

UNIVERSITY OF OLDENBURG

Investigating the Specificity of Pre-Speech Auditory Modulation - From Global Gating to Selective Silence?

Author:

TIM DREßLER
Bachelor of Science
Student ID: 6530387

Examiners:

PROF. DR. STEFAN DEBENER
Ammerländer Heerstr. 114-118
26129 Oldenburg

PROF. DR. ANDREA HILDEBRANDT
Ammerländer Heerstr. 114-118
26129 Oldenburg

Submitted for the degree of
Master of Science
August 7, 2025

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Prof. Dr. Stefan Debener and Prof. Dr. Andrea Hildebrandt, for believing in me and in this project. I am truly thankful for their guidance, insightful feedback, and constructive criticism throughout its course. I am especially grateful to Prof. Dr. Stefan Debener for granting me access to the resources of the Neuropsychology Lab to carry out my experiment - and also for taking me along to bike races.

My heartfelt thanks also go to all members of the Neuropsychology Lab and the PMuS Lab for their support. I am particularly thankful to Dr. Manuela Jäger, Reiner Emkes, Micha Burkhardt, Dr. Mareike Daeglau, and Melanie Klapprott for their invaluable assistance throughout the project. I would like to extend my sincere thanks to Dr. Heiko Stecher from the Experimental Psychology Lab, Prof. Dr. Roozbeh Behroozmand from the University of Texas at Dallas and Sean Kinahan from Arizona State University for their valuable comments and insights, which greatly enriched this work. I would also like to thank the Wübben Stiftung Wissenschaft for their support throughout my studies.

A big thank-you goes to the student assistants involved in this project - Gubse Aydin, Roxana Seeger, and Sila Simsek - for their help with data collection. I'm also deeply grateful to my fellow students and friends for their unwavering support - whether it was assisting with data collection, keeping an eye on my endlessly running scripts, proofreading, or simply offering emotional support. Thank you, Alex, Jana, Linn, Malin, Shari and Sophie.

Finally, I want to thank my family for their continuous support - not only during this project but throughout my life. Thank you to my dad, Sandra, grandma Maria and grandpa Otto, grandma Waltraud, grandpa Heinz, and Harry. A very special thank-you goes to my girlfriend Rebekka, for her patience, understanding, and encouragement - even when I stayed up late or when we had to pilot the experiment for a 3rd time. Last but by no means least, I want to thank my mum for being the incredibly supportive and awesome human being she is.

Für Mama.

Contents

Acknowledgements	i
Contents	ii
List of Figures	v
List of Tables	vi
Acronyms	vii
Abstract	viii
Zusammenfassung	ix
1 Introduction	1
1.1 Internal Forward Models	1
1.2 Speaking-induced Suppression	3
1.2.1 Precision of Speaking-induced Suppression	4
1.2.2 Behavioural & Clinical Relevance of Speaking-induced Suppression	6
1.3 Pre-Speech Processing	7
1.4 Pre-Speech Auditory Modulation	8
1.4.1 Precision of Pre-Speech Auditory Modulation	8
1.4.2 Behavioural & Clinical Relevance of Pre-Speech Auditory Modulation	11
1.5 The Current Study	12
2 Methods	15
2.1 Participants	15
2.2 Materials	15
2.2.1 EEG System	15
2.2.2 Audiovisual System	16
2.2.3 Questionnaires	17
2.2.4 Stimuli	17
2.3 Task	18
2.4 Procedure	20
2.5 Analysis	21
2.5.1 Behavioural Preprocessing	21
2.5.2 EEG Preprocessing	21

2.5.3	Exclusion Criteria	22
2.5.4	Behavioural Analysis	23
2.5.5	ERP Analysis	24
2.5.6	Classification Analysis	26
2.6	Use of Artificial Intelligence	31
2.7	Preregistration & Code Availability	31
3	Results	33
3.1	Vocal Responses & Behavioural Data	33
3.2	Pre-Speech Auditory Modulation	34
3.3	Single-Trial Classification	41
4	Discussion	43
4.1	Revisiting Pre-Speech Auditory Modulation	43
4.1.1	Attention as a Confounding Factor	46
4.1.2	Load as a Modulator of Attentional Effects	47
4.1.3	Concluding Remarks on Attentional Modulation	49
4.1.4	Lack of Specificity in Pre-Speech Auditory Modulation	50
4.2	Inter- and Intraindividual Differences During Speech Preparation	51
4.3	Limitations	53
4.4	Conclusion	54
References		56
Appendix A	N1 Amplitude by Task Condition, Probe Type and Probe Onset	75
Appendix B	PSAM Effect by Probe Type and Probe Onset	76
Appendix C	N1 Latency by Task Condition, Probe Type and Probe Onset	77
Appendix D	Individual Hyperparameter Accuracy Heatmaps	79
Appendix E	Filter Characteristics	86
Appendix F	Grand Average Uncorrected, Corrected and Control ERPs	87
Appendix G	SAM Ratings over Time	88
Appendix H	NASA-TLX Ratings	89

Appendix I Assessment of Initial State	90
Appendix J SAM	92
Appendix K NASA-TLX	93
Appendix L Participant Information	95
Appendix M Consent Form	102
Eigenständigkeitserklärung	105

List of Figures

1	EEG channel layout	16
2	Experimental task	19
3	Single-trial classification pipeline	27
4	Standardised probe F0 values for altered and unaltered probes & Average F0 value by probe type and probe onset	35
5	Standardised F0 distributions and probe F0 values across participants	35
6	N1 amplitude by task condition, probe type and probe onset	37
7	Grand average corrected ERPs and topographies for all conditions	38
8	PSAM Effect by probe type and probe onset	39
9	N1 latency by task condition, probe type and probe onset	40
10	Percentage of possible hyperparameter-pairs leading to an above-chance classification for early and late time-windows	41
11	Grand average hyperparameter accuracy heatmaps	42
D.1	Individual hyperparameter accuracy heatmaps for sub-01, sub-03, sub-04 and sub-05	79
D.2	Individual hyperparameter accuracy heatmaps for sub-06, sub-08, sub-09 and sub-10	80
D.3	Individual hyperparameter accuracy heatmaps for sub-11, sub-12, sub-13 and sub-14	81
D.4	Individual hyperparameter accuracy heatmaps for sub-15, sub-16, sub-17 and sub-19	82
D.5	Individual hyperparameter accuracy heatmaps for sub-20, sub-21, sub-22 and sub-23	83
D.6	Individual hyperparameter accuracy heatmaps for sub-24, sub-25, sub-27 and sub-28	84
D.7	Individual hyperparameter accuracy heatmaps for sub-29, sub-30, sub-31 and sub-33	85
E.8	Filter characteristics of the preregistered pipeline	86
E.9	Filter characteristics of the final pipeline	86
F.10	Grand average uncorrected, corrected and control ERPs	87
G.11	SAM Ratings over time	88
H.12	NASA-TLX Ratings	89

List of Tables

A.1	N1 amplitude by task condition, probe type, and probe onset	75
A.2	Planned post hoc tests - N1 amplitude	75
B.3	PSAM Effect by probe type and probe onset	76
B.4	Planned post hoc tests - PSAM Effect	76
C.5	N1 latency by task condition, probe type, and probe onset	77
C.6	Planned post hoc tests - N1 latency	78

Acronyms

ATS	ANOVA-Type-Statistic
CD	Corollary Discharge
CNV	Contingent Negative Variation
EC	Efference Copy
ECoG	Electrocorticography
EEG	Electroencephalogram
EOG	Electrooculography
ERP	Event Related Potential
F0	Fundamental Frequency
FAL	Fragebogen zur Ausgangslage
FIR	Finite Impulse Response
ICA	Independent Component Analysis
ICC	Intraclass Correlation Coefficient
IFG	Inferior Frontal Gyrus
ITC	Inter-Trial-Coherence
LMM	Linear Mixed-Effects Model
MEG	Magnetoencephalography
NASA-TLX	National Aeronautics and Space Administration Task Load Index
PCA	Principal Component Analysis
PSAM	Pre-Speech Auditory Modulation
RBF	Radial Basis Function
rMANOVA	Repeated Measures Analysis of Variance
RMS	Root Mean Square
RP	Readiness Potential
SAM	Self-Assessment Manikin
SIS	Speaking-induced Suppression
SNR	Signal-to-Noise-Ratio
SVM	Support Vector Machine
SZ	Schizophrenia
tDCS	Transcranial Direct Current Stimulation
TMS	Transcranial Magnetic Stimulation

Abstract

Background: The theory of internal forward models suggests that self-generated actions are accompanied by an efference copy (EC) which is used to predict the sensory consequences of actions, leading to a corollary discharge (CD) in the corresponding sensory areas. As a result, neural responses to self-generated stimuli are suppressed. In the speech-auditory domain this phenomenon is known as speaking-induced suppression (SIS). Crucially, SIS depends on a match between expected and actual auditory feedback. Recent findings suggest that a similar suppression occurs during speech preparation, termed pre-speech auditory modulation (PSAM). However, the characteristics of PSAM remain mostly unknown.

Objective: The current study investigated whether PSAM is specific to the expected vocal outcomes, or whether it reflects a general gating mechanism. The study also investigated whether the specificity depends on the phase of the speech preparation and how speech preparation processes change over time.

Methods: Neural activity was recorded via Electroencephalogram (EEG) while participants ($n = 28$) were presented with individually recorded vocal probes - either unaltered or pitch-shifted - during early or late stages of speech preparation. The PSAM effect, defined as the difference in probe-evoked N1 amplitudes between speech preparation and passive listening, was examined for its dependence on probe type and onset. Additionally, EEG data from early and late preparation phases were used to classify active versus passive trials using Support Vector Machines (SVMs).

Results: In contrast to the hypothesised positive PSAM effect (i.e., suppressed N1 amplitudes during speech preparation), results revealed enhanced N1 amplitudes during speech preparation compared to passive listening - a negative PSAM effect that was not significantly influenced by probe type or probe onset. The analysis of the SVM model complexity suggested no difference between early and late preparation phases, indicating relatively stable speech preparation processes within the investigated time-window.

Conclusion: The findings indicate a lack of specificity in the PSAM effect, suggesting it may not reflect prediction-dependent mechanisms. Furthermore, the results emphasise the possible impact of attentional biases inherent in active-passive paradigms and the potential influence of inter- and intraindividual variability on neural response patterns.

Zusammenfassung

Hintergrund: Die Theorie interner Vorwärtsmodelle besagt, dass selbstinitiierte Handlungen von einer Efferenzkopie (EC) begleitet werden, die genutzt wird um sensorische Konsequenzen vorherzusagen und zu einem Corollary Discharge (CD) in sensorischen Arealen führt. Dadurch werden neuronale Reaktionen auf selbstgenerierte Reize unterdrückt - im Sprachbereich als Speaking-induced Suppression (SIS) bekannt. Diese Suppression hängt von der Übereinstimmung zwischen erwartetem und tatsächlichem Feedback ab. Eine vergleichbare Suppression tritt bereits während der Sprechvorbereitung auf und wird als Pre-Speech Auditory Modulation (PSAM) bezeichnet. Ihre Eigenschaften sind bisher kaum untersucht.

Zielsetzung: Diese Studie untersucht, ob die PSAM einen unspezifischen Gating-Mechanismus widerspiegelt oder spezifisch für das erwartete Feedback ist, und ob diese Spezifität von der Phase der Sprechvorbereitung abhängt. Außerdem wird untersucht, wie sich Sprechvorbereitungsprozesse im Zeitverlauf verändern.

Methoden: Den Teilnehmenden ($n = 28$) wurden in frühen oder späten Phasen der Sprechvorbereitung individuell aufgezeichnete Sprachproben - unverändert oder tonhöhenverändert - präsentiert, während Elektroenzephalogramm (EEG) aufgezeichnet wurde. Der PSAM-Effekt, definiert als die Differenz in den N1-Amplituden zwischen Sprechvorbereitung und passivem Hören, wurde auf seine Abhängigkeit vom Proben-Typ und Proben-Onset untersucht. Zusätzlich wurden EEG-Daten aus frühen und späten Vorbereitungsphasen genutzt, um aktive versus passive Trials mithilfe von Support Vector Machines (SVMs) zu klassifizieren.

Ergebnisse: Im Gegensatz zum erwarteten positiven PSAM-Effekt (d.h. unterdrückte N1-Amplituden während der Sprechvorbereitung) zeigten die Ergebnisse eine verstärkte N1-Amplitude während der Sprechvorbereitung im Vergleich zum passiven Hören – ein negativer PSAM-Effekt, der weder signifikant vom Proben-Typ noch vom Proben-Onset beeinflusst wurde. Die Analyse der Komplexität der SVM Modelle zeigte keinen Unterschied zwischen früher und später Vorbereitung, was auf relativ stabile Sprechvorbereitungsprozesse innerhalb des untersuchten Zeitfensters hinweist.

Zusammenfassung: Die Ergebnisse deuten auf eine fehlende Spezifität des PSAM-Effekts hin, was nahelegt, dass er keine vorhersageabhängigen Mechanismen widerspiegelt. Darüber hinaus unterstreichen die Resultate den möglichen Einfluss aufmerksamkeitsbedingter Verzerrungen in Aktiv-Passiv Paradigmen sowie die potenzielle Rolle inter- und intraindividueller Variabilität.

1 Introduction

When reading this thesis, you likely experience an inner voice articulating the words in your mind. But how do you know that the voice you 'hear' is your own? And why does your voice always sound the same every time you speak? Though these questions may seem trivial at first, conditions such as stuttering and Schizophrenia (SZ) reveal the consequences when these mechanisms malfunction. In stuttering, disrupted vocal control is thought to arise from difficulties in the timing and coordination between speech planning and auditory feedback, leading to repetitions, prolongations, or blocks in speech (Bradshaw et al., 2021; Max & Daliri, 2019). In SZ, a failure to properly tag self-generated speech may result in auditory hallucinations - perceiving one's own (internal) voice as coming from an external source (Frith, 2019). These phenomena highlight the complexity of speech, which demands not only motor processing but also the continuous modulation and integration of the auditory system. The interaction between motor and sensory systems has long been of scientific interest (Sperry, 1950; von Holst & Mittelstaedt, 1950), yet the precise mechanisms remain incompletely understood. One influential framework that seeks to explain this interaction is the theory of internal forward models (Blakemore, Wolpert, & Frith, 2000; Wolpert, 1997). Building on the assumptions of this framework, the present work aims to further explore the complex interplay between motor planning and auditory processing.

1.1 Internal Forward Models

In the framework of internal forward models, it is proposed that motor actions are accompanied by an efference copy (EC; von Holst and Mittelstaedt, 1950), which is used to predict the sensory consequences (corollary discharge; CD; Sperry, 1950) of the motor act (Blakemore, Wolpert, & Frith, 2000; Wolpert, 1997). The predicted sensory consequences are then compared to the actual sensory feedback. In case of self-generated actions, the sensory consequences can be accurately predicted. Therefore, the difference between predicted and actual sensory consequences is minimal, and the resulting sensory input can be suppressed (e.g., Arjmandi and Behroozmand, 2024; Ford and Mathalon, 2019; Frith, 2019; Guenther et al., 2006; Hickok, 2012; Straka et al., 2018; Wolpert, 1997). It is important to note that, although animal studies indicate that suppression of sensory responses (also) involves EC-based mechanisms in the periphery, the suppression observed in human studies using Electroencephalogram (EEG) and Electrocorticography (ECoG) (see below) is more likely to originate from cen-

tral mechanisms (Brooks & Cullen, 2019; Frith, 2019; Horváth, 2015; Horváth & Burgýán, 2013). Interestingly, the neurobehavioural relevance of such forward models seem to be multifaceted. By using internal forward models, the brain possibly makes computation more efficient (Ford & Mathalon, 2005). Furthermore, sensory systems often face the problem that self-generated actions interfere with sensory processing. The suppression of self-generated sensations is therefore necessary in order to perceive the world as we perceive it (Aitchison & Lengyel, 2017; Reznik & Mukamel, 2019; Straka et al., 2018). For example, notice that the eyes 'jump' when looking out of the window of a moving train while the vision itself remains stable (Frith, 2019). The inclusion of internal forward models in sensorimotor processing also solves the problem associated with an inevitable delay between a movement and the sensory feedback associated with the movement (Frith, 2019). When grasping an object, the visual feedback is delayed (i.e., the limb has already reached the target, but the visual perception to confirm the action arrives after it has happened). If the feedback is predicted, the destabilizing effect of feedback delay can be overcome. Notably, such internal forward models could also be used to prevent sensory desensitization (Poulet & Hedwig, 2002; Reznik & Mukamel, 2019; Straka et al., 2018) and enhance sensory processing (Enikolopov et al., 2018; Singla et al., 2017). Moreover, internal forward models have been linked to two additional phenomena. In the auditory/speech domain, internal forward models are associated with online error monitoring of speech (Chang et al., 2013; Guenther et al., 2006; Houde & Chang, 2015). When presented with modified (e.g., pitch-shifted) feedback, participants adjust their voice within milliseconds. For example, when participants' vocal feedback was altered (e.g., pitched down), they automatically adjusted their voices to speak at a higher pitch to compensate for the feedback manipulation (Arjmandi & Behroozmand, 2024; Chang et al., 2013; Greenlee et al., 2013; Houde & Jordan, 1998; Jones & Munhall, 2000; H. Liu & Larson, 2007; Scheerer & Jones, 2018; Subramaniam et al., 2018; Tourville et al., 2008; Zhang et al., 2024). In cases of altered feedback, the expected auditory feedback does not match the prediction, resulting in an error. This error might subsequently be used to update the motor plan, resulting in a compensatory response (D. Liu et al., 2023; H. Liu et al., 2011; Scheerer & Jones, 2018). Further, the described mechanism might be used to distinguish between self-generated and external sensations (Behroozmand & Larson, 2011; Behroozmand et al., 2011; Beño-Ruiz-de-la-Sierra et al., 2023; Blakemore, Wolpert, & Frith, 2000; Chang et al., 2013; Ford & Mathalon, 2005; Heinks-Maldonado et al., 2006, 2007; Reznik & Mukamel, 2019; Straka et al., 2018; Tian & Poeppel, 2012; von Holst & Mittelstaedt, 1950). During self-

generated actions, the sensory consequences can be accurately predicted, 'tagging' the action and its sensory consequences as self-generated and resulting in sensory suppression. One might argue that external events can be accurately predicted as well and thus further argue that this property alone is not sufficient to distinguish self-generated sensations from external ones. However, mere predictability seems to not solely explain the effects of sensory suppression, indicating a special role of self-generated actions with respect to sensory suppression (Mifsud et al., 2016; Oestreich et al., 2015; Weiss & Schütz-Bosbach, 2012).

It is worth noting that with alternative explanations emerging (e.g., Brown et al., 2013), the presence of internal forward models is still under debate (Dogge et al., 2019; Kiepe et al., 2021; Korka et al., 2022; Press et al., 2023; Roussel et al., 2013). Nevertheless, despite differing perspectives, one core phenomenon associated with the theory of internal forward models - and others - remains well-established: The suppression of sensory responses associated with self-generated actions (Frith, 2019).

1.2 Speaking-induced Suppression

In the speech/auditory domain, the suppression of sensory responses associated with self-generated actions can be demonstrated in so-called talk/listen paradigms (Ford et al., 2010). In these paradigms, participants run through two conditions: In the talk condition, participants are asked to produce speech whilst hearing themselves. In the listen condition, participants passively listen to recordings of their speech. Although variations of this paradigm are used (e.g., Whitford et al., 2011), the bottom line remains the same. The theory now predicts that the sensory response to self-generated sounds (i.e., the talk condition) will be suppressed compared to the sensory response to externally generated sounds (i.e., the listen condition). Human neurophysiological evidence for the described mechanism comes primarily from Event Related Potential (ERP) research. A wide range of studies showed suppressed N1 ERP components during trials in which participants had to talk compared to trials in which they merely listened to auditory stimuli (Behroozmand and Larson, 2011; Behroozmand et al., 2011; Beño-Ruiz-de-la-Sierra et al., 2023; Ford et al., 2001; Ford and Mathalon, 2005; Ford et al., 2007, 2021; Heinks-Maldonado et al., 2005, 2007; Kudo et al., 2004; Mathalon and Ford, 2008; Sitek et al., 2013; J. Wang et al., 2014, for review, see Whitford, 2019). Results from Magnetoencephalography (MEG) show a similar pattern (Beal et al., 2010; Heinks-Maldonado et al., 2006; Houde et al., 2002; Niziolek et al., 2013; Ventura et al., 2009). This effect was often named speaking-induced suppression (SIS) and is believed to represent the ac-

tion of EC/CD-related mechanisms (Whitford, 2019). In a modification of the talk/listen paradigm during which participants had to produce the sound not by talking but by pressing a button, similar results were shown (Aliu et al., 2009; Bäß et al., 2008; Cao et al., 2017; Duggirala et al., 2024; Pinheiro et al., 2018, 2020; Timm et al., 2014; Whitford et al., 2011). Interestingly, the N1 suppression also occurred when participants used inner speech (Ford & Mathalon, 2004; Jack et al., 2019). Results from imaging studies show that the N1 ERP component is generated in the auditory cortex (Zouridakis et al., 1998), thus the suppression of the N1 component may reflect the suppression of the auditory cortex during self-generated sounds as proposed by the theory of internal forward models. This interpretation is supported by animal work in mice (Rummell et al., 2023) and primates (Eliades & Wang, 2003, 2008, 2019; Tsunada & Eliades, 2020) that showed attenuated responses in the auditory cortex to self-generated sounds. Lastly, also invasive work in humans using ECoG showed similar suppression effects in auditory regions during self-generated speech (Behroozmand et al., 2016; Flinker et al., 2010; Ozker et al., 2024; for review, see Schneider and Mooney, 2018).

1.2.1 Precision of Speaking-induced Suppression

With the finding of SIS being very common across species and methodological approaches, one might interpret this finding as strong evidence for the presence of an internal forward model as described above. However, one might also argue that the mentioned results could be explained by a general attenuation mechanism instead of a predictive one. In other words, one would have to show that, in fact, the suppression only occurs in case of a match between the expected sensory consequences and the actual feedback. To test this hypothesis, one prominent approach is to alter the auditory feedback presented to the participants (e.g., Behroozmand et al., 2011). In line with the hypothesis of internal forward models, Heinks-Maldonado et al. (2005) showed that the N1 SIS is most pronounced for unaltered feedback as compared to pitch-shifted feedback and alien feedback. Further studies showed similar results (Heinks-Maldonado et al., 2006, 2007; Houde et al., 2002). Next to altering the spectral characteristics of the feedback, other studies used delayed feedback. For example, Aliu et al. (2009) and Whitford et al. (2011) showed that suppression occurs only if the feedback is not delayed. A 50 ms delay was enough to cancel the suppression. Interestingly, even the N1 SIS during inner speech seems to be timing- and content-specific (Jack et al., 2019). Together, these results hint at a precise forward model that is used to predict the sensory consequences of self-generated actions. In case of a match between prediction and feedback, the sensory response is suppressed

(Heinks-Maldonado et al., 2005, 2007; for application to the tactile domain, see Blakemore, Smith, et al., 2000; Blakemore, Wolpert, and Frith, 2000). The N1 component is therefore thought to reflect the comparison between prediction and feedback (Behroozmand & Larson, 2011; Scheerer & Jones, 2018; Tian & Poepel, 2015). Using ECoG, Ozker et al. (2022) showed that there is an error signal in the auditory cortex that scales with the amount of feedback delay, suggesting that the auditory error signal encodes not only a match or a mismatch, but also the amount of mismatch. Similarly, Tourville et al. (2008) suggested that the expected consequences of articulations and the actual feedback are compared in the posterior superior temporal cortex. However, it was also argued that while the suppression may be predictive, it might be relatively coarse (Duggirala et al., 2024). Behroozmand et al. (2011) found significant SIS effects for 0 cent (unaltered), 50 cent, 100 cent, 200 cent, but not 400 cent pitch-shifts. Thus, the errors generated by the internal forward model seem to be less sensitive to small deviations potentially introduced by trial-by-trial variability. Others argue that the sensory prediction does not accurately track feedback variability over the same repeated motor task (Niziolek et al., 2013). Instead, Niziolek et al. (2013) propose that the EC resembles a sensory goal, not a sensory prediction. They found that the amount of SIS was dependent on the spectral distance of the produced vocalisation to the median vocalisation. Somewhat similar, Sitek et al. (2013) showed that the SIS is reduced in trials that are more different in their formant structure compared to the previous trial. However, Sitek et al. (2013) did not find a median-centred effect as compared to Niziolek et al. (2013). Recent work in primates has shown that while neural activity is sensitive to natural variability in the animals' vocal acoustics, this sensitivity is much greater when feedback is altered, also suggesting that the sensory prediction is relatively coarse (Eliades & Tsunada, 2024). More importantly, the authors found a correlation between the neural activity and the deviation of a vocal production from an animal's mean production.

In summary, while there is evidence for a precise forward model rather than general suppression (e.g., Heinks-Maldonado et al., 2005), recent work suggests that sensory prediction may be more of a sensory target, relating to an average vocalisation (Eliades & Tsunada, 2024; Niziolek et al., 2013). In other words, while SIS is not due to a general suppression, it is also not due to a highly precise sensory prediction accounting for trial-by-trial variability. Instead, the prediction seems to be relatively coarse, relying on average vocal acoustics. Such a mechanism has several advantages compared to a highly precise one. For example, it could simplify neural connectivity needed to mediate the described effects (Eli-

ades & Tsunada, 2024). Further, more coarse predictions might make the system more stable in case of noisy environments. While the debate about the precision of SIS is still ongoing, recent animal work suggests the presence of multiple mechanisms (Tsunada et al., 2024). Specifically, the authors proposed two roles associated with the CD mechanism, a temporally precise but unspecific suppression that is time-locked to movement onset, and a predictive suppression that is based on specific expected consequences of motor actions. The existence of predictive attenuation and general gating are therefore not mutually exclusive (Kilteni & Ehrsson, 2022; Tsunada et al., 2024).

1.2.2 Behavioural & Clinical Relevance of Speaking-induced Suppression

Although the behavioural relevance of internal forward models seems to be multifaceted, one might question the specific behavioural relevance of the SIS effect. As mentioned above, the SIS effect is thought to mediate the internal distinction between self-generated and external sounds (Behroozmand & Larson, 2011; Behroozmand et al., 2011; Beño-Ruiz-de-la-Sierra et al., 2023; Chang et al., 2013; Ford & Mathalon, 2004, 2005; Heinks-Maldonado et al., 2006, 2007; Jack et al., 2019; Reznik & Mukamel, 2019; Straka et al., 2018; Tian & Poepel, 2012; Yang et al., 2024). Importantly, Timm et al. (2014) found that a movement elicited with Transcranial Magnetic Stimulation (TMS) does not result in suppressed sensory responses. Thus, the authors concluded suppression effects are only present in case of there being an intention, linking the suppression effects to the sense of agency. Based on this assumption, the SIS effect has been of great interest when studying SZ. Multiple studies have shown that the SIS effect is altered in SZ (Ford et al., 2001; Ford and Mathalon, 2004, 2005; Ford et al., 2007, 2021; Heinks-Maldonado et al., 2007; Stephane et al., 2022, for reviews, see Ford and Mathalon, 2019; Frith, 2019; Whitford, 2019; Whitford et al., 2012). Similar results have been shown in button-press tasks (Whitford et al., 2011) and in the tactile domain (Blakemore, Smith, et al., 2000). Also, Pinheiro et al. (2018) showed reduced suppression in individuals with high hallucination proneness using a button-press task. Recently, there have also been endeavours to design treatment strategies based on the results mentioned above. Roach et al. (2019) investigated whether an auditory training could improve the N1 SIS in SZ and found this to be the case. Furthermore, the authors argued that the auditory training may have improved the transmission of the EC due to improved fronto-temporal connectivity. Other studies used Transcranial Direct Current Stimulation (tDCS) to treat symptoms of psychosis (Bose et al., 2019;

Brunelin et al., 2012; Mondino et al., 2015, 2016; Nawani, Bose, et al., 2014; Nawani, Kalmady, et al., 2014; Nayok et al., 2022). Remarkably, the treatment resulted in reduced hallucinations, reduced covert/overt speech misattributions (e.g., Mondino et al., 2015) and intact SIS (e.g., Nawani, Bose, et al., 2014). Furthermore, it has been suggested that the internal prediction is not only used for distinguishing self-generated sound and external ones but also to allow for online vocal monitoring (Chang et al., 2013). When a pitch-shift is applied during a vocalisation, individuals (and primates) tend to correct for the shift (e.g., Eliades and Tsunada, 2023; Jones and Munhall, 2000; Schneider et al., 2018; Tourville et al., 2008). Thus, if the voice is pitched down, participants tend to increase their pitch to compensate for the shift. The association between the SIS effect and this behavioural phenomenon remains mostly unclear, but it has been suggested that the suppression enhances the error sensitivity of the auditory system (Behroozmand et al., 2009, 2015; Eliades and Wang, 2003, 2008; Ozker et al., 2024, but also see, Chang et al., 2013; Tsunada et al., 2024).

1.3 Pre-Speech Processing

Critically for this study, speech not only modulates auditory processing related to the sensory consequences of speech, i.e. the processing of the sound produced by the speech itself, but also affects processing prior to the onset of speech (or movement in general). For example, a study conducted by Chen et al. (2011) showed increased gamma band synchrony between frontal and temporal sites 50ms before speech onset. Further, Ford et al. (2007) reported increased Inter-Trial-Coherence (ITC) in the beta band before the onset of speech. The pre-onset ITC was further related to the N1 SIS seen after onset. Again, similar results were shown in button-press tasks (Mathalon & Ford, 2008). In line with the theories described above, it was shown that there is an information flow approximately 100 ms before speech onset from motor areas to auditory areas, potentially resembling the transmission of the EC (Khalilian-Gourtani et al., 2024). Sites that received more information inflow later showed increased suppression. Similarly, J. Wang et al. (2014) reported increased activity in the inferior frontal gyrus (IFG) and increased coherence between the IFG and auditory areas before the onset of speech, which correlated with the subsequent suppression. Animal studies have shown that the modulation seen during a vocalisation actually begins before the onset of the vocalisation (Clayton et al., 2021; Eliades & Wang, 2003; Tsunada & Eliades, 2020; Tsunada et al., 2024), suggesting that auditory areas may receive copies of motor commands prior to their execution in order to modulate sensory input (Rummell et al., 2023). Taken together, these findings

suggest that the processes associated with the modulation of auditory processing during active speech begin before the onset of speech. However, while the described findings imply the presence of pre-onset processes, they do not answer the question of whether auditory processing is also altered during the pre-onset phase, similar to the described post-onset effects.

1.4 Pre-Speech Auditory Modulation

To investigate this open question, Max et al. (2008) asked participants to perform the following task: At the start of every trial, a word was presented. After 600 ms, the colour of the word turned green. In the speaking condition, participants were asked to speak after this cue occurred (causing them to prepare to speak during the delay period). In the reading condition, participants were asked to silently read the word, whereas in the seeing condition, the participants only had to watch a row of symbols. In one-third of the trials, a (pure) tone probe was presented during the preparation phase (i.e., prior to the onset of the go-signal). Analyses of probe-evoked N1 responses showed that the N1 amplitude is, compared to the other conditions, suppressed during speech preparation. This result was confirmed in subsequent studies (Daliri and Max, 2015a, 2015b, 2016; also see, Mock et al., 2011). The observed effect was called pre-speech auditory modulation (PSAM). Next to affecting the N1 amplitude, speech planning seems to also modulate auditory perception during speech preparation by increasing perceptual thresholds for intensity discrimination (Merrikhi et al., 2018). However, there is also evidence suggesting that auditory modulation may be more nuanced. S. Li et al. (2020) and Yang et al. (2024) differentiate between general preparation, where one prepares to speak without knowing what to say, and specific preparation, where the speech content is already known. The former is associated with suppression of auditory responses (but also see, Zheng et al., 2022), while the latter appears to enhance auditory responses, reflected in increased N1 amplitudes. These partially conflicting findings highlight the need for further investigation of PSAM.

1.4.1 Precision of Pre-Speech Auditory Modulation

Again, one might ask whether the modulation of auditory processing during speech preparation is independent of the auditory probe or whether it has predictive properties, similar to its post-onset counterpart. In other words, is the degree of suppression seen in the preparation phase dependent on the match between the presented probe and the subsequently produced vocalisation? So far, studies in-

vestigating this question directly are very limited. In a study by Daliri and Max (2016), participants completed a similar paradigm to the one described above (Max et al., 2008). However, instead of the probes being pure tones only, the authors used either a 40 ms pure tone or a prerecorded syllable shortened to 40 ms. The results showed a similar N1 modulation for both probe types, in line with a general attenuation account of PSAM (Daliri & Max, 2016; H. Wang et al., 2024). However, Daliri and Max (2016) found a suppressed P2 amplitude for the syllable probe but not for the tone probe. Thus, the authors suggested that speech preparation has a general effect on the early auditory processing but a speech-specific one for later processing stages. Potentially, speech preparation modulates both lower-level populations of auditory neurons and higher-order speech-processing ones. When using a tone probe, only the former modulation is seen. In comparison, when using speech-related probes (such as a syllable), both modulatory influences can be revealed. Alternatively, and most importantly for this study, the authors suggested that the timing may also play a role. The P2 component arises roughly 100 ms after the N1 component. Thus, the P2 component is closer to the movement onset. Therefore, Daliri and Max (2016) suggested that motor commands are progressively refined during the movement planning phase, so that the modulating signals sent to the auditory cortex may also become more specific over time. The neural generators of the P2 component may receive more refined information about the expected input, allowing for a more specific modulation. In line with the latter explanation, the modulation of auditory responsiveness is greatest and more specific in the time period closest to the target and the subsequent vocal response (Mock et al., 2011). However, one has to consider the following when interpreting the results from Daliri and Max (2016). While the syllable probe was speech-related, it was still heavily shortened so that its 'speech properties' were somewhat limited. Moreover, the syllable probe was not related to the later produced word. Thus, the EC did not match either the pure tone probe or the syllable probe. Results from post-onset modulation show that while the N1 suppression is reduced in case of a mismatch, there does not seem to be a significant difference between different levels of mismatch (e.g., Behroozmand et al., 2011). Further, as noted above, the involved mechanisms are suggested to be relatively coarse (Eliades & Tsunada, 2024; Niziolek et al., 2013). Therefore, one might hypothesise that both types of probes were mismatches, therefore leading to a similar (reduced) suppression. Further, the two types of probes had widely varying acoustical properties, making it hard to interpret the effects as the P2 modulation could have also been influenced by basic stimulus properties (Daliri & Max, 2016). These shortcomings are addressed in the present study. In

another recent study conducted by Zheng et al. (2022) investigating this effect, participants were asked to prepare to speak (without presenting them the target). When a cue was shown, the participants were asked to vocalise. Again, during the preparation phase, different probes were presented. The probes included a pure tone, a syllable, and a cough. The authors suggested that the motor signal during the intention phase does not contain specific information about the sound since the auditory neural responses were suppressed to all types of probes. However, the suppression was stronger for speech vs. non-speech sounds. All together, the authors proposed the idea of a generic inhibition that is focused on human speech sounds. The generic effect in this study is not surprising, as the participant could not prepare for a specific sound to produce. Zheng et al. (2022) speculated further that when motor signals become more concrete, the suppression precisely affects the specific auditory target. Using a similar paradigm, S. Li et al. (2020) showed suppressed N1 responses during general preparation when syllable probes were used but enhanced N1 responses during preparation when tones were used. Interestingly, S. Li et al. (2020) also showed that when participants engaged in specific preparation (i.e., preparing to speak while knowing what to say), N1 amplitudes were enhanced (not suppressed) for auditory probes that matched the upcoming vocalisation, but not for those that did not match. This result indicates a specific modulation of auditory responses, but in the opposite direction of the PSAM effect reported in the studies mentioned above (e.g., Daliri and Max, 2016). While concerning latency instead of amplitude, Mock et al. (2011) showed that N1 latencies were shorter when an auditory probe, presented during speech preparation, matched the subsequent vocalisation. This was, however, only the case when the probe was presented during a later preparation phase.

All together, the modulatory effects seen during speech preparation seem to be at least partly dependent on the probe being used. Furthermore, they appear to be influenced by the specificity of preparation. While direct empirical evidence is, to our knowledge, limited to the studies described above, it is still worth taking into account theoretical considerations that are based on related paradigms and results. Further, evidence from other sensory domains is discussed.

When preparing to perform an action, a specific ERP component can be identified shortly before the movement is executed - the Readiness Potential (RP, Shibasaki and Hallett, 2006) (also see Contingent Negative Variation, CNV, Rohrbaugh et al., 1976). It has been suggested that even in inner speech, this signal resembles specific action-effect predictions during the preparation phase that may begin to suppress the activity of sensory responses in anticipation of the predicted stimuli (Chung et al., 2023). In line with that, Ford et al. (2014)

and Reznik et al. (2018) argue that the RP represents both motor planning and the anticipations of sensory consequences, making it a candidate signature for the forward model embedding the expected sensory consequences. Cao et al. (2017) found an increase in alpha power starting 400 ms before onset in an active condition compared to a passive condition, supporting the idea of active inhibition of auditory areas during motor preparation. This is likely to result from top-down mediated predictions in anticipation of the self-generated (and expected) sensory stimulus. The modulation of pre-onset alpha power in the auditory cortex associated with suppressed processing is thought to be triggered by the medial prefrontal cortex, potentially resembling top-down modulations associated with expected sensory consequences (Müller et al., 2015). In the tactile domain, vibrotactile probes that were congruent with the predicted sensations were shown to be more suppressed than incongruent probes, suggesting that a predictive mechanism regulates the strength of suppression based on specific sensorimotor predictions about the future sensory states generated by a movement (Fuehrer et al., 2022). Similarly, Gale, Areshenkoff, et al. (2021), Gale, Flanagan, and Gallivan (2021), Vercillo et al. (2018), and Voss et al. (2008) suggested that motor system tunes the sensory system during the preparation phase based on expected sensory outcomes.

Overall, while theoretical interpretations linked to effects like pre-onset alpha modulation and the RP suggest the availability of specific expectations about sensory consequences during the preparation phase, evidence regarding the specificity of sensory modulation during this phase is sparse and inconclusive. Another factor worth considering is the functional and clinical relevance of the described pre-onset effects, as it potentially offers further evidence regarding their specificity.

1.4.2 Behavioural & Clinical Relevance of Pre-Speech Auditory Modulation

In addition to the unknown specificity of the PSAM effect, the behavioural relevance of this effect is also unclear. As suggested above, one purpose of PSAM may be to prepare the auditory cortex for incoming feedback (Daliri & Max, 2015a, 2016, 2018). In particular, the modulation may play a role in the preparation of the auditory system for its role in vocal feedback control. Strikingly, the PSAM effect was shown to be atypical in stutterers, who are known to show atypical patterns of vocal responses to altered feedback (Daliri & Max, 2015a, 2015b, 2018; Max & Daliri, 2019; Sares et al., 2018, 2020). Further, the preparatory motor activity as indexed by the CNV is significantly higher preceding a flu-

ent word as compared to a stuttered one, hinting at a compensatory mechanism (Vanhoutte et al., 2016). This finding is complemented by a study from Kinahan et al. (2024), who found that a more complex decision boundary is needed for stutterers when classifying trials based on pre-onset activity. The authors suggested that this might be due to an underactivation of the left motor cortex during the preparation phase (Kinahan et al., 2024; Vanhoutte et al., 2016). This may hint at a deficit in predicting sensory consequences via the EC, potentially caused by delays in the motor system or white matter abnormalities affecting the EC transmission, leading to maladaptive corrections that contribute to stuttering (Daliri & Max, 2015a). However, recently, an alternative approach has been suggested. Instead of tuning the auditory cortex to the expected sensory feedback, the modulation seen during the preparation phase might also be due to the auditory cortex being involved in controlling feedforward commands (J. J. Li et al., 2024; H. Wang et al., 2024). In other words, during movement planning, the response of the auditory cortex to probe tones may be attenuated (i.e., PSAM occurs) because certain neuronal populations are involved in evaluating auditory predictions prior to movement onset, aiding to select the most effective motor commands (J. J. Li et al., 2024). In addition to stuttering, auditory modulation and other pre-onset effects during speech preparation have also been shown to be altered in SZ, similar to the abnormalities seen in post-onset effects (Pinheiro et al., 2018, 2020; Yang et al., 2024).

Taking the clinical effects and the assumed behavioural relevance described above into account, the question regarding the specificity of the PSAM remains. If the PSAM is due to the auditory cortex being primed for the expected sensory consequences to enhance vocal feedback control, one could imagine that the effect is specific, suppressing probes that match the subsequent vocalisation (and therefore the EC) the most. Supporting this claim, both vocal feedback control and PSAM are altered in stuttering individuals. However, as the assumptions regarding the behavioural relevance of the PSAM effect remain speculative, one could also imagine the suppression to be general, i.e., unrelated to the expected sensory consequences of the subsequent action.

1.5 The Current Study

Taken together, former studies investigating the specificity of pre-onset auditory modulation show inconsistent results. The present work therefore aims to provide additional empirical evidence using a more refined and controlled approach as compared to previous studies. By presenting individuals with auditory probes that were assumed to match or mismatch the EC during different phases

of the speech preparation, we aim to answer multiple open questions stated in the literature (Daliri & Max, 2016). In the used paradigm participants were required to prepare to speak (and subsequently to speak) or to stay passive. The presented probes were either assumed to match the EC of the subsequent vocalisation, or they were altered by applying a pitch shift, in a manner similar to that used in studies investigating the specificity of (post-onset) SIS (e.g., Heinks-Maldonado et al., 2005). This enabled a controlled analysis of the specificity of the PSAM effect. A previous study (Daliri & Max, 2016) relied on using stimuli that differed in many aspects of their psychophysical properties, making the paradigm less controllable. Next to investigating the PSAM effect by analysing the N1 responses evoked by the auditory probes, we also aim to directly investigate the speech preparation itself, as it has been suggested that its specificity might in turn influence the specificity of the PSAM effect (Daliri & Max, 2016). Together, this study aims to further understand the PSAM effect and its underlying mechanisms. This, in turn, could help to refine treatment approaches based on the framework of internal forward models for disorders such as stuttering or SZ.

Based on the defined goals and the reviewed literature, the following hypotheses were formulated.

H1a: Replicating prior work (e.g., Daliri and Max, 2015b), we hypothesise to show the PSAM effect as indexed by suppressed (i.e., more positive) N1 amplitudes evoked by auditory probes presented during speech preparation compared to passive listening.

H1b: We hypothesise that the PSAM effect will be more pronounced for probes presented during a late preparation phase compared to probes presented during an early preparation phase (Mock et al., 2011).

H1c: Similar to effects seen in post-onset SIS (e.g., Heinks-Maldonado et al., 2005), we assume the PSAM effect to be, to some extent, specific for the expected sensory consequences of the subsequent vocalisation. Thus, we assume the PSAM effect to be more pronounced when the probe is unaltered compared to when it is altered. This effect is thought to be increased (or even only present) when the probe is presented during a late preparation phase.

H2: We hypothesise that the latencies of the N1 ERP component will be shorter for unaltered probes compared to altered probes during speech preparation. No such difference is expected during passive listening.

H3: We hypothesise that speech preparation is more specific in the later stages of preparation. This is hypothesised to be reflected in a more complex decision boundary being required when classifying trials based on EEG data recorded during an earlier stage of preparation compared to EEG data recorded during a later stage of preparation. This, in turn, is indicated by a lower percentage of possible hyperparameter-pairs (γ, C), identified by a grid search approach, leading to above-chance classification of single-trial data when using features that were extracted from the early preparation stage (for details, see Kinahan et al., 2024).

2 Methods

2.1 Participants

Data from 35 healthy adults were collected between 7 May 2025 and 19 June 2025 at the University of Oldenburg. After excluding one participant due to technical problems during the recording and six participants due to too many excluded trials, the final sample included $n = 28$ participants (20 female, 8 male). The mean age was 24.68 years ($SD = 3.26$). Most participants had completed the German Abitur (96.43 %), and the majority were students (92.86 %). Exclusion criteria included uncorrected visual impairments, hearing loss, and psychiatric or neurological disorders, including both current and past stuttering (Daliri & Max, 2018). It was further ensured that participants had obtained sufficient sleep (at least five hours), had not used illegal drugs, had not consumed alcohol on the day of testing, and were not taking medications affecting the central nervous system. To minimize the risk of including individuals with undiagnosed hearing impairments, only participants between the ages of 18 and 40 were included (Arvin et al., 2013). In addition, a minimum German language proficiency of level C1 was required. Participants were recruited via the university's StudIP system. Written informed consent was obtained from all participants prior to the experiment. Afterwards, they received monetary compensation (12 € per hour). The ethics proposal was approved by the Ethics Committee of the University of Oldenburg under reference number Drs.EK/2025/027.

2.2 Materials

2.2.1 EEG System

EEG data were recorded using a stationary setup including a subset of 30 Ag/AgCl electrodes embedded in a 94-channel cap (Easycap, Herrsching, Germany), arranged in a custom equidistant layout. Electrodes E1 to E28 were used for EEG recording, while electrodes E29 and E30 served as Electrooculography (EOG) channels. Although the layout does not follow the standard 10–20 system, electrode E01 corresponds to the Cz location as defined in the 10–20 system. The channel configuration is illustrated in Figure 1. All signals were recorded using BrainAmp amplifiers (Brainproducts GmbH, Gilching, Germany), with a nose tip reference. Data were sampled at 1000 Hz and band-pass filtered between 0.0159 and 250 Hz. Electrodes were prepared with electrolyte gel (Abralyt HiCl, Easy-cap GmbH, Herrsching, Germany), and impedances were reduced to below $20k\Omega$ prior to the start of recording.

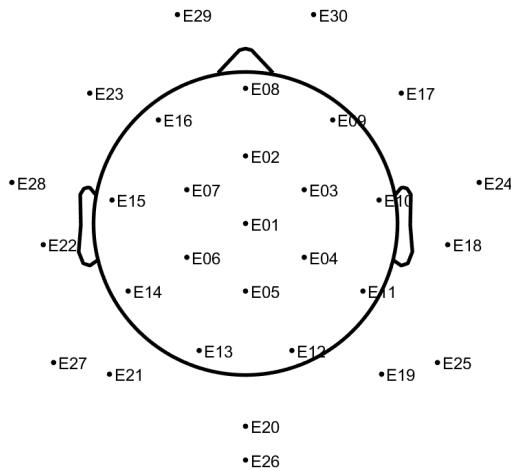


Figure 1: EEG channel layout.

Electrodes E29 and E30 were used as EOG channels.

2.2.2 Audiovisual System

To record the vocal responses made by the participants, a Sennheiser MD 43 (Sennheiser electronic KG, Wedemark, Germany) dynamic microphone was used, which was placed approximately 5 cm in front of each participant's mouth. The microphone was connected via an XLR cable to a windows PC using a Scarllet 2i2 3rd Gen (Focusrite, High Wycombe, United Kingdom) audio-interface. The sampling rate for the recording was set to 44.1 kHz. This set up was used for both the initial stimuli recordings and the experiment itself (see Chapter 2.4). To present the auditory stimuli during the experiment, the set up included an internal HDSP 9632 sound card (RME, Haimhausen, Germany), an external ADI 8 DS MK III D/A converter (RME, Haimhausen, Germany), two PA5 attenuators (Tucker-Davis Technologies, Alachua, United States) and a HB7 Headphone buffer (Tucker-Davis Technologies, Alachua, United States). The HB7 output was routed to a pair of Sirocco S30 loudspeakers (Cambridge Audio, London, United Kingdom) via a C245BEE amplifier (NAD, Pickering, Canada). The loudspeakers were positioned at ear level, approximately 115 cm in front of the participants, angled 45 degrees to the left and right. The sample rate was set to 44.1 kHz with a buffer size of 256. The loudness of the stimuli was individually calibrated in order to reach a pleasant intensity level for each participant (see Chapter 2.4). Visual stimuli were presented on a 24 inch screen with a resolution of 1920 x 1080 and a refresh rate of 60 Hz, which was placed approximately 110 cm in front of the participants. Both the recording and the subsequent presentation of auditory stimuli, as well as the presentation of visual stimuli, was handled using PsychoPy (Version 2024.2.4; Peirce et al., 2019).

2.2.3 Questionnaires

Multiple questionnaires were used throughout the experiment to assess various information. The Assessment of Initial State questionnaire (Fragebogen zur Ausgangslage, FAL) was used to obtain information about the participant's biographical data. It contained questions on general information such as age, sex, sleep, education, and medical history. To measure the emotional state of the participants during the experiment, a custom version of the Self-Assessment Manikin (SAM; Bradley and Lang, 1994) questionnaire was used. Participants rated their emotional state on a scale from 0 (very negative) to 10 (very positive) and their tiredness on a scale from 0 (very alert) to 10 (very tired). Both scales used a step size of one. The SAM includes cartoon manikins to describe the different scale levels. Furthermore, a modified version of the National Aeronautics and Space Administration Task Load Index (NASA-TLX; Flägel et al., 2019) was used to assess workload or perceived strain after the task. Five dimensions were rated on a scale from 0 to 10 in 0.5-point steps. For the dimensions 'mental demand', 'physical demand', 'effort', and 'frustration', 0 corresponded to low and 10 to high levels. The dimension 'performance' was rated differently, with 0 indicating good and 10 indicating bad performance. 'Mental demand' referred to the cognitive demand required for processing information and decision-making. 'Physical demand' described the level of physical activity involved (e.g., pulling, pushing, steering). 'Performance' captured perceived success and satisfaction with task performance. 'Effort' reflected how much effort was needed to meet task demands, and 'frustration' referred to feelings of stress, irritation, or frustration during the task. All used questionnaires are displayed in Appendix I, Appendix J, and Appendix K, respectively.

2.2.4 Stimuli

Each participant's stimuli were individually recorded. Before the main task, the participants were asked to vocalize a short syllable (/ga/) 21 times while being recorded. To ensure that the vocal responses during the stimulus recordings matched those during the experiment as closely as possible, the following trial structure was used for both the initial stimulus recordings and the actual task (see Chapter 2.3). At the beginning of each trial, a fixation cross was presented for 0.5–1.5 s (jittered). Then a circle appeared with '/ga/' written inside. After 3 s, the fill colour of the circle changed to green. This was the go-signal for participants to produce the syllable. As with the main task, participants were instructed to prepare to produce the syllable during the delay period prior to the go-signal.

The software Praat (Version 6.4.27; Boersma and Weenink, 2025) was used to detect on- and offsets, trimming the recording. Subsequently, two probe versions were generated, including an unaltered version with a 0 semitone pitch-shift and an altered version with a -4 semitone pitch-shift. The pitch-shifts were applied using the Praat Vocal Toolkit (Corretge, 2024). The 0 semitone pitch-shift was applied to ensure both versions completed the same number of processing steps. After that, the probes were trimmed to a length of 80 ms, and a 10 ms fade-in and a 10 ms fade-out were applied. Finally, the Fundamental Frequency (F0) of each version was extracted using the filtered-autocorrelation method and both probes were normalized to 70 dB SPL. Next, the (unaltered) probe (out of the 21) with the median F0 was selected. This was done to ensure that the selected (unaltered) probe represents a typical vocalisation of the participant, which was in return assumed to match the EC associated with the vocalisation (Eliades & Tsunada, 2024; Niziolek et al., 2013). The altered (i.e., pitch-shifted) probe, on the other hand, was assumed to result in a mismatch when being compared to the EC. The experimenter listened to the selected probe to ensure it contained no artefacts. If the probe was deemed unusable, the adjacent probe closest to the median was selected - first the one below the median, then the one above. If that probe was also unsuitable, the process was repeated with probes located two steps from the median. If none of these alternatives were usable, no further substitutions were allowed and the recording had to be repeated. Upon successfully completing the stimuli selection process, each participant was left with two individual stimuli, one unaltered probe and one altered (i.e., pitch-shifted) probe. The loudness of the stimuli was individually calibrated for each participant (see Chapter 2.4).

2.3 Task

The main task of the experiment consisted of 8 blocks with 120 trials each. Each trial included the following elements. After a fixation cross was shown for 0.5 - 1.5 s (jittered), a circle appeared on the screen. Within the circle, an instruction was written. The instruction '/ga/' indicated an active trial, while the instruction '/xx/' indicated a passive trial. After a 3 s delay, the fill colour of the circle changed to green. This constituted the go-signal. The green circle remained on the screen for 1.5 s. After that, there was a 0.5 s delay and then a new trial began. Depending on the instruction, participants were either asked to only observe this process passively (passive task condition) or they were required to produce a vocalisation (/ga/) (active task condition) once the circle turned green. Importantly, in the active condition, participants were also asked to prepare to speak during the delay period. In 50 % of the trials, no further stimuli were presented

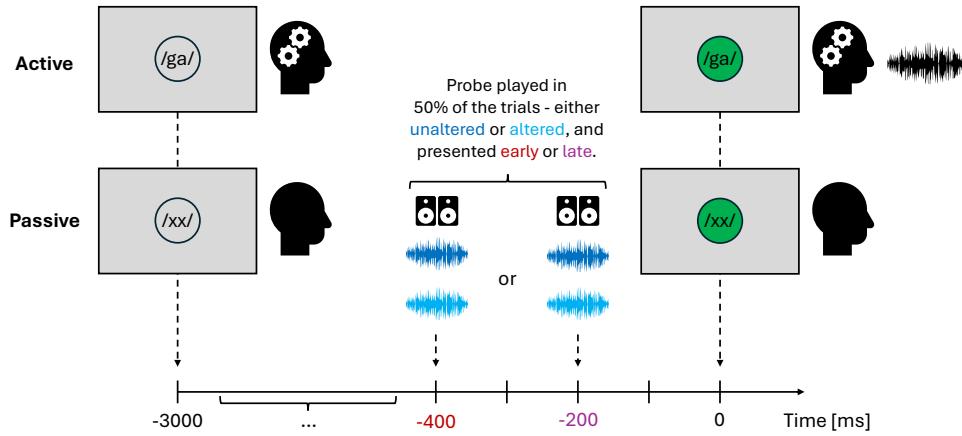


Figure 2: Experimental task.

In the active condition, participants were instructed to prepare to produce the syllable /ga/. When the go-signal - indicated by the circle turning green - appeared, they vocalised the syllable /ga/. In the passive condition, participants remained passive, observing the go-signal without responding or preparing. In 50 % of trials, a previously recorded probe was presented, either in its unaltered or altered form, and presented at either an early or late time point.

(no-probe trials). In the other 50 % of the trials, the previously recorded probe (see Chapter 2.2.4) was presented. The probe onset was early (-400 ms before the go-signal) or late (-200 ms before the go-signal). The probe type was either unaltered or altered. In total, participants completed 960 trials. 240 'active' 'no-probe' trials, 240 'passive' 'no-probe' trials, 60 'active' 'early' 'unaltered' probe trials, 60 'active' 'late' 'unaltered' probe trials, 60 'active' 'early' 'altered' probe trials, 60 'active' 'late' 'altered' probe trials, 60 'passive' 'early' 'unaltered' probe trials, 60 'passive' 'early' 'altered' probe trials, 60 'passive' 'late' 'unaltered' probe trials, 60 'passive' 'late' 'altered' probe trials, and 60 'passive' 'late' 'altered' probe trials. Thus, in summary, the experiment utilized a 2 (Task Condition: active, passive) \times 2 (Probe Type: unaltered, altered) \times 2 (Probe Onset: early, late) within-subject design. The trial structure followed a pseudo-randomization approach. To minimize the risk of participants forming expectations about the presence or absence of probes, a constraint was imposed on how many probe or no-probe trials could appear in sequence. To implement this constraint, trials were organized into miniblocks, each containing 16 trials: eight probe trials and eight no-probe trials. Within each miniblock, trial order was randomized while enforcing a repetition limit, ensuring that no more than four consecutive probe or no-probe trials occurred. If a sequence violated

this constraint, the randomization was repeated until a valid configuration was obtained. Once a valid sequence was established, trials were further categorized into active and passive conditions, with half of the probe and half of the no-probe trials randomly assigned to each. For probe trials, additional parameters were assigned: Probe type (unaltered or altered) and probe onset (early or late). The miniblock generation process was repeated iteratively until the required total number of trials was reached. Each participant was then left with a unique conditions file, which was subsequently divided into eight equal-sized experimental blocks. This structured yet flexible design ensured a balanced and unpredictable distribution of trial types and conditions. By limiting repetition and randomizing trial features, the design reduced the potential for pattern learning and maintained the experimental manipulation's integrity.

2.4 Procedure

Upon arrival, participants were informed about the study and gave their written consent. Participants then completed the FAL and washed their hair. The EEG cap was then fitted and prepared with electrode gel (Abralyt HiCl, Easy-cap GmbH, Herrsching, Germany). Impedances were brought below $20k\Omega$ before the start of the recording. During the preparation, participants received written instructions concerning the upcoming task and were encouraged to ask questions if anything was unclear. After taking seat in a sound-proof chamber, the initial stimuli recordings took place (see Chapter 2.2.4). Once completed, each participant was left with two auditory probes, an unaltered one and an altered (pitch-shifted) one. Before starting the main task of the experiment, the unaltered probe was repeatedly played back in order to calibrate a pleasant loudness level. Specifically, the experimenter successively reduced the loudness until reaching a pleasant level. Next, the main task of the experiment started (see Chapter 2.3). After each block, participants completed a custom version of the SAM questionnaire (Appendix J). If needed, the participants also had the opportunity to drink water during the break. Each break lasted approximately 3 min. Once all 960 trials were completed, participants completed the NASA-TLX questionnaire (Appendix K), and the EEG cap was removed. Finally, participants completed the reimbursement form and were given the opportunity to wash their hair. Including preparation, stimuli recordings, main task, and follow-up procedures, the experiment took approximately 210 min.

2.5 Analysis

2.5.1 Behavioural Preprocessing

The audio files recorded during each trial were preprocessed using Praat (Version 6.4.27; Boersma and Weenink, 2025) and MATLAB (Version R2024a; The MathWorks Inc., 2024). At first, every trial was classified based on whether it included a vocal response using a custom algorithm. A trial was defined as including a vocal response if the maximum intensity was above 47.5 dB SPL and if the intensity range was larger than 12 dB SPL. Further, to be defined as a vocal response, the Praat function *To TextGrid (silences)* had to be able to detect more than one interval. Lastly, both onset and F0 had to be defined for trials to be defined as including a vocal response. If the used Praat functions failed to extract the onset and/or the F0, a trial was defined as including no vocal response. This algorithm was developed using pilot data. For trials that were classified as including a vocal response, the onset-time (relative to the go-signal) was extracted. Further, the F0 for each vocal response was extracted using the filtered-autocorrelation method. Next, trials were flagged as being invalid if a vocal response was made during a passive trial and if no vocal response was made during an active trial. Additionally, trials were also flagged if the *z*-value of the vocal onset-time and/or the F0 exceeded ± 3 .

2.5.2 EEG Preprocessing

The EEG data were preprocessed using MATLAB (Version R2024a; The MathWorks Inc., 2024) in combination with the EEGLAB toolbox (Version 2023.0; Delorme and Makeig, 2004), the ICLabel plugin (Pion-Tonachini et al., 2019) and the BeMoBIL toolbox (Klug et al., 2022). Since the EEG data were used for both an ERP analysis and a single-trial classification analysis, two different preprocessing pipelines were used. To maximize the quality of the Independent Component Analysis (ICA), another ICA-specific preprocessing approach was used. For the ICA-specific preprocessing, the raw EEG data were high-pass filtered using a windowed sinc Finite Impulse Response (FIR) filter of order 1650, designed with a Hamming window, with a -6 dB cut-off frequency at 1 Hz. Next, bad channels were automatically identified and removed using the *bemo_bil_detect_bad_channels()* function from the BeMoBIL toolbox (Klug et al., 2022). The continuous data was then segmented into consecutive, non-overlapping 1-second epochs. Bad epochs were rejected based on joint probability and kurtosis criteria, using a threshold of ± 3 standard deviations. ICA weights were then computed using the *runica* algorithm and saved for subsequent artifact correction.

The ICLLabel plugin (Pion-Tonachini et al., 2019) was used to automatically label components representing non-brain activity, such as eye-blanks or muscle artifacts. Following classification, components were flagged if they had a probability above 0.7 in any non-brain class (e.g., muscle, eye, heart, line noise, channel noise, or other), regardless of their brain class probability. For the ERP-specific preprocessing, the raw EEG data were re-loaded. After renaming the relevant event markers and removing the previously identified bad channels, a low-pass filter was applied using a windowed-sinc FIR filter of order 440, implemented with a Hamming window, with a -6 dB cut-off frequency at 30 Hz. According to the preregistration (<https://doi.org/10.17605/OSF.IO/23EYF>), a high-pass FIR filter with a -6 dB cut-off at 0.3 Hz (order 5500) was set to be applied prior to low-pass filtering. However, this filter introduced artifacts in the data (see Appendix E) and was therefore omitted. Thus, only the low-pass filter was applied in this processing step. Next, previously calculated ICA weights were applied and flagged components removed. After this step, previously removed channels were interpolated using the *pop_interp()* function. Lastly, epochs were extracted from -200 ms to 400 ms relative to probe onset (see Chapter 2.3), and a baseline correction was applied. The baseline period was set to -200 ms to 0 ms, relative to the probe onset. Bad epochs were flagged based on joint probability and kurtosis criteria, using a threshold of ± 3 standard deviations. In addition to the epochs time-locked to the auditory probes, epochs time-locked to the (hypothetical) probe onset-time during the no-probe trials were also extracted in order to apply a correction (see Chapter 2.5.5). The preprocessing steps applied for the single-trial-based classification analysis were similar compared to the ones used for the ERP-specific preprocessing. However, instead of epoching the data relative to the onset of the auditory probes, the data were epoched from -1000 ms to 100 ms relative to the onset of the go-signal (see Chapter 2.3). The baseline period included the time-frame between -1000 ms to -800 ms, relative to the onset of the go-signal. Epochs were flagged based on the same criteria as described above. All other procedures used for the ERP-specific preprocessing were also applied to the preprocessing for the single-trial classification analysis. To ensure that the responses to probes did not impact the classification, only no-probe trials were used for the single-trial classification analysis.

2.5.3 Exclusion Criteria

To ensure a high data quality, trials which were flagged were excluded from the subsequent analyses. Since multiple data types (EEG data and vocal data) and different EEG preprocessing pipelines (ERP-specific preprocessing and single-

trial-specific preprocessing) were used, the flagged trials had to be merged in a systematic way. Since the ERP analysis and the behavioural analysis were linked in the sense that the vocal responses being made were thought to influence the ERP analysis, the flagged trials were merged. In other words, if a trial was flagged based on the behavioural data, it was also removed from the ERP analysis. Also, if a trial was flagged based on the ERP-specific preprocessing, it was also removed from the behavioural analysis. Thus, the ERP analysis and the behavioural analysis included the exact same trials. Since the classification analysis did not rely on any quantities associated with the vocal responses themselves, a different approach was used. A trial was excluded from the classification analysis if it was either flagged based on the single-trial-specific preprocessing and/or if it was flagged based on the behavioural preprocessing. However, the single-trial-specific preprocessing did not affect the trials included in the behavioural analysis. Thus, if a trial was flagged based on the behavioural preprocessing, it was also excluded from the classification analysis. However, if a trial was (only) flagged based on the single-trial-specific preprocessing, it was not excluded from the behavioural analysis. This approach was used to maximize the number of usable trials in either analysis. If the flagged trials were matched across all preprocessing pipelines, this could have led to trials being unnecessarily excluded. Lastly, participants were only included if more than 25 trials per condition remained for the ERP analysis, and more than 100 trials per condition remained for the classification analysis. Participants who did not meet these criteria were excluded from all analyses. Further, if data from any data set were missing, the participant was also excluded from all analyses¹. Thus, while individual trials were matched between the ERP and behavioural analyses - but not with the classification analysis - the same set of participants was included across all analyses.

2.5.4 Behavioural Analysis

Although not the focus of the present work, the vocal responses were analysed with regards to their F0 and their onset-time relative to the go-signal. Both quantities were extracted using Praat (Version 6.4.27; Boersma and Weenink, 2025) and further processed using MATLAB (Version R2024a; The MathWorks Inc., 2024) (see Chapter 2.5.1). Statistical analyses were performed using R (Version 4.5.0; R Core Team, 2025). The F0 values of the auditory probes (unaltered and altered) were z -standardised relative to the mean and the standard deviation of each participant's F0 distribution (based on the vocal responses made

¹This practice does only apply for missing EEG data and behavioural data, not missing questionnaire data.

during the experiment, not during the stimuli recordings). Paired *t*-tests and repeated measures Analysis of Variance (rmANOVAs) were used to compare F0 values and vocal onset-times across conditions. As recommended by Bakeman (2005), generalized eta squared (η_G^2 ; Olejnik and Algina, 2003) was used as effect size for rmANOVAs while Cohen's *d* was reported for paired *t*-tests. To address potential violations of the sphericity assumption, the Greenhouse-Geisser correction was applied when necessary. When the Greenhouse-Geisser ϵ exceeded 0.75, the Huynh-Feldt correction was used instead. Bonferroni-corrected *t*-tests were used for post hoc comparisons when required. Given violations of the normality assumption, paired *t*-tests were replaced by Wilcoxon signed-rank tests with rank-biserial correlations being reported as an effect size measure. rmANOVAs were replaced by non-parametric ANOVA-type tests using the nparLD package (Noguchi et al., 2012), with Bonferroni-corrected Wilcoxon signed-rank tests used for post hoc comparisons when required. For non-parametric ANOVA-type tests, the ANOVA-Type-Statistic (ATS) was reported as a test statistic (Noguchi et al., 2012). Lastly, questionnaire data were analysed as well using the same procedures as reported above. An alpha level of 0.05 was used to determine statistical significance.

2.5.5 ERP Analysis

The EEG data were processed using MATLAB (Version R2024a; The MathWorks Inc., 2024) in combination with the EEGLAB toolbox (Version 2023.0; Delorme and Makeig, 2004) while the statistical analysis was performed using R (Version 4.5.0; R Core Team, 2025). After preprocessing, the extracted epochs were averaged to extract the ERP. To ensure that the ERPs best reflect auditory processing, the following correction was applied. In addition to the ERPs time-locked to the auditory probes, control ERPs time-locked to the (hypothetical) probe onset-time during the no-probe trials were also extracted. The latter are thought to reflect only non-auditory processing (e.g., processing related to motor, visual, linguistic, or cognitive operations), whereas the former are thought to reflect both auditory and non-auditory processing. To isolate the signal reflecting auditory processing, the (no-probe) control ERPs were subtracted from the (probe) ERPs (see Appendix F). To obtain the control ERPs, no-probe trials were randomly assigned to either an early or a late control condition. Epochs were then extracted relative to the hypothetical onset-time of the corresponding (early or late) probe, simulating the timing structure of probe trials. To maximize the number of trials available for each control condition, no distinction was made between unaltered and altered probe trials during the correction

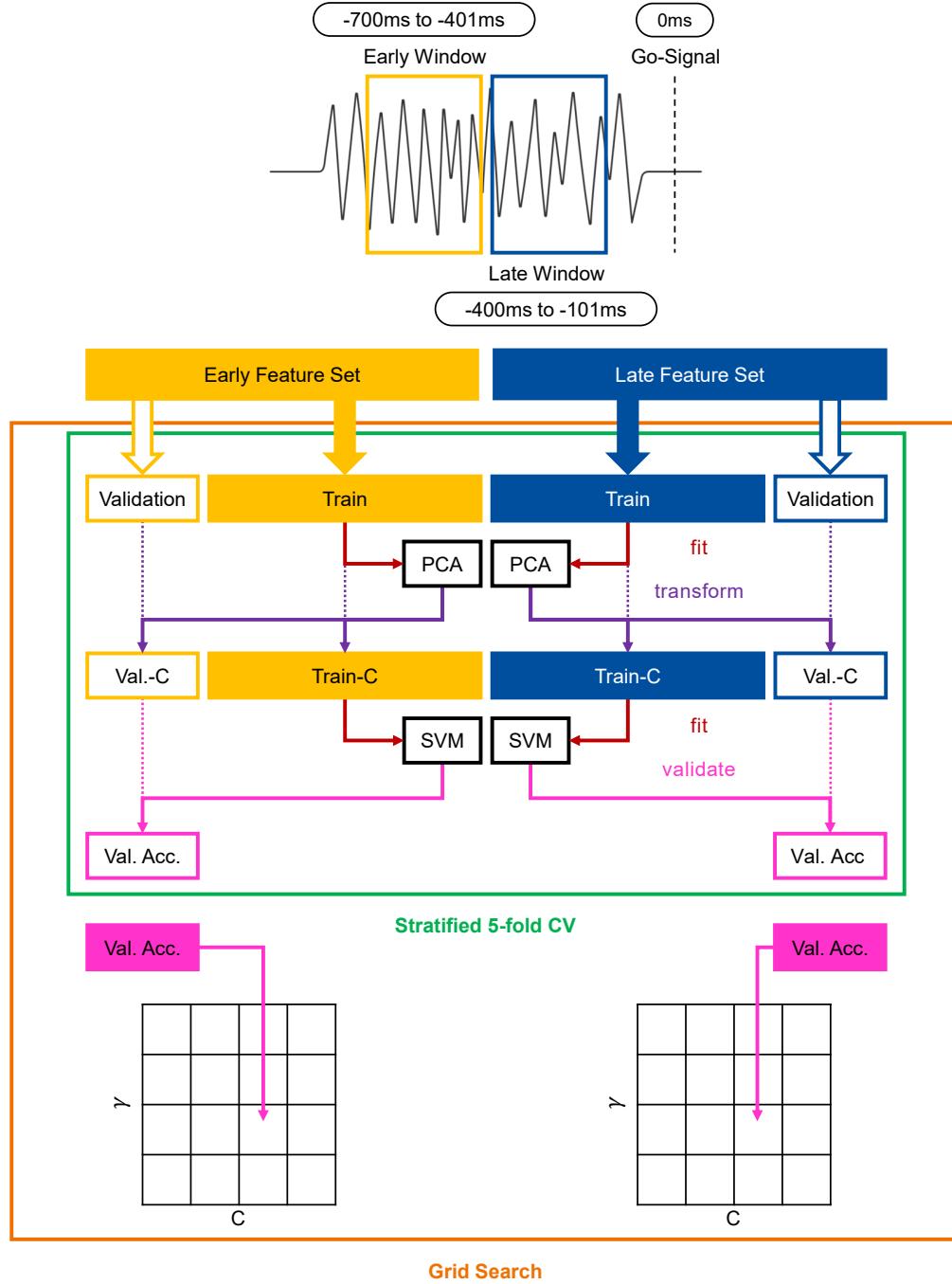
process. Consequently, early-unaltered and early-altered probe trials were both corrected using the same early control ERP, and vice versa. We acknowledge that this procedure relies on purely addictive effects and therefore is not able to correct any interactions between auditory and non-auditory processing. Still, this procedure provided the best estimate of the true auditory response. Furthermore, previous studies investigating the PSAM effect (Daliri & Max, 2015a, 2015b, 2016) have used this procedure, so its use in this study is necessary to allow comparisons. The resulting ERP was analysed with a focus on the N1 component, defined as the most negative peak occurring between 90 ms and 140 ms relative to the onset of the auditory probe. According to the preregistration (<https://doi.org/10.17605/OSF.IO/23EYF>), the time-window for the N1 component was set to range from 70 ms to 130 ms. However, due to late occurring N1 peaks for some participants, the window was changed to 90 ms to 140 ms. Electrode E01, which corresponds to electrode Cz in the 10-20 system, was used exclusively for the subsequent analysis due to it capturing the substantial part of the auditory N1 component (Sur & Sinha, 2009). The N1 ERP component was chosen as it is thought to be generated in the auditory cortex, representing auditory processing (Zouridakis et al., 1998). Furthermore, both the investigated PSAM effect and also the (potentially) related SIS effect are primarily observed in the N1 ERP component. To compare the N1 amplitudes and latencies across conditions, linear mixed-effects models (LMMs) were applied using the lme4 package (Bates et al., 2018) with participants being treated as random effects. LMMs provide a flexible and statistically robust approach for analysing hierarchical data structures, making them well-suited for repeated-measures designs (Schielzeth et al., 2020). Recent work suggests using a maximal random effects structure, i.e. random intercepts and random slopes for all predictors (Barr et al., 2013; Oberauer, 2022). However, fitting too complex models is also accompanied by some disadvantages (Bates et al., 2018), for example, a significant loss of statistical power (Matuschek et al., 2017). Thus, although all predictors were within-subject factors, random slopes were not included in the fitted models. This decision was based on the assumption that the effects of the predictors would be consistent across participants, with no theoretical expectation of substantial individual variability in the size or direction of these effects. Moreover, including random slopes in models with limited data can lead to overparameterization, reduced statistical power, and convergence issues (Bates et al., 2018; Matuschek et al., 2017). As shown by Matuschek et al. (2017), simplified random-effects structures - such as models with random intercepts only - can maintain appropriate Type I error control and improve power when random slopes account for little variance. Therefore,

a parsimonious model structure using only random intercepts was adopted to balance model complexity, interpretability, and statistical reliability. Furthermore, the PSAM effect was defined as the N1 amplitude in the active task condition – (minus) the N1 amplitude in the (matching) passive task condition. A positive PSAM effect refers to the suppression of N1 amplitudes during the active task condition (e.g., Daliri and Max, 2016). This index was calculated to reduce the complexity and enhance the interpretability of the fitted models. The emmeans package (Lenth, 2025) was used for pairwise comparisons. To correct for multiple comparisons, the Bonferroni correction was applied. An alpha level of 0.05 was used to determine statistical significance.

2.5.6 Classification Analysis

The aim of the classification analysis was to classify trials based on whether they were active trials (i.e., participants preparing to speak) or passive trials. Support Vector Machines (SVMs) with a Radial Basis Function (RBF) kernel were used as a binary classifier. Multiple features were extracted from the EEG data recorded during the preparation period (i.e., before the go-signal, see Chapter 2.3). Importantly, the features were extracted twice, once during the early preparation phase and once during the late preparation phase (Figure 3). Subsequently, the characteristics of models using data from the early preparation period were compared to the characteristics of models using data from the late preparation period. Only no-probe trials were used for this analysis. The used approach was adapted from Kinahan et al. (2024). The SVMs were fitted using the scikit-learn package (Pedregosa et al., 2011) in Python (Version 3.13.4). Further processing was performed using MATLAB (Version R2024a; The MathWorks Inc., 2024) and R (Version 4.5.0; R Core Team, 2025).

Feature Extraction & Feature Selection As described above, the preprocessed and epoched data were segmented into two time-windows - an early window (-700 ms to -401 ms) and a late window (-400 ms to -101 ms), relative to the onset of the go-signal (Figure 3). These windows were chosen as they correspond to the probe onsets, analysed in the ERP analysis (see Chapter 2.5.5). For each window, multiple features were extracted separately for every channel and epoch. The choice of extracted features was guided by earlier studies (Dhole, 2024; Kinahan et al., 2024). EOG channels were not used. For example, the mean amplitude was extracted twice for each epoch and channel - once using the early window and once using the late window. Time-domain features included mean amplitude, Root Mean Square (RMS), standard deviation of the amplitude, maximum

**Figure 3:** Single-trial classification pipeline.

Several features were extracted from the preprocessed single trial EEG data. One feature set was extracted from an early time-window (yellow) and one feature set was extracted from a late time-window (blue). Different hyperparameter-pairs (C & γ) were evaluated using a stratified 5-fold cross-validation. Before fitting the SVMs, the dimensionality of the feature sets was reduced using a PCA. For each iteration of the cross-validation, only the training split was used to fit the PCA to avoid data leakage. The validation accuracy for each hyperparameter-pair was determined by averaging the accuracies obtained in each iteration of the cross-validation. The process was repeated for every participant, leaving all participants with two validation accuracy matrices.

and minimum amplitude, kurtosis, skewness, and zero-crossing rate. Frequency-domain features were derived using a Hilbert Transformation on the bandpass filtered data for alpha (8 - 13 Hz), beta (14 - 30 Hz), and (low) gamma (31 - 37 Hz) frequency bands. Before applying the Hilbert Transformation, bad channels (as previously identified) were removed, the previously calculated ICA weights were applied, and flagged components were removed (see Chapter 2.5.2). After this step, previously removed channels were interpolated using the *pop_interp()* function. Next, the continuous data was filtered using separate low- and high-pass FIR filters and the Hilbert Transformation was applied. Lastly, the data were epoched from -1000 ms to 100 ms relative to the onset of the go-signal (see Chapter 2.3) and the band power in each window (early, late) was extracted. Importantly, epochs, which were previously identified as containing artifacts, were also removed from this processing step. Additionally, the Hjorth parameters - Activity, Mobility, and Complexity - were computed according to Alawee et al. (2023), with a minor modification concerning Equation (3)². Activity was defined as the variance of the signal:

$$\text{Activity} = \text{Var}(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (1)$$

Mobility was calculated based on the first derivative of the signal. First, the derivative was defined as:

$$d_i = x_{i+1} - x_i \quad (2)$$

Then, the variance of the derivative was computed:

$$\text{Var}(d) = \frac{1}{N} \sum_{i=1}^N (d_i - \bar{d})^2 \quad (3)$$

Finally, mobility was given by:

$$\text{Mobility} = \sqrt{\frac{\text{Var}(d)}{\text{Var}(x)}} \quad (4)$$

Complexity was defined as the ratio of the mobility of the derivative to the mobility of the original signal:

$$\text{Complexity} = \frac{\text{Mobility of } d}{\text{Mobility of } x} \quad (5)$$

²Alawee et al. (2023) inconsistently used N and $N - 1$ as denominators for equations (1) and (3). Here, we consistently use N as a denominator.

After extracting all features, each participant was left with two feature sets, one using data from the early window and one using data from the late window. Each feature set contained 392 features (28 electrodes \times 14 measures). To reduce the dimensionality of the feature sets, a Principal Component Analysis (PCA) was applied after standardizing each feature set. For each feature set, 100 components were extracted, leaving each participant with two final feature sets, each containing 100 features. The number of components was determined using pilot data. Importantly, the PCA weights were estimated only by using the training split for each iteration of the k-fold cross-validation (see below). The estimated PCA weights were then applied to the test split to obtain the PCA-transformed validation split for each iteration of the k-fold cross-validation.

Hyperparameter Tuning To evaluate performance across different model configurations, a grid search was conducted separately for each participant and feature set (Figure 3). SVMs with a RBF kernel were trained using a range of hyperparameter combinations. Specifically, 30 values of the regularization parameter C were logarithmically sampled between 10^{-2} and 10^{10} , and 30 values of the kernel coefficient γ were logarithmically sampled between 10^{-14} and 10^{-1} , yielding 900 unique parameter combinations. The hyperparameter ranges included in the search space were defined on the basis of pilot data. Model performance was assessed using stratified 5-fold cross-validation. In k-fold cross-validation, the dataset is divided into k equal parts (folds). The model is trained on k-1 folds and tested on the remaining one, repeating this process k times to obtain an average performance estimate across folds. Compared to regular k-fold cross-validation, stratified k-fold cross-validation ensures that each fold maintains a class distribution similar to that of the entire dataset (Allen et al., 2021; Prusty et al., 2022; Zeng & and Martinez, 2000). To reduce the dimensionality of the feature sets, a PCA was applied after z-standardization, retaining 100 components per feature set. Importantly, PCA weights were estimated using only the training split in each cross-validation iteration, and then applied to the corresponding validation split to avoid data leakage. The validation accuracy was computed using the validation split in each fold. The final validation accuracy was calculated as the average across all iterations. This procedure was repeated independently for each participant and time-window, resulting in one accuracy matrix per participant and time-window. Grand average accuracy matrices were obtained by averaging across participants.

Analysis of the Decision Boundary Following Kinahan et al. (2024), we propose that the learned decision boundary of a binary classifier is directly related to the difficulty of the classification problem. In the case of SVMs with a RBF kernel, the decision boundary is influenced by two key hyperparameters: the regularization parameter C and the kernel coefficient γ . The parameter C controls the trade-off between maximizing the margin and minimizing classification errors. Lower values of C allow for a wider margin by tolerating more misclassifications, resulting in a smoother and simpler decision boundary. In contrast, higher values of C place greater emphasis on correctly classifying each training example, which can lead to a more complex boundary that closely follows the training data. The parameter γ defines the influence of individual data points. Low values of γ produce a broad and more global decision surface, whereas high values result in narrow, highly localized regions of influence, enabling the classifier to capture fine-grained patterns but increasing the risk of overfitting. Taken together, small values for C and γ lead to a more linear decision boundary, while large values for both hyperparameters lead to a more complex model. If a relatively simple model is able to perform well, one can infer that the classification problem is also relatively simple (Kinahan et al., 2024). To assess whether the classification problem is less complex when using features extracted from the late time-window compared to the early time-window (see H3), the following approach was employed. First, the proportions of hyperparameter combinations that yielded above-chance classification performance were quantified. Since the number of excluded trials varied across participants and statistical significance depends on the number of retained trials (Combrisson & Jerbi, 2015), the chance level was determined individually for each participant. For example, given the present 2-class classification problem and using a α -level of 0.05, the chance-level is 54.7 % when including 300 trials (Combrisson & Jerbi, 2015). To determine whether classification accuracy exceeded chance level at a statistically significant level ($\alpha = 0.05$), the following procedure was applied. For each participant, the number of included trials was determined from the available data. Based on Combrisson and Jerbi (2015), the threshold accuracy was then calculated using the following formula:

$$\text{Threshold\%} = \frac{\text{binoinv}(1 - \alpha, N_{\text{trials}}, \frac{1}{N_{\text{classes}}})}{N_{\text{trials}}} \quad (6)$$

α represents the type I error, N_{trials} is the number of trials, N_{classes} is the number of classes (here $N_{\text{classes}} = 2$) and $\text{binoinv}(1 - \alpha, N_{\text{trials}}, \frac{1}{N_{\text{trials}}})$ denotes the inverse cumulative distribution function of the binomial distribution. This threshold reflects the minimum accuracy needed to conclude that a classifier's performance

is significantly better than chance, based on the sample size and the used α -level. The proportion of hyperparameter-pairs resulting in an above-chance classification was extracted for the models using the feature set derived from the early window and for the models using the feature set derived from the late window. Thus, each participant was left with two values. A higher proportion of successful combinations suggests that a robust decision boundary can be learned across a wide range of parameter settings, indicating a relatively simple classification problem. Conversely, a lower proportion implies that accurate classification is only achievable under narrowly constrained parameter configurations, suggesting a more complex or ambiguous feature space. To statistically compare classification complexity between the two conditions, the proportion of above-chance hyperparameter pairs was compared using a Wilcoxon signed-rank test.

2.6 Use of Artificial Intelligence

ChatGPT (OpenAI, 2025) & Gemini (Google, 2025) were employed to revise previously written text by shortening, rephrasing, and correcting grammatical errors. These revisions were selectively applied across all chapters. Additionally, ChatGPT & Gemini were used to generate text based on bullet points in chapters 2 and 3. To further enhance the manuscript's language quality, DeepL (DeepL GmbH, 2024) was used for spelling and grammar correction in selected passages throughout all chapters. ChatGPT & Gemini were also utilized for code-related tasks, including debugging and translations between programming languages. Debugging involved analysing error messages and identifying problematic script sections. In rare instances, ChatGPT & Gemini were used to generate code from scratch; in such cases, the output was thoroughly reviewed and, if necessary, modified. Lastly, ChatGPT & Gemini were employed to improve the efficiency and performance of previously written scripts. All AI-generated or corrected code was carefully inspected to ensure accuracy and correctness.

2.7 Preregistration & Code Availability

This study was preregistered under the following link.

<https://doi.org/10.17605/OSF.IO/23EYF>

Any deviations from the preregistration were reported as such in the present manuscript and under <https://osf.io/9quhn>. All scripts and functions used are available in the project's GitHub repository. The repository also includes relevant documents such as instructions and questionnaires.

<https://github.com/timdressler/psam>

While the GitHub repository might be subject to further changes and optimizations, the repository was further published as a time-stamped, immutable, and permanent version under the following link.

<https://doi.org/10.5281/zenodo.16632078>

3 Results

3.1 Vocal Responses & Behavioural Data

The mean F0 for male participants was 112.09 Hz ($SD = 15.16$), while female participants had an average of 209.28 Hz ($SD = 18.26$). Both the mean values and the variability fall within expected ranges reported in the literature (Berg et al., 2017; Pépiot, 2014). To investigate how well the recorded auditory probes matched the vocal responses made during the experiment with respect to their F0, the F0s of the probes were standardised relative to the individual M and SD of each participant's F0 distribution (based on the vocal responses made during the experiment, not during the stimuli recordings). The F0 distribution itself was also standardised relative to its M and SD (Figure 5). A paired t -test comparing the standardised F0s of altered and unaltered auditory probes revealed a significant difference, $t(27) = -13.84$, $p < .001$, $d = -2.62$ (Figure 4). Altered probes exhibited significantly lower standardised F0 values ($M = -4.26$, $SD = 1.74$) than unaltered probes ($M = 1.02$, $SD = 1.52$). This difference reflects the manipulation applied to the altered probes, which systematically reduced their F0, making them less representative of natural vocal production. In contrast, the unaltered probes retained their original F0 characteristics, closely resembling typical vocalizations. As such, they are assumed to match the EC while the altered probes introduced a mismatch (Eliades & Tsunada, 2024; Niziolek et al., 2013). However, for some participants, even the unaltered probe did not represent the vocal responses made during the experiment well (see Figure 5). To assess whether the presence of a probe affected vocal production, comparisons were conducted between probe and no-probe trials. A Wilcoxon signed-rank test indicated no significant difference in vocal F0s between the two conditions, $V = 189$, $p = .762$, $r = -0.07$, suggesting that the mere presence of a probe did not modulate vocal pitch. Including only probe trials, further analyses examined the influence of probe type and probe onset on vocal F0s. A non-parametric rmANOVA revealed a significant main effect of probe type, $ATS(1) = 16.83$, $p < .001$, whereas neither probe onset, $ATS(1) = 0.67$, $p = .414$, nor the interaction between probe type and probe onset, $ATS(1) = 0.77$, $p = .382$, reached significance (Figure 4). Follow-up Wilcoxon signed-rank tests confirmed that both early and late altered probes were followed by vocal responses with significantly lower F0 values compared to unaltered probes ($p < .001$). In addition, the effect of probe presence on vocal onset-times was also examined. A Wilcoxon signed-rank test showed that onset-times were significantly earlier during probe trials ($M = 0.48$, $SD = 0.10$) compared to no-probe trials ($M = 0.55$, $SD = 0.09$), $V = 404$,

$p < .001$, $r = 0.99$. Using only probe trials, a rmANOVA on vocal onset-times revealed a significant main effect of probe onset, $F(1, 27) = 65.06$, $p < .001$, $\eta_G^2 = .058$, as well as a significant interaction between probe type and probe onset, $F(1, 27) = 5.89$, $p = .022$, $\eta_G^2 = .001$. However, the main effect of probe type did not reach statistical significance, $F(1, 27) = 3.57$, $p = .070$, $\eta_G^2 = .001$. The effect of probe onset reflected longer vocal onset-times following late probes ($M = 0.51$, $SD = 0.11$) compared to early probes ($M = 0.46$, $SD = 0.10$), consistent with findings reported by Mock et al. (2011). Post-hoc comparisons indicated no significant difference between probe types in early trials ($p = .998$), whereas in late trials, altered probes were associated with significantly shorter onset-times than unaltered probes ($p = .001$). Finally, to evaluate time-related changes, vocal F0s and onset-times were analysed across experimental blocks. No significant block effect was observed for vocal F0s, $ATS(3.16) = 1.00$, $p = .395$, nor for vocal onset-times, $F(2.27, 61.16) = 1.06$, $p = .358$, $\eta_G^2 = .004$. In addition to vocal responses, participants provided self-reported data on mood and tiredness using standardised questionnaires. To assess whether these behavioural measures changed over time, responses were analysed across experimental blocks using rmANOVAs. This approach allowed for the examination of changes across the session, providing insight into potential temporal effects related to fatigue or affective state. For mood, the effect of experimental block was not statistically significant, $F(3.39, 91.44) = 1.60$, $p = .188$, $\eta_G^2 = .010$. Similarly, for tiredness, the effect of block was non-significant, $F(2.61, 70.54) = 1.04$, $p = .374$, $\eta_G^2 = .010$. These results indicate that self-reported mood and tiredness remained relatively stable across the experimental session (Appendix G). NASA-TLX ratings suggested that the participants experienced relatively low task demands overall (Appendix H). Mental demand was rated with a mean of $M = 3.11$, $SD = 2.04$, and physical demand was even lower, $M = 1.98$, $SD = 2.12$. Performance was rated at $M = 2.18$, $SD = 1.23$, indicating that participants generally perceived their performance as good. Effort received the highest ratings among all dimensions ($M = 4.07$, $SD = 2.28$), suggesting that participants exerted a moderate amount of effort during the task. Frustration levels were also low ($M = 2.05$, $SD = 2.00$), indicating that participants did not find the task particularly frustrating.

3.2 Pre-Speech Auditory Modulation

A baseline model for N1 amplitudes revealed an overall negativity, with a mean amplitude of $-7.50 \mu\text{V}$. The Intraclass Correlation Coefficient (ICC) was .767, indicating substantial between-subject variability. The ERPs and topographies for each condition are displayed in Figure 7. To investigate whether the prepa-

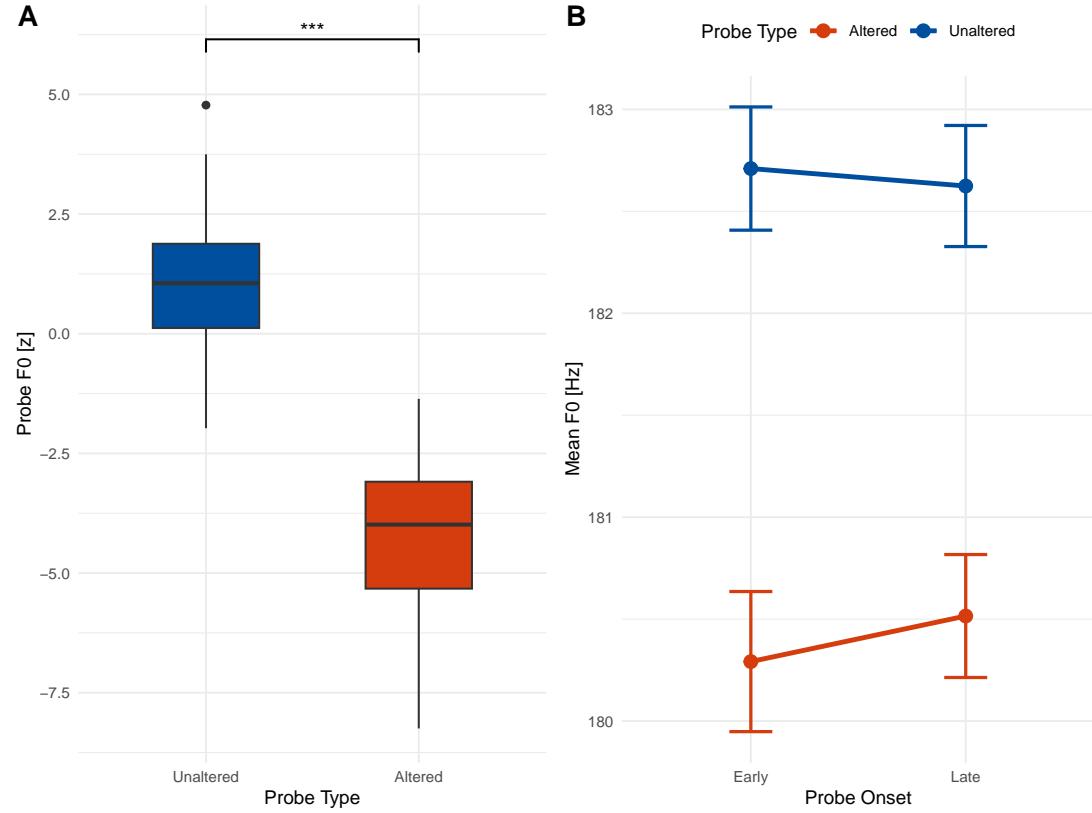


Figure 4: Standardised probe F0 values for altered and unaltered probes (A) & Average F0 value by probe type and probe onset (B). Error bars in B represent within-subject standard errors according to Morey (2008).

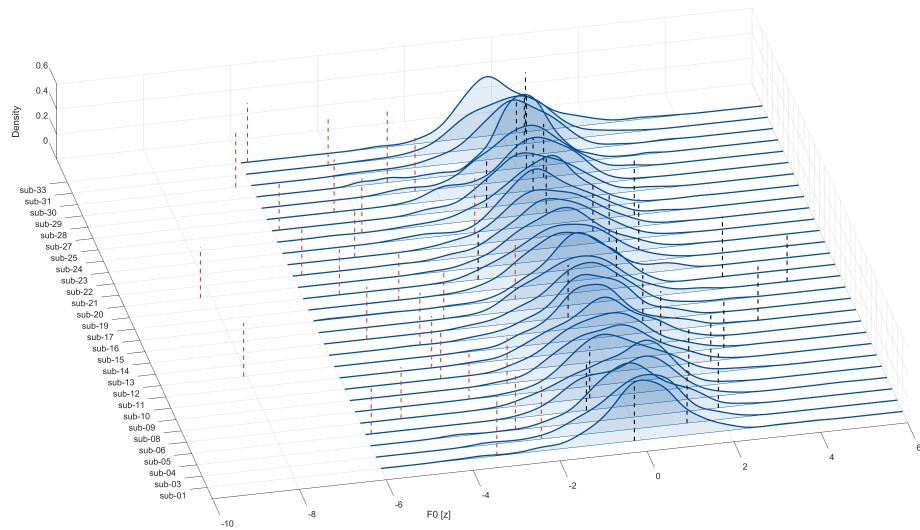


Figure 5: Standardised F0 distributions and probe F0 values across participants. Probe F0 values and F0 distributions were standardised using the M and SD of each participant's own F0 distribution. Black lines indicate the standardised F0s of unaltered probes. Red lines indicate the standardised F0s of altered probes.

ration to speak modulated auditory processing as indexed by N1 amplitudes, a LMM was fitted with task condition, probe type, and probe onset as predictors. A significant main effect of task condition emerged, $\beta = 1.21$, $SE = 0.53$, $t(189) = 2.28$, $p = .024$, reflecting more negative N1 amplitudes in the active condition than in the passive condition (Figure 6). This finding runs counter to hypothesis H1a, which predicted N1 suppression (i.e., more positive amplitudes) during speech preparation. No significant main effects were found for probe type or probe onset, and no interactions reached significance, all $p > .05$ (Appendix A). A planned contrast confirmed the task-related amplitude difference, $t(189) = -3.76$, $p < .001$. Thus, although the task condition significantly influenced N1 amplitudes, the direction of the effect was opposite to the predicted pattern described in Daliri and Max (2015a, 2015b, 2016), with amplitudes becoming more negative - not more positive - during speech preparation. In other words, while a PSAM effect was observed, it was negative: N1 amplitudes were larger (i.e., more negative) in the active condition, contrary to the predicted suppression pattern, which would have manifested as less negative amplitudes during speech preparation (i.e., a positive PSAM effect). As this pattern deviates from the definition of the PSAM effect as proposed in Daliri and Max (2015a, 2015b, 2016), all amplitude-related hypotheses (H1a, H1b, and H1c) were not supported. Nevertheless, the planned analyses were carried out to further examine the characteristics of this task-related modulation of the N1 amplitude. The PSAM effect was computed according to the analysis plan by calculating the difference in N1 amplitudes between active and passive conditions for each probe type and probe onset combination. The resulting scores were then entered into a LMM with probe type, probe onset, and their interaction as fixed effects. The model showed no significant effects for probe onset ($\beta = -0.55$, $SE = 0.69$, $t(81) = -0.80$, $p = .426$), probe type ($\beta = 0.68$, $SE = 0.69$, $t(81) = 0.98$, $p = .328$), or their interaction ($\beta = 0.60$, $SE = 0.98$, $t(81) = 0.61$, $p = .541$) (Figure 8). Planned contrasts confirmed that there were no significant differences in the PSAM effect between altered and unaltered probes, neither at the early onset ($t(81) = -0.98$, $p = .328$), nor at the late onset ($t(81) = -1.85$, $p = .068$). However, it is noteworthy that, descriptively, the PSAM effect was even more negative for altered probes, particularly when those probes were presented at the later time point. In other words, although the N1 amplitudes were generally enhanced (i.e., more negative) in the active task condition, this enhancement appeared to be, at least descriptively, more pronounced for altered probes presented later in the trial. Finally, to examine N1 latency, an additional model assessed the effects of task condition, probe type, and probe onset. All main effects and interactions failed to

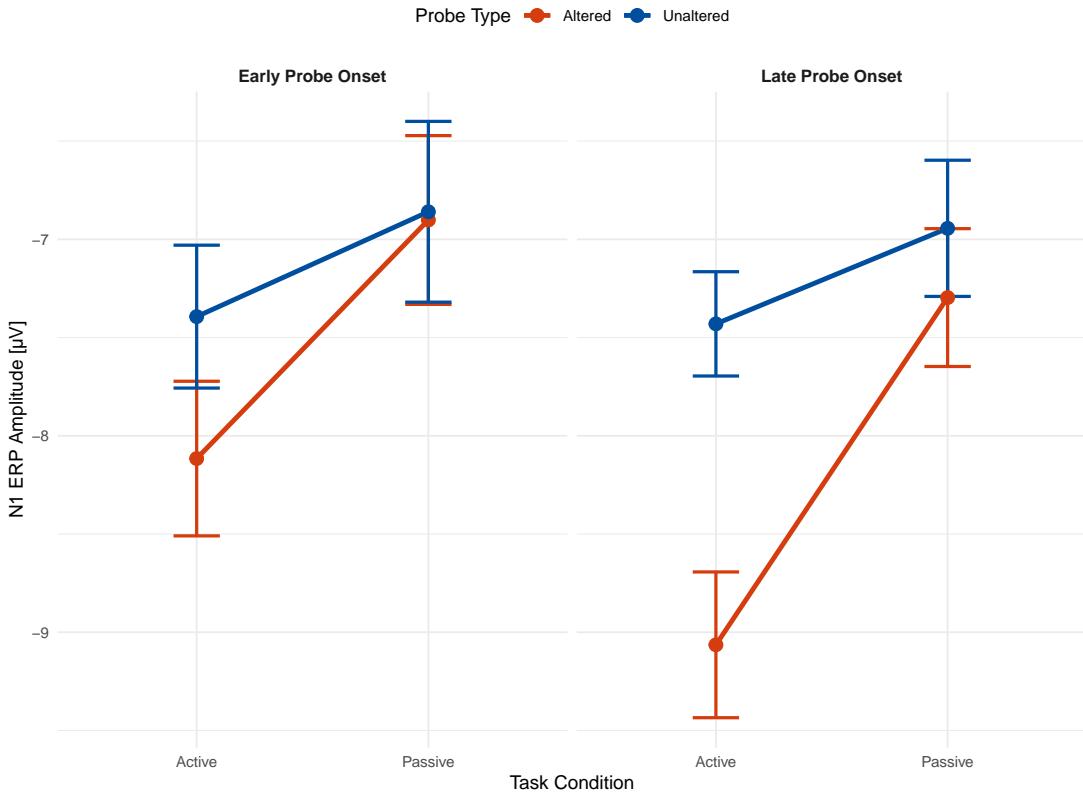


Figure 6: N1 amplitude by task condition, probe type and probe onset.
Error bars represent within-subject standard errors according to Morey (2008).

reach significance (all $p > .05$) (Figure 9). Specifically, there were no significant effects of task condition ($\beta = 2.93$, $SE = 2.74$, $t(189) = 1.07$, $p = .287$), probe onset ($\beta = 0.18$, $SE = 2.74$, $t(189) = 0.07$, $p = .948$), or probe type ($\beta = 0.79$, $SE = 2.74$, $t(189) = 0.29$, $p = .775$). Similarly, the interaction terms were not significant: task condition by probe onset ($\beta = -6.86$, $SE = 3.88$, $t(189) = -1.77$, $p = .079$), task condition by probe type ($\beta = -7.29$, $SE = 3.88$, $t(189) = -1.88$, $p = .062$), probe onset by probe type ($\beta = -1.93$, $SE = 3.88$, $t(189) = -0.50$, $p = .620$), and the three way interaction ($\beta = 7.64$, $SE = 5.49$, $t(189) = 1.39$, $p = .165$). Planned contrasts further indicated no significant N1 latency difference between altered and unaltered probes, neither in the active condition ($t(189) = 0.09$, $p = .927$), nor in the passive condition ($t(189) = 1.88$, $p = .062$). These results indicate that N1 latencies did not reliably vary as a function of task condition, probe onset, or probe type. This pattern contradicts hypothesis H2, which predicted shorter latencies for unaltered probes during speech preparation, with no difference expected during passive listening. Thus, hypothesis H2 was also not supported by the data.

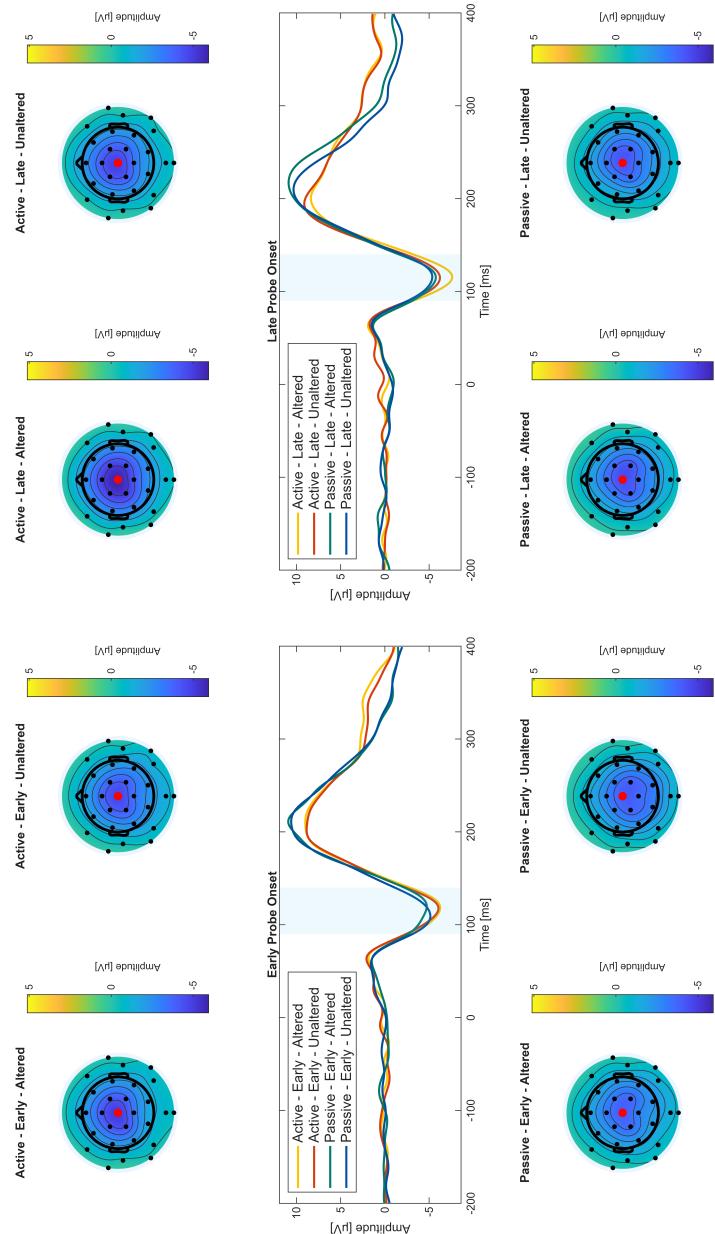


Figure 7: Grand average corrected ERPs and topographies for all conditions. Topographies are averaged over 90 ms to 140 ms.

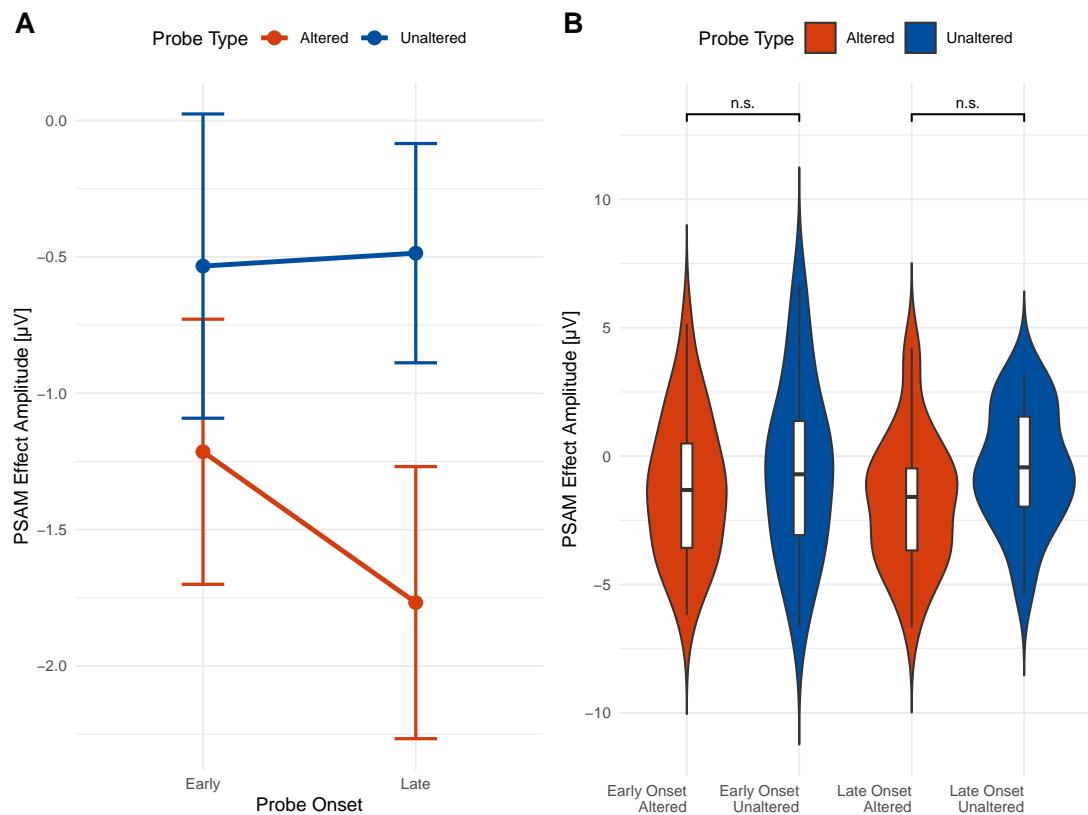


Figure 8: PSAM Effect by probe type and probe onset.
Error bars in A represent within-subject standard errors according to Morey (2008).

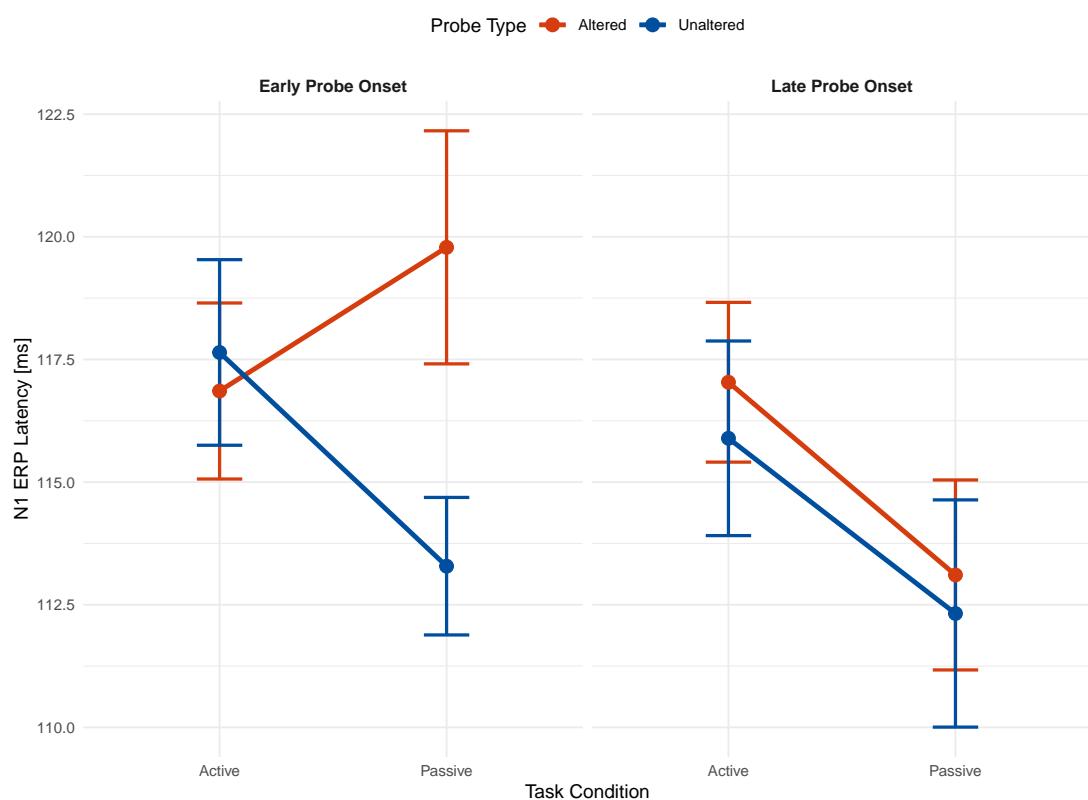


Figure 9: N1 latency by task condition, probe type and probe onset.
Error bars represent within-subject standard errors according to Morey (2008).

3.3 Single-Trial Classification

To assess whether speech preparation becomes more specific over time, we used SVMs with a RBF kernel to classify active and passive trials and compared the percentage of hyperparameter-pairs - combinations of the regularization parameter C and the kernel coefficient γ - that led to above-chance classification performance when features were extracted from either an early or a late temporal window (see Chapter 2.5.6). The underlying rationale was that a higher proportion of successful hyperparameter-pairs indicates a simpler classification problem (Kinahan et al., 2024), which in turn suggests that the neural signatures are more distinct, potentially hinting at a more specific preparation. On average, a slightly higher proportion of successful pairs was observed when using features from the late window ($M = 0.30$, $SD = 0.24$) compared to the early window ($M = 0.27$, $SD = 0.29$), consistent with hypothesis H3 (Figure 10). However, a Wilcoxon signed-rank test revealed no significant difference, $V = 183$, $p = .662$, $r = -0.10$, indicating relatively stable preparation processes in the investigated time-window. Thus, hypothesis H3 was not supported. Grand average accuracy heatmaps for early and late windows are shown in Figure 11. Individual accuracy heatmaps for early and late windows are shown in Appendix D. Notably, the observed patterns showed substantial variability across participants (Figure 10 and Appendix D).

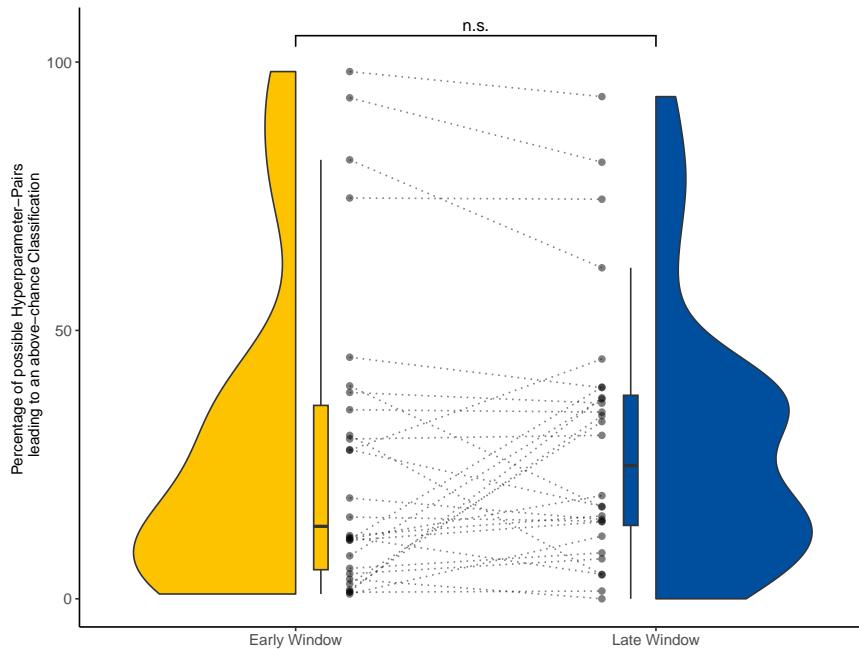


Figure 10: Percentage of possible hyperparameter-pairs leading to an above-chance classification for early and late time-windows.

Dotted lines represent individual trajectories.

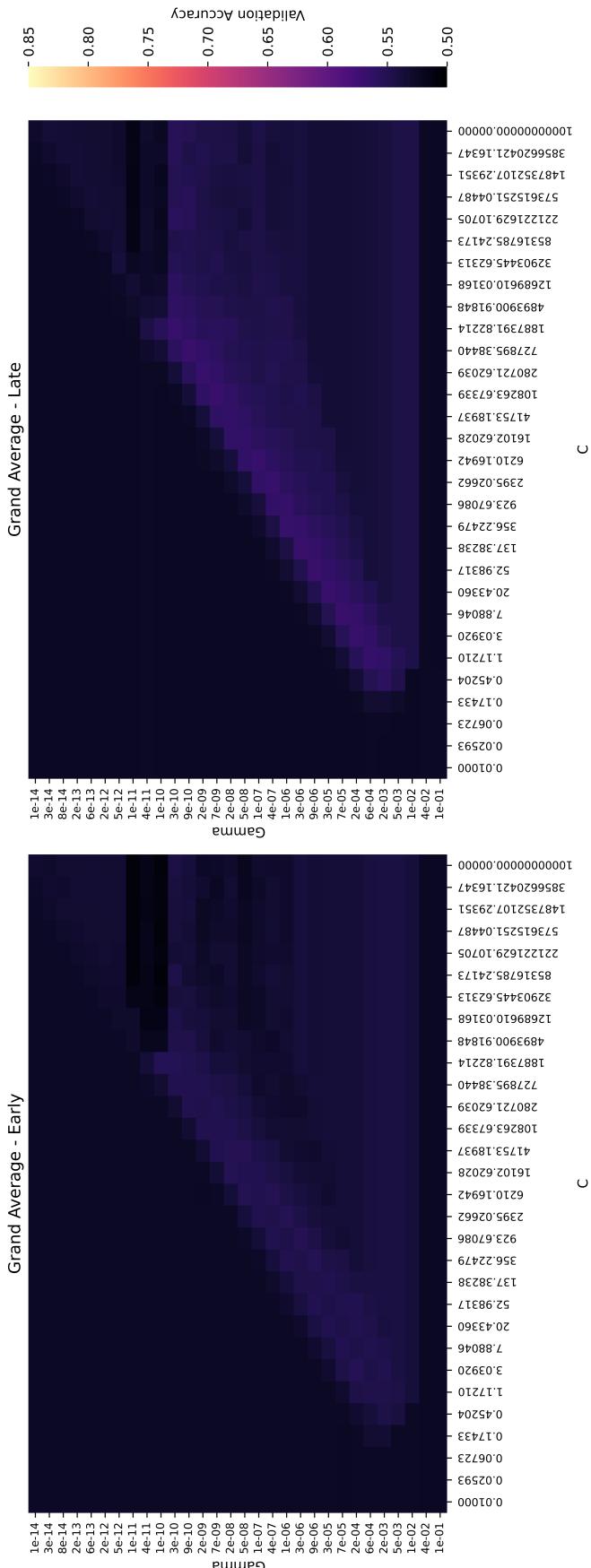


Figure 11: Grand average hyperparameter accuracy heatmaps. Each matrix element corresponds to one hyperparameter-pair. Gamma values are displayed in scientific notation, rounded to the nearest power-of-ten with one digit in the mantissa for improved readability.

4 Discussion

The present study investigated the specificity of the PSAM effect, examining whether auditory processing during speech preparation is modulated by predictions about the upcoming sensory consequences of speech. To this end, we examined whether modulation of the auditory N1 component during speech preparation varied depending on the congruence of the auditory probe with the assumed EC and its timing within the preparation phase. While PSAM was observed as a modulation of N1 amplitudes during speech preparation, the direction of the effect was opposite to that seen in previous studies, showing enhanced rather than suppressed responses during speech preparation (H1a not supported). Moreover, there was no evidence that modulation was greater during the late compared to the early preparation phase (H1b not supported), nor did the PSAM effect differ between congruent and incongruent probes (H1c not supported). N1 latencies were also unaffected by probe type and probe onset (H2 not supported). To better understand the nature of the preparation process itself, rather than only its effects on auditory processing, we investigated whether speech preparation becomes more specific over time. This was achieved by comparing the complexity of classification models trained on features extracted from an early versus a late preparation phase. However, no differences in model complexity were found (H3 not supported), suggesting no measurable increase in specificity. Although not the focus of the present study, vocal responses were also analysed to assess whether the auditory probes influenced speech production. Vocal F0s were not affected by the mere presence of a probe, but were significantly lower following altered compared to unaltered probes, irrespective of timing. In addition, effects on vocal onset-times were observed, suggesting that auditory probes can influence response timing.

4.1 Revisiting Pre-Speech Auditory Modulation

The primary objective of the present study was to investigate the PSAM effect, which is typically indexed as suppressed N1 amplitudes evoked by auditory probes presented during speech preparation compared to passive listening. However, contrary to the expected suppression, enhanced N1 amplitudes were observed during speech preparation. This enhancement was neither influenced by the timing of the probe presentation nor by the congruence of the probe and the assumed EC. One could argue that, although the unaltered probe reflected a typical vocalisation for most participants, based on its F0, it may have introduced a mismatch with the assumed EC for some participants, thereby leading to reduced suppression.

sion or even enhancement of neural responses. However, this explanation appears unlikely for several reasons. First, while the unaltered probe may not have represented a typical vocalisation for some individuals and thus created a mismatch with the EC (Niziolek et al., 2013), it did accurately represent a typical vocalisation for the majority of participants. Moreover, even pure tones used in previous studies (e.g., Daliri and Max, 2016), which clearly introduced a mismatch with the EC, still resulted in suppressed N1 amplitudes. Taken together, these findings suggest that the manipulation of EC congruence was largely successful and that the absence of suppression, or rather the observed enhancement, is unlikely to be caused by a mismatch between the (unaltered) probe and the EC. That said, we acknowledge that EC congruence in the current study was defined solely by the F0, which likely constitutes an oversimplification, given that the EC probably also encodes additional features such as formant structure and other acoustic properties (Niziolek et al., 2013; Tian et al., 2018). A pattern similar to the one observed in the present study was reported by S. Li et al. (2020). In their study, participants were instructed to engage in either general preparation (preparing to speak without knowing what to say) or specific preparation (preparing to speak while knowing what to say). Following the preparation phase, participants were cued to produce a specific syllable (e.g., /ba/). Tones or syllables were presented as probes during the preparation phases. For syllable probes, results indicated suppressed N1 amplitudes during general preparation but enhanced N1 amplitudes during specific preparation when the probe was congruent with the prepared syllable. S. Li et al. (2020) interpreted this pattern as a potential reflection of different modulatory signals being available at distinct preparation stages, leading to varying modulations in the auditory system. Specifically, they hypothesised the presence of a general (suppressive) CD signal during general preparation and a specific EC signal during specific preparation, which would enhance the sensory response in case there is a match between the prepared and the presented vocalisation. However, subsequent studies have shown inconsistent results regarding the general preparation effect, with the specific preparation not being tested (Zheng et al., 2022). Furthermore, the findings from S. Li et al. (2020) may be subject to alternative interpretations (see below), and it is not believed that the present results, despite showing similar patterns (increased N1 amplitudes during specific preparation), are attributable to the mechanisms described by S. Li et al. (2020). Nevertheless, other studies also reported motor-based enhancement of auditory responses. Reznik et al. (2014) reported enhanced activity in the auditory cortex during active sound generation compared to passive listening. Findings from other domains, such as vision, also demonstrate motor-based enhancement (Cao

& Händel, 2019; Mifsud et al., 2016). However, Mifsud et al. (2016) ultimately interpreted the observed enhancement as reflecting attentional effects rather than purely motor-based mechanisms (see below). Additionally, Scheerer and Jones (2018) and Wu et al. (2025) suggest that motor-based predictions enhance the brain's sensitivity to predicted acoustic features, which is crucial for vocal control. However, increased sensitivity does not invariably equate to enhanced neural responses. Indeed, the opposite has been observed, where greater suppression leads to higher sensitivity (Eliades and Tsunada, 2024; Eliades and Wang, 2008; Ozker et al., 2024; but also see Chang et al., 2013). Further evidence for EC-based enhancement stems from repetition studies, which demonstrated increased neural responses when a stimulus was presented shortly after imagined production, provided it was congruent (Tian & Poeppel, 2013). Consequently, it is plausible that participants engaged in imaginary articulation as a preparation strategy, which could have led to enhanced responses to probes presented after the imagined articulation (Tian & Poeppel, 2013). However, this explanation appears highly unlikely, as probes presented during imagined articulation have been shown to lead to attenuated responses (Jack et al., 2019). Thus, for imaginary speech to account for the observed enhancement, one would have to assume that participants engaged in imagined speech prior to, but not during, probe presentation, a scenario highly unlikely to occur across participants without explicit instructions. While some of these findings, particularly those reported by S. Li et al. (2020), offer potential explanations for the enhancement of auditory responses during speech preparation, they, like the results of the present study, stand in direct conflict with the findings of Daliri and Max (2015a, 2015b, 2016). In addition to alternative theoretical interpretations (e.g., S. Li et al., 2020), the conflicting results might also be attributed to methodological differences. Differences between the present study and Daliri and Max (2015a, 2015b, 2016) include the trial structure. Daliri and Max (2015a, 2015b, 2016) utilized a block design, whereas the present study employed pseudo-randomized trials. Blocked designs might facilitate better participant engagement in preparation, allowing for refinement across trials within a block. Conversely, randomized trials may necessitate greater resource allocation for identifying trial types, potentially reducing resources available for preparation. Another difference lies in the longer preparation phase employed in the current study. This extended duration could have led participants to either delay preparation, knowing ample time was available, or to initiate preparation at the beginning of the phase, leading to high inter- and intraindividual variability (see below). Furthermore, the longer preparation phase afforded participants more time to identify and process the visual infor-

mation shown on the screen. Most importantly, Daliri and Max (2015a, 2015b, 2016) used words from a word list instead of merely syllables. One could argue that preparing different words is more demanding than preparing the same syllable, which is inherently less complex than a word. Therefore, a less complex preparation might lead to reduced suppression. Beal et al. (2010) observed SIS effects for the M50 component for words but not for vowels, suggesting this difference might be due to increased motor plan complexity and a consequent increase in the number or magnitude of EC-messages transmitted to the auditory cortex. However, no such difference was observed for the M100 component, which was suppressed for both vowels and words and is arguably more relevant to the present work. Conversely, Ventura et al. (2009) demonstrated that the amount of suppression decreases with increasing utterance complexity, meaning more suppression for simpler utterances, presumably because it is easier for the brain to predict and match the sensory consequences of an easier utterance compared to a complex one. It is important to note that both Beal et al. (2010) and Ventura et al. (2009) employed a talk/listen paradigm rather than presenting probes during a preparation phase, rendering their results not directly comparable to the current findings. Also, crucially, the present study did not find reduced or omitted suppression but rather observed enhancement.

4.1.1 Attention as a Confounding Factor

A potential reason for the observed enhancement could be a significant confounding factor that is often overlooked in active-passive experiments: Attention (Dogge et al., 2019; Horváth, 2015; Kiepe et al., 2021). It is highly probable that participants are more attentive during active trials than during passive ones. Consequently, active trials involve increased (selective) attention and motor preparation, while passive trials involve neither. Early work by Okita (1979) and Okamoto et al. (2007) already demonstrated that N1 amplitudes are enhanced during phases of increased attention. In an fMRI study, Reznik et al. (2014) also found enhanced activity in the auditory cortex during active sound generation compared to passive listening, potentially reflecting increased allocation of attentional resources. One might argue that the attention effect observed in the N1-P2 time-window is not caused by modifying the components themselves but by 'adding' another negative slow wave component (Okita, 1979), which should also be present in active no-probe trials. Since the participants were unaware of when probe or no-probe trials would occur, it is reasonable to assume similar attentional characteristics. Due to the applied correction (see Chapter 2 and Appendix F), one could argue that the slow wave attention effect (Okita, 1979)

should not be present in the final corrected ERP. However, it is emphasized once again that the subtraction only accounts for additive effects. It is possible that the effect is an interaction occurring solely during probe trials, leaving it unaffected by the applied correction (Horváth, 2015). It is also noteworthy that while Mock et al. (2011) did not find any effect on N1 amplitudes evoked by auditory probes during speech preparation compared to passive listening, they did observe a negative slow wave ERP component similar to what was described in early work by Okita (1979). This finding hints at attentional mechanisms having influenced the ERP results from Mock et al. (2011), as the authors also claim themselves. Although the attentional effects in the present study appear to also influence the N1 amplitude, the mechanisms underlying the observed effects might be similar to those described by Mock et al. (2011). In this regard, Mock et al. (2011) acknowledge that the observed slow wave component may also reflect a CNV, indicating motor preparation processes instead of reflecting solely attention. However, current density analyses and topographies in Mock et al. (2011) argue against slow wave differences between conditions being solely a manifestation of the CNV.

4.1.2 Load as a Modulator of Attentional Effects

Following the interpretation above, one might question why Daliri and Max (2015a, 2015b, 2016) observed suppression effects, given that attention should also be increased during active trials compared to passive ones in their studies. This discrepancy could be explained by the load theory (Lavie et al., 2004). The load theory of attention proposes that perceptual processing has a limited capacity but proceeds automatically and involuntarily on all information within its capacity. Specifically, high perceptual load during tasks exhausts capacity, leading to no awareness of unattended information. In low-load tasks, spare capacity allows for the perception of irrelevant information despite attempts to ignore it. In other words, task-irrelevant stimuli should elicit a brain response, even with a top-down bias to ignore them, in tasks involving low perceptual load (Molloy et al., 2015; Neelon et al., 2011). Conversely, high perceptual load should significantly reduce the brain response to task-irrelevant stimuli. For example, Molloy et al. (2015) showed decreased auditory M100 responses to task-irrelevant tones when visual load was high versus when it was low. Here, and in Daliri and Max (2015a, 2015b, 2016), the task-relevant 'stimuli' are motor preparation and the visual cue, while the auditory probes are irrelevant. Importantly, the load during the preparation phase is considerably higher in the studies by Daliri and Max (2015a, 2015b, 2016) compared to the present one. While participants in the present study had a relatively long preparation phase (3 s), they needed to be

much quicker (0.6 s) in the studies from Daliri and Max (2015a, 2015b, 2016). Additionally, and crucially, in the studies from Daliri and Max (2015a, 2015b, 2016), participants had to prepare alternating words instead of the same syllable, making both the perception of the visual cue (the word) (c.f., Molloy et al., 2015) and the preparation itself more complex. Therefore, although participants likely directed more attention towards motor preparation and the visual cue during active trials - both in the present study and in those by Daliri and Max (2015a, 2015b, 2016) - the low load in the current study may have allowed this attentional focus to 'spill over' to the irrelevant auditory cues, thereby enhancing N1 responses (Molloy et al., 2015; Neelon et al., 2011; Okamoto et al., 2007; Okita, 1979). During high load, as in the studies from Daliri and Max (2015a, 2015b, 2016), no attentional resources were left as the more complex and time-restricted preparation phase exhausted capacity, leading to reduced responses to irrelevant information (the auditory probes). Horváth and Burgýán (2013) and Horváth et al. (2012) also mentioned the possibility that pressing a button or performing an action draws attention away from task-irrelevant auditory stimulation for a short period, resulting in attenuated N1 amplitudes for tones presented close to that action (Horváth, 2015). Lavie et al. (2014) emphasize that load theory primarily focuses on perceptual, not cognitive, load. Therefore, it is argued that the observed effects are mostly driven by the more complex visual stimulus (changing words) in the studies from Daliri and Max (2015a, 2015b, 2016) compared to the same simple syllable in the present study. Related to this, Forster and Lavie (2009) showed that task load is negatively related to mind-wandering. Thus, the reduced load in the present study could have led to increased task-unrelated thoughts, affecting the preparation processes. In a worst-case scenario, it could have even caused participants to stop complying with the task instruction to prepare, focusing instead on task-unrelated matters. In this case, the attentional confound would still be present, as participants are thought to be more attentive during active trials since they needed to respond to the go-signal, but without motor preparation, resulting in enhanced auditory responses without any (assumed and seen in other studies) 'counterplay' from motor processes. However, regardless of whether only attentional effects were present or if attentional effects outweighed motor effects, the outcome is the same - enhanced neural responses during active trials. The same principle might also explain the results seen in S. Li et al. (2020), where specific preparation (i.e., preparing to speak while knowing what to say) led to increased N1 amplitudes as responses to congruent probes, while general preparation (i.e., preparing to speak without knowing what to say) led to suppressed N1 responses. One could argue that the load was relatively low

during the specific preparation stage, since, as in this study, only syllables were used and the delay period was relatively long (1.5 – 2 s). This could have caused the increased selective attention during the active trials to 'spill over' to the irrelevant auditory probes. One could further argue that during general preparation, the load is higher due to the uncertainty and increased abstractness, leading to suppression effects as no cognitive resources are left to process the irrelevant audio probes. It must be noted that Zheng et al. (2022) also found enhanced N1 responses for syllable probes during general preparation. Although the authors claim minor methodological reasons to account for the inconsistent results seen in Zheng et al. (2022) and S. Li et al. (2020), the robustness of the effect seen during general preparation is yet to be investigated. It is acknowledged that the assumption of increased load during general preparation is highly speculative, and it is stressed that 'true' motor-based effects in any of the mentioned studies are not ruled out.

4.1.3 Concluding Remarks on Attentional Modulation

Moving away from assumptions regarding load during the preparation/delay period, another attention-based mechanism may also account for the present findings. While in the account presented above, probes were categorized as task-irrelevant, as only motor preparation and the visual cue were to be focused on, one could also argue that the probes were relevant. This is because, if they were played, they were always presented before the visual go-signal, thus acting as a warning cue for the subsequent visual go-signal. This would render them task-relevant, consequently eliciting enhanced responses due to increased attention in active trials, regardless of the load (Lavie et al., 2004). Importantly, in the present study, auditory probes were presented in 50% of the trials, whereas in the studies by Daliri and Max, probes were presented in only one-third of the trials (Daliri & Max, 2015a) or in 40% of the trials (Daliri & Max, 2015b, 2016). Therefore, one could argue that probes could be used as a warning signal more reliably in the present study, thereby increasing their task relevance. However, it is questionable whether this small deviation significantly alters task relevance. Furthermore, in the present study, participants were explicitly instructed not to attend to the probes, as they had no relevance for the task. On the other hand, vocal onsets were significantly earlier during probe trials, indicating that the probe did indeed have some 'warning effect'. Importantly, it is not claimed that all studies using an active-passive paradigm of any sort misinterpreted their (suppression) results as motor-related processes. Even the results from Daliri and Max (2015a, 2015b, 2016) could be caused by motor/EC-related processes, as they claim. In fact,

studies investigating the effects of attention in active-passive paradigms repeatedly concluded that while attention effects are present, there is definitely a 'true' motor-induced suppression (Saupe et al., 2013; Timm et al., 2013). Therefore, the common interpretation of the suppression effect being related to motor processes (e.g., Chung et al., 2023; J. J. Li et al., 2024; Merrikhi et al., 2018; Reznik et al., 2018; Voss et al., 2008) appears to be valid. Most convincingly, Niziolek and Guenther (2013) showed that suppression effects differ between typical and atypical vocalisations. This difference cannot be attributed to task-related attentional shifts, as both cases represent active trials. However, it is argued that the effects seen in the present study are rather attributable to attentional instead of motor/EC-related processes. While only the N1 ERP component was investigated, it is worth noting that the general morphology of the ERPs is consistent with what one would expect from an attentional manipulation (Polich, 2007). The hypothesised attention-based effects highlight the critical need to control for this confounding factor in future research of this kind. While attention-related influences can confound all types of active-passive paradigms, studies examining modulatory effects during the preparation phase, where motor-based effects are likely to be weaker or even absent compared to execution, may be particularly vulnerable (Mock et al., 2011). This is because the relatively small effect during preparation might be more easily masked by confounding factors such as attentional shifts. However, it is also emphasized that the interpretation of motor processes and attention having opposite effects (suppression and enhancement, respectively), and thus suppression being solely related to motor processes (e.g., Pinheiro et al., 2018; Voss et al., 2008), is flawed, as the effects of attention depend on task-relevance and load (Lavie et al., 2004, 2014; Molloy et al., 2015; Neelon et al., 2011). Action-related auditory ERP modulation is likely a collection of separate effects (e.g., motor-based effects, attention, and predictability), probably with different underlying causes (Horváth, 2015). Heterogeneity in experimental parameters (e.g., varying task instructions) may lead to differential engagement of underlying mechanisms, leading to variable effects.

4.1.4 Lack of Specificity in Pre-Speech Auditory Modulation

Regardless of whether and how attention or motor/EC-related processes influenced N1 amplitudes, the findings are consistent with non-specific modulation. Early, late, altered, and unaltered probes were affected similarly, with no significant main effects or interactions observed. These findings align with results from Daliri and Max (2016), who also suggested rather unspecific N1 effects, as tone probes and syllable probes were similarly affected. Importantly, these results con-

trast with effects seen when investigating motor-based suppression effects during execution, where altered feedback leads to less suppression (e.g., Behroozmand and Larson, 2011; Heinks-Maldonado et al., 2005). This further suggests that the present results are mainly influenced by unspecific processes that affect any type of feedback, regardless of its congruence with the assumed EC. S. Li et al. (2020) argue that they observed specific effects during specific preparation, in that they found enhanced N1 amplitudes when the probe matched the prepared vocalisation but no enhanced N1 amplitudes when there was a mismatch. However, two points should be noted here. First, S. Li et al. (2020) introduced a mismatch by changing the syllable (i.e., preparing /ga/ and being presented /ba/) instead of altering a basic feature such as the F0. This introduced a phonemic mismatch instead of only a phonetic one, which may differently affect ERP components (Tian & Poeppel, 2013). Second, the pattern of results is important to note. While there was a significant difference between congruent preparation and passive listening, and no significant difference between incongruent preparation and passive listening, there was no significant difference between the congruent and incongruent preparation (S. Li et al., 2020). Therefore, there was also no significant difference in the enhancement effect between incongruent and congruent trials, making the observed and interpreted specificity at least questionable. Overall, it is suggested that the observed modulation in this study is dominantly caused by non-specific attention effects, unrelated to any motor-based predictive processes.

Although the explanations above may account for the N1 amplitude effects observed in the present study, they do not explain the lack of N1 latency effects observed in earlier research (Mock et al., 2011). The absence of an N1 latency effect, compared to Mock et al. (2011), might be attributed to the fact that, in the present study, mismatch and match were determined solely by a basic feature such as pitch, whereas Mock et al. (2011) used a congruence manipulation that was similar to S. Li et al. (2020). Potentially, while there appears to be some specificity in PSAM (Mock et al., 2011), it might only apply to higher-level, phonemic characteristics, instead of lower-level properties such as pitch.

4.2 Inter- and Intraindividual Differences During Speech Preparation

Lastly, the present study aimed to investigate speech preparation itself, independent of auditory responses. To this end, active and passive trials (all no-probe) were classified based on EEG data preceding the go-signal. Crucially, all features were extracted twice, once during an earlier time-window and once during

a later time-window, resulting in two feature vectors per trial. After dimension reduction via PCA, SVMs were used to perform the classification. It was hypothesised that speech preparation would become more specific over time. This would be reflected in a reduced complexity of the classification problem, indexed by a higher percentage of hyperparameter-pairs leading to an above-chance classification when using the late feature set. While descriptively, the observed pattern reflected the expected one, the difference was not significant, indicating that the preparation did not become more specific over the investigated time course. Multiple accounts may explain the present results. Overall, substantial variability was observed in the results. For some subjects, the predicted pattern occurred, for others, the opposite pattern was visible, and for some, no effect at all (see Figure 10 and Appendix D). One reason for the observed pattern could be inter- and intraindividual variability due to non-specific instructions or fatigue during the experiment. While the instructions specified that the participants should engage in speech preparation during active trials while staying passive during passive trials, they did not specify how to perform either. Therefore, it is possible that various preparation strategies were employed across participants, such as imagined speech, increased alertness, or other strategies, leading to unique activation patterns (Seghier & Price, 2018; Zorowitz & Niv, 2023). Furthermore, even within a participant, multiple different strategies could have been used. Given the length of the experiment, approximately 115 minutes, it is plausible that participants became unmotivated and consequently reduced the intensity of their preparation, changed their preparation strategy, or abandoned it altogether. In the beginning, when participants were highly motivated and had yet to learn the delay time, they potentially engaged in preparation right away. However, upon learning that the delay period was relatively long (3 s), they might have started to engage during a later phase, thus changing their strategy as their understanding of the task evolved (Zorowitz & Niv, 2023). Although participants were instructed to prepare during the preparation phase, it was not explicitly stated that they should begin immediately, although this was arguably implied. Nonetheless, even if this had been explicitly stated, there remains a high probability of participants forgetting or ignoring this instruction over the course of the experiment. Along these lines, data from reaction time studies show that intraindividual variability increases over the course of an experiment (C. Wang et al., 2014). Furthermore, mind-wandering is known to increase as a function of time (Thomson et al., 2014). Given the length of the experiment, it is plausible that changes in mind-wandering over time affected the investigated processes. Additionally, while the task itself was relatively simple, it is likely that some

fatigue occurred nonetheless, also affecting intraindividual variability known to influence both neural and behavioural measures (J. Liu et al., 2024). In addition to inconsistent preparation strategies, it was also not specified how to maintain passivity. Consequently, high inter- and intraindividual variability might also be present here (Horváth, 2015), leading to unique activation patterns (Seghier & Price, 2018; Zorowitz & Niv, 2023). Some participants may have engaged in mind-wandering or even imagined speech, while others attempted to 'clear their mind'. Additionally, due to the lengthy experiment, some may have even changed their 'passivity strategy' during the experiment, adding intraindividual variability that potentially affected the results. In summary, both interindividual differences, stemming from divergent strategies and interpretations of the task, and intraindividual changes, attributable to fatigue and fluctuations during a long experiment, are likely to have influenced the present results. Future studies could mitigate these issues by providing more detailed instructions to constrain strategy variability. Furthermore, post-experiment questionnaires assessing participants' strategies should be incorporated to better understand observed (null) results. While these factors likely also affected the ERP analyses, the higher Signal-to-Noise-Ratio (SNR) potentially led to milder influences of variability, whereas the single-trial classification analysis was more affected, potentially contributing to the observed null results in the latter.

4.3 Limitations

This study highlights several important limitations that may also be present in other studies investigating sensorimotor integration. Due to the extensive length of this paradigm, necessitated by the many conditions introduced by three two-way factors, participants may have disengaged from the task, ceasing any motor preparation and thereby giving more weight to the assumed confounding factor of attention. While the general mood of the participants was investigated after the experiment, no specific questionnaire investigated actual task compliance. Furthermore, as a specific preparation strategy was not defined, participants may have engaged in several different strategies, leading to high interindividual variability. Due to the long experiment, participants potentially even switched strategies during the experiment, also contributing to intraindividual variance. Future studies should specifically assess and/or control preparation strategies, as this might also explain differences across various studies. Following this argument, there might also be high variability in the passive conditions, as there were no detailed instructions regarding 'how to stay passive', which potentially also affected the results (Horváth, 2015). Furthermore, while we would argue that

the present study successfully achieved the manipulation of the match between auditory probes and the assumed EC of a yet-to-be-executed vocalisation, we acknowledge that the mismatch was introduced only via a F0 shift. Other factors likely also represented in the EC, such as formant structure, remained unchanged. While this was a deliberate decision to tightly control the acoustic properties of the probes by limiting the manipulation to a single dimension, it also results in a limitation, as other manipulations (potentially more 'natural' ones) may have led to different results. By far the largest limitation concerns the lack of a control condition for attentional effects. As stated above, it is highly probable that participants were generally more attentive during active trials compared to passive trials, introducing a significant confounding factor often shown to influence sensory processing and sensorimotor integration (Okamoto et al., 2007; Okita, 1979; Saupe et al., 2013; Zhang et al., 2024, 2025). Therefore, future research investigating sensorimotor integration using active-passive paradigms - particularly when examining the presumably weaker effects occurring during motor preparation compared to execution - should either employ an 'active-like' condition that does not involve motor processes, or investigate attentiveness during active and passive trials, in order to account for potential effects. The latter could be achieved by presenting additional stimuli that need to be detected, thereby increasing attentiveness in both conditions while simultaneously tracking it. When doing so, it might be valuable to investigate differences depending on the domain by presenting both, for example, visual and auditory stimuli, measuring attentiveness for both domains. However, a drawback of this approach would be that it dramatically increases the load, potentially altering the observed effects (Horváth, 2015; Lavie et al., 2004).

4.4 Conclusion

In conclusion, the present study observed enhanced N1 amplitudes during speech preparation, a finding that deviates from the hypothesised suppression effect. While motor-based enhancement accounts exist in the literature (e.g., S. Li et al., 2020), the comparison between our results and those showing suppression (e.g., Daliri and Max, 2016) strongly suggests that attentional mechanisms played a significant role in the observed enhancement. We propose that participants were more attentive during active trials than passive ones, which affected auditory processing. Specifically, we hypothesise that the relatively low load during the preparation phase in our paradigm, compared to previous studies, allowed attentional resources to 'spill over' to task-irrelevant auditory probes, thereby amplifying N1 responses (Lavie et al., 2004). This interpretation is further supported

by the non-specific modulation of N1 amplitudes across different probe types and probe onsets, indicating a general attentional influence rather than a precise, motor/EC-based mechanism. The single-trial classification analysis suggested that, contrary to our hypothesis, speech preparation did not get more specific over time. However, inconsistent patterns across participants suggest high inter- and (potentially) intraindividual variability. This outcome underscores the impact of factors such as unspecific task instructions, leading to diverse preparation strategies, and intraindividual changes due to fatigue or mind-wandering during the experimental session. These findings highlight the critical need for future research to implement more stringent controls over participants' strategies and compliance.

Ultimately, this study highlights a crucial methodological consideration in sensorimotor integration research, particularly within active-passive paradigms - the confounding influence of attention. While genuine motor-based suppression and enhancement effects are generally well-supported, the present findings advocate for the careful design of control conditions that dissociate attentional effects from motor-related modulations. Addressing these limitations in future studies, perhaps through explicit attentional manipulation or detailed post-experimental assessments, will be vital for a more precise understanding of PSAM and the complex interplay between motor systems and sensory processing. Maybe you're still 'hearing' your voice as you read this final paragraph - only now, perhaps, with a heightened awareness not only of the complex mechanisms that make this seemingly simple experience possible, but also of the challenges involved in studying them.

References

- Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and bayesian inference in the brain. *Current opinion in neurobiology*, 26, 219–227. <https://doi.org/10.1016/j.conb.2017.08.010>
- Alawee, W. H., Basem, A., & Al-Haddad, L. A. (2023). Advancing biomedical engineering: Leveraging hjorth features for electroencephalography signal analysis. *Journal of electrical bioimpedance*, 14(1), 66. <https://doi.org/10.2478/joeb-2023-0009>
- Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of cognitive neuroscience*, 21(4), 791–802. <https://doi.org/10.1162/jocn.2009.21055>
- Allen, J., Liu, H., Iqbal, S., Zheng, D., & Stansby, G. (2021). Deep learning-based photoplethysmography classification for peripheral arterial disease detection: A proof-of-concept study. *Physiological Measurement*, 42(5), 054002. <https://doi.org/10.1088/1361-6579/abf9f3>
- Arjmandi, M. K., & Behroozmand, R. (2024). On the interplay between speech perception and production: Insights from research and theories. *Frontiers in Neuroscience*, 18, 1347614. <https://doi.org/10.3389/fnins.2024.1347614>
- Arvin, B., Prepageran, N., & Raman, R. (2013). “high frequency presbycusis”—is there an earlier onset? *Indian Journal of Otolaryngology and Head & Neck Surgery*, 65(3), 480–484. <https://doi.org/10.1007/s12070-011-0356-x>
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3), 379–384. <https://doi.org/10.3758/BF03192707>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bäß, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory n1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology*, 70(2), 137–143. <https://doi.org/10.1016/j.ijpsycho.2008.06.005>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2018). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beal, D. S., Cheyne, D. O., Gracco, V. L., Quraan, M. A., Taylor, M. J., & de Nil, L. F. (2010). Auditory evoked fields to vocalization during passive

- listening and active generation in adults who stutter. *NeuroImage*, 52(4), 1645–1653. <https://doi.org/10.1016/j.neuroimage.2010.04.277>
- Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. R. (2009). Vocalization-induced enhancement of the auditory cortex responsiveness during voice f0 feedback perturbation. *Clinical Neurophysiology*, 120(7), 1303–1312. <https://doi.org/10.1016/j.clinph.2009.04.022>
- Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, 12(1), 54. <https://doi.org/10.1186/1471-2202-12-54>
- Behroozmand, R., Liu, H., & Larson, C. R. (2011). Time-dependent neural processing of auditory feedback during voice pitch error detection. *Journal of cognitive neuroscience*, 23(5), 1205–1217. <https://doi.org/10.1162/jocn.2010.21447>
- Behroozmand, R., Oya, H., Nourski, K. V., Kawasaki, H., Larson, C. R., Brugge, J. F., Howard, M. A., & Greenlee, J. D. (2016). Neural correlates of vocal production and motor control in human heschl's gyrus. *The Journal of Neuroscience*, 36(7), 2302. <https://doi.org/10.1523/JNEUROSCI.3305-14.2016>
- Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., & Greenlee, J. D. (2015). Sensory–motor networks involved in speech production and motor control: An fmri study. *NeuroImage*, 109, 418–428. <https://doi.org/10.1016/j.neuroimage.2015.01.040>
- Beño-Ruiz-de-la-Sierra, R. M., Arjona-Valladares, A., Fondevila Estevez, S., Fernández-Linsenbarth, I., Díez, Á., & Molina, V. (2023). Corollary discharge function in healthy controls: Evidence about self–speech and external speech processing. *European Journal of Neuroscience*, 58(7), 3705–3713. <https://doi.org/10.1111/ejn.16125>
- Berg, M., Fuchs, M., Wirkner, K., Loeffler, M., Engel, C., & Berger, T. (2017). The speaking voice in the general population: Normative data and associations to sociodemographic and lifestyle factors. *Journal of Voice*, 31(2), 257.e13–257.e24. <https://doi.org/10.1016/j.jvoice.2016.06.001>
- Blakemore, S.-J., Smith, J., Steel, R., Johnstone, C. E., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological medicine*, 30(5), 1131–1139. <https://doi.org/10.1017/s0033291799002676>
- Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *NeuroReport*, 11(11). <https://doi.org/10.1097/00001756-200008030-00002>

- Boersma, P., & Weenink, D. (2025). Praat: Doing phonetics by computer. <http://www.praat.org/>
- Bose, A., Nawani, H., Agarwal, S. M., Shivakumar, V., Kalmady, S. V., Shenoy, S., Sreeraj, V. S., Narayanaswamy, J. C., Kumar, D., & Venkatasubramanian, G. (2019). Effect of fronto-temporal transcranial direct current stimulation on corollary discharge in schizophrenia: A randomized, double-blind, sham-controlled mediation analysis study. *Schizophrenia Research*, 204, 411–412. <https://doi.org/10.1016/j.schres.2018.07.040>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Bradshaw, A. R., Lametti, D. R., & McGettigan, C. (2021). The role of sensory feedback in developmental stuttering: A review. *Neurobiology of Language*, 2(2), 308–334. https://doi.org/10.1162/nol_a_00036
- Brooks, J. X., & Cullen, K. E. (2019). Predictive sensing: The role of motor signals in sensory processing. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(9), 842–850. <https://doi.org/10.1016/j.bpsc.2019.06.003>
- Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14(4), 411–427. <https://doi.org/10.1007/s10339-013-0571-3>
- Brunelin, J., Mondino, M., Gassab, L., Haesebaert, F., Gaha, L., Suaud-Chagny, M.-F., Saoud, M., Mechri, A., & Poulet, E. (2012). Examining transcranial direct-current stimulation (tdcs) as a treatment for hallucinations in schizophrenia. *American Journal of Psychiatry*, 169(7), 719–724. <https://doi.org/10.1176/appi.ajp.2012.11071091>
- Cao, L., & Händel, B. (2019). Walking enhances peripheral visual processing in humans. *PLoS biology*, 17(10), e3000511. <https://doi.org/10.1371/journal.pbio.3000511>
- Cao, L., Thut, G., & Gross, J. (2017). The role of brain oscillations in predicting self-generated sounds. *NeuroImage*, 147, 895–903. <https://doi.org/10.1016/j.neuroimage.2016.11.001>
- 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. <https://doi.org/10.1073/pnas.1216827110>
- Chen, C.-M. A., Mathalon, D. H., Roach, B. J., Cavus, I., Spencer, D. D., & Ford, J. M. (2011). The corollary discharge in humans is related to synchronous

- neural oscillations. *Journal of cognitive neuroscience*, 23(10), 2892–2904. <https://doi.org/10.1162/jocn.2010.21589>
- Chung, L. K.-h., Jack, B. N., Griffiths, O., Pearson, D., Luque, D., Harris, A. W. F., Spencer, K. M., Le Pelley, M. E., So, S. H.-w., & Whitford, T. J. (2023). Neurophysiological evidence of motor preparation in inner speech and the effect of content predictability. *Cerebral Cortex*, 33(24), 11556–11569. <https://doi.org/10.1093/cercor/bhad389>
- Clayton, K. K., Williamson, R. S., Hancock, K. E., Tasaka, G.-I., Mizrahi, A., Hackett, T. A., & Polley, D. B. (2021). Auditory corticothalamic neurons are recruited by motor preparatory inputs. *Current Biology*, 31(2), 310–321. <https://doi.org/10.1016/j.cub.2020.10.027>
- Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, 250, 126–136. <https://doi.org/10.1016/j.jneumeth.2015.01.010>
- Corretge, R. (2024). Praat vocal toolkit. <https://www.praatvocaltoolkit.com>
- Daliri, A., & Max, L. (2015a). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain and language*, 150, 37–44. <https://doi.org/10.1016/j.bandl.2015.08.008>
- Daliri, A., & Max, L. (2015b). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and language*, 143, 59–68. <https://doi.org/10.1016/j.bandl.2015.03.002>
- Daliri, A., & Max, L. (2016). Modulation of auditory responses to speech vs. nonspeech stimuli during speech movement planning. *Frontiers in human neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00234>
- Daliri, A., & Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex*, 99, 55–68. <https://doi.org/10.1016/j.cortex.2017.10.019>
- DeepL GmbH. (2024). DeepL Translator [Accessed: 2025-08-06]. <https://www.deepl.com>
- Delorme, A., & Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Dhole, P. V. (2024). A review on: Eeg-based brain-computer interfaces for imagined speech classification. *Educational Administration: Theory and Practice*, 6019–6036. <https://doi.org/10.53555/kuey.v30i5.3894>

- Dogge, M., Custers, R., & Aarts, H. (2019). Moving forward: On the limits of motor-based forward models. *Trends in cognitive sciences*, 23(9), 743–753. <https://doi.org/10.1016/j.tics.2019.06.008>
- Duggirala, S. X., Schwartze, M., Goller, L. K., Linden, D. E. J., Pinheiro, A. P., & Kotz, S. A. (2024). Hallucination proneness alters sensory feedback processing in self-voice production. *Schizophrenia bulletin*, 50(5), 1147–1158. <https://doi.org/10.1093/schbul/sbae095>
- Eliades, S. J., & Tsunada, J. (2023). Effects of cortical stimulation on feedback-dependent vocal control in non-human primates. *The Laryngoscope*, 133(S2), 1–10. <https://doi.org/10.1002/lary.30175>
- Eliades, S. J., & Tsunada, J. (2024). Vocal error monitoring in the primate auditory cortex. *bioRxiv*, 2024.06.15.599151. <https://doi.org/10.1101/2024.06.15.599151>
- Eliades, S. J., & Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*, 89(4), 2194–2207. <https://doi.org/10.1152/jn.00627.2002>
- Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *NATURE*, 453(7198), 1102–1106. <https://doi.org/10.1038/nature06910>
- Eliades, S. J., & Wang, X. (2019). Corollary discharge mechanisms during vocal production in marmoset monkeys. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(9), 805–812. <https://doi.org/10.1016/j.bpsc.2019.06.008>
- Enikolopov, A. G., Abbott, L. F., & Sawtell, N. B. (2018). Internally generated predictions enhance neural and behavioral detection of sensory stimuli in an electric fish. *Neuron*, 99(1), 135–146.e3. <https://doi.org/10.1016/j.neuron.2018.06.006>
- Flägel, K., Galler, B., Steinhäuser, J., & Götz, K. (2019). Der „national aeronautics and space administration-task load index“ (nasa-tlx) – ein instrument zur erfassung der arbeitsbelastung in der hausärztlichen sprechstunde: Bestimmung der psychometrischen eigenschaften. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 147-148, 90–96. <https://doi.org/10.1016/j.zefq.2019.10.003>
- Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. *The Journal of Neuroscience*, 30(49), 16643. <https://doi.org/10.1523/JNEUROSCI.1809-10.2010>

- Ford, J. M., Mathalon, D. H., Kalba, S., Whitfield, S., Faustman, W. O., & Roth, W. T. (2001). Cortical responsiveness during talking and listening in schizophrenia: An event-related brain potential study. *Biological psychiatry*, 50(7), 540–549. [https://doi.org/10.1016/s0006-3223\(01\)01166-0](https://doi.org/10.1016/s0006-3223(01)01166-0)
- Ford, J. M., & Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of psychiatric research*, 38(1), 37–46. [https://doi.org/10.1016/S0022-3956\(03\)00095-5](https://doi.org/10.1016/S0022-3956(03)00095-5)
- Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? *International Journal of Psychophysiology*, 58(2-3), 179–189. <https://doi.org/10.1016/j.ijpsycho.2005.01.014>
- Ford, J. M., & Mathalon, D. H. (2019). Efference copy, corollary discharge, predictive coding, and psychosis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(9), 764–767. <https://doi.org/10.1016/j.bpsc.2019.07.005>
- Ford, J. M., Palzes, V. A., Roach, B. J., & Mathalon, D. H. (2014). Did i do that? abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophrenia bulletin*, 40(4), 804–812. <https://doi.org/10.1093/schbul/sbt072>
- Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: Auditory hallucinations in schizophrenia. *American Journal of Psychiatry*, 164(3), 458–466. <https://doi.org/10.1176/ajp.2007.164.3.458>
- Ford, J. M., Roach, B. J., & Mathalon, D. H. (2010). Assessing corollary discharge in humans using noninvasive neurophysiological methods. *Nature Protocols*, 5(6), 1160–1168. <https://doi.org/10.1038/nprot.2010.67>
- Ford, J. M., Roach, B. J., & Mathalon, D. H. (2021). Vocalizing and singing reveal complex patterns of corollary discharge function in schizophrenia. *International Journal of Psychophysiology*, 164, 30–40. <https://doi.org/10.1016/j.ijpsycho.2021.02.013>
- Forster, S., & Lavie, N. (2009). Harnessing the wandering mind: The role of perceptual load. *Cognition*, 111(3), 345–355. <https://doi.org/10.1016/j.cognition.2009.02.006>
- Frith, C. D. (2019). Can a problem with corollary discharge explain the symptoms of schizophrenia? *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(9), 768–769. <https://doi.org/10.1016/j.bpsc.2019.07.003>

- Fuehrer, E., Voudouris, D., Lezkan, A., Drewing, K., & Fiehler, K. (2022). Tactile suppression stems from specific sensorimotor predictions. *Proceedings of the National Academy of Sciences*, 119(20), e2118445119. <https://doi.org/10.1073/pnas.2118445119>
- Gale, D. J., Areshenkoff, C. N., Honda, C., Johnsrude, I. S., Flanagan, J. R., & Gallivan, J. P. (2021). Motor planning modulates neural activity patterns in early human auditory cortex. *Cerebral Cortex*, 31(6), 2952–2967. <https://doi.org/10.1093/cercor/bhaa403>
- Gale, D. J., Flanagan, J. R., & Gallivan, J. P. (2021). Human somatosensory cortex is modulated during motor planning. *The Journal of Neuroscience*, 41(27), 5909. <https://doi.org/10.1523/JNEUROSCI.0342-21.2021>
- Google. (2025). Gemini (july 2025 version) [Accessed: 2025-08-06]. <https://gemini.google.com/app?hl=de>
- Greenlee, J. D. W., Behroozmand, R., Larson, C. R., Jackson, A. W., Chen, F., Hansen, D. R., Oya, H., Kawasaki, H., & Howard, Matthew A., III. (2013). Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. *PLoS one*, 8(4), e60783. <https://doi.org/10.1371/journal.pone.0060783>
- 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. <https://doi.org/10.1016/j.bandl.2005.06.001>
- Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, 42(2), 180–190. <https://doi.org/10.1111/j.1469-8986.2005.00272.x>
- Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of general psychiatry*, 64(3), 286–296. <https://doi.org/10.1001/archpsyc.64.3.286>
- Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoecephalographic evidence for a precise forward model in speech production. *NeuroReport*, 17(13). <https://doi.org/10.1097/01.wnr.0000233102.43526.e9>
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145. <https://doi.org/10.1038/nrn3158>
- Horváth, J. (2015). Action-related auditory erp attenuation: Paradigms and hypotheses. *Brain Research*, 1626, 54–65. <https://doi.org/10.1016/j.brainres.2015.03.038>

- Horváth, J., & Burgýán, A. (2013). No evidence for peripheral mechanism attenuating auditory erps to self-induced tones. *Psychophysiology*, 50(6), 563–569. <https://doi.org/10.1111/psyp.12041>
- Horváth, J., Maess, B., Baess, P., & Tóth, A. (2012). Action–sound coincidences suppress evoked responses of the human auditory cortex in eeg and meg. *Journal of cognitive neuroscience*, 24(9), 1919–1931. https://doi.org/10.1162/jocn_a_00215
- Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current opinion in neurobiology*, 33, 174–181. <https://doi.org/10.1016/j.conb.2015.04.006>
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. <https://doi.org/10.1126/science.279.5354.1213>
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An meg study. *Journal of cognitive neuroscience*, 14(8), 1125–1138. <https://doi.org/10.1162/089892902760807140>
- Jack, B. N., Le Pelley, M. E., Han, N., Harris, A. W., Spencer, K. M., & Whitford, T. J. (2019). Inner speech is accompanied by a temporally-precise and content-specific corollary discharge. *NeuroImage*, 198, 170–180. <https://doi.org/10.1016/j.neuroimage.2019.04.038>
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of f0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3), 1246–1251. <https://doi.org/10.1121/1.1288414>
- Khalilian-Gourtani, A., Wang, R., Chen, X., Yu, L., Dugan, P., Friedman, D., Doyle, W., Devinsky, O., Wang, Y., & Flinker, A. (2024). A corollary discharge circuit in human speech. *bioRxiv*, 2022.09.12.507590. <https://doi.org/10.1101/2022.09.12.507590>
- Kiepe, F., Kraus, N., & Hesselmann, G. (2021). Sensory attenuation in the auditory modality as a window into predictive processing. *Frontiers in human neuroscience*, 15. <https://doi.org/10.3389/fnhum.2021.704668>
- Kilteni, K., & Ehrsson, H. H. (2022). Predictive attenuation of touch and tactile gating are distinct perceptual phenomena. *iScience*, 25(4), 104077. <https://doi.org/10.1016/j.isci.2022.104077>
- Kinahan, S. P., Saidi, P., Daliri, A., Liss, J., & Berisha, V. (2024). Electroencephalographic classification reveals atypical speech motor planning in stuttering adults. *Journal of Speech, Language, and Hearing Research*, 67(7), 2053–2076. https://doi.org/10.1044/2024_JSLHR-23-00635

- Klug, M., Jeung, S., Wunderlich, A., Gehrke, L., Protzak, J., Djebbara, Z., Argubi-Wollesen, A., Wollesen, B., & Gramann, K. (2022). The bemoobil pipeline for automated analyses of multimodal mobile brain and body imaging data. *bioRxiv*, 2022.09.29.510051. <https://doi.org/10.1101/2022.09.29.510051>
- Korka, B., Widmann, A., Waszak, F., Darriba, Á., & Schröger, E. (2022). The auditory brain in action: Intention determines predictive processing in the auditory system—a review of current paradigms and findings. *Psychonomic Bulletin & Review*, 29(2), 321–342. <https://doi.org/10.3758/s13423-021-01992-z>
- Kudo, N., Nakagome, K., Kasai, K., Araki, T., Fukuda, M., Kato, N., & Iwanami, A. (2004). Effects of corollary discharge on event-related potentials during selective attention task in healthy men and women. *Neuroscience Research*, 48(1), 59–64. <https://doi.org/10.1016/j.neures.2003.09.008>
- Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: Attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1641), 20130205. <https://doi.org/10.1098/rstb.2013.0205>
- Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of experimental psychology. General*, 133(3), 339–354. <https://doi.org/10.1037/0096-3445.133.3.339>
- Lenth, R. V. (2025). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.11.1, <https://rvlenth.github.io/emmeans/>]. <https://rvlenth.github.io/emmeans/>
- Li, J. J., Daliri, A., Kim, K. S., & Max, L. (2024). Does pre-speech auditory modulation reflect processes related to feedback monitoring or speech movement planning? *Neuroscience Letters*, 843, 138025. <https://doi.org/10.1016/j.neulet.2024.138025>
- Li, S., Zhu, H., & Tian, X. (2020). Corollary discharge versus efference copy: Distinct neural signals in speech preparation differentially modulate auditory responses. *Cerebral cortex (New York, N.Y. : 1991)*, 30(11), 5806–5820. <https://doi.org/10.1093/cercor/bhaa154>
- Liu, D., Chang, Y., Dai, G., Guo, Z., Jones, J. A., Li, T., Chen, X., Chen, M., Li, J., Wu, X., Liu, P., & Liu, H. (2023). Right, but not left, posterior superior temporal gyrus is causally involved in vocal feedback control. *NeuroImage*, 278, 120282. <https://doi.org/10.1016/j.neuroimage.2023.120282>

- Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice f level on the pitch-shift reflex. *The Journal of the Acoustical Society of America*, 122(6), 3671–3677. <https://doi.org/10.1121/1.2800254>
- Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology*, 122(5), 951–957. <https://doi.org/10.1016/j.clinph.2010.08.010>
- Liu, J., Zhu, Y., Cong, F., Björkman, A., Malesevic, N., & Antfolk, C. (2024). Analysis of modulations of mental fatigue on intra-individual variability from single-trial event related potentials. *Journal of Neuroscience Methods*, 406, 110110. <https://doi.org/10.1016/j.jneumeth.2024.110110>
- Mathalon, D. H., & Ford, J. M. (2008). Corollary discharge dysfunction in schizophrenia: Evidence for an elemental deficit. *Clinical EEG and neuroscience*, 39(2), 82–86. <https://doi.org/10.1177/155005940803900212>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Max, L., & Daliri, A. (2019). Limited pre-speech auditory modulation in individuals who stutter: Data and hypotheses. *Journal of Speech, Language, and Hearing Research*, 62(8S), 3071–3084. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0358
- Max, L., Daniels, J. C., Curet, K. M., & Cronin, K. L. (2008). Modulation of auditory and somatosensory processing during the planning of speech movements. *Proceedings of the 8th International Seminar on Speech Production*, 2, 41–44. https://www.researchgate.net/publication/289434900_Modulation_of_auditory_and_somatosensory_processing_during_the_planning_of_speech_movements
- Merrikhi, Y., Ebrahimpour, R., & Daliri, A. (2018). Perceptual manifestations of auditory modulation during speech planning. *Experimental Brain Research*, 236(7), 1963–1969. <https://doi.org/10.1007/s00221-018-5278-3>
- Mifsud, N. G., Oestreich, L. K. L., Jack, B. N., Ford, J. M., Roach, B. J., Mathalon, D. H., & Whitford, T. J. (2016). Self-initiated actions result in suppressed auditory but amplified visual evoked components in healthy participants. *Psychophysiology*, 53(5), 723–732. <https://doi.org/10.1111/psyp.12605>
- Mock, J. R., Foundas, A. L., & Golob, E. J. (2011). Modulation of sensory and motor cortex activity during speech preparation. *European Journal*

- of Neuroscience*, 33(5), 1001–1011. <https://doi.org/10.1111/j.1460-9568.2010.07585.x>
- Molloy, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses. *The Journal of Neuroscience*, 35(49), 16046. <https://doi.org/10.1523/JNEUROSCI.2931-15.2015>
- Mondino, M., Haesebaert, F., Poulet, E., Suaud-Chagny, M.-F., & Brunelin, J. (2015). Fronto-temporal transcranial direct current stimulation (tdcs) reduces source-monitoring deficits and auditory hallucinations in patients with schizophrenia. *Schizophrenia Research*, 161(2), 515–516. <https://doi.org/10.1016/j.schres.2014.10.054>
- Mondino, M., Jardri, R., Suaud-Chagny, M.-F., Saoud, M., Poulet, E., & Brunelin, J. (2016). Effects of fronto-temporal transcranial direct current stimulation on auditory verbal hallucinations and resting-state functional connectivity of the left temporo-parietal junction in patients with schizophrenia. *Schizophrenia bulletin*, 42(2), 318–326. <https://doi.org/10.1093/schbul/sbv114>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in quantitative methods for psychology*, 4(2), 61–64. <https://doi.org/10.20982/tqmp.01.1.p042>
- Müller, N., Leske, S., Hartmann, T., Székely, S., & Weisz, N. (2015). Listen to yourself: The medial prefrontal cortex modulates auditory alpha power during speech preparation. *Cerebral Cortex*, 25(11), 4029–4037. <https://doi.org/10.1093/cercor/bhu117>
- Nawani, H., Bose, A., Agarwal, S. M., Shivakumar, V., Chhabra, H., Subramaniam, A., Kalmady, S., Narayanaswamy, J. C., & Venkatasubramanian, G. (2014). Modulation of corollary discharge dysfunction in schizophrenia by tDCS: Preliminary evidence. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, 7(3), 486–488. <https://doi.org/10.1016/j.brs.2014.01.003>
- Nawani, H., Kalmady, S. V., Bose, A., Shivakumar, V., Rakesh, G., Subramaniam, A., Narayanaswamy, J. C., & Venkatasubramanian, G. (2014). Neural basis of tDCS effects on auditory verbal hallucinations in schizophrenia: A case report evidence for cortical neuroplasticity modulation. *The Journal of ECT*, 30(1). <https://doi.org/10.1097/YCT.0b013e3182a35492>
- Nayok, S. B., Bose, A., Bagali, K. B., Maity, K., Sreeraj, V. S., Shivakumar, V., & Venkatasubramanian, G. (2022). Accelerated tDCS improves corollary discharge deficit & prediction error signalling in schizophrenia: A case

- report. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, 15(5), 1218–1220. <https://doi.org/10.1016/j.brs.2022.08.014>
- Neelon, M. F., Williams, J. C., & Garell, P. C. (2011). Elastic attention: Enhanced, then sharpened response to auditory input as attentional load increases. *Frontiers in human neuroscience*, 5. <https://doi.org/10.3389/fnhum.2011.00041>
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience*, 33(29), 12090. <https://doi.org/10.1523/JNEUROSCI.1008-13.2013>
- Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? evidence from speech production. *The Journal of Neuroscience*, 33(41), 16110. <https://doi.org/10.1523/JNEUROSCI.2137-13.2013>
- Noguchi, K., Gel, Y. R., Brunner, E., & Konietzschke, F. (2012). Nparld: An r software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, 50(12), 1–23. <https://doi.org/10.18637/jss.v050.i12>
- Oberauer, K. (2022). The importance of random slopes in mixed models for bayesian hypothesis testing. *Psychological Science*, 33(4), 648–665. <https://doi.org/10.1177/09567976211046884>
- Oestreich, L. K., Mifsud, N. G., Ford, J. M., Roach, B. J., Mathalon, D. H., & Whitford, T. J. (2015). Subnormal sensory attenuation to self-generated speech in schizotypy: Electrophysiological evidence for a ‘continuum of psychosis’. *International Journal of Psychophysiology*, 97(2), 131–138. <https://doi.org/10.1016/j.ijpsycho.2015.05.014>
- Okamoto, H., Stracke, H., Wolters, C. H., Schmael, F., & Pantev, C. (2007). Attention improves population-level frequency tuning in human auditory cortex. *The Journal of Neuroscience*, 27(39), 10383. <https://doi.org/10.1523/JNEUROSCI.2963-07.2007>
- Okita, T. (1979). Event-related potentials and selective attention to auditory stimuli varying in pitch and localization. *Biological Psychology*, 9(4), 271–284. [https://doi.org/10.1016/0301-0511\(79\)90027-9](https://doi.org/10.1016/0301-0511(79)90027-9)
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434–447. <https://doi.org/10.1037/1082-989X.8.4.434>

- OpenAI. (2025). Chatgpt (july 2025 version) [Accessed: 2025-08-06]. <https://chat.openai.com>
- Ozker, M., Doyle, W., Devinsky, O., & Flinker, A. (2022). A cortical network processes auditory error signals during human speech production to maintain fluency. *PLoS biology*, 20(2), e3001493. <https://doi.org/10.1371/journal.pbio.3001493>
- Ozker, M., Yu, L., Dugan, P., Doyle, W., Friedman, D., Devinsky, O., & Flinker, A. (2024). Speech-induced suppression and vocal feedback sensitivity in human cortex. *bioRxiv*. <https://doi.org/10.1101/2023.12.08.570736>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior research methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pépiot, E. (2014). Male and female speech: A study of mean f0, f0 range, phonation type and speech rate in parisian french and american english speakers. *Speech Prosody*, 305–309. <https://doi.org/10.21437/SpeechProsody.2014-49>
- Pinheiro, A. P., Schwartze, M., Amorim, M., Coentre, R., Levy, P., & Kotz, S. A. (2020). Changes in motor preparation affect the sensory consequences of voice production in voice hearers. *Neuropsychologia*, 146, 107531. <https://doi.org/10.1016/j.neuropsychologia.2020.107531>
- Pinheiro, A. P., Schwartze, M., & Kotz, S. A. (2018). Voice-selective prediction alterations in nonclinical voice hearers. *Scientific Reports*, 8(1), 14717. <https://doi.org/10.1038/s41598-018-32614-9>
- Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198, 181–197. <https://doi.org/10.1016/j.neuroimage.2019.05.026>
- Polich, J. (2007). Updating p300: An integrative theory of p3a and p3b. *Clinical Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>

- Poulet, J. F. A., & Hedwig, B. (2002). A corollary discharge maintains auditory sensitivity during sound production. *NATURE*, 418(6900), 872–876. <https://doi.org/10.1038/nature00919>
- Press, C., Thomas, E. R., & Yon, D. (2023). Cancelling cancellation? sensorimotor control, agency, and prediction. *Neuroscience & Biobehavioral Reviews*, 145, 105012. <https://doi.org/10.1016/j.neubiorev.2022.105012>
- Prusty, S., Patnaik, S., & Dash, S. K. (2022). Skcv: Stratified k-fold cross-validation on ml classifiers for predicting cervical cancer. *Frontiers in Nanotechnology*, 4. <https://doi.org/10.3389/fnano.2022.972421>
- R Core Team. (2025). *R: A language and environment for statistical computing version 4.5.0*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Reznik, D., Henkin, Y., Schadel, N., & Mukamel, R. (2014). Lateralized enhancement of auditory cortex activity and increased sensitivity to self-generated sounds. *Nature Communications*, 5(1), 4059. <https://doi.org/10.1038/ncomms5059>
- Reznik, D., & Mukamel, R. (2019). Motor output, neural states and auditory perception. *Neuroscience & Biobehavioral Reviews*, 96, 116–126. <https://doi.org/10.1016/j.neubiorev.2018.10.021>
- Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia*, 119, 302–307. <https://doi.org/10.1016/j.neuropsychologia.2018.08.028>
- Roach, B. J., Ford, J. M., Biagioli, B., Hamilton, H. K., Ramsay, I. S., Fisher, M., Loewy, R., Vinogradov, S., & Mathalon, D. H. (2019). Efference copy/corollary discharge function and targeted cognitive training in patients with schizophrenia. *International Journal of Psychophysiology*, 145, 91–98. <https://doi.org/10.1016/j.ijpsycho.2018.12.015>
- Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1976). Brain wave components of the contingent negative variation in humans. *Science*, 191(4231), 1055–1057. <https://doi.org/10.1126/science.1251217>
- Roussel, C., Hughes, G., & Waszak, F. (2013). A preactivation account of sensory attenuation. *Neuropsychologia*, 51(5), 922–929. <https://doi.org/10.1016/j.neuropsychologia.2013.02.005>
- Rummell, B. P., Bikas, S., Babl, S. S., Gogos, J. A., & Sigurdsson, T. (2023). Altered corollary discharge signaling in the auditory cortex of a mouse model of schizophrenia predisposition. *Nature Communications*, 14(1), 7388. <https://doi.org/10.1038/s41467-023-42964-2>

- Sares, A. G., Deroche, M. L. D., Ohashi, H., Shiller, D. M., & Gracco, V. L. (2020). Neural correlates of vocal pitch compensation in individuals who stutter. *Frontiers in human neuroscience*, 14. <https://doi.org/10.3389/fnhum.2020.00018>
- Sares, A. G., Deroche, M. L. D., Shiller, D. M., & Gracco, V. L. (2018). Timing variability of sensorimotor integration during vocalization in individuals who stutter. *Scientific Reports*, 8(1), 16340. <https://doi.org/10.1038/s41598-018-34517-1>
- Saupe, K., Widmann, A., Trujillo-Barreto, N. J., & Schröger, E. (2013). Sensorial suppression of self-generated sounds and its dependence on attention. *International Journal of Psychophysiology*, 90(3), 300–310. <https://doi.org/10.1016/j.ijpsycho.2013.09.006>
- Scheerer, N. E., & Jones, J. A. (2018). Detecting our own vocal errors: An event-related study of the thresholds for perceiving and compensating for vocal pitch errors. *Neuropsychologia*, 114, 158–167. <https://doi.org/10.1016/j.neuropsychologia.2017.12.007>
- Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11(9), 1141–1152. <https://doi.org/10.1111/2041-210X.13434>
- Schneider, D. M., & Mooney, R. (2018). How movement modulates hearing. *Annual review of neuroscience*, 41, 553–572. <https://doi.org/10.1146/annurev-neuro-072116-031215>
- Schneider, D. M., Sundararajan, J., & Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. *NATURE*, 561(7723), 391–395. <https://doi.org/10.1038/s41586-018-0520-5>
- Seghier, M. L., & Price, C. J. (2018). Interpreting and utilising intersubject variability in brain function. *Trends in cognitive sciences*, 22(6), 517–530. <https://doi.org/10.1016/j.tics.2018.03.003>
- Shibasaki, H., & Hallett, M. (2006). What is the bereitschaftspotential? *Clinical Neurophysiology*, 117(11), 2341–2356. <https://doi.org/10.1016/j.clinph.2006.04.025>
- Singla, S., Dempsey, C., Warren, R., Enikolopov, A. G., & Sawtell, N. B. (2017). A cerebellum-like circuit in the auditory system cancels responses to self-generated sounds. *Nature Neuroscience*, 20(7), 943–950. <https://doi.org/10.1038/nn.4567>

- Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., & Ford, J. M. (2013). Auditory cortex processes variation in our own speech. *PloS one*, 8(12), e82925. <https://doi.org/10.1371/journal.pone.0082925>
- Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology*, 43(6), 482. <https://doi.org/10.1037/h0055479>
- Stephane, M., Dzemidzic, M., & Yoon, G. (2022). Altered corollary discharge in the auditory cortex could reflect louder inner voice experience in patients with verbal hallucinations, a pilot fmri study. *Schizophrenia Research*, 243, 475–480. <https://doi.org/10.1016/j.schres.2022.02.007>
- Straka, H., Simmers, J., & Chagnaud, B. P. (2018). A new perspective on predictive motor signaling. *Current Biology*, 28(5), R232–R243. <https://doi.org/10.1016/j.cub.2018.01.033>
- Subramaniam, K., Kothare, H., Mizuiri, D., Nagarajan, S. S., & Houde, J. F. (2018). Reality monitoring and feedback control of speech production are related through self-agency. *Frontiers in human neuroscience*, 12. <https://doi.org/10.3389/fnhum.2018.00082>
- Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, 18(1), 70–73. <https://doi.org/10.4103/0972-6748.57865>
- The MathWorks Inc. (2024). *Matlab version: 24.1.0.2628055 (r2024a)*. Natick, Massachusetts, United States. <https://www.mathworks.com>
- Thomson, D. R., Seli, P., Besner, D., & Smilek, D. (2014). On the link between mind wandering and task performance over time. *Consciousness and Cognition*, 27, 14–26. <https://doi.org/10.1016/j.concog.2014.04.001>
- Tian, X., Ding, N., Teng, X., Bai, F., & Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nature Human Behaviour*, 2(3), 225–234. <https://doi.org/10.1038/s41562-018-0305-8>
- Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in human neuroscience*, 6, 314. <https://doi.org/10.3389/fnhum.2012.00314>
- Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of cognitive neuroscience*, 25(7), 1020–1036. https://doi.org/10.1162/jocn_a_00381
- Tian, X., & Poeppel, D. (2015). Dynamics of self-monitoring and error detection in speech production: Evidence from mental imagery and meg. *Journal*

- of cognitive neuroscience, 27(2), 352–364. https://doi.org/10.1162/jocn_a_00692
- Timm, J., SanMiguel, I., Keil, J., Schröger, E., & Schönwiesner, M. (2014). Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *Journal of cognitive neuroscience*, 26(7), 1481–1489. https://doi.org/10.1162/jocn_a_00552
- Timm, J., SanMiguel, I., Saupe, K., & Schröger, E. (2013). The n1-suppression effect for self-initiated sounds is independent of attention. *BMC Neuroscience*, 14(1), 2. <https://doi.org/10.1186/1471-2202-14-2>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- Tsunada, J., & Eliades, S. J. (2020). Dissociation of unit activity and gamma oscillations during vocalization in primate auditory cortex. *The Journal of Neuroscience*, 40(21), 4158. <https://doi.org/10.1523/JNEUROSCI.2749-19.2020>
- Tsunada, J., Wang, X., & Eliades, S. J. (2024). Multiple processes of vocal sensory-motor interaction in primate auditory cortex. *Nature Communications*, 15(1), 3093. <https://doi.org/10.1038/s41467-024-47510-2>
- Vanhoutte, S., Cosyns, M., van Mierlo, P., Batens, K., Corthals, P., de Letter, M., van Borsel, J., & Santens, P. (2016). When will a stuttering moment occur? the determining role of speech motor preparation. *Neuropsychologia*, 86, 93–102. <https://doi.org/10.1016/j.neuropsychologia.2016.04.018>
- Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*, 10(1), 58. <https://doi.org/10.1186/1471-2202-10-58>
- Vercillo, T., O’Neil, S., & Jiang, F. (2018). Action–effect contingency modulates the readiness potential. *NeuroImage*, 183, 273–279. <https://doi.org/10.1016/j.neuroimage.2018.08.028>
- von Holst, E., & Mittelstaedt, H. (1950). Das reafferenzprinzip. *Naturwissenschaften*, 37(20), 464–476. <https://doi.org/10.1007/BF00622503>
- Voss, M., Ingram, J. N., Wolpert, D. M., & Haggard, P. (2008). Mere expectation to move causes attenuation of sensory signals. *PLoS one*, 3(8), e2866. <https://doi.org/10.1371/journal.pone.0002866>
- Wang, C., Ding, M., & Kluger, B. M. (2014). Change in intraindividual variability over time as a key metric for defining performance-based cognitive fatigability. *Brain and Cognition*, 85, 251–258. <https://doi.org/10.1016/j.bandc.2014.01.004>

- Wang, H., Ali, Y., & Max, L. (2024). Perceptual formant discrimination during speech movement planning. *PloS one*, 19(4), e0301514. <https://doi.org/10.1371/journal.pone.0301514>
- Wang, J., Mathalon, D. H., Roach, B. J., Reilly, J., Keedy, S. K., Sweeney, J. A., & Ford, J. M. (2014). Action planning and predictive coding when speaking. *NeuroImage*, 91, 91–98. <https://doi.org/10.1016/j.neuroimage.2014.01.003>
- Weiss, C., & Schütz-Bosbach, S. (2012). Vicarious action preparation does not result in sensory attenuation of auditory action effects. *Consciousness and Cognition*, 21(4), 1654–1661. <https://doi.org/10.1016/j.concog.2012.08.010>
- Whitford, T. J., Mathalon, D. H., Shenton, M. E., Roach, B. J., Bammer, R., Adcock, R. A., Bouix, S., Kubicki, M., de Siebenthal, J., & Rausch, A. C. (2011). Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. *Psychological medicine*, 41(5), 959–969. <https://doi.org/10.1017/S0033291710001376>
- Whitford, T. J. (2019). Speaking-induced suppression of the auditory cortex in humans and its relevance to schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(9), 791–804. <https://doi.org/10.1016/j.bpsc.2019.05.011>
- Whitford, T. J., Ford, J. M., Mathalon, D. H., Kubicki, M., & Shenton, M. E. (2012). Schizophrenia, myelination, and delayed corollary discharges: A hypothesis. *Schizophrenia bulletin*, 38(3), 486–494. <https://doi.org/10.1093/schbul/sbq105>
- Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in cognitive sciences*, 1(6), 209–216. [https://doi.org/10.1016/S1364-6613\(97\)01070-X](https://doi.org/10.1016/S1364-6613(97)01070-X)
- Wu, Y., Han, Z., & Tian, X. (2025). Motor-based prediction mediates implicit vocal imitation. *NeuroImage*, 310, 121169. <https://doi.org/10.1016/j.neuroimage.2025.121169>
- Yang, F., Zhu, H., Cao, X., Li, H., Fang, X., Yu, L., Li, S., Wu, Z., Li, C., Zhang, C., & Tian, X. (2024). Impaired motor-to-sensory transformation mediates auditory hallucinations. *PLoS biology*, 22(10), e3002836. <https://doi.org/10.1371/journal.pbio.3002836>
- Zeng, X., & Martinez, T. R. (2000). Distribution-balanced stratified cross-validation for accuracy estimation. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(1), 1–12. <https://doi.org/10.1080/095281300146272>

- Zhang, Y., Sarmukadam, K., Nejati, V., & Behroozmand, R. (2025). Functional neural oscillatory activities reveal the impact of attentional instructions on speech auditory feedback control. *Neuropsychologia*, 216, 109202. <https://doi.org/10.1016/j.neuropsychologia.2025.109202>
- Zhang, Y., Sarmukadam, K., Wang, Y., & Behroozmand, R. (2024). Effects of attentional instructions on the behavioral and neural mechanisms of speech auditory feedback control. *Neuropsychologia*, 201, 108944. <https://doi.org/10.1016/j.neuropsychologia.2024.108944>
- Zheng, X., Zhu, H., Li, S., & Tian, X. (2022). The generic inhibitory function of corollary discharge in motor intention: Evidence from the modulation effects of speech preparation on the late components of auditory neural responses. *eneuro*, 9(6), ENEURO.0309–22.2022. <https://doi.org/10.1523/ENEURO.0309-22.2022>
- Zorowitz, S., & Niv, Y. (2023). Improving the reliability of cognitive task measures: A narrative review. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 8(8), 789–797. <https://doi.org/10.1016/j.bpsc.2023.02.004>
- Zouridakis, G., Simos, P. G., & Papanicolaou, A. C. (1998). Multiple bilaterally asymmetric cortical sources account for the auditory n1m component. *Brain Topography*, 10(3), 183–189. <https://doi.org/10.1023/A:1022246825461>

Appendix A N1 Amplitude by Task Condition, Probe Type and Probe Onset

Supplementary Table A.1: N1 amplitude by task condition (TC), probe type (PT), and probe onset (PO). LMM fixed effects estimates.

	Estimate	Std. Error	df	t value	p value	
(Intercept)	-8.1159	0.8131	40.61	-9.982	1.72×10^{-12}	***
TC	1.2144	0.5326	189.00	2.280	0.0237	*
PO	-0.9481	0.5326	189.00	-1.780	0.0766	.
PT	0.7224	0.5326	189.00	1.356	0.1766	
TC × PO	0.5532	0.7532	189.00	0.735	0.4635	
TC × PT	-0.6806	0.7532	189.00	-0.904	0.3673	
PO × PT	0.9113	0.7532	189.00	1.210	0.2278	
TC × PO × PT	-0.6007	1.0652	189.00	-0.564	0.5735	

Significance codes: *** 0.001, ** 0.01, * 0.05, . 0.1

Abbreviations: TC = Task Condition, PT = Probe Type, PO = Probe Onset

Supplementary Table A.2: Planned post hoc tests - N1 amplitude. Estimated Marginal Means and Pairwise Contrast.

Estimated Marginal Means					
Task Condition	Emmean	SE	df	Lower CL	Upper CL
Active	-8	0.745	28.8	-9.52	-6.48
Passive	-7	0.745	28.8	-8.52	-5.48
Pairwise Contrast					
Contrast	Estimate	SE	df	t ratio	p value
Active - Passive	-1	0.266	189	-3.757	0.0002

Results are averaged over the levels of Probe Type and Probe Onset.

Degrees-of-freedom method: Kenward-Roger.

Confidence level: 0.95.

P-value adjustment: Bonferroni.

Appendix B PSAM Effect by Probe Type and Probe Onset

Supplementary Table B.3: PSAM effect by probe type (PT) and probe onset (PO). LMM fixed effects estimates.

	Estimate	Std. Error	df	t value	p value	
(Intercept)	-1.2144	0.5281	101.88	-2.300	0.0235	*
PO	-0.5532	0.6920	81.00	-0.799	0.4263	
PT	0.6806	0.6920	81.00	0.984	0.3282	
PO × PT	0.6007	0.9786	81.00	0.614	0.5411	

Significance codes: *** 0.001, ** 0.01, * 0.05, . 0.1

Abbreviations: PT = Probe Type, PO = Probe Onset

Supplementary Table B.4: Planned post hoc tests - PSAM Effect. Estimated Marginal Means and Pairwise Contrasts.

Estimated Marginal Means					
Probe Onset = Early					
Probe Type	Emmean	SE	df	Lower CL	Upper CL
Altered	-1.214	0.528	102	-2.26	-0.17
Unaltered	-0.534	0.528	102	-1.58	0.51
Probe Onset = Late					
Probe Type	Emmean	SE	df	Lower CL	Upper CL
Altered	-1.768	0.528	102	-2.82	-0.72
Unaltered	-0.486	0.528	102	-1.53	0.56
Pairwise Contrasts					
Probe Onset = Early					
Contrast	Estimate	SE	df	t ratio	p value
Altered - Unaltered	-0.681	0.692	81	-0.984	0.3282
Probe Onset = Late					
Contrast	Estimate	SE	df	t ratio	p value
Altered - Unaltered	-1.281	0.692	81	-1.852	0.0677

Degrees-of-freedom method: Kenward-Roger.

Confidence level: 0.95.

P-value adjustment: Bonferroni.

Appendix C N1 Latency by Task Condition, Probe Type and Probe Onset

Supplementary Table C.5: N1 latency by task condition (TC), probe type (PT), and probe onset (PO). LMM fixed effects estimates.

	Estimate	Std. Error	df	t value	p value	
(Intercept)	116.8571	2.5689	94.19	45.489	$< 2 \times 10^{-16}$	***
TC	2.9286	2.7433	189.00	1.068	0.2871	
PO	0.1786	2.7433	189.00	0.065	0.9482	
PT	0.7857	2.7433	189.00	0.286	0.7749	
TC × PO	-6.8571	3.8796	189.00	-1.767	0.0788	.
TC × PT	-7.2857	3.8796	189.00	-1.878	0.0619	.
PO × PT	-1.9286	3.8796	189.00	-0.497	0.6197	
TC × PO × PT	7.6429	5.4866	189.00	1.393	0.1653	

Significance codes: *** 0.001, ** 0.01, * 0.05, . 0.1

Abbreviations: TC = Task Condition, PT = Probe Type, PO = Probe Onset

Supplementary Table C.6: Planned post hoc tests - N1 latency. Estimated Marginal Means and Pairwise Contrasts.

Estimated Marginal Means					
Task Condition = Active					
Probe Type	Emmean	SE	df	Lower CL	Upper CL
Altered	117	2.170	53.6	113	121
Unaltered	117	2.170	53.6	112	121
Task Condition = Passive					
Probe Type	Emmean	SE	df	Lower CL	Upper CL
Altered	116	2.170	53.6	112	121
Unaltered	113	2.170	53.6	108	117
Pairwise Contrasts					
Task Condition = Active					
Contrast	Estimate	SE	df	t ratio	p value
Altered - Unaltered	0.179	1.940	189	0.092	0.9268
Task Condition = Passive					
Contrast	Estimate	SE	df	t ratio	p value
Altered - Unaltered	3.643	1.940	189	1.878	0.0619

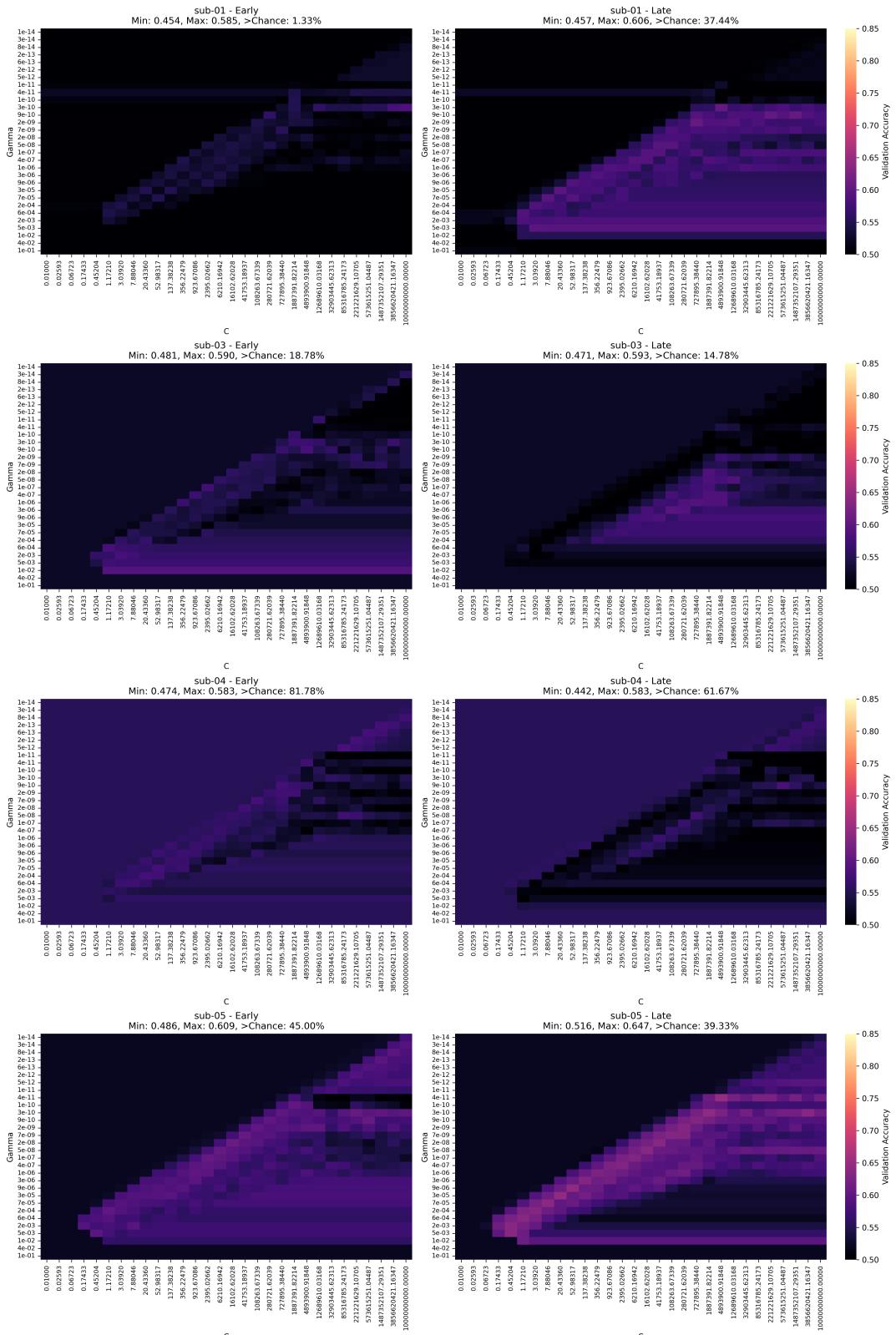
Results are averaged over the levels of Probe Onset

Degrees-of-freedom method: Kenward-Roger.

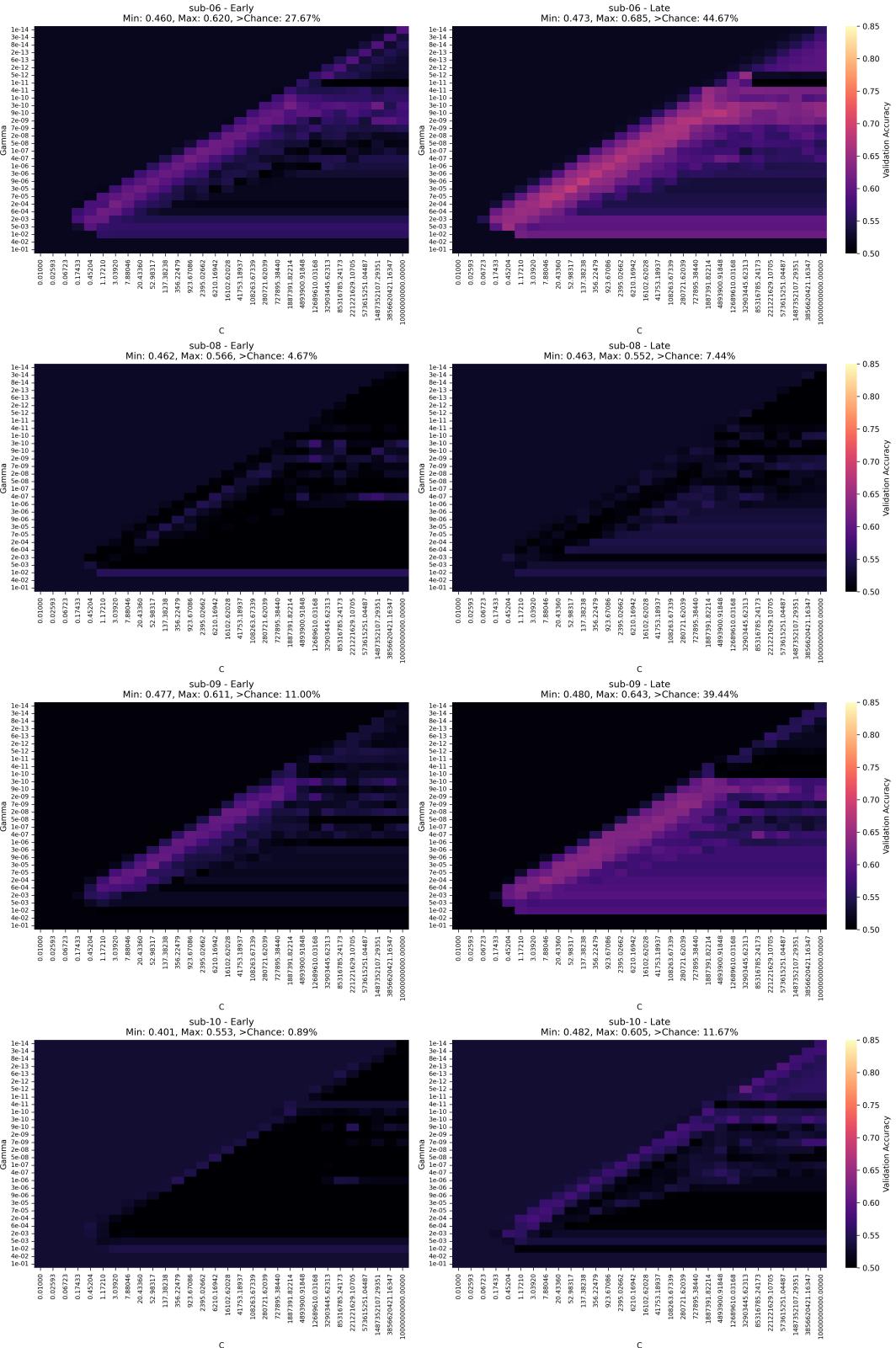
Confidence level: 0.95.

P-value adjustment: Bonferroni.

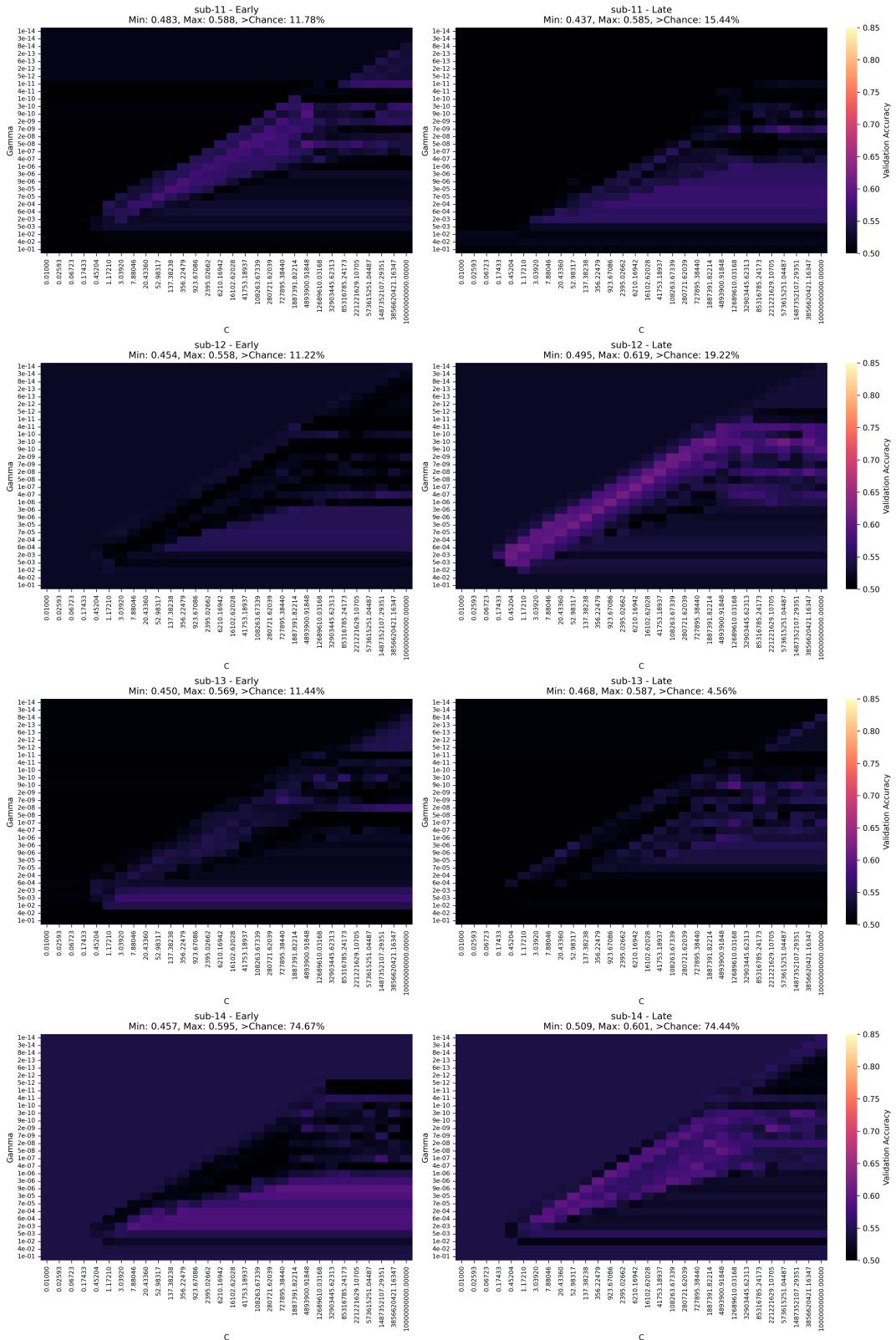
Appendix D Individual Hyperparameter Accuracy Heatmaps



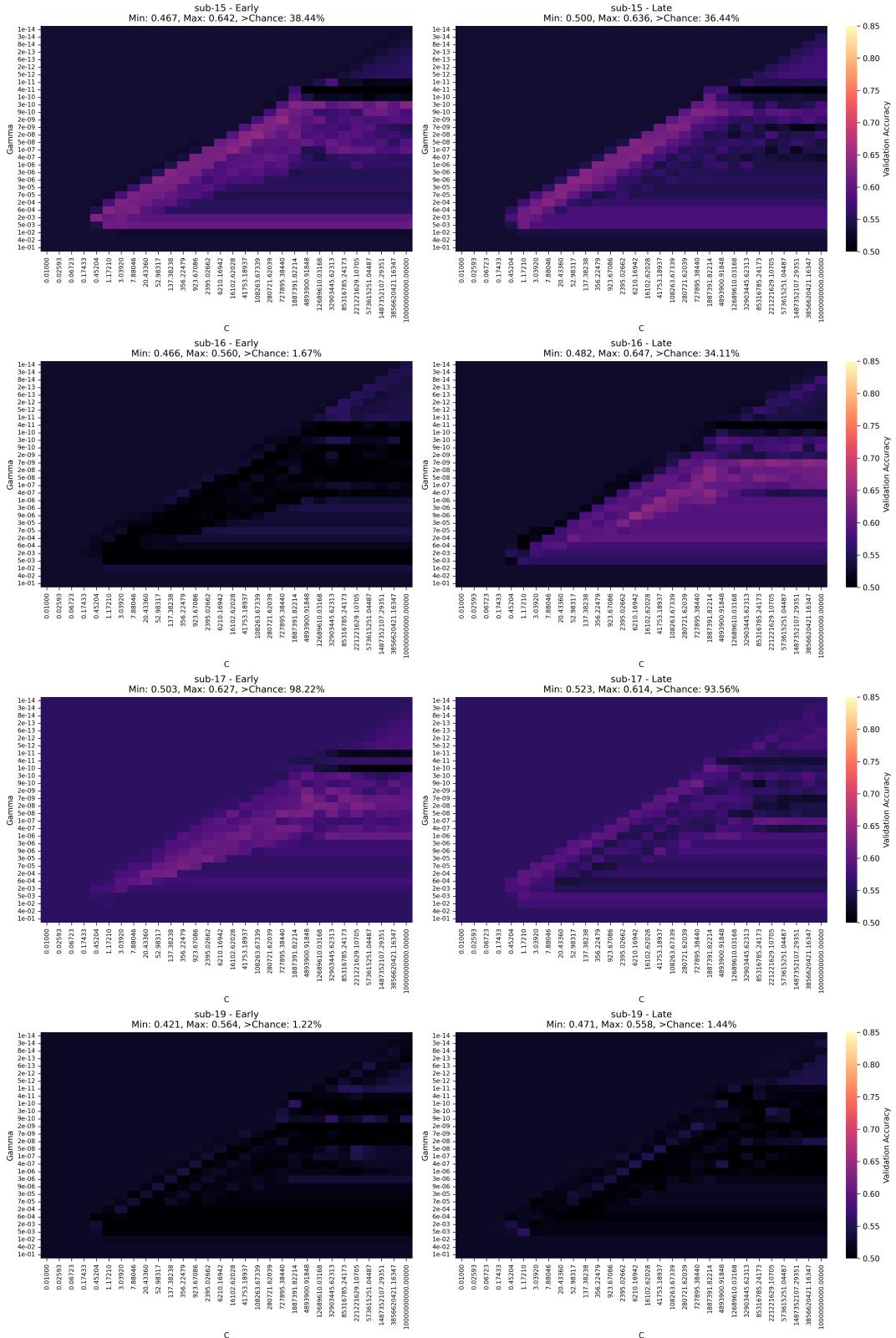
Supplementary Figure D.1: Individual hyperparameter accuracy heatmaps for sub-01, sub-03, sub-04 and sub-05. Min/Max indicate the lowest and highest validation accuracy in the search space. >Chance shows the proportion of hyperparameter-pairs yielding in an above-chance classification. Gamma values shown in scientific notation, rounded to one-digit mantissas for readability.



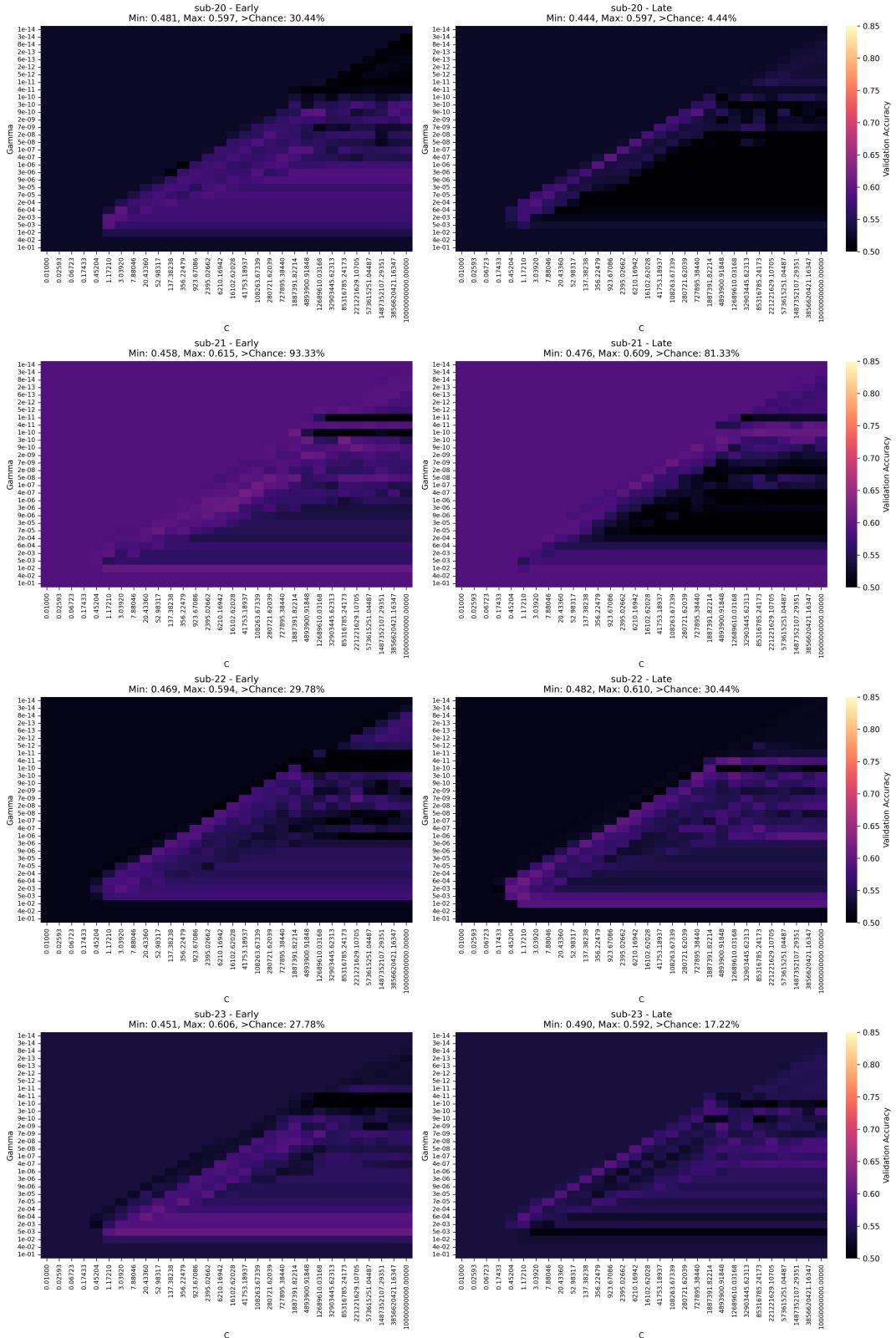
Supplementary Figure D.2: Individual hyperparameter accuracy heatmaps for sub-06, sub-08, sub-09 and sub-10. Min/Max indicate the lowest and highest validation accuracy in the search space. >Chance shows the proportion of hyperparameter-pairs yielding in an above-chance classification. Gamma values shown in scientific notation, rounded to one-digit mantissas for readability.



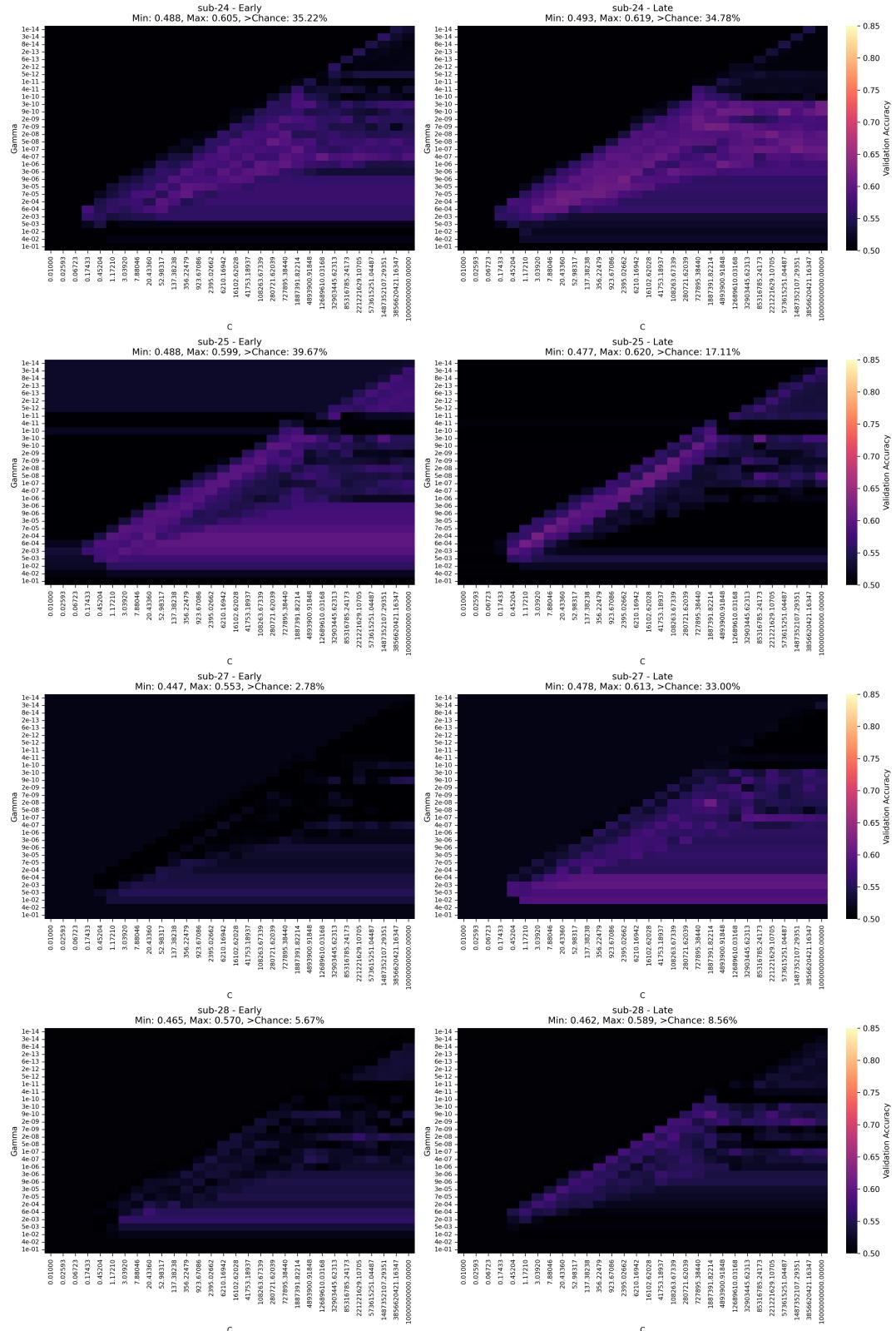
Supplementary Figure D.3: Individual hyperparameter accuracy heatmaps for sub-11, sub-12, sub-13 and sub-14. Min/Max indicate the lowest and highest validation accuracy in the search space. >Chance shows the proportion of hyperparameter-pairs yielding in an above-chance classification. Gamma values shown in scientific notation, rounded to one-digit mantissas for readability.



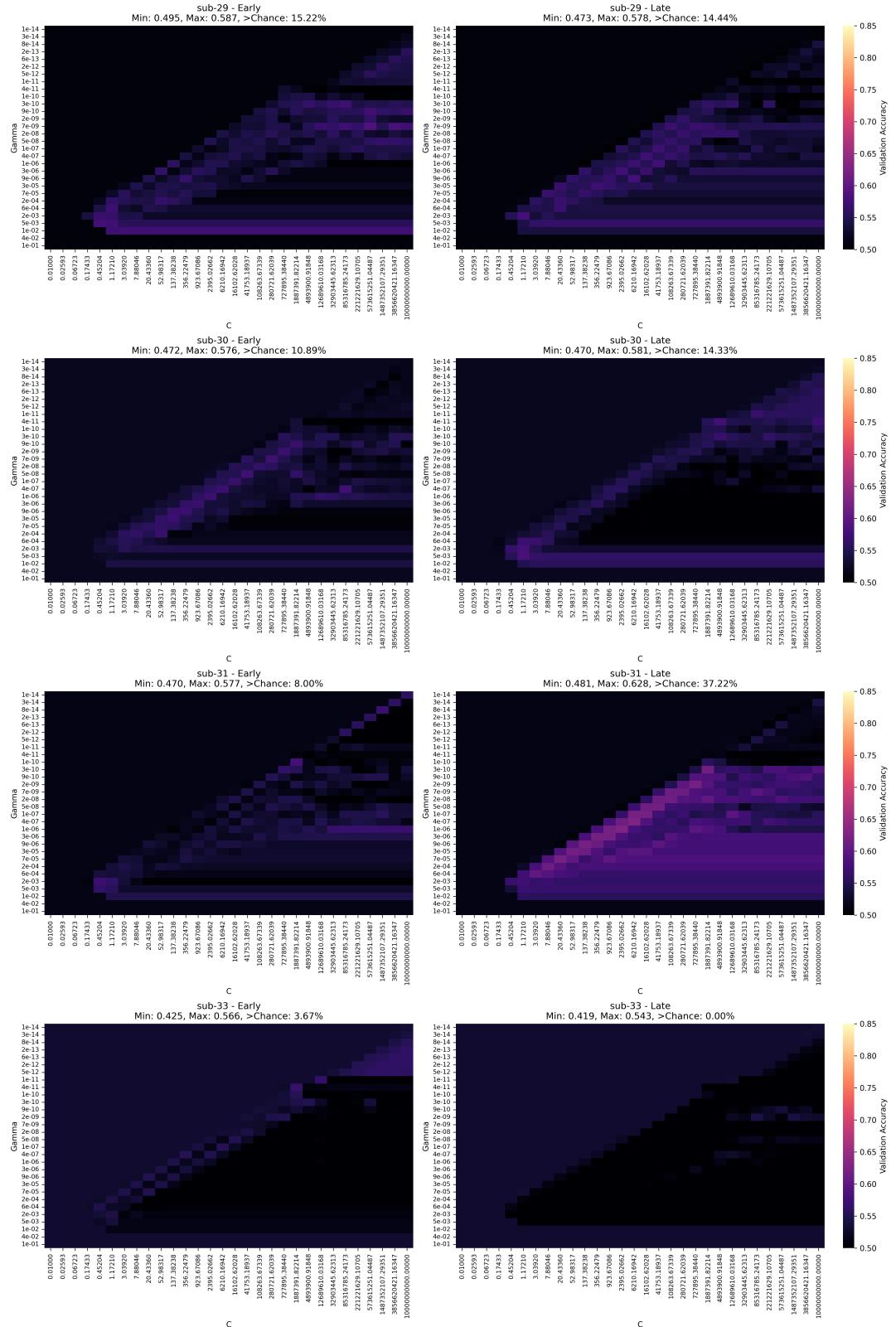
Supplementary Figure D.4: Individual hyperparameter accuracy heatmaps for sub-15, sub-16, sub-17 and sub-19. Min/Max indicate the lowest and highest validation accuracy in the search space. >Chance shows the proportion of hyperparameter-pairs yielding in an above-chance classification. Gamma values shown in scientific notation, rounded to one-digit mantissas for readability.



Supplementary Figure D.5: Individual hyperparameter accuracy heatmaps for sub-20, sub-21, sub-22 and sub-23. Min/Max indicate the lowest and highest validation accuracy in the search space. >Chance shows the proportion of hyperparameter-pairs yielding in an above-chance classification. Gamma values shown in scientific notation, rounded to one-digit mantissas for readability.

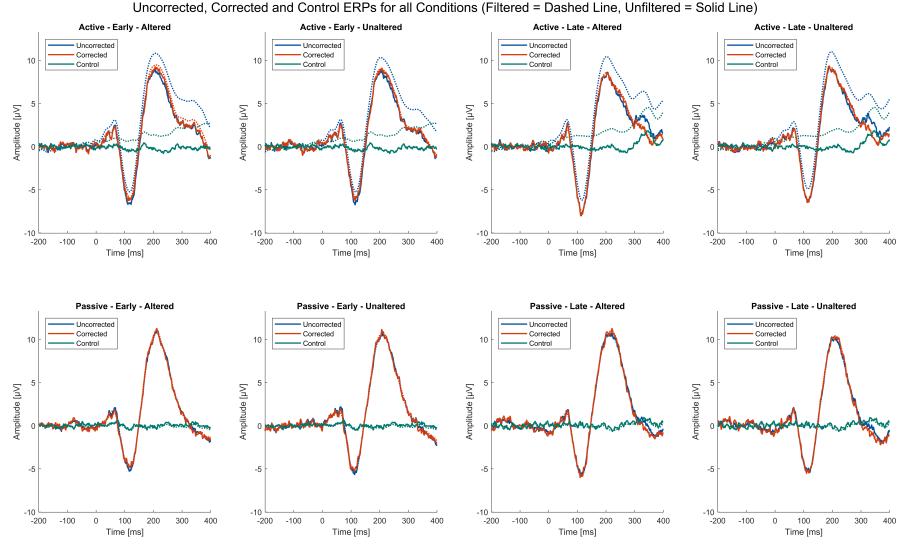


Supplementary Figure D.6: Individual hyperparameter accuracy heatmaps for sub-24, sub-25, sub-27 and sub-28. Min/Max indicate the lowest and highest validation accuracy in the search space. >Chance shows the proportion of hyperparameter-pairs yielding in an above-chance classification. Gamma values shown in scientific notation, rounded to one-digit mantissas for readability.

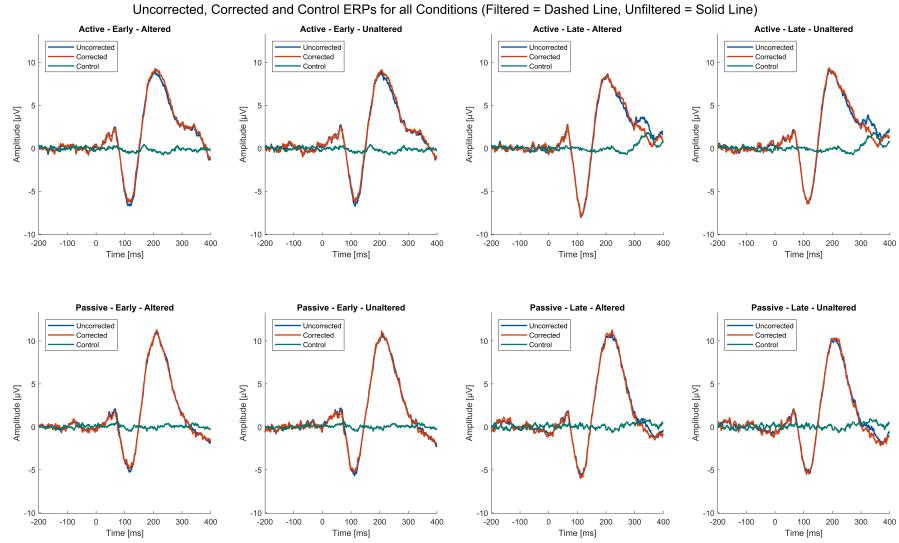


Supplementary Figure D.7: Individual hyperparameter accuracy heatmaps for sub-29, sub-30, sub-31 and sub-33. Min/Max indicate the lowest and highest validation accuracy in the search space. >Chance shows the proportion of hyperparameter-pairs yielding in an above-chance classification. Gamma values shown in scientific notation, rounded to one-digit mantissas for readability.

Appendix E Filter Characteristics

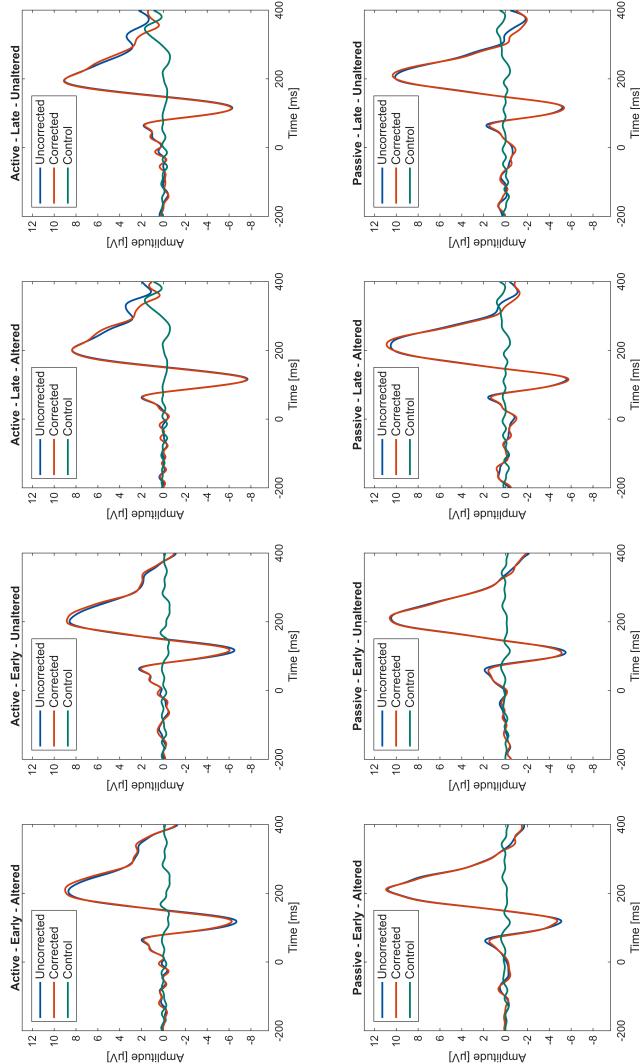


Supplementary Figure E.8: Filter characteristics of the preregistered pipeline. Comparison of non-filtered and filtered signals using the preregistered pipeline, which included both a high-pass filter (windowed sinc FIR, -6 dB cut-off at 0.3 Hz, order 5500, Hamming window) and a low-pass filter (windowed sinc FIR, -6 dB cut-off at 30 Hz, order 440, Hamming window). The filtered signal shows artifacts that are likely introduced by the high-pass filter.



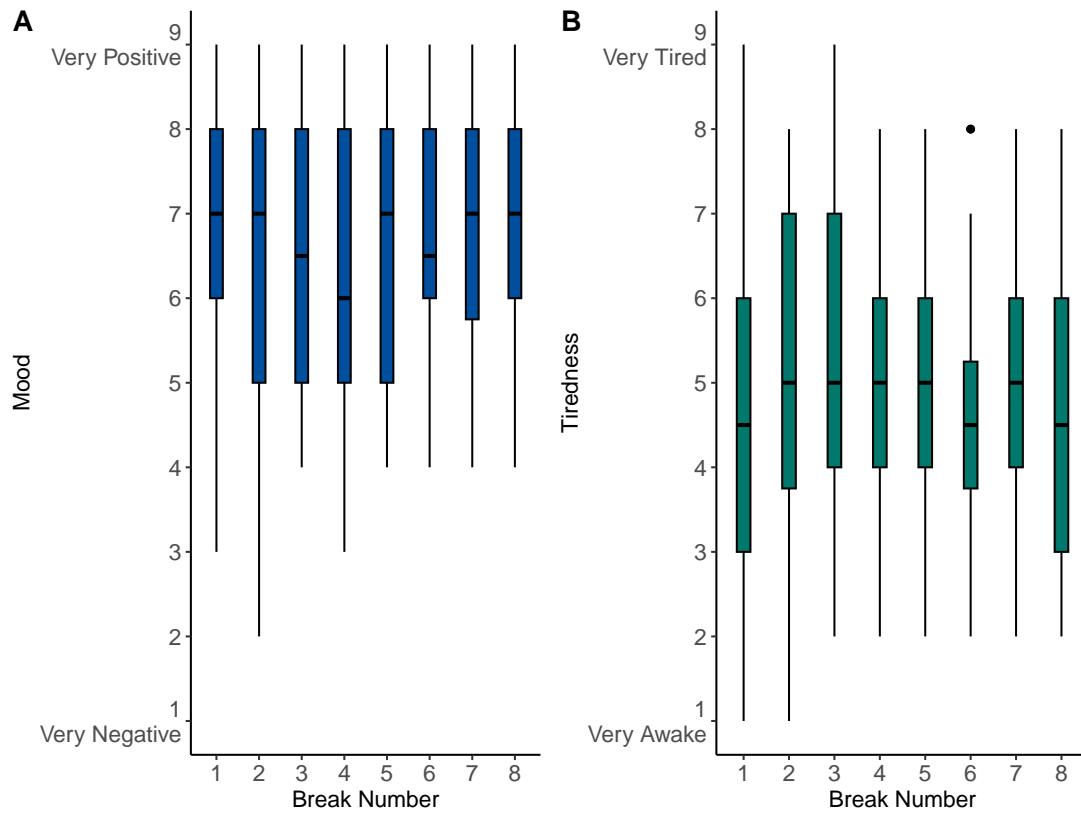
Supplementary Figure E.9: Filter characteristics of the final pipeline. Characteristics of the low-pass filter applied in the final preprocessing pipeline. The high-pass filter was omitted to prevent the introduction of artifacts. The final low-pass filter (windowed sinc FIR, -6 dB cut-off at 30 Hz, order 440, Hamming window) preserves signal integrity without introducing distortions.

Appendix F Grand Average Uncorrected, Corrected and Control ERPs



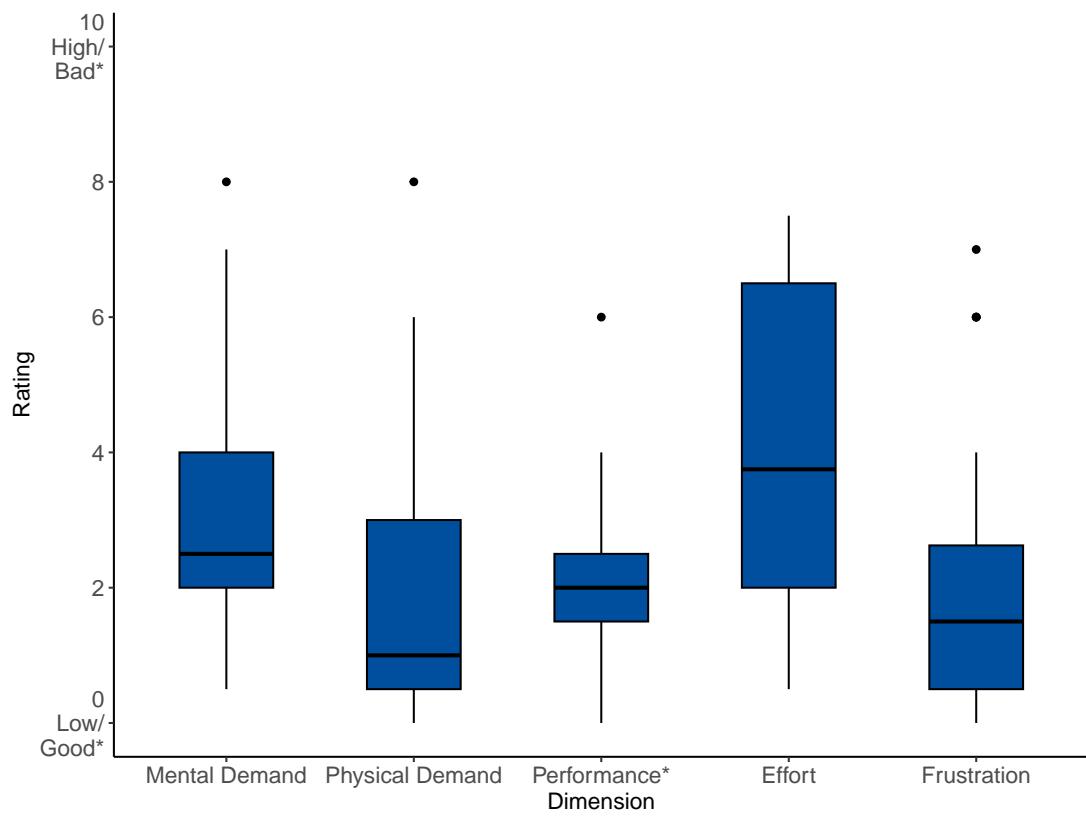
Supplementary Figure F.10: Grand average uncorrected, corrected and control ERPs. Auditory ERPs were isolated by subtracting control ERPs (green) - extracted from no-probe trials time-locked to hypothetical early or late probe onsets - from probe ERPs (blue). This procedure removes non-auditory components and approximates the true auditory response (orange), based on the assumption of additive effects (also see Daliri and Max, 2015a, 2015b, 2016).

Appendix G SAM Ratings over Time



Supplementary Figure G.11: SAM ratings over time. Mood (A) and tiredness (B) over time as measured with the SAM questionnaire. Values for the mood dimension range from 1 (very negative) to 9 (very positive). Values for the tiredness dimension range from 1 (very awake) to 9 (very tired).

Appendix H NASA-TLX Ratings



Supplementary Figure H.12: NASA-TLX Ratings. All dimensions except for performance are rated from 0 (low) to 10 (high), whereas the performance dimension is rated from 0 (good) to 10 (bad).

Appendix I Assessment of Initial State

tid_psam	FAL sub-__	Tim Dreßler
	INSTITUT FÜR PSYCHOLOGIE ABTEILUNG NEUROPSYCHOLOGIE	<p>Institut für Psychologie AG Neuropsychologie Prof. Dr. Stefan Debener</p> <p>Ansprechpartner für eventuelle Rückfragen: Prof. Dr. Stefan Debener Telefon: +49 (0)441 798 4271 E-mail: stefan.debener@uol.de</p>

Fragebogen zur Ausgangslage (FAL) – sub-__

Bitte beantworten Sie die folgenden Fragen gewissenhaft. Ihre Angaben werden vertraulich behandelt und nicht an Dritte weitergegeben!

1. Wie alt sind Sie? _____ Jahre
2. Sind Sie:
 - rechtshändig
 - linkshändig
 - beidhändig.
3. Geschlecht:
 - männlich
 - weiblich
 - divers.
4. Welchen Schulabschluss haben Sie? _____
5. Welcher Tätigkeit gehen Sie z.Zt. nach?
 - Studium der Fachrichtung _____
 - berufstätig als _____
 - nicht berufstätig, sondern berentet/arbeitslos/sonstiges (bitte zutreffendes unterstreichen)
6. Leiden Sie unter einer Beeinträchtigung des Hörvermögens?
 - Nein Ja, _____ (bitte Grund/Diagnose angeben)
7. Leiden Sie unter Ohrgeräuschen?
 - Nein
 - Ja
8. Wie viele Stunden haben Sie letzte Nacht geschlafen? _____ Stunden
9. Diese Schlafdauer war für mich:
 - normal
 - eher lang
 - viel zu kurz, ich bin ziemlich müde.
10. Haben Sie gestern Alkohol getrunken?
 - Nein Ja, _____ (bitte Menge angeben)

Bitte wenden!

tid_psam	FAL sub-_____	Tim Dreßler
----------	---------------	-------------

11. Haben Sie heute Alkohol getrunken?

Nein
 Ja, _____ (bitte Menge angeben)

12. Falls Sie Raucher sind, haben Sie heute:

sehr viel
 normal
 wenig / gar nicht geraucht?

13. Falls Sie Kaffee, Tee oder Cola trinken, haben Sie heute:

sehr viel
 normal
 wenig/keine

14. Vor wie vielen Stunden haben Sie zum letzten Mal gegessen? _____ Stunden

15. Sind Sie z.Zt. in neurologischer (inkl. z.B. wegen Stotterns), psychiatrischer oder psychologischer Behandlung?

Nein
 Ja, wegen _____
(bitte Grund/Diagnose und Dauer der Behandlung angeben)

16. Waren Sie bereits früher einmal in neurologischer (inkl. z.B. wegen Stotterns), psychiatrischer/psychologischer Behandlung?

Nein
 Ja, wegen _____
(bitte Grund/Diagnose und Zeitraum angeben)

17. Sind Sie z.Zt. in ärztlicher Behandlung (außer beim Neurologen/Psychiater)?

Nein
 Ja, wegen _____

18. Nehmen Sie z.Zt. Medikamente ein?

Nein
 Ja, _____
(bitte Namen des Medikamentes angeben)

19. Nehmen Sie Drogen oder haben Sie schon mal regelmäßig Drogen genommen?

Nein
 Ja, _____
(bitte Namen des Medikamentes angeben)

Vielen Dank!

Appendix J SAM

tid_psam

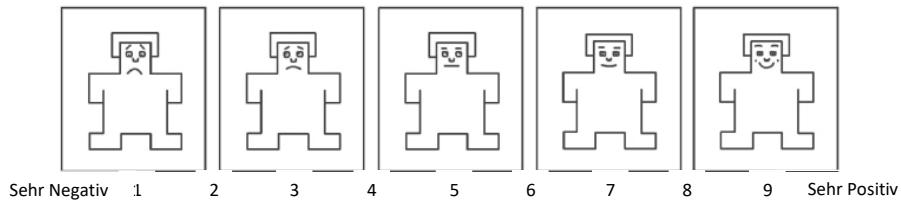
SAM sub-____

Tim Dreßler

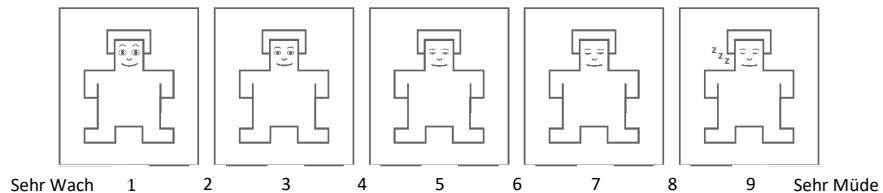


Fragebogen zur Gefühlslage und Müdigkeit – Pause: 1 sub-____

Wie würden Sie Ihre aktuelle Gefühlslage beschreiben. Bitte kreuzen Sie eine Zahl an.



Wie würden Sie Ihre aktuelle Müdigkeit beschreiben. Bitte kreuzen Sie eine Zahl an.



Appendix K NASA-TLX

tid_psam

NASA-TLX sub-__

Tim Dreßler



Fragebogen zur Stimmung und Anstrengung – sub-__

Geben Sie jetzt für jede der unten stehenden Dimensionen an, wie hoch die Beanspruchung war. Markieren Sie dazu bitte auf den folgenden Skalen, in welchem Maße Sie sich in den sechs genannten Dimensionen von der Aufgabe beansprucht oder gefordert gesehen haben:

Geistige Anforderung

Wie viel geistige Anforderung war bei der Informationsaufnahme und bei der Informationsverarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen ...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erfordert sie hohe Genauigkeit oder ist sie fehler tolerant?



Körperliche Anforderung

Wie viel körperliche Aktivität war erforderlich (z.B. ziehen, drücken, drehen, steuern, aktivieren ...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?



Leistung

Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?



Bitte wenden!

Sub-__

tid_psam

NASA-TLX sub-___

Tim Dreßler

Anstrengung

Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?



Frustration

Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?



Vielen Dank!

Appendix L Participant Information

<u>tid_psam</u>	<u>Teilnehmerinformation</u>	<u>Tim Dreßler</u>
 <p>Institut für Psychologie AG Neuropsychologie Prof. Dr. Stefan Debener A7 - 0-056 Ammerländer Heerstraße 114-118 26129 Oldenburg</p> <p>Ansprechpartner für eventuelle Rückfragen: Prof. Dr. Stefan Debener Telefon: +49 (0)441 798 4271 E-Mail: stefan.debener@uol.de</p>		

Allgemeine Teilnehmerinformation über die Untersuchung *Institut für Psychologie*

Titel der Studie: Auditive Verarbeitung während der Sprechvorbereitung

Herzlich willkommen bei unserer Studie zum Thema "Auditive Verarbeitung während der Sprechvorbereitung"! Wir danken Ihnen für Ihr Interesse an dieser Studie. In dieser Studie soll untersucht werden, wie die Vorbereitung zu Sprechen die Verarbeitung auditorischer Reize verändert. Die Teilnahme ist freiwillig und kann jederzeit ohne Angabe von Gründen widerrufen oder abgebrochen werden, ohne dass Ihnen daraus Nachteile entstehen. Die Untersuchung dient nicht der Diagnostik oder Therapie. Es handelt sich um eine wissenschaftliche Untersuchung mit Hilfe der Elektroenzephalographie (EEG).

Kurze Erläuterung der Studie

Selbstgenerierte und externe Reize werden im Gehirn unterschiedlich verarbeitet. Ziel dieser Studie ist es, die zugrunde liegenden Prozesse im Gehirn besser zu verstehen. Wir wollen erfahren, inwieweit sich die sensorische Verarbeitung von selbstgenerierten Stimuli von der sensorischen Verarbeitung externer Stimuli unterscheidet. Dazu werden in der vorliegenden Studie zuvor aufgenommene Audioaufnahmen Ihrer Stimme präsentiert, während Sie sich auf das Sprechen vorbereiten. Die Dauer der Aufgabe beläuft sich auf ca. 95 min. Mit Vor- und Nachbereitung nimmt die Studie ca. 210 Minuten in Anspruch.

Was bedeutet EEG?

Im Rahmen dieser Studie werden wir bei Ihnen ein Elektroenzephalogramm (EEG) aufzeichnen. Hierbei handelt es sich um die elektrische Aktivität des Gehirns, die an der Kopfoberfläche gemessen werden kann. Das EEG wird mit Hilfe von Elektroden, die direkt / mit Hilfe einer elastischen Kappe auf der Kopfoberfläche befestigt werden, aufgezeichnet. Die Aufzeichnung des EEGs ist beim Menschen mit keinen Risiken verknüpft. Da die Potentialfelder des Gehirns an der Kopfoberfläche sehr schwach sind, ist es erforderlich, dass jede Stelle, an der eine Elektrode angebracht wird, mit Hilfe einer speziellen Paste und Alkohol gereinigt wird. Der Kontakt zwischen Elektrode und Kopfoberfläche wird über ein Elektrodengel hergestellt. Die verwendeten Chemikalien sind klinisch getestet und lassen sich nach Abschluss des Experiments leicht auswaschen.

tid_psam	Teilnehmerinformation	Tim Dreßler
----------	-----------------------	-------------

In seltenen Fällen können trotzdem Hautirritationen auftreten. Manchmal bleiben noch für eine Weile Druckstellen an den Orten zurück, an denen die Elektroden bzw. die Elektrodenkappe befestigt wurde; in ganz seltenen Fällen sind die Stellen, an denen die Elektroden saßen, noch für ein paar Tage sichtbar (z. B. Rötungen). Bitte teilen Sie uns mit, falls Sie an bestimmten Hautallergien oder Überempfindlichkeiten der Haut leiden. Sollten Sie noch Fragen haben, wenden Sie sich damit bitte an den Versuchsleiter.

Ablauf der Studie

Vor der Untersuchung werden Sie von der Versuchsleitung ausführlich über die geplanten Messungen und Ziele schriftlich informiert. Danach werden Sie gebeten kurze Laute in ein Mikrofon einzusprechen (z.B. /da/). Dabei befinden Sie sich in einer Schallschutzkammer, in der auch später das eigentliche Experiment stattfindet. Die eingesprochenen Laute werden später im Hauptexperiment verwendet. Im Anschluss werden über Fragebögen studienrelevante Daten abgefragt. Dazu gehören z.B. allgemeine Angaben zur Person. Danach beginnen die Vorbereitungen zur EEG-Messung.

Die Vorbereitung der Messung dauert etwa 15-30 Minuten und umfasst die folgenden Schritte: Anbringen einer elastischen EEG-Haube mit eingearbeiteten Sensoren am Kopf, Reinigung der Haut unterhalb der Sensoren mit einem in Alkohol getränkten Wattestäbchen und Herstellung des Kontaktes zwischen Kopfoberfläche und Sensoren mithilfe eines speziellen hautfreundlichen Kontaktgels. Nach dem Experiment wird die EEG-Haube entfernt und Sie erhalten die Gelegenheit sich die Haare zu waschen. Durch die Vorbereitung der EEG-Sensoren (Kontaktgel und Wattestäbchen) kann es in seltenen Fällen zu leichten lokalen Rötungen der Haut kommen. Diese klingen normalerweise innerhalb weniger Stunden wieder ab. Häufig kann die Verwendung einer Hautcreme die Rückbildung von Hautreizungen noch beschleunigen.

Nach Abschluss der Vorbereitungen beginnt das eigentliche Experiment. Hierfür werden Sie wieder in der Schallschutzkammer Platz nehmen. Ihre Aufgabe ist es einen Bildschirm zu beobachten auf dem Ihnen ein Bild präsentiert wird. Über die Zeit verändert sich das Bild, was sie darauf hinweist, einen Laut zu produzieren. In manchen Durchläufen wird in der Vorbereitungsphase (also vor der Produktion des Lautes) eine der zuvor aufgenommenen Audioproben über Lautsprecher präsentiert. Die Dauer der Aufgabe beläuft sich auf ca. 95 Minuten und umfasst 480 Durchläufe, in denen Sie einen Laut produzieren sollen, sowie weitere 480 Durchläufe, in denen Sie passiv bleiben sollen. In regelmäßigen Abständen finden kurze Erholungsausen statt, in denen Sie z.B. etwas trinken können.

Nach Beendigung der EEG-Aufnahme wird die EEG-Kappe entfernt und Sie haben Gelegenheit, sich die Haare zu waschen. Die Nachbereitung (Aufklärung über Studienziele, Haare waschen) beträgt ca. 20 Minuten.

Insgesamt beläuft sich der Zeitaufwand für die Durchführung des Experiments inklusive Vor- und Nachbereitung auf ungefähr 210 Minuten. Sie haben zu allen Zeitpunkten die Möglichkeit, Fragen zu stellen. Sollten Sie noch Fragen zum Ablauf haben, wenden Sie sich damit bitte an den Versuchsleiter.

Zusätzliche Informationen

Es werden generell keine schmerzhaften Reize verwendet. Sie werden vor jedem Experiment grundsätzlich aufgeklärt und können das Experiment jederzeit ohne Nennung von Gründen beenden. Ein*e Versuchsleiter*in wird Sie jederzeit während der EEG-Aufnahme beobachten. Sollten Sie noch Fragen haben, wenden Sie sich damit bitte an die Versuchsleitung. Während einer laufenden Messung können Sie jederzeit mit der Versuchsleitung Kontakt aufnehmen.

Freiwilligkeit und Anonymität

Die Teilnahme an der Studie ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen die Teilnahme an dieser Studie beenden, ohne dass Ihnen daraus Nachteile entstehen. Auch wenn Sie die Studie vorzeitig abbrechen, haben Sie Anspruch auf eine entsprechende Vergütung / entsprechende Versuchspersonenstunden für den bis dahin erbrachten Zeitaufwand.

Die im Rahmen dieser Studie erhobenen, oben beschriebenen Daten und persönlichen Mitteilungen werden vertraulich behandelt. So unterliegen diejenigen Projektmitarbeiter*innen, die durch direkten Kontakt mit Ihnen über personenbezogene Daten verfügen, der Schweigepflicht. Des Weiteren wird die Veröffentlichung der Ergebnisse der Studie in anonymisierter Form erfolgen, d. h. ohne dass Ihre Daten Ihrer Person zugeordnet werden können.

Allgemeine Hinweise

Es ist wichtig, dass Sie ausgeschlafen zum vereinbarten Termin kommen. Bitte trinken Sie am Tag vor der Untersuchung keinen Alkohol. Falls Sie regelmäßig Medikamente einnehmen, bringen Sie den Beipackzettel mit.

Wieviel Zeit erfordert die Teilnahme und wo findet die Studie statt?

Die Gesamtdauer der Untersuchung beträgt ca. 210 Minuten (siehe unter Ablauf der Studie) und findet im Labor der Abteilung für Neuropsychologie in Oldenburg statt (Campus Haarentor, Gebäude A7). Die genauen Termine werden mit der Versuchsleitung nach Ihren Wünschen abgesprochen.

Auffällige Befunde

Die Untersuchung dient ausschließlich Forschungszwecken. Eine medizinische oder psychologische Beurteilung Ihrer Daten erfolgt nicht. Es könnte uns jedoch ein ungewöhnliches Untersuchungsergebnis auffallen. In diesem Fall werden wir Sie darüber informieren und Ihnen empfehlen, dieses Ergebnis bei Ihrem Hausarzt diagnostisch weiter abklären zu lassen. Nur wenn Sie damit einverstanden sind, dass wir Sie ggf. über einen auffälligen Befund informieren, können Sie an dieser Studie teilnehmen. Sofern bei dieser diagnostischen Abklärung eine Erkrankung festgestellt werden sollte, könnten Ihnen daraus unter Umständen Nachteile entstehen, z. B. der Abschluss einer privaten Krankenversicherung oder einer Lebensversicherung erschwert werden.

Vergütung

Für die Teilnahme an der Untersuchung erhalten Sie eine Vergütung in Höhe von 12 € pro Stunde. Die Vergütung wird Ihnen bargeldlos per Überweisung ausgezahlt. Bei der Überweisung der Vergütung müssen Sie Ihre Kontoverbindung angeben. Alle diesbezüglichen Informationen werden separat von den Untersuchungsdaten aufbewahrt.

Kategorien personenbezogener Daten, die verarbeitet werden

Von der Datenverarbeitung sind folgende personenbezogene Daten umfasst:

- **Allgemeine Kategorien personenbezogener Daten**
 - o Kontaktdaten (Name, Anschrift, E-Mail-Adresse, Telefonnummern)
 - o Demografische Daten (Alter, Geschlecht, Tätigkeit)
 - o Bankverbindung (Name, IBAN, Kreditinstitut, BIC)
 - o Abrechnungsdaten (Name, Anschrift, das Datum, gezahlter Betrag, Zweck der Zahlung, ggf. Bankverbindung)
- **Besondere Kategorien personenbezogener Daten**
 - o Gesundheitsdaten (Krankheitsgeschichte, Schlafgewohnheiten, Alkohol- und Drogenkonsum, Medikamenteneinnahme)
 - o Tonaufnahmen der Stimme

Verfahren der Datenverarbeitung

Die Erhebung und Verarbeitung Ihrer oben beschriebenen persönlichen Daten erfolgt pseudonymisiert im Institut für Psychologie unter Verwendung einer Nummer und ohne Angabe Ihres Namens. Es existiert eine Kodierliste auf Papier, die Ihren Namen mit der Nummer verbindet. Die Kodierliste ist nur den Versuchsleitern und dem Projektleiter zugänglich; das heißt, nur diese Personen können die erhobenen Daten mit Ihrem Namen in Verbindung bringen. Die Kodierliste wird in einem abschließbaren Schrank aufbewahrt und nach Abschluss der Datenerhebung, spätestens aber am 31.12.2025 vernichtet. Ihre Daten sind dann anonymisiert. Damit ist es niemandem mehr möglich, die erhobenen Daten mit Ihrem Namen in Verbindung zu bringen. Die anonymisierten Daten werden mindestens 10 Jahre gespeichert (gute wissenschaftliche Praxis). Solange die Kodierliste existiert, können Sie die Löschung aller von Ihnen erhobenen Daten verlangen. Ist die Kodierliste aber erst einmal gelöscht, können wir Ihren Datensatz nicht mehr identifizieren. Deshalb können wir Ihrem Verlangen nach Löschung Ihrer Daten nur solange nachkommen, wie die Kodierliste existiert. Dementsprechend können wir Ihnen unten genannten Rechten (1-7) nur bis zur Löschung der Kodierliste entsprechen.

Ihre Abrechnungsdaten bleiben bis zu ihrer Löschung nach zehn Jahren nur noch für das Buchhaltungs-/Abrechnungssystem und dessen Mitarbeiter sichtbar (10-Jährige Aufbewahrungsfrist des Dezernats 2 gem. § 147 Abgabenordnung).

tid_psam	Teilnehmerinformation	Tim Dreßler
Dauer der Verarbeitung		
Nach Aufzeichnung aller Daten der Studie werden Ihre Daten schnellstmöglich –		
insbesondere bevor eine Veröffentlichung zu wissenschaftlichen Zwecken (z.B. Fachartikel, Ta-		
gungsbeiträge, wissenschaftliche Datenbanken [Open Data Repositories]) stattfindet – ano-		
nymisiert. Hierzu ist der Verantwortliche nach § 13 Absatz 2 Satz 1 Niedersächsisches Da-		
tenschutzgesetz (NDSG) verpflichtet. Anonymisierung bedeutet, dass niemand mehr Ihre		
Daten Ihrer Person zuordnen kann (siehe oben). Ihre Daten sind dann nicht mehr „personen-		
bezogen“ im Sinne der datenschutzrechtlichen Rechtsvorschriften.		
Verwendung der Daten		
Diese Studie dient ausschließlich Forschungszwecken. Die Sie betreffenden personenbezo-		
genen Daten werden ohne Ihre Einwilligung nicht an Dritte weitergegeben.		
Aufbewahrungsfrist und Nachnutzung Ihrer anonymisierten Daten		
Zur Sicherung guter wissenschaftlicher Praxis werden Daten wissenschaftlicher Studien oft		
in anonymisierter Form in Forschungs-Datenbanken (z.B. https://openneuro.org/) abgelegt.		
Dies ermöglicht es anderen Forschenden die Auswertung nachzuvollziehen oder eine alter-		
native Auswertung zu testen. Die Daten können auch für neue Fragestellungen genutzt wer-		
den, die über das Ziel dieser Studie hinausgehen. Bitte beachten Sie jedoch, dass jedes Gehirn		
einzigartig ist und dadurch eine Identifizierung von Personen nicht vollständig ausgeschlos-		
sen werden kann. Durch die Bereitstellung in Forschungsdatenbanken ist es auch anderen		
Wissenschaftlerinnen und Wissenschaftlern im In- und Ausland – auch in Ländern außerhalb		
Europas mit einem niedrigeren Datenschutzniveau möglich diese Forschungsdaten zu erhal-		
ten. Sobald die Daten in diesen Forschungsdatenbanken abgelegt sind, kann die Nutzung		
dieser Daten nicht mehr kontrolliert, nachträglich eingeschränkt oder überall gelöscht wer-		
den. Sie können einer solchen Datenweitergabe in der Einwilligungserklärung gesondert zu-		
stimmen oder widersprechen.		
Kontaktdaten der Verantwortlichen und des Datenschutzbeauftragten		
Verantwortliche		Datenschutzbeauftragter
Carl von Ossietzky Universität Oldenburg		Carl von Ossietzky Universität Oldenburg
(KdöR),		Der Datenschutzbeauftragte
gesetzlich vertreten durch den Präsidenten		Ammerländer Heerstr. 114-118
Ammerländer Heerstr. 114-118		26129 Oldenburg
26129 Oldenburg		Tel.: 0441-798-4196
Telefon: +49 441 798-0		E-Mail: dsuni@uol.de
Telefax: +49 441 798-3000		Internet: https://uol.de/datenschutz/
E-Mail: internet@uol.de		
Internet: https://uol.de		

Ansprechpartner

Wir würden uns sehr freuen, wenn Sie bereit wären an unserem Forschungsprojekt teilzunehmen. Falls Sie weitere Rückfragen haben, wenden Sie sich bitte an Prof. Dr. Stefan Debe
ner (stefan.debener@uol.de, Tel: (0441) 798-4271).

Rechtsgrundlage

Die Rechtsgrundlage für die Erhebung Sie betreffenden personenbezogener Daten ist:

Rechte als Betroffener

- Sie haben ein **Recht auf Auskunft** über die Sie betreffenden personenbezogenen Daten (Art. 15 DSGVO).
- Sie können unverzüglich von dem Verantwortlichen **Berichtigung** Sie betreffender unrichtiger oder **Vervollständigung** unvollständiger personenbezogener Daten verlangen (Art. 16 DSGVO).
- Sie sind hiermit darüber informiert worden, dass Sie jederzeit eine **Lösung** der Sie betreffenden personenbezogenen Daten verlangen können (Art. 17 DSGVO).
- Sie können die **Einschränkung der Verarbeitung** verlangen, soweit die gesetzlichen Voraussetzungen vorliegen (Art. 18 DSGVO).
- Sie haben das Recht, die Sie betreffenden personenbezogenen Daten, **in einem strukturierten, gängigen und maschinenlesbaren Format zu erhalten** und diese Daten einem anderen Verantwortlichen zu übermitteln (Art. 20 DSGVO).
- Sie können jederzeit gegen die Verarbeitung Sie betreffender personenbezogener Daten **Widerspruch einlegen**, die aufgrund von Artikel 6 Abs. 1 lit. e oder f DSGVO erfolgt (Art. 21 DSGVO).
- Sie können die erteilte **Einwilligung jederzeit mit Wirkung für die Zukunft widerrufen**, ohne, dass die Rechtmäßigkeit der aufgrund der Einwilligung bis zum Widerruf erfolgten Verarbeitung berührt wird (Art. 7 Abs. 3 DSGVO) *Sofern Rechtsgrundlage auf Einwilligung basiert.*

tid_psam

Teilnehmerinformation

Tim Dreßler

Bereitstellung der Daten und Folgen der Nichtbereitstellung

Die Bereitstellung der Sie betreffenden personenbezogenen Daten ist weder vertraglich noch gesetzlich vorgeschrieben. Sie sind nicht dazu verpflichtet, Sie betreffende personenbezogene Daten bereitzustellen. Die Nichtbereitstellung hätte zur Folge, dass Sie nicht an der Studie teilnehmen können.

Beschwerderecht bei einer Aufsichtsbehörde

Falls Sie der Ansicht sind, dass die Verarbeitung Ihrer personenbezogenen Daten gegen Datenschutzvorschriften verstößt, wenden Sie sich bitte an die/den Datenschutzbeauftragte/n der Verantwortlichen (s.o.). Unabhängig hiervon haben Sie ein Recht auf **Beschwerde** bei der zuständigen Aufsichtsbehörde. Die zuständige Aufsichtsbehörde ist:

Die Landesbeauftragte für den Datenschutz Niedersachsen

Prinzenstraße 5

30159 Hannover

Telefon: 0511 120-4500

Telefax: 0511 120-4599

Email: poststelle@lfd.niedersachsen.de

Appendix M Consent Form

<u>tid_psam</u>	<u>Einwilligungserklärung</u>	<u>Tim Dreßler</u>
-----------------	-------------------------------	--------------------



*Institut für Psychologie
AG Neuropsychologie
Prof. Dr. Stefan Debener*

A7-0-056
Ammerländer Heerstraße 114-118
26129 Oldenburg

Ansprechpartner für eventuelle Rückfragen:
*Prof. Dr. Stefan Debener
Telefon: +49 (0)441 798 4271
E-Mail: stefan.debener@uol.de*

Einwilligungserklärung

Institut für Psychologie

Titel der Studie: Auditive Verarbeitung während der Sprechvorbereitung

Ich (Name des Teilnehmers /der Teilnehmerin in Blockschrift)

bin mündlich und schriftlich über die Studie und den Versuchsablauf aufgeklärt worden. Ich willige ein, an der hier beschriebenen EEG-Studie, die die Eigenschaften auditiver Verarbeitung während der Sprechvorbereitung untersucht, teilzunehmen.

Sofern ich Fragen zu dieser vorgesehenen Studie hatte, wurden sie von Herrn/Frau _____ vollständig und zu meiner Zufriedenheit beantwortet.

Mit der beschriebenen Erhebung und Verarbeitung der EEG Daten sowie der Audiaufzeichnungen bin ich einverstanden. Die Aufzeichnung und Auswertung dieser Daten erfolgt pseudonymisiert im Institut für Psychologie, unter Verwendung einer Nummer und ohne Angabe meines Namens. Es existiert eine Kodierliste auf Papier, die meinen Namen mit dieser Nummer verbindet. Diese Kodierliste ist nur den Versuchsleitern und dem Projektleiter zugänglich, das heißt, nur diese Personen können die erhobenen Daten mit meinem Namen in Verbindung bringen. Nach Abschluss der Datenerhebung, spätestens am 31.12.2025, wird die Kodierliste gelöscht. Meine Daten sind dann anonymisiert. Damit ist es niemandem mehr möglich, die erhobenen Daten mit meinem Namen in Verbindung zu bringen. Mir ist bekannt, dass ich mein Einverständnis zur Aufbewahrung bzw. Speicherung dieser Daten widerrufen kann, ohne dass mir daraus Nachteile entstehen. Ich bin darüber informiert worden, dass ich jederzeit eine Löschung all meiner Daten verlangen kann. Wenn allerdings die Kodierliste bereits gelöscht ist, kann mein Datensatz nicht mehr identifiziert und also auch nicht mehr gelöscht werden. Meine Daten sind dann anonymisiert. Ich bin einverstanden, dass meine anonymisierten Daten zu Forschungszwecken weiterverwendet werden können und mindestens 10 Jahre gespeichert bleiben.

Sollten behandlungsbedürftige Auffälligkeiten im EEG erkannt werden, bin ich damit einverstanden, dass mir diese mitgeteilt werden, so dass ich diese ggf. weiter abklären lassen kann. Ich wurde darüber informiert, dass die Information über auffällige Befunde u.U. mit versicherungsrechtlichen Konsequenzen verbunden sein kann.

tid_psam

Einwilligungserklärung

Tim Dreßler

ZUSATZ EINWILLIGUNGSERKLÄRUNG für Bild- und Tonaufnahmen

Ich bin darüber informiert, dass Tonaufnahmen gemacht werden. Die Aufzeichnung und Auswertung der Tonaufnahme erfolgen pseudonymisiert, d. h. unter Verwendung einer Nummer und ohne Angabe meines Namens und dass eine Kodierliste auf Papier existiert, die meinen Namen mit der Nummer verbindet. Die Kodierliste ist nur dem Versuchsleiter zugänglich und wird nach Abschluss der Datenerhebung gelöscht, spätestens aber am 31.12.2025. Es besteht die sehr geringe Wahrscheinlichkeit, dass eine an der Datenauswertung beteiligte Person mich erkennt. Aus diesem Grund unterliegen alle an der Auswertung beteiligten Personen einer absoluten Schweigepflicht und dürfen unter keinen Umständen vertrauliche Informationen an Dritte weitergeben. Die Audioaufnahmen werden zur evtl. Nachnutzung auf Universitätsservern aufbewahrt. Alle Aufnahmen werden aber spätestens am 31.12.2035 gelöscht.

Mir ist bekannt, dass ich mein Einverständnis zur Aufbewahrung bzw. Speicherung dieser Daten widerrufen kann, ohne dass mir daraus Nachteile entstehen. Ich bin darüber informiert worden, dass ich jederzeit eine Löschung meiner Aufnahmen verlangen kann, solange die Kodierliste existiert. Die Aufnahmen werden aber in jedem Fall nach Abschluss der Auswertung vernichtet.

Mit der beschriebenen Handhabung der erhobenen Aufnahmen bin ich einverstanden.

Die Einverständniserklärung für die Tonaufnahme ist freiwillig. Ich kann diese Erklärung jederzeit widerrufen. Im Falle einer Ablehnung oder eines Rücktritts entstehen für mich keinerlei Kosten oder anderweitige Nachteile; eine Teilnahme an der Studie ist dann allerdings nicht möglich.

tid_psam

Einwilligungserklärung

Tim Dreßler

Ich hatte genügend Zeit für eine Entscheidung und bin bereit, an der o.g. Studie teilzunehmen. Ich weiß, dass die Teilnahme an der Studie freiwillig ist und ich die Teilnahme jederzeit ohne Angaben von Gründen beenden kann. Ich weiß, dass ich in diesem Fall Anspruch auf eine Vergütung für die bis dahin erbrachten Stunden habe.

Eine Ausfertigung der Teilnehmerinformation über die Untersuchung und eine Ausfertigung der Einwilligungserklärung habe ich erhalten. Die Teilnehmerinformation ist Teil dieser Einwilligungserklärung.

Ich willige ein, dass im Rahmen der Studie EEG-Aufzeichnungen sowie Verhaltensdaten (inkl. Tonaufnahmen) von mir erhoben und wissenschaftlich ausgewertet werden.

ja nein (bitte Zutreffendes ankreuzen)

Ich willige ein, dass die EEG-Aufzeichnungen und Verhaltensdaten in Forschungsdatenbanken (ohne Nennung meines Namens) veröffentlicht werden. Aufgrund der Einzigartigkeit des Gehirns kann eine Identifizierung meiner Person niemals vollständig ausgeschlossen werden.

ja nein (bitte Zutreffendes ankreuzen)

Ort, Datum & Unterschrift des Teilnehmers: _____ Name des Teilnehmers in Druckschrift:

Ort, Datum & Unterschrift des Versuchsleiters: _____ Name des Versuchsleiters in Druckschrift:

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.

07.08.2025, Oldenburg

Datum, Ort



Unterschrift