and the syntactic properties of sentences; it rather suggests that the link between the prosodic and syntactic structure is not general (i.e., present regardless of the syntactic composition of the individual sentences) but likely context-specific, e.g., as is shown in Uhmann's work (1991) relating the breadth of the focus domain to the FA-carrying PoS within the sentence.

4.3.5 Interpretation of the neurocorrelation with the main pcs

We correlated the loadings of the pcs with the neural activity in the same was as we did with the individual linguistic parameters. For reasons of time and effort, this analysis has, so far, been conducted in one subject (S5) only (Fig. 19), but it yielded interesting and promising results. Unlike the individual parameters, pcs 1-4 yielded strong, robust effects which were, in agreement with our neurological expectations, stretched over multiple frequencies and time points. These effects were located in the pericentral cortex along the central sulcus and they took place at electrodes implicated in mouth motor processes (exemplified in Fig. 19 using one electrode in the parietal operculum). Interestingly, although the first pc as dominated by linguistic and not by speech-duration-related parameters, as was the case with pc2, both pcs showed correlations with high gamma activity, with the time and not the location of correlation being informative of these differences: while pc2 showed effects prior to and about a second after clause duration, pc1 showed effects shortly after clause production. This speaks for a possible contribution of this electrode in feedback-related processing of syntax-relevant information and is in line with the assumption that motor-cortical areas may contribute not only to motor but also to higher-order, syntactic processing (Fogassi and Ferrari 2007). The author of this thesis, however, would like to formulate this as a careful speculation, since validation of this interesting finding in the other subjects is desired and will be conducted out of the scope of the thesis at hand.

5 Peculiarities and limitations of our approach

5.1 Sample sizes

The present thesis reports on three studies that were conducted to address the neural correlates on linguistic processes during spontaneous speech production which were conducted based on a dedicated multimodal neurolinguistic corpus built up to this end. While the analysed ECoG data possess numerous attractive properties such as the excellent temporal and a good spatial resolution (Ball et al. 2009) and allow capturing clear changes of speech-related high gamma activity (Crone et al. 2001a,b), their collection and analysis are associated with certain challenges that need to be mentioned. Since ECoG data emerge as a by-product of pre-neurosurgical evaluation, which is carried out in comparatively few institutions in Europe, such recordings are relatively rare. As far as we are concerned, about ten patients per year have received such implantations over the last years in Freiburg; not all of them had implantations which would meet the requirements to be included in our corpus (i.e., that the neural data come from fronto-temporo-parietal areas, that the seizure onset zone lies outside major language areas, and that the subjects provide written informed consent for retrospective evaluation of the recorded materials). Furthermore, current diagnostic procedures are moving away from large electrode grids, as analysed in the present thesis, toward increasingly spatially focalized approaches, such as by targeting seizure onset areas with the help of small depth electrodes (Miller et al. 2013), which reduces the number of large-scale implantations. This inevitably limits the sample of subjects who can be studied. In comparison with non-invasive methods, ECoG studies thus usually investigate smaller populations. Note that the sample sizes in the here reported work are not uncommon in such research: Crone and colleagues, e.g., investigated high-gamma activity patterns during speech production in one (2001a) and during speech perception in four subjects (2001b), Bouchard et al. (2013) studied the functional organization of the sensorimotor cortex during speech articulation in three subjects, and the study by Herff et al. (2016), which attempted to reconstruct speech from ECoG activity recorded from temporal areas during speech perception, reports on results from a single subject. Our neurolinguistic results based on four (Study 1), five (Study 2), and one (Study 3) subjects thus fall within the scope of a usual cohort size for ECoG-based research on language. Comparisons of our findings with results obtained using other, non-invasive methods, as well as validation of our findings by follow-up ECoG studies into the neural representation of natural language will be of interest.

5.2 Reproducibility of the neural results in the light of data quality and quantity

Given our relatively small samples in comparison with non-invasive research, the neurolinguistic findings of the here reported work need to be commented with regard to their reproducibility and its relation to the quality of the acquired data.

5.2.1 The inter-rater agreement in our emotionality ratings is common in the literature

In Study 1 addressing the difference between FA and nFA and also between emoFA and NemoFA word groups, we were able to observe most differences between these groups in the postcentral cortex (Fig. 7), and a single-trial decoding analysis confirmed their predominantly postcentral location (Fig. 8). These effects, however, took place in several frequency components and in different 100-ms time windows either before or after word production (Tab. 2), indicating their moderate reproducibility with regard to the temporo-frequential characteristics of the ECoG signal. Several interpretations of this lack of reproducibility are conceivable: (i) either the linguistic data in the available amount and quality do not allow seeing a picture clear enough or (ii) these temporo-frequentially narrow effects likely point to functionally specific activation of small local populations of neurons, whose signal properties indeed vary between linguistic parameters and subjects. The numbers of trials that were available to us for the analyses in Study 1 were limited by constraints of data harvesting: in the emoFA/NemoFA contrast, for instance, we implemented a rating procedure and selected words at a 100%-agreement for emotionality assignments to create maximally contrastive samples, and we matched the words with regard to a number of control parameters to obtain controlled data (see Methods).

Given the fact that the inter-rater agreement in the assignment of emotionality proved only "fair" (21-40% in Landis and Koch (1977)) in all subjects (37% in S1, 35% in S2, 37% in S3, and 38% in S4), one may suspect that the lack of correspondence between raters in our data is due to poor instruction or the inferior quality of the linguistic information available to the raters to enable clear judgment. The author of the present thesis, however, is convinced that this is not the case for the following reasons. First, high-quality transcriptions were made continuously and context information was available to the raters (see Methods). Second, all raters were native speakers of German, trained in linguistics, experienced in transcription, and familiar with the acoustic data behind the transcribed speech. Third and foremost, our measures of inter-rater agreement for the emotional content of speech fall within the average range of what has been reported in the literature: Siegert et al. (2014), e.g., investigated, among other sources, inter-rater agreements in the assignment of emotions ("valence," "arousal," and "dominance") using the "The Vera am Mittag" audio-visual emotional speech database consisting of dyadic conversations from a German talk show, which contained high-quality emotional speech. They found out that inter-rater agreements were “quite poor. When evaluated with the agreement interpretations […], the nominal values are poor to slight, whereas the ordinal values are fair to moderate. But with a smallest value of 0.086 and a highest value of 0.478 they are far away from a good or substantial reliability” (ibid.: 22). These authors replicated this result based on other corpora relying on different emotionality annotations and containing a distinction between emotional vs. non-emotional speech as well as a variety of emotions (see their Table 2 on page 21). They generally “confirm the assumption of a low inter-rater agreement for emotional annotation. The values for the reliability utilizing a nominal metric distance are between 0.165 and 0.217, which means a poor to lower fair agreement” and “suppose that the specific method used does not affect the inter-rater reliability and therefore, the choice is only a matter of personal preferences and of the investigated scientific question” (ibid.: 23). Since our values for inter-rater agreement correspond to the average range reported in the literature (see Siegert et al. 2014 for a detailed summary), we do not attribute the moderate reproducibility of the temporo-frequential features of the neural findings to caveats in the quality of the linguistic data or rater instruction.

Previous (psycho-)linguistic literature agrees on the fact that judgments on the emotionality of speech are generally difficult because emotions cannot necessarily be perceived by external observers (Truong et al. 2012): Truong et al. (2008), for instance, found that the agreement between self-rater assignments of emotionality was higher among external raters than between external raters and self-raters. A plausible way to improve the quality of emotional assignment in future research may be to ask the speakers to rate their own speech. Whether better reproducibility of neural signals can be achieved by implementing such a rating procedure, is an interesting question for follow-up ECoG research.

5.2.2 Limited evidence for clear effects of lexical frequency

The results of Study 2 indicated that, contrary to our expectation, lexical frequency did not show the clearest effects in comparison with other parameters. Besides the strong effects of the temporal distance between speech and word start as well as speech and word end along the course of the central sulcus (Fig. 10), it was the CVR that yielded most robust results that could be observed in correlations of this parameter with the neural signal both before and after residualization in the fronto-parietal cortex (Figs. 11, 13). The fact that the negative correlations of FRQ with cortical activity that were observable with high gamma activity, which was in agreement with our hypothesis, could be observed prior but not after residualization, may be related to the presence of strong correlations between this parameter with other parameters of the linguistic data (see Results and Baayen 2010). The extent to what variance in lexical frequency can be explained by other linguistic factors is an interesting question for investigation, which may further extend current understanding of this phenomenon (ibid.). With regard to our scarce findings on the neural representation of word frequency, a comparison with a previous study from our lab is possible. In her doctoral thesis based on our corpus, Diekmann (2019) arrived at the same conclusion, although she used a different approach to speech segmentation (automated and manually corrected setting of word borders in the acoustic data with post-hoc co-registration with the ECoG signal) and implemented a different corpus for the extraction of lexical frequencies (sDEWaC; Faaß and Eckart 2013). We thus think that attribution of the limited evidence pointing to clear effects of lexical frequency in our spontaneously spoken data to errors in speech segmentation and frequency assignment is unlikely[18]. It is conceivable that data from a larger cohort of subjects and from additional brain areas may yield a clearer picture than is possible based on our data. Addressing this question with the help of ECoG, however, will require larger datasets than those currently available in "The Freiburg/First Neurolinguistic Corpus".

5.2.3 First interesting evidence on syntax and syntactic frequency

Study 3 has shown some interesting results with regard to syntactic parameters including the frequency of syntactic constellations: linguistic factors such as clause length in words, syntactic hierarchy levels, clause duration, and the frequencies of PoS- and sentence constituent constellations dominated the first principal component of the linguistic data (Fig. 16.), and a neurocorrelation analysis of cortical activity with this component showed meaningful effects in the articulatory cortex (Fig. 19). This result agrees with the idea that syntactic processing may be supported by the articulatory cortex and not only by the classical higher-order areas (Fogassi and Ferrari 2007). Since this latter analysis was, so far, conducted on data from one subject only, validation based on the remaining data sets will be of interest.

5.3 Importance of both experimental and non-experimental approaches

The data analysed in the thesis at hand were collected in conditions of spontaneous, uninstructed communication. As is explained in detail in the Introduction, our choice of such data was motivated by concerns of ecological validity of the spoken material and of the associated neural recordings. A challenge one has to meet when dealing with such data, however, is the presence of collinearity between linguistic parameters, as well as possible contributions to EMG and auditory signal properties to the observed neural effects. While we were able to account for a number of variables, influence of other factors, such as medication or some additional aspects of behaviour that have not been captured, cannot be excluded. The experimental approach has an important advantage inasmuch as it enables rigid control of potentially confounding factors, which may not always be possible in non-experimental studies.

By setting the scope of the present research on natural communication, we do not seek to undermine the importance of previous evidence obtained using experimental psycho- and

---

[18] Also note that we have implemented extensive checking procedures to prevent errors in annotation and co-registration of materials form different modalities (see Methods).

neurolinguistic research. Instead, we believe that systematic comparisons of results obtained using both approaches can lead to a better understanding of the neural and behavioural processes that are typical of human communication.

6 Conclusions and outlook

To sum up, a key innovation of the present thesis is that it undertook an attempt to study linguistic processing in conditions of experimentally-unconstrained, real-life speech production and at several levels of linguistic abstraction.

To this end, we created and analysed a multimodal neurolinguistic corpus which consisted of audio, video, EMG, and neural recordings of data from seven epilepsy patients, comprising ca. 450 simple clauses and 1000 content words per subject. The speech data were transcribed and segmented using established linguistic conventions (Selting et al. 2009) and annotated with regard to prosodic, lexical, and (morpho-)syntactic features. With the help of colleague linguists, the author of the present thesis maintained and updated the corpus continuously throughout this work, making sure that coherence between linguistic annotations and the associated neural recordings, as well as between linguistic analyses on different levels of abstraction, was granted. The data were subsequently used for neurolinguistic analyses which are reported in this doctoral thesis, in the doctoral thesis by Diekmann (2019) and in several other academic works affiliated with out lab (e.g., Hader 2016; Klumpp 2016).

The main methodological challenge related to the analysis of the gathered data dealt with the question of how to extract controlled material out of the variable and complex behavioural and linguistic data which are representative of natural verbal communication. We conducted three studies, each dedicated to a distinctive aspect of linguistic processing. Study 1 was dedicated to prosody, Study 2 dealt with questions related to word complexity, and Study 3 investigated the syntactic processing underlying clause production. The methodological approach we opted for to meet this challenge consisted of (i) application of a matching procedure to a priori selected word categories for further comparisons of neural activity between them (Study 1), (ii) orthogonalization of the individual linguistic parameters in an attempt to overcome the problem of collinearity between correlated linguistic parameters, (iii) the usage of a PCA to extract most informative components of the linguistic material which would also be mutually-orthogonal.

Our findings in Study 1 showed that effects in the postcentral cortex could be observed in relation to prosodic processing. These effects, however, were little reproducible with regard to the timing and frequency components of high gamma activity. In Study 2, we were able to identify CVR (i.e., the proportional relation between vowels and consonants) as the most informative parameter which showed meaningful correlations with high gamma activity at the junction of the premotor cortex with BA 44 and the prefrontal cortex and in the ventral postcentral region. These effects occurred prior to as well as and after residualization, which may be an indication of their robustness. In terms of frequency and time, however, they were also little reproducible, similar to the effects observed in Study 1. The temporo-frequential characteristics of these effects contradict our expectation of temporo-frequentially extended effects, as are well-known for event-related cortical processing (e.g., Derix et al. 2014). This may be either an indication of the fact that the studied linguistic parameters are only moderately represented in the neural signals we have investigated. Since the spectrum of gamma frequencies is a composite phenomenon which relies on multiple cell types (Buzsáki, Anastassiou, and Koch 2012), however, it is also conceivable that these temporo-frequentially narrow effects may indicate the functionally specific activation of narrow populations of neurons whose signal properties vary between linguistic parameters and subjects. Since we are not aware of any published ECoG works investigating neural effects of linguistic processing during natural speech production, further validation of these observations will be required. Study 3, which was dedicated to syntactic processing, yielded an interesting finding of a temporo-frequentially extended correlation of a linguistically-dominated pc1 with neural activity in the mouth motor cortex. This is in line with the assumption that motor-cortical areas may engage not only in low-level motor but also in higher-order operations such as syntax. Since we have been able to complete this analysis in one subject by the time point of completion of this thesis, validation of this interesting finding will be of interest in the author’s future work.

A further interesting step of analysis would be to combine statistical approaches from Study 2 and Study 3 and calculate a PCA on the word-level data. This would enable interesting comparisons with the work by Baayen (2010) and possibly allow creating a model of word complexity which would account for the individual contributions of the studied linguistic factors in our spontaneously spoken material. Considering that previous modelling work by Ziegler and Aichert (2015) is based on evidence from a clinically different group of subjects, a comparison of such an endeavour to their results will be of interest.

In general terms, this thesis demonstrates that it is possible to achieve control over numerous potentially confounding variables and to obtain meaningful neural effects from ECoG recordings underlying spontaneously spoken language. These observations speak in favour of the emergent non-experimental approach to neurolinguistic (Derix et al. 2012; 2014) and neuroscientific (Ruescher et al. 2013; Wang et al. 2016) research, and they open up interesting and still little-explored possibilities to study the neural correlates of linguistic processing during real-world conversations.

The fact that the neural effects observed in the here reported investigations were moderately reproducible between subjects in terms of temporal and frequential properties indicates that further work may be needed which would allow accounting for larger data sets. The annotation procedure and the maintenance of “The Freiburg/First Neurolinguistic Corpus,” however, were associated with substantial investment of time and effort. Automation of transcription, annotation of neural and linguistic data would therefore be welcome. Automated tagging of the linguistic data proved frequently erroneous in our experience, which is likely due to the fact that currently available automated tools have been developed and optimized based on conventions for written German language (e.g., Foth 2006). For this reason, it was a deliberate decision to conduct the annotation and correction of “The Freiburg/First Neurolinguistic Corpus” largely manually. As is illustrated in the Introduction, we encountered a number of challenges, multiple linguistic decisions were made to standardize the annotations throughout the corpus (Neuromedical AI Lab 2019, online), and numerous Matlab-based programs had to be written to manage the corpus’s quality. An interesting further step would be to use these sources of information, combining them with already established automated procedures, e.g., those implemented in Weblicht (e.g., Hinrichs et al. 2010), in order to be able to meet challenges associated with the annotation of spontaneously spoken language with the help of adequately designed automated tools.

Neural recordings from the cortical surface of human subjects are extremely rare, and collection of large data sets and their annotation using a common standard may be difficult to achieve. Since current neurosurgical developments are moving away from large-surface grids to increasingly compact designs of recording tools (Wang et al. 2017), the scientific value of the recordings in “The Freiburg/First Neurolinguistic Corpus” and of their alike cannot be overestimated. While linguistic research has been pursuing the strategy of creating language corpora for common use since decades, this practice is relatively new in neurolinguistics. So far, a handful of data sets from neurolinguistic experiments are available online (e.g., on www.brainsignals.de), and coordination of efforts from multiple labs may be helpful to produce large-scale neurolinguistic corpora. We hope that the neuroscientific community further uses the wealth of the documented linguists’ experiences with spontaneous spoken speech to meet this challenge, as attempted in the present thesis.

References

1. Abelin, Å., & Allwood, J. (2000). Cross linguistic interpretation of emotional prosody. In: *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 1*, pp. 110-113.
2. Aertsen, A. M. H. J., & Johannesma, P. I. M. (1981). A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biological cybernetics*, *42*(2), 145-156.
3. Aichert, I., & Ziegler, W. (2004). Segmental and metrical encoding in aphasia: Two case reports. *Aphasiology, 18(*12), 1201-1211.
4. Altmann, H. (1993). Fokus-Hintergrund-Gliederung und Satzmodus. In: Reis, M. (Ed.): *Wortstellung und Informationsstruktur*, pp. 1-37. Tübingen: Niemeyer.
5. Auer, P. (2005). Projection in interaction and projection in grammar. *Text-Interdisciplinary journal for the study of discourse*, *25*(1), 7-36.
6. Auer, P. (2009). On-line syntax: Thoughts on the temporality of spoken language. *Language Sciences*, *31*(1), 1-13.
7. Auer, P. (2010). Zum Segmentierungsproblem in der gesprochenen Sprache. *InLiSt-Interaction and Linguistic Structures, 49*, 1-19.
8. Baayen, R. H. (2008 [repr. 2012]). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
9. Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon, 5*(3), 436-461.
10. Ball, T., Kern, M., Mutschler, I., Aertsen, A., & Schulze-Bonhage, A. (2009). Signal quality of simultaneously recorded invasive and non-invasive EEG. *Neuroimage*, *46*(3), 708-716.
11. Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production?. *Journal of Memory and Language*, *24*(1), 89-106.
12. Bartlett, S. F. C. (1995). *Remembering: A Study in Experimental and Social Psychology*, Cambridge: Cambridge University Press.
13. Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7), 1-23.
14. Batliner, A. (1989). Fokus, Modus Und Die Grosse Zahl. Zur Intonatorischen Indizierung Des Fokus Im Deutschen. In: Altmann, H. et al. (Eds.), *Zur Intonation von Modus Und Fokus Im Deutschen*, pp. 21-70. Tübingen: Niemeyer.
15. Belyk, M., & Brown, S. (2013). Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *9*(9), 1395-1403.
16. Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex*, *19*(12), 2767-2796.
17. Blakely, T., Miller, K. J., Rao, R. P., Holmes, M. D., & Ojemann, J. G. (2008). Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 4964-4967. IEEE.
18. Bloomfield, L. (1926). A set of postulates for the science of language. *Language*, *2*(3), 153-164.

19. Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9/10), 341-347.
20. Bose, A., van Lieshout, P., & Square, P. A. (2007). Word frequency and bigram frequency effects on linguistic processing and speech motor performance in individuals with aphasia and normal speakers. *Journal of Neurolinguistics*, *20*(1), 65-88.
21. Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327.
22. Brådvik, B., Dravins, C., Holtås, S., Rosen, I., Ryding, E., & Ingvar, D. H. (1991). Disturbances of speech prosody following right hemisphere infarcts. *Acta Neurologica Scandinavica*, *84*(2), 114-126.
23. Braze, D., Mencl, W. E., Tabor, W., Pugh, K. R., Constable, R. T., Fulbright, R. K., ... & Shankweiler, D. P. (2011). Unification of sentence processing via ear and eye: An fMRI study. *Cortex, 47*(4), 416-431.
24. Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and language*, *120*(2), 163-173.
25. Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature reviews neuroscience*, *13*(6), 407.
26. Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
27. Caplan, D. (2007). Functional neuroimaging studies of syntactic processing in sentence comprehension: A critical selective review. *Language and linguistics compass*, *1*(1-2), 32-47.
28. Caplan, D. (2015). The neural basis of syntactic processing: a critical review. In: Hillis, A. (Ed.), *The Handbook of Adult Language Disorders, 2nd edition*, pp. 355-374. New York, London: Psychology Press.
29. Caplan, D., Alpert, N., & Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience*, *10*(4), 541-552.
30. Caplan, D., Alpert, N., & Waters, G. (1999). PET studies of syntactic processing with auditory sentence presentation. *NeuroImage*, *9*(3), 343-351.
31. Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
32. Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, *110*(7), 2653-2658.
33. Cheung, H., & Kemper, S. (1992). Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics, 13*(1), 53-76.
34. Chomsky, N. (2006). *Language and mind*. Cambridge: Cambridge University Press.
35. Chomsky, N. (2013). Problems of projection. *Lingua*, *130*, 33-49.
36. Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
37. Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084.
38. Crone, N. E, Hao, L., Hart, J., Boatman, D., Lesser, R. P., Irizarry, R., & Gordon, B. (2001a). Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology*, *57*(11), 2045-2053.

39. Crone, N. E., Boatman, D., Gordon, B., & Hao, L. (2001b). Induced electrocorticographic gamma activity during auditory perception. *Clinical Nneurophysiology*, *112*(4), 565-582.
40. Crone, N. E., Korzeniewska, A., & Franaszczuk, P. J. (2011). Cortical gamma responses: searching high and low. *International journal of psychophysiology*, *79*(1), 9-15.
41. Crone, N. E., Miglioretti, D. L., Gordon, B., & Lesser, R. P. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain*, *121*(12), 2301-2315.
42. Crystal, D. (1987). *The Cambridge encyclopaedia of language*. Cambridge: Cambridge University Press.
43. Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.
44. Davitz, J. R., & Davitz, L. J. (1959). The communication of feelings by content-free speech. *Journal of Communication, 9*, 6-13.
45. Demandt, E., Mehring, C., Vogt, K., Schulze-Bonhage, A., Aertsen, A., & Ball, T. (2012). Reaching movement onset- and end-related characteristics of EEG spectral power modulations. *Frontiers in neuroscience*, *6*, 65.
46. Derix, J., Iljina, O., Schulze-Bonhage, A., Aertsen, A., & Ball, T. (2012). “Doctor” or “darling”? Decoding the communication partner from ECoG of the anterior temporal lobe during non-experimental, real-life social interaction. *Frontiers in human neuroscience*, *6*, 251.
47. Derix, J., Iljina, O., Weiske, J., Schulze-Bonhage, A., Aertsen, A., & Ball, T. (2014). From speech to thought: the neuronal basis of cognitive units in non-experimental, real-life communication investigated using ECoG. *Frontiers in human neuroscience*, *8*, 383.
48. de Zubicaray, G., McMahon, K., Eastburn, M., & Pringle, A. (2006). Top-down influences on lexical selection during spoken word production: A 4T fMRI investigation of refractory effects in picture naming. *Human brain mapping*, *27*(11), 864-873.
49. Diaz, M. T., & McCarthy, G. (2009). A comparison of brain activity evoked by single content and function words: an fMRI investigation of implicit word processing. *Brain research*, *1282*, 38-49.
50. Diekmann, B. C. (2019). *Wortproduktion im Satzkontext: Eine ECoG-Untersuchung*. [Doctoral dissertation, University of Freiburg], FreiDoc plus, Universitätsbibliothek Freiburg, doi: 10.6094/UNIFR/149963.
51. Diekmann, B., Iljina, O., Wiest, P., Schulze-Bonhage, A., Auer, P., & Ball, T. (2016, August*). A novel neurolinguistic corpus of spontaneous speech production and concurrent electrocorticographic (ECoG) data.* [Poster presented at the Society for the Neurobiology of Language, London, UK].
52. Dieminger, L. (2017). **“**An ECoG investigation of articulatory complexity and word frequency of real-world speech production” [Bachelor thesis in liberal arts and sciences, University of Freiburg]
53. Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., Riecker, A., & Wildgruber, D. (2002). The Speaking Brain: A Tutorial Introduction to fMRI Experiments in the Production of Speech, Prosody and Syntax. *Journal of Neurolinguistics, 15*(1), 59-90.
54. *Duden Online*. (no year). Online dictionary on www.duden.de (last visited on 11.04.2019).
55. Eisenberg, P. (2004). Der Satz: Grundriß der deutschen Grammatik, Bd. 2. Stuttgart: Verlag J. B. Metzler.

56. Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, *24*(2), 143-188.
57. Faaß, G., & Eckart, K. (2013). Sdewac – a corpus of parsable sentences from the web. In: *Language processing and knowledge in the Web*, pp. 61-68. Berlin / Heidelberg: Springer.
58. Féry, C. (1993). German intonational patterns, vol. 285 of *Linguistische Arbeiten*. Niemeyer, Tübingen.
59. Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, *76*(5), 378.
60. Fogassi, L., & Ferrari, P. F. (2007). Mirror neurons and the evolution of embodied language. *Current directions in psychological science*, *16*(3), 136-141.
61. *FOLK*. (2012). Forschungs- und Lehrkorpus Gesprochenes Deutsch. Retrieved on March 25, 2016 from http://agd.ids-mannheim.de/folk.shtml (last visited on 11.04.2019).
62. Foth, K. A. (2006). Eine umfassende Constraint-Dependenz-Grammatik des Deutschen. [Technical report, Fachbereich Informatik der Universität Hamburg].
63. Fox, A. (2000). *Prosodic features and prosodic structure: The phonology of suprasegmentals*. New York/Oxford: Oxford University Press.
64. Frazier, L. (1985). Syntactic complexity. In: Dowty, D. R., Karttunen, L., & Zwicky, A. M. (Eds.), *Natural Language Parsing*. Cambridge University Press, Cambridge, UK, 129-187.
65. Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*(3), 412.
66. Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological reviews*, *91*(4), 1357-1392.
67. Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. *Neuroimage*, *20*, 8-17.
68. Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, *2*(4), 189-210.
69. Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33), 10336-10341.
70. Gaona, C. M., Sharma, M., Freudenburg, Z. V., Breshears, J. D., Bundy, D. T., Roland, J., ... & Leuthardt, E. C. (2011). Nonuniform high-gamma (60-500 Hz) power changes dissociate cognitive task and anatomy in human cortex. *Journal of Neuroscience*, *31*(6), 2091-2100.
71. Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *Journal of Speech, Language, and Hearing Research*, *51*, 1183-1202.
72. Gibson, J. J. (1950). The perception of the visual world. Boston, MA: Houghton Mifflin.
73. Glanz (Iljina), O., Derix, J., Kaur, R., Schulze-Bonhage, A., Auer, P., Aertsen, A., & Ball, T. (2018). Real-life speech production and perception have a shared premotor-cortical substrate. *Scientific reports, 8*(1), 8898.
74. Grande, M., Meffert, E., Huber, W., Amunts, K., & Heim, S. (2011). Word frequency effects in the left IFG in dyslexic and normally reading children during picture naming and reading. *Neuroimage*, *57*(3), 1212-1220.
75. Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and

length in overt picture naming. *Journal of cognitive neuroscience*, *19*(4), 617-631.

76. Grodzinsky, Y., & Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Current opinion in neurobiology*, *16*(2), 240-246.
77. Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and language*, *96*(3), 280-301.
78. Gussenhoven, C. (1999). On the limits of focus projection in English. *Focus: Linguistic, cognitive, and computational perspectives*, 43-55.
79. Hader, M. (2016). *The Effects of Articulatory Complexity in Neural Activity - An ECoG Study*. [Master thesis in cognitive science, University of Freiburg].
80. Hayashi, Y. (1999). Recognition of vocal expression of emotions in Japanese: using the interjection “eh.”. In: *Proceedings of ICPhS’99, the 14th International Congress of Phonetic Sciences*, pp. 2355-2358.
81. Heilemann, M. (2008). Informationsstrukturierung für die syntaktische Annotation eines diachronen Korpus des Deutschen [Diplomarbeit, Universität Regensburg].
82. Herff, C., Johnson, G., Diener, L., Shih, J., Krusienski, D., & Schultz, T. (2016, August). Towards direct speech synthesis from ECoG: A pilot study. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1540-1543). IEEE.
83. Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135.
84. Hinrichs, E., Hinrichs, M., & Zastrow, T. (2010). WebLicht: Web-based LRT services for German. In: *Proceedings of the ACL 2010 System Demonstrations*, pp. 25-29. Association for Computational Linguistics. https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page (last visited on 11.04.2019).
85. Hirschberg, J., & Pierrehumbert, J. (1986). The Intonational Structuring of Discourse. *Proceedings of the 24th Annual Meeting on Association for Computational Linguistics*, pp. 136-144. Association for Computational Linguistics.
86. Hugdahl, K., Thomsen, T., Ersland, L., Rimol, L. M., & Niemi, J. (2003). The effects of attention on speech perception: an fMRI study. *Brain and language*, *85*(1), 37-48.
87. Iljina, O., Derix, J., Schirrmeister, R. T., Schulze-Bonhage, A., Auer, P., Aertsen, A., & Ball, T. (2017). Neurolinguistic and machine-learning perspectives on direct speech BCIs for restoration of naturalistic communication. *Brain-computer interfaces*, *4*(3), 186-199.
88. Indefrey, P., Hagoort, P., Herzog, H., Seitz, R. J., & Brown, C. M. (2001). Syntactic processing in left prefrontal cortex is independent of lexical meaning. *Neuroimage*, *14*(3), 546-555.
89. Indefrey, P., Hellwig, F., Herzog, H., Seitz, R. J., & Hagoort, P. (2004). Neural responses to the production and comprehension of syntax in identical utterances. *Brain and language*, *89*(2), 312-319.
90. Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101-144.
91. Jacobs, J. (1988). Fokus-hintergrund-gliederung und grammatik. In: Altmann, H. (Ed.): *Intonationsforschungen 200*, pp. 89-134. Tübingen: Niemeyer.
92. Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824.
93. Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274(5284), 114-116.

94. Kang, A. M., Constable, R. T., Gore, J. C., & Avrutin, S. (1999). An event-related fMRI study of implicit phrase-level syntactic and semantic processing. *Neuroimage*, *10*(5), 555-561.
95. Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic inquiry*, *8*(1), 63-99.
96. Kang, S. (2012). *Neural basis of the word frequency effect and its relation to lexical processing*. [UC Berkeley Phonology Lab Annual Report (2012), http://linguistics.berkeley.edu/phonlab/documents/2012/Kang_word_freq.pdf (last visited on 11.04.2019)].
97. Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. *Cerebral cortex*, *11*(3), 223-237.
98. Kelly, M. H., Bock, J. K., & Keil, F. C. (1986). Prototypicality in a linguistic context: Effects on sentence structure. *Journal of memory and language*, *25*(1), 59-74.
99. Kern, M., Aertsen, A., Schulze-Bonhage, A., & Ball, T. (2013). Heart cycle-related effects on event-related potentials, spectral power changes, and connectivity patterns in the human ECoG. *Neuroimage*, *81*, 178-190.
100. Kern, M., Bert, S., Glanz, O., Schulze-Bonhage, A., & Ball, T. (2019). Human motor cortex relies on sparse and action-specific activation during laughing, smiling and speech production. *Communications biology, 2*(1), 118.
101. Kipp, A., Wesenick, B., & Schiel, F. (1997). Pronunciation Modeling Applied to Automatic Segmentation of Spontaneous Speech. *Fifth European Conference on Speech Communication and Technology*, pp. 1023-1026. Rhodes.
102. Kirchner, R. (1998). *An effort-based approach to consonant lenition*. [Doctoral dissertation, University of California, Los Angeles].
103. Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean *length of utterance in morphemes. Journal of Child Language, 12*(2), 251-269.
104. *Klumpp, A.-K. (2016). Effekte von Frequenzen und Übergangswahrscheinlichkeiten bei Bigrammen in der spontanen Sprachproduktion - Eine ECoG-Studie*. [Master thesis in German linguistics, University of Freiburg].
105. Kortmann, B., & Szmrecsanyi, B. (2009). “World Englishes between simplification and complexificationeenIn: Siebers, L. & Hoffmann, T. (Eds.): *World Englishes. Problems – Properties – Prospects*, pp. 265 ems. Philadelphia: Benjamins.
106. Kreitewolf, J., Friederici, A. D., & von Kriegstein, K. (2014). Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *Neuroimage*, *102*, 332-344.
107. Kusters, W. (2008). Complexity in linguistic theory, language learning and language change. In: Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.), *Language Complexity: Typology, Contact, Change*, pp. 3-22. Amsterdam: Benjamins.
108. Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*, 159-74.
109. Lau, D. (2016). *Investigating the neuronal correlates of prosody during real-world speech production using ECoG*. [Bachelor thesis in biology, University of Freiburg].
110. Levelt, W. J. (1999). Models of word production. *Trends in cognitive sciences*, *3*(6), 223-232.
111. Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, *74*, 431-461.

112. Levelt, W. J., & Meyer, A. S. (2000). Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology*, *12*(4), 433-452.
113. Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22(*1), 1-38.
114. Maitz, P. (2014). Sprachwandel und sprachliche Komplexität. In: Agel, V. & Gardt, A. (Eds.): *Jahrbuch für Germanistische Sprachgeschichte*, *5*(1), 94-108, Berlin, Boston: de Gruyter.Maitz, P., & Németh, A. (2014). Language contact and morphosyntactic complexity: evidence from German. *Journal of Germanic Linguistics, 26*, 1-29.
115. McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44*(3), 295-322.
116. McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746.
117. Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006-1010.
118. Miestamo, M. (2006a). On the feasibility of complexity metrics. In: Kerge, K., and Sepper, M.-M. (Eds.), *Finest Linguistics. Proceedings of the Annual Finnish and Estonian Conference of Linguistics. Tallinn, May 6-7, 2004*, Tallinn: TLÜ.
119. Miestamo, M. (2006b). On the complexity of standard negation. In: Suominen, M. et al. (Eds.), *A Man of Measure: Festschrift in Honour of Fred Karlsson on His 60th Birthday* [Special Supplement to SKY Journal of Linguistics 19], pp. 345-356.
120. Miller, J. F., Neufang, M., Solway, A., Brandt, A., Trippel, M., Mader, I., ... & Kahana, M. J. (2013). Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science, 342*(6162), 1111-1114.
121. Mugler, E. M., Patton, J. L., Flint, R. D., Wright, Z. A., Schuele, S. U., Rosenow, J., ... & Slutzky, M. W. (2014). Direct classification of all American English phonemes using signals from functional speech motor cortex. *Journal of neural engineering*, *11*(3), 035015.
122. Murray, W. S., & Forster, K. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, 111, 721-756.
123. Münte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed class words: class and frequency effects. *Neuropsychologia*, *39*(1), 91-102.
124. Navarrete, E., Basagni, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production?. *Quarterly Journal of Experimental Psychology*, *59*(10), 1681-1690.
125. Neuromedical AI Lab (2019). Linguistische Konventionen zur Erstellung des “The Freiburg/First Neurolinguistic Corpus“. https://www.ieeg.uni-freiburg.de/FNLC%20corpus (last visited on 17.03.2021)].
126. Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*, 273-281.
127. Oppenrieder, W. (1987). Aussagesätze im Deutschen. *Satzmodus zwischen Grammatik und Pragmatik*, In: Meibauer, J. (Ed.), pp. 161-189. Tübingen: Niemeyer.
128. Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: an fMRI study based on individual differences. *Neuroimage*, *21*(2), 623-631.
129. Paeschke, A., Kienast, M., & Sendlmeier, W. F. (1999). F0-contours in emotional speech. In: *Proc. 14th Int. Congress of Phonetic Sciences, 2*, pp. 929-932.

130. Pei, X., Barbour, D. L., Leuthardt, E. C., & Schalk, G. (2011). Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *Journal of neural engineering*, *8*(4), 046028.
131. Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*(4), 389-443.
132. Percival, D. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge, New York: Cambridge University Press.
133. Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [Doctoral dissertation, Massachusetts Institute of Technology].
134. Pfänder, S., Behrens, H., & Auer, P. (2013). Erfahrung zählt. Frequenzeffekte in der Sprache: ein Werkstattbericht. *Zeitschrift für Literaturwissenschaft und Linguistik*, *43*(1), 7-32.
135. Pfurtscheller, G. (1992). Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. *Electroencephalography and clinical neurophysiology*, *83*(1), 62-69.
136. Pfurtscheller, G., & Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, *110*(11), 1842-1857.
137. Pheby, J. (1975). *Intonation Und Grammatik Im Deutschen*. Berlin: Akademischer Verlag.
138. Pistohl, T., Schulze-Bonhage, A., Aertsen, A., Mehring, C., & Ball, T. (2012). Decoding natural grasp types from human ECoG. *Neuroimage*, *59*(1), 248-260.
139. Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing: activation interacts with task demands. *Neuroimage*, *17*(1), 401-410.
140. Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816-847.
141. Prabhakaran, R., Blumstein, S. E., Myers, E. B., Hutchison, E., & Britton, B. (2006). An event-related fMRI investigation of phonological – lexical competition. *Neuropsychologia*, *44*(12), 2209-2221.
142. Prince, E. (1981). Toward a taxonomy of Given-New information. In: Cole, P. (Ed.), *Radical pragmatics*, pp. 223-255. New York: Academic Press.
143. Przyrembel, M., Smallwood, J., Pauen, M., & Singer, T. (2012). Illuminating the dark matter of social neuroscience: considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Frontiers in human neuroscience*, *6*, 190.
144. Pulvermüller, F., & Fadiga, L. (2016). Brain language mechanisms built on action and perception. In: Hickok, G., & Small, S. L. (Eds.), *Neurobiology of Language*, pp. 311-324. Amsterdam: Elsevier.
145. Ramsey, N. F., Salari, E., Aarnoutse, E. J., Vansteensel, M. J., Bleichner, M. G., & Freudenburg, Z. V. (2018). Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *Neuroimage*, *180*, 301-311.
146. Rice, G. A., & Robinson, D. O. (1975). The role of bigram frequency in the perception of words and nonwords. *Memory & Cognition*, *3*(5), 513-518.
147. Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural neurology*, *12*(4), 191-200.
148. Ruescher, J., Iljina, O., Altenmüller, D. M., Aertsen, A., Schulze-Bonhage, A., & Ball, T. (2013). Somatotopic mapping of natural upper-and lower-extremity movements and speech production with high gamma electrocorticography. *Neuroimage*, *81*, 164-177.
149. Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, *18*(4), 367-379.

150. Sagae, K., Lavie, A., & MacWhinney, B. (2005). Automatic measurement of syntactic development in child language. In: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 197-204, Association for Computational Linguistics.
151. Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, *25*(23), 3079-3085.
152. Shankweiler, D., & Harris, K. S. (1966). An experimental approach to the problem of articulation in aphasia. *Cortex*, *2*(3), 277-292.
153. Schiel, F. (2015). A Statistical Model for Predicting Pronunciation. *Proceedings of the International Conference on Phonetic Sciences, 195*. Glasgow, UK.
154. Schiller, A., Teufel, S., & Thielen, C. (1995). *Guidelines für das Tagging deutscher Textkorpora mit STTS*. [Technical report, Universität Stuttgart and Universität Tübingen, http://ylvi.sfs.uni-tuebingen.de/fileadmin/static/ascl/resources/stts-1995.pdf (last visited on 11.04.2019)].
155. Scholz, U. (1987). Wunschsätze im Deutschen - formale und funktionale Beschreibung. In: Meibauer, J. (Ed.): *Satzmodus zwischen Grammatik und Pragmatik*, pp. 234-258. Tübingen: Niemeyer.
156. Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J. R., Bergmann, P., Birkner, K., Couper-Kuhlen, E., Deppermann, A., Gilles, P., & Günthner, S. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung: Online-Zeitschrift Zur Verbalen Interaktion, 10*, 353-402.
157. Senkowski, D., & Herrmann, C. S. (2002). Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. *Clinical Neurophysiology*, *113*(11), 1742-1753.
158. Sheskin, D. J. (2007). *Handbook of Parametric and Nonparametric Statistical Procedures, 4th edition*. Boca Raton: Chapman & Hall.
159. Siegert, I., Böck, R., & Wendemuth, A. (2014). Inter-rater reliability for emotion annotation in human-computer interaction: comparison and methodological improvements. *Journal on Multimodal User Interfaces, 8*(1), 17-28.
160. Sinai, A., Bowers, C. W., Crainiceanu, C. M., Boatman, D., Gordon, B., Lesser, R. P., ... & Crone, N. E. (2005). Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. *Brain*, *128*(7), 1556-1570.
161. Sinnemäki, K. (2008). Complexity trade-offs between the subsystems of language. In: Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.), *Language Complexity: Typology, Contact, Change*, pp. 67-88. Amsterdam: Benjamins.
162. Solari, M. E. (1969). The “maximum likelihood solution” of the problem of estimating a linear functional relationship. *Journal of the Royal Statistical Society: Series B (Methodological)*, *31*(2), 372-375.
163. Steinschneider, M., Nourski, K. V., & Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special?. *Hearing research*, *305*, 57-73.
164. Strijkers, K., Costa, A., & Thierry, G. (2009). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral cortex*, *20*(4), 912-928.
165. Szmrecsanyi, B. (2004). On operationalizing syntactic complexity. In: *Le poids des mots. Proceedings of the 7th international conference on textual data statistical analysis, Louvain-la-Neuve, 2*, pp. 1032-1039.
166. Szmrecsanyi, B., & Kortmann, B. (2012). Introduction: Linguistic complexity e Second Language Acquisition, indigenization, contact, In: Kortmann, B. & Szmrecsanyi, B. (Eds.),

Linguistic Complexity: Second Language Acquisition, Indigenization, Contact, pp. 6-34. Berlin / Boston: Walter de Gruyter.

167. Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2005). Neural Circuitry Underlying Sentence-Level Linguistic Prosody. *Neuroimage*, 28(2), 417-428.
168. Towle, V. L., Yoon, H. A., Castelle, M., Edgar, J. C., Biassou, N. M., Frim, D. M., ... & Kohrman, M. H. (2008). ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain*, *131*(8), 2013-2027.
169. Truong, K. P., Neerincx, M. A., & Van Leeuwen, D. A. (2008). Assessing agreement of observer-and self-annotations in spontaneous multimodal emotion data. In: *Interspeech 2008 – Eurospeech*, pp. 318-321. Brisbane, Australia.
170. Truong, K. P., Van Leeuwen, D. A., & De Jong, F. M. (2012). Speech-based recognition of self-reported and observed emotion in a dimensional space. *Speech communication*, *54*(9), 1049-1063.
171. Uhmann, S. (1991). *Fokusphonologie*. Tübingen: Niemeyer.
172. Van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, *39*(4), 973-978.
173. Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Research on Language & Social Interaction*, *13*(2), 201-277.
174. Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & cognition*, *18*(4), 380-393.
175. von Heusinger, K. (1999). *Intonation and Information Structure*. [Habilitation thesis, University of Konstanz].
176. von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research*, *17*(1), 48-55.
177. Walker, M. B., & Trimboli, C. (1982). Smooth transitions in conversational interactions. *The Journal of Social Psychology*, *117*(2), 305-306.
178. Wang, X., Gkogkidis, C. A., Iljina, O., Fiederer, L. D., Henle, C., Mader, I., ... & Ball, T. (2017). Mapping the fine structure of cortical activity with different micro-ECoG electrode array geometries. *Journal of neural engineering*, *14*(5), 056004.
179. Wang, N. X., Olson, J. D., Ojemann, J. G., Rao, R. P., & Brunton, B. W. (2016). Unsupervised decoding of long-term, naturalistic human neural recordings with automated video and audio annotations. *Frontiers in human neuroscience*, *10*, 165.
180. Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech, 46*(1), 53-81.
181. Wellmer, J., Weber, C., Mende, M., Von Der Groeben, F., Urbach, H., Clusmann, H., ... & Helmstaedter, C. (2009). Multitask electrical stimulation for cortical language mapping: hints for necessity and economic mode of application. *Epilepsia*, *50*(10), 2267-2275.
182. Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., & Ackermann, H. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral cortex*, *14*(12), 1384-1389.
183. Wise, R. J. S., Greene, J., Büchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *The Lancet*, *353*(9158), 1057-1061.
184. Wittgenstein, L. (1922 [repr. online (no year)]). *Tractatus Logico-Philosophicus*, http://tractatus-online.appspot.com/Tractatus_en.html (last visited on 06.08.2019).

185. www.brainsignals.de (no year), [online source of shared neural data (last visited on 22.08.2019)].
186. Ziegler, W., & Aichert, I. (2015). How much is a word? Predicting ease of articulation planning from apraxic speech error patterns. *Cortex*, *69*, 24-39.
187. Zipf, G. K. (1935). *The psycho-biology of language: an introduction to dynamic philology*. Boston: Houghton Mifflin.

Appendix 1:
Supplementary Tables 1-3

The here listed tables provide information about additional analyses conducted in terms of Study 1 which are not presented in the text of the thesis proper but which may nevertheless be helpful to gain a detailed picture of the conducted analyses and of the statistical properties of the linguistic data analyzed.

Supplementary Table 1: The success of matching words with a focus accent (FA) and words with no focus accent (nFA) for each subject and PoS. S1-4: subjects 1-4, PoS: part of speech, ADV: adverbs, FV: full verbs, NN: nouns, № w.: number of words, contr.: contrast, max: the number of words in each condition after the first matching attempt using the maximum possible number of trials, m(n)FA: the number of matched words in each of the FA and nFA conditions from all three PoS combined; grey fields with numbers of trials for PoS indicate significant differences between FA and nFA conditions prior to matching (column titles FA and nFA, respectively) and after matching (column titles starting with "max"), green fields at the respective PoS indicate that successful matching could be achieved using the number of trials in these fields. # indicates that no further matching was necessary, since FA and nFA categories could be matched before the matching iteration indicated by the title of the respective column.

contr.	PoS	total	FA	nFA	max	max-10	max-20	max-30	max-40	max-50	m(n)FA
FA/nFA						**S1**					
	ADV	286	76	210	76	66	#	#	#	#	
	FV	389	133	256	133	123	113	103	93	83	
	NN	127	64	63	63	53	#	#	#	#	
	№. w.	802	273	529							202
						S2					
	ADV	248	48	200	48	#	#	#	#	#	
	FV	302	122	180	122	112	102	#	#	#	
	NN	100	57	43	43	#	#	#	#	#	
	№. w.	650	227	423							193
						S3					
	ADV	240	59	181	59	49	#	#	#	#	
	FV	327	110	217	110	100	#	#	#	#	
	NN	120	62	58	58	48	#	#	#	#	
	№. w.	687	231	456							197
						S4					
	ADV	535	160	375	160	150	140	#	#	#	
	FV	741	221	520	221	211	201	191	#	#	
	NN	317	171	146	146	136	#	#	#	#	
	№. w.	1593	552	1041							467

Supplementary Table 2: The success of matching words with an emotional FA (emoFA) and words with no emotional FA (NemoFA) for each subject and PoS. All conventions as in Suppl. Tab. 1.

contr.	PoS	FA	emoFA	NemoFA	max	max-1	max-2	max-3	max-5	m(N)emoFA
emoFA/NemoFA					**S1**					
	ADV	66	31	8	8	#	#	#	#	
	FV	83	11	26	11	#	#	#	#	
	NN	53	12	16	12	#	#	#	#	
	№. w.	202	54	50						31
					S2					
	ADV	48	10	8	8	#	#	#	#	
	FV	102	22	24	22	#	#	#	#	
	NN	43	9	19	9	#	#	#	#	
	№. w.	193	41	51						39
					S3					
	ADV	49	20	11	11	#	#	#	#	
	FV	100	21	32	21	#	#	#	#	
	NN	48	12	7	7	#	#	#	#	
	№. w.	197	53	50						39
					S4					
	ADV	140	50	17	17	#	#	#	#	
	FV	191	35	49	35	#	#	#	#	
	NN	136	21	39	21	20	19	18	16	
	№. w.	467	106	105						68

Supplementary Table 3: Details on the success of matching FA and nFA words for each subject, PoS, and control parameter. Grey fields with a "T" ("true") show that differences between the FA and nFA conditions with regard to the given parameter were significant (Wilcoxon rank sum test, p<0.05). Green fields with an "F" ("false"), on the contrary, indicate that no significant differences were present (successful matching). Bold font highlights the table lines with successful matching iterations. Abbreviations: subj.: subject, FRQ: lemma frequency, med.: median, № ph.: number of phonemes, № occ.: number of word occurrences in the subcorpus of the given subject, w. str.: position of the syllable with the word stress, ss_ws: time interval between speech start and word start in ms, ws_we: same for word duration, we_se: same for word end to speech end duration, p: p-value, H: hypothesis. All numbers were rounded to the second decimal. "m0": statistical comparison before matching, other conventions as in Tab. 2.

subj.		ADV										NN											VV						
	m. iterat.	FRQ (p)	FRQ (H)	med. FRQ FA	med. FRQ nFA	№ (p)	№ (H)	med. № FA	med. № nFA	№ occ. (p)	№ occ. (H)	med. № FA	med. № nFA	w. str. (p)	w. str. (H)	med. w. FA	med. w. nFA	ss_ws (p)	ss_ws (H)	med. ss_ws FA	med. ss_ws nFA	ws_we (p)	ws_we (H)	med. ss_ws FA	med. ss_ws nFA	we_se (p)	we_se (H)	med. we_se FA	med. we_se nFA
S1	m0	0	T	371.5	3657	0	T	4	3	0	T	2	9	0	T	1	1	0.65	F	739.75	848.63	0	T	426.76	239.26	0	T	472.17	1308.59
	max	0.29	F	371.5	664	0.88	F	4	4	0.25	F	2	3	0.27	F	1	1	0.69	F	739.75	781.74	0.03	T	426.76	337.4	0.07	F	472.17	693.85
	max-	**0.26**	**F**	**364**	**557.5**	**0.75**	**F**	**4**	**4**	**0.38**	**F**	**2**	**3**	**0.23**	**F**	**1**	**1**	**0.97**	**F**	**800.29**	**815.43**	**0.11**	**F**	**426.76**	**351.07**	**0.06**	**F**	**472.17**	**693.85**
S2	m0	0.04	T	717.5	8882	0.7	F	4	3	0	T	4.5	9	0	T	1	1	0.93	F	718.26	709.47	0	T	292.48	231.93	0.11	F	582.52	800.78
	max	**0.36**	**F**	**717.5**	**1254**	**0.26**	**F**	**4**	**4**	**0.4**	**F**	**4.5**	**5**	**0.55**	**F**	**1**	**1**	**0.84**	**F**	**718.26**	**696.29**	**0.8**	**F**	**292.48**	**275.39**	**0.71**	**F**	**582.52**	**586.43**
S3	m0	0	T	431	4792	0.01	T	3	3	0	T	2	8	0.64	F	1	1	0.99	F	643.55	719.73	0	T	471.68	263.67	0	T	525.39	1123.05
	max	0.09	F	431	1338	0.78	F	3	3	0.01	T	2	4	0.46	F	1	1	0.79	F	643.55	690.43	0.01	T	471.68	336.91	0.35	F	525.39	605.47
	max-	**0.36**	**F**	**779**	**1338**	**0.66**	**F**	**3**	**3**	**0.11**	**F**	**3**	**4**	**0.65**	**F**	**1**	**1**	**0.8**	**F**	**678.71**	**680.66**	**0.35**	**F**	**406.25**	**352.54**	**0.49**	**F**	**586.91**	**608.4**
S4	m0	0	T	195.5	4298	0	T	4	3	0	T	2	15	0	T	1	1	0.04	T	676.76	628.91	0	T	447.27	237.3	0	T	402.83	982.42
	max	0	T	195.5	460	0.28	F	4	4	0.11	F	2	3	0.72	F	1	1	0.36	F	676.76	806.64	0	T	447.27	340.33	0.02	T	402.83	599.61
	max-	0.02	T	248	460	0.69	F	4	4	0.28	F	2	3	0.65	F	1	1	0.39	F	678.71	791.02	0.02	T	397.95	356.45	0.18	F	476.07	594.24
	max-	**0.05**	**F**	**312**	**460**	**0.72**	**F**	**4**	**4**	**0.25**	**F**	**3**	**3**	**0.73**	**F**	**1**	**1**	**0.43**	**F**	**687.01**	**795.41**	**0.09**	**F**	**365.72**	**356.45**	**0.26**	**F**	**480.96**	**594.24**
		FV																											
S1	m0	0	T	407	5527.5	0	T	5	4	0	T	2	4	0.01	T	1	1	0.73	F	594.73	633.3	0	T	500	280.27	0	T	375	841.31
	max	0.02	T	407	1246	0.01	T	5	5	0.77	F	2	2	0.45	F	1	1	0.24	F	594.73	837.89	0	T	500	384.77	0.13	F	375	578.13
	max-	0.06	F	435	1348	0.06	F	5	5	0.87	F	2	2	0.95	F	1	1	0.4	F	599.61	832.03	0	T	479.49	404.3	0.15	F	352.54	550.78
	max-	0.1	F	543	1348	0.11	F	5	5	0.9	F	2	2	0.82	F	1	1	0.57	F	626.95	822.27	0.01	T	470.7	406.25	0.25	F	352.54	456.05
	max-	0.23	F	543	1246	0.26	F	5	5	0.92	F	2	2	0.87	F	1	1	0.65	F	626.95	806.64	0.02	T	470.7	404.3	0.38	F	376.95	420.9
	max-	0.28	F	650	1246	0.38	F	5	5	0.87	F	2	2	0.85	F	1	1	0.73	F	626.95	792.97	0.04	T	464.84	407.23	0.39	F	352.54	419.92
	max-	**0.28**	**F**	**771**	**1491**	**0.5**	**F**	**5**	**5**	**0.92**	**F**	**2**	**2**	**1**	**F**	**1**	**1**	**0.73**	**F**	**626.95**	**792.97**	**0.07**	**F**	**458.98**	**399.41**	**0.59**	**F**	**376.95**	**416.02**
S2	m0	0	T	1047	6096	0	T	5	3	0	T	2	5	0.56	F	1	1	0.32	F	532.71	479.98	0	T	343.75	223.14	0	T	508.79	751.95
	max	0.02	T	1047	1402.5	0.06	F	5	4	0.49	F	2	2	0.52	F	1	1	0.42	F	532.71	493.16	0	T	343.75	257.32	0.11	F	508.79	718.26
	max-	0.11	F	1047	1348	0.22	F	5	4	0.75	F	2	2	0.5	F	1	1	0.27	F	519.04	527.83	0.04	T	309.08	260.25	0.47	F	577.15	688.48
	max-	**0.14**	**F**	**1047**	**1297**	**0.28**	**F**	**5**	**4**	**0.5**	**F**	**2**	**2**	**0.56**	**F**	**1**	**1**	**0.34**	**F**	**519.04**	**527.83**	**0.06**	**F**	**301.76**	**257.81**	**0.98**	**F**	**602.54**	**643.55**
S3	m0	0	T	1457	4959	0	T	4	3	0	T	2	6	0.73	F	1	1	0.14	F	637.21	590.82	0	T	424.8	236.33	0	T	522.46	1194.34
	max	0.23	F	1457	1509	0.24	F	4	4	0.96	F	2	2	0.18	F	1	1	0.44	F	637.21	776.86	0.02	T	424.8	343.26	0.01	T	522.46	898.93
	max-	**0.35**	**F**	**1509**	**1509**	**0.48**	**F**	**4**	**4**	**0.92**	**F**	**2**	**2**	**0.22**	**F**	**1**	**1**	**0.46**	**F**	**687.99**	**884.28**	**0.27**	**F**	**396.97**	**385.74**	**0.1**	**F**	**595.7**	**802.73**
S4	m0	0	T	753	4959	0	T	5	3	0	T	2	5	0	T	1	1	0.19	F	557.62	613.77	0	T	482.42	277.34	0	T	517.58	813.96
	max	0.14	F	753	849	0.27	F	5	5	0.81	F	2	2	0.88	F	1	1	0.24	F	557.62	677.73	0	T	482.42	395.51	0.86	F	517.58	600.59
	max-	0.12	F	753	849	0.29	F	5	5	0.99	F	2	2	0.72	F	1	1	0.48	F	571.29	628.91	0.01	T	478.52	393.55	0.84	F	521.48	594.73
	max-	0.24	F	753	830	0.34	F	5	5	0.92	F	2	2	0.69	F	1	1	0.58	F	589.84	650.39	0.03	T	478.52	397.46	0.66	F	543.95	558.59
	max-	**0.32**	**F**	**753**	**849**	**0.46**	**F**	**5**	**5**	**0.57**	**F**	**2**	**2**	**0.83**	**F**	**1**	**1**	**0.67**	**F**	**589.84**	**614.26**	**0.1**	**F**	**441.41**	**397.46**	**0.69**	**F**	**551.76**	**564.45**
		NN																											
S1	m0	0	T	30.5	93	0.03	T	6	5	0.09	F	1	1	0.76	F	1	1	0.11	F	930.18	1315.43	0	T	570.8	476.56	0.46	F	723.14	786.13
	max	0	T	34	93	0.04	T	6	5	0.12	F	1	1	0.79	F	1	1	0.08	F	923.83	1315.43	0	T	573.24	476.56	0.49	F	725.59	786.13
	max-	**0.25**	**F**	**47**	**62**	**0.34**	**F**	**6**	**5**	**0.23**	**F**	**1**	**1**	**0.69**	**F**	**1**	**1**	**0.16**	**F**	**898.44**	**1357.42**	**0.09**	**F**	**553.71**	**482.42**	**0.85**	**F**	**759.77**	**697.27**
S2	m0	0.52	F	20	37	0.03	T	6	5	0.79	F	1	1	0.19	F	1	1	0.48	F	745.12	771.48	0.01	T	475.59	333.98	0.32	F	477.54	823.24
	max	**0.41**	**F**	**31**	**37**	**0.14**	**F**	**5**	**5**	**0.37**	**F**	**1**	**1**	**0.3**	**F**	**1**	**1**	**0.38**	**F**	**695.31**	**771.48**	**0.08**	**F**	**419.92**	**333.98**	**0.51**	**F**	**524.41**	**823.24**
S3	m0	0.02	T	34	139	0.35	F	5	5	0.14	F	1	2	0.68	F	1	1	0.69	F	977.05	1101.07	0.11	F	554.69	466.31	0	T	360.84	948.24
	max	0.02	T	35.5	139	0.53	F	5	5	0.25	F	1.5	2	0.98	F	1	1	0.82	F	1017.58	1101.07	0.23	F	546.39	466.31	0	T	424.8	948.24
	max-	**0.14**	**F**	**36**	**96.5**	**0.87**	**F**	**5**	**5**	**0.57**	**F**	**1.5**	**2**	**1**	**F**	**1**	**1**	**0.95**	**F**	**1053.71**	**1098.14**	**0.83**	**F**	**486.33**	**487.3**	**0.14**	**F**	**618.16**	**811.04**
S4	m0	0.02	T	15	61	0	T	6	5	0.64	F	1	1	0.79	F	1	1	0.18	F	872.07	930.66	0	T	545.9	477.05	0.01	T	549.8	797.36
	max	0.19	F	21.5	61	0.05	F	5	5	0.37	F	1	1	0.54	F	1	1	0.2	F	874.51	930.66	0.01	T	519.53	477.05	0.09	F	607.91	797.36
	max-	**0.76**	**F**	**20.5**	**35.5**	**0.39**	**F**	**5**	**5**	**0.26**	**F**	**1**	**1**	**0.37**	**F**	**1**	**1**	**0.49**	**F**	**889.16**	**930.66**	**0.12**	**F**	**517.09**	**477.54**	**0.41**	**F**	**663.57**	**712.4**

Appendix 2:
Calculation of the ease-of-articulation index

The methodological approach documented in this appendix describes our approach to calculation of the ease-of-articulation (EoA) index. Regardless of the bleak neuroanatomical findings related to this parameter, our calculation is presented here for the reader to be able to follow, how exactly the parameter was generated. Note that the calculation of the EoA index in this thesis was not conducted from the scratch but that it builds upon Marina Hader's Master Thesis (2016), who made first attempts toward investigating the neural infrastructure supporting this phenomenon in collaboration with the author of the present thesis[1]. The description below contains borrowed and modified examples, tables and formulae from Marina Hader's work. Marina Hader gave her consent for the presentation of these materials in this Appendix.

---

[1] See „Methods" for details on how EoA calculations in the present thesis differ from those by Hader (2016).

Application of the model of articulatory complexity by Ziegler and Aichert (2015)

We estimated the articulatory complexity of the words using a mathematical model by Ziegler and Aichert (2015). This is a tree-structure model that describes the hierarchical embedding of vocal-tract gestures in single words. It was developed to predict the accuracy of word articulation in patients with aphasia of speech (AoS). Ziegler and Aichert (2015) calculated the likelihood of correct word articulation by accounting for the number of accurate articulations out of the total number of articulations of a set of words in 33 AoS patients. They identified a set of linguistic parameters relevant to the accuracy of word production on several levels of linguistic abstraction: consonant clusters, syllabic and prosodic structure levels. Based on them, these authors constructed a non-linear regression model, and trained it in predicting the likelihood of correct word articulation. In a cross-validation trial with a different group of 40 AoS patients, the model was able to predict this likelihood with a high accuracy ($R^2_{adj.} = 0.67$). Its application to estimate articulatory complexity is therefore plausible. AoS is “an impairment of the capacity to program the movements of the articulators for the purpose of speaking” (Ziegler, 2008, p. 269). One can therefore assume that the probability of correct articulation estimated with help of Ziegler and Aichert’s model, further referred to as the *ease-of-articulation index* (EoA), reflects the requirements of motor planning of words in speech production.

The model by Ziegler and Aichert (2015) builds on the theory of articulatory phonology (e.g., Ohala et al., 1986; Goldstein and Fowler, 2003). This theory sees vocal-tract gestures, or discrete actions of articulatory organs (the lips, the tongue, the velum and the glottis), as basic units of articulation that are combined during the production of segmental components of speech. Columns 2-6 of Suppl. Tab. 4 provide an overview of what articulatory organs are involved in the production of German consonants (summarized based on Kortmann (2005) and Ziegler and Aichert, (2015)).

ph.	lips	t. tip	t. back	vel. apert.	glot. apert.	compl. constr.
p	1				1	
b	1					
t		1			1	
d		1				
k		1			1	
g		1				
N			1	1		
m	1			1		
n		1		1		
l		1				1
R			1			1
r		1				1
f	1				1	1
v	1					1
s		1			1	1
z		1				1
S		1			1	1
Z		1				1
j			1			
x			1		1	1
h						1

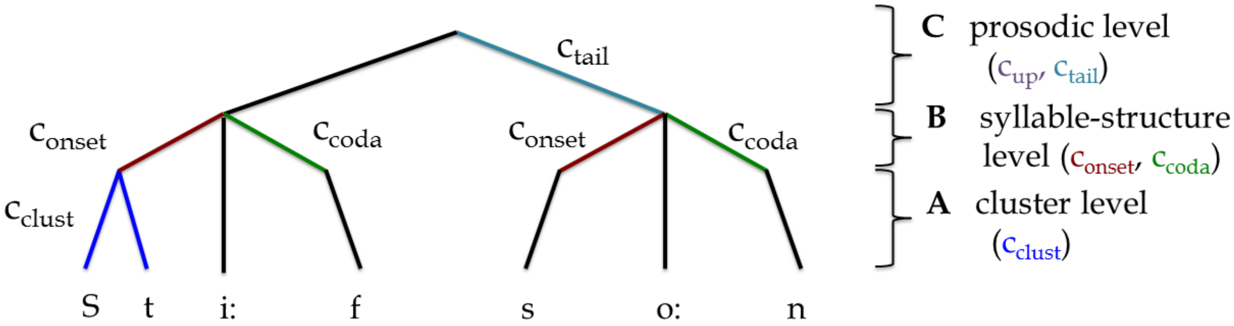
Supplementary Table 4: Articulatory features relevant to the production of German consonants. Abbreviations: ph.: phoneme, t.: tongue, vel./glot. apert.: velar/glottal aperture, compl. constr.: complex constriction. The consonants are listed using the conventions of the Speech Assessment Methods Phonetic Alphabet (SAMPA; Wells, 1995). “1” indicates the presence of the articulatory feature in the articulation of the phoneme; an empty field indicates its absence. Reproduced with permission from Hader (2016).

Execution of articulatory gestures in the stream of speech does not start from an invariant, zero position of articulators in the vocal tract. Rather, it depends on the immediate phonological context. A new articulatory gesture can require different movements for the production of one and the same phoneme, depending on its antecedent (the so-called "co-articulation effect" (Liberman, Delattre, and Cooper (1952)). Several lines of evidence from psycholinguistic research (summarized in Ziegler and Aichert (2015)) suggest that such variability in speech is possible owing to the phonological planning of articulatory gestures in hierarchical, non-linear constellations, which Ziegler and Aichert (2015) assume as the basis of their EoA model. This model adheres to the following other principles. It (1) accommodates two kinds of tongue gestures: those of the tongue tip and those of the tongue back, (2) treats all vowels and diphthongs as involving exactly one gesture of the vocal tract, (3) treats all consonants as involving one gesture of either the lips, the tongue tip, or the tongue body plus a possible additional glottal aperture in voiceless phonemes or a velar aperture in nasal phonemes. Finally, it (4) treats fricatives and lateral and rhotic sounds as "complex" and plosives, nasals and vowels as "simple" constriction types (the last column of Suppl. Tab. 4).

Ziegler and Aichert's model (2015) assumes a probability *p* of correct articulation for all gestures of the vocal tract. It uses a number of weighting coefficients to account for the characteristics of the individual gestures as well as for gestural embedding (described below). Glottal apertures are weighted with the coefficient c_{glot} and velar apertures with the coefficient c_{vel}. When involved in the production of complex-constriction-type phonemes, the primary articulators, namely, the lips, the tongue tip, and the tongue back, are weighted with the coefficient c_{cnstr}. For the phoneme /S/, e.g., the gesture of the tongue tip is weighted with its occurrence in a fricative (c_{cnstr}), and the glottal aperture is weighted with c_{glot}. The probability of correct articulation of a combination of *n* gestures is obtained by multiplying the weighted probabilities $p_1, p_2, \dots p_n$ for each gesture. The probability $p_{/S/}$ of correct articulation of /S/ is therefore (reproduced with permission from Hader, 2016):

$$p_{/S/} = \underbrace{(p_1 \times c_{cnstr})}_{\text{tongue-tip gesture}} \times \underbrace{(p2 \times \; c_{glot})}_{\text{glottal aperture}}$$

Ziegler and Aichert's (2015) model specifies the relations between vocal-tract gestures on three higher levels of gestural embedding: (A) the level of consonant clusters, (B) the level of syllable structure, and (C) the level of prosody (Suppl. Fig. 1).



Supplementary Figure 1. A schematic illustration of the hierarchical relations between vocal-tract gestures relevant to the calculation of the ease of articulation index (EoA) on three levels of gestural embedding in Ziegler and Aichert's model (2015), illustrated on the example of the word "Stiefsohn" (Engl. "stepson"). **(A)**, Articulatory gestures that are part of a consonant cluster are weighted with c_{clust} on the cluster level. **(B)** They are weighted on the level of syllable structure for the occurrence in the onset or coda position within the syllable. **(C)** On the prosodic level, articulatory gestures are weighted with the coefficients c_{up} or c_{tail}, depending on their accentual-syllabic meter. See Suppl. Tab. 5 for definitions of these coefficients. Colored bars of the tree structure indicate that the respective coefficient needs to be applied in the EoA calculation of "Stiefsohn". Reproduced with permission from Hader (2016).

(A) Articulatory gestures that occur in a consonant cluster are additionally weighted with c_{clust}. Note that the model by Ziegler and Aichert (2015) accounts for the fact that articulation of a phoneme depends on its phonological neighborhood: If two adjacent phonemes involve the same primary articulatory organ (columns 2-4 of Suppl. Tab. 4) and have the same constriction complexity (last column of Suppl. Tab. 4), the model counts the shared articulatory gesture not twice but once (a half for each phoneme). The likelihood of accurate articulation for /St/ in the word “Stiefsohn” (Engl. “stepson”) is thus weighted as follows (modified with permission from Hader, 2016):

$$p_{/\mathrm{St}/} = \underbrace{\underbrace{(p_1 \times c_{cnstr} \times c_{clust})}_{\text{tongue-tip gesture}} \times \underbrace{(p_2 \times c_{glot} \times c_{clust} \times 0.5)}_{\text{glottal aperture}}}_{/\mathrm{S}/} \times \underbrace{\underbrace{(p_3 \times c_{glot} \times c_{clust} \times 0.5)}_{\text{glottal aperture}} \times \underbrace{(p_4 \times c_{clust})}_{\text{tongue-tip gesture}}}_{/\mathrm{t}/}$$

The gesture for devoicing, the glottal aperture gesture, is shared between /S/ and /t/, and it therefore multiplied by 0.5 for each phoneme. The tongue-tip gestures of /S/ and /t/ are not shared, since the type of constriction complexity changes from /S/ (complex) to /t/ (simple). The vocal-tract gestures that are not part of a consonant cluster do not receive additional weights on this level. On the level of syllable structure (B), all articulatory gestures in the syllable onset position, i.e., before the nucleus of the syllable, are weighted with c_{onset}. The probability of correct articulation of the phoneme /S/ in the word “Stiefsohn” is therefore (modified with permission from Hader, 2016):

$$p_{/S/} = \underbrace{(p_1 \times c_{cnstr} \times c_{clust} \times c_{onset})}_{\text{tongue-tip gesture}} \times \underbrace{(p_2 \times c_{glot} \times c_{clust} \times c_{onset} \times 0.5)}_{\text{glottal aperture}}$$

All gestures that occur in the coda position of a syllable are weighted with c_{coda}. The nucleus does not receive an additional weight on the level of syllable structure. On the prosodic level (C), the weighting of articulatory gestures depends on the word’s accentual-syllabic meter. The gestures that occur in the stressed syllable of an iambic, trochaic or amphibrach foot are never weighted. Articulatory gestures that appear in the weak syllable of a trochaic foot are weighted with c_{tail}, and gestures in the weak syllable of an iambic foot are weighted with c_{up}. In an amphibrach foot, the first syllable is analyzed as the upbeat syllable and the last syllable as the tail syllable with the respective c_{up} and c_{tail} coefficients. Once the contributions of each articulatory gesture within a word have been accounted for on all three levels of linguistic abstraction (Suppl. Fig. 1), the weighted probabilities are multiplied to calculate the probability of correct articulation for the entire word.

The weighting coefficients used to specify the relationships described above are provided in Suppl. Tab. 5, reproduced with permission from Ziegler and Aichert (2015). These authors obtained the coefficients using a model trained to predict the probability of correct articulation in AoS patients and validated them in a different group of 40 AoS patients based on a set of 48 words and 48 non-words; The model yielded a good estimate for accurate articulations, reaching an $R^2_{\text{adj.}}$ of 0.67 in the cross-validation trial. The suitability of their model and of these coefficients to describe the EoA of words is thus plausible.

coefficient	estimate	standard error	significance level[a]
$\mathbf{p}$	0.970	0.019	ns
$\mathbf{c}_{\mathbf{onset}}$	0.859	0.039	0.01
$\mathbf{c}_{\mathbf{coda}}$	0.896	0.036	0.05
$\mathbf{c}_{\mathbf{glot}}$	1.191	0.051	0.01
$\mathbf{c}_{\mathbf{vel}}$	1.188	0.058	0.01
$\mathbf{c}_{\mathbf{clust}}$	0.984	0.021	ns
$\mathbf{c}_{\mathbf{cnstr}}$	0.902	0.026	0.01
$\mathbf{c}_{\mathbf{up}}$	0.916	0.018	0.01
$\mathbf{c}_{\mathbf{tail}}$	1.010	0.017	ns

[a] 0.05/0.01: 95%/99% confidence interval (CI) excludes "1"; ns: 95% CI includes "1".

Supplementary Table 5: Coefficient estimates for the full regression model, reproduced with permission from Hader (2016), there adapted with permission from Ziegler and Aichert (2015).

References:

1. Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, *88*(1), 148-159.
2. Hader, Marina (2016). *The Effects of Articulatory Complexity in Neural Activity - An ECoG Study*. [Master thesis in Cognitive science at the University of Freiburg].
3. Kortmann, B. (2005). *English linguistics: essentials*. Berlin: Cornelsen.
4. Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The Role of Selected Stimulus Variables in the Perception of the Unvoiced Stop Consonants. *American Journal of Psychology, 65*, 497-516.
5. Ohala, J. J., Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *3*(1), 219-252.
6. Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. In: Gibbon, D., Moore, R., & Winski, R. (Eds.), *Handbook of standards and resources for spoken language systems*, part IV, section B. Berlin / New York: Mouton de Gruyter.
7. Ziegler, W. (2008). Apraxia of speech. *Handbook of Clinical Neurology*, *88*, 269-285.
8. Ziegler, W., & Aichert, I. (2015). How much is a word? Predicting ease of articulation planning from apraxic speech error patterns. *Cortex*, *69*, 24-39.