UNIVERSITAT DE
BARCELONA

Auditory processing around actions: Evidence from psychophysics, electroencephalography, and pupillometry

Nadia Paraskevoudi

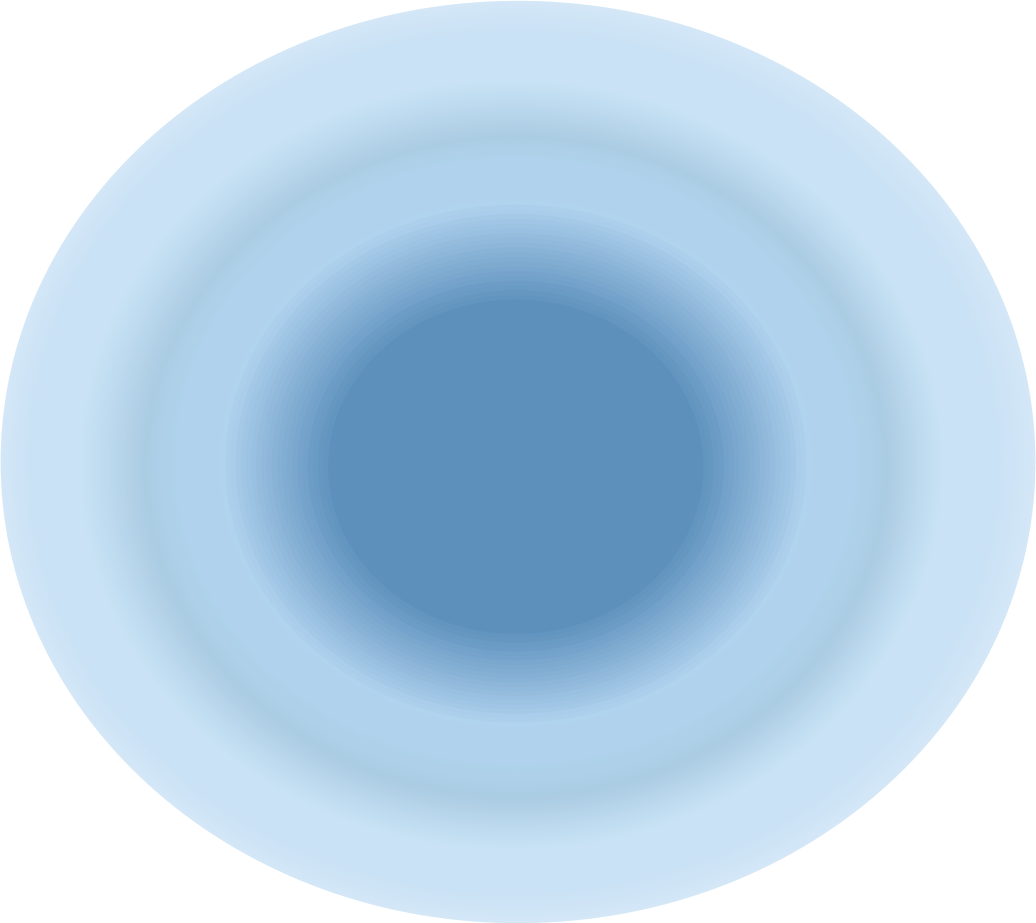
ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Auditory processing around actions:
Evidence from psychophysics, electroencephalography, and pupillometry.

Nadia Paraskevoudi



Doctoral programme in Biomedicine

UNIVERSITAT DE
BARCELONA

Department of Clinical Psychology and Psychobiology
Faculty of Psychology
University of Barcelona

Department of Clinical Psychology and Psychobiology
Faculty of Psychology
University of Barcelona

Auditory processing around actions: Evidence from psychophysics, electroencephalography, and pupillometry.

Thesis presented by

Nadia Paraskevoudi

to obtain the degree of doctor from the Programme of Biomedicine at the University of Barcelona

Supervised by
Dr. Iria San Miguel Insua
Dr. Concepción Clemente Lapena

To those I have missed most during these four years

ACKNOWLEDGMENTS

The present thesis would not have started (nor ended) without the contribution of many people. Evidently, I would need many more pages to thank you all properly, but for the sake of space (and your time) I will try to limit myself. First of all, thanks to all these students that with admirable patience participated in the studies presented in this thesis.

Mis supervisoras. Iria. Sin ti, esta tesis ni siquiera habría empezado. Gracias por darme la oportunidad de hacer este doctorado pandémico a tu lado. Gracias por haberme dado tanta libertad y tanta confianza, por creer tanto en mi desde el primer mes que estuve en el lab. Imma, gracias de verdad por todo el cariño, todo el apoyo constante que me has dado estos años. Obviamente, esta tesis no hubiese acabado sin el apoyo de mi laboratorio. Carles, gracias por creer en mi desde el momento zero, por aguantar todas las presentaciones interminables, por estar siempre dispuesto a ayudar y escucharme. Marc, Judith, Paco, Jose, gracias por todo el humor absurdo que me habéis regalado, por todas las charlas random en el pasillo del departamento que hacían las mañanas en Mundet mucho más interesantes. Marta Turró, de verdad no sé qué hubiese hecho sin tu ayuda durante los primeros meses en Barcelona. Gracias por toda la ayuda con burocracia, por estar siempre dispuesta a ayudarme y siempre con una gran sonrisa. Jordi! Estos años hubiesen sido mucho menos interesantes sin ti. Gracias por contagiarme tu curiosidad y por todo el apoyo dentro y fuera del laboratorio, y por creer tanto en mi. Gracias por todos los abrazos, todas las fiestas, y todas las risas histéricas que hemos compartido. Y gracias por traer el Prince en mi vida!

Mis colegas – primero, gracias al grupo que se formó random en el comedor del departamento. Un grupo que empezó con solo brainlab y se perfeccionó cuando nos dimos cuenta que hay algo interesante también por el otro lado del pasillo. Todxs vosotrxs habéis hecho la vida en Barcelona mucho más fácil de lo que pensaba cuando me hacía las maletas para mudarme. Natàlia, gracias por estar a mi lado, por acompañarme hasta a médicos, por ser de las primeras personas que me hicieron sentir parte de esta familia. Sonia, entre dibujos de peces y palabras cortadas a lo andaluz, me has dado mucha risa durante estos años. Gracias por ser tan transparente, por estar dispuesta a ayudar en cualquier momento. Gracias por to! Giannina, has sido un placer compartir despacho contigo. Gracias por todas la birras en la enfadada, por todos los abrazos que me dabas cuando me veías estresada, por estar siempre sonriendo. Fran! Han pasado ya tres años sin verte, pero para mi siempre formarás parte de esta etapa. Gracias por todos los momentos bonitos en Barcelona y Nueva York. Montse, eres de las personas más sweet que he conocido en este doctorado. Gracias por trasmitirme tu tranquilidad, por estar siempre allí a escuchar todos los pequeños dramas que ocurieron durante estos años. Maria. Aún no sé cómo agradecerte por todo el cariño que me has regalado estos años. Gracias por cuidarme tanto, por traerme hasta comida al dept, por llevarme a tu casa a conocer a tu familia. Te admiro por muchas razones, sobre todo por no dejar de sonreir incluso en momentos duros. Joni, my friend in black. Has sido una gran fuente de apoyo dentro y fuera del departamento. Tantas historias y eso en menos de un año que coincidimos en Barcelona. Gracias por el paseo en Sabadell, las listas de música, los vermuts, la Nueva York, la Mágia Roja y mucho mucho más. Más que nada, gracias for sticking around aunque estamos en diferentes puntas del mundo. Vitt, amori, las horas en el despacho no habrían pasado tan fácil si no fuera por ti. Gracias por todo el cariño, todo el apoyo durante estos años, gracias por cuidarme como si fuera tu hermana pequeña. Marta. Una de las fuentes de surealismo puro durante estos años: el viaje catastrófico en Marsella, el bird-watching de palomas intentando identificar si tienen viruela o no. Has sido un gran apoyo. Gracias por todas estas historias y por todo el cariño que me has dado. Hablando de surealismo. Tere. Miles de historias incluso en otros continentes. Gracias por bailar conejo y callejear conmigo hasta la madrugada, por estar a mi lado en todo lo bueno y todo lo malo. Te lo he dicho varias veces y te lo vuelvo a decir aquí: Te admiro por ser tan transparente, fuerte y sensible a la vez, por poder querer tanto a tus amigxs y hacer todo lo que puedas para ayudarles. Mil gracias por todo. Gracias a tus padres, a Maria y a Toni, por hacerme sentir que tengo una familia en Ibiza.

Voy a las más nuevas del lab, las que tuvieron la gran suerte de empezar un doctorado en el pico de la pandemia. Marta Puertollano, Sam, Martina, Raquel. Gracias amoris por todos los momentos que compartimos en aulas virtuales o no. Special thanks to Martina y Raquel: Ha sido un placer ver todo vuesto camino desde el grado hasta el doctorado. He tenido mucha suerte poder trabajar con vosotras. Martina, gracias por preguntarme siempre de los pequeños detalles de un análisis o de MATLAB, he aprendido mucho de ti. Raquel, tú y Camila fuisteis una ayuda enorme durante los primeros años, y sabes que el primer estudio de esta tesis no hubiese acabado sin tu ayuda. Gracias de verdad. Stef, Schatzi! You´ve been one of the greatest additions in this life of the PhD. I cannot thank you enough for all the moments we´ve shared as colleagues and friends, for all these times that you were sharing with me the feeling of desperation when it came to more and more analyses. Thanks, pouláki mou, for being so kind, so unique, so sarcastic.

Laura, nada sería lo mismo si no fuese por ti. Gracias por darme tanto cariño desde el principio cuando aún apenas me conocías bien, por quererme tanto, por estar a mi lado, por compartir conmigo literalmente cada momento de estos cuatro años: en el piso, en las calles, en los bares. En Nueva York, Nápoles, y Atenas. Gracias por ser la mejor compañera en conocer espontánexs, por animarme a empezar conversaciones random hasta con las moscas. Gracias Lauraki por ser como eres. Aitor, agápi mu, tú en Urano y yo en Hades, o al revés. Ya sabes de lo que te estoy hablando. Has sido de las mejores personas que he conocido en esta tierra. Gracias por todo el cariño y todas las risas histéricas. Alberto, my only Prince, otro fan del baile de conejo (eso lo compartes con Tere!). Gracias por toda la absurdidad, por acompañarme hasta a islas en la mitad de la nada, por todas las noches de cine (y elevated terror!) que apenas podíamos acabar la peli con los mil comentarios que hacías. En fin, vosotrxs tres habéis ido by far una de las cosas más interesantes que me regaló esta ciudad. Gracias!

Hablando de cosas interesantes. Λος supervivientes: Alex, Fotis, Panos, Makis, Manos, Mitsos, Marcella, Anna, Sofi y more recently Laura y Stefanos. Qué hubiese hecho sin vosotrxs. Obviamente necesitaría muchas más páginas para agradeceros todxs. Gracias por las miles de historias, el callejeo interminable, los sleep-overs durante la pandemia. Gracias por animarme a seguir la noche cuando bajaban las persianas y gracias por los miles de millones de abrazos. Gracias por ser como sois.

Jon, my partner in crime. Tak for all the unconditional support and love, for all these moments you were there while I was trying to cope with everything. Thanks for being so genuinely weird, for giggling together when the most absurd events were unfolding in front of our eyes. Thanks for standing by my side when the Danish railway system was collapsing and Greek ferries were crashing in the middle of the Mediterranean, thanks for causing me laugh attacks during the most stressful moments of the last two years.

Obviously, none of this would have started if it wasn´t for all these people supporting me in my way out of Athens. First, the three professors that stood by my side as supervisors and friends. John Pezaris, thanks for believing in me so much, for giving me the opportunity to learn from you and for trusting me with projects that were new and challenging to me. Irini Skaliora and Argiro Vatakis, you are the sweetest, strongest, and most multidimensional representations of science in my life. Irini, you´ve been one of the most interesting professors I had. Thanks for helping me to decide to leave, for all the support during and after the master´s. Argiro. You´ve been more than a supervisor to me. Thanks for supporting me when I was planning to stay but also when I decided to go. Thanks for believing in me so much, for transmitting me your assertiveness and enthusiasm, for inspiring me to continue with a PhD despite all the difficulties that this decision entailed.

Thanks to these two groups of people that have been sticking around for years. First, my partners in suffering, the ones that coincided with me at our very first and scary encounter with neuroscience, but also at the difficult (and sometimes painful) decision of continuing with a PhD: Penny, Kotsari, Lydia, Thanopoule, Koko! Thanks for all the days and nights that we kept talking for hours about everything and nothing, for all the collective moments of self-doubt, sarcasm, and hopelessness that we shared, while devouring cheap whiskeys and disgusting pizzas. Finally, the few heroes that stuck around for the past 20 years! Alexia, Dafni, Olympia, Kats. Thanks αγάπες μου for all the absurd memories and giggles we've shared since 2002, for all the collective Heimweh (and sometimes Fernweh!).

●

Finally, the people I missed so much during this PhD, those supporting me in their own special way daily, constantly, and unconditionally for the biggest part of my life. This PhD is dedicated to you.

Ilektra and Kostis. Nothing said here can even remotely describe how thankful I am and how lucky I feel for having grown up with you. Ilektra, thanks for the more than 7k days that you have been loving and supporting me, and especially for the past four years that you were always there to give me a hug, a glass of wine, and a cigarette to scare my stress away. Kostis, my brother, my closest friend, my biggest love. Thanks for all your trust, love, constant support, thanks for never doubting in me. Thank you both for being a ceaseless constant no matter what.

My parents, my points of reference and my biggest source of support. There are not enough words to say how thankful I am for all the unconditional support you have been giving me. Thank you for supporting me to follow this path even when I was overwhelmed and scared of leaving Athens behind. Thanks for teaching me to laugh away any unfortunate event and for inspiring me try things out, instead of regretting not doing them. Thank you both for all the strength that you've been giving me, all the stress-releasing laughs and wines, thanks for being my parents and my friends at the same time.

Timo. Nothing would have been possible without you. Thanks for growing up with me, for being a source of inspiration and creativity, for supporting me in trying things out even when this implied me being thousands of kilometers away from you. Thanks for always reminding me of the important things after all.

TABLE OF CONTENTS

ACKNOWLEDGMENTS 3
ABSTRACT.. 7
RESUM... 9
FOREWORD 11
LIST OF ORIGINAL STUDIES... 12
ABBREVIATIONS 13
CHAPTER 1: Introduction.. 14
1.1. Immediate effects of actions on sensory processing 15
    1.1.1. Action effects on perception 15
    1.1.2. Action effects on neurophysiological responses 17
    1.1.3. Models and theories of action effects 18
    1.1.4. Specific predictions or unspecific modulations? 23
    1.1.5. Subcortical neuromodulation 26
1.2. Action effects on memory processing 28
    1.2.1. Production effect 28
    1.2.2. Predictive models of memory.. 28
CHAPTER 2: Objectives and hypotheses 30
CHAPTER 3: General methodology 32
3.1. Participants... 32
3.2. Experimental design and stimuli 32
3.3. Apparatus 32
3.4. Data analysis... 33
    3.4.1. Behavioural measures 33
    3.4.2. EEG preprocessing and ERP analysis 33
    3.4.3. Pupil preprocessing.. 34
    3.4.4. Cluster-based permutation tests 35
CHAPTER 4: STUDIES 36
STUDY I... 37
STUDY II.. 63
STUDY III... 93
CHAPTER 5: General discussion 136
5.1. Modulatory effects of self-generation on perception and memory 136
5.2. Modulatory effects of actions on auditory responses.. 141
5.3. Modulatory effects of actions on subcortical neuromodulation 144
5.4. Relationships between the action effects on memory, auditory responses, and subcortical neuromodulation 145
5.5. Strengths and limitations 148
5.6. Future directions 149
CHAPTER 6: Conclusions.. 150
CHAPTER 7: References 151

ABSTRACT

Biological organisms are constantly bombarded with a flux of sensory information that needs to be adequately processed for optimal interaction with the environment. Yet, what we perceive is not just a mere reproduction of the signals reaching our sensory organs. Instead, we interpret the external world through an interaction between the self and our environment. A specific instance of this interaction is how we perceive, process, and memorize the sensory outcomes of our own actions. Although substantial work has been done in this domain, mainly showing attenuated perception and sensory processing for self-generated information, several issues remain unknown, often leading to heated debates as for the mechanisms underlying the differential processing of self-generated stimuli. At the core of this debate stand questions related to the direction of the action effects on behaviour (i.e., suppression or enhancement of perception and memory) and sensory processing (i.e., cancellation or sharpening of sensory responses), the nature of the effects (i.e., stimulus-specific motor-predictions or unspecific mechanisms possibly driven by neuromodulatory processes), but also the influence of other factors that are often confounded with self-generation (e.g., stimulus intensity and predictability).

The present thesis attempts to elucidate the mechanisms underlying the effects of actions on auditory processing. In three original studies, we examined the self-generation effects from the angles of behavioural responses, namely perceptual processing and memory encoding, and basic physiology, namely electrophysiological responses and neuromodulatory processes (i.e., measured with pupillometry). Study I tested for possible interactions between actions and stimulus intensity on sound detection and loudness discrimination, while examining which aspect of perception is modulated (sensitivity or bias). Study II addressed the effects of cued actions on sensory processing and memory encoding of concurrent, but unpredictable, sounds, while exploring for the first time the involvement of neuromodulatory systems in the action-induced modulations of auditory responses. Study III disentangled self-generation and predictability – two factors that have been conflated in previous work – and assessed how they interact in shaping auditory responses, subcortical neuromodulation, and memory encoding of sounds. Finally, Study II and Study III also tested for possible links between memory encoding, sensory attenuation, and subcortical neuromodulation.

Related to the behavioural findings, the present thesis tapped onto two main research lines, namely low-level perceptual processing and memory encoding. Study I showed that actions interact with stimulus intensity on perceptual bias, but not sensitivity measures. Study II and Study III showed that actions modulate memory performance, but the direction of the effects depends on predictability confounds and the type of action. Specifically, when actions are cued and afford temporal predictability, memory performance drops and relates to the magnitude of the attenuation effects for self-generated sounds (Study II). In contrast, in the absence of predictability confounds and when action alternatives are provided, the mere presence of an action enhances memory performance of sounds, but this enhancement does not relate to the suppression effects (Study III).

Related to the effects of actions on sensory processing, the present thesis assessed the specificity of the action-induced attenuation effects and clarified the role of neuromodulatory processes in sensory suppression. Study II showed that cued actions attenuated sensory responses and increased pupil diameter despite the absence of a predictive action-sound relationship, but sensory suppression and subcortical neuromodulation were not related. Study III employed a predictable and an unpredictable session where both self- and externally-generated sounds were equally predictable in identity, timing, and probability of occurrence and were

presented shortly after or farther away from the action or cue, respectively. With short delays, predictability mattered since attenuation was only obtained in the predictable session. With longer delays, self-generation and predictability no longer interacted: N1-attenuation was driven by self-generation only, while Tb- and P2-attenuation were driven by predictability only. Crucially, pupil diameter increased during actions, interacted with predictability, and was linked to the Tb-attenuation when the effects were strongest (i.e., with fully contingent action-sound relationship and short action-sound delays).

Taken together, the present thesis disentangled the contribution of factors other than self-generation in modulating perception, memory, and neurophysiological responses for self-generated inputs (i.e., sound intensity, predictability, and action-sound delay), and showed that actions trigger a cascade of stimulus-specific and unspecific processes – presumably driven by subcortical neuromodulatory processes – that collaboratively orchestrate auditory processing and memory encoding.

RESUM

El que percebem no és una simple reproducció dels senyals que arriben als nostres òrgans sensorials, sinó que interpretem el món a través de la nostra interacció amb l'entorn. Un exemple d'aquesta interacció és el com percebem, processem i memoritzem les conseqüències sensorials de les nostres accions. Tot i que treballs previs han mostrat que el processament sensorial esta atenuat per a la informació autogenerada en comparació a la informació generada externament, hi ha diverses qüestions respecte al processament diferencial d'estímuls autogenerats que continuen sent debatudes. En el nucli d'aquest debat es troben les qüestions relacionades amb la direcció dels efectes de l'acció en el comportament (empobriment o millora de la percepció i la memòria), el processament sensorial (cancel·lació o augment de les respostes sensorials), la naturalesa dels efectes (prediccions motores específiques d'estímul o mecanismes no específics) i la influència d'altres factors que sovint es confonen amb l'autogeneració (p. ex., la intensitat de l'estímul i la predictibilitat).

Aquesta tesi tracta d'elucidar els mecanismes subjacents als efectes de les accions en el processament auditiu. En tres estudis, hem examinat els efectes d'autogeneració des dels angles de les respostes conductuals, en concret, el processament perceptiu i la codificació de la memòria, i la fisiologia bàsica, en concret, les respostes electrofisiològiques i els processos neuromodulatoris. L'estudi I va explorar possibles interaccions entre les accions i la intensitat d'estímul en la detecció del so i la discriminació de la sonoritat, mentre examinava quin aspecte de la percepció està sent modulat (sensibilitat o biaix). L'estudi II va abordar els efectes de les accions en el processament sensorial i la memòria dels sons concurrents, però impredictibles, mentre explorava per primera vegada la implicació dels sistemes neuromoduladors en les modulacions induïdes per l'acció. L'estudi III aïlla l'autogeneració i la predictibilitat – dos factors que han estat barrejats prèviament– i avalua com modulen interactivament les respostes auditives, la neuromodulació subcortical i la memòria. Finalment, els estudis II i III també van provar possibles relacions entre la memòria, l'atenuació sensorial i la neuromodulació subcortical.

Les troballes conductuals han contribuït al coneixement respecte al processament perceptiu de baix nivell i la memòria. L'estudi I va demostrar que les accions interaccionen amb la intensitat de l'estímul en el biaix perceptiu, però no en les mesures de sensibilitat. Els estudis II i III van demostrar que les accions modulen la memòria, però la direcció dels efectes depèn de les contribucions de la predictibilitat i del tipus d'acció. Quan les accions van guiades i proporcionen predictibilitat temporal, el rendiment de la memòria cau i correlaciona amb la magnitud dels efectes d'atenuació per als sons autogenerats. En canvi, quan la contribució de la predictibilitat es controla i es proporcionen alternatives d'acció, l'acció millora la memòria, però aquesta millora no correlaciona amb la supressió.

Respecte als efectes de les accions en el processament sensorial, aquesta tesi ha avaluat l'especificitat dels efectes d'atenuació i ha aclarit el paper dels processos neuromoduladors en la supressió sensorial. L'estudi II va demostrar que les accions guiades atenuaven les respostes sensorials i augmentaven el diàmetre de la pupil·la malgrat l'absència d'una relació predictiva entre so i acció, però la supressió sensorial i la neuromodulació subcortical no estaven relacionades. L'estudi III va emprar sons autogenerats i externament generats que eren igualment predictibles o impredictibles en identitat, temps i probabilitat d'ocurrència i que es presentaven a prop o lluny de l'acció. Amb retards curts, l'atenuació només es va obtenir per a sons predictibles. Amb retards llargs, l'autogeneració i la predictibilitat no interaccionaven: l'autogeneració atenuava la N1, mentre que la predictibilitat atenuava Tb i P2. Crucialment, el diàmetre de la pupil·la va

augmentar durant les accions, va interaccionar amb la predictibilitat, i va corelacionar amb la atenuació de Tb quan els efectes eren més forts (amb predictibilitat i retards curts).

En conjunt, aquesta tesi desembrolla la contribució de factors més enllà de l'autogeneració (com ara intensitat de so, predictibilitat i retard entre so i acció) en la modulació de la percepció, la memòria i les respostes neurofisiològiques per a estímuls autogenerats, i demostra que les accions desencadenen una cascada de processos estímul-específics i no específics -presumiblement impulsats per processos neuromoduladors subcorticals- que col·laboren per a orquestrar el processament auditiu i la codificació de memòria.

FOREWORD

This work has been carried out at the Brainlab-Cognitive Neuroscience Research Group (Centre of Excellence established by the Generalitat de Catalunya, 2009SGR11) at the Department of Clinical Psychology and Psychobiology, Faculty of Psychology, University of Barcelona (UB; Barcelona, Catalonia, Spain). This work has been supported by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the European Social Fund with the predoctoral fellowship FI-DGR 2019 awarded to Nadia Paraskevoudi, and additional funds awarded to Iria San Miguel Insua (PSI201452573P, PSI201785600P and RYC201312577, funded by the Spanish MINECO MCIN/AEI/ 10.13039/501100011033, “ERDF A way of making Europe”, and “ESF Investing in your future”). This work has been additionally supported by the MDM-2017-0729-18-2M Maria de Maeztu Center of Excellence UBNeuro, funded by MCIN/AEI/ 10.13039/501100011033, and by the Excellence Research Group 2017SGR-974 funded by the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya.

LIST OF ORIGINAL STUDIES

STUDY I

Paraskevoudi, N., & SanMiguel, I. (2021). Self-generation and sound intensity interactively modulate perceptual bias, but not perceptual sensitivity. *Scientific Reports***,** 11**,** 17103. https://doi.org/10.1038/s41598-021-96346-z

STUDY II

Paraskevoudi, N., & SanMiguel, I. (2021). Sensory suppression and increased neuromodulation during actions disrupt memory encoding of unpredictable self-initiated stimuli [Preprint]. bioRxiv, Neuroscience. **Under review in** *Psychophysiology* (submitted on December 14th, 2021). https://doi.org/10.1101/2021.12.15.472750

STUDY III

Paraskevoudi, N., & SanMiguel, I. (in preparation). Semi self-paced actions enhance memory of sounds but interact with predictability in shaping auditory and neuromodulatory responses.

ABBREVIATIONS

A	Auditory-only
ACtx	Auditory cortex
AN	Active near-threshold
ANOVA	Analysis of Variance
AS	Active supra-threshold
BF	Bayes Factor
BOLD	Blood oxygen level dependent
E	Empty
EEG	Electroencephalography
ERP	Event related potentials
ICA	Independent Component Analysis
JND	Just noticeable difference
LC-NE	Locus coeruleus norepinephrine system
M	Motor-only
MCtx	Motor cortex
MA	Motor-auditory
MEG	Magnetoencephalography
MMN	Mismatch Negativity
PN	Passive near-threshold
POE	Point of Objective Equality
PS	Passive supra-threshold
PSE	Point of Subjective Equality
SNR	Signal-to-noise ratio
1T	“One test sound at encoding” sequences
2AFC	Two alternative forced choice
2T	“Two tests sounds at encoding” sequences
V	Visual-only

CHAPTER 1: Introduction

Maybe one of the most crucial abilities for living organisms is to optimally adapt to the noisy information in their environments. However, biological organisms do not perceive their environment just as a mere and truthful reproduction of the inputs reaching their sensory apparatus. Like other species, humans as well interpret the external world through an interaction between the self and its surroundings.

Over the past few years, theories of neuroscience and experimental psychology have come to the agreement that perception is not a passive stimulus-driven process, but instead it constitutes a quite complex interaction between bottom-up, stimulus-driven processes triggered by the inputs to our eyes, skin, ears, and top-down predictions about what we expect to happen (Press et al., 2020; Friston, 2009; Korka et al., 2021). Predictions might have multiple sources and might provide information about different aspects related to the predicted stimulus (e.g., identity, timing, or probability of occurrence). For example, sensory-based predictions can be formed based on general statistical regularities in the environment or multisensory associations (e.g., predicting a sound based on a visual stimulus), while a different set of predictions can be formed through our own actions (e.g., motor- or intention-based predictions) or through learning as in the case of music or language (for reviews see Press et al., 2020; Bendixen et al., 2012; Schröger et al., 2015; Korka et al., 2021).

Specifically, in the auditory domain, sensory-based predictions have been typically studied using the oddball paradigm. In this paradigm, participants are presented with a series of a frequent, standard tones and occasionally a rare, deviant sound is presented (i.e., mismatching the established expectation) that differs from the standard one in at least one physical feature (e.g., pitch or intensity; for a review see Näätänen et al., 2007). Deviant, novel stimuli elicit two responses, the mismatch negativity (MMN) and the P3. The former occurs at around 200 ms post-stimulus and indicates the detection of a stimulus deviating from an expectation (Näätänen et al., 2007; Escera & Malmierca, 2014), while the latter occurs at around 300 ms (Escera et al., 1998) and reflects the updating of the internal models about auditory regularities (Donchin & Coles, 1988; Polich, 2007). Conversely, standard stimuli elicit a response, termed repetition positivity (Haenschel et al., 2005; Baldeweg, 2007; Costa-Faidella et al., 2011) which reflects the neural response to inputs that match an established expectation. Predictions made in this paradigm are considered sensory-based, in that the expectation is formed by statistical regularities within a sequence.

However, as briefly mentioned before, another specific instance of predictive processing is the ability to anticipate the sensory consequences of our own actions. Pioneering animal work was the first to assume that there must be a mechanism generating predictions about the sensory outcome of a motor act (i.e., reafference principle; von Holst & Mittelstaedt, 1950; Sperry, 1950). This idea was partly inspired by the fact that although eye movements cause the visual representation of our environment to move across the retina, the brain can create a stable visual image of our surroundings, using a mechanism that can compensate for the motion effects during saccadic movements. Later, the nominal study by Schafer & Marcus (1973) brought this idea into the auditory domain, by showing that self-initiated auditory stimulation elicits smaller amplitudes of the auditory N1 - a response elicited about 100 ms after sound onset - than the N1 amplitudes to identical but machine-delivered inputs. This differential processing for self-induced stimuli is the main focus of the present thesis.

Since the first description of the effect by Schafer and Marcus (1973), several lines of research have shown that actions suppress the processing of the self-generated reafferent input (e.g., action-induced blindness, saccadic suppression, self-generation of stimuli), as proposed by the reafference principle (Sperry, 1950; von Holst & Mittelstaedt, 1950), and that the attenuation effects are modality-independent and widespread throughout the animal kingdom (Weiskrantz et al., 1971; Crapse & Sommer, 2008; Cullen, 2004; Straka et al., 2018; Brooks & Cullen, 2019). For example, studies with primates have shown that the vestibular nucleus activity is suppressed in response to vestibular reafference during active compared to passive head movements (Cullen, 2012; Brooks et al., 2015; Roy, 2004). In mice, auditory cortical responses to self-generated sounds are suppressed compared to externally-generated inputs and the effect is specific for tone frequencies that were associated with the animal's movement (Schneider et al., 2018). In crickets, auditory processing is inhibited in phase with the animal's chirps, allowing the animal to maintain sensitivity to external, environmental sounds (Poulet & Hedwig, 2006; Poulet & Hedwig, 2003). In the electrosensory domain, electric fish attenuates its electrosensory reafference, thereby responding only to externally-generated electrical stimulation (Fukutomi & Carlson, 2020; Sawtell, 2017).

Inspired by the studies by Sperry and von Holst and Mittelstaedt in the 50's, the predominant view attributes these attenuation effects to a predictive mechanism that allows the organism to anticipate the sensory consequences of its own actions. Motor control theories have further refined this idea by implicating forward and inverse models in sensorimotor behaviour and sensory attenuation (Miall & Wolpert, 1996; Wolpert et al., 1995): Forward models are thought to estimate the current and future state of the system by combining the predicted sensory consequences of an action and the actual sensory input, while inverse models, on the other hand, serve to estimate a motor plan (and its associated motor commands) that can contribute in achieving a desired state. This line of work has led the dominant cancellation theories (Blakemore, Wolpert et al., 1998; Frith et al., 2000; Wolpert & Flanagan, 2001) to explain the attenuation effects as the operation of a forward model that generates prediction signals before or during an action and sends them from the motor to the corresponding sensory cortices to attenuate responses to self-initiated and predictable stimulation when the prediction matches the actual sensory input (Sperry, 1950; von Holst & Mittelstaedt, 1950). Nevertheless, it is worth noting that although sensory attenuation has been reported in most studies, it is also known that movement-related signals influence sensory processing in myriad ways besides cancellation of reafference, with some studies reporting also enhanced responses to self-initiated stimulation (Reznik et al., 2014; Müller-Preuss & Ploog, 1981; Eliades & Wang, 2003) that in humans might be limited in the posterior superior temporal gyrus (Flinker et al., 2010). The following chapters present an overview of the action effects on perceptual and sensory processing in the auditory modality.

1.1. Immediate effects of actions on sensory processing

1.1.1. Action effects on perception

Undoubtedly, expectations shape how we perceive the world, allowing us to extract statistical regularities to interact with the environment in an optimal way. Evidence from the wider sensory literature points to facilitatory effects of expectation on perceptual processing (Jaramillo & Zador, 2011; Pinto et al., 2015; Stein & Peelen, 2015), suggesting that we are more likely to detect what we expect. However, these findings usually conflict with work from the action domain (for a review on this conflict see Press et al., 2020).

It is known that actions affect perception in myriad ways, shaping subjective time and sensory processing. For example, voluntary action modulates our perception of time, as shown by the fact that the temporal interval between an action and its sensory consequence is perceptually compressed (i.e., intentional binding; Haggard et al., 2002). Additional evidence has shown that action execution can interfere with concurrent encoding of visual information linked with the motor act (i.e., identifying arrowheads while performing spatially compatible button presses; action-induced blindness; Kunde & Wühr, 2004) and that visual sensitivity is largely suppressed during saccades (i.e., saccadic suppression; Ross et al., 2001), pointing to suppression effects as proposed by the reafference principle (Sperry, 1950; von Holst & Mittelstaedt, 1950). But maybe the most well-known and intuitive observation of how actions affect perception is the fact that we cannot tickle ourselves, but we can be tickled by others (e.g., Blakemore, Wolpert et al., 1998; Claxton, 1975).

Over the last few years, a series of studies has examined perceptual attenuation mainly in the auditory (Sato, 2008; Weiss et al., 2011a, 2011b), and tactile (e.g., Kilteni et al., 2020; Claxton, 1975; Blakemore, Wolpert et al., 1998) modality, by employing discrimination tasks (Sato, 2008; Weiss et al., 2011a, 2011b) to assess whether perception is attenuated (but there is also work with detection tasks in the visual domain; e.g., Cardoso-Leite et al., 2010). In paradigms with auditory stimulation, participants are typically presented with two sounds, a first standard one of fixed intensity, and a second comparison one of varying intensities (typically ± 3 dB in steps of 1 dB relative to the standard one). In active trials, the standard tone is self-generated, that is participants had to press a button to generate it, while in passive trials, the standard tone is passively presented by the computer. The comparison sound is always passively presented to the participants, and they have to respond which of the two sounds is louder. Two parameters of interest are typically extracted from this task: the point of subjective equality (PSE), which is a measure of perceptual bias and represents the intensity at which the standard tone was perceived as loud as the comparison one, and the just noticeable difference (JND), which reflects the participants' sensitivity for the sound discrimination. Although JND has not been found to be affected by actions (Sato, 2008; Weiss et al., 2011a, 2011b, but see Endo et al., 2021 for evidence showing lower JND values in a motor compared to a non-motor condition), most studies have indeed provided evidence for lower PSE during self-generation compared to passive listening (Sato, 2008; Weiss et al., 2011a, 2011b) pointing to attenuated loudness perception for sounds produced by actions.

However, more recent work has raised the possibility of the perceptual attenuation being dependent on the intensity of the input. Reznik and colleagues (2015) employed a discrimination task as done by previous work, but in addition to the self-generation (i.e., self- or externally-generated) they also manipulated the sound intensity. Specifically, participants had to judge the perceived intensity of self- and externally-generated sounds that were presented at a supra- or a near-threshold intensity. Nevertheless, unbeknownst to the participants, the standard and the comparison tones were always presented at the exact same intensity, which did not allow for the calculation of PSE and JND due to the lack of varying comparison intensities. Instead, they calculated how often participants judged the first standard tone as being louder than the comparison one. Therefore, lower frequency of "Standard louder" responses in the self-generation trials would reflect a perceptual attenuation for these sounds, while high frequency of "Standard louder" responses in these trials would reflect enhanced perception. In line with their hypotheses, their results were the first to show an interaction between intensity and sound source, that is, they replicated the attenuation effect when the self-generated sounds were loud (cf. Sato, 2008; Weiss et al., 2011a, 2011b), but the same sounds were perceptually enhanced when presented at near-threshold intensities.

In a closer look, it seems that perceptual suppression is reported in human studies with supra-threshold stimulation only and is mostly reflected in modulations in perceptual bias (PSE) rather than sensitivity (JND). Reznik et al.'s (2015) study fit with these findings but also provide an interim explanation to why human and animal detection studies (that necessarily use near-threshold stimulation) have been reporting the opposite effect, that is enhanced detection ability for stimuli tied with actions (in both the auditory and visual domain: Reznik et al., 2014; Yon & Press, 2017; Cao & Gross, 2015a; Carcea et al., 2017; Neske et al., 2019; but see Myers et al., 2021 for no effects). Nevertheless, it should be noted that behavioural paradigms in human and animal work are not directly comparable due to the differences in type of actions employed (i.e., voluntary/self-paced or cued button-presses, locomotion, or motor response to a stimulus, e.g., Go-stimulus), the relationship between action and sound (i.e., sound identity and timing could be predicted by the action in human studies, but not in animal studies that present a sound during a general active state such as locomotion), and the conditions compared (e.g., predictable vs. unpredictable or predicted vs. mispredicted).

1.1.2. Action effects on neurophysiological responses

Animals

Evidently, actions do not only modulate behavioural responses, but they also affect sensory processing. Animal studies (most typically with rodents) have assessed these effects by examining auditory processing during locomotion or when animals are trained to generate a sound by pressing a lever. Most of these studies have reported suppression of auditory cortical responses during a wide range of movements (McGinley et al., 2015; Zhou et al., 2014; Kelley & Bass, 2010; Carcea et al., 2017; Buran et al., 2014; Singla et al., 2017; Schneider & Mooney, 2018; Schneider et al., 2014; Schneider et al., 2018, but see Eliades & Wang, 2003 for the opposite effect), but also during active task engagement (Otazu et al., 2009; Kuchibhotla et al., 2017). Interestingly, this suppression is not limited to sensory-specific cortical areas: Rummell and colleagues (2016) recorded responses to self-generated stimuli in the auditory thalamus, the auditory cortex, and the hippocampus, and they found that despite being evident in all these areas, the attenuation effect was most pronounced in the hippocampus.

Despite the mounting evidence supporting the suppression effect of movement on sensory responses, the exact mechanism underlying this effect remains unknown and seems to be modality dependent. For example, actions enhance visually evoked responses in the visual cortex (Niell & Stryker, 2010), but attenuate sound-evoked activity in auditory cortex (Zhou et al., 2014; Schneider et al., 2014; McGinley et al., 2015). Focusing on the auditory domain, it has been proposed that the attenuation effect may be partly driven by neurons in the secondary motor cortex that extend axon collaterals to auditory cortex, where they make fast excitatory synapses on inhibitory interneurons and pyramidal cells (McGinley et al., 2015; Nelson, Schneider et al., 2013; Schneider et al., 2014). It should be noted that although movement in general seems to suppress responsiveness to auditory stimulation overall (Schneider et al., 2014), different types of movement and the relevance for task performance may exert distinct effects on cortical responsiveness. For example, Clayton et al. (2021) showed that Layer 6 corticothalamic neurons and other Layer 6 units began spiking hundreds of milliseconds prior to orofacial movements linked to sound presentation and reward, but not to other movements such as locomotion, which were not linked to an explicit behavioural task. Most importantly, it should be noted that to date direct evidence for motor-induced inhibition of the auditory cortex comes from work with rodents, but there is no evidence in primates for direct anatomical connections between primary motor cortex (or supplementary motor area) and auditory cortex (specifically the superior temporal gyrus).

However, the functional coupling might suggest that motor pathways may exert modulations through indirect anatomical connections on auditory areas (for a comprehensive review see Reznik & Mukamel, 2019).

Humans

In humans, most studies have assessed the effects of actions on auditory processing by using the so-called contingent self-generation paradigm (e.g., Baess et al., 2008; Knolle et al., 2012; Sowman et al., 2012; Ott & Jäncke, 2013; Timm et al., 2013, 2014). In this paradigm, participants have to perform button presses that always generate a fully predictable sound (self-generated) or listen passively to sounds presented by the computer (externally-generated). In a separate block, participants perform button presses that do not generate any sound (motor-only control). Subsequently, the sensory responses to externally-generated sounds are compared to the responses elicited by the self-generated sounds after correcting them for motor activity (by subtracting the motor-only responses from the responses to the self-generated ones). Typically, most of these studies assessed the N1 and P2 components of the stimulus-evoked response, which are the negative and positive deflections elicited around ~100 and ~200 ms after sound onset, respectively. Independently of whether self- and externally-generated sounds are presented in the same or in separate blocks, there is compelling evidence showing that both of these components are strongly attenuated for the self- compared to the externally-generated inputs (Baess et al., 2011; Baess et al., 2009; Baess et al., 2008; Martikainen et al., 2004; Mifsud & Whitford, 2017; Saupe et al., 2013; Schafer & Marcus, 1973; Timm et al., 2013; Klaffehn et al., 2019; Reznik et al., 2021; Bolt & Loehr, 2021; for extensive reviews see Schröger et al., 2015; Korka et al., 2021). Action-induced modulations have been also found in earlier responses (Pa and Nb; 23-33 and 40-46 ms post-stimulus; Baess et al., 2009), as well as in hippocampal and parahippocampal activity (Halgren, 1991; Mukamel et al., 2010, in line with the animal study by Rummell and colleagues, 2016), pointing to widespread effects of actions across the cortical hierarchy. Note that there are also studies that instead of having participants press buttons, they used participants' own speech (e.g., Gunji et al., 2000; Kudo et al., 2004; Baess et al., 2011; Mock et al., 2011) and found the same suppression effect. However, the present thesis will mainly focus on button-press studies since the associations established between actions and sounds differ from the natural situation speech represents, but also because of the many factors that need to be considered in the case of speech, such as differences in sound intensity and distortion of the auditory input due to bone conduction (Schröger et al., 2015).

1.1.3. Models and theories of action effects

The increasing interest in understanding how animals process self-produced stimulation as presented in the previous sections has given rise to attempts of constructing theoretical models that could provide a plausible explanatory framework for the self-generation effects on perceptual and sensory processing. Several models have been proposed, however as summarised in this section, they often disagree on how predictions shape the processing of the sensory consequences of our actions. It should be noted that the differences between these models might stem from differences in the conditions to be compared (e.g., self-generated predictable vs. externally-generated unpredictable; Baess et al., 2008 or self-generated predicted vs. self-generated mispredicted; Yon et al., 2019) and different types of predictions (e.g., arbitrary association between an action and its perceptual effect; Roussel et al., 2013 or sensory reafferent predictions based on an efference copy of motor commands; e.g., Baess et al., 2008; Kilteni et al., 2020).

Cancellation / motor forward models

The most known model for explaining the action-induced sensory and behavioural attenuation has been the cancellation model (alternatively also called comparator model; Blakemore, Wolpert et al., 1998), which also assumes that perceptual and neurophysiological attenuation are related (e.g., Kilteni et al., 2020; Sato, 2008). This model was largely inspired by animal physiological work (Sperry, 1950; von Holst & Mittelstaedt, 1950) that was the first to suggest that a copy of the motor command (i.e., efference copy) is sent to the sensory cortices, allowing the system to calculate the predicted consequences of the movement (i.e., corollary discharge). These predictions are, subsequently, compared to the actual (i.e., received) sensory input, and if there is a match between the two, prediction error is reduced, effectively dampening sensory responses (see Figure 1). This idea has later captured the attention of motor control theories that have highlighted that sensory prediction is a critical part of motor control (Miall & Wolpert, 1996; Wolpert et al., 1995) and that this is achieved by the operation of forward and inverse modelling: One the one hand, forward models estimate the future state of the system and the sensory consequences of a motor act, by considering the current state of the system, as well as the motor commands sent. On the other hand, internal inverse models allow the system to estimate the motor commands required to achieve a desired state, for example by performing an online adjustment of motor commands to test the alternative outcomes of an action.

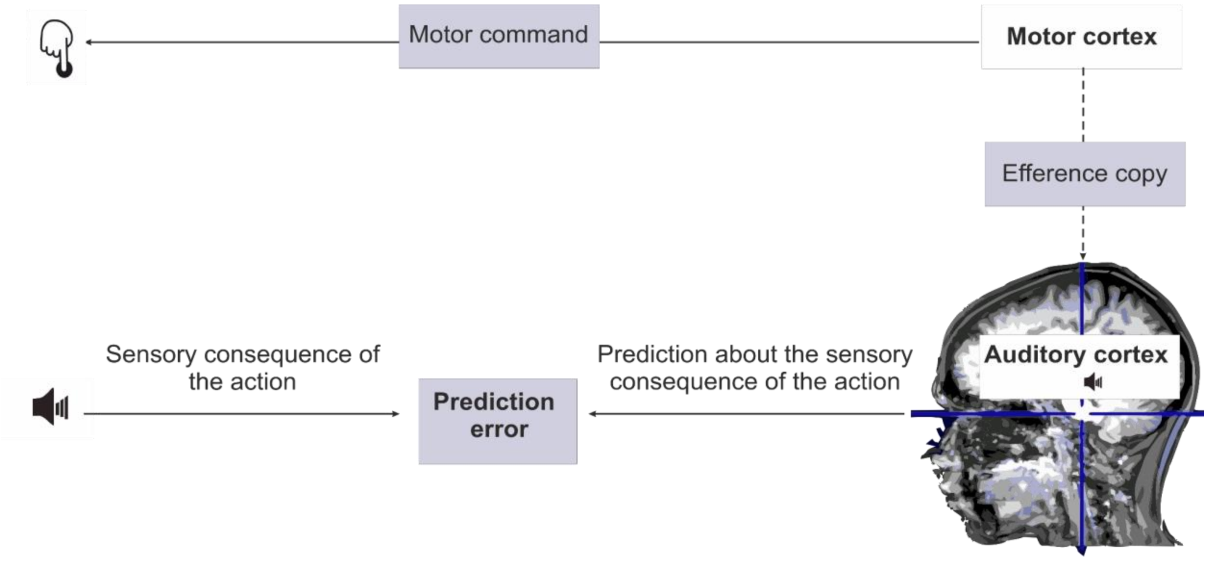


Figure 1. Schematic representation of the cancellation model. During actions, the motor cortex sends a motor command that allows motor execution (e.g., button press), but it also generates a copy of this motor command (i.e., efference copy) that is sent to the auditory cortex. This efference copy effectively represents the prediction about the sensory consequence of the action which is compared to the actual sensory input (e.g., the sound) and the difference between the two is called prediction error. When there is a match between prediction and sensory feedback, prediction error is reduced, which is reflected in attenuated sensory responses to the self-generated stimulus.

Theory of event coding / Ideomotor theory

Another framework for explaining the effects of actions on perception has been provided by the ideomotor theory, which was later further refined by the theory of event coding (Hommell et al., 2001). According to this theory, organisms learn that their actions have certain sensory consequences and use these action-effect couplings to select actions according to the desired outcome (Hommell et al., 2001). The main idea of this theory is that actions are represented by the codes of their sensory effects, and activation of such sensory codes automatically leads to activation of the related motor patterns, that is we internally activate the sensory outcome of our own action (Hommel et al., 2001). This link between perception and action has been supported by several studies (for a review see Shin et al., 2010), with the most known one being the study by Elsner & Hommel (2001). In this study, participants performed left and right keypresses that were associated with different sounds (high and low pitch, respectively). In the subsequent test phase, participants had to perform keypresses in response to these high- and low-pitch tones. In some cases, the stimulus-response mapping at test was compatible with the mapping learned at acquisition. They showed that participants' reaction times were faster with compatible mappings than with incompatible ones, and that they more frequently chose the mapping compatible with the acquisition phase than the incompatible one. These findings point to a bidirectional link between motor and sensory events established through learned action-effect associations that can later affect action selection and initiation. Supporting evidence to this idea comes also from imaging studies showing an activation of brain areas involved in the planning of endogenously controlled (i.e., not stimulus triggered) motor events, when presented with previously learned action-effects couplings (Melcher et al., 2008). It is worth noting that ideomotor accounts are compatible with cancellation theories and forward modelling in that they both highlight the critical role of action-effect anticipation. However, a critical conceptual difference between the two is the origin of prediction: Cancellation theories suggest that predictions are driven by efference copies, while according to ideomotor theories, predictions are dependent on one's intentions (Dogge et al., 2019).

Active inference / predictive coding

The predictive coding framework has also been used in explaining the self-generation effects, constituting one of the most influential models for explaining how perception arises by the complex interaction of predictions and sensory input (Clark, 2013; Friston, 2001; Brown et al., 2013). The main idea of this framework is that the brain uses prior beliefs to generate predictions about the causes of sensory events in our environment to minimise the level of uncertainty and surprise that usually accompanies unexpected events (Friston, 2001). In fact, according to this framework the goal of every biological organism is the minimization of surprise or uncertainty which is reflected in a reduction of prediction error. Specifically, at the core of this model stands the assumption of the brain as a hierarchical system, consisting of representational units encoding the predicted causes of sensory input and prediction error units: Representational units send prediction from higher levels of the cortical hierarchy via backward, top-down projections. Prediction error units compare these predictions with the actual input that is sent from lower levels of the cortical hierarchy via feedforward, bottom-up projections. Critically, at each level of the cortical hierarchy, only the prediction error, that is the residual signal when comparing predicted and actual input, is moving upwards to higher levels. More importantly, prediction error is thought to be reflected in the activity of superficial pyramidal cells that are behind the generation of EEG signals (Feldman & Friston, 2010). This inevitably implies that the attenuated evoked responses, typically seen in self-generation studies, reflect a reduction of prediction error thanks to the well-formed predictions from higher areas that allow the organism to anticipate the sensory consequences of its motor act (for a review see Schröger et al., 2015). For the sake of completeness, it should be noted that a second explanation within the predictive coding framework attributes the suppression effects to attentional mechanisms that reduce the precision of sensory evidence during movement, thereby resulting in a decrease in neuronal gain (Brown et al., 2013). This is equivalent to attending away from the consequences of self-initiated acts. This account has been recently proposed to explain the tactile gating effect; i.e., the reduction in the precision of somatosensory input on the moving limb during the movement (Kilteni & Ehrsson, 2022), but in the auditory domain the self-generation effects have been suggested to be independent of attentional allocation (Saupe et al., 2013; Timm et al., 2013; for comprehensive reviews on the interactions between prediction and attention see Lange, 2013 and Schröger et al., 2015).

Preactivation account

Another model compatible with the predictive coding framework is the preactivation account (Cardoso-Leite et al., 2010; Roussel et al., 2013; Waszak et al., 2012). According to this account, action execution activates the representation of the sensory effects associated with the motor act, leading to increased activation of the network representing the sensory event, even before the processing of the actual input has started. The increased activation of the predicted sensory consequence of an action is conceptualised as increases in baseline activity (i.e., noise levels), thereby rendering the actual motor-triggered input less discriminable from baseline, and consequently reducing detection and/or discrimination sensitivity (typically examined using the d’ sensitivity index as proposed by signal detection theory; Roussel et al., 2013). In contrast, discrimination of externally-generated stimulation is facilitated given the lower levels of noise before the start of sensory processing (since the activation of the networks responsible for stimulus processing is at baseline level). This account has found support in studies showing reduced contrast sensitivity for self-generated stimulation that is compatible with previously learned action-effect couplings (Roussel et al., 2013). However, an important limitation of this account is that although it provides a plausible explanatory framework for the behavioural effects of sensory attenuation, it fails in elucidating the neurophysiological mechanisms that could explain how the motor-induced preactivation of sensory networks would result in attenuating sensory responses.

Sharpening accounts

The core tenet of the models presented so far, namely that sensory processing of predicted action consequences is suppressed, contradicts predictive processing models developed in the wider sensory cognition literature, outside of the action domain. Specifically, sharpening models (also called Bayesian accounts of perception), mainly inspired by studies in the visual domain (e.g., Yuille & Kersten, 2006; for a review see Press et al., 2020; Summerfield & de Lange, 2014), underline that it is adaptive to perceptually prioritise what we expect. The proposed mechanism has been suggested to ‘turn up’ the volume on units that encode the expected (rather than unexpected) input (Summerfield & de Lange, 2014), which has been transferred to the action domain to suggest that the motor-driven suppression proposed by cancellation theories is limited to units tuned away from the expected input, while units encoding the expected self-generated input increase their activity. This, consequently, results in a sharpened population response and higher signal-to-noise ratio which behaviourally would be reflected in enhanced detection sensitivity for the expected (Yon & Press, 2017). The sharpening model has found support from studies of the wider sensory cognition literature, but also from paradigms employing self-initiated stimulation, showing that we are more biased in reporting the presence of expected over unexpected stimuli (Chalk et al., 2010; Wyart et al., 2012; Yon et al., 2019). This has been supported by neuroimaging work as well, demonstrating that suppression for expected events is limited to voxels tuned away from the expected input, and that the expected events were better decoded from visual brain activity than unexpected ones, pointing to higher fidelity representations for what we expect (Yon et al., 2018). However, it should be noted that usually studies supporting the sharpening account in the action domain differ from studies reporting sensory cancellation for reafferent input in one important aspect: The former typically compare predicted vs. mispredicted (i.e., deviant) stimuli (i.e., visual stimuli congruent or incongruent with concurrently performed actions; e.g., Yon et al., 2018), while the latter have mainly compared responses to predictable vs. unpredictable inputs (i.e., self- compared to externally-generated sounds; e.g., Baess et al., 2008).

Opposing process

More recently, there has been an attempt to reconcile the sharpening and cancellation models of the effects of actions on sensory processing (Press et al., 2020). Inspired by the need to create an explanatory framework that would capture the effects of predictions on perception within and outside the action domain, Press and colleagues developed the opposing process theory. According to this theory, perception is in principle biased towards what we expect, as proposed by sharpening (Bayesian) models, which could be mediated through preactivation of expected units (e.g., Kok et al., 2017). However, if a highly surprising (i.e., unexpected) event appears, later processes will be triggered, allowing us to update our models in the case of informative, unexpected events. Critically, this theory suggests that these two opposing processes co-exist, but operate on different timescales. Support for this idea comes from work showing that intensity perception is stronger for expected events shortly after their presentation (i.e., 50 ms post-stimulus), but this effect is reversed at a later time point (i.e., intensity perception is stronger for unexpected events 200 ms post-stimulus; Yon & Press, 2017). Press and colleagues argue that this distinction of sharpening or cancellation as a function of timescale could fit with M/EEG studies showing that the dampened responses to self-initiated stimulation are typically found from ~100 and onwards (i.e., within the N1 and P2 windows; for a review see Schröger et al., 2015). However, it should be underlined that this is only partial support to their theory given the lack of evidence of enhanced evoked responses at earlier time points (or evidence for suppressed early auditory responses such as the Pa and Nb as reported by Baess et al., 2009).

Reznik & Mukamel model

Another model that sought out to reconcile the contrasting findings of action-induced attenuation and enhancement has been recently proposed by Reznik and Mukamel (2019), focusing specifically on the auditory modality. Contrary to the opposing process theory that pointed to the existence of opposing processes operating at different timescales, Reznik and Mukamel suggested that the direction of the action effects on perceptual and sensory processing depends on the environmental context, namely the stimulus' intensity. Inspired by evidence from animal physiology studies (e.g., Schneider et al., 2014), they suggested that the inhibitory effects of motor cortex on auditory areas during actions may dampen or enhance processing of self-generated auditory stimulation depending on its intensity. According to their model, the action-induced cancellation of auditory responses (Carcea et al., 2017; Buran et al., 2014) results in reducing activity at the population level, but meanwhile leads to more selective responses, thereby increasing the signal-to-noise ratio. This inevitably implies that population activity is always reduced during movement (Nelson, Schneider et al., 2013; Zhou et al., 2014), that is irrespective of stimulus intensity, a finding replicated in animal studies employing both faint (i.e., 10-30 dB sound pressure level in Buran et al., 2014) and loud auditory stimulation (e.g., 40–50 and 80 dB sound pressure level in Zhou et al., 2014 and Nelson, Schneider et al., 2013, respectively). However, their assumption has another, more important, implication: If movement results in more selective responses as they propose, then the signal-to-noise ratio for self-initiated stimulation should be higher with soft than loud stimulus intensities. This is a valid assumption considering that loud auditory stimulation elicits responses on "best-frequency" neurons, but also on neurons tuned to neighbouring frequencies. In contrast, with faint auditory stimulation, only "best-frequency" neurons will respond, with a few scarce responses from nearby frequency neurons (for a review see Reznik & Mukamel, 2019). Faint and salient contexts (i.e., near- or supra-threshold self-triggered stimulation) will, therefore, differ in how they will be affected by the global inhibition exerted by the motor cortex during movement. In either case, spontaneous activity in the auditory cortex will be inhibited, resulting in a relative boost of sound-evoked activity compared to the background noise (Carcea et al., 2017; Buran et al., 2014). With faint stimulation, this relative boost will be noticeable, since the motor-induced inhibition will completely silence the activity of nearby-frequency neurons, leaving responses only from the "best-frequency" neurons, thereby improving detection sensitivity due to the higher signal-to-noise ratio. In contrast, sensitivity is already at ceiling when the auditory consequence of our actions is supra-threshold. Therefore, this proposal suggests that in salient contexts, actions affect loudness perception, resulting in attenuated perception which is driven by the reduced population activity in the auditory cortex during movement. In contrast, in faint contexts, actions affect sensitivity, allowing for better detectability for self-initiated and near-threshold stimulation due to the boost in signal-to-noise ratio.

1.1.4. Specific predictions or unspecific modulations?

As we have seen so far, extant models disagree on how actions affect sensory processing (suppression, enhancement, or both), while the mechanism driving the effects remains under debate. Despite the apparent debate, most studies in the action domain report suppression for self-initiated stimulation and the dominant models attribute the suppression effects to predictive stimulus-specific processes during an action (e.g., cancellation models as shown in Figure 1). Importantly, the dominant cancellation account has one important implication that stands at the core of the present thesis: For the effects to be specific, they should be specific to the stimulus (i.e., only predictable self-generated stimuli should be suppressed) and mediated by sensory-specific areas (i.e., the effect should reflect modulation of activity in the sensory-specific areas).

Specificity of the effects

In fact, there is sufficient evidence that the suppression for self-generated stimulation is (at least partly) specific. For example, functional imaging studies have reported reduced activity in the sensory specific cortices for self-generated input compared to externally-generated signals (Blakemore, Rees et al., 1998; Christoffels et al., 2007; for a review see Schröger et al., 2015). Specifically in the auditory modality, MEG studies have examined the magnetic counterpart of the N1 directly at its auditory cortex sources and have provided evidence of reduced amplitude to self-generated sounds compared with passive listening, both when the self-generated input was participants' own speech (Numminen & Curio, 1999; Numminen et al., 1999; Curio et al., 2000; Gunji et al., 2001; Ventura et al., 2009; Niziolek et al., 2013; Tian & Poeppel, 2013) and sounds triggered by button presses (Martikainen et al., 2004; Aliu et al., 2009). More direct evidence comes from an electrocorticography study showing that multiunit activity in the superior temporal gyrus was suppressed for spoken words compared to passive listening of the same words, with the effect peaking within the N1 latency range, at 150 ms post-stimulus (Flinker et al., 2010).

Further evidence supporting the specificity of the effects comes from studies examining responses when the received input did not match the predictions. For example, both speech studies suggest that the suppression effect is only specific to the predicted unaltered feedback (McGuire et al., 1996; Hirano et al., 1997; Houde et al., 2002; Hashimoto & Sakai, 2003; Heinks-Maldonado et al., 2005; Fu et al., 2006; Ott & Jäncke 2013), while it is reduced or even reversed when the self-generated auditory stimulus does not match the prediction (i.e., with altered auditory feedback; Heinks-Maldonado et al., 2005; Behroozmand et al., 2011). Similarly, Ott & Jäncke (2013) showed N1 enhancement, rather than suppression, when the pitch of the self-generated input was not predictable, indicating that the N1-suppression effect is highly dependent on a specific prediction about the identity of the upcoming sound. Conversely, button press studies have confirmed the specificity of the effect with non-speech stimuli, by showing suppressed responses to tones when their pitch was congruent with the hand-specific prediction compared to responses to tones with incongruence between pitch and hand-specific prediction (Hughes et al., 2013b).

Although these studies have focused on sound identity to manipulate the match between prediction and input, actions afford temporal predictions as well: Pressing a button could allow one to predict when the stimulus will be generated. Baess and colleagues (2008) assessed whether timing and identity predictability both result in N1 suppression by asking participants to perform button presses that would generate a sound of either predictable or unpredictable frequency at a predictable or unpredictable time point. Attenuation was obtained under both predictable and unpredictable contexts (both for identity and timing), but the attenuation effect was largest when both identity and timing could be predicted. This finding supports partly the specificity of the effects; however, it also shows that the system can anticipate the sensory consequences of an action even under some uncertainty as long as the action is voluntary and self-paced.

Non-specificity of the effects

Despite the evidence that the suppression effects are specific - at least to some extent -, there is also work showing that responses to sounds may be unspecifically gated during movement, thereby partly challenging the dominant cancellation models that attribute the suppression to stimulus-specific predictive mechanisms. For example, dampening of auditory responses has been found also for stimuli merely coinciding with finger movements (Hazemann et al., 1975; Horváth et al., 2012; Horváth, 2013a, b; Makeig et al., 1996; Tapia et al., 1987) and for unrelated auditory inputs during speech (Houde et al., 2002; Kudo et al., 2004) or locomotion (McGinley et al., 2015; Schneider et al., 2014). This unspecific suppression of auditory responses is reminiscent of the generalized attenuation found in other sensory modalities during movements (Crapse & Sommer, 2008; Williams et al., 1998), with the most known instances being the suppression of visual processing during saccades (Ross et al., 2001) and the dampened sensitivity on moving body parts (Chapman et al., 1987; Cohen & Starr, 1987; Williams et al., 1998; Williams & Chapman, 2000, 2002; see also Kilteni & Ehrsson, 2022 for a discussion on the distinction between somatosensory attenuation and tactile gating).

To directly examine the possibility of unspecific mechanisms contributing to the well-described self-generation effect, Horváth and colleagues (2012) employed a coincidence paradigm, where participants had to press a button several times and simultaneously, but independently from the actions, a sound sequence with varying between-sound intervals was presented. That is, contrary to the typical contingent self-generation paradigm, the action could not predict the identity, timing, or occurrence of the sound. Nevertheless, despite the absence of a contingent relationship between button press and sound, they obtained the N1 suppression effects for sounds merely coinciding with a motor act, thereby questioning the specificity of the suppression effects, and raising the possibility of the N1 suppression being partly driven by the temporal proximity between action and stimulus.

Further evidence challenging the specificity of the self-generation effects came from the study by SanMiguel and colleagues (2013) that raised the question of whether the N1-suppression effects reflect a genuine modulation of activity within the auditory cortex. It is known that the N1 is not a unitary phenomenon and early work has shown that the N1-suppression may reflect the overlap of several components (Näätänen & Picton, 1987), which cannot be captured by most studies that have examined N1 only at Cz or a group of frontocentral electrodes. Two of the components of N1 are proposed to be sensory-specific and to reflect processing in primary and secondary auditory areas (Näätänen & Picton, 1987). The first N1 subcomponent is generated by tangentially oriented sources in the auditory cortex, therefore inverting polarity on electrode locations below the Sylvian fissure (i.e., mastoids) when using a nose reference. Nevertheless, the few studies that have analyzed N1 amplitudes at the mastoid electrodes have reported no suppression (Timm et al., 2013) or even enhanced amplitude in response to self-generated sounds (Horváth et al., 2012), which inevitably questions the specificity of the effect. The second subcomponent, usually referred to as the "T complex", is generated by radial sources in the superior temporal gyrus and is typically identified as the first and second negative peaks (i.e., Na and Tb, respectively) on anterior temporal sites (Tonnquist-Uhlen et al., 2003; Wolpaw & Penry, 1975). The few studies that have analyzed this component have only shown attenuated Tb during self-generation, but Tb attenuation has been also suggested to be driven by mere temporal contiguity, rather than stimulus-specific predictions (Horváth et al., 2012; Horváth, 2013b; SanMiguel et al., 2013). Therefore, if the N1 attenuation was to be attributed to stimulus-specific motor predictions suppressing activity in auditory areas, it should mainly reflect the attenuation of the sensory components of N1, that is the ones generated in the auditory cortex. However, SanMiguel and colleagues (2013) showed that self-generation may mostly affect the stimulus-unspecific component of N1 which is thought to be the cortical projection of a reticular process that facilitates motor activity (Näätänen & Picton, 1987), presumably reflecting an attenuated orienting response rather than a prediction about the specific characteristics of the sound. This agrees with the coincidence studies by Horváth and colleagues (Horváth et al., 2012; Horváth, 2013a, 2013b) and highlights the need to differentiate the specific suppression of predicted sounds in the auditory cortex from unspecific gating mechanisms (see also Kilteni & Ehrsson, 2022 for a similar discussion in the somatosensory domain).

Given the evidence that auditory responses may be unspecifically gated during movement, one would inevitably raise the question of whether such gating is mediated by peripheral or central mechanisms. In fact, the possibility of peripheral mechanisms driving this unspecific gating during movements has been already raised by evidence showing that actions may cause the coactivation of the stapedius muscle with the task-relevant effector (Carmel & Starr, 1963; Salomon & Starr, 1963). This activation could lower the signal transmission efficiency in the middle ear, and consequently dampen responses to concurrently presented tones. However, this possibility has been ruled out since the stapedius muscle contraction is highly dependent on tone frequency (i.e., it applies only for frequencies less than 2kHz, Horváth & Burgyán, 2013), whereas the N1-suppression effect prevails independently of sound frequency. Therefore, the suppression effect during self-generation can be safely attributed to purely central mechanisms. Schneider and colleagues (2014) have indeed provided some evidence implicating a central mechanism as the driving force of the attenuation of auditory responses during movement. Using in vivo intracellular recordings in mice, they showed that excitatory neurons in the auditory cortex were suppressed by secondary motor cortex neurons innervating the auditory cortex during movement. Critically, movements reduced the responsiveness of the auditory cortex to sounds overall, suggesting that the suppression effects could be partly at least attributed to nonspecific predictions, that is, the organism expects some consequence of the motor act (since actions are rarely silent), but does not necessarily generate a prediction on the specific effect of the action.

In sum, the first chapters of the present thesis have presented evidence pointing to both highly specific mechanisms that depend on the match between prediction and input that attenuate sensory responses in sensory-specific areas, but also unspecific gating mechanisms during movement (i.e., akin to an unspecific halo of modulation) that affects processing of sounds presented close in time with a motor act even in the absence of a causal relationship between the action and the sensory stimulus (e.g., Hazemann et al., 1975; Makeig et al., 1996; Horváth, 2013a, b; Horváth et al., 2012).

1.1.5. Subcortical neuromodulation

An intriguing possibility is that rather than being exclusively driven by stimulus-specific predictions, the well-described action-induced suppression effects might result - at least partly - from unspecific processes: Actions might create a halo of neuromodulation around them which might dampen responses to sounds falling inside it. It has been proposed that such attenuation might reflect an attenuated orienting response towards self-generated stimulation (SanMiguel et al., 2013), independently of whether the sensory input can be predicted by the action or not. Whatever falls inside the halo is suppressed, allowing us to not be alerted by stimulation caused by ourselves. This idea remains largely unexplored and constitutes one of the core objectives of the present thesis. As we describe in this section, there is already some evidence supporting the existence of a halo of neuromodulation around actions.

There is already evidence that broader, unspecific processes during actions also modulate perceptual and neural responses (Press et al., 2020; Press & Cook, 2015; Schröger et al., 2015; Korka et al., 2021), suggesting that the attenuation effects could be driven by the fact that during actions the system expects some sensory consequences, without necessarily generating a specific prediction about their identity. Early work has proposed that the unspecific effects of actions on sensory processing could be driven by reticular mechanisms (Näätänen & Picton, 1987; Starr, 1964), and one could assume that such mechanisms are mediated by arousal-related neuromodulatory processes. The involvement of neuromodulatory systems in the motor-induced suppression effects seems plausible considering that actions are known to trigger a cascade of neuromodulatory processes (Vinck et al., 2015; Eggerman et al., 2014; McGinley et al., 2015; Aston-Jones & Cohen, 2005), and that during movement, the auditory cortex receives motor inputs, as well as neuromodulatory inputs by subcortical areas (Nelson & Mooney, 2016; Schneider & Mooney, 2018).

For example, cholinergic inputs from the basal forebrain innervate the auditory cortex (both excitatory and inhibitory neurons), often synapsing on the exact same auditory neurons that receive monosynaptic input from motor cortex (Nelson & Mooney, 2016). It is, though, worth noting that cholinergic and motor cortical afferents to the auditory cortex do not share the same activity profile (Nelson & Mooney, 2016). Specifically, the cholinergic inputs arise from the caudal part of the basal forebrain and receive input from brainstem regions implicated in movement and arousal, whereas motor cortical neurons that innervate auditory areas receive information mostly from forebrain regions, including those implicated in motor planning. This suggests that the auditory cortex integrates bottom-up arousal-related cholinergic signals during movements with top-down information concerning impending movements and motor planning, raising the possibility of simultaneous, but possibly independent, effects of motor and subcortical inputs on auditory processing, that could result in a diverse set of motor and neuromodulatory influences in the auditory cortex (Nelson & Mooney, 2016).

Despite the evidence pointing to the involvement of neuromodulatory process in the action-induced suppression effects, few attempts have been made to elucidate which mechanism could be - at least partly -

driving sensory attenuation. So far, the most likely reticular candidate to mediate the unspecific gating of stimulus processing around motor acts has been proposed to be the Locus Coeruleus Norepinephrine system.

The Locus Coeruleus Norepinephrine system

The Locus Coeruleus Norepinephrine (henceforth LC-NE) system is a small subcortical structure (i.e., only ~3000 neurons in rodents), located bilaterally in the brainstem under the cerebellum and lateral to the fourth ventricle, and is the main source of norepinephrine to the cortex (Poe et al., 2020; Aston-Jones & Cohen, 2005). The LC-NE sends wide projections of both excitatory and inhibitory nature to almost the entire brain (Berridge & Waterhouse, 2003; Foote et al., 1983), and sometimes its individual cells innervate multiple brain areas simultaneously (Sara, 2009; Loughlin et al., 1982). Given the broad projections and the convergent input on multiple areas, the LC-NE has been proposed as a critical regulator of a diverse set of cognitive processes, such as arousal, attention, sensory processing, synaptic plasticity, and memory consolidation (for reviews see Berridge & Waterhouse, 2003; Sara 2009; Poe et al., 2020; Aston-Jones & Cohen, 2005), with the first demonstration of LC-NE modulations on the sensory cortex of a behaving primate being dated already back in the 70's (Foote et al., 1975).

The effects of LC-NE on sensory processing have been further supported by animal studies showing a wide spectrum of effects of LC-NE activation on the magnitude of stimulus-evoked neural responses across thalamic and cortical, sensory-specific areas (Foote et al., 1975; Rogawski & Aghajanian, 1980; Manunta & Edeline, 2004; McCormick, 1989). For example, LC-NE stimulation in time with a sound presentation can result in norepinephrine-induced selective frequency tuning, effectively shifting the peak of the frequency tuning curve of auditory neurons to the frequency paired with LC-NE stimulation (Manunta & Edeline, 2004; Edeline et al., 2011), thereby boosting performance in a frequency discrimination task (i.e., when comparing tones with tuned frequencies and other tones in neighbouring frequencies; Martins & Froemke, 2015; Glennon et al., 2019), in line with the improvements reported in other detection and discrimination paradigms (Navarra et al., 2017; Rodenkirch et al., 2019). However, it should be noted that LC-NE effects on performance and sensory responses have been suggested to follow an inverted-U function (Berridge & Waterhouse, 2003; Poe et al., 2020; Aston-Jones & Cohen, 2005; McGinley et al., 2015), such that the facilitating effects reach their maximum as norepinephrine level increases from a relatively low level towards intermediate levels and then gradually decreases with further increases in LC-NE activity. Critically, LC-NE activity is known to correlate with pupil diameter (Aston-Jones & Cohen, 2005; Vinck et al., 2015; Murphy et al., 2014, Joshi et al., 2016), raising the possibility to address the contributions of LC-NE in sensory processing using the non-invasive method of pupillometry.

Along with this evidence implicating LC-NE in sensory processing, further work supports the idea of the LC-NE as the mechanism driving the neuromodulation halo around actions. This evidence comes from work showing a close association between pupil diameter – as a proxy of LC-NE activity – (Aston-Jones & Cohen, 2005; Vinck et al., 2015; Murphy et al., 2014, Joshi et al., 2016) and actions (e.g., whisking or button press; Lee & Margolis, 2016; McGinley et al., 2015; Vinck et al., 2015; Yebra et al., 2019; Bornert & Bouret, 2021; Lubinus et al., 2021). Such halo could act as a temporal orienting filter that produces temporally specific but spatially widespread modulations of cortical responsivity to stimuli during actions (Aston-Jones & Cohen, 2005), allowing us to either ignore predictable and irrelevant, albeit unavoidable, sensory consequences of our actions, or pay particular attention to the consequences of goal-driven intentional acts. However, to date, there have been no attempts to test for possible links between sensory attenuation and neuromodulation (i.e., as reflected in pupil diameter).

1.2. Action effects on memory processing

Meanwhile, although most studies have focused on the effects of self-generation on the immediate sensory processing, previous work has reported modulatory effects of movements on hippocampal and parahippocampal activity (Halgren, 1991; Mukamel et al., 2010; Rummell et al., 2016), raising the possibility that the differential processing of self- and externally-generated stimulation may have consequences for memory encoding as well.

1.2.1. Production effect

One line of evidence shows that self-initiation has a beneficial impact on memory encoding, which is known as the production effect. The production effect has been typically found in studies using words (MacDonald & MacLeod, 1998; MacLeod et al., 2010; Ekstrand et al., 1966; Hopkins & Edwards, 1972; Conway & Gathercole, 1987; Gathercole & Conway, 1988) or melodies (Brown & Palmer, 2012; Mathias et al., 2015) and have shown that memory performance is greater for words read aloud than for silently read ones, and that melodies rehearsed actively with motor-auditory feedback are better remembered than melodies that were listened to passively. In line with this beneficial effect of actions on memory, there is also work with musicians showing enhanced accuracy in detecting changes in pitch in previously produced melodies compared to passively heard ones (Mathias et al., 2015). Critically, Mathias and colleagues (2015) showed that pitch alterations amplified the electrophysiological potentials arising from cortical motor structures, and that this enhancement was associated with more accurate recognition of alterations in the produced melodies in comparison with the non-produced ones. These memory improvements for self-produced stimuli have been attributed to the increased distinctiveness of those items because producing them provides extra mnemonic information (e.g., motor movements associated with word production and the auditory feedback of our own voice) that is not present for silently read words (Conway & Gathercole, 1987; Mama & Icht, 2016; Ozubko et al., 2012). In line with, and complementing, the production effect studies, recent work suggests that the action-induced memory enhancement may be driven by the engagement of the noradrenergic system, as shown by the increased pupil dilation and locus coeruleus activity in response to stimuli tied with – but not produced by – actions (i.e., Go-events in a Go/No-Go task; Yebra et al., 2019).

1.2.2. Predictive models of memory

On the other hand, there is another line of work under the predictive coding framework which - indirectly - makes opposite predictions about the effect of self-initiation on memory encoding. Predictive coding has the concept of prediction error at its core, and inevitably the assumptions of this framework as to how we learn and remember events is largely based on the prediction error elicited by an item. Under this model, learning and memory are driven by the amount of surprise associated with an item, that is, prediction error is the critical factor determining what and how much is learned (Bar, 2009; Krawczyk et al., 2017; Pine et al., 2018). This means that items that elicit large prediction errors (i.e., reflected in evoked potentials in M/EEG research or BOLD signal in functional neuroimaging studies) are remembered better than items with reduced prediction errors (Greve et al., 2017; Heilbron & Chait, 2018; Henson & Gagnepain, 2010; Pine et al., 2018). This translates to the idea that unpredictable items (e.g., externally-generated stimulation) should be better encoded in memory than predictable stimulation (e.g., self-generated, reafferent input). This idea has found support in studies showing that items leading to high prediction error tend to produce greater hippocampus fMRI signal in the study (i.e., encoding) phase of recognition memory paradigms, and that this increase in hippocampal activity is linked to better recollection in the subsequent test phase (Gagnepain et al., 2011; Henson & Gagnepain, 2010; Pine et al., 2018). Therefore, applying the assumptions of this framework to the domain of self-generation, one would expect memory enhancements for the unpredictable, externally-generated sounds, but only in typical contingent paradigms where they would inherently elicit larger prediction errors compared to the more predictable self-generated stimuli.

CHAPTER 2: Objectives and hypotheses

The present thesis aimed at exploring the effects of actions on auditory processing from the angles of basic physiology (i.e., electrophysiological responses and neuromodulatory processes), perceptual processing (i.e., sound detection and loudness perception), and high-level cognition (i.e., memory encoding). Specifically, the aim of the present thesis was twofold, namely to assess the effects of actions on 1) perceptual processes, including sensory responses and 2) memory. Therefore, in all three studies we employed self- and externally-generated sounds.

Related to the first aim (i.e., action-effects on perceptual processes and sensory responses), the present thesis aimed to assess the contribution of other factors (i.e., sound intensity, predictability, and action-sound delay) to the typical self-generation effects on perception and sensory processing. These factors have been shown to modulate the direction and the magnitude of the self-generation effects (e.g., Reznik & Mukamel, 2019; Press et al., 2020; Hughes et al., 2013a; Timm et al., 2016), but to date, none of the extant models has accounted for their contribution in the perceptual and sensory processing for self-generated stimulation. Additionally, disentangling the contribution of these factors would allow us to examine the specificity of the sensory attenuation effects (i.e., stimulus-specific prediction mechanisms or unspecific effects during movement), and clarify the role of unspecific subcortical neuromodulatory processes during actions. The second main aim of the present thesis was to elucidate the possible neurophysiological mechanisms behind the effects of actions on memory and assess whether memory encoding for self-generated stimulation is related to the action-induced suppression effects and subcortical neuromodulation. The specific aims of each study are the following:

Study I

Study I aimed to elucidate the modulatory effects of intensity on the perceptual processing of self-generated sounds, while assessing whether the expected effects drive changes in sensitivity and/or perceptual bias. Previous work has already raised the possibility of a critical dependence of perceptual processing of self-generated sounds on stimulus intensity, suggesting that perception is attenuated when self-generated sounds are presented at supra-threshold intensities, but perceptual enhanced when sounds are presented at near-threshold intensities (e.g., Reznik et al., 2015; Reznik & Mukamel, 2019). Interim evidence supporting such interaction has been provided only by one study (Reznik et al., 2015), which nonetheless did not address which aspect of perception is modulated (sensitivity and/or bias). To this end, we employed a behavioural study consisting of a sound detection and a loudness discrimination task and compared the detection and discrimination sensitivity, as well as the possible bias in perceived loudness for self- vs. externally-generated sounds at both supra- and near-threshold intensities. Based on previous work (for a review see Reznik & Mukamel, 2019), we expected to observe higher detection sensitivity for low-intensity self-generated sounds, and interactive effects of self-generation and intensity on perceived loudness (i.e., enhanced perceived loudness for soft, self-generated sounds, but attenuated perceived loudness for the same sounds when presented at supra-threshold intensities).

Study II

Study II is the first – to our knowledge – aiming to simultaneously assess the self-generation effects on both sensory processing and memory in a non-contingent paradigm. Specifically, we aimed to replicate the sensory attenuation effects in a paradigm where the action is not fully predictive of the sound (Hórvath et al., 2012), while clarifying the contribution of subcortical neuromodulation in the action-induced attenuation effects. Meanwhile, Study II also aimed to assess the effects of actions on memory encoding of concurrent, but unpredictable, sounds. More importantly, we examined whether the action-induced effects on sensory processing relate to the memory encoding of self-generated events and tested for possible links between the attenuation of sensory responses and the subcortical neuromodulation during actions. To this end, we employed a combination of a self-generation and memory task, concurrently with EEG and pupillometry recordings. We expected to observe the typical suppression effects on auditory evoked potentials for sounds that could not be fully predicted by the action (cf. Horváth et al., 2012). Additionally, neuromodulatory activity (i.e., reflected in pupil diameter; Aston-Jones & Cohen, 2005) was expected to increase during actions (cf. McGinley et al., 2015) and we hypothesised that action-induced pupil dilation would correlate with the attenuation effects. Further, we expected to observe differences in memory performance between passively encoded sounds compared to sounds that coincided with an action but given the mixed evidence (enhanced vs. reduced memory for self-initiated stimulation, as proposed by the production effect and predictive coding, respectively, e.g., MacDonald & MacLeod, 1998; Henson & Gagnepain, 2010), we did not have clear hypotheses as to the direction of the effects. Finally, the potential differences in the memory encoding of sounds presented with or without a concomitant action were expected to be driven by, and therefore correlate with the differential neurophysiological responses (i.e., event-related potentials and pupil diameter) for sounds that either coincided with an action or were presented passively.

Study III

Following the research line of Study II, Study III aimed to assess the self-generation effects under fully predictable and unpredictable action-sound relationships. The aims of this study were largely inspired by the need to disentangle the effects of predictability and self-generation that have been conflated in previous work (Hughes et al., 2013a). To this end, we employed a combination of self-generation and memory task, concurrently with EEG and pupillometry recordings, and orthogonally controlled predictability and self-generation. With this design, Study III sought out to disentangle the contributions of self-generation and predictability to the sensory attenuation effects, replicate the engagement of neuromodulatory processes during actions, and assess whether neuromodulation and sensory attenuation are related. Additionally, Study III examined the effects of actions and predictability on memory encoding and finally, it explored whether the potential behavioural effects of actions are related to the self-generation effects and/or the neuromodulatory processes during movements. We expected sensory attenuation for self-generated sounds, with strongest effects when the action is fully predictive of the sound. Additionally, pupil diameter was expected to be largest when participants perform button presses, but no specific hypothesis was formulated as to whether this effect would interact with predictability. Further, memory performance was expected to differ between self- and externally-generated sounds but given the mixed evidence we remained agnostic as to the direction of the effects, but also as to the contributions of predictability on memory encoding.

CHAPTER 3: General methodology

3.1. Participants

The participants in all studies were typically undergraduate university students at the University of Barcelona. All participants had normal hearing, did not report any psychiatric or neurological illness, and did not consume drugs or pharmaceuticals acting on the central nervous system. All participants gave written informed consent for their participation after the nature of the study was explained to them and they were monetarily compensated (10€/hour). Additional materials included a personal data questionnaire and a data protection document (in all studies) and five personality trait questionnaires (in Study II and Study III). The studies were approved by the Bioethics Committee of the University of Barcelona. Across the three studies, the overall sample size was 71 participants (49 women) with mean age of 24.12 years, and age range between 18 and 43 years old.

3.2. Experimental design and stimuli

Study I consisted of two two-alternative forced-choice (2AFC) tasks: a detection and a discrimination task. In both tasks, we employed self- and externally-generated sounds (factor: Source), while in the discrimination task we employed supra- and near-threshold intensities (factor: Intensity). The auditory stimuli consisted of pure tones of 1 kHz, 300 ms duration, and sampling rate of 44.1 kHz.

Study II employed a combination of a self-generation and memory recognition task where participants had to encode and later recall self- or externally-generated sounds. The auditory stimuli consisted of non-identifiable sounds that were edited to all have 250 ms of duration, a sampling rate of 44.1 kHz, and with 70 dB of intensity.

Study III employed a combination of a self-generation and memory task where participants had to encode and later recall a series of sounds. The sounds were either self- or externally-generated (factor: Source), and either predictable or unpredictable (factor: Predictability), and were presented shortly after or farther away from the button press or a visual cue, respectively (factor: Delay). Auditory stimuli consisted of identifiable sounds from four categories (animals, musical instruments, environmental sounds, and human-related sounds) that were edited to all have 500 ms of duration, a sampling rate of 44.1 kHz, and with 75 dB of intensity.

3.3. Apparatus

In all studies, the visual stimuli were presented on an ATI Radeon HD 2400 monitor. The auditory stimuli were presented via Sennheiser KD 380 PRO noise cancelling headphones. To record participants' button presses and behavioural responses, we used a Korg nanoPAD2. The buttons of this device do not produce any mechanical noise when pressed, and, thus, do not interfere with our auditory stimuli. The presentation of the stimuli and recording of participants' button presses and responses were controlled using MATLAB R2017a (The Mathworks Inc., 2017), the Psychophysics Toolbox extension (Brainard, 1997; Kleiner et al., 2007).

EEG recordings were performed only in Study II and Study III. EEG activity was acquired with Neuroscan SynAmps RT amplifier (NeuroScan, Compumedics, Charlotte, NC, USA) and the software used for the recordings was Neuroscan 4.4 software in Study II and CURRY 8 Neuroscan in Study III. We recorded continuously with Ag/AgCl electrodes from 64 standard locations according to the 10% extension of the International 10–20 system (Chatrian et al., 1985; Oostenveld & Praamstra, 2001) mounted in a nylon cap (Quick-Cap; Compumedics, Charlotte, NC, USA). An additional electrode was placed at the tip of the nose (serving as online reference). The vertical electrooculogram (EOG) was measured with two electrodes placed above and below the left eye, and the horizontal EOG with two electrodes placed on the outer canthi of the eyes referenced to the common reference (unipolar montage). The ground electrode was placed at AFz. All impedances were kept below 10 kΩ during the whole recording session and data was sampled at 500 Hz.

Pupil recordings were performed only in Study II and Study III, concurrently with the EEG recordings. Horizontal and vertical gaze position, and the pupil area, were recorded using EyeLink 1000 desktop mount (SR Research, sampling rate: 1,000 Hz). Visual stimulation was presented at the center of the screen (i.e., within central vision) to eliminate any effects of gaze position on pupil diameter (Gagl et al., 2011). The pupil was assessed in the centroid mode of the eye tracker, using a center-of-mass algorithm. This algorithm detects the pupil area by identifying the number of black pixels and its center on the video image. Importantly, in contrast to methods using ellipse fitting for the measurement of the pupil, this method is not affected by noise (S-R Research Eyelink-CL Manual, p. 71).

3.4. Data analysis

In all studies, statistical analyses for the behavioural and ERP data were performed using R (version 3.6.0). For all the *t*-tests performed, we first confirmed that the assumption of normality was not violated (Shapiro–Wilk normality test $p > .05$). Post-hoc comparisons following significant interactions in ANOVAs were performed using the Bonferroni correction. Cluster-based permutation statistics were used for the pupillometric data, but also for the EEG data as an additional, exploratory analysis. Post-hoc comparisons for significant interactions in the cluster-based permutation analyses were performed within the significant window of the interaction. Finally, all the correlation analyses were conducted using the Pearson correlation coefficient. Correlation analyses in Study III was complemented by Bayesian correlations (Jeffreys, 1961). A summary of the factors, aims, and measures for each study is presented in Table 1.

3.4.1. Behavioural measures

Behavioural measures were obtained in all three studies. Perceptual measures were obtained only in Study I (both of sensitivity and bias), while Studies II and III only obtained memory measures both of bias (Study II) and performance (Studies II and III). All three studies contrasted behavioural measures between self- and externally-generated sounds, and assessed interactions with the factors studied in each study (e.g., sound intensity, predictability, and action-sound delay), using *t*-tests and/or ANOVAs. Only for Study I, some of the measures introduced in the statistical analyses (e.g., detection thresholds, Point of Subjective Equality, Just Noticeable Difference) were obtained by a psychometric function fitting procedure using the quickpsy package (Linares & López-Moliner, 2016).

3.4.2. EEG preprocessing and ERP analysis

EEG data in Study II and Study III was analyzed with EEGLAB (Delorme & Makeig, 2004) and plotted with EEProbe (ANT Neuro). Data were high-pass filtered (0.5 Hz high-pass, Kaiser window, Kaiser β 5.653, filter order 1812), manually inspected to reject atypical artifacts and identify malfunctioning electrodes, and corrected for eye movements with Independent Component Analysis, using the compiled version of runica (binica) that uses the logistic infomax ICA algorithm (Onton & Makeig, 2006). After visual inspection, we rejected components capturing eye movement artifacts and the remaining components were projected back into electrode space. Data was then low-pass filtered (30 Hz low-pass, Kaiser window, Kaiser β 5.653, filter order 1812), remaining artifacts were rejected by applying a 75 µV maximal signal-change per epoch threshold, and malfunctioning electrodes were spherically interpolated. Data was then epoched for each event of interest, baseline corrected (only in Study II), and corrected for motor and visual activity (see Study II and III for details). In each study, we calculated the average wave for each event of interest for each subject.

Subsequently, we analyzed the following auditory-event related potentials: N1 and P2 at Cz (N1, P2) and at the mastoids (henceforth, $N1_{mast}$, $P2_{mast}$), the P3 component at Pz, and the N1 subcomponents Na and Tb at the electrodes T7 and T8. The windows were defined after visual inspection of the data by locating the highest negative or positive (depending on the component of interest) peak in the usual latencies for each component as reported by previous work (SanMiguel et al., 2013). Specifically, time windows for N1 (and $N1_{mast}$), P2 (and $P2_{mast}$), Na, and Tb were defined on the grand-averaged waveforms of the externally-generated sounds (cf. SanMiguel et al., 2013). Na and Tb were identified as the first and second negative peaks, respectively, identifiable after sound onset on electrodes T7 and T8, as recommended by Tonnquist-Uhlen et al. (2003). N1/$N1_{mast}$ and P2/$P2_{mast}$ were identified as the negative and positive peaks occurring in the window ~70 to 150 ms, and ~150 to 250 ms after stimulus onset on Cz, respectively, showing reversed polarity at the mastoid electrodes. P3 (only for Study II) was identified as the peak of the difference wave (externally-generated – motor-corrected self-generated waveforms) in the P3 window range based on previous work (e.g., Baess et al., 2008). Given variations in peak latencies across the conditions in each study, the width of the windows was defined such that it could capture the peak both waveforms (externally- and self-generated sound responses), and it was proportional to the width of the component. Statistical analyses were performed using paired samples *t*-tests and/or within-subjects ANOVAs. For components identified in two electrodes (e.g., $N1_{mast}$, $P2_{mast}$, Na, and Tb), an additional factor of Laterality was introduced.

3.4.3. Pupil preprocessing

In both Study II and Study III, the preprocessing of the pupil data was the following. Missing data and blinks, as detected by the EyeLink software, were padded by 100 ms and linearly interpolated. Additional blinks were identified using peak detection on the velocity of the pupil signal and linearly interpolated (Urai et al., 2017). Blinks separated by less than 250 ms were aggregated to a single blink. The interpolated pupil data was bandpass filtered (0.05–4 Hz third-order Butterworth filter). We also performed a deconvolution analysis to estimate the effect of blinks and saccades on the pupil response and removed these responses from the data using linear regression using a procedure detailed in previous work (Knapen et al., 2016; Urai et al., 2017). The residual bandpass filtered pupil data was used for the evoked analyses (van Slooten et al., 2019). After zscoring per trial, we epoched the data, baseline corrected each trial by subtracting the mean pupil diameter 500 ms before onset of the event and resampled to 100 Hz. Subsequently, for each participant, the average evoked response for the main events of interest was obtained. No motor correction was performed in the pupillometry analysis since we were interested in assessing the effects of motor acts on pupil response. The pupil preprocessing and analysis was performed with custom software based on previous work (Urai et al., 2017) using Fieldtrip (Oostenveld et al., 2011).

3.4.4. Cluster-based permutation tests

Cluster-based permutation statistics were performed for the EEG and pupillometric data in Studies II and III. For the EEG data, we decided to perform these additional data-driven analysis to explore the possibility of further effects that would not have been captured with traditional ERP analyses. For the pupillometric data, cluster-based permutation tests constituted the main analysis to assess the effects of interest, based on previous work using the same approach (Urai et al., 2017). Specifically, we used cluster-based non-parametric correction to account for multiple comparisons and determined whether there were statistical differences between the contrasting conditions within participants. Each cluster was constituted by the samples that consecutively passed a specified threshold (in this case sample *p*-value of 0.05). The cluster statistics was chosen as the sum of the paired *t*-values of all the samples in the cluster. For each statistical test, this procedure was performed by randomly switching labels of individual observations between these paired sets of values. We repeated this procedure 10,000 times and computed the difference between the group means on each permutation. The obtained *p*-value was the fraction of permutations that exceeded the observed difference between the means (i.e., two-sided dependent samples tests). This analysis was done using Fieldtrip (Oostenveld et al., 2011).

Table 1

Summary of the factors, aims, and measures for each one of the three studies compiling the present thesis.

Studies	**Factors**	**Aims and measures**			
		Perception & Memory		**Sensory processing**	
		Perception	*Memory*	*Sensory response modulation*	*Subcortical neuromodulation*
Study I	Source	Sound detection	-	-	-
	Source Intensity	Loudness discrimination			
Study II	Source	-	Memory bias and performance	Auditory evoked responses	Pupil diameter
Study III	Source Predictability Delay				

CHAPTER 4: STUDIES

In this chapter, we present the three original studies compiling the present thesis.

Study I has been already published in *Scientific Reports*.

- Paraskevoudi, N., & SanMiguel, I. (2021). Self-generation and sound intensity interactively modulate perceptual bias, but not perceptual sensitivity. *Scientific Reports***,** 11**,** 17103. https://doi.org/10.1038/s41598-021-96346-z

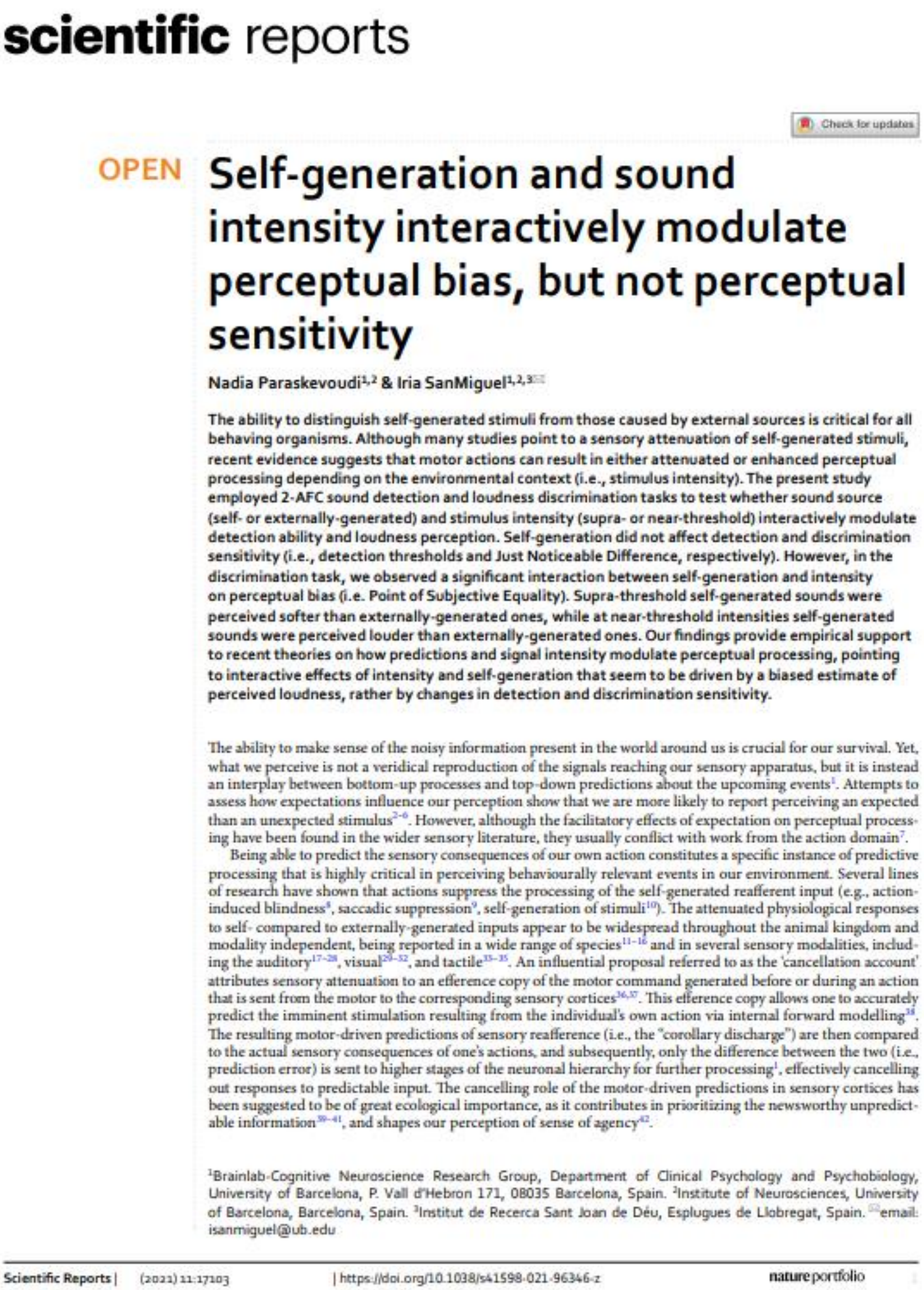
Study II has been already published as preprint in the bioRxiv platform and it is currently under review in *Psychophysiology* (submitted on December 14th, 2021).

- Paraskevoudi, N., & SanMiguel, I. (2021). Sensory suppression and increased neuromodulation during actions disrupt memory encoding of unpredictable self-initiated stimuli [Preprint]. bioRxiv, Neuroscience. **Under review in** *Psychophysiology.* https://doi.org/10.1101/2021.12.15.472750

Study III is expected to be submitted for publication in the upcoming months.

- Paraskevoudi, N., & SanMiguel, I. (in preparation). Semi self-paced actions enhance memory of sounds but interact with predictability in shaping auditory and neuromodulatory responses.

STUDY I



scientific reports

Check for updates

OPEN

Self-generation and sound intensity interactively modulate perceptual bias, but not perceptual sensitivity

Nadia Paraskevoudi[1,2] & Iria SanMiguel[1,2,3]

The ability to distinguish self-generated stimuli from those caused by external sources is critical for all behaving organisms. Although many studies point to a sensory attenuation of self-generated stimuli, recent evidence suggests that motor actions can result in either attenuated or enhanced perceptual processing depending on the environmental context (i.e., stimulus intensity). The present study employed 2-AFC sound detection and loudness discrimination tasks to test whether sound source (self- or externally-generated) and stimulus intensity (supra- or near-threshold) interactively modulate detection ability and loudness perception. Self-generation did not affect detection and discrimination sensitivity (i.e., detection thresholds and Just Noticeable Difference, respectively). However, in the discrimination task, we observed a significant interaction between self-generation and intensity on perceptual bias (i.e. Point of Subjective Equality). Supra-threshold self-generated sounds were perceived softer than externally-generated ones, while at near-threshold intensities self-generated sounds were perceived louder than externally-generated ones. Our findings provide empirical support to recent theories on how predictions and signal intensity modulate perceptual processing, pointing to interactive effects of intensity and self-generation that seem to be driven by a biased estimate of perceived loudness, rather by changes in detection and discrimination sensitivity.

The ability to make sense of the noisy information present in the world around us is crucial for our survival. Yet, what we perceive is not a veridical reproduction of the signals reaching our sensory apparatus, but it is instead an interplay between bottom-up processes and top-down predictions about the upcoming events[1]. Attempts to assess how expectations influence our perception show that we are more likely to report perceiving an expected than an unexpected stimulus[2–6]. However, although the facilitatory effects of expectation on perceptual processing have been found in the wider sensory literature, they usually conflict with work from the action domain[7].

Being able to predict the sensory consequences of our own action constitutes a specific instance of predictive processing that is highly critical in perceiving behaviourally relevant events in our environment. Several lines of research have shown that actions suppress the processing of the self-generated reafferent input (e.g., action-induced blindness[8], saccadic suppression[9], self-generation of stimuli[10]). The attenuated physiological responses to self- compared to externally-generated inputs appear to be widespread throughout the animal kingdom and modality independent, being reported in a wide range of species[11–16] and in several sensory modalities, including the auditory[17–28], visual[29–32], and tactile[33–35]. An influential proposal referred to as the 'cancellation account' attributes sensory attenuation to an efference copy of the motor command generated before or during an action that is sent from the motor to the corresponding sensory cortices[36,37]. This efference copy allows one to accurately predict the imminent stimulation resulting from the individual's own action via internal forward modelling[38]. The resulting motor-driven predictions of sensory reafference (i.e., the "corollary discharge") are then compared to the actual sensory consequences of one's actions, and subsequently, only the difference between the two (i.e., prediction error) is sent to higher stages of the neuronal hierarchy for further processing[1], effectively cancelling out responses to predictable input. The cancelling role of the motor-driven predictions in sensory cortices has been suggested to be of great ecological importance, as it contributes in prioritizing the newsworthy unpredictable information[39–41], and shapes our perception of sense of agency[42].

[1]Brainlab-Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, University of Barcelona, P. Vall d'Hebron 171, 08035 Barcelona, Spain. [2]Institute of Neurosciences, University of Barcelona, Barcelona, Spain. [3]Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain. ✉email: isanmiguel@ub.edu

Scientific Reports | (2021) 11:17103 | https://doi.org/10.1038/s41598-021-96346-z nature portfolio

Abstract

The ability to distinguish self-generated stimuli from those caused by external sources is critical for all behaving organisms. Although many studies point to a sensory attenuation of self-generated stimuli, recent evidence suggests that motor actions can result in either attenuated or enhanced perceptual processing depending on the environmental context (i.e., stimulus intensity). The present study employed 2-AFC sound detection and loudness discrimination tasks to test whether sound source (self- or externally-generated) and stimulus intensity (supra- or near-threshold) interactively modulate detection ability and loudness perception. Self-generation did not affect detection and discrimination sensitivity (i.e., detection thresholds and Just Noticeable Difference, respectively). However, in the discrimination task, we observed a significant interaction between self-generation and intensity on perceptual bias (i.e. Point of Subjective Equality). Supra-threshold self-generated sounds were perceived softer than externally-generated ones, while at near-threshold intensities self-generated sounds were perceived louder than externally-generated ones. Our findings provide empirical support to recent theories on how predictions and signal intensity modulate perceptual processing, pointing to interactive effects of intensity and self-generation that seem to be driven by a biased estimate of perceived loudness, rather by changes in detection and discrimination sensitivity.

1. Introduction

The ability to make sense of the noisy information present in the world around us is crucial for our survival. Yet, what we perceive is not a veridical reproduction of the signals reaching our sensory apparatus, but it is instead an interplay between bottom-up processes and top-down predictions about the upcoming events (Friston, 2005). Attempts to assess how expectations influence our perception show that we are more likely to report perceiving an expected than an unexpected stimulus (Chalk et al., 2010; Jaramillo & Zador, 2011; Pinto et al., 2015; Stein & Peelen, 2015; Wyart et al., 2012). However, although the facilitatory effects of expectation on perceptual processing have been found in the wider sensory literature, they usually conflict with work from the action domain (Press et al., 2020).

Being able to predict the sensory consequences of our own action constitutes a specific instance of predictive processing that is highly critical in perceiving behaviourally relevant events in our environment. Several lines of research have shown that actions suppress the processing of the self-generated reafferent input (e.g., action-induced blindness, saccadic suppression, self-generation of stimuli; Kunde & Wühr, 2004; Ross et al., 2001; Straka et al., 2018). The attenuated physiological responses to self- compared to externally-generated inputs appear to be widespread throughout the animal kingdom and modality independent, being reported in a wide range of species (Chagnaud et al., 2015; Kelley & Bass, 2010; Kim et al., 2015; Requarth & Sawtell, 2011; Roy & Cullen, 2001; Schneider et al., 2014) and in several sensory modalities, including the auditory (Baess et al., 2011; Horváth, 2013a, 2013b; Martikainen et al., 2004; Mifsud & Whitford, 2017; SanMiguel et al., 2013; Saupe et al., 2013; Schafer & Marcus, 1973; Timm et al., 2013; Klaffehn et al., 2019; Weller et al., 2017; Pyasik et al., 2018), visual (Hughes & Waszak, 2011; Mifsud et al., 2018; Roussel et al., 2013, 2014), and tactile (Blakemore, Wolpert et al., 1998; Hesse et al., 2010; Kilteni et al., 2020). An influential proposal referred to as the 'cancellation account' attributes sensory attenuation to an efference copy of the motor command generated before or during an action that is sent from the motor to the corresponding sensory cortices (Sperry, 1950; von Holst, 1954). This efference copy allows one to accurately predict the imminent stimulation resulting from the individual's own action via internal forward modelling (Wolpert et al., 1995). The resulting motor-driven predictions of sensory reafference (i.e., the "corollary discharge") are then compared to the actual sensory consequences of one's actions, and subsequently, only the difference between the two (i.e., prediction error) is sent to higher stages of the neuronal hierarchy for further processing (Friston, 2005), effectively cancelling out responses to predictable input. The cancelling role of the motor-driven predictions in sensory cortices has been suggested to be of great ecological importance, as it contributes in prioritizing the newsworthy unpredictable information (Blakemore et al., 2000; Poulet & Hedwig, 2002; Barron et al., 2020), and shapes our perception of sense of agency (Gallagher, 2000).

However, in the animal kingdom corollary discharge has been found to influence sensory processing in myriad ways besides cancellation of reafference (Crapse & Sommer, 2008). Contrary to cancellation theories, recent sharpening models propose that perception is biased towards the expected input (Yon & Press, 2017; Yon et al., 2019) in line with evidence showing enhanced BOLD responses to self-generated stimuli (Reznik et al., 2014; Simões-Franklin et al., 2011) and increased discharges in some neurons during self-initiated vocalizations (Eliades & Wang, 2003). The discrepancy between cancellation and sharpening accounts is also reflected in human studies attempting to assess the behavioural correlates of the neurophysiological effects of self-generation on stimulus processing. While self-initiated action effects have been typically found to be perceived as less ticklish (Blakemore, Wolpert et al., 1998; Claxton, 1975; Weiskrantz et al., 1971), less forceful (Kilteni et al., 2020; Bays et al., 2005), or less loud (Sato, 2008; Weiss et al., 2011a, 2011b) than equivalent stimuli initiated by another person or by a computer, recent findings show enhanced perception for action-expected outcomes (Reznik et al., 2014; Desantis et al., 2016). Collectively, the discrepancy in the results reported so far points to factors other than self-generation that may interactively modulate sensory processing during motor actions.

In a closer look, the mixed findings reported so far as concerns the neurophysiological and behavioural effects of motor predictions on sensory processing may be due to critical differences in the experimental paradigm, stimulus features, and obtained measures (see Table 1). On the one hand, animal studies with perceptual measures have reported both attenuation (McGinley et al., 2015; Neske et al., 2019) and enhancement (Carcea et al., 2017), but assess perceptual processing during locomotion compared to quiescence (McGinley et al., 2015; Neske et al., 2019; Bennett et al., 2013) or in Go-No/Go tasks (Carcea et al., 2017). However, sensory processing during action may differ from processing of stimuli resulting from action as assessed in contingent paradigms with humans that typically compare action-predicted vs. unpredictable stimuli (i.e., self- vs. externally-generated; Kilteni et al., 2020; Sato, 2008; Weiss et al., 2011a, 2011b) or predicted vs. mispredicted stimuli (action-congruent vs. action-incongruent; Yon & Press, 2017), thus rendering it difficult to disentangle whether the observed effects are driven by specific motor-driven predictions or by unspecific arousal mechanisms (McGinley et al., 2015). Additionally, studies also differ in the task and stimulus intensities that they employ. Human studies reporting suppression typically use supra-threshold stimuli in discrimination paradigms and show modulations in perceptual bias (Point of Subjective Equality; PSE) rather than sensitivity measures (Just Noticeable Difference; JND; Kilteni et al., 2020; Sato, 2008; Weiss et al., 2011a, 2011b). In contrast, evidence supporting sharpening accounts has been reported mostly in detection paradigms that obligatorily need to use near-threshold stimuli (Yon & Press, 2017; Reznik et al., 2014; Desantis et al., 2016; Cao & Gross, 2015a). This line of work has reported changes in sensitivity in both directions (Reznik et al., 2014; Cao & Gross, 2015a; Cardoso-Leite et al., 2010, but see Schwarz et al., 2018 and Cao & Gross, 2015b for no effects), but also in decision processes (Desantis et al., 2016). Collectively, these findings raise the possibility that the conflicting findings on the nature of the effects of action on the perceptual processing of self-initiated stimuli may depend on a handful of specific factors (i.e., action/no action comparisons vs. action-predicted/action-unpredicted comparisons; stimulus intensity) that may selectively affect certain aspects of perception (i.e., detection or discrimination ability; sensitivity or bias).

Recent work has indeed provided some evidence showing that sensory attenuation may be dependent on the stimulus intensity (Burin et al., 2017; Reznik et al., 2015; but see Majchrowicz & Wierzchoń, 2021). Reznik and colleagues (2015) had participants judge the perceived intensity of self- and externally-generated sounds presented at a supra- or a near-threshold intensity. Unbeknownst to the participants, the two sounds were always presented at the exact same intensity, but they were asked to report which one of them was louder. Their results showed a significant interaction between intensity and sound source. While the supra-threshold self-generated sounds were perceived as less loud than the passive comparisons, the opposite effect was obtained for near-threshold intensities. That is, when the sensory consequences of participants' movements were of low intensity, a significant sensory enhancement was observed, with the self-generated tones being judged as louder than the comparison passive tones. However, due to the experimental design of this study (i.e., no varying comparison intensities), no psychophysical measures (e.g., PSE, JND) could be obtained to further examine whether the modulatory effects of intensity on perceptual processing for self-initiated sounds are driven by changes in bias or sensitivity, respectively.

Table 1

Human studies assessing the behavioural effects of self-generation on auditory processing.

Self-generation effects	Study	Task	Intensity	Bias / sensitivity
Attenuation	Sato, 2008; Weiss et al., 2011a, 2011b	Loudness discrimination	L	Bias (PSE)
	Reznik et al., 2015			Bias (% 1st sound louder)
	Cao & Gross, 2015a	Detection of attended frequencies	NT	Sensitivity (d’)
Enhancement	Reznik et al., 2015	Loudness discrimination	NT	Bias (% 1st sound louder)
	Reznik et al., 2014	Detection	NT	Sensitivity (d’, thresholds)
	Myers et al., 2020	Loudness discrimination	L	Sensitivity (% correct)
No effect	Sato, 2008; Weiss et al., 2011a, 2011b	Loudness discrimination	L	Sensitivity (JND)
	Myers et al., 2020	Detection	NT	Sensitivity (thresholds)
	Cao & Gross, 2015b	Loudness discrimination	L	Bias (PSE)
	Cao & Gross, 2015a	Detection of nonattended frequencies	NT	Sensitivity (d’)

Note. Studies have reported either attenuation, enhancement, or no effects in detection or discrimination tasks with either loud (L) or near-threshold (NT) sounds by obtaining various measures that are used as a proxy of either bias or sensitivity (Point of Subject Equality, PSE; Just Noticeable Difference, JND; d’, d-prime).

Taken together, the evidence reported so far suggests that the direction of self-generation effects may be dependent on the intensity and therefore the amount of sensory noise in the signal. Indeed, recent work has highlighted the role of sensory noise in driving perceptual processing, suggesting that enhanced sensory processing for unexpected events is dependent on the ‘newsworthiness’ of the signal, such that the less the sensory noise (i.e., high intensities), the higher the sensory precision of the signal, and thus the more informative the unexpected stimulus (Press et al., 2020; Barron et al., 2021). Yet, we reason that the findings obtained from the previous self-generation studies cannot provide solid conclusions on this matter, due to the use of a small range of intensities (either supra-threshold only; Sato, 2008; Weiss et al., 2011a, 2011b, near-threshold only; Reznik et al., 2014, or only one of each; Reznik et al., 2015). More importantly, the inconsistency between the studies raises the possibility of differential effects of self-generation on different aspects of perceptual processing. Although expectations have been found to yield differential effects on perceptual bias and sensitivity measures in the literature outside the action domain (Wyart et al., 2012; Bang & Rahnev, 2017), no systematic attempts have been made to date to assess whether motor actions alter our sensitivity or whether they bias the estimate of stimulus’ perceived loudness as a function of sound intensity.

The aim of the present study is twofold: We sought to elucidate the modulatory effects of intensity on the perceptual processing of self-generated sounds across the auditory intensity range, while systematically assessing whether the expected effects drive changes in sensitivity and/or perceptual bias. To this end, we employed a sound detection and a loudness discrimination task and compared the detection and discrimination sensitivity, as well as the possible bias in perceived loudness for self- vs. externally-generated sounds at both supra- and near-threshold intensities.

Based on previous studies with self-initiated sounds of high and low intensities, we expected to observe i) sensory attenuation for self- compared to externally-generated sounds at supra-threshold intensities and ii) sensory enhancement for self- compared to externally-generated sounds at near-threshold intensities. This interaction would be evident by better detection performance for the self- as compared to the externally-generated sounds (Reznik et al., 2014). Similarly, in the discrimination task, this interaction would be reflected in i) lower PSE for self- compared to externally-generated sounds at supra-threshold intensities (Sato, 2008; Weiss et al., 2011a, 2011b; Reznik et al., 2015) and ii) higher PSE for self- compared to externally-generated sounds at near-threshold intensities (Reznik et al., 2015). Finally, based on previous studies (Sato, 2008; Weiss et al., 2011a, 2011b), we did not expect any significant differences in the JND, at least for the supra-threshold conditions.

The hypotheses and planned analyses for this study were preregistered on the Open Science Framework (https://osf.io/ypajr/). The Method and Results sections follow the preregistered plan.

2. Methods

The present study consisted of two two-alternative forced-choice (2AFC) tasks: a detection and a discrimination task. In the detection task, participants were presented with one sound at varying intensities and had to indicate whether it was presented in a first or a second interval of time, while in the discrimination task two sounds were presented in two different consecutive intervals of time and participants had to indicate whether the first sound (standard) or the second sound (comparison) was louder. The order of tasks was counterbalanced across participants.

2.1. Participants

Thirty-one healthy, normal-hearing subjects, participated in the present study. Participants were typically undergraduate university students at the University of Barcelona. Participants with hearing thresholds above 20 dB, psychiatric or neurological illness, aged below 18 or above 50 years old and who consumed drugs or pharmaceuticals acting on the central nervous system were excluded. Data from three participants (i.e., participants 2, 19, 25) had to be excluded due to technical problems or inability to comply with the task instructions, leaving data from twenty-eight participants (6 men, 22 women, M_{age} = 23, age range: 18–33 years). The sample size was defined based on the preregistered a priori power analysis. All participants gave written informed consent for their participation after the nature of the study was explained to them and they were monetarily compensated (10 euros per hour). Additional materials included a personal data questionnaire and a data protection document. The study was approved by the Bioethics Committee of the University of Barcelona and all provisions of the Declaration of Helsinki were followed.

2.2. Apparatus

The visual stimuli were presented on an ATI Radeon HD 2400 monitor. The auditory stimuli were presented via the Sennheiser KD 380 PRO noise cancelling headphones. To record participants' button presses and behavioural responses, we used the Korg nanoPAD2. The buttons of this device do not produce any mechanical noise when pressed, and, thus, do not interfere with our auditory stimuli. The presentation of the stimuli and recording of participants' button presses and responses were controlled using MATLAB R2007a (The Mathworks Inc., 2017), and the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997).

2.3. Stimuli

In the detection task we used pure tones presented binaurally with durations of 300 ms at a frequency of 1000 Hz (created using MATLAB R2007a; The Mathworks Inc., 2017). The sampling frequency was 44,100 Hz, the ramp duration (duration of the onset and offset ramps) was 25 ms and a number of 16 bits per sample (Reznik et al., 2014; Reznik et al., 2015). The tone intensity ranged from 0 dB to 28 dB in steps of 4 dB for passive and active conditions.

For the discrimination task, we created pure tones with the same characteristics as those used in the detection task, except for the intensities. The intensities for the standard and comparison tones were partly based on those used in previous studies (Sato, 2008; Weiss et al., 2011a, 2011b; Reznik et al., 2015). The standard tone was always presented at a fixed intensity, while the comparison intensities varied. Specifically, the standard tones had a fixed intensity of 74 dB for supra-threshold conditions, while for the near-threshold conditions we used a fixed intensity of 5 dB above the threshold as obtained from the audiometry for the 1000 Hz sounds (Reznik et al., 2015). The comparison supra-threshold stimuli varied randomly between 71 and 77 dB in steps of 1 dB, thereby resulting in seven possible comparison intensities: 71, 72, 73, 74, 75, 76, 77 (Sato, 2008; Weiss et al., 2011a, 2011b). For near-threshold conditions, the comparison intensities were presented at intensities starting from 3 dB below to 3 dB above the standard intensity in steps of 1 dB, so as to match the comparison intensities of the supra-threshold conditions.

2.4. Procedure

Participants were seated in a soundproof chamber and auditory stimuli were presented to both ears via headphones. Visual stimuli were presented by a computer screen located in front of the participants. Prior to each task, hearing thresholds were assessed with a standard pure-tone audiometry. Additionally, practice blocks were used so that participants could familiarize themselves with each task, which also allowed us to obtain the stimulus-onset-asynchrony (SOA) between interval-cue presentation and button press in order to introduce the same visual-to-sound delay in the first passive trials.

2.4.1. Detection task

Participants performed a 2-Alternative Forced Choice auditory detection task, where they had to report whether a sound of varying intensities was presented in interval one or two (Figure 1a). The sounds were either self-generated (active trials) or passively presented by the computer (passive trials). Every trial started with a fixation cross with a duration of 500 ms followed by two consecutive intervals with a duration of 800 ms each. In the active trials, the sound presentation was contingent on participants' button press. That is, participants had to press a button with their right hand once the visual cues "PRESS 1" and "PRESS 2" appeared in order to generate a sound that was triggered by the button press in either the 1st or the 2nd interval. For the intervals containing the sound (either 1st or 2nd), the participants' button press triggered the sound only if he/she pressed the button up to 300 ms prior to the interval offset. This allowed us to control that the sound had always a 300-ms duration in case a participant delayed the button press. In the passive trials, participants were passively presented with a sound in one of the two intervals indicated by the visual cues "LISTEN 1" and "LISTEN 2". To match the timing of the sound in the active conditions, the sound was presented after an interval that was randomly selected from the participants' distribution of press times in the active trials performed until the current trial. Thus, the timing of the stimulus presentation was equal for the two types of trials, thereby minimizing any effects of differences in sound timing on the ability to detect self- and externally-generated sounds (Horváth, 2015; Hughes et al., 2013a). After the offset of the second interval, the question "Did you hear the sound in the 1st or 2nd interval?" appeared on the screen for 1500 ms and participants had to press a button with their left hand within this time window to respond. For both trials, once a response was provided the question displayed on the screen disappeared immediately. The next trial started always after the 1500 ms response window was over.

The whole task was divided into 25 blocks consisting of 40 trials, resulting in 1000 trials in total (500 active and 500 passive trials). Active and passive conditions were presented randomly intermixed within each block (20 active and 20 passive trials). The intensities were presented using the method of constant stimuli. Intensities from 0 dB to 24 dB were presented a total of 70 times each for each condition, while we only presented the sound at 28 dB 10 times for each condition to save experimental time, given that pilot data showed ceiling performance at this intensity level. The interval containing the sound (interval 1 or 2) was random.

2.4.2. Discrimination task

In the discrimination task two sounds were presented in two different consecutive intervals and participants had to indicate whether the first (standard) or the second sound (comparison) was louder (Figure 1b). Similarly to the detection task, there were two types of trials, passive and active. However, there were two additional intensity conditions, supra- and near-threshold, thereby resulting in 4 possible types of trials in total: Active and Supra-threshold (AS), Passive and Supra-threshold (PS), Active and Near-threshold (AN) and Passive and Near-threshold (PN).

Each trial started with a fixation cross with a duration of 500 ms followed by two consecutive intervals with a duration of 800 ms each. In the active trials, participants had to press a button with their right hand in the first interval, instructed by the cue "PRESS: sound 1", in order to generate the standard tone. The comparison sound was passively presented in the second interval of time following the visual cue "LISTEN: sound 2". The interval between visual cue and comparison sound onset was randomly selected from the participants' distribution of press times in the first interval. For the standard self-generated sound, the participants' button press triggered the sound only if he/she pressed the button up to 300 ms prior to the interval offset. This allowed us to control that the sound had always a 300-ms duration in case a participant delayed the button press. In the passive trials, participants were passively presented with two sounds in the 1st and the 2nd interval, respectively, indicated by the visual cues "LISTEN: sound 1" and "LISTEN: sound 2". The sounds were presented after an interval that was randomly selected from the participants' distribution of press times in the active trials. The interval between the two sounds was therefore random depending on the timing of the button press (active conditions) or the random delay drawn by the distribution of press times (passive conditions). Unbeknownst to the subjects, the standard tone was always presented at the same intensity within each intensity condition: 74 dB for supra-threshold conditions and 5 dB above the threshold obtained from the audiometry for near-threshold conditions. In contrast, the comparison sound ranged from 71 dB to 77 dB in steps of 1 dB for supra-threshold conditions and ±3 dB in steps of 1 dB relative to the standard tone for near-threshold conditions. After the offset of the second comparison interval, the question "Which sound was louder: Sound 1 or Sound 2?" appeared on the screen for 1500 ms and participants had to press a button with their left hand to indicate whether the first (left button) or the second (right button) sound was louder. To control for the possibility that participants did not hear the near-threshold sounds, a third control button was used, and participants were instructed to press it only if they did not hear the two sounds. After participants' response, the question disappeared immediately. The next trial started always after the 1500 ms response window was over.

The task was divided in 25 blocks, each one consisting of 28 trials. Each of the seven possible comparison tone intensities was presented 25 times per condition using the method of constant stimuli, as it yields a better estimation of the Point of Subjective Equality (PSE) and Just Noticeable Difference (JND) values compared to other methods (Guilford, 1954). This resulted in 175 trials per experimental condition (active/passive and supra-/near-threshold) and 700 trials in total for each participant. The conditions (i.e., sound-source: active vs. passive, and intensity: supra- vs. near-threshold) were intermixed within each block and the order of presentation was randomized for each participant.

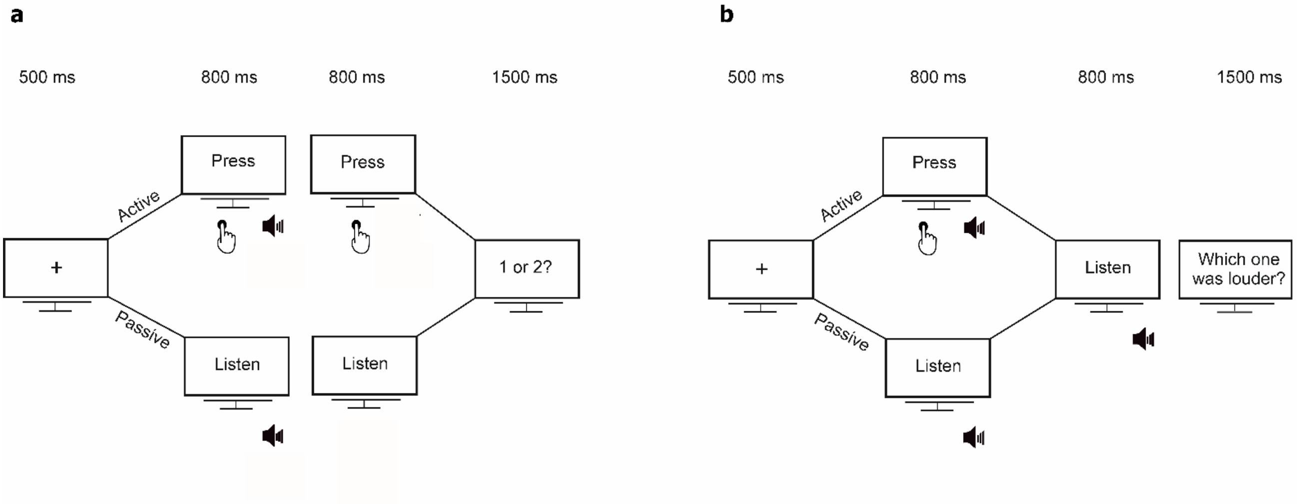


Figure 1. Schematic illustration of the experimental design. **a**) *Detection task*: Each trial started with a fixation cross, followed by two intervals. In active trials, participants were instructed to press a button in each interval ("Press" cue) and a sound was triggered either in 1st or in the 2nd one (in the example shown here, the sound is presented in the 1st interval). In passive trials, the sound was passively presented ("Listen" cue). Participants had to respond whether they heard the sound in the 1st or in the 2nd interval. **b**) *Discrimination task*: Each trial started with a fixation cross, followed by two sounds. The first sound was either self- (active trials; "Press" cue) or externally-generated (passive trials; "Listen" cue) and was presented at an intensity of either 74 dB (supra-threshold intensity) or 5 dBs above each participant's audiometric threshold (near-threshold intensity). The second sound was always externally-generated ("Listen" cue) and ranged ±3 dB in steps of 1 dB relative to the first one. Participants had to respond which one was louder.

2.5. Modifications from the preregistered plan

This experiment was preregistered on the Open Science Framework (https://osf.io/ypajr/). Relative to our preregistered plan, we made one modification: Instead of fitting the psychometric function with the Palamedes Toolbox (Kingdom & Prins, 2016) as reported in the preregistration of this study, we decided to use the quickpsy package in R (Linares & López-Moliner, 2016) for better visualization of the data and in order to directly introduce the values obtained from the fitting procedure to statistical analysis in R. The change in the toolbox used is not expected to have affected the results, as we kept all the parameters as predefined in the preregistration.

2.6. Data analysis

Data analysis follows the preregistered plan. All analysis code will be publicly released with the data upon publication (https://osf.io/ypajr/).

2.6.1. Detection task

For each participant, the percentage of correct answers were calculated for each intensity and condition − active and passive −. Subsequently, for each condition, the percentage of correct responses was fitted with a normal cumulative function (see average psychometric functions in Figure 2 and the individual psychometric functions in Supplementary Fig. S1) according to the maximum likelihood procedure, using the quickpsy package in R (Linares & López-Moliner, 2016). For each participant and condition, two parameters were extracted from the model: alpha (i.e., values for thresholds in the range of the intensity levels we used) and beta (i.e., values for slope in the range of 0 to 10 in steps of .1). The lower asymptote of the psychometric function (i.e., gamma) was set to 0.5 as in previous 2-AFC detection tasks, while the upper asymptote (i.e., lambda), which corresponds to the lapse rate, was set to .001 (Kingdom & Prins, 2016). For each participant and condition, goodness-of-fit and the 95% confidence intervals for thresholds were calculated by a parametric bootstrap procedure (n = 1000; Efron & Tibshirani, 1993), using the quickpsy package in R (Linares & López-Moliner, 2016).

The second part of the analysis consisted in calculating the d’ sensitivity index and criterion in order to directly compare our results with previous studies using this measure (Reznik et al., 2014). This analysis was performed using the Palamedes toolbox (version 1.10.3; Kingdom & Prins, 2016). Given that here we employed a 2-AFC task, we first calculated the hit and false alarm rate for one of the two intervals (interval 1 as target). As hit for interval 1 were defined the trials, where the sound was in interval 1 and the participant responded that the sound was indeed presented in this interval. As false alarm for interval 1 were defined the trials, where the participant incorrectly detected the sound in interval 1, while the stimulus was actually presented in interval 2. Subsequently, we calculated the hit rate (= number of hits divided by the number of signal trials, i.e., trials where the sound was presented in the 1st interval) and the false alarm rate (= number of false alarms divided by the number of noise trials, i.e., trials where the sound was presented in the 2nd interval). After z-transforming the hit and false alarm rates, we calculated the d’ (i.e., z(Hit) – z(False Alarm)) and criterion (i.e., -0.5 * (z(Hit) + z(False Alarm))) for active and passive trials. Finally, we calculated the mean interval between the cue presentation and participants’ button press (henceforth SOAs) in the active trials.

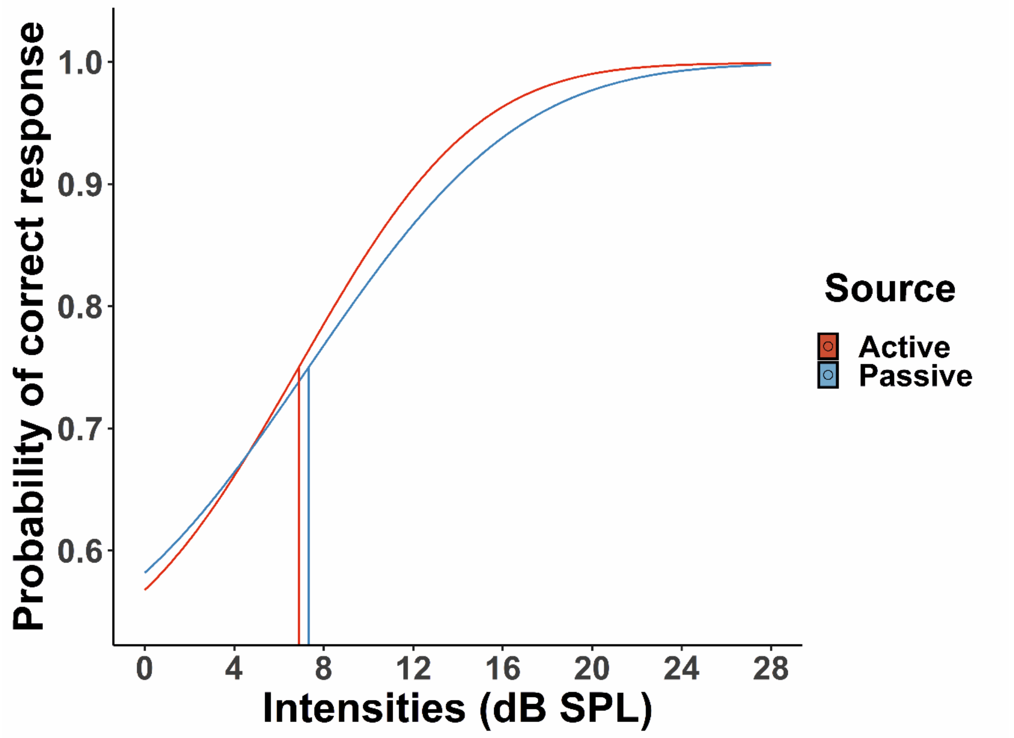


Figure 2. Group psychometric functions per source condition (active and passive) in the detection task. Vertical lines represent the mean thresholds per source condition (i.e., defined as the intensity accurately detected at 75% of the trials). Detection thresholds did not differ between active and passive trials (*p* > .050).

2.6.2. Discrimination task

For each participant, the proportion of "second sound louder" responses was calculated for each condition (active/passive, supra-/near-threshold) and for the seven comparison intensities. Data from the trials where participants did not hear the near-threshold sounds (as indicated by the third control button; see Procedure) were excluded from the analysis. In order to directly compare performance across supra- and near-threshold conditions, we defined the comparison intensities as the difference in dB from the standard stimulus: -3, -2, -1, 0, 1, 2, 3. The "second sound louder" responses for each condition were, then, fitted with a normal cumulative function (see average psychometric functions in Figure 3 and the individual psychometric functions in Supplementary Fig. S2) according to the maximum likelihood procedure, using the quickpsy package in R (Linares & López-Moliner, 2016). For each participant and condition, two parameters were extracted from the model: alpha (i.e., values in the range of the comparison intensity levels we used) and beta (i.e., values for slope in the range of 0 to 10 in steps of .1). The lower asymptote of the psychometric function (i.e., gamma) was set to 0 as in previous 2-AFC discrimination tasks, while the upper asymptote (i.e., lambda), which corresponds to the lapse rate, was set to .001 (Kingdom & Prins, 2016). Thus, for each participant and condition, two measures were obtained. First, the Point of Subjective Equality (PSE), which corresponds to the alpha values of the model, and is defined as the intensity, where the comparison stimulus was reported as louder than the standard one on 50% of the trials. This value is used to estimate the comparison tone intensity that would make the standard and comparison tones perceptually equal and is considered an index of perceptual bias (Bausenhart et al., 2018). Higher PSE values would indicate that the standard first tone is perceived as louder, while lower PSE values would reflect an attenuated perceived loudness for this sound. Thus, shifts of the PSE values from the Point of Objective Equality (i.e., the point indexing the physical equality of the two sounds, which is 0 dBs here) would reflect a biased estimate of perceived loudness. Second, we extracted the just noticeable difference (JND), which corresponds to the beta values of the model (i.e., the standard deviation extrapolated from the fit) and is considered a measure of precision associated with the estimate. Higher JND values would reflect lower precision in discriminating the loudness of the two sounds (i.e., lower differential sensitivity; Gescheider, 1997). For each participant and condition, goodness-of-fit and the 95% confidence intervals for PSE were calculated by a parametric bootstrap procedure (n = 1000; Efron & Tibshirani, 1993), using the quickpsy package in R (Linares & López-Moliner, 2016). Finally, we calculated the mean interval between the cue presentation and participants' button press (henceforth SOAs) in the active trials.

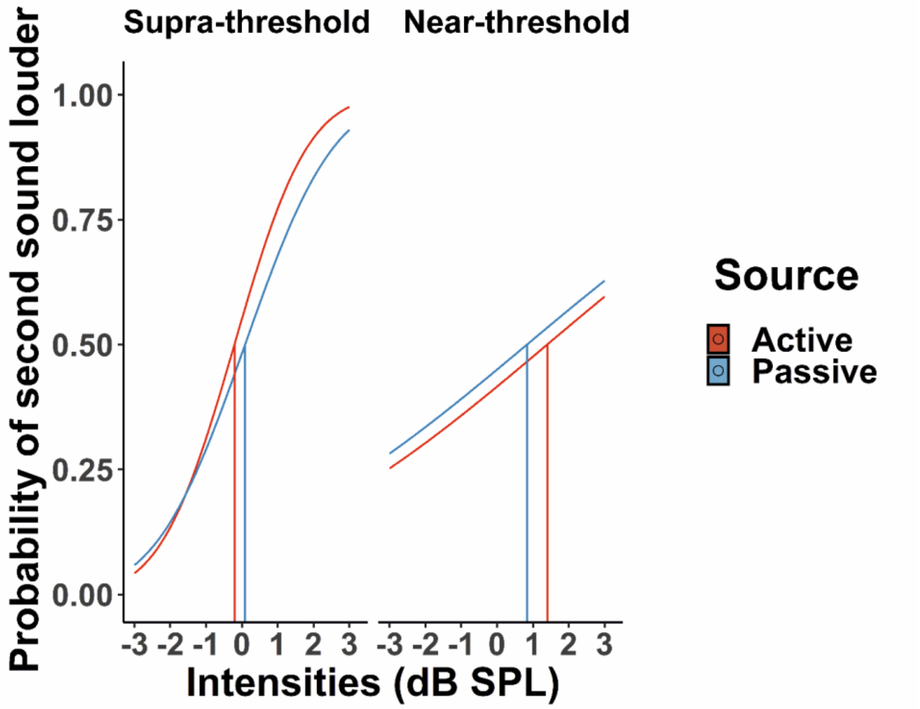


Figure 3. Group psychometric functions per source and intensity conditions in the discrimination task. Vertical lines represent the PSE values per source and intensity (i.e., defined as the intensity, where the second comparison stimulus was reported as louder than the first standard one on 50% of the trials). Supra-threshold active sounds had lower PSE values than the supra-threshold passive ones, suggesting attenuated perceived loudness for the former. In contrast, at near-threshold intensities, active sounds had higher PSE values than the passive ones, pointing to enhanced perceived loudness for the former.

3. Results

All statistical analyses were performed using R (version 3.6.0). For all the significant results in the ANOVA, we report the η_G^2 effect size and the η_p^2, since the η_G^2 is less biased than η_p^2 (Bakeman, 2005; Olejnik & Algina, 2003), but we also wanted to compare our findings with other studies that usually report the η_p^2 effect size. Participants' audiometric thresholds did not differ across the two audiometric sessions (see Supplementary Fig. S3). Significant subject-wise correlations between the measures across the two tasks are reported in Supplementary Table S1. Error bars in Figures 4 and 5 depict the within-subjects confidence intervals (Pfister & Janczyk, 2013; Cousineau, 2005; Morey, 2008), calculated using the *summarySEwithin* function in R (Morey, 2008).

3.1. *Modifications from the preregistered plan*

Relative to our preregistered analyses (https://osf.io/ypajr/), we made one modification: For the detection task, we initially planned to perform a paired-samples *t*-test to test for differences in the slope of the psychometric function. However, considering that the normality test was violated, we performed a non-parametric Wilcoxon test.

3.2. *Detection Task*

The thresholds, slopes, d', and criterion values, were analyzed using paired samples *t*-tests with the factor sound source – active (A) or passive (P). Trials with erroneous presses (i.e., late onset time of button press and no presses) were excluded from all analyses (M_A= 28.26%, SD_A = 20.37 M_P = 2.35%, SD_P = 3.3). For the active trials, the mean interval between cue onset and button press was 0.39 s (*SD* = .07) for Interval 1 and 0.16 s (*SD* = .14) for Interval 2.

First, we performed statistical analyses for the measures obtained from the psychometric fitting procedure (Figure 4). To test for differences between the thresholds in the active and passive conditions, we used a paired samples one-tailed *t*-test with the hypothesis of expecting lower detection thresholds in the active compared to passive trials (Reznik et al., 2014; Shapiro-Wilk normality test *p* > .050). The analysis did not show any significant differences between active and passive conditions (*t*(27) = -1.09, *p* > .050, M_A = 7.46, M_P = 7.85, SD_A = 3.7, SD_P = 3.66). Subsequently, we tested for possible differences in the slope of the psychometric function. Considering that the assumption of normality was violated (Shapiro-Wilk normality test *p* = .020), we performed a nonparametric Wilcoxon's signed rank test for paired data on the beta values obtained from the psychometric functions. The analysis did not show any significant difference between active and passive slopes (*W* = 146, *p* > .050, M_A = 4.65, M_P = 5.05, SD_A = 2.48, SD_P = 3.11). Finally, to further test for possible effects of self-generation and intensity level on detection performance, we also analyzed the percent of correct responses for both the active and passive trials for each one of the intensity levels, but did not find any significant interactions (see Supplementary Fig. S4).

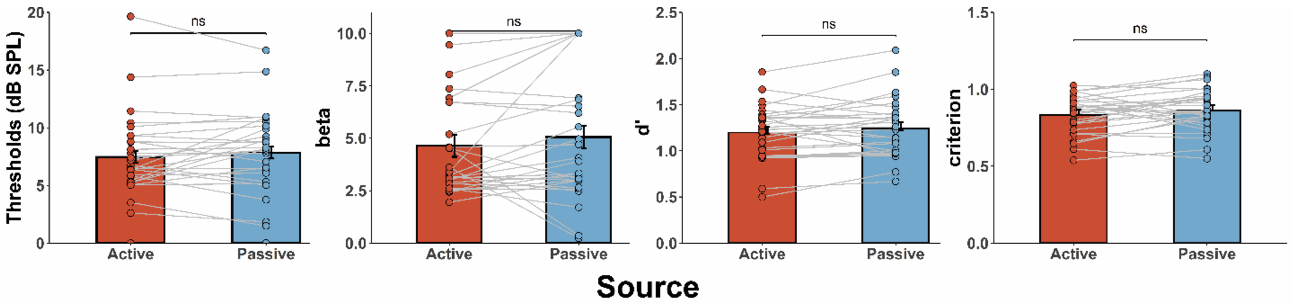


Figure 4. Summary of the results from the detection task. Mean threshold, beta value for slope, d' score, and criterion. Error bars depict within-subjects confidence intervals (Pfister & Janczyk, 2013; Cousineau, 2005; Morey, 2008). There were no significant differences between active and passive in the threshold (one-tailed paired samples *t*-test, *p* > .050), slope (i.e., beta values from the psychometric fitting procedure; nonparametric Wilcoxon test due to violation of normality assumption, *p* > .050), d' score (one-tailed paired samples *t*-test, *p* > .050), or criterion (two-tailed paired samples *t*-test, *p* > .050).

To analyze the differences in the thresholds between the two conditions, we also calculated a 95% confidence interval for the difference in thresholds based on the simulations from the bootstrapping procedure ($n = 1000$). For 23 out of the 28 subjects no significant differences were observed between the active and the passive trials. For one of them, the comparison between observed and simulated thresholds showed a significantly higher threshold for the active compared to the passive trials, while for the other four, a significantly lower threshold was obtained for the active trials. The goodness-of-fit routine showed that for the active trials, 26 out of the 28 psychometric curves resulted in acceptable goodness-of-fit statistics, while the fitting procedure for the passive trials showed acceptable goodness-of-fit statistics for 25 out of the 28 psychometric curves. Despite the non-acceptable goodness-of-fit for some subjects and conditions, we kept these subjects in the analyses, after confirming that results would remain the same when excluding them.

Subsequently, we performed a signal detection analysis for the d' and criterion values (Figure 4; Shapiro-Wilk normality test, $p > .050$). The d' values were analyzed using a paired samples one-tailed *t*-test with the hypothesis of expecting higher d' in active compared to passive trials (Reznik et al., 2014). Contrary to previous work (Reznik et al., 2014), the analysis did not show any significant differences between the active vs. passive d' values ($M_A = 1.2$, $SD_A = 0.3$, $M_P = 1.24$, $SD_P = .32$, $p > .050$). Similarly, the criterion values were analyzed using a paired-samples two-tailed *t*-test (Reznik et al., 2014). Similar to previous work (Reznik et al., 2014), we did not observe any significant difference in the criterion values between active and passive trials ($M_A = .83$, $SD_A = .12$, $M_P = .86$, $SD_P = .13$, $p > .050$). Collectively, although these findings suggest that self-generation does not affect detection sensitivity or response bias in a 2-AFC detection task, the lack of a contingent press-sound relationship (i.e., participants pressed twice in every active trial but only one button press generated the sound) may have also minimized any possible effects of motor-related predictions on detection performance.

3.3. Discrimination Task

The PSE and JND values were analyzed using a repeated measures ANOVA with two factors: sound source − active (A) or passive (P) − and sound intensity − supra- (S) or near-threshold (N) −. Trials with erroneous presses (i.e., late onset time of button press and no presses) were excluded from all analyses ($M_{AS} = 22.9\%$, $SD_{AS} = 19.1$, $M_{PS} = 0.96\%$, $SD_{PS} = 2.74$, $M_{AN} = 23.29\%$, $SD_{AN} = 18.88$, $M_{PN} = 1.51\%$, $SD_{PN} = 3.09$). For the active trials, the mean interval between cue onset and button press was 0.37 s ($SD = .06$).

The analysis for the PSE values revealed that there was not a main effect of source ($F(1,27) = .8$, $p > .050$; $M_A = .39$, $M_P = .25$, $SD_A = 1.65$, $SD_P = 1.65$) or a main effect of intensity ($F(1,27) = 2.62$, $p > .050$; $M_N = .65$; $M_S = -.008$, $SD_N = 2.12$, $SD_S = .86$). However, there was a significant interaction between source and intensity ($F(1,27) = 12.10$, $p = .002$, $\eta_p^2 =.31$ and $\eta_G^2= .15$; Figure 5). The Bonferroni corrected post-hoc tests revealed a higher PSE for the AN condition compared with the AS condition ($M_{AN} = .92$, $M_{AS} = -.13$, $SD_{AN} = 2.04$, $SD_{AS} = .9$, $t(27) = -2.48$, $p = .020$, $d = .47$; two-tailed post-hoc *t*-test), a lower PSE for the AS compared to the PS condition ($M_{AS} = -.13$, $M_{PS} = .12$, $SD_{AS} = .9$, $SD_{PS} = .83$, $t(27) = -2.41$, $p = .012$, $d = .45$; one-tailed post-hoc *t*-test), and a higher PSE for the AN compared to the PN condition ($M_{AN} = .92$, $M_{PN} = .39$, $SD_{AN} = 2.04$, $SD_{PN} = 2.19$, $t(27) = 2.09$, $p = .020$, $d = .39$; one-tailed post-hoc *t*-test). The post-hoc analysis did not show differences between the PS and the PN condition ($M_{PS} = .12$, $M_{PN} = .39$, $\mathrm{SD}_{PS} = .83$, $SD_{PN} = 2.19$, $t(27) = -.64$, $p > .050$; two-tailed post-hoc *t*-test). Thus, we replicate the findings obtained by previous discrimination studies with supra-threshold sounds (Sato, 2008; Weiss et al., 2011a, 2011b), but more importantly we extend previous work by showing that self-generated near-threshold sounds are perceived louder compared to the passively presented ones.

The analysis for the JND values revealed that there was a significant main effect of intensity ($F(1,27) = 119.45$, $p < .001$, $\eta_p^2 =.82$ and $\eta_G^2= .49$), with a significantly higher JND (i.e., lower discrimination sensitivity) for the near- compared to the supra-threshold conditions ($M_S = 1.93$, $M_N = 5.79$, $SD_S = 1.5$, $SD_N = 2.39$; Figure 5). The analysis did not show a significant main effect of source ($F(1,27) = 2.75$, $p > .050$; $M_A = 3.68$, $M_P = 4.03$, $SD_A = 2.7$, $SD_P = 2.9$) or a significant interaction between source and intensity ($F(1,27) = .77$, $p > .050$). Collectively, the results obtained by these analyses are consistent with previous work with both auditory (Sato, 2008; Weiss et al., 2011a, 2011b) and tactile self-generated stimuli (Kilteni et al., 2020) and further show that the interactive effects of intensity and self-generation are not dependent on participants' differential sensitivity in discriminating the loudness of two sounds (as indexed by the JND values).

For analyzing differences in the PSE between the four conditions, the 95% confidence intervals were calculated for each condition based on the simulations from the bootstrapping procedure (n = 1000). For 9 subjects we found significant differences between the active and passive supra-threshold conditions (for 8 subjects, lower PSE in the AS compared to the PS), while for the near-threshold intensities only 3 subjects had significantly higher PSE values in the active compared to the passive condition. Within the active condition, significant differences were obtained between the supra- and near-threshold intensities for 16 subjects (for 13 subjects, lower PSE in the AS compared to the AN), while for 18 subjects we found significant differences between the passive supra- and passive near-threshold conditions (for 12 subjects, lower PSE in PS compared to PN). The goodness-of-fit routine showed that for 26, 27, 26, and 26 psychometric curves out of the 28 total curves fitted per condition, the fitting procedure resulted in acceptable goodness of-fit statistics (for the AN, AS, PN, and PS, respectively). Despite the non-acceptable goodness-of-fit for some subjects and conditions, we kept these subjects in the analyses, after confirming that results would remain the same when excluding them.

Finally, we aimed to directly compare our results with the findings obtained by Reznik et al. (2015), where they employed a similar discrimination task where the standard and comparison tone were always presented at the same intensity (either supra- or near-threshold). Thus, in this analysis we only included the trials where the comparison sound was presented at the same intensity as the standard one (i.e., 74 dB for the supra-threshold and 5 dBs above each participant's audiometric threshold for near-threshold conditions). In order to directly compare with Reznik et al.'s study, we calculated the "1st sound louder" responses and performed a repeated measures ANOVA with the factors sound source (active/passive) and sound intensity (supra-/near-threshold). The results showed a significant main effect of source ($F(1,27) = 13.54$, $p < .001$, $\eta_p^2 = .33$ and $\eta_G^2 = .04$), with less "1st sound louder" responses for active compared to passive trials ($M_A = 46.1$, $SD_A = 20.62$, $M_P = 53.63$, $SD_P = 19.86$). The main effect of intensity did not reach significance ($F(1,27) = 3.26$, $p > .050$, $M_N = 53.77$, $SD_N = 21.57$, $M_S = 45.98$, $SD_S = 18.76$). However, consistent with Reznik et al. (2015), we obtained a significant interaction between source and intensity ($F(1,27) = 8.94$, $p < .010$, $\eta_p^2 = .25$ and $\eta_G^2 = .04$; Figure 5). The post-hoc *t*-tests showed that while there were significantly less "1st sound louder" responses for AS compared to PS trials ($M_{AS} = 38.12$, $M_{PS} = 53.82$, $SD_{AS} = 16.56$, $SD_{PS} = 17.75$, $t(27) = -5.19$, $p < .001$, $d = .98$; one-tailed paired samples *t*-test), no differences were observed between active and passive trials at near-threshold intensities ($M_{AN} = 54.09$, $M_{PN} = 53.45$, $SD_{AN} = 21.43$, $SD_{PN} = 22.10$, $t(27) = .17$, $p = .570$; one-tailed paired samples *t*-test). We also observed significantly more "1st sound louder responses" for the AN compared to the AS condition ($M_{AN} = 54.09$, $SD_{AN} = 21.43$, $M_{AS} = 38.12$, $SD_{AS} = 16.56$, $t(27) = -3.03$, $p = .010$, $d = .01$; two-tailed paired samples *t*-test), while no differences were obtained between the PS and PN conditions ($M_{PS} = 53.82$, $SD_{PS} = 17.75$, $M_{PN} = 53.45$, $SD_{PN} = 22.10$, $t(27) = .08$, $p = .840$; two-tailed paired samples *t*-test). Collectively, the comparison analysis we performed replicates the significant interaction reported by Reznik et al. (2015) with an even larger effect size ($\eta_p^2 =.25$ here compared to $\eta_p^2 =.21$ in their study), but the follow-up analyses demonstrate that when the standard and comparison tones are presented at the same intensity, the differences between self- and externally-generated sounds are limited to supra-threshold intensities.

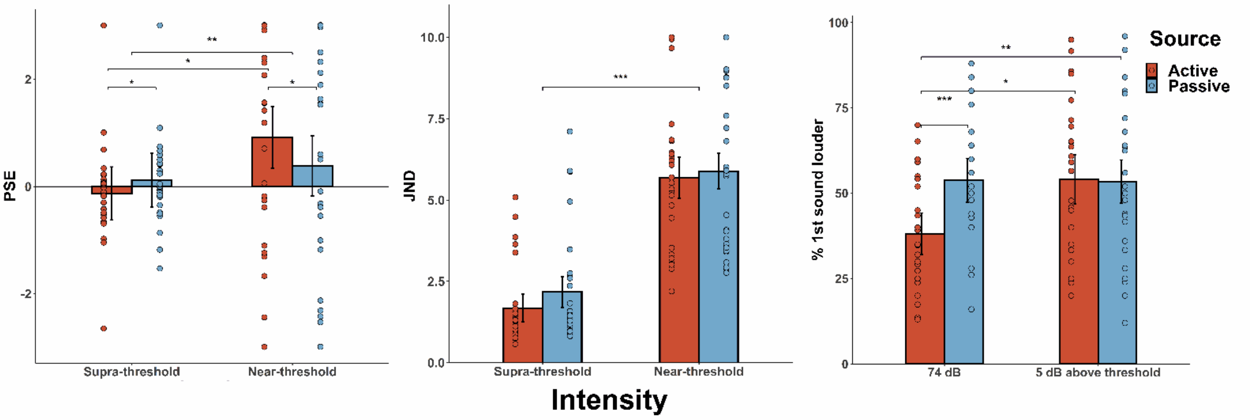


Figure 5. Summary of the results from the discrimination task. Mean PSE, JND, and percent of "1st sound louder responses" (Reznik et al., 2015). Error bars depict within-subjects confidence intervals (Pfister & Janczyk, 2013; Cousineau, 2005; Morey, 2008). From left to right: Significant interaction between source and intensity on PSE ($p < .010$), with the post-hoc comparisons showing lower PSE for the active supra-threshold compared to the passive supra-threshold condition (one-tailed paired samples post-hoc t-test; $p < .050$), significantly higher PSE for the active near-threshold compared to the passive near-threshold condition (one-tailed paired samples post-hoc t-test; $p < .050$), and significantly higher PSE for the active near-threshold compared to active supra-threshold (two-tailed paired samples post-hoc t-test; $p < .050$). Significant main effect of intensity on JND, with lower JND (i.e., better discrimination sensitivity) for the supra- compared to the near-threshold condition ($p < .001$). For the "1st sound louder responses", we only included trials where the standard and the comparison sounds were presented at the same intensity (i.e., 74 dB as a supra-threshold intensity and 5 dB above each participant's threshold as a near-threshold intensity; Reznik et al., 2015). There was a significant interaction between source and intensity ($p < .010$), with the post-hoc comparisons showing less "1st sound louder" responses for active compared to passive trials when the sound was presented at 74 dB (one-tailed paired samples post-hoc t-test; $p < .001$), less "1st sound louder" responses for active trials when presented at 74 dB compared to when presented at 5 dB above each participant's threshold ($p < .050$), and no differences between active and passive trials when the sounds were presented at 5 dB above each participant's threshold (one-tailed paired samples post-hoc t-test; $p > .050$).

4. Discussion

To-date, many different models have attempted to elucidate the effects of motor acts on perceptual processing. Yet, empirical evidence as to the exact direction and nature of these effects remain mixed. We hypothesized that the mixed findings may be related to the modulatory effects of stimulus intensity and to differences regarding the exact aspect of perceptual processing that is being tested. Here, we present a preregistered study with a priori power estimations (https://osf.io/ypajr/), where we utilized a wide range of intensities to test for possible differences between self- and externally-generated sounds in detection and discrimination ability. Contrary to previous work (Reznik et al., 2014), we did not observe enhancements in the detection sensitivity for near-threshold self-generated sounds. However, in the discrimination task we found a significant interaction between self-generation and intensity on perceptual bias (i.e., PSE) that replicates and extends previous work (Sato, 2008; Weiss et al., 2011a, 2011b; Reznik et al., 2015) by showing that perceived intensity is reduced for self-generated sounds when they are presented at supra-threshold intensities, but enhanced when presented at near-threshold intensities.

Extant models disagree about how motor predictions affect the perceptual processing for expected action consequences. On one hand, consistent with ideomotor theories proposing that we internally activate the sensory outcome of our own action (Hommel et al., 2001), dominant cancellation models in the action literature have suggested that behavioural and neurophysiological responses to expected action effects are suppressed (Kilteni et al., 2020; von Holst, 1954; Blakemore et al., 2000). Such attenuation is also predicted by preactivation accounts postulating that expectations preactivate representations of the predicted effects, increasing their baseline activity, thereby rendering the actual input less discriminable from baseline and reducing detection sensitivity (Roussel et al., 2013). In contrast, according to sharpening models, the motor-driven suppression proposed by cancellation theories is limited to units tuned away from the expected input, resulting in a sharpened population response and higher signal-to-noise ratio (SNR) that ultimately boosts detection sensitivity for what we expect (Yon & Press, 2017). However, none of these models can account for our findings: The cancellation account would predict lower perceived intensity irrespective of signal strength, while according to the preactivation and sharpening models we should have found significant differences in detection sensitivity (lower or higher for self-generated sounds, respectively). Critically, these models cannot explain the enhanced perceived intensity for expected near-threshold sounds. Although this enhancement may be partly driven by multisensory integration processes that are known to boost processing when the unimodal signal is of low strength like the near-threshold self-generated sounds (e.g., inverse effectiveness; Stein & Meredith, 1993), two recent models have raised the possibility that the signal strength interacts with motor predictions in determining whether the processing of the expected events will be enhanced or cancelled out (Press et al., 2010; Reznik & Mukamel, 2019).

Reznik and Mukamel (2019) recently proposed that the inhibitory influence exerted by the motor cortex on auditory areas during motor acts (Schneider & Mooney, 2018) may either dampen or enhance perceptual processing of self-generated sounds depending on the environmental context. According to their model, the motor-driven suppression of the auditory cortex (Carcea et al., 2017; Buran et al., 2014) leads to reduced activity at the population level, but also to more selective responses and thus higher SNR. Crucially, while net activity should be always reduced during motor engagement irrespective of intensity, the resulting SNR is proposed to be higher in faint compared to salient contexts. Faint stimulation is known to elicit responses only on “best-frequency” neurons, while louder stimuli also stimulate the neurons tuned to nearby frequencies (Reznik & Mukamel, 2019). Thus, Reznik and Mukamel propose that in faint contexts, the global inhibition during motor engagement may result in “best-frequency” responses only, with almost complete silence of the activity in nearby frequencies thanks to the inhibition of the spontaneous activity, relatively enhancing the sound-evoked activity compared to the background noise (Carcea et al., 2017; Buran et al., 2014).

This proposal has two important implications as concerns the consequences of motor engagement in perceptual processing: First, salient environments would be characterized by reductions in the loudness perception that are proposed to be driven by reduced population activity. Yet, no predictions are made as to whether perceived intensity for near-threshold sounds would be also attenuated or not, thus leaving unexplained our finding that perceived intensity is enhanced for self-generated near-threshold sounds. Second, the increased SNR would boost the detectability of near-threshold sounds only, since in salient contexts sensitivity is already at ceiling. The authors found support for this claim in a study (Reznik et al., 2014) where self-generation significantly enhanced sound detectability. However, this finding was not replicated in the present study.

A caveat to the model proposed by Reznik and Mukamel is that it is largely based on animal studies that compared auditory responses in active vs. passive states (i.e., locomotion vs. quiescence or Go/No-Go tasks; McGinley et al., 2015; Carcea et al., 2017; Buran et al., 2014), rather than comparing self- vs. externally-generated sounds. It is very likely that active states and contingent action-stimulus relationships do not have the same underlying mechanisms, and that they in turn do not modulate perception in the same way. The modulations found in active states may be mostly driven by unspecific neuromodulatory processes (i.e., arousal; McGinley et al., 2015), while in the presence of a contingent action-stimulus relationship specific prediction mechanisms may dominate (i.e., corollary discharge). This critical difference may explain why we did not find any significant effects in the detection task that lacked a contingent press-sound relationship (only 50% of the presses generate a sound). However, previous detection paradigms have also reported no such enhancement (McGinley et al., 2015; Neske et al., 2019; Myers et al., 2020), thus raising the possibility that the low power of the only human study reporting lower detection thresholds for self-generated sounds (n = 10; Reznik et al., 2014) may have reduced the likelihood of their statistically significant result reflecting a true effect (Button et al., 2013). Collectively, although Reznik and Mukamel were the first to attempt to explain how sound intensity may modulate neural and behavioural responses during motor engagement, their model cannot fully explain our findings, and in particular it also cannot explain why the interactive effects we observed here are limited to perceptual bias, rather than perceptual sensitivity.

We believe that our findings can be best explained by the opposing process theory which highlights the role of signal strength in enhancing or suppressing the processing of predictable stimuli (Press et al., 2020). According to this theory, perception is in principle biased towards expected stimuli, such as self-generated and thus more predictable stimuli. However, if the presentation of an unpredicted stimulus generates a high level of surprise after the initial stages of sensory processing, then the perceptual processing of this surprising stimulus is boosted. In terms of self-generation effects, this would imply enhanced processing of externally-generated, and thus unpredictable (surprising) stimuli. Critically, however, the level of surprise is closely related to signal strength, as surprise reflects both the distance between the prior and posterior distributions, as well as their precision (Kullback-Leibler divergence; Kullback, 1997; Itti & Baldi, 2009), and weaker signals are inherently less precise. For example, the sound of a horn in the middle of the night would elicit surprise but only if it is loud, and thus clearly audible. In sum, according to this view, supra-threshold externally-generated stimuli are inherently more surprising than the self-generated ones, shifting perception toward the unexpected (i.e., enhanced perceived loudness for the externally-generated sounds at supra-threshold intensities). Conversely, when sounds are presented at a near-threshold intensity, the increased uncertainty and higher level of noise in the signal renders externally-generated sounds unsurprising and perception is shifted towards the expected (i.e., enhanced perceived intensity for the self- compared to the externally-generated sounds at near-threshold intensities). Thus, the surprise-driven mechanism operates only for highly precise and therefore task-relevant unexpected signals, triggering a process that boosts their perception by driving attention away from the consequences of self-made acts as proposed by the active inference framework (Brown et al., 2013). Therefore, the shifts in perceived intensity in either direction may be related to surprise-induced attentional mechanisms that have been suggested to modulate the precision of the prediction error, rather than the prediction error itself (Barron et al., 2020; Brown et al., 2013). Nevertheless, one would expect that this mechanism would also operate in detection paradigms, contrary to the null findings obtained in the detection task. While these findings may be due to the lack of a contingent action-sound relationship as mentioned before, an alternative explanation is that the attentional nature of these effects results in affecting certain aspects of perceptual processing.

The studies conducted so far have not systematically assessed the effects of self-generation (and their interaction with stimulus intensity) on the different perceptual measures. Discrimination studies have reported shifts in the PSE, a measure of *perceptual bias*, while JND – a measure of *perceptual sensitivity* – remains unaffected by self-generation (Kilteni et al., 2020; Sato, 2008; Weiss et al., 2011a, 2011b). Conversely, detection studies have typically measured perceptual thresholds or the d' score (*perceptual sensitivity*) and criterion (*response bias;* Reznik et al., 2014; Cardoso-Leite et al., 2010). Here, we provide a more complete picture of how motor actions may affect perception by having two tasks that allowed us to obtain all these measures within subjects and show that the effects of self-generation and their interaction with stimulus intensity are driven by shifts in *perceptual bias*, rather than *sensitivity* measures.

In sum, the present study showed that the intensity of the sensory feedback biases perception for self-initiated stimuli in a differential manner, with attenuated perceived loudness at loud intensities, but perceptual enhancement for near-threshold ones. These findings provide empirical support to the opposing process theory (Press et al., 2020) by showing that the behavioural difference between self- and externally-generated sounds interacts with the noise of the sensory outcome in driving perceptual processing. The strength of this study is that it extends previous work by demonstrating that self-generation and its interaction with intensity only affects *perceptual bias*, rather than *perceptual sensitivity* (Sato, 2008; Weiss et al., 2011a, 2011b; Myers et al., 2020) or *response bias* (Reznik et al., 2014). Although the opposing process theory does not clarify whether expectation effects are driven by perceptual or later decisional processes (Press et al., 2020), we argue that the proposed bias in perception as a function of signal strength implies a competition between two percepts, which was only the case in the discrimination task and may point to attentional processes that are known to reverse the effects of prediction on sensory processing (Kok et a., 2012). We believe that further work is required to replicate these findings, assess the neurophysiological correlates of these effects, as well as the influence of other factors such as arousal, that are also known to affect behavioural performance (McGinley et al., 2015; Kuchibhotla et al., 2017), and ultimately provide a comprehensive account of how motor predictions and signal strength shape the perception of our environment.

STUDY I
SUPPLEMENTARY MATERIAL

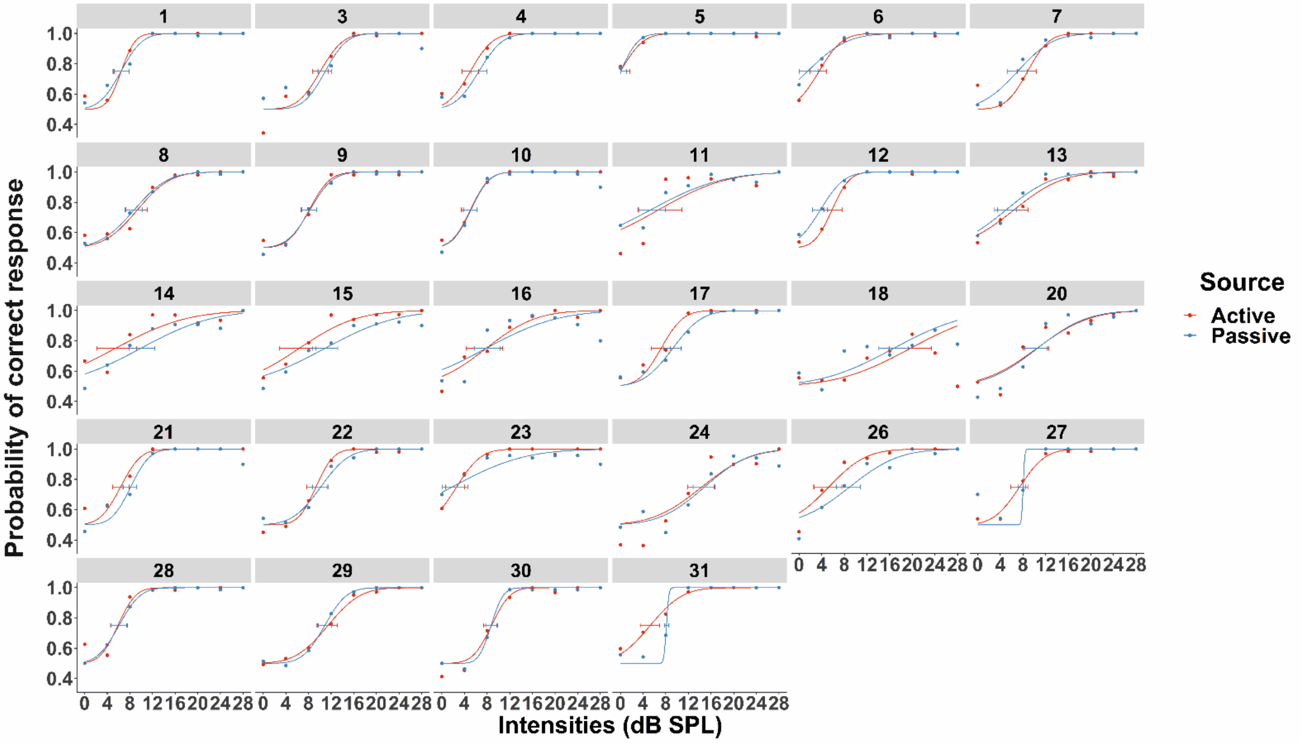


Figure S1. Psychometric functions for 28 participants from the detection task fitted to the percent correct responses as a function of sound intensity. Number in the legend above each plot corresponds to each participant's number (participants with numbers 2, 19, 25 were excluded; see Methods). The small horizontal segments represent the 95% confidence intervals for thresholds (parametric bootstrap procedure with n = 1000). The threshold is defined as the intensity accurately detected at 75% of the trials (as derived from the psychometric function fitted for each participant) and is represented by the intersection of the confidence interval with the psychometric function.

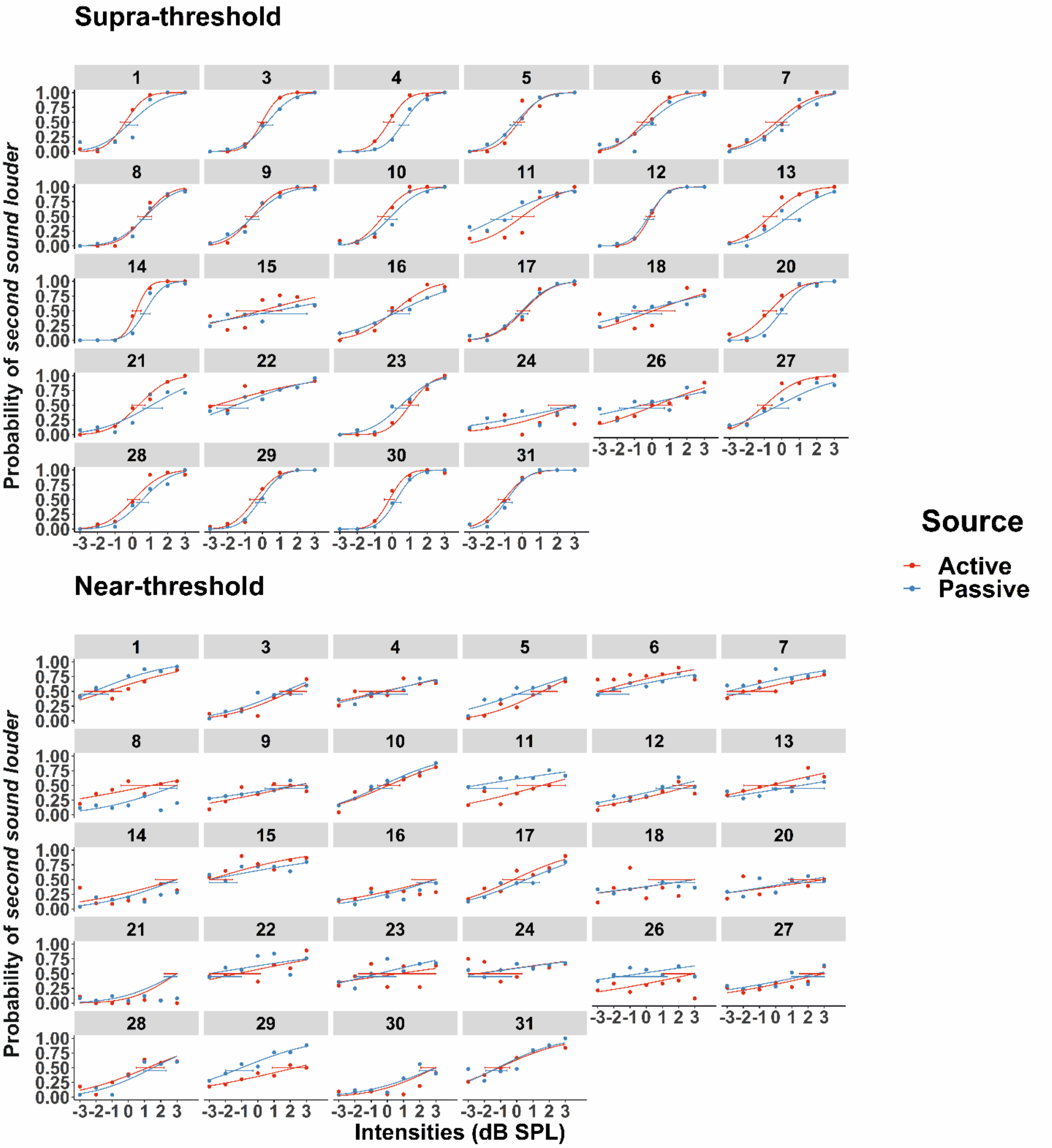


Figure S2. Psychometric functions for 28 participants from the discrimination task fitted to the probability of judging the comparison sound as louder as a function of its difference in dB from the first standard tone (±3 dB in steps of 1) for the supra- and the near-threshold intensities, respectively. Number in the legend above each plot corresponds to each participant's number (participants with numbers 2, 19, 25 were excluded; see Methods). The small horizontal segments represent the 95% confidence intervals (parametric bootstrap procedure with n = 1000) for the point of subjective equality (PSE), which is defined as the intensity, where the comparison stimulus was reported as louder than the standard one on 50% of the trials.

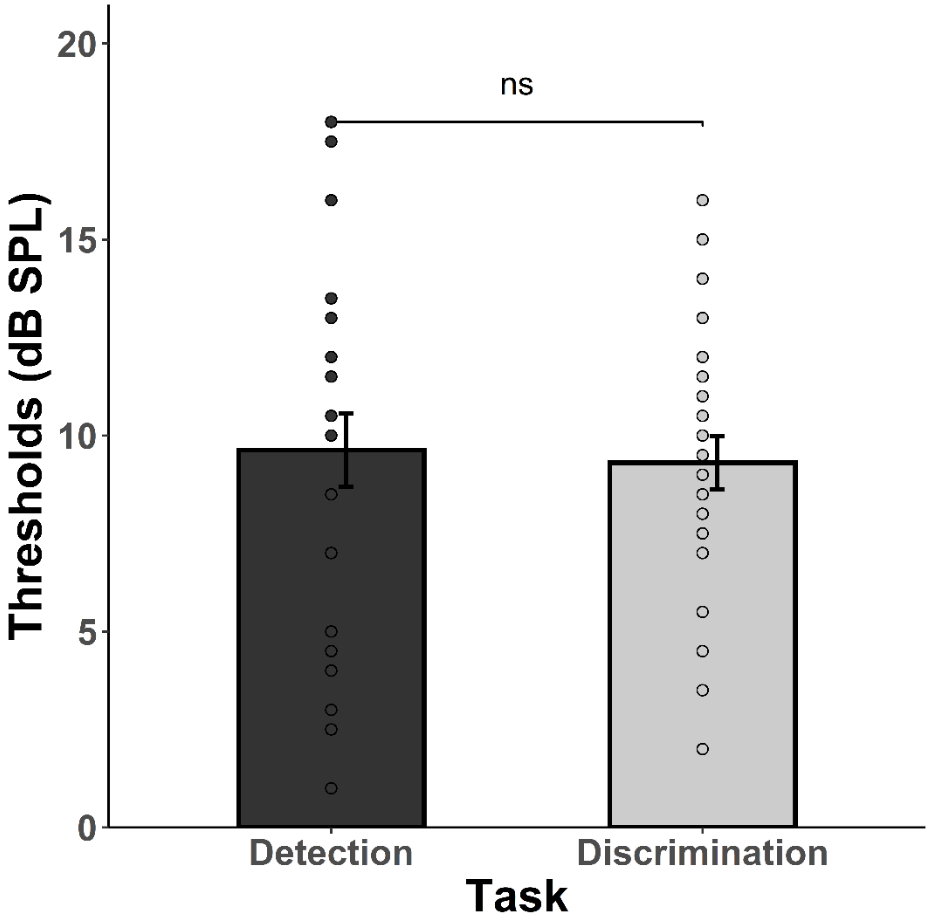


Figure S3. Mean audiometric thresholds prior to each task. From each audiometry, we obtained the thresholds for both the left and right ear. For all subjects, the thresholds were below 20 dB. Considering that in both tasks, we utilized a pure tone of 1000 Hz, in this analysis we only considered the thresholds for the 1000-Hz sounds. The mean thresholds across ears for each condition were introduced in a statistical analysis using a paired-sampled two-sided *t*-test to test for differences in audiometric thresholds prior to each task. The analysis did not show any significant differences ($M_{AM\_Detection}$ = 9.63, $M_{AM\_Discrimination}$ = 9.3, $SD_{AM\_Detection}$ = 4.94, $SD_{AM\_Discrimination}$ = 3.59, p > .050; Shapiro-Wilk normality test, p > .050).

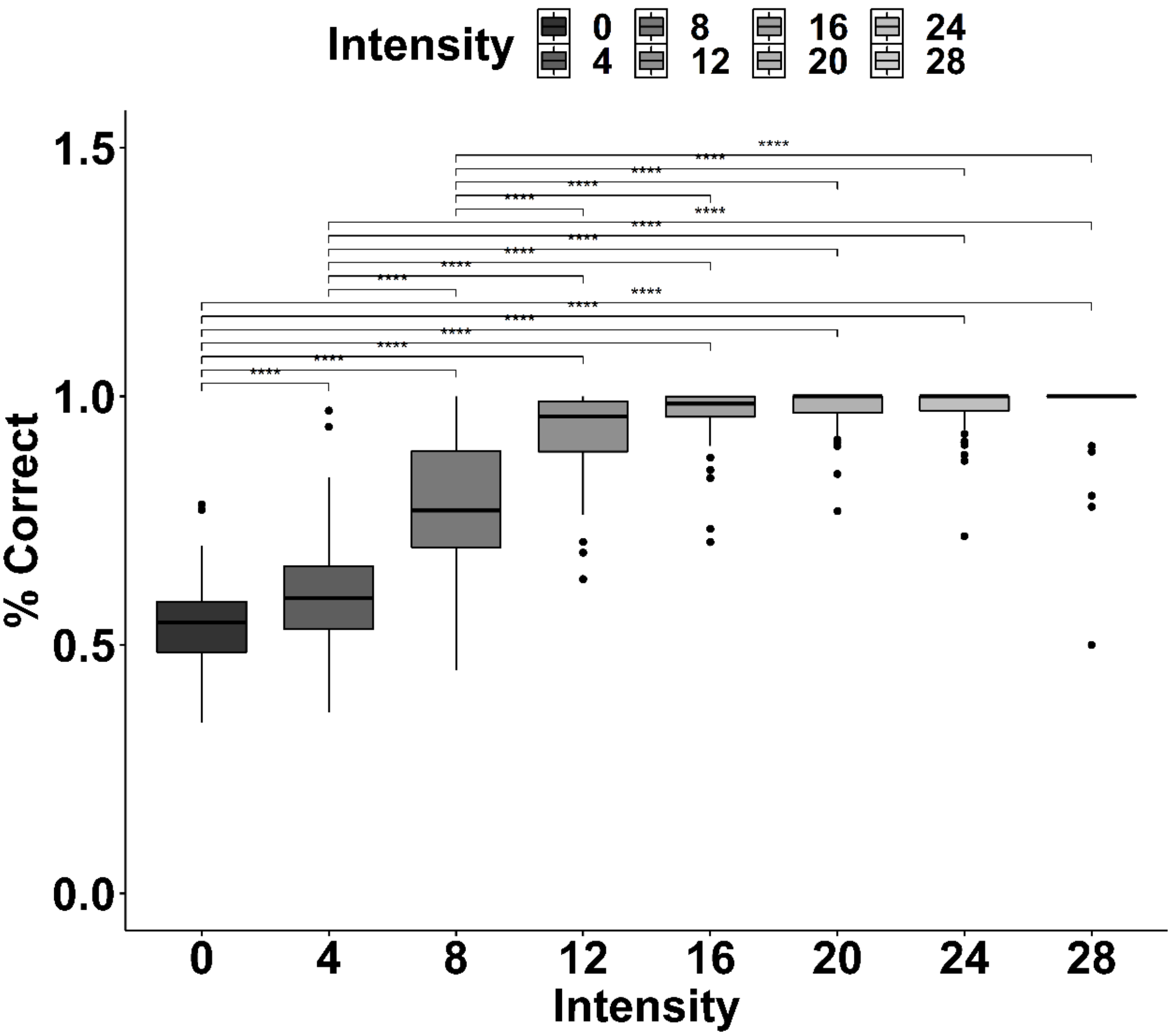


Figure S4. Repeated measures ANOVA with factors Intensity (0, 4, 8, 12, 16, 20, 24, 28) and Source (active and passive) on detection accuracy. The Greenhouse-Geisser correction was applied where sphericity was violated. The analysis did not show any significant main effect of source ($F(1,27) = 1.64$, $p > .050$), but we obtained a significant main effect of intensity, $F(2.35,63.44) = 228.79$, $p < .001$, $\eta_p^2 = .89$ and $\eta_G^2 = .78$. Specifically, irrespective of whether the sound was self- or externally-generated, participants' accuracy was significantly lower at 0 dBs compared to the rest of the intensities, at 4 dBs compared to the intensities above 8 dBs, at 8 dBs compared to the intensities above 12 dBs, and at 12 dBs compared to intensities above 16 dBs (all $p < .001$; $M_0 = 54.65$, $SD_0 = 8.85$, $M_4 = 61.19$, $SD_4 = 11.9$, $M_8 = 77.99$, $SD_8 = 13.33$, $M_{12} = 92.53$, $SD_{12} = 8.88$, $M_{16} = 96.47$, $SD_{16} = 6.03$, $M_{20} = 97.4$, $SD_{20} = 4.36$, $M_{24} = 97.37$, $SD_{24} = 4.79$, $M_{28} = 97.26$, $SD_{28} = 8.09$). Comparisons between higher intensities (i.e., 16 – 28 dBs) did not show any significant differences in participants' accuracy. The interaction between source and intensity did not reach significance ($F(4.10,110.68) = .62$, $p > .050$).

Table S1

Significant correlation for the values obtained between the two tasks (thresholds/slopes, PSE/JND values).

Correlations between mesures in		***r***	***p***
Detection	**Discrimination**		
slope_passive	JND_AN	0.502218293833961	0.006
slope_passive	JND_PS	0.458809161204315	0.014
slope_passive	JND_PN	0.458580361105521	0.014
slope_active	JND_PN	0.456946049747953	0.015
slope_active	JND_PS	0.447376137712047	0.017
slope_active	JND_AN	0.400560869306606	0.035
slope_passive	JND_AS	0.393510389324828	0.038
slope_active	JND_AS	0.37480689145703	0.049

Note. Significant correlations were only obtained between the slopes at the detection task and the JND values at the discrimination task for all conditions (A = Active, P = Passive, S = Supra-threshold, N = Near-threshold). No significant correlations between a) detection thresholds and PSE values, b) slopes at the detection task and PSE values at the discrimination task, c) detection thresholds and JND values at the discrimination task (all $p >$ .050).

STUDY II

Abstract

Actions modulate sensory processing by attenuating responses to self- compared to externally-generated inputs, which is traditionally attributed to stimulus-specific motor predictions. Yet, suppression has been also found for stimuli merely coinciding with actions, pointing to unspecific processes that may be driven by neuromodulatory systems. Meanwhile, the differential processing for self-generated stimuli raises the possibility of producing effects also on memory for these stimuli, however, evidence remains mixed as to the direction of the effects. Here, we assessed the effects of actions on sensory processing and memory encoding of concomitant, but unpredictable sounds, using a combination of self-generation and memory recognition task concurrently with EEG and pupil recordings. At encoding, subjects performed button presses that half of the time generated a sound (motor-auditory; MA) and listened to passively presented sounds (auditory-only; A). At retrieval, two sounds were presented and participants had to respond which one was present before. We measured memory bias and memory performance by having sequences where either both or only one of the test sounds were presented at encoding, respectively. Results showed worse memory performance – but no differences in memory bias –, attenuated responses, and larger pupil diameter for MA compared to A sounds. Critically, the larger the sensory attenuation and pupil diameter, the worse the memory performance for MA sounds. Nevertheless, sensory attenuation did not correlate with pupil dilation. Collectively, our findings suggest that sensory attenuation and neuromodulatory processes coexist during actions, and both relate to disrupted memory for concurrent, albeit unpredictable sounds.

1. Introduction

Forming predictions about upcoming events in the environment is crucial for all behaving organisms. A critical instance of such predictive processing is our ability to anticipate the sensory consequences of our own actions, which is essential for building a sense of self and shapes our perception of sense of agency (Gallagher, 2000). Although predictions have been suggested to facilitate perceptual processing in the wider sensory literature, in the action literature most studies report that the processing of predicted self-produced stimuli is attenuated (Schröger et al., 2015; Press et al. 2020), with only a few exceptions showing the opposite effect (e.g., Reznik et al., 2014; Eliades & Wang, 2003). Thus, several lines of research have shown that actions suppress the processing of the self-generated reafferent input (so-called self-generation effect) in a wide range of species (Chagnaud et al., 2015; Kelley & Bass, 2010; Kim et al., 2015; Requarth & Sawtell, 2011; Roy et al., 2001; Schneider et al., 2014) and irrespective of sensory modality (auditory; Baess et al., 2011; Horváth, 2013a; Horváth, 2013b; Martikainen et al., 2004; Mifsud & Whitford, 2017; SanMiguel et al., 2013; Saupe et al., 2013; Schafer & Marcus, 1973; Timm et al., 2013; Klaffehn et al., 2019, visual; Hughes & Waszak, 2011; Mifsud et al., 2018; Roussel et al., 2013; Roussel et al., 2014, and tactile; Blakemore, Wolpert, et al., 1998; Hesse et al., 2010; Kilteni et al., 2020). Nevertheless, the exact mechanisms underlying the suppression of sensory responses to self-initiated stimuli is still a matter of debate (for reviews see Schröger et al., 2015; Horváth, 2015; Hughes et al., 2013a). Interestingly, beyond modulating sensory responses, self-generation also appears to have consequences for memory encoding. The so-called "production effect" (MacDonald & MacLeod, 1998; Brown & Palmer, 2012) refers to memory benefits reported for stimuli that are self-generated in a predictive context; e.g., it is easier to remember a piano melody that was learnt by playing vs. listening to it (Brown & Palmer, 2012). Nevertheless, the bulk of the evidence for the production effect on memory comes from behavioral studies, and thus the underlying neurophysiological mechanisms remain largely unexplored. Crucially, given that memory relies on the sensory representation (e.g., Nyberg et al., 2000; Wheeler et al., 2000), the production effect could be a direct consequence of the differential sensory processing of self-initiated stimulation. However, to date, the possible relationship between the effects of self-initiation on sensory processing and memory has not been investigated. Here we sought out to bridge the gap between these two different research lines that have evolved separately over the last decades, aiming to identify the possibly shared neurophysiological mechanisms involved in each of these effects, focusing on the auditory modality. In the following paragraphs, we summarize findings that have inspired the present work, in an attempt to highlight the need of examining in detail the processes driving the self-generation effects and their possible links with the encoding of self-generated stimulation in memory.

1.1. Sensory processing of self-initiated stimuli

To date, most studies assessing the effects of actions on sensory processing, have attributed the attenuation effects to a predictive mechanism that predicts the sensory consequences of our actions (e.g., Bays et al., 2006; Blakemore, Wolpert et al., 1998; Martikainen et al., 2004). This view was inspired to a great extent by classic physiology research and the reafference principle (Sperry, 1950; von Holst & Mittelstaedt, 1950), and later by motor control theories that have further refined this idea by suggesting that sensory attenuation is an integral part of our motor abilities (Miall & Wolpert, 1996; Wolpert et al., 1995). This line of work was the first to point to the involvement of forward and inverse models in sensorimotor behaviour: The former estimate the future state of the system by comparing the predicted to the actual sensory consequences of the action, while the latter allow the system to estimate a motor plan (and its associated motor commands) so as to achieve a desired state (Miall & Wolpert, 1996; Wolpert et al., 1995). Especially forward models have been at the core of the dominant cancellation account that has been widely used to explain the self-generation effects (also known as the comparator model; Blakemore, Wolpert et al., 1998; Frith et al., 2000; Wolpert & Flanagan, 2001). According to this account, the suppression effects result from the operation of a forward model that generates stimulus-specific prediction signals before or during an action and sends them from the motor to the corresponding sensory cortices (Sperry, 1950; von Holst & Mittelstaedt, 1950). These motor-induced predictions of sensory reafference (i.e., corollary discharge) are compared to the sensory input generated by one's actions, and only the difference between the two (i.e., prediction error) is sent to higher stages of the neuronal hierarchy for further processing, ultimately suppressing the processing of the anticipated event in order to prioritize the most informative unexpected inputs (Friston, 2005). An important implication of the cancellation model is that the effects should be specific to the predicted stimulus, and thus mediated by sensory-specific areas (i.e., the effect should reflect attenuation of the predicted stimulus' representation in the sensory-specific areas).

In fact, there is mounting evidence showing that the attenuation effects for self-generated stimuli are (at least partly) stimulus-specific (Martikainen et al., 2004; Aliu et al., 2009; Houde et al., 2002; Hashimoto & Sakai, 2003; Heinks-Maldonado et al., 2005; Fu et al., 2006; Ott & Jäncke, 2013). Most studies supporting the specificity of the effects have employed the contingent self-generation paradigm, where neural responses to sounds generated by the participants in a fully predictable fashion are compared to the responses elicited by externally-generated sounds (e.g., Baess et al., 2009; Baess et al., 2011; Martikainen et al., 2004; Mifsud & Whitford, 2017) and have shown attenuated auditory N1 and P2 event-related potential (ERP) amplitudes (for a review see Schröger et al., 2015). Crucially, suppression is larger when the match between predicted and actual sensory input is more precise (Fu et al., 2006; Hashimoto & Sakai, 2003; Heinks-Maldonado et al., 2005; Houde et al., 2002; Baess et al., 2008), and it is suggested to occur within the auditory cortex (Martikainen et al., 2004; Aliu et al., 2009; Flinker et al., 2010), providing further support to the stimulus-specificity of the effects.

However, additional, stimulus-unspecific processes are also known to modulate sensory and perceptual responses during actions (Press et al., 2020; Press & Cook, 2015; Schröger et al., 2015; Korka et al., 2021). For example, there is evidence for generalized unspecific attenuation during movements (e.g., saccadic suppression and somatosensory gating, Crapse & Sommer, 2008; Ross et al., 2001; Chapman et al., 1987; Cohen & Starr, 1987; Williams et al., 1998; Williams & Chapman, 2000, 2002), which suggests that during actions the system might expect some action-related consequence, without necessarily generating a specific sensory prediction regarding the effect of the motor act. Supporting evidence to this idea comes also from studies showing suppression of responses when the stimulus merely coincides with an action (Hazemann et al., 1975; Makeig et al., 1996; Horváth, 2013a, 2013b; Horváth et al., 2012; Tapia et al., 1987), that is in the absence of a predictive relationship between action and stimulus. A recent study further examined the specificity of the attenuation effects, by assessing whether the self-generation effects (measured as auditory N1 attenuation) reflect a genuine modulation within the auditory cortex (SanMiguel et al., 2013). Based on evidence showing that the N1 response reflects the overlap of several components (Näätänen & Picton, 1987), SanMiguel and colleagues (2013) assumed that if the attenuation effects reflect stimulus-specific prediction mechanisms, then the effects should be observable in the sensory-specific components, namely the N1 at the mastoids (i.e., generated by tangentially oriented sources in the auditory cortex) and the "T complex" (i.e., the first and second negative peaks, known as Na and Tb, identified on anterior temporal sites, and generated by radial sources in the superior temporal gyrus; Tonnquist-Uhlen et al., 2003; Wolpaw & Penry, 1975). However, they showed that sensory attenuation mainly reflects the modulation of the unspecific N1 component, which is suggested to be the cortical projection of a reticular process facilitating motor activity, related to the orienting response (Näätänen & Picton, 1987). In contrast, they could not find a clear attenuation of the specific N1 components (cf. Timm et al., 2013; Horváth et al., 2012). Collectively, the findings reviewed this far point to a complex picture of possibly coexisting specific and unspecific effects of actions on sensory responses and suggest that the effects cannot be fully attributed to stimulus-specific prediction mechanisms.

As we have seen so far, converging evidence suggests that stimulus-unspecific processes might partly drive the sensory attenuation effects, thereby raising the need to identify the mechanism driving the unspecific attenuation during movement. One intriguing possibility is that the suppression effects may be mediated by a halo of neuromodulation surrounding actions, which would unspecifically gate auditory processing for stimuli presented close in time with the motor act. Neuromodulatory influences on the action-induced suppression effects seem plausible considering that rodent studies show that actions trigger a cascade of neuromodulatory processes (Vinck et al., 2015; Eggerman et al., 2014; McGinley et al., 2015), and that motor and neuromodulatory inputs overlap in auditory areas during movement (Nelson & Mooney, 2016; for a review see Schneider & Mooney, 2018). A possible candidate for creating a halo of neuromodulation that could mediate unspecific effects during movement could be the locus coeruleus norepinephrine system (LC-NE). This possibility has received substantial support recently, mainly by data showing a close association between pupil diameter – a proxy of LC-NE activity – (Aston-Jones & Cohen, 2005; Vinck et al., 2015; Murphy et al., 2014, Joshi et al., 2016) and actions in rodents (Lee & Margolis, 2016; McGinley et al., 2015; Vinck et al., 2015), monkeys (Bornert & Bouret, 2021), and humans (Yebra et al., 2019; Lubinus et al., 2021). However, to the best of our knowledge, there have been no attempts to test for possible links between sensory attenuation for self-generated sounds and neuromodulation (i.e., as reflected in pupil diameter) during actions.

1.2. Memory encoding of self-initiated stimuli

Meanwhile, actions might also affect other high-level processes beyond the immediate sensory processing. Strikingly, despite the mounting evidence pointing to a differential processing of self- and externally-generated stimuli (e.g., for a review see Schröger et al., 2015), but also to modulatory effects of movements on areas supporting memory processes (e.g., hippocampal and parahippocampal activity; Halgren, 1991; Mukamel et al., 2010; Rummell et al., 2016), there have been only few attempts to assess the effects of actions on memory encoding of self-generated stimulation. One research line – known as the "production effect" – has shown improved memory for self-produced stimuli compared to their passively listened comparisons (e.g., spoken words or played melodies compared to passively listened ones; MacDonald & MacLeod, 1998; Brown & Palmer, 2012), which has been attributed to the increased distinctiveness afforded by the extra mnemonic information of having produced these items that is not present for silently read words (Conway & Gathercole, 1987; Mama & Icht, 2016; Ozubko et al., 2012). This line of work, however, contrasts with the predictions made by predictive coding theories of memory. According to this account, memory is driven by the amount of surprise (i.e., prediction error) associated with an item, such that items eliciting larger prediction error responses (as reflected in larger evoked potentials or fMRI signal) should be encoded better in memory than the less surprising, predictable ones (Bar, 2009; Krawczyk et al., 2017; Pine et al., 2018; Greve et al., 2017; Heilbron & Chait, 2018; Henson & Gagnepain, 2010). This framework would, therefore, predict memory enhancements for the externally-generated sounds in a typical contingent paradigm where they inherently elicit larger prediction error (and enhanced sensory responses) compared to the more predictable self-generated stimuli.

1.3. The present study

To the best of our knowledge, there have been no attempts to simultaneously address the effects of self-generation on sensory processing and memory encoding of sounds and assess the possible relationship between these two phenomena and their underlying neurophysiological mechanisms. Based on the evidence indicating that self-generation effects might not be solely attributed to stimulus-specific motor predictions (e.g., Horváth et al., 2012; SanMiguel et al., 2013), we hypothesize that performing an action may trigger the activation of stimulus-unspecific neuromodulatory mechanisms, namely the LC-NE system. We hypothesize that this motor-driven noradrenergic activity may modulate the processing of sounds presented during the performance of the action, leading to suppressed sensory responses and altered memory encoding as a consequence of the latter.

In order to test these hypotheses, here we examine whether and how motor actions affect sensory processing and memory encoding of concomitant, but unpredictable sounds, by employing a combination of a self-generation and memory recognition task, while monitoring the brain's and the pupil's responses to sounds that are either presented passively or that coincide in time with a motor act. The aim of this study is twofold: Our first aim is to investigate the role of the neuromodulatory LC-NE system in the motor-driven modulation of auditory processing of self-generated sounds. Related to this first aim, we have specific hypotheses about the effects of actions on sensory responses and pupil diameter. As for the sensory responses, we hypothesize that event-related potentials (i.e., N1 at vertex but not at the mastoids, P2, and Tb) should be attenuated for sounds coinciding with an action (cf. Horváth et al., 2012; Hazemann et al., 1975; Makeig et al., 1996; Horváth, 2013a, 2013b). As for the pupil diameter, we hypothesize that neuromodulatory activity (i.e., reflected in pupil diameter; Aston-Jones & Cohen, 2005) should increase during actions (cf. Lee & Margolis, 2016; McGinley et al., 2015; Vinck et al., 2015; Simpson, 1969; Yebra et al., 2019) and that it should correlate with the sensory attenuation effects measured on the auditory event-related potentials. Our second aim is to assess whether the differential sensory processing of stimuli paired with an action affects their encoding in memory. We expect to observe differences in memory performance between passively encoded sounds and sounds that coincide with an action at encoding. However, given the lack of contingency between actions and sounds in the present paradigm as compared to the typical production effect studies, as well as the mixed evidence (memory for self-initiated stimulation is enhanced in previous production effect studies but should be reduced according to predictive coding views; MacDonald & MacLeod, 1998; Brown & Palmer, 2012; Henson & Gagnepain, 2010), we remain agnostic as to the direction of the effect. However, critically, we hypothesize that the potential differences in the memory encoding of sounds presented with or without a concomitant action should be driven by, and thus correlate with, the differential neurophysiological responses (i.e., event-related potentials and pupil diameter) at encoding for sounds that were either paired with an action or not.

2. Methods

All the scripts for the experimental stimulation and data analysis are available on Open Science Framework, along with the detailed experimental protocol:
https://osf.io/238xe/?view_only=4b6d8fdc2a2f4982bac76a72dc78e0ec.

2.1. Participants

Twenty-six healthy, normal-hearing subjects, participated in the present study. Participants were typically undergraduate university students at the University of Barcelona. Data from three participants had to be excluded due to technical problems, inability to comply with the task instructions, or excessive artifacts in the EEG recording, leaving data from twenty-three participants (6 men, 17 women, M_{age} = 24.82, age range: 21-36). None of them had any hearing impairments, suffered, or had suffered from psychiatric disorders or had taken substances affecting the central nervous system the 48 hours prior to the experiment. All participants gave written informed consent for their participation after the nature of the study was explained to them and they were monetarily compensated (10 euros per hour). Additional materials included a personal data questionnaire, a data protection document, and five personality trait questionnaires. The study was accepted by the Bioethics Committee of the University of Barcelona.

2.2. General experimental design

Each trial consisted of three phases: the encoding phase, the retention phase, and the retrieval phase (Fig.1).

Encoding phase: At the start of each trial, subjects were presented with a row of six vertical lines on the screen, separated in semi-random distances from each other. The positions of vertical lines were distributed based on the sequence presented in each trial. During the whole duration of the encoding period (12 seconds), a horizontal line moved at a stable pace across the screen from left to right, intersecting each of the vertical lines as it advanced. Participants pressed a button with their right thumb every time the horizontal line reached one of the vertical ones. Only half of these presses produced a sound (Motor-auditory event; MA). The other half did not result in the presentation of a sound (Motor-only event; M). Additionally, three more sounds were presented passively to the participants without being triggered by a button press (Auditory-only event; A). Thus, in every trial, the encoding set consisted of six different sounds to be remembered, delivered within nine events (three motor-only (M), three Motor-auditory sounds (MA), and three Auditory-only (A) sounds). The interval between any two events (MA, M or A) varied from 0.8 s to 2.4 s, in steps of 0.05 s, while the interval between any two sounds varied between 1.6 and 2.4 seconds in steps of 0.05. The encoding phase finished when the horizontal line had intersected all the vertical ones and arrived at the right of the screen. If the task was performed correctly (i.e., all required button presses were performed), the trial continued into the retention phase. Otherwise, an error message appeared on the screen indicating that the participant did not press the button every time the horizontal line reached a vertical one, and a new trial began.

Retention phase: During the subsequent retention phase, participants were presented with a fixation cross on the screen for 3 s and they were asked to remember the six individual sounds that had been presented in the encoding phase.

Retrieval phase: In the retrieval phase, participants were presented with two test sounds with a 2 s sound-to-sound onset asynchrony (indicated by the visual stimuli “Sound 1” and “Sound 2”, respectively). Subsequently, a question mark appeared on the screen, prompting participants to respond whether the first or the second test sound was presented during the encoding phase. We employed two different types of sequences (see Section 2.3.1. Sequences for more details) that differed only in this retrieval phase: “Two Test Sounds at Encoding” sequences (henceforth 2T; Figure 1, left panel) and “One Test Sound at Encoding” sequences (henceforth 1T; Figure 1, right panel). Unbeknownst to the participants, in the 2T sequences, both test sounds had been presented at encoding, while in the 1T sequences only one of the test sounds had been presented at encoding. Nevertheless, participants made a forced choice between the two sounds on every trial. The response window was 2 seconds. After the end of the response window, a fixation was presented for 2 seconds (inter-trial interval) before the start of the next trial.

2.3. Stimuli

2.3.1. Sequences

Two types of sequences were created, differing in whether both or only one of the test sounds presented at retrieval were also present during the encoding phase. In the "Two Test Sounds at Encoding" (2T; Figure 1, left panel) and unbeknownst to the participants, the two test sounds presented passively at retrieval were also presented in the encoding sequence, one as a motor-auditory (Encoded as MA) and the other one as an auditory-only event (Encoded as A). These sequences were intended to measure memory bias. In the "One Test Sound at Encoding" (1T; Figure 1, right panel), only one of the test sounds at retrieval was presented at encoding, either as a MA (Encoded as MA) or as an A event (Encoded as A), while the other sound was not presented at encoding (New sound). The 1T sequences were intended to measure memory performance. They were introduced only for the behavioural data and were not used for the EEG and pupillometry analyses. This design allowed us to have enough trials for Encoded as A and Encoded as MA sounds at retrieval, keep the experiment's duration within a reasonable time, and obtain an additional objective measure of memory performance in the 1T sequences besides the measure of memory bias obtained in the 2T sequences. Five of the 1T sequences were randomly chosen to be used during the practice block.

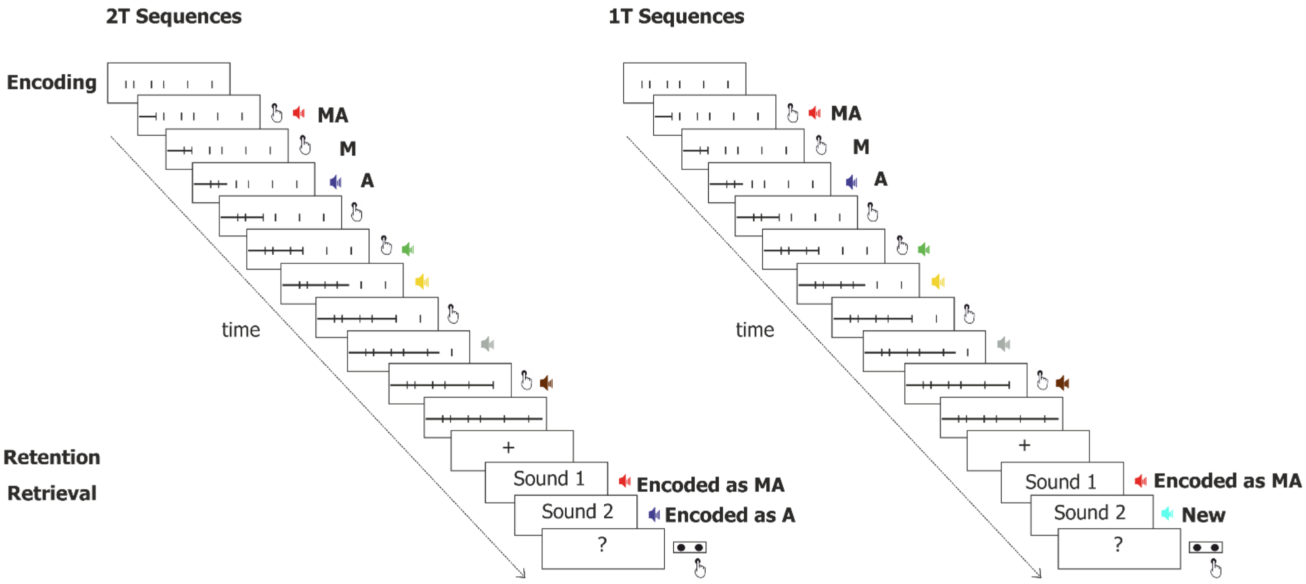


Figure 1. Schematic representation of the experimental design, showing an example trial for the two types of sequences employed: 2T sequences (left) and 1T sequences (right). Each trial consisted of three phases: encoding, retention, and retrieval. The two types of sequences differed only in the retrieval phase. The different boxes represent the visual stimulation as a function of time. Each trial started with the encoding phase, where six vertical lines were initially presented (top box), and subsequently a horizontal line started moving across the screen from left to right, intersecting each of the six vertical lines as it advanced. Participants were asked to press a button every time the horizontal line reached one of the vertical ones. Only half of these presses produced a sound (Motor-auditory; MA). The other half did not result in the presentation of a sound (Motor-only; M). Additionally, three more sounds were presented passively to the participants without being triggered by a button press (Auditory-only condition; A). Therefore, six different sounds (shown by the

different colours of the sounds in the figure) were presented during encoding and had to be maintained in memory for a 3s retention period (box with fixation cross). In the retrieval phase, participants were presented with two sounds, indicated by the visual cues "Sound 1" and "Sound 2". In the 2T sequences (left), the sounds at retrieval were both presented at encoding, one Encoded as MA and the other Encoded as A. In the 1T sequences (right), only one of the two sounds was presented at encoding that was either Encoded as A or Encoded as MA (in the figure an "Encoded as MA" sound is shown), while the other sound was not presented at encoding (New). After the presentation of the test sounds, a question mark appeared on the screen, prompting participants to respond whether the first or the second test sound was presented during the encoding phase.

Importantly, the same sounds in the same encoding sequence positions were used as either A or MA in different trials, which allowed us to compare between physically identical auditory sequences that only differed in the actions performed. Additionally, we counterbalanced the order of the sounds at encoding that would be later used as test at retrieval for the 2T sequences, the order of the two retrieval sounds, and the position of the test sounds in the encoding sequence. Related to the latter, we discarded the first and last position of the encoding sequence for placing test sounds to avoid primacy and recency effects, which refer to an improvement in memory retention for stimuli that have been presented first or last in a list, respectively (Mondor & Morin, 2004). However, we included 20 catch trials with test sounds in those positions, which were randomly interleaved with the experimental sequences described above and discarded from all analyses.

2.3.2. Auditory stimuli

For the main experiment, 255 different, environmental, natural, complex, and non-identifiable sounds were gathered from the libraries of McDermott Sound Library (http://mcdermottlab.mit.edu/svnh/Natural-Sound/Stimuli.html) and Adobe (https://offers.adobe.com/en/na/audition/offers/audition_dlc.html). These sounds were then edited to all have 250 ms of duration, a sampling rate of 44.1 kHz and to be played at 16 bits, mono and with 70 dB of intensity. Subsequently, six volunteers that did not participate in the main experiment rated the 255 sounds based on their identifiability (i.e., how easy it was to assign a name to them). All sounds were presented to them in a randomized order and each sound was presented twice. The volunteers rated them in a scale from 1-3 (1 = identifiable, 2 = not sure, 3 = not identifiable), and the mean score for each sound was calculated. The 108 less identifiable sounds were selected to construct the unique experimental sound sequences. The sounds used in the practice block consisted of 35 pure tones of different frequencies, ranging from 300 Hz to 3700 Hz in steps of 100.

2.4. Apparatus

The visual stimuli were presented on an ATI Radeon HD 2400 monitor. The auditory stimuli were presented via the Sennheiser KD 380 PRO noise cancelling headphones. To record participants' button presses at encoding (right hand button press) and behavioural responses at retrieval (left hand button presses), we used the Korg nanoPAD2. The buttons of this device do not produce any mechanical noise when pressed, and, thus, do not interfere with our auditory stimuli. The presentation of the stimuli and recording of participants' button presses and responses were controlled using MATLAB R2017a (The Mathworks Inc., 2017), the Psychophysics Toolbox extension (Brainard, 1997; Kleiner et al., 2007), and the Eyelink add-in toolbox for eyetracker control.

EEG activity was acquired with Neuroscan 4.4 software and Neuroscan SynAmps RT amplifier (NeuroScan, Compumedics, Charlotte, NC, USA). We recorded continuously with Ag/AgCl electrodes from 64 standard locations according to the 10% extension of the International 10–20 system (Chatrian et al., 1985; Oostenveld & Praamstra, 2001) mounted in a nylon cap (Quick-Cap; Compumedics, Charlotte, NC, USA). An additional electrode was placed at the tip of the nose (serving as online reference). The vertical electrooculogram (EOG) was measured with two electrodes placed above and below the left eye, and the horizontal EOG with two electrodes placed on the outer canthi of the eyes referenced to the common reference (unipolar montage). The ground electrode was placed at AFz. All impedances were kept below 10 kΩ during the whole recording session and data was sampled at 500 Hz.

Concurrently with the EEG recording, horizontal and vertical gaze position, as well as the area of the pupil, were recorded using EyeLink 1000 desktop mount (SR Research, sampling rate: 1,000 Hz; left eye recordings except for three participants for whom the right eye was recorded instead). The pupil was assessed in the centroid mode of the eye tracker, which uses a center-of-mass algorithm. This algorithm detects the pupil area by identifying the number of black pixels and its center on the video image. Importantly, in contrast to methods using ellipse fitting for the measurement of the pupil, this method is hardly affected by noise (S-R Research Eyelink-CL Manual, p. 71).

2.5. Procedure

Prior to the start of the experiment, participants were asked to complete several questionnaires and were given written and verbal instructions about the task. Specifically, they were told that at the start of every trial they would first see six vertical lines and that a horizontal line would start to move from left to right, intersecting each vertical line as it advanced. They were explicitly instructed to press the predefined button every time the horizontal line crossed one of the vertical ones (not too early / late and no more than one button press per vertical line). They were told that they would hear several sounds while the line advanced, some of them might coincide with the button press and some of them not and that they should try to memorize all the sounds presented because later they would be tested in memory. Finally, they were told that once the horizontal line had crossed all the vertical lines, a fixation cross would appear and subsequently two sounds would be presented (indicated by the visual stimuli “Sound 1” and “Sound 2”, respectively) and that they would have only two seconds to reply which one of the two sounds was presented during the encoding period of the trial. They were asked to make a choice on each trial between the two test sounds within the response window.

After explaining them the instructions, participants were seated in an electrically and acoustically shielded room and were asked to place their head on a chinrest at approximately 60 cm from the screen. Eye-tracker calibration was performed first at the start of the experiment and then every six blocks. In order to familiarize themselves with the task, participants first completed a practice block of 5 trials and repeated it as many times as they needed to make sure they understood how to perform the task. During the main experiment, participants completed a total of 236 trials: 56 1T trials, 160 2T trials and 20 catch trials. These were divided in 24 blocks, 20 of them comprised of 10 trials (9 experimental and 1 catch trial) and the remaining 4 comprised of 9 trials (all of them experimental trials). At the end of each block, a message appeared informing participants about the number of errors (i.e., not pressing the button when required) and extra-presses (i.e., more than the required button presses) at the encoding phase, as well as the number of missed responses at retrieval for this block. Catch trials, as well as trials including errors in button-pressing and missed responses were discarded from further analyses. Participants took a break of approximately 5 minutes every six blocks to prevent fatigue. The experiment lasted for approximately 1.5 hours excluding the EEG preparation.

2.6. Data analysis

2.6.1. Behavioral analysis

To test for differences in memory bias (2T sequences) and memory performance (1T sequences) for sounds encoded as A or MA, we performed two different analyses. For the 1T sequences, we calculated the percent correct for the sounds at retrieval (i.e., memory performance), separately for those that were Encoded as MA and Encoded as A, which was subsequently submitted to a two-sided paired sample *t*-test. For the 2T-trial sequences, we calculated the percent recall for sounds Encoded as MA and Encoded as A and tested for differences in memory bias, using a two-sided paired samples *t*-test. We complemented the frequentist *t*-tests with corresponding Bayesian *t*-tests, separately for the 1T and 2T sequences. For both Bayesian comparisons, the Bayes factor (BF_{10}) for the alternative hypothesis (i.e., the difference of the means is not equal to zero) was calculated (using the function *ttestBF* of the BayesFactor package in R). Specifically, the null hypothesis was specified as a point-null prior, corresponding to a standardized effect size $\delta = 0$, and the alternative hypothesis was defined as a Cauchy prior distribution centered around 0 with a scaling factor of $r = 0.707$ (Rouder et al., 2009). In line with the Bayes Factor interpretation (Lee & Wagenmakers, 2013) and with previous work reporting Bayes Factors (Korka et al., 2019; Korka et al., 2020; Marzecová et al., 2018), data were taken as moderate evidence for the alternative hypothesis if the BF_{10} was greater than 3, while values close to 1 were considered only weakly informative. Values greater than 10 were considered strong evidence for the alternative (or null) hypothesis.

2.6.2. EEG preprocessing

EEG data was analyzed with EEGLAB (Delorme & Makeig, 2004) and plotted with EEProbe (ANT Neuro). Data were high-pass filtered (0.5 Hz high-pass, Kaiser window, Kaiser β 5.653, filter order 1812), manually inspected so as to reject atypical artifacts and identify malfunctioning electrodes, and corrected for eye movements with Independent Component Analysis, using the compiled version of runica (binica) that uses the logistic infomax ICA algorithm (Onton & Makeig, 2006). Components capturing eye movement artifacts were rejected by visual inspection and the remaining components were then projected back into electrode space. Data was then low-pass filtered (30 Hz low-pass, Kaiser window, Kaiser β 5.653, filter order 1812), remaining artifacts were rejected by applying a 75 µV maximal signal-change per epoch threshold, and malfunctioning electrodes were interpolated (spherical interpolation). A −100 to +500 ms epoch was defined around each event both at the encoding and the retrieval phase. The data was subsequently baseline corrected (100 ms prior to the event). We calculated the average wave for each event of interest, as well as the grand average for the whole sample. Specifically, we obtained the averages for the MA, A, and M events at encoding, while for the retrieval data, we binned the responses to motor-auditory and auditory-only sounds as a function of memory (i.e., Encoded as MA and Encoded as A at retrieval that were remembered or forgotten). For each condition of interest the number of remaining trials used for the analyses after trial rejection were: Auditory-only (*M* = 424.9, *SD* = 46.9), Motor-auditory (*M* = 427.2, *SD* = 40.6), Motor-only (*M* = 429, *SD* = 40.8), Encoded as A and forgotten (*M* = 68, *SD* = 11.7), Encoded as A and remembered (*M* = 64, *SD* = 14.7), Encoded as MA and forgotten (*M* = 64.1, *SD* = 14.2), Encoded as MA and remembered (*M* = 67.7, *SD* = 11.9).

To assess self-generation effects at encoding, MA sound responses were corrected for motor activity by subtracting the motor-only (M) averages from the motor-auditory (MA) averages, since the signal obtained in the MA condition represents the brain signal elicited by the sound, but also by the planning and execution of the finger movement to press the button. We, therefore, obtained a motor-corrected wave that only included the brain signal elicited by the MA sound. Self-generation effects at encoding were then assessed by comparing responses to MA sounds corrected for motor activity (MA–M) with the responses elicited by the auditory-only sounds (A). Self-generation effects are presented in all figures as the difference wave between the motor-auditory (corrected) sound responses and the auditory-only sound responses (A–[MA–M]). No motor correction was performed at retrieval since both test sounds were presented passively.

2.6.3. ERP analysis

For all the effects of interest at encoding, we examined responses separately for the N1 and P2 at Cz (N1, P2) and at the mastoids (henceforth, $N1_{mast}$, $P2_{mast}$), the P3 component at Pz, and the N1 subcomponents Na and Tb at T7 and T8. The same components except for P3 were examined at retrieval. The windows were defined after visual inspection of the data by locating the highest negative or positive (depending on the component of interest) peak in the usual latencies for each component as reported by previous work (SanMiguel et al., 2013). Specifically, time windows for N1 (and $N1_{mast}$), P2 (and $P2_{mast}$), Na, and Tb were defined on the grand-averaged waveforms of the auditory-only sounds as previously reported (e.g., SanMiguel et al., 2013). Na and Tb were identified as the first and second negative peaks, respectively, identifiable after sound onset on electrodes T7 and T8, as recommended by Tonnquist-Uhlen et al. (2003). $N1/N1_{mast}$ and $P2/P2_{mast}$ were identified as the negative and positive peaks occurring in the window ~70 to 150 ms, and ~150 to 250 ms after stimulus onset on Cz, respectively, showing reversed polarity at the mastoid electrodes. P3 at encoding was identified as the peak of the difference wave (A – [MA-M]) in the P3 window range based on previous work (e.g., Baess et al., 2008). The time windows for the $N1/N1_{mast}$, $P2/P2_{mast}$, P3, Na, and Tb peaks were centered on the identified peaks ± 13, 25, 15, 10, and 15 ms, respectively and were the following: Na 72–92 ms, Tb 120–150 ms, $N1/N1_{mast}$ 94–120 ms, P2/ $P2_{mast}$ 174–224 ms, P3 256–286 ms. Given variations in peak latencies across the conditions, the width of the windows was defined such that it could capture the peak of the MA sound waveform as well, and it was proportional to the width of the component. For the encoding data, we performed paired samples *t*-tests with the factor Sound Type (A vs. MA) to test for differences in N1, P2 and P3, and a repeated measures ANOVA with factors Sound Type (A vs. MA) x Laterality (M1 vs. M2 or T7 vs. T8) to test for differences in $N1_{mast}$, $P2_{mast}$ and Na, Tb, respectively. For the retrieval data we performed 2x2 ANOVAs with the factors Sound Type (Encoded as A vs. Encoded as MA) and Memory (Remembered vs. Forgotten) on N1 and P2, while for the $N1_{mast}$, $P2_{mast}$, Na, and Tb an additional factor Laterality was introduced in the ANOVAs (i.e., M1 vs. M2 or T7 vs. T8).

2.6.4. Pupillometry analysis

Missing data and blinks, as detected by the EyeLink software, were padded by 100 ms and linearly interpolated. Additional blinks were found using peak detection on the velocity of the pupil signal and linearly interpolated (Urai et al., 2017). Blinks separated by less than 250 ms were aggregated to a single blink. The interpolated pupil time series were bandpass filtered using a 0.05–4 Hz third-order Butterworth filter. Low-pass filtering reduces measurement noise not likely to originate from physiological sources, as the pupil functions as a low-pass filter on fast inputs (Binda et al., 2013; Hoeks & Levelt, 1993). High-pass filtering removes slow drifts from the signal that are not accounted for by the model in the subsequent deconvolution analysis. First, we estimated the effect of blinks and saccades on the pupil response through deconvolution and removed these responses from the data using linear regression using a procedure detailed in previous work (Knapen et al., 2016; Urai et al., 2017). The residual bandpass filtered pupil time series was used for the evoked analyses (van Slooten et al., 2019). After zscoring per trial, we epoched trials (epoching window -0.5 to 1.5 post-event), baseline corrected each trial by subtracting the mean pupil diameter 500 ms before onset of the event and resampled to 100 Hz.

For each participant, we first obtained the average evoked response for the main events of interest. Specifically, we obtained the averages for the A and MA events at encoding, while at retrieval we obtained the averages for the Encoded as A and Encoded as MA sounds, separately for the remembered and the forgotten ones. We used non-parametric permutation statistics to test for the group-level significance of the individual averages, separately for encoding and retrieval. Specifically, we computed *t*-values of the difference between the two conditions of interest and thresholded these *t*-values at a *p*-value of 0.05. Each cluster was constituted by the samples that passed the threshold of the *p*-value. The cluster statistics was chosen as the sum of the paired *t*-values of all the samples in the cluster. First, we compared the pupil response to motor-auditory and auditory-only events at encoding. For the retrieval data, we aimed to test for possible main effects of Sound Type (Encoded as A vs. Encoded as MA) and Memory (Remembered vs. Forgotten), as well as for possible interactions between the two. For the main effects of Sound Type and Memory at retrieval, the permutation statistics were performed between Encoded as A and Encoded as MA sounds (irrespective of their memory) and between Remembered and Forgotten sounds (irrespective of how they were encoded before), respectively. To test for possible interactions, the cluster-permutation test was performed on the difference waves ([Encoded as A and remembered – Encoded as MA and remembered] and [Encoded as A and forgotten – Encoded as MA and forgotten]). For each statistical test, this procedure was performed by randomly switching labels of individual observations between these paired sets of values. We repeated this procedure 10,000 times and computed the difference between the group means on each permutation. The obtained *p*-value was the fraction of permutations that exceeded the observed difference between the means (i.e., two-sided dependent samples tests). The pupil preprocessing and analysis was performed with custom software based on previous work (Urai et al., 2017) using Fieldtrip (Oostenveld et al., 2011).

2.6.5. Correlations

Finally, we hypothesized that the electrophysiological and neuromodulatory effects at encoding (i.e., sensory suppression and pupil dilation for MA events) might be driving any memory encoding differences between A and MA sounds, and that neuromodulation might be behind the suppression of ERP responses to MA sounds. To assess these relationships, we tested for possible correlations between the behavioural, electrophysiological and neuromodulatory (i.e., pupil diameter) effects of actions. Only those differences between MA and A events that were found to be significant in the previous analyses were introduced in the correlation analyses. For all the behavioural and the electrophysiological effects, we first calculated the difference by subtracting the MA from A values (i.e., difference in memory and ERP amplitude for each component of interest between A and MA). Regarding the ERPs identified in two electrodes (e.g., Na, Tb, $N1_{mast}$, $P2_{mast}$), we calculated the mean amplitude across the two (T7/T8 and M1/M2, respectively). For the pupil data, we used the peak of the difference wave between A and MA events at encoding. We then submitted these values to a Pearson correlation coefficient to test for correlations between a) the effects on ERPs at encoding and memory performance/bias (1T and 2T sequences, respectively), b) the neuromodulatory effects at encoding and memory performance/bias (1T and 2T sequences, respectively), and c) the effects on the ERPs and the neuromodulatory effects at encoding. In all correlations, for the ERPs, the larger the attenuation effects for the negative (N1, $P2_{mast}$, Na, Tb) and positive ($N1_{mast}$, P2, P3) components, the more negative and positive the values, respectively. Conversely, for the pupil and the behavioural data, the more negative the value, the larger the pupil diameter and the worse the memory performance for MA sounds.

3. Results

All statistical analyses were performed using R (version 3.6.0). For all the *t*-tests performed, we first confirmed that the assumption of normality was not violated (Shapiro–Wilk normality test $p > .05$). As we mentioned before (see Methods), the 1T sequences were introduced to be used only for the behavioural analyses. For the EEG and pupil analyses, we only included the data from the 2T sequences, after confirming that the results would remain the same when including the 1T sequences as well.

3.1. Behavioural performance

For the analysis of the behavioural data, we calculated the percent correct (i.e., memory performance in the 1T sequences) and the percent recall (memory bias in the 2T sequences) for sounds that were encoded as motor-auditory or auditory-only (see Figure 2). For the 1T sequences, we obtained significantly better memory performance for sounds that were encoded as auditory-only compared to those that coincided with participants' motor acts in the previous encoding phase, $t(22) = 3.15$, $p = .005$, $d = 0.66$ (M_{MA}= .757, M_{A} = .799, SD_{MA} = .108, SD_{A} = .0924). This difference, however, was not reflected in memory bias since we did not find significant differences between the Encoded as A and Encoded as MA sounds in the 2T sequences, where both of the test sounds were presented at encoding, $t(22) = 1.14$, $p = .267$ (M_{MA} = .509, M_{A} = .491, SD_{MA} = .0395, SD_{A} = .0395). The absence of significant differences in memory bias may suggest that they remembered both sounds as evident by the generally high accuracy (i.e., mean performance in the 1T sequences = 0.78 with standard deviation of 0.1) which led them to choose randomly between A and MA sounds in 2T sequences. We complemented the frequentist *t*-tests with corresponding Bayesian *t*-tests, separately for memory performance (1T sequences) and memory bias (2T sequences). The Bayesian *t*-tests for the 1T and 2T sequences yielded similar results as the ones obtained from the frequentist *t*-tests. Specifically, this analysis brought strong evidence for the alternative hypothesis in the case of 1T sequences (BF_{10} = 9.375), while the Bayesian *t*-test for the 2T sequences, brought weak evidence for the alternative hypothesis (BF_{10} = 0.389).

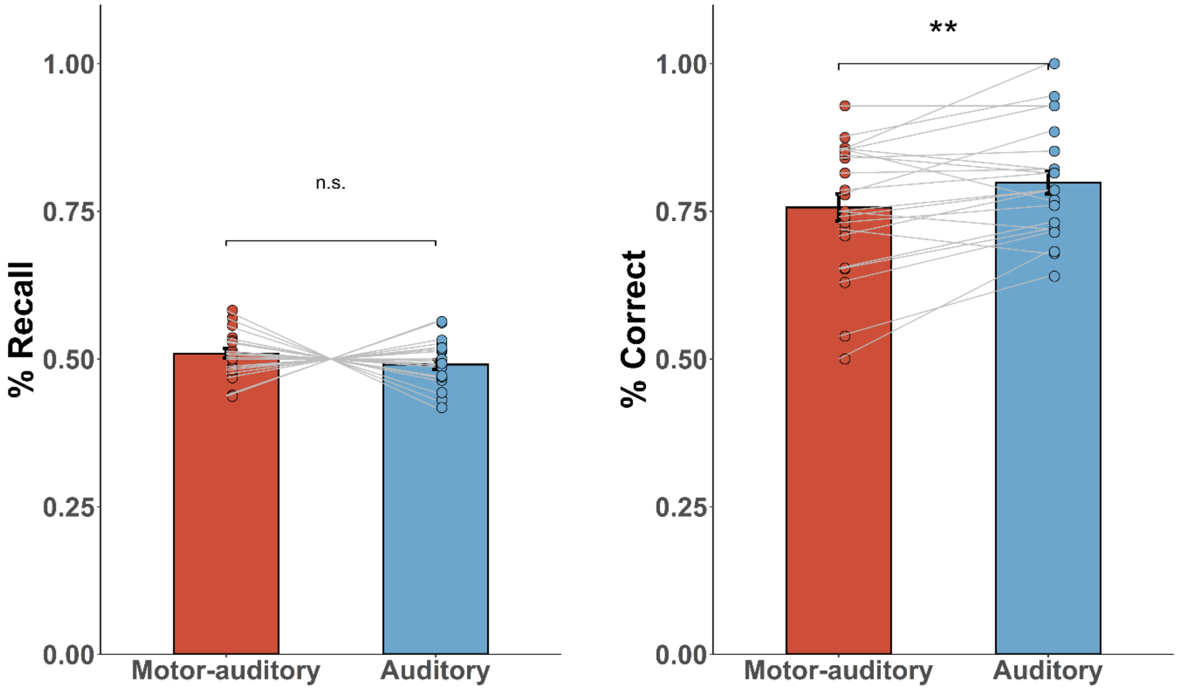


Figure 2. Summary of the behavioural results, separately for memory bias in the 2T sequences (left) and memory performance in the 1T sequences (right). Error bars depict standard error of the mean. Gray lines connect the data points of each subject, showing the response (% Recall and % Correct, respectively) to MA and A sounds for each individual. For memory bias (i.e., percent recall in 2T sequences), there were no significant differences between motor-auditory and auditory-only sounds (two-tailed paired samples *t*-test, $p > .050$, $M_{MA} = .509$, $M_{A} = .491$, $SD_{MA} = .0395$, $SD_{A} = .0395$), in line with the Bayesian analysis that provided weak evidence for the alternative hypothesis ($BF_{10} = 0.389$). For memory performance (i.e., percent correct in 1T sequences), there was a significant difference between motor-auditory and auditory-only sounds (two-tailed paired samples *t*-test, $t(22) = 3.15$, $p = .005$, $d = 0.66$; indicated by two asterisks), with higher accuracy for the latter ($M_{MA}= .757$, $M_{A} = .799$, $SD_{MA} = .108$, $SD_{A} = .0924$), which was also supported by the Bayesian analysis that brought strong evidence in favor of the alternative hypothesis ($BF_{10} = 9.375$).

3.2. Electrophysiological responses at encoding

Figure 3a shows all the studied peaks identified on the passive sound responses for the encoding conditions at the relevant electrodes for each peak. The motor-auditory sounds at encoding were motor corrected (see Methods). The time windows defined for each peak were the following: Na 72–92 ms, Tb 120–150 ms, N1/N1$_{mast}$ 94–120 ms, P2/ P2$_{mast}$ 174–224 ms, P3 256–286 ms.

First, we performed a one-sided *t*-test to test for possible differences in N1 amplitude between A and MA sounds at encoding, with the hypothesis of attenuated responses for the latter. Indeed, we obtained a significant attenuation for the N1, $t(22) = -1.89$, $p = .036$, $d = -0.39$, with lower amplitudes for sounds that coincided with a motor act, compared to those that were passively presented to the participants (Figure 3a-b, see Table 1 for all the mean amplitudes per condition). We also tested for differences in N1 (with reversed polarity) at the mastoids (N1$_{mast}$) using a repeated measures ANOVA with factors Sound Type (MA vs. A) and Laterality (M1 vs. M2). We obtained a significant enhancement for the MA sounds $F(1, 22) = 15.68$, $p < .001$, $\eta_p^2 = .42$, suggesting that besides the attenuation for MA sounds observed at vertex, further modulatory effects of sound-action coincidence occur (Figure 3). We also found a significant main effect of Laterality, *F*(1, 22) = 5.96, *p* = .023, η_p^2 = .21, with lower amplitudes at M1 compared to M2, while the interaction between Sound Type and Laterality did not reach significance, *F*(1, 22) = 3.55, *p* = .073.

Table 1

Mean amplitudes and standard deviation per component and condition across 23 participants.

		Auditory-only (A)		Motor-auditory (MA)		Encoded as MA and forgotten		Encoded as MA and remembered		Encoded as A and forgotten		Encoded as A and remembered	
ERPs	*Electrodes*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
N1	Cz	-3.14	1.79	-2.66	1.98	-3.89	2.01	-4.51	2.28	-4.13	2.45	-4.19	2.22
P2	Cz	4.95	2.49	3.83	2.01	7.16	4.38	7.37	3.51	7.33	3.96	7.76	4.25
P3	Pz	-0.08	1.29	1.49	1.43	-	-	-	-	-	-	-	-
$N1_{mast}$	M1	0.26	0.87	0.67	0.84	0.51	1.33	0.27	1.02	0.59	0.95	0.53	1.29
	M2	0.43	0.99	1.03	0.98	0.65	1.12	0.61	1.41	0.83	1.38	0.86	1.33
$P2_{mast}$	M1	-0.75	0.79	-0.19	0.81	-1.88	1.51	-2.53	1.71	-2.03	1.32	-2.24	1.42
	M2	-0.56	1.01	0.05	0.87	-2.24	1.43	-2.63	1.64	-2.18	1.57	-2.45	1.55
Na	T7	-0.89	0.94	-0.97	1.18	-1.23	1.37	-1.48	1.19	-1.11	1.02	-0.86	1.09
	T8	-0.47	0.76	-0.45	1.03	-0.89	1.36	-1.21	1.30	-0.82	1.68	-0.59	1.12
Tb	T7	-1.91	1.01	-1.75	1.12	-2.89	1.73	-3.26	1.94	-2.97	1.66	-2.34	1.53
	T8	-2.18	1.40	-1.54	1.56	-3.68	2.25	-3.62	1.94	-3.40	2.19	-2.81	1.63

Next, we examined the attenuation effects at the N1 subcomponents at temporal sites, using a 2x2 ANOVA, with factors Sound Type (A vs. MA) and Laterality (T7 vs. T8) on Na and Tb (Figure 3a). For Na, only a significant main effect of Laterality was obtained, with lower amplitudes at T8 compared to T7, *F*(1, 22) = 4.82, *p* = .039, η_p^2 = .18, while the main effect of Sound Type and the interaction did not reach significance, *F*(1, 22) = 0.05, *p* = .828 and *F*(1, 22) = 0.35, *p* = .563, respectively. For Tb, however, we obtained significantly lower amplitudes for sounds coinciding with a motor act compared to the auditory-only ones, *F*(1, 22) = 9.03, *p* = .007, η_p^2 = .29, while the main effect of Laterality did not reach significance, *F*(1, 22) = 0.03, *p* = .871. However, we also found a significant interaction, *F*(1, 22) = 8.63, *p* = .008, η_p^2 = .28, reflecting that the attenuation for MA sounds was only significant in T8 but not in T7 (post-hoc *t*-tests, *t*(22) = -4.06, *p* < .001, *d* = -0.85 and *t*(22) = -1.04, *p* = .311, respectively).

Subsequently, we performed a one-sided *t*-test to test for possible differences in P2 amplitudes between A and MA sounds at encoding, with the hypothesis of attenuated responses for the latter. We obtained a significant P2 attenuation at Cz, *t*(22) = 3.98, *p* < .001, *d* = 0.83, with lower amplitudes for sounds that coincided with a motor act, compared to those that were passively presented to the participants (Figure 3a-b). We also tested for differences in this component (with reversed polarity) at the mastoids ($P2_{mast}$) using a repeated measures ANOVA with factors Sound Type (MA vs. A) and Laterality (M1 vs. M2). We observed a significant attenuation for the MA sounds, replicating the attenuation observed at Cz, *F*(1, 22) = 34.23, *p* < .001, η_p^2 = .61, as well as a main effect of Laterality, *F*(1, 22) = 4.66, *p* = .042, η_p^2 = .17, with more negative amplitudes at M1 compared to M2. The interaction of Sound Type and Laterality on $P2_{mast}$ did not reach significance, *F*(1, 22) = 0.54, *p* = .470. Finally, we also tested for differences in P3 at Pz, which yielded a significantly larger P3 amplitude for sounds coinciding with a motor act, *t*(22) = -6.57, *p* < .001, *d* = -1.37 (Figure 3). Finally, we decided to examine our data using a more data-driven approach to test for further effects that may have not been captured in the hypotheses-driven ERP analysis (cluster-based permutation analyses; see Supplementary Material; Maris & Oostenveld, 2007). We found a negative cluster (p < .001; 56 – 344 ms post-stimulus) and one positive cluster (p = 0.01; 122 – 232 ms post-stimulus), in line with the findings obtained in the ERP analysis (see Supplementary Material).

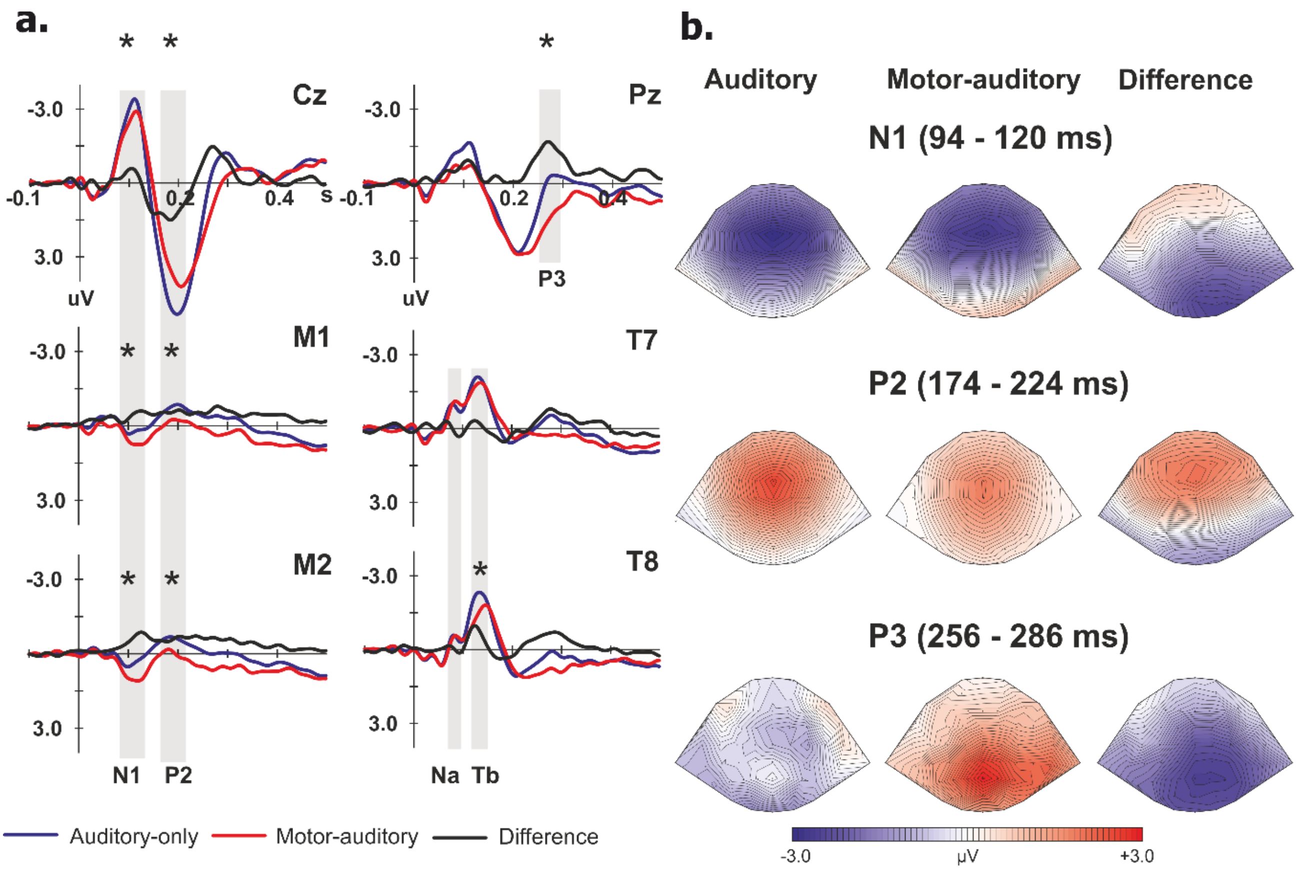


Figure 3. **a)** Group-average event-related potentials across 23 participants for the corrected motor-auditory (red) and auditory-only (blue), analyzed in the corresponding electrodes. Difference waves (A–[MA–M]) depicting the self-generation effects are represented in black. Time windows used for the analyses are indicated in gray (Na: 72–92 ms, Tb: 120–150 ms, N1: 94–120 ms, P2: 174–224 ms, P3: 256–286 ms). Significant differences in the event-related potentials are indicated by asterisks. **b)** N1, P2, and P3 scalp topographies in the time windows for: (1) the auditory-only condition (left); (2) the corrected motor-auditory condition (middle); and (3) the (A–[MA–M]) difference waves, reflecting suppression (N1, P2) and enhancement (P3) effects.

3.3. Electrophysiological responses at retrieval

Next, we performed exploratory analyses for the retrieval data, by subdividing it depending on whether the sound was encoded as A or MA and whether this sound was recalled or not. This allowed us to assess whether auditory evoked responses were affected by how the sound was encoded and whether it was remembered or forgotten. To this end, we ran an ANOVA with Sound Type (Encoded as MA vs. Encoded as A) and Memory (Remembered vs. Forgotten) as within-subject factors on N1/N1$_{\text{mast}}$, P2/P2$_{\text{mast}}$, Na, and Tb. An electrode factor (Laterality) was included in the ANOVA for the components identified in the mastoids and temporal electrodes. Figure 4 shows all the studied peaks for the remembered (a) and the forgotten (b) sounds at retrieval in the time windows 72–92 ms, 120–150 ms, 94–120 ms, 174–224 ms, for the Na, Tb, N1/N1$_{\text{mast}}$, and P2/P2$_{\text{mast}}$, respectively at the relevant electrodes for each peak.

We did not observe any significant effects (all *ps* > .05) on the N1 at Cz and N1$_{\text{mast}}$. However, significant results were obtained when we analyzed the modulatory effects of Sound Type and Memory on the N1 subcomponents at temporal sites. We obtained a significant main effect of Sound Type on Na, *F*(1, 22) = 7.39, *p* = .013, η_p^2 = .25, and Tb, *F*(1, 22) = 7.28, *p* = .013, η_p^2 = .25, reflecting an enhanced amplitude for sounds that were previously encoded as MA. Additionally, we found a significant interaction between Sound Type and Memory on Na, *F*(1, 22) = 5.08, *p* = .035, η_p^2 = .19, where post-hoc comparisons showed significantly larger Na amplitude for sounds that were Encoded as MA and were remembered compared to sounds that were Encoded as A and were remembered, *t*(45) = 3.73, *p* < .001, *d* = 0.55. In contrast, the post-hoc comparisons did not show significant differences for forgotten sounds as a function of how they were encoded, *t*(45) = 0.67, *p* = .504. No significant differences were found between remembered and forgotten sounds that were Encoded as A, *t*(45) = -1.34, *p* = .187, or between remembered and forgotten sounds that were Encoded as MA, *t*(45) = 1.64, *p* = .109. Similarly, we obtained a significant interaction between Sound Type and Memory on Tb, *F*(1, 22) = 4.85, *p* = .038, η_p^2 = .18. Post-hoc comparisons showed significantly larger Tb amplitude for sounds that were Encoded as MA and were remembered compared to sounds that were Encoded as A and were remembered, *t*(45) = 4.31, *p* < .001, *d* = 0.64, which is in line with the differences we obtained in the Na window. The post-hoc comparisons also showed lower Tb amplitudes for the Encoded as A sounds when they were remembered compared to when they were forgotten, *t*(45) = -3.23, *p* = .002, *d* = -0.48. Nevertheless, no significant differences were observed between remembered and forgotten sounds that were encoded as MA, *t*(45) = 0.64, *p* = .523, or between the Encoded as MA and Encoded as A sounds that were forgotten, *t*(45) = 0.47, *p* = .640. For both Na and Tb, we did not observe any significant main effects of Laterality, nor any significant interactions between Laterality and Sound Type and/or Memory (all *p*s > 0.05). Finally, we did not observe any significant effects on P2 at Cz and P2$_{\text{mast}}$ (all *ps* > .05), except for a significant main effect of Memory on P2$_{\text{mast}}$, *F*(1, 22) = 7.65, *p* = .011, η_p^2 = .26, that showed lower amplitudes for sounds that were forgotten ($M_{Forgotten}$ = -2.08, $M_{Remembered}$ = -2.46, $SD_{Forgotten}$ = 1.44, $SD_{Remembered}$ = 1.56). Similar to the approach we followed for the encoding data, we also conducted exploratory analyses using cluster-based permutation statistics (Maris & Oostenveld, 2007), but we did not find any significant clusters for any of the effects (see Supplementary Material).

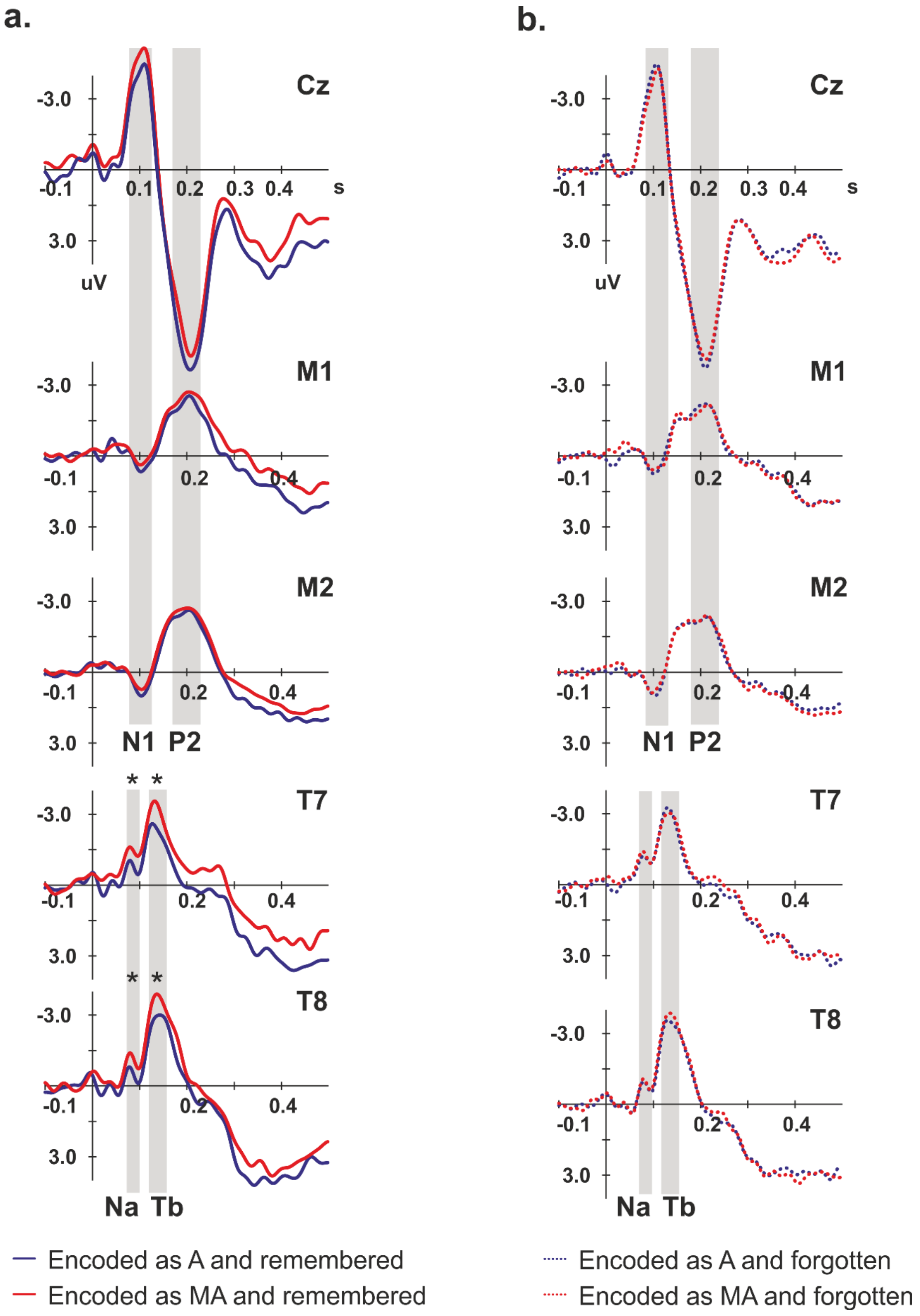


Figure 4. Group-average event-related potentials across 23 participants for the Encoded as MA (red) and Encoded as A (blue), analyzed in the corresponding electrodes and presented separately for the remembered (left) and the forgotten sounds (right). Time windows used for the analyses are indicated in gray (Na: 72–92 ms, Tb: 120–150 ms, N1: 94–120 ms, P2: 174–224 ms). Significant differences in the event-related potentials are indicated by asterisks.

3.4. Pupil responses at encoding and retrieval

Cluster-based permutation statistics were used to test for possible differences in pupil diameter between the conditions of interest. First, we tested for differences in the pupil response between motor-auditory and auditory-only events at encoding and we obtained significantly larger pupil diameter for motor-auditory events (starting 180 ms before sound onset and lasting up to 1,230 ms after sound onset; $p < .05$; Figure 5a) in line with previous work in rodents (e.g., McGinley et al., 2015). Interestingly, the effect of action started already in the pre-stimulus period, that is before the button press (which immediately triggered the sound), in agreement with previous work showing that LC activity and pupil diameter start increasing before the onset of movement (Aston-Jones & Cohen, 2005; Reimer et al., 2016). Subsequently, we conducted an exploratory analysis to test for possible main effects of Sound Type (Encoded as A vs. Encoded as MA) and Memory (Remembered vs. Forgotten), as well as for interactions between Sound Type and Memory on the pupil responses at retrieval. This analysis showed only a significant main effect of Memory, with larger diameter for forgotten sounds at retrieval compared to the remembered ones, irrespective of how they were encoded (starting 170 ms after sound onset and lasting until 830 ms after sound onset (*p* < .05; Figure 5b). Note that the morphology of the responses differs between the encoding (Figure 5a) and the retrieval (Figure 5b) data, most likely due to differences in the visual stimulation between the two phases (i.e., dynamic visual stimulation with the moving line at encoding vs. brief and static visual stimuli at retrieval, namely the cues “Sound 1” and “Sound 2”).

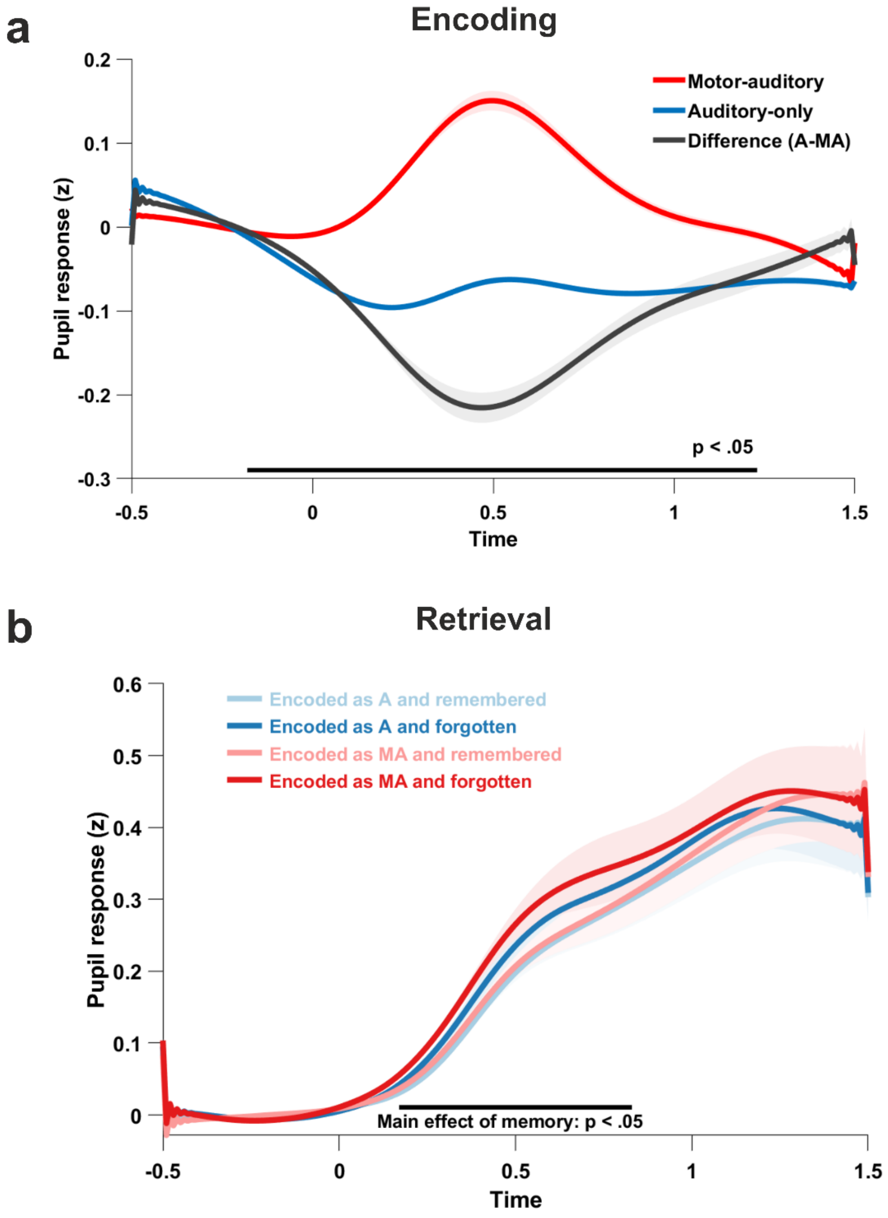


Figure 5. **a)** The group-average evoked pupil responses at encoding to auditory-only (blue) and motor-auditory (red) events. The effect is depicted as the difference between auditory-only and motor-auditory events (black). Black bar indicates a significant Auditory-only vs. Motor-auditory effect in the window 180 pre-stimulus to 1,230 ms post-stimulus, $p < .05$ (cluster-based permutation test). **b)** The group-average evoked pupil responses at retrieval to encoded as auditory (A) and encoded as motor-auditory (MA), separately for the remembered and forgotten sounds. Black bar indicates a significant main effect of memory for Remembered vs. Forgotten sounds in the window 170 – 830 ms post-stimulus, $p < .05$ (cluster-based permutation test).

3.5. Correlations

Next, we tested for possible correlations between the behavioural performance, pupillometric and electrophysiological data. For the correlation analyses, we focused on the significant neurophysiological effects at encoding (i.e., ERPs and pupil diameter) and the significant behavioural effect on memory performance. The effects were introduced in the correlation analyses as the difference between A and MA events (see Methods). For the components identified in two electrodes, we calculated the mean amplitude across the two, except for the Tb at encoding, where we introduced only the amplitudes at T8 given the significant interaction between Sound Type and Laterality that showed that attenuation was lateralized. For the pupil data, we calculated the peak of the difference wave (A – MA) within the window of significance (180 ms pre-stimulus until 1,230 ms post-stimulus). All the planned correlations are reported in Table 2.

Table 2

Correlations between the significant self-generation effects. a) electrophysiological effects at encoding (N1, P2, $N1_{mast}$, $P2_{mast}$, P3, and Tb amplitudes) and memory performance (1T sequences), b) neuromodulatory effects at encoding (pupil diameter) and memory performance (1T sequences), c) electrophysiological (N1, P2, $N1_{mast}$, $P2_{mast}$, P3, and Tb amplitudes) and neuromodulatory (pupil diameter) effects at encoding.

Correlations between		***r***	***p***
(a) Memory performance (1T sequences)	N1	-0.43	**0.041** *
	Tb (at T8 only)	-0.55	**0.007****
	P2	-0.19	0.383
	$N1_{mast}$	-0.41	0.055
	$P2_{mast}$	-0.10	0.657
	P3	-0.35	0.098
(b) Memory performance (1T sequences)	Pupil diameter	0.46	**0.029***
(c) Pupil diameter	N1	-0.36	0.091
	Tb (at T8 only)	-0.25	0.251
	P2	0.27	0.209
	$N1_{mast}$	-0.23	0.291
	$P2_{mast}$	-0.16	0.507
	P3	-0.08	0.702

First, we tested whether the significant self-generation effects at encoding (on N1, P2, $N1_{mast}$, $P2_{mast}$, P3, and Tb amplitudes) correlated with the significant self-generation effects on memory performance (1T sequences). This analysis showed a negative correlation between N1 suppression and memory performance (r = -0.43, p =

.041; Figure 6a), and a negative correlation between Tb suppression (at T8) and memory performance (*r* = -0.55, *p* = .007; Figure 6b), that is, the larger the N1 and Tb suppression, the greater the memory impairment for motor-auditory compared to auditory-only sounds. The remaining correlations did not reach significance (all *ps* > .05). Second, we assessed whether the difference in pupil diameter between auditory-only and motor-auditory events was related to memory performance, and we obtained a significant positive correlation between the two (*r* = 0.46, *p* = 0.029; Figure 6c), that is, the larger the pupil dilation for the motor-auditory events, the greater the memory impairment for these sounds. Third, we tested for possible links between the self-generation effects obtained in the ERP analyses (i.e., N1, P2, $N1_{mast}$, $P2_{mast}$, P3 and Tb) and the larger pupil diameter for motor-auditory events. None of these correlations reached significance (all *ps* > .05), but we observed a non-significant trend towards a correlation between N1 attenuation at Cz and pupil dilation for MA events (Figure 6d).

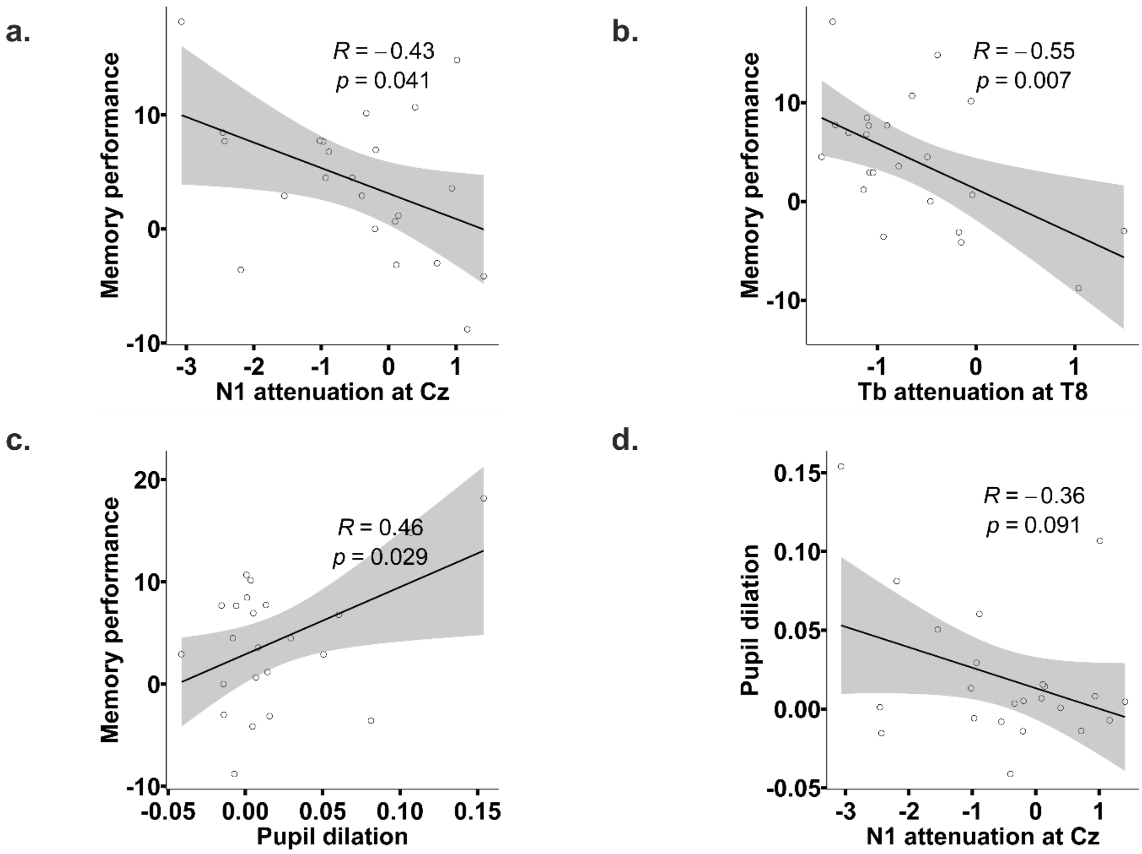


Figure 6. Planned correlations between the behavioural, electrophysiological, and pupil data using the Pearson correlation coefficient. **a-b)** Significant negative correlations between N1 suppression (at Cz) and memory performance (*r* = -0.43, *p* = .041), and Tb suppression (at T8) and memory performance (*r* = -0.55, *p* = .007), showing that the larger the N1 and Tb suppression, the greater the memory impairment for motor-auditory compared to the auditory-only sounds. More negative values indicate larger suppression effects for N1 and Tb and worse memory performance for motor-auditory sounds. **c)** Significant positive correlation between pupil dilation and memory performance (*r* = 0.46, *p* = 0.029), that is, the larger the pupil dilation for the motor-auditory events, the greater the memory impairment for these sounds. **d)** The correlation between N1 attenuation at Cz and pupil dilation at encoding for the MA events did not reach significance (*r* = -0.36, *p* = 0.091). The shaded gray areas represent the confidence interval (95% confidence level).

Finally, we performed an exploratory correlation analysis to test whether the significant differences in sensory processing we obtained at retrieval between Encoded as A and Encoded as MA sounds were related to the magnitude of the self-generation effects at encoding. To this end, we performed a correlation analysis between the A – MA difference in peaks of the Na and Tb amplitudes (only for the remembered sounds due to the significant interaction) and the effects at encoding (for the N1, P2, $N1_{mast}$, $P2_{mast}$, P3, and Tb amplitudes). We obtained a significant positive correlation between the P2 suppression at encoding and the Na enhancement at retrieval for the remembered sounds, reflecting that the larger the attenuation for P2 at encoding, the larger the Na enhancement for the Encoded as MA sounds that were remembered at retrieval (r = 0.51, p = .012). Similarly, we also obtained a significant negative correlation between Tb at encoding (at T8) and Na for the remembered sounds at retrieval (r = -0.42, p = .04), showing that the larger the attenuation for Tb at encoding, the greater the Na enhancement for motor-auditory sounds that were remembered at retrieval.

4. Discussion

In this study, we assessed the effects of motor actions on sensory processing and memory encoding of concomitant, but unpredictable sounds, by employing a combination of a self-generation and memory recognition task, while monitoring the brain's and the pupil's responses to sounds that were either presented passively or that coincided in time with a motor act. The aim of the present work was to assess how motor acts affect first sensory processing and second memory encoding of concomitant sounds, and the possible relationships between these two types of effects of actions.

Related to the first aim, regarding the effects of actions on sensory processing, we examined whether a) attenuation of sensory processing (i.e., measured by ERPs) prevails even in the absence of a contingent action-sound relationship (e.g., Horváth et al., 2012), b) actions create a halo of subcortical neuromodulation around them that could be reflected in the pupil diameter (e.g., McGinley et al., 2015), and c) sensory processing (i.e., measured by ERPs) and subcortical neuromodulation (i.e., measured by pupil diameter) during actions were related. Our findings showed N1, P2, $P2_{mast}$, and Tb attenuation for motor-auditory sounds even when they merely coincide with the action, as well as enhancement of P3 and $N1_{mast}$ (cf. Horváth et al., 2012). These findings suggest that self-generation effects are at least partly stimulus-unspecific and driven by alternative mechanisms to the cancellation of predicted sensory reafference via motor forward modelling. Additionally, our data replicated previous work (e.g., McGinley et al., 2015; Lee & Margolis, 2016; Vinck et al., 2015; Simpson, 1969; Yebra et al., 2019) showing that pupil diameter increases dramatically during actions providing evidence for an alternative stimulus-unspecific mechanism that could partly underlie sensory suppression for self-generated sounds, namely the activation of subcortical neuromodulation during motor actions. However, contrary to our initial hypothesis, the data did not provide clear evidence for a correlation between sensory attenuation and pupil dilation for motor-auditory events. The second aim of the present study was to investigate how actions affect memory encoding of concomitant sounds and whether the potential differences in the memory encoding of motor-auditory and passively presented sounds correlate with sensory suppression and/or subcortical neuromodulation during encoding. We found a significant impairment in memory performance for sounds that were encoded as motor-auditory compared to the auditory-only ones demonstrating that the mere presence of an action affects memory encoding of simultaneously presented stimuli. Most importantly, worsened memory performance for motor-auditory events correlated with increased sensory suppression (i.e., N1 and Tb attenuation) and larger pupil dilation for motor-auditory events at encoding. These findings fit well with the predictive coding framework suggesting that prediction errors (i.e., reflected in ERPs) drive learning and memory (Henson & Gagnepain, 2010) and further support previous work showing that high arousal (i.e., reflected in pupil diameter) may worsen behavioural performance (McGinley et al., 2015). In the following, we discuss each of these effects in detail.

The first aim of the present study was to assess the effects of actions on auditory processing and subcortical neuromodulation, as well as the relationship between the two. First, we provide evidence that the self-generation effects are at least partly unspecific by showing that N1 attenuation prevails even for mere action-sound coincidences and that it partly reflects the modulation of the unspecific N1 component, since for the suppression to be specific to the auditory cortex, N1 should be suppressed at vertex but also at the mastoids, which was not found here (cf. Horváth et al., 2012, Horváth, 2013b). This finding goes along with previous work pointing to partly unspecific mechanisms behind the action-induced suppression effects (e.g., Horváth et al., 2012; SanMiguel et al., 2013). For example, attenuation of auditory responses occurs also for stimuli merely coinciding with finger movements (Hazemann et al., 1975; Horváth et al., 2012; Makeig et al., 1996; Tapia et al., 1987) or for unrelated auditory inputs during speech (Numminen et al., 1999). Similarly, previous work has suggested that N1 (and Tb) attenuation can be driven by mere temporal contiguity (Horváth et al., 2012; Hazemann et al., 1975; Han et al., 2021) or by temporal predictability (Schafer & Marcus, 1973; Lubinus et al., 2021; Kaiser & Schütz-Bosbach, 2018, but see also Klaffehn et al., 2019 for evidence showing that attenuation prevails when controlling for temporal predictions), rather than stimulus-specific predictions, and that it mostly reflects modulations of the unspecific component of the auditory N1 (SanMiguel et al., 2013). Meanwhile, there is also mounting evidence supporting the stimulus-specificity of the effects by showing more pronounced suppression when predictions match more precisely with the sensory input (Fu et al., 2006; Hashimoto & Sakai, 2003; Heinks-Maldonado et al., 2005; Houde et al., 2002; Baess et al., 2008). Collectively, we believe that our findings point to the involvement of unspecific processes in the action-induced suppression of auditory responses that can, nevertheless, co-exist with stimulus-specific predictive mechanisms as suggested by previous work (Horváth, 2015; Schröger et al., 2015; Flinker et al., 2010).

In addition to the N1-attenuation effects, we observed attenuated P2 and enhanced P3 responses for the sounds coinciding with actions. Although a functional interpretation of P2 is missing (Crowley & Colrain, 2004), empirical evidence has shown that the P2 component originates in secondary auditory areas (Bosnyak et al., 2004; Pantev et al., 1996), reflecting the processing of the specific features of auditory stimuli (Shahin et al., 2005), and it correlates with the sense of agency (i.e., the feeling of control over actions and their consequences; Gallagher, 2000) contrary to the N1 that does not (Ford et al., 2014; Kühn et al., 2011; Timm et al., 2016). These characteristics along with our data showing P2 attenuation in both vertex and mastoids may point to a functional dissociation between N1 and P2 as suggested by previous work (Knolle et al., 2013b; Schröger et al., 2015; Chen et al., 2012). Following the P2 attenuation, we found enhanced P3 amplitude at Pz for sounds coinciding with actions. Interestingly, a P3 effect was also evident – although not discussed – in previous work with action-sound coincidences (Horváth et al., 2012). Recently, this effect has been suggested to reflect violations in action-related predictions (Darriba et al., 2021) which may occur in tasks where the self-generated sound is unexpected (e.g., in coincidence tasks where the action does not always result in a sound; Horváth et al., 2012). Although previous work has already described P3 modulations in self-generation paradigms, the posterior distribution and later peak of our effect differentiates it from the fronto-central P3a effect reported for unexpected externally-generated sounds (Baess et al., 2011) or self-generated deviant sounds (Knolle et al., 2013a). Based on previous theories, we speculate that the posterior P3 effect may be related to context updating (Donchin & Coles, 1988), event categorization (Kok, 2001) or decision making

(Twomey et al., 2015) and may reflect an evaluative process of the stimulus (i.e., self/external categorization) that ultimately updates the internal model about the sensory consequences of the button press (Polich, 2007).

The second important finding related to our first aim is that neuromodulatory processes take place concomitantly to the modulatory effects of action-sound coincidence on evoked electrophysiological responses. We obtained pupil dilation measures that are known to track the activity of the LC-NE system (Aston-Jones & Cohen, 2005, Murphy et al., 2014; Joshi et al., 2016) and in line with our hypothesis, we showed a remarkable increase in pupil diameter for the motor-auditory events that started even before the action (cf. McGinley et al., 2015; Aston-Jones & Cohen, 2005; Reimer et al., 2016), supporting previous work reporting pupil dilation during finger movements (Lubinus et al., 2021; Yebra et al., 2019), and locomotion (Reimer et al., 2014; Vinck et al., 2015; McGinley et al., 2015) even in the absence of visual stimulation (Richer & Beatty, 1985; Hupe et al., 2009). We also hypothesized that these neuromodulatory processes might be behind the stimulus-unspecific effects of actions on the auditory evoked responses. However, pupil dilation did not correlate with the sensory suppression effects for self-generated sounds. Although this may suggest that motor-induced sensory suppression and arousal-related neuromodulation during actions operate independently, there was a non-significant trend towards a link between N1 attenuation at vertex and pupil dilation, and both of these measures correlated significantly with memory performance. Taken together, these findings raise the need of future work to further test for relationships between action-induced suppression effects and neuromodulatory mechanisms operating during movement.

The second aim of the present study was to assess how the differential processing for sounds coinciding with actions might affect their encoding in memory. While the links between sensorimotor processing of auditory stimuli and memory processes remain largely unexplored, there is evidence that actions attenuate responses in areas supporting memory processes (i.e., Rummell et al., 2016; Mukamel et al., 2010), raising the possibility of a link between self-generation and memory. In our study, motor actions affected the memory encoding of concurrent sounds, but the effects were reflected only in memory performance and not in memory bias. The null effects on memory bias might suggest that participants could recognize that both test sounds at retrieval were presented before, which is supported by the general high level of objective accuracy as well as by reports during an informal debriefing suggesting that many participants thought that most times all sounds at retrieval were presented before. The memory benefit for the more surprising externally-generated sounds fits well with predictive coding theories postulating that items eliciting larger prediction errors at encoding will be encoded better in memory (Henson & Gagnepain, 2010; Greve et al., 2017; Heilbron & Chait, 2018; Pine et al., 2018; Krawczyk et al., 2018; Rescorla & Wagner, 1972; Exton-McGuinness et al., 2015). Yet, one would expect to observe this effect only in contingent paradigms where self-generated sounds are inherently more predictable than the externally-generated ones. However, although in our study actions were not predictive of sound identity or occurrence, they afforded better temporal predictability, which might have rendered motor-auditory sounds less salient, thereby compromising their encoding in memory (but not in 2T sequences where participants clearly remembered both sounds). We, therefore, acknowledge that our study cannot completely disentangle whether the effects observed on memory encoding are due to the neurophysiological effects of motor acts at encoding (e.g., attenuation and increased neuromodulation as indexed by pupil dilation), temporal predictability, or both.

Related to the second aim of the present study, we also hypothesized that the memory encoding of sounds paired with actions should be related to the neurophysiological effects of actions on sensory processing of sounds, namely the suppression effects and the pupil dilation for action-sound coincidences. First, we showed that the self-generation effects (i.e., N1 and Tb attenuation) are related to the performance decrements for sounds produced by actions as suggested by previous work in rodents (Schneider et al., 2014; McGinley et al., 2015; for a review see Schneider, 2020). These findings support the idea that the larger prediction error responses to unexpected items (as indexed by enhanced ERPs to A compared to MA events at encoding) initiate a cascade of synaptic changes, allowing for more distinctive representations at encoding (Kirwan & Stark, 2007; Norman, 2010) and thus better recollection at retrieval. Our findings could also fit with the compelling evidence for hippocampal involvement in learning from prediction errors (Schiffer et al., 2012) and expecting upcoming events (Davachi & DuBrow, 2015; Hindy et al., 2016; Schapiro et al., 2017): The reduced prediction errors at hippocampus to self-initiated stimulation (Rummell et al., 2016; Mukamel et al., 2010) could translate to memory decrements for these items. Second, we showed that memory performance correlated with pupil diameter as well, such that the larger the pupil diameter for motor-auditory events the worse the memory performance for these sounds at retrieval. To date, there have been no direct attempts to test for possible links between motor-induced pupil dilation and memory performance for stimuli triggered by actions. Some interim evidence points to a negative relationship between pupil dilation and detection performance during locomotion (McGinley et al., 2015), suggesting that performance may follow the classically described, inverted U-shaped dependence on arousal (Yerkes & Dodson, 1908): Intermediate levels of arousal – as indexed by pupil diameter – occur in states of quiet wakefulness and are characterized by optimal performance. In contrast, performance during high-arousal states such as movement drops dramatically. Collectively, we showed that sensory attenuation and pupil dilation independently correlate with memory performance, supporting the predictive account of memory (i.e., memory enhancements for items eliciting larger prediction errors at encoding) and providing yet another piece of evidence supporting the detrimental effects of high arousal (i.e., as indexed by pupil diameter) on behavioural performance.

The present study had clear hypotheses about the effects of actions on sensory and pupil responses at encoding, yet, exploratory analyses of the retrieval data revealed further effects. First, we obtained higher Na and Tb amplitudes for the sounds encoded as motor-auditory and remembered compared to the remembered and encoded as auditory-only ones. Since the sounds encoded as motor-auditory were presented passively at retrieval (i.e., without the motor representation that they were encoded with), the higher Na and Tb amplitudes may reflect a form of contextual prediction error (Exton-McGuinness et al. 2015; Kim et al., 2014; Sinclair & Barense, 2019) due to the mismatch between encoding and retrieval contexts for these sounds. This interpretation can be partly supported by the exploratory correlation analyses that showed that the larger the P2 and Tb attenuation for motor-auditory sounds at encoding, the greater the Na enhancement for these sounds at retrieval when they were remembered. Thus, the greater the effect of the action at encoding, the greater the contextual prediction error when the sound is presented without the action at retrieval. Second, we found larger pupil responses for the forgotten compared to the remembered sounds at retrieval irrespective of how they were encoded. While previous work has reported an old/new pupil effect (i.e., increased pupil responses for the remembered items; Kafkas & Montaldi, 2015; Naber et al., 2013, but see Beukema et al., 2019 for the opposite effect), in our study both sounds at retrieval were presented before. The increase in pupil diameter for the forgotten sounds at retrieval could be instead related to selection or decision uncertainty (Geng et al., 2015; Richer & Beatty, 1987; Nassar et al., 2012; Preuschoff et al., 2011) when participants experienced greater difficulty to decide whether a given sound was presented before or not.

In sum, the overarching aim of the present study was to investigate how motor acts affect both sensory processing and the memory encoding of concomitant sounds. To the best of our knowledge, there have been no previous attempts to simultaneously assess the specificity of the self-generation effects and their possible link with neuromodulatory processes while also looking into the effects of actions on memory encoding of sounds. Here, in a combination of self-generation and memory task, we show that actions affect auditory responses, pupil diameter, and memory encoding of sounds. Actions suppressed sensory responses for concomitant sounds and increased pupil diameter, but these effects were not related, pointing to simultaneous, but probably independent processes. However, sensory suppression and pupil dilation both correlated with memory performance independently, such that the memory performance for sounds coinciding with actions decreased with larger sensory attenuation and greater pupil dilation. Collectively, our findings show self-generation effects even in the absence of a predictive action-sound relationship, replicate previous work showing that pupil diameter increases during actions, and finally point to differentiated internal memory representations for stimuli triggered by ourselves compared to externally presented ones. More importantly, the present study shows that subcortical neuromodulatory systems, along with cortical processes, simultaneously orchestrate auditory processing and memory encoding.

STUDY II

SUPPLEMENTARY MATERIAL

Cluster-based permutation tests for exploratory analyses

We decided to examine our data using a more data-driven approach to test for further effects that may have not been captured in the ERP analysis (Maris & Oostenveld, 2007). To this end, cluster-based non-parametric correction was used to account for multiple comparisons and determined whether there were statistical differences between the contrasting conditions within participants. Each cluster was constituted by the samples that consecutively passed a specified threshold (in this case sample *p*-value of 0.05). The cluster statistics was chosen as the sum of the paired *t*-values of all the samples in the cluster.

For each statistical test, this procedure was performed by randomly switching labels of individual observations between these paired sets of values. We repeated this procedure 10,000 times and computed the difference between the group means on each permutation. The obtained *p*-value was the fraction of permutations that exceeded the observed difference between the means (i.e., two-sided dependent samples tests). The cluster-based permutation tests were done using Fieldtrip (Oostenveld et al., 2011).

For the encoding data, we compared responses to auditory and motor-auditory sounds at encoding in the entire epoch window (-100 to +500). This analysis yielded two significant clusters that mirror the findings obtained in the ERP analysis: a negative cluster ($p < .001$) starting at 56 ms and lasting up to 344 ms post-stimulus, with the effect being most pronounced over parietal sensors and one positive cluster ($p = 0.01$) starting 122 ms and lasting up to 232 ms being most evident over frontocentral electrodes (See Figure S1).

For the retrieval data, we tested for main effects of Sound Type and Memory but also for possible interactions between Sound Type (Encoded as A vs. Encoded as MA) and Memory (Remembered vs. Forgotten), using the responses at retrieval in the entire epoch window (-100 to +500). Specifically for the interaction, the cluster-permutation test was performed on the difference waves ([Encoded as A and remembered – Encoded as MA and remembered] and [Encoded as A and forgotten – Encoded as MA and forgotten]). The analyses did not show any significant clusters for the main effects nor for the interaction.

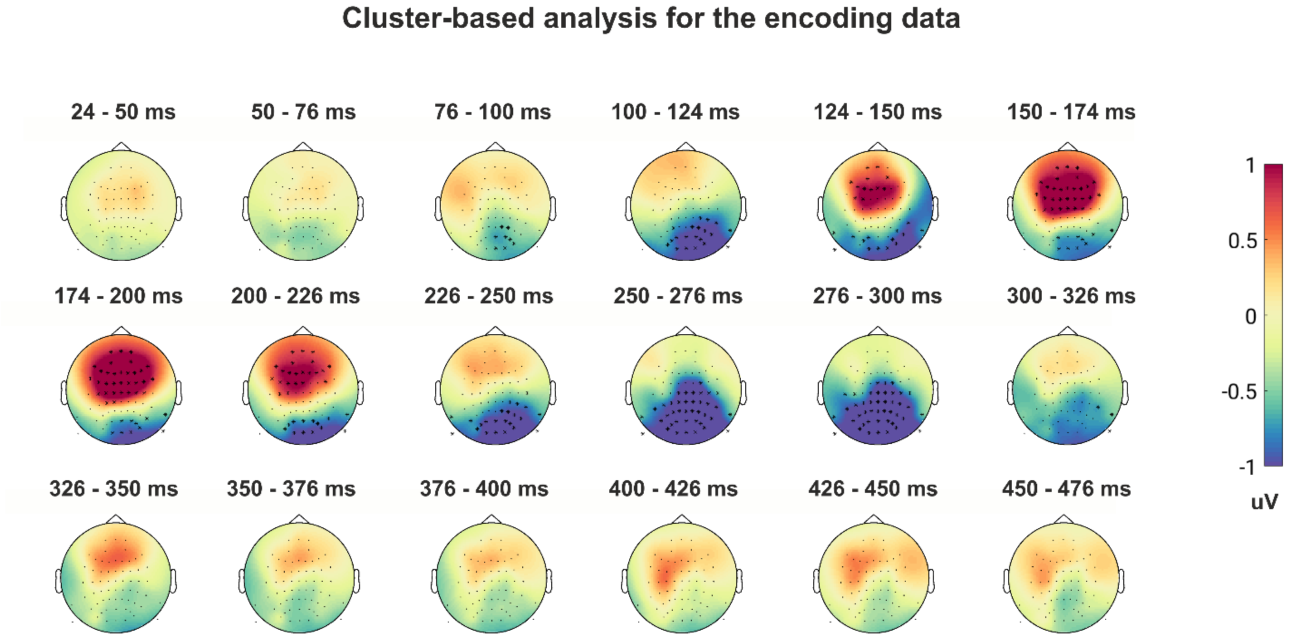


Figure S1. Non-parametric cluster-based permutation test comparing the average EEG signal in the auditory-only and the corrected motor-auditory condition (A–[MA–M]). Topographical maps denote the positive (red) and negative (blue) effects. The topography is shown for segments of 25 ms. The black dots indicate the electrodes over which the difference between the two conditions reaches significance. There were two significant clusters, a negative cluster (*p* < .001; 56 – 344 ms post-stimulus) and one positive cluster (*p* = 0.01; 122 – 232 ms post-stimulus).

STUDY III

Abstract

Distinguishing self-generated stimuli from those caused by external sources is critical for behaving organisms. Although actions have been shown to attenuate responses to self-produced sounds, several issues related to predictability confounds and the mechanism driving these effects (i.e., specific motor-predictions or unspecific processes) remain under debate. Meanwhile, the effects of self-generation on memory encoding – and their possible interactions with predictability – remain also largely unexplored. Here, we recorded behavioural, EEG, and pupil responses during a combined self-generation and memory paradigm to assess the effects of self-generation (motor-auditory vs. auditory-only; MA and A) and predictability (predictable vs. unpredictable sessions), and their interactions, on the sensory processing and memory recall of sounds. Predictability was controlled such that MA and A sounds were equally predictable (or unpredictable) in identity, timing, and probability of occurrence. In each trial, participants memorized a series of either predictable or unpredictable MA or A sounds, each sound presented either 150 or 750 ms after a semi self-paced and self-chosen action or a visual cue (for MA and A, respectively), and were later tested for recall. Results showed better memory performance for MA sounds, irrespective of predictability, and larger pupil diameter during actions especially when actions had predictable sensory consequences. Self-generation also attenuated N1, P2, and Tb responses with short action/cue-sound delays, but only in the predictable session. However, with longer action/cue-sound delays, self-generation attenuated N1 responses (irrespective of predictability), while P2 and Tb responses were suppressed by predictability (irrespective of self-generation). Crucially, sensory attenuation correlated with pupil diameter when the effects were strongest (predictable MA sounds presented shortly after the action). In sum, we show beneficial effects of actions on memory encoding and provide evidence for a cascade of action-induced stimulus-specific and unspecific influences on auditory responses, that might be linked to subcortical neuromodulation when the effects are more strongly elicited.

1. Introduction

The ability to distinguish between the two possible sources of sensory input (self or environment) is crucial for all organisms. Such distinction allows one to perceptually prioritize the newsworthy unpredictable – and potentially threatening – information (Blakemore et al., 2000; Poulet & Hedwig, 2006; Press et al., 2020) and to shape our sense of agency (Gallagher, 2000), and it highlights that rather than being a veridical reproduction of our external world, perception is scaffolded by our own predictions about the identity and timing of any upcoming event (Press et al., 2020). Such predictions can easily be afforded by our own actions: We cannot tickle ourselves, but we can be tickled by others, and similarly, we will be alerted by the sound of a horn, unless we are the ones generating it. This perceptual attenuation of self-generated reafferent input has been already reported (Sato, 2008) and agrees with neurophysiological data showing that indeed responses to self-generated input are suppressed compared to identical, albeit passively presented signals (Schröger et al., 2015).

Strikingly, the exact mechanisms driving the attenuation effects are still a matter of debate (for reviews see Schröger et al., 2015; Hughes et al., 2013a), and evidence suggests that sensory responses are dampened when we can predict the sensory consequences of our actions (e.g., Baess et al., 2011), but also when we cannot (i.e., attenuation occurs also for stimuli merely coinciding with movement, Horváth et al., 2012). Meanwhile, there is also evidence on effects of actions on memory processes (e.g., the production effect; MacDonald & MacLeod, 1998; Brown & Palmer, 2012), but again in these studies the action and predictability effects are usually conflated. These findings raise the need to disentangle the effects of predictability and self-generation on sensory processing and memory encoding, which constitutes the overarching aim of the present study. Specifically, we contrasted the attenuation and memory effects under predictable and unpredictable contexts, to isolate the contribution of motor acts on sensory processing and memory encoding and its interaction with predictability. In the following paragraphs, we provide an overview of the literature, summarizing the findings related to the action-induced modulations in sensory responses and memory performance as a function of predictability.

1.1. Sensory processing of self-produced stimuli

In the auditory domain, the self-generation effects on sensory processing have been mainly assessed using a *contingent* paradigm, where participants either listen passively to sounds presented by the computer or perform button presses that always result in the generation of a fully predictable sound (e.g., Baess et al., 2011). Most studies have reported attenuated auditory N1 and P2 event-related potential amplitudes and – inspired by early animal physiology work (Sperry, 1950; von Holst & Mittelstaedt, 1950) and computational models of motor control (Miall & Wolpert, 1996; Wolpert et al., 1995) – have explained the effect under the dominant cancellation models (also known as the comparator model; Blakemore, Wolpert et al., 1998; Frith et al., 2000; Wolpert & Flanagan, 2001). According to this model, the attenuation effects result from the operation of an internal forward model that allows one to predict the sensory consequences of her own action (corollary discharge) using a copy of the motor command (efference copy) (Blakemore et al., 2000; Wolpert & Flanagan, 2001). The resulting motor-driven predictions of sensory reafference are compared to the actual sensory consequences of one's actions, and subsequently, only the difference between the two (i.e., prediction error) is sent to higher stages of the neuronal hierarchy for further processing (Friston, 2005), effectively cancelling out responses to predictable, self-generated signals (Blakemore, Wolpert et al., 1998).

Inevitably, according to the cancellation model one would expect that suppression of responses would occur only when the stimulus can be predicted by the action and that the suppression effects would reflect modulations in sensory-specific areas. There is indeed evidence supporting the stimulus-specificity of the effects: For example, attenuation seems to be largest when the stimulus is fully predictable (Fu et al., 2006; Houde et al., 2002; Baess et al., 2008), and the locus of the effect has been found to be in areas within the auditory cortex (Martikainen et al., 2004; Aliu et al., 2009). Although these studies have provided strong support to the specificity of the effects, they assessed the effects only under fully predictable contexts for the self-generated stimulation, but not necessarily for the externally-generated one. That is, while the self-generated sounds were fully predictable in their identity, timing, and probability of occurrence, the timing and probability of occurrence for the externally-generated sounds could not be predicted (since they were presented passively without any cue). Therefore, the majority of these contingent paradigms have consistently conflated self-generation and predictability and have also confounded different types of predictions (i.e., predictions related to the identity, timing, and probability of occurrence of the stimulus; Hughes et al., 2013a), that can possibly exert distinct effects on sensory processing.

However, there is also evidence that sensory attenuation prevails even in the absence of a contingent action-sound relationship. For example, sensory responses are suppressed for stimuli that could not be predicted by the action (Horváth et al., 2012; Horváth, 2013a, 2013b; Paraskevoudi & SanMiguel, 2021; Hazemann et al., 1975; Makeig et al., 1996; Numminen et al., 1999), in line with the stimulus-unspecific generalized attenuation during certain actions reported outside of the domain of self-generation (e.g., saccadic suppression and somatosensory gating on moving body parts, Crapse & Sommer, 2008; Ross et al., 2001; Williams et al., 1998). Further support to the contribution of stimulus-unspecific processes in the suppression effects comes from work showing that part of the attenuation effects, specifically the auditory N1 attenuation, might not reflect a genuine modulation within the auditory cortex (SanMiguel et al., 2013). Specifically, SanMiguel and colleagues showed that instead of attenuating the sensory-specific N1 components (i.e., N1 at the mastoids and the “T complex” that are generated by tangentially oriented sources in the auditory cortex and radial sources in the superior temporal gyrus, respectively; Näätänen & Picton, 1987; Tonnquist-Uhlen et al., 2003; Wolpaw & Penry, 1975), actions reduced mainly the unspecific component of N1, which is proposed to be the cortical projection of a reticular process facilitating motor activity (Näätänen & Picton, 1987). Altogether, these findings raise the possibility of the self-generation effects not being solely a consequence of stimulus-specific predictions as proposed by cancellation theories, but they might rather reflect the operation of a broader unspecific mechanism (Korka et al., 2021) that allows the system to expect some action-related consequence, without necessarily generating a prediction on the specific effect of the action. However, to date, no attempts have been made to assess this possibility by manipulating all aspects of predictability (identity, timing, and probability of occurrence) in one experimental paradigm.

Meanwhile, despite the converging evidence pointing to stimulus-unspecific processes behind the action-induced modulations of sensory processing (Press et al., 2020; Press & Cook, 2015), the exact mechanism mediating unspecific effects remains largely unexplored. It is possible that actions create a halo of neuromodulation around them which might unspecifically gate auditory processing for stimuli presented in close temporal proximity with the action. Neuromodulatory influences are a plausible candidate for mediating – at least partly – the action-induced attenuation effects given that actions initiate a series of neuromodulatory processes (Vinck et al., 2015; Eggerman et al., 2014; McGinley et al., 2015), that are known to send inputs to the auditory cortex (sometimes even overlapping with inputs from motor areas; Nelson & Mooney, 2016; for a review see Schneider & Mooney, 2018). Specifically, a possible candidate for creating a halo of neuromodulation that could mediate stimulus-unspecific effects during movement could be the locus coeruleus norepinephrine system (LC-NE). This subcortical structure provides diffuse and widespread neuromodulatory inputs to the entire cortex, and is part of the ascending reticular activating system, which has been shown to modulate arousal and cortical responsivity to sensory stimulation (Aston-Jones & Cohen, 2005). In fact, the LC-NE has connections with both the auditory (Budinger et al., 2008; Foote et al., 1975) and primary motor cortex (Foote & Morrison, 1987), and many studies have shown a close association between pupil diameter – a proxy of LC-NE activity (Aston-Jones & Cohen, 2005; Vinck et al., 2015; Murphy et al., 2014; Joshi et al., 2016) – and actions (e.g., whisking or button press; McGinley et al., 2015; Vinck et al., 2015; Lubinus et al., 2021; Paraskevoudi & SanMiguel, 2021). Based on this evidence, the link between action-induced attenuation and increased neuromodulation might seem plausible, however, the only study that has assessed this relationship only showed a trend, but not a clear link between the two (Paraskevoudi & SanMiguel, 2021).

1.2. Memory encoding for self-produced stimuli

Given the differential sensory processing of self-generated stimuli, one would expect effects of self-generation also on memory. This assumption seems valid considering evidence reporting modulatory effects of movement on hippocampal and parahippocampal activity (Halgren, 1991; Mukamel et al., 2010; Rummell et al., 2016), hippocampal involvement in predicting upcoming events (Davachi & DuBrow, 2015; Hindy et al., 2016; Schapiro et al., 2017), as well as sensitivity to mismatches between expected and observed information (Fonken et al., 2019; Kumaran & Maguire, 2009). However, the direction of the effects of movement on memory remain mixed. One line of evidence points to a beneficial effect of actions on memory encoding: Memory enhancements have been observed for *fully* predictable self-initiated stimuli such as spoken words or played melodies compared to words read silently (i.e., production effect, MacDonald & MacLeod, 1998; Brown & Palmer, 2012). In stark contrast with the production effect studies, predictive coding theories postulate that the prediction error elicited by surprising stimuli results in better memory performance than memory performance for *fully* predictable inputs (Bar, 2009; Pine et al., 2018), reminiscent of the von Restorff effect (von Restorff, 1933). Indeed, evidence points to a positive relationship between prediction error and hippocampus activity at encoding for unpredictable inputs, which ultimately results in better memory performance (Gagnepain et al., 2011; Henson & Gagnepain, 2010; Pine et al., 2018). To our knowledge, our recent study is the only one providing supporting evidence for a link between sensory processing and memory encoding in the domain of self-generation (Paraskevoudi & SanMiguel, 2021): Sensory attenuation for sounds that coincided with button presses correlated with worsened memory performance for these stimuli. Interestingly, memory performance also correlated negatively with pupil diameter, suggesting that sensory suppression and pupil dilation for self-initiated stimulation independently relate to memory weakening. However, in all the above-mentioned studies, self-generated stimuli were to some extent more predictable than the passive comparisons (e.g., increased temporal predictability in Paraskevoudi & SanMiguel, 2021 and fully predictable in identity, timing, and probability of occurrence in production effect studies; MacDonald & MacLeod, 1998; Brown & Palmer, 2012), which raises the need to disentangle predictability- and action-effects on memory encoding.

1.3. The present study

In sum, as we outlined in the first two sections of the introduction, the paradigms attempting to explore the effects of actions on sensory processing and memory differed in the relationship between the action and the sound, usually conflating self-generation with predictability, but also in the types of actions performed. In many paradigms, actions were self-paced and goal-directed to generate the sounds and they provided either stimulus-specific identity predictions ("what" stimulus will be generated), or only temporal predictions ("when" the stimulus will be generated; for reviews see Hughes et al., 2013a; Schröger et al., 2015). Meanwhile, many other paradigms presented the sounds close in time to, but independently of, the actions and showed that movement dampens sound processing in an indiscriminate unspecific manner (as in human coincidence studies; Horváth et al., 2012; Horváth, 2013a, 2013b), which may point to a diffuse temporal expectation tied to the action (since actions typically have some auditory consequences). Additionally, despite the evidence of the effects of actions on memory, the two dominant lines of research (i.e., production effect and predictive coding) have observed these effects (either enhancement or weakening, respectively) only in fully contingent designs, leaving unexplored the possibility of other factors (i.e., action-related temporal control over the stimulus or unspecific temporal expectation about the stimulus) affecting the encoding of sounds in memory.

The present study constitutes a first attempt to orthogonally control for predictability and self-generation in order to assess whether the self-generation effects on sensory processing and memory encoding of sounds are due to stimulus-specific predictions, due to stimulus-unspecific effects of action, or a mixture of both. Specifically, the aim of this study was twofold: First, we aimed to assess the effects of actions and predictability, as well as the interactions between the two, on sensory responses (i.e., measured by auditory evoked potentials) and subcortical neuromodulation mediated by the LC-NE system (i.e., measured by pupil diameter), and assess whether sensory attenuation and subcortical neuromodulation during actions are related. Second, we sought out to examine in what way semi self-paced actions affect the encoding of predictable and unpredictable sounds in memory, and whether the effects of actions on memory performance are related to the sensory attenuation effects and/or the neuromodulatory processes during movements (as shown in our previous work; Paraskevoudi & SanMiguel, 2021).

Related to our first aim, we hypothesized that electrophysiological responses to self-generated sounds would be attenuated, with the attenuation being strongest when all aspects of the sound (i.e., identity, timing, and probability of occurrence) are predictable (e.g., Baess et al., 2008). We also hypothesized that button presses would increase pupil diameter (e.g., Paraskevoudi & SanMiguel, 2021; Lubinus et al., 2021), but we remained agnostic as to whether this effect would interact with predictability. Finally, we hypothesized that sensory attenuation and subcortical neuromodulation would be linked. Related to our second aim, namely the effects of actions and predictability on memory performance, we expected differences in recall rates between self- and externally-generated sounds but given the mixed evidence we did not have a specific hypothesis about the direction of the memory effects: Based on production effect studies, we would expect better memory for the self-generated sounds, however, predictive coding theories would predict the opposite effect (i.e., higher memory performance for externally-generated sounds). Critically, these two alternative hypotheses stemming from different lines of memory research only apply when the self-generated stimuli are more predictable than the externally-generated ones. Therefore, we aimed to assess whether indeed such differences would interact with predictability or whether the mere effect of action and temporal control over a stimulus (irrespective of predictability) can also modulate memory encoding.

To tackle these questions, we manipulated the predictability (predictable vs. unpredictable sessions) and source (self- vs. externally-generated) of sounds in a paradigm, where participants had to encode self- or externally-generated sounds and later recall them. In the predictable session, self- and externally-generated sounds were predictable in a) identity (i.e., fixed mapping between action and sound category), b) timing (all sounds were presented after a fixed delay), and c) probability of occurrence (i.e., all actions resulted in a sound). In contrast, in the unpredictable session, the self-generated sounds were unpredictable in all these aspects. Critically, in our design, externally-generated sounds were equally predictable or unpredictable in all these aspects (see Methods), allowing us to isolate the effects of motor prediction (in the predictable session) or motor proximity (in the unpredictable session) on sensory processing and memory encoding.

2. Methods

2.1. Participants

Twenty-six healthy, normal-hearing subjects, participated in the present study. Participants were typically undergraduate university students at the University of Barcelona. Data from six participants had to be excluded due to technical problems, inability to comply with the task instructions, or excessive artifacts in the EEG recording, leaving data from twenty participants (10 women, M_{age} = 24.55, age range: 18-43, 18 right-handed). None of them had any hearing impairments, had suffered from psychiatric disorders or had taken substances affecting the central nervous system the 48 hours prior to the experiment. All participants gave written informed consent for their participation after the nature of the study was explained to them and they were monetarily compensated (10 euros per hour). Additional materials included a personal data questionnaire and a data protection document. The study was accepted by the Bioethics Committee of the University of Barcelona.

2.2. Experimental design

The experiment consisted of a memory task (Figure 1) where, in each trial, participants were presented with a series of sounds that were either self- or externally-generated (motor-auditory, MA and auditory-only, A) at encoding. At retrieval, the same sounds, along with a set of new ones, were passively presented and participants had to indicate whether the sound was presented during encoding ('Old') or not ('New'). We manipulated predictability during the encoding phase, by employing an unpredictable and a predictable session, which were performed on separate days (2-4 days apart and order counterbalanced). Predictability for self-generated sounds was afforded by the action chosen by the participant, while predictability for the externally-generated sounds was afforded by externally-controlled visual cues. The sounds to be remembered were either fully predictable or fully unpredictable with respect to their identity (i.e., *what* sound category would be presented), their timing (i.e., *when* they would be presented after the action or cue) and probability of occurrence (i.e., *whether* they would be presented after the action or cue).

Encoding phase

In both sessions, every trial started with an encoding phase that consisted of 20 events of different types, each type of event presenting a different combination of presence or absence of actions, visual cues, and auditory stimuli. Specifically, of the 20 events of the encoding phase of each trial, 10 included sounds, half of them self-generated and the other half externally-generated (5 Motor-auditory and 5 Auditory-only events; MA and A, respectively). The remaining 10 events were secondary control events (5 Motor-only and 5 Visual-only events in the unpredictable session and 10 Empty events in the predictable session; M, V, and E, respectively) whose role is described below. The different event types were presented in random order within each trial. See Figure

1 for a detailed timeline of the encoding phase events and Table 1 for a summary of the number of events in each trial per predictability session.

Each event started with the presentation of the letter “P” or “L” centered on the screen along with four unfilled circles of different colours below it. When the letter “P” (i.e., “PRESS”) was presented, participants had to press one of the four pre-specified buttons. Button pressing was (semi-)self-paced (within a 1-3 s window) and self-chosen, similar to previous self-generation studies (e.g., Baess et al., 2011), thus giving a stronger feeling of temporal control or agency. Once participants pressed the button, one of the circles was filled immediately and the letter “P” disappeared. In the unpredictable session, the button press and colour-filling of the circle was followed by a sound only half of the times (5 unpredictable MA events), while in the other half no sound was presented (5 M events). Additionally, there was a random mapping between the button press, circle to be colour-filled, and sound category. Thus, in the unpredictable session, the sound category was randomly chosen and could not be predicted by the button pressed or the circle filled, and the probability of occurrence of the sound after the press was 50%. In contrast, in the predictable session, the button press and colour-filling of the circle was always followed by a sound (5 predictable MA events) and there was a fixed mapping between all three: For example, if the left most button was pressed, the left most circle on the screen would be colour-filled, and a human sound would be presented. Thus, in the predictable session, sound category for the MA sounds could be predicted by the button that was self-chosen and the circle that was subsequently filled, and there was a 100% probability of hearing a sound after the press.

When the letter “L” (i.e., “LISTEN”) was presented along with the four unfilled circles, participants had to remain still without pressing any button. In the unpredictable session, after a random delay of up to 3 seconds (drawn from the button press delays from the “P” trials within session), one of the circles – chosen randomly – was colour-filled. Immediately, the letter “L” disappeared and in half of the occasions a sound was presented – chosen randomly among the four possible sound categories (5 unpredictable A events), while in the other half of the occasions no sound was presented (5 V events). Thus, similar to the MA sounds of the unpredictable session, sound category for the A sounds could not be predicted by the circle that was colour-filled, and the probability of occurrence of the sound after the visual cue was 50%. In contrast, in the predictable session, after the presentation of the letter “L” on the screen, a circle was not always filled. This was necessary to achieve 100% contingency between cues or actions and sounds in this session, while keeping the trials equal in length and memory load (i.e., same sounds per second ratio) as in the unpredictable session (that contained also V and M events in which no sounds were presented after the cues or actions). Thus, in the predictable session, only 5 events within each trial were similar to the unpredictable session: the letter “L” was presented and after a random delay of up to 3 seconds (drawn from the button press delays from the “P” trials of this session) one of the circles was filled. Immediately, the letter “L” disappeared, but here a sound from the sound category corresponding to the filled circle was always presented (5 predictable A events). Thus, similar to the MA sounds of the predictable session, sound category for the A sounds could be predicted by the circle that was colour-filled and there was a 100% probability of hearing a sound after the visual cue. The remaining 10 events within each trial of the predictable session started with the letter “L”, but no circle was filled and no sound was played. Thus, the full cue-sound contingency was not disturbed by these empty events (10 E events).

With this design, we could manipulate identity predictability and probability of occurrence, by providing fixed or random mapping between button press, circle to be colour-filled, and sound category, and by having 100% or 50% contingency between button presses and sounds and colour-filled cues and sounds, for the predictable and unpredictable sessions, respectively. We additionally manipulated temporal predictability, by introducing two delays (150 and 750 ms) between the button press and/or the colour-filling of the circle and the sound, that were intermixed in a random fashion within each trial in the unpredictable session, while they alternated block-wise in the predictable session. Thus, the button press or filling of the circle predicted the exact timing of the sound in the predictable session, while the timing (either 150 or 750 ms delay) could not be predicted in the unpredictable session.

Table 1

Number of events of each type per trial for each predictability session.

Event type	**Session**	
	Unpredictable	*Predictable*
Motor-auditory (MA)	5	5
Auditory-only (A)	5	5
Visual-only (V)	5	-
Motor-only (M)	5	-
Empty (E)	-	10

Finally, to make sure that the event duration was similar for all event types, for the no-sound events (i.e., M, V, and E), we presented an empty sound of the same duration as the sounds presented as A or MA. The interval between events (e.g., offset of sound or no-sound and start of the next P or L letter on the screen) was set to 350 ms. Therefore, each event consisted of the interval between letter appearance and button press or cue-filling (1–3 s), the delay between button press/cue-filling and sound (150 or 750 ms), the duration of the sound (or no sound for M, V, E events; 500 ms), and the inter-event interval (350 ms), resulting in a total event duration in the range between 2 and 4.6 s, and a total encoding duration of 40–92 s, depending on the above.

Retention and retrieval phases

Following the encoding phase, a short retention phase of 3 seconds followed, where participants were presented with a fixation cross on the screen. Following the retention phase, the retrieval phase started. A series of 20 sounds were presented passively to the participants. During the presentation of the sound (500 ms), a fixation cross was shown in the screen. Each sound was followed 350 ms after sound offset by the question O/N (i.e., Old/New?) and participants had to perform an Old/New judgment. Half of the sounds at retrieval were new (i.e., they were not presented at the encoding phase of the trial), and the other half were old (i.e., they were presented in the preceding encoding phase). Among the old sounds, half of them were encoded as MA, and the other half as A, and of each half with 150 and half with 750 ms delay. Response buttons were counterbalanced across participants and the response window was 1 s. Once participants responded, or the response window was over, the question disappeared, and a fixation cross was displayed for 1.15 s and the message “NEXT” was presented for 2 s (inter-trial interval) informing participants that the following trial was about to start.

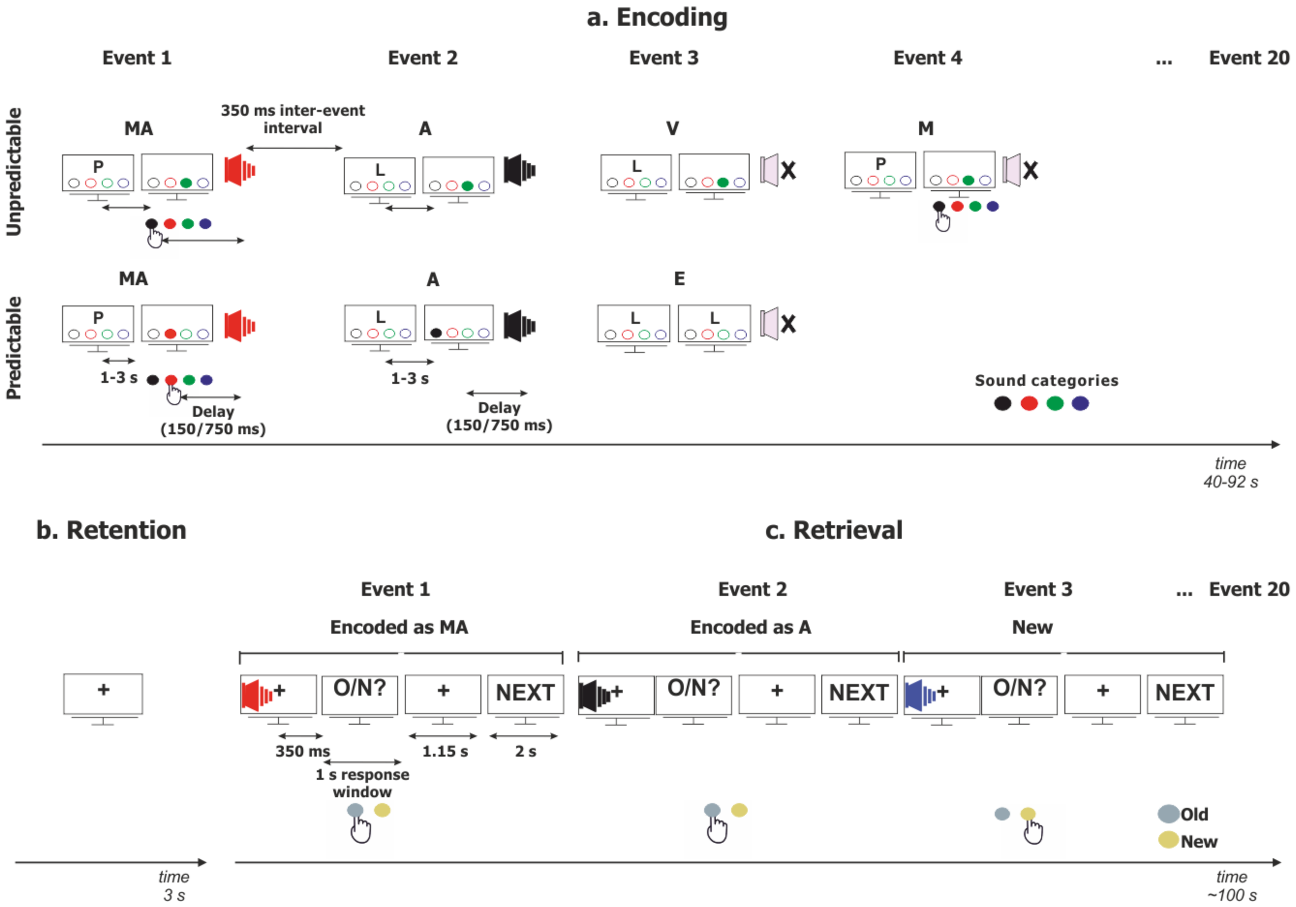


Figure 1. Schematic representation of the design. Two sessions were employed (predictable and unpredictable) that differed only in the encoding. Each trial within each session consisted of three phases: encoding, retention, and retrieval. **a)** At encoding, participants were presented with 20 events starting with the letter “P” or “L”, along with four unfilled circles, and they had to press one of the four predefined buttons (1–3 s) or not perform any button pressing and wait for a circle to be colour filled (except for the E events), respectively. In the unpredictable session, only half of the button presses and colour-filled cues resulted in unpredictable Motor-auditory (MA) and Auditory-only (A) sounds presented 150 or 750 ms (in mixed fashion) after the cue or button press. The other half of the button presses and colour-filled cues did not result in a sound (i.e., an empty sound was presented to match the duration with the sound events; Motor-only and Visual-only events; M and V). In the predictable session, fully predictable Motor-auditory (MA) and Auditory-only (A) sounds were presented 150 or 750 ms (in blocked fashion) after the button press and the colour-filling of the cue. The sound category (animal, human, environmental, or musical) was contingent on the button that participants pressed and the circle that was colour-filled. Empty events (i.e., no colour-filling of the cue and no sound presentation; E) were added to make the predictable and unpredictable trials equal in length and had the same timing as the rest of the events. The inter-event interval (i.e., sound offset to upcoming letter) was set to 350 ms. **b)** At retention, in both predictable and unpredictable sessions, participants were presented with a fixation cross and were instructed to maintain the previously presented sounds in memory. **c)** At retrieval, a series of 20 sounds (10 old and 10 new) were presented passively to the participants. Among the old sounds, half of them were encoded as MA, and the other half as A (either with 150 or 750 ms delay). During sound presentation, a fixation was shown on the screen. The question O/N (i.e., Old/New?) appeared 350 ms after sound offset and subjects had 1 s to reply. Following the end of the response window, a fixation cross appeared for 1.15 s, followed by the cue “NEXT” (2 s) indicating that the next sound was about to be presented.

2.3. Auditory stimuli

The auditory stimuli consisted of a pool of identifiable sounds from four categories (animals, musical instruments, environmental sounds, and human-related sounds), which allowed us to have well-established identity mappings between button type and sound category for the predictable session. These sounds were first drawn from several freely available sound repositories (the Adobe and FreeSound databases, and those of Norman-Haignere et al., 2015; Gygi & Shafiro, 2010; Hocking et al., 2013; Belin et al., 2000), and were then edited to have a 500 ms duration including 0.01 s exponential ramps and a sampling rate of 44.1 kHz, and to be played at 16 bits, mono and 75 dB intensity. Sound identifiability was assessed by the ratings of three volunteers that were presented with a series of sounds and indicated whether each sound could be easily assigned to one of the four categories. An additional option was provided to them for the sounds that they could not be assigned in none of the above-mentioned categories. The most identifiable sounds, as assessed by this rating, were used as the auditory stimulation in the present work. In each predictability session, we used 340 sounds at encoding (A and MA that were also presented at retrieval) and 340 additional and different sounds that were used as New at retrieval. The same sounds were used in both predictability sessions.

2.4. Apparatus

The visual stimuli were presented on an ATI Radeon HD 2400 monitor. The auditory stimuli were presented via Sennheiser KD 380 PRO noise cancelling headphones. To record participants' button presses and behavioural responses, we used a Korg nanoPAD2. The buttons of this device do not produce any mechanical noise when pressed, and, thus, do not interfere with our auditory stimuli. The presentation of the stimuli and recording of participants' button presses and responses were controlled using MATLAB R2017a (The Mathworks Inc., 2017), the Psychophysics Toolbox extension (Brainard, 1997; Kleiner et al., 2007), and the Eyelink add-in toolbox for eyetracker control.

EEG activity was acquired at a 500 Hz sampling rate with a Neuroscan SynAmps RT amplifier (NeuroScan, Compumedics, Charlotte, NC, USA), using the CURRY 8 Neuroscan software. We recorded continuously with Ag/AgCl electrodes from 64 standard locations according to the 10% extension of the International 10–20 system (Chatrian et al., 1985; Oostenveld & Praamstra, 2001) mounted in a nylon cap (Quick-Cap; Compumedics, Charlotte, NC, USA). An additional electrode was placed at the tip of the nose (serving as online reference). The vertical electrooculogram (EOG) was measured with two electrodes placed above and below the left eye, and the horizontal EOG with two electrodes placed on the outer canthi of the eyes referenced to the common reference (unipolar montage). The ground electrode was placed at AFz. All impedances were kept below 10 kΩ during the whole recording.

Concurrently with the EEG recording, horizontal and vertical gaze position, as well as the area of the pupil, were recorded using EyeLink 1000 desktop mount (SR Research, sampling rate: 1,000 Hz; left eye recordings). Visual stimulation was presented at the center of the screen (i.e., visual angle 2.3°) to eliminate any effects of gaze position on pupil diameter (Gagl et al., 2011). The pupil was assessed in the centroid mode of the eye tracker, which uses a center-of-mass algorithm. This algorithm detects the pupil area by identifying the number of black pixels and its center on the video image.

2.5. Procedure

Prior to the start of the experiment, participants were asked to complete several questionnaires. Subsequently, participants were seated in an electrically and acoustically shielded room and were asked to place their head in a chinrest at approximately 60 cm from the screen. Eye-tracker calibration was performed at the start of the experiment. In order to familiarize themselves with the task and the button pressing, participants completed three training blocks at the start and in the middle of the experiment. During the first two training blocks participants were presented with the letter "P" and four unfilled circles and were instructed to press a button within the predefined pressing window (up to 3 seconds). They had to produce a uniform distribution of letter-press delays between 1 and 3 s, as well as a uniform distribution of the presses over the 4 buttons. Once they pressed, a circle was colour-filled and a sound was presented either based on a fixed button-circle-sound category mapping (predictable session) or based on a random one (unpredictable session). The delay between button press and sound was either 150 or 750 ms, presented randomly intermixed in the unpredictable session and in a blocked fashion in the predictable session. The third training block consisted of two trials that had the same structure as the experimental trials within predictability session, with the only difference being that the sounds were drawn from a different pool so that during training participants would not be exposed to the sounds used in the main experiment.

The main experiment consisted of 34 trials per predictability session, divided in eight experimental blocks (six blocks of four trials and two blocks of five trials). At the end of each block, a message appeared informing participants about the number of missed responses, the mean letter-press delay, and the percentage of pressing each one of the four buttons for this block. Participants took a break of approximately 5 minutes every 2 blocks to prevent fatigue. Each session of the experiment lasted for approximately 2 hours excluding the EEG preparation.

Finally, given that participants performed the two predictability sessions in two separate days, we confirmed that there were no differences in the cap/EEG setup between the two sessions that might compromise our main findings. To this end, 9 participants of our sample performed three additional blocks in each session (at the start, in the middle, and at the end of each session), where they passively listened to sounds of different frequencies and we compared N1 responses (at Cz) between the two sessions. The auditory stimuli consisted of pure tones of 30 different frequencies (300 to 3200 Hz in steps of 100) and of 250 ms duration. In each block, the 30 sounds were presented five times, resulting in 150 sounds per block. The sound-to-sound interval was set to 500 ms. The preprocessing of this data followed the typical preprocessing (see Data Analysis) and epochs were created –100 up to 400 ms around each sound, separately for the first and the second session. Baseline correction was done using the 100 ms pre-stimulus interval. After averaging, we defined the N1 window (90–110 ms post-stimulus) and we performed a two-sided t-test to test for differences between sessions. There were no differences in N1 responses between the two sessions: $t(8) = 1.32$, $p = 0.22$.

2.6. Data analysis

2.6.1. Behavioral analysis

To test for differences in memory performance, we calculated the percent correct for the sounds at retrieval and we ran an 3x2 ANOVA with factors Delay (150 ms vs. 750 ms), Predictability (Predictable vs Unpredictable), and Source (Auditory vs. Motor-auditory).

2.6.2. EEG preprocessing

EEG data was analyzed with EEGLAB (Delorme & Makeig, 2004) and plotted with EEProbe (ANT Neuro). Data were high-pass filtered (0.5 Hz high-pass, Kaiser window, Kaiser β 5.653, filter order 1812), manually inspected so as to reject atypical artifacts and identify malfunctioning electrodes, and corrected for eye movements with Independent Component Analysis, using the compiled version of runica (binica) that uses the logistic infomax ICA algorithm (Onton & Makeig, 2006). Components capturing eye movement artifacts were rejected by visual inspection and the remaining components were then projected back into electrode space. Data was then low-pass filtered (30 Hz low-pass, Kaiser window, Kaiser β 5.653, filter order 1812), remaining artifacts were rejected by applying a 75 µV maximal signal-change per epoch threshold, and malfunctioning electrodes were interpolated (spherical interpolation).

For the encoding data, we were interested in comparing responses to A and MA sounds as a function of predictability, separately for each delay condition. Typically, the comparison between self- and externally-generated sounds is done after subtracting the motor-only averages from the motor-auditory averages to correct for motor activity (for a review see Schröger et al., 2015), since the signal obtained in the motor-auditory condition represents the brain signal elicited by the sound, but also by the planning and execution of the finger movement to press the button. Although such correction is not necessary with long press-sound delays such as in the 750 ms delay condition, correction for both visual and motor activity was needed in the 150 ms delay condition. Since motor activity differs as a function of predictability context (Neszmelyi & Horváth, 2017) and given that we did not have a motor-only condition in the predictable session, the correction was, therefore, performed with the following procedure: Using the data from the 750 ms delay only, we first locked each trial to the color-filling of the circle which coincided with a button press in the case of the MA sounds, separately for each predictability session (predictable and unpredictable) and source (A and MA). Subsequently, we created subjects' averages locked to the sound for the 150 ms events (epoching window: −250 to +550) and we subtracted from these averages the signal from the 750 ms events prior to sound onset (i.e., −100 to +700 ms after the color-filling of the circle). This allowed us to correct both for motor and visual activity in all sounds presented only 150 ms after the color-filling of the cue and the button press. No baseline was used for the epoching performed in this procedure. Finally, we also created epochs locked to the sound for the 750 ms events in the same window as for the 150 ms events (i.e., −250 to +550), but no correction was required with the long delay. For the retrieval data, a −100 to +500 ms epoch was defined around each event (Encoded as A and Encoded as MA, separately for each predictability and delay condition) which was baseline corrected (100 ms prior to the event). No motor correction was performed at retrieval since the Encoded as MA sounds were presented passively. Finally, we calculated the average wave for each event of interest, as well as the grand average for the whole sample (A and MA sounds for each predictability and delay condition, separately for encoding and retrieval). After trial rejection, each individual average for each of the event types was obtained averaging a mean of 69.88 epochs for the encoding data (*SD* = 14.97) and 69.12 for the retrieval data (*SD* = 19.37).

2.6.3. ERP analysis

At encoding, we aimed to assess the self-generation effects as a function of predictability, separately for each Delay condition, by comparing responses to MA sounds (i.e., corrected for motor/visual activity or not, for the 150 ms and 750 ms delay, respectively) with the responses elicited by the A sounds (i.e., corrected for visual activity or not, for the 150 ms and 750 ms delay, respectively). At retrieval, we sought to examine whether responses to sounds were modulated as a function of predictability (Predictable vs. Unpredictable) and how they were encoded (Encoded as A or Encoded as MA).

For all the effects of interest at encoding, we examined responses separately for the N1 and P2 at Cz, N1 at the mastoids ($N1_{mast}$), and the N1 subcomponents Na and Tb at temporal sites. The same components were examined at retrieval. The windows were defined after visual inspection of the data by locating the highest negative or positive (depending on the component of interest) peak in the usual latencies for each component as reported by previous work (SanMiguel et al., 2013). Specifically, time windows for N1, $N1_{mast}$, P2, Na, and Tb were defined on the grand-averaged waveforms of the auditory-only sounds in the predictable session as previously reported (e.g., SanMiguel et al., 2013). Na and Tb were identified as the first and second negative peaks, respectively, identifiable after sound onset on electrodes T7 and T8, as recommended by Tonnquist-Uhlen et al. (2003). N1 and P2 were identified as the negative and positive peaks occurring in the window ~70 to 150 ms, and ~150 to 250 ms after stimulus onset on Cz, respectively. Conversely, $N1_{mast}$ was identified as the positive peak in the window ~70 to 150 ms after stimulus onset on electrodes M1 and M2. The time windows for the N1, $N1_{mast}$, P2, Na, and Tb peaks were centered on the identified peaks ± 12, 15, 25, 10, and 15 ms for the encoding and ± 10, 15, 20, 10, and 15 ms for the retrieval, respectively. Given variations in peak latencies across the conditions, the width of the windows was defined such that it could capture the peak of the MA sound waveform as well, and it was proportional to the width of the component. For each delay condition, we performed a 2x2 ANOVA with factors Source (A vs. MA or Encoded as A vs. Encoded as MA for the encoding and the retrieval data, respectively) and Predictability (Predictable vs. Unpredictable) to test for differences in evoked potentials. For the components identified in two electrodes (i.e., Na, Tb, and $N1_{mast}$) an additional factor Laterality was introduced in the ANOVAs (i.e., T7 vs. T8 and M1 vs. M2).

In addition to the auditory ERP analyses, our design partly offered us the possibility to assess whether self-generation and predictability interact on visual ERPs. Note that a direct comparison between self- and externally-generated visual stimuli was not possible since we could not correct the self-generated stimuli for motor activity (because in both the 150 and 750 ms delay conditions, participants pressed a button that was followed immediately by the colour-filling of the circle). Therefore, these analyses focused on whether the visual responses after the button press are modulated by predictability context. To this end, we locked responses to the button press that was immediately followed by the filling of the cue for predictable and unpredictable MA sounds, separately for 150 and 750 ms delay. Given that the visual responses should not differ between delays (since the delay was introduced for the press-sound interval, whereas the visual cue was filled right after the button press), we took the mean N1 amplitudes between the 150 and 750 ms delay conditions for each participant (for MA predictable and MA unpredictable events separately) and performed a two-sided *t*-test. Visual N1 was identified by visual inspection as the first negative peak after stimulus onset on PO8 (cf. Kimura, 2021) with a window of ± 20 ms around the peak.

2.6.4. Cluster-based permutation tests

In addition to the ERP analyses, we opted for a data-driven analysis in order to explore the possibility of further effects of self-generation and predictability on sensory processing during encoding that would not have been captured by the ERP analysis on targeted components. To this end, cluster-based non-parametric correction was used to account for multiple comparisons and determined whether there were statistical differences between the contrasting conditions within participants. Each cluster was constituted by the samples that consecutively passed a *p*-value of 0.05. The cluster statistics was chosen as the sum of the paired *t*-values of all the samples in the cluster. Using the entire epoch of our data (i.e., -250 +550), we tested for main effects of Source (A vs. MA) and Predictability (predictable vs. unpredictable) as well as for possible interactions between the two factors on responses across all electrodes. Specifically for the interaction, the cluster-permutation test was performed on the difference waves ([A predictable – MA predictable] and [A unpredictable – MA unpredictable]). For the significant interactions, we performed post-hoc comparisons within the window of significance. For each statistical test, this procedure was performed by randomly switching labels of individual observations between these paired sets of values. We repeated this procedure 10,000 times and computed the difference between the group means on each permutation. The obtained p-value was the fraction of permutations that exceeded the observed difference between the means (i.e., two-sided dependent samples tests). The cluster-based permutation tests were done using Fieldtrip (Oostenveld et al., 2011).

2.6.5. Pupillometry analysis

Missing data and blinks, as detected by the EyeLink software, were padded by 100 ms and linearly interpolated. Additional blinks were found using peak detection on the velocity of the pupil signal and linearly interpolated (Urai et al., 2017). Blinks separated by less than 250 ms were aggregated to a single blink. The interpolated pupil time series were bandpass filtered using a 0.05–4 Hz third-order Butterworth filter. Low-pass filtering reduces measurement noise not likely to originate from physiological sources, as the pupil functions as a low-pass filter on fast inputs (Binda et al., 2013; Hoeks & Levelt, 1993). High-pass filtering removes slow drifts from the signal that are not accounted for by the model in the subsequent deconvolution analysis. First, we estimated the effect of blinks and saccades on the pupil response through deconvolution and removed these responses from the data using linear regression using a procedure detailed in previous work (Knapen et al., 2016; Urai et al., 2017). The residual bandpass filtered pupil time series was used for the evoked analyses (van Slooten et al., 2019). After zscoring per trial, we epoched the data, baseline corrected each trial by subtracting the mean pupil diameter 500 ms before onset of the event and resampled to 100 Hz. For the encoding data, we locked the data to the color-filling of the circle (epoching window -0.5 to 2 s post-event) that coincided with the button press in MA events, since we were interested in the effect of action on the pupil response which might not have been as clear to detect if the response was locked to the sound (especially for the 750 ms delay condition). For the retrieval data, the data was locked to the sound onset (epoching window: -0.5 to 2 s post-stimulus).

For each participant, we first obtained the average evoked response for the main events of interest. Specifically, we obtained the averages for the A and MA events at encoding locked to the color-filling of the cue, separately for each predictability and delay condition, while at retrieval we obtained the averages for the Encoded as A and Encoded as MA sounds, separately for each predictability and delay condition. We used non-parametric permutation statistics to test for the group-level significance of the individual averages, separately for encoding and retrieval. Specifically, we computed *t*-values of the difference between the two conditions of interest and thresholded these t values at a *p*-value of 0.05. Each cluster was constituted by the samples that passed the threshold of the *p*-value. The cluster statistics was chosen as the sum of the paired *t*-values of all the samples in the cluster. For each delay condition and separately for the encoding and the retrieval data, we aimed to test for possible main effects of Source (A vs. MA events and Encoded as A vs. Encoded as MA sounds, for encoding and retrieval, respectively) and Predictability (Predictable vs. Unpredictable), as well as for possible interactions between the two. For the main effects of Source and Predictability, the permutation statistics were performed between A and MA events (or Encoded as A and Encoded as MA sounds at retrieval) irrespective of the predictability session and between predictable and unpredictable events irrespective of how their source (A/MA or Encoded as A/Encoded as MA). To test for possible interactions, the cluster-permutation test was performed on the difference waves (e.g., [A predictable – MA predictable] and [A unpredictable – MA unpredictable]). For each statistical test, this procedure was performed by randomly switching labels of individual observations between these paired sets of values. We repeated this procedure 10,000 times and computed the difference between the group means on each permutation. The obtained *p*-value was the fraction of permutations that exceeded the observed difference between the means (i.e., two-sided dependent samples tests). The pupil preprocessing and analysis was performed with custom software based on previous work (Urai et al., 2017) using Fieldtrip (Oostenveld et al., 2011).

2.6.6. Correlations

As in our previous work (Paraskevoudi & SanMiguel, 2021), we hypothesized that the electrophysiological and neuromodulatory effects at encoding (i.e., sensory suppression and pupil dilation for MA events) might be driving any memory encoding differences between A and MA sounds, and that neuromodulation might be behind the suppression of ERP responses to MA sounds. To assess these relationships, we tested for possible correlations between the behavioural, electrophysiological and neuromodulatory (i.e., pupil diameter) effects of actions, separately for each delay condition and predictability session. Only those differences between MA and A events that were found to be significant in the previous analyses were introduced in the correlation analyses. For all the behavioural and the electrophysiological effects, we first calculated the difference by subtracting the MA from A values (i.e., difference in memory performance and ERP amplitude for each component of interest between A and MA, separately for each delay condition and predictability session). Regarding the ERPs identified in two electrodes (e.g., Na, Tb, and $N1_{mast}$), we calculated the mean amplitude across the two (T7/T8 and M1/M2, respectively). For the pupil data, we used the peak of the difference wave between A and MA events at encoding for each condition separately (150/750 ms and predictable/unpredictable session). We then submitted these values to a Pearson correlation coefficient to test for correlations between a) the effects on ERPs at encoding and memory performance, b) the neuromodulatory effects at encoding and memory performance, and c) the effects on the ERPs and the neuromodulatory effects at encoding. In all correlations, for the ERPs, the larger the attenuation effects for the negative (N1, Na, Tb) and positive ($N1_{mast}$, P2) components, the more negative and positive the values, respectively. Conversely, for the pupil and the behavioural data, the more negative the value, the larger the pupil diameter and the worse the memory performance for MA sounds.

We complemented theses analyses with corresponding Bayesian correlation analyses. For all Bayesian correlations, the Bayes factor (BF_{10}) for the alternative hypothesis (i.e., correlation is not equal to zero and can be either positive or negative) was calculated (using the function *correlationBF* of the *BayesFactor* package in R). The Bayes factor, and its interpretation, was based on Jeffreys' (1961) test for linear correlation. Specifically, Bayes factors below 1 are thought to provide evidence against the alternative hypothesis, while weak, moderate, strong, and very strong evidence for the alternative hypothesis is provided by Bayes factors in the ranges 1 – 3, 3 – 10, 10 – 30, 30 – 100, respectively (Jeffreys, 1961). Bayes factors above 100 provide extreme evidence in support of the alternative hypothesis.

3. Results

All statistical analyses were performed using R (version 3.6.0). For all the post-hoc comparisons following significant ANOVAs, we used the Bonferroni correction.

3.1. *Behavioural performance*

For the analysis of the behavioural data, we calculated the percent correct (i.e., memory performance) for sounds that were encoded as MA or A, separately for each Predictability session and Delay condition. The 3x2 ANOVA (Source: A vs. MA, Predictability: Predictable vs. Unpredictable, Delay: 150 vs 750 ms) showed only a main effect of Source, $F(1, 19) = 8.98$, $p = .007$, $\eta_p^2 = .32$, with better performance for sounds encoded as MA compared to those encoded as A ($M_{MA} = .72$, $M_A = .69$, $SD_{MA} = .12$, $SD_A = .14$). The rest of the effects did not reach significance (all *ps* > .05).

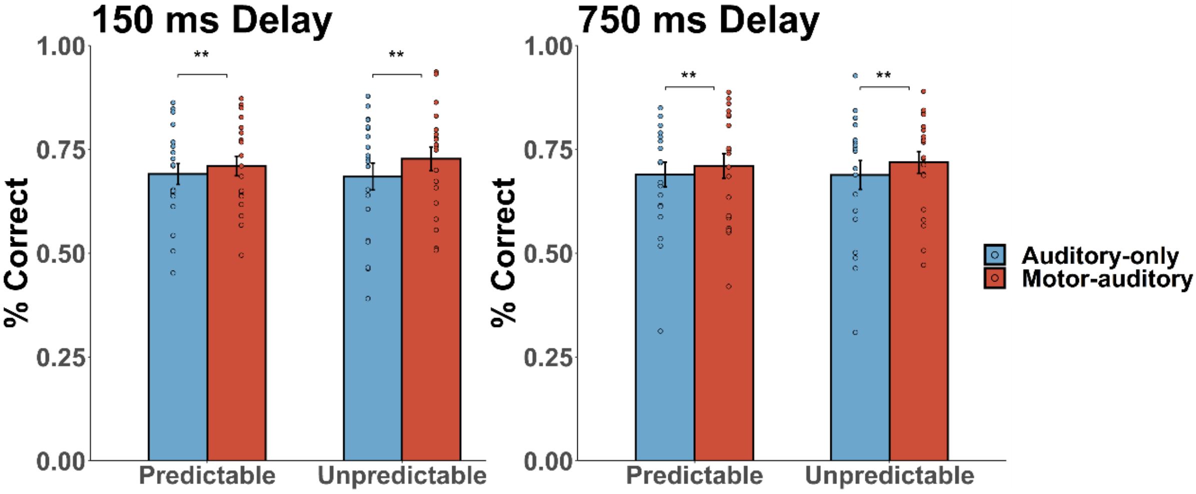


Figure 2. Summary of the behavioural findings showing a significant main effect of Source, with better performance for sounds that were encoded as MA compared to those encoded as A, irrespective of predictability and delay. Error bars depict the standard error of the mean. Individual data points are shown for each predictability, source, and delay condition.

3.2. *Evoked responses at encoding*

3.2.1. *Auditory responses as a function of source and predictability for each delay*

Figures 3 and 5 show all the studied peaks for the encoding data at the relevant electrodes for each peak, for the 150 and the 750 ms delay, respectively. For the 150 ms delay, we performed visual and motor-visual correction for the A and MA sounds, respectively (see Methods). The corrected data is presented in Figure 3. For the 750 ms, no correction was required due to the longer delay between button press and/or colour-filling of the circle that did not affect auditory responses. The time windows defined for each peak were the following: Na 70–90 ms, Tb 120–150 ms, N1 92–116 ms, $N1_{mast}$ 88–118 ms, P2 180–230 ms. Separately for each delay condition, we performed within-subjects ANOVAs (Source x Predictability) on N1 and P2 (at Cz), N1 at mastoids ($N1_{mast}$), as well as on Na and Tb responses (at T7 and T8) to test for possible interactions between self-generation and predictability. For the Na, Tb, and $N1_{mast}$, an additional factor Laterality was introduced in the ANOVA. Table 2 summarizes the main effects and interactions obtained in the analyses presented in this section. The mean amplitudes for the ERPs at encoding are presented in Supplementary Material (Tables S1, S2, S3).

<u>150 ms delay</u>

First, we assessed whether self-generation and predictability interact on N1 at Cz and at the mastoids (N1$_{\text{mast}}$), as well as the P2 at Cz. Related to the N1, we obtained a significant interaction between Source and Predictability. The post-hoc comparisons showed a significant attenuation for MA vs. A sounds in the predictable session (*t*(19) = -1.91, *p* = .036, *d* = -0.43; one-sided *t*-test with the hypothesis of observing an attenuation based on the well-established finding of attenuation for predictable self-generated sounds), and a trend towards the opposite effect for the unpredictable session (*t*(19) = -1.89, *p* = .075, *d* = 0.42; two-sided *t*-test). The comparisons between predictable and unpredictable sounds showed a significant attenuation for unpredictable A vs. predictable A sounds (*t*(19) = -2.92, *p* = .009, *d* = -0.65; two-sided *t*-test), while MA sounds did not differ as a function of predictability (*p* > .05, two-sided *t*-test). As for the N1$_{\text{mast}}$, the main effects of interest did not reach significance (but see Table 2). Related to the P2, we obtained a significant interaction between Source and Predictability, with the post-hoc comparisons showing a highly significant attenuation for MA vs. A in the predictable session (*t*(19) = 4.85, *p* < .001, *d* = 1.08; one-sided *t*-test with the hypothesis of finding lower amplitudes for MA sounds), while the same comparison in the unpredictable session did not reach significance (*p* > .05; two-sided *t*-test). Additionally, we found attenuated P2 responses for predictable compared to the unpredictable MA sounds (*t*(19) = -2.80, *p* = .011, *d* = -0.6; two-sided *t*-test), while the P2 responses to A sounds did not differ as a function of predictability (*p* > .05; two-sided *t*-test).

Table 2

Summary of the significant main effects (Source, Predictability) and interactions at encoding separately for each delay and ERP component.

ERPs	**150 ms delay**
N1	Source x Predictability: *F*(1, 19) = 7.03, *p* = .016, η_p^2 =.27
P2	Source: *F*(1, 19) = 18.32, *p* < .001, η_p^2= .49 Source x Predictability: *F*(1, 19) = 9.38, *p* = .006, η_p^2= .33
N1$_{\text{mast}}$	Laterality: *F*(1, 19) = 12.08, *p* = .003, η_p^2 = .39
Na	Non-significant.
Tb	Source x Predictability: *F*(1,19) = 7.67, *p* = .012, η_p^2= .29
	750 ms delay
N1	Source: *F*(1, 19) = 17.36, *p* < .001, η_p^2 = .48
P2	Predictability: *F*(1, 19) = 7.05, *p* = .016, η_p^2= .27
N1$_{\text{mast}}$	Laterality: *F*(1, 19) = 11.11, *p* = .003, η_p^2 = .37
Na	Laterality: *F*(1, 19) = 6.58, *p* = .019, η_p^2= .26
Tb	Predictability: *F*(1, 19) = 6.46, *p* = .02, η_p^2 = .25

Next, we assessed whether Na and Tb are modulated by self-generation and/or predictability. Related to the Na, we did not obtain any significant results. In contrast, the analysis on the Tb showed only one significant interaction between Source and Predictability, with post-hoc tests showing a significant attenuation for MA compared to A sounds in the predictable session (*t*(39) = -2.01, *p* = .026, *d* = -0.32; one-sided *t*-test with the hypothesis of finding lower amplitudes for MA sounds), while the opposite effect (i.e., enhanced Tb for MA vs. A) was observed in the unpredictable session (*t*(39) = 3.66, *p* < .001, *d* = -0.71; two-sided *t*-test). Additionally, A sounds differed as a function of predictability (i.e., lower for unpredictable, *t*(39) = -4.48, *p* < .001, *d* = -0.71; two-sided *t*-test), which was not found in the case of MA sounds (*p* > .05; two-sided *t*-test).

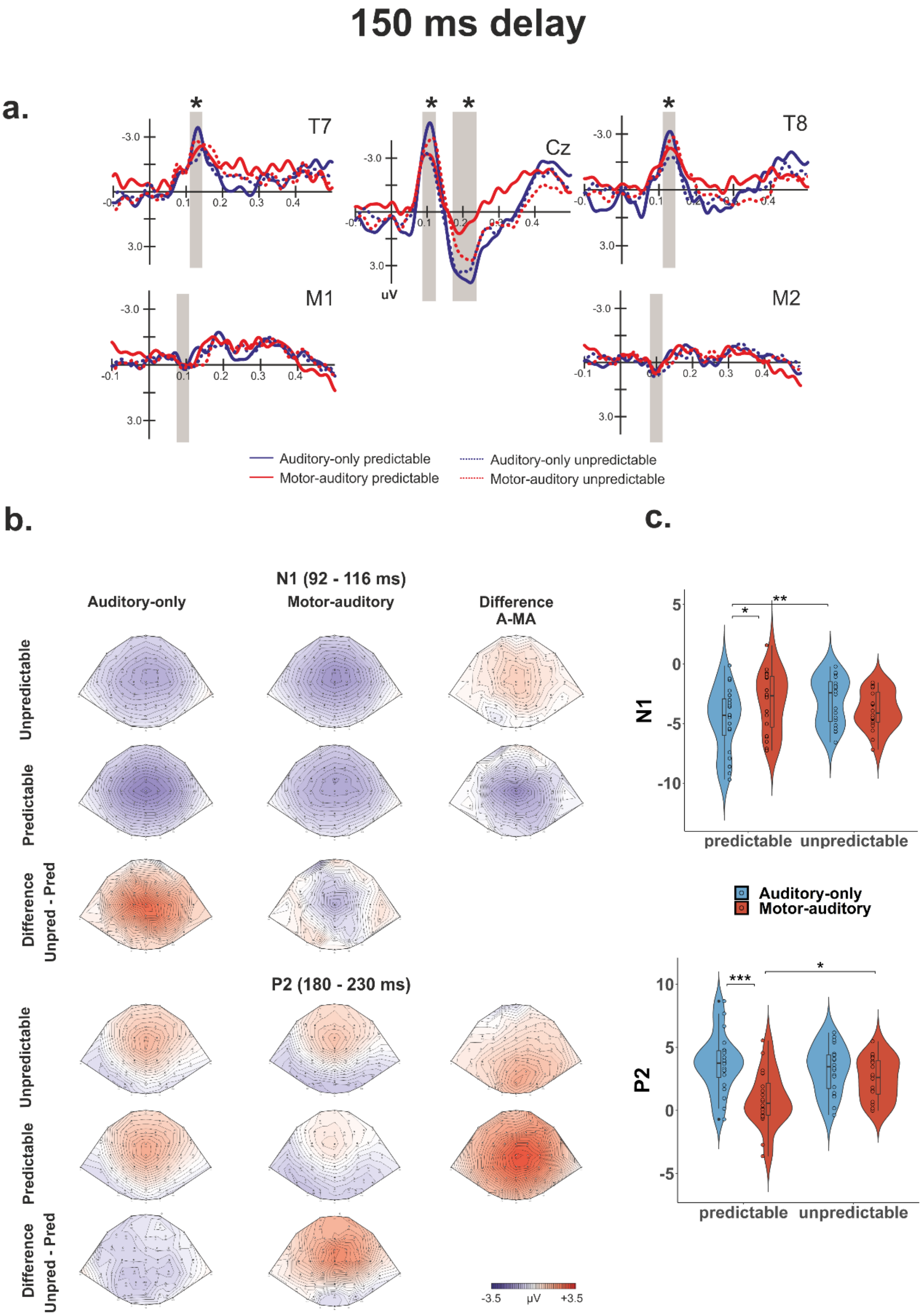


Figure 3. **a)** Group-average event-related potentials across 20 participants for the motor-auditory (red) and auditory-only (blue), in the predictable and unpredictable sessions (solid and dashed lines, respectively) for sounds presented 150 ms after the button press and/or colour-filling of the cue. The motor- and visually-

corrected averages are depicted. Time windows used for the analyses are indicated in gray (Na: 70–90 ms, Tb: 120–150 ms, N1: 92–116 ms, $N1_{mast}$: 88–118 ms, P2: 180–230 ms). Significant differences in the event-related potentials are indicated by an asterisk. **b)** N1 and P2 scalp topographies in the time windows for: the auditory-only condition (left); the corrected motor-auditory condition (middle); and the (A–[MA–M]) difference waves, separately for the predictable and the unpredictable session (right). The third row represents the difference (Unpredictable – Predictable) for the auditory-only (left) and motor-auditory sounds (right). **c)** Violin plots representing the data and differences between conditions for the N1 and P2 at encoding. The width of each curve corresponds with the approximate frequency of data points in each region. In the middle of each density curve is a small box plot, with the rectangle showing the ends of the first and third quartiles and central dot the median. Asterisks (‘*’, ‘**’, ‘***”) indicate *p* values smaller than .05, .01, and .001, respectively.

In addition to the ERP analyses, we tested for further effects during encoding using a cluster-based permutation approach (Figure 4). Specifically, we tested for a main effect of Source (A – MA), a main effect of Predictability (Unpredictable – Predictable), as well as an interaction between the two factors by comparing the difference waves of the self-generation effects across sessions (i.e., [A – MA in Unpredictable] – [A – MA in Predictable]). For the short delay data, these analyses showed two significant clusters for the main effect of Source, one positive one in the window 150 – 312 ms post-stimulus (*p* = .004), showing central-parietal distribution and reminiscent of the topography of the P2/P3 complex, and a later negative one between 396 and 548 ms post-stimulus (*p* = .011), showing fronto-parietal distribution. No significant clusters were found for the main effect of Predictability; however, we obtained two significant negative clusters for the interaction between Source and Predictability, the first one -178 to -16 ms pre-stimulus (*p* = .002) and the second one in the window 196 to 390 ms post-stimulus (*p* = .001). Post-hoc comparisons within the significant windows of interactions did not show significant differences between A and MA sounds in the unpredictable session, nor between A sounds as a function of predictability. However, A and MA sounds differed significantly in the predictable session, as shown by the two significant positive clusters. The first cluster was found between -178 up to -20 pre-stimulus (*p* = .001), showing a central distribution, while the second one was found in the window 196 to 380 post-stimulus (*p* = .001), mainly over central and parietal sensors. Additionally, we obtained two significant positive clusters when comparing MA unpredictable vs. MA predictable sounds. The first one lasted from -150 until -22 ms pre-stimulus (*p* = .003) and was mainly observable over parietal and occipital areas, probably reflecting differences in processing the self-generated visual stimulus that was colour-filled right after the button press (i.e., 150 ms prior to sound onset). The second cluster lasted from 198 until 390 ms post-stimulus (*p* < .001) and was observable over fronto-central areas.

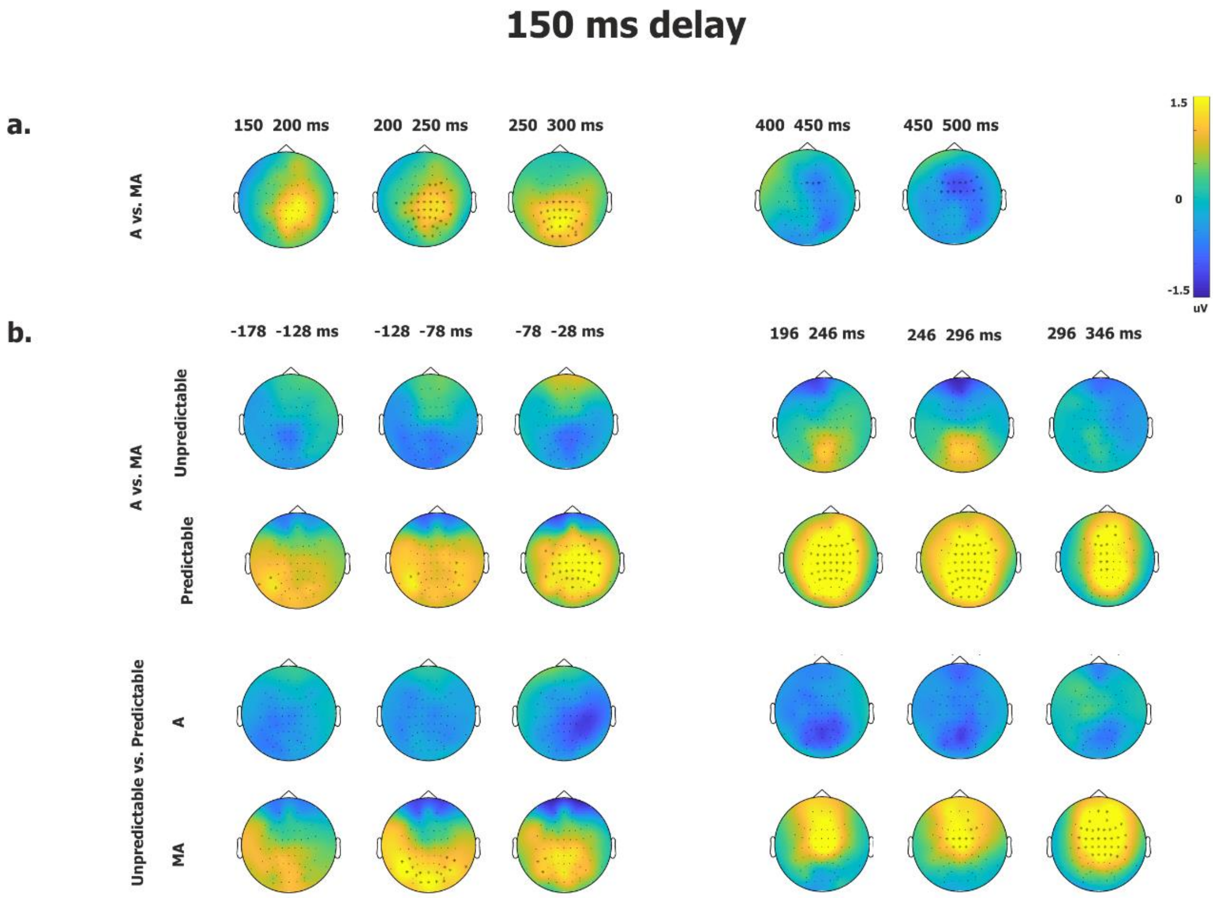


Figure 4. Summary of the main findings derived from the cluster-based permutation analyses for the 150 ms delay. Clusters are averaged within the time interval of 50 ms (time intervals in ms are shown). Topographical maps are color-coded according to the amplitude of the difference. Clusters of electrodes with significant difference between the two conditions are marked in '*' sign ($p < .05$). **a)** Two significant clusters for the main effect of Source (A vs. MA), in the windows 150 – 312 ($p = .004$) and 396 – 548 ($p = .011$). **b)** Two significant clusters for the interaction between Source and Predictability in the windows -178 up to -20 pre-stimulus ($p = .001$) and 196 – 380 post-stimulus ($p = .001$). Post-hoc comparisons showed that the significant differences were limited in the comparisons between A vs. MA in the predictable session (in the windows -178 up to -20 pre-stimulus and 196 to 380 post-stimulus; both *ps* = .001) and MA unpredictable vs. MA predictable sounds (in the windows -150 until -22 ms pre-stimulus and 198 until 390 ms post-stimulus; $p = .003$ and $p < .001$, respectively).

750 ms delay

Related to the N1, we only obtained a highly significant main effect of Source on N1, showing attenuation for MA compared to A sounds. As for the P2, we obtained only a significant main effect of Predictability, showing suppressed responses for predictable compared to unpredictable sounds, irrespective of whether they were self-initiated or not. Lastly, as for the $N1_{mast}$, the effects of interest did not reach significance (but see Table 2). Further analyses on the N1 subcomponents at temporal sites (i.e., Na and Tb) showed only a main effect of Predictability on Tb, with lower amplitudes for the predictable compared to unpredictable sounds.

750 ms delay

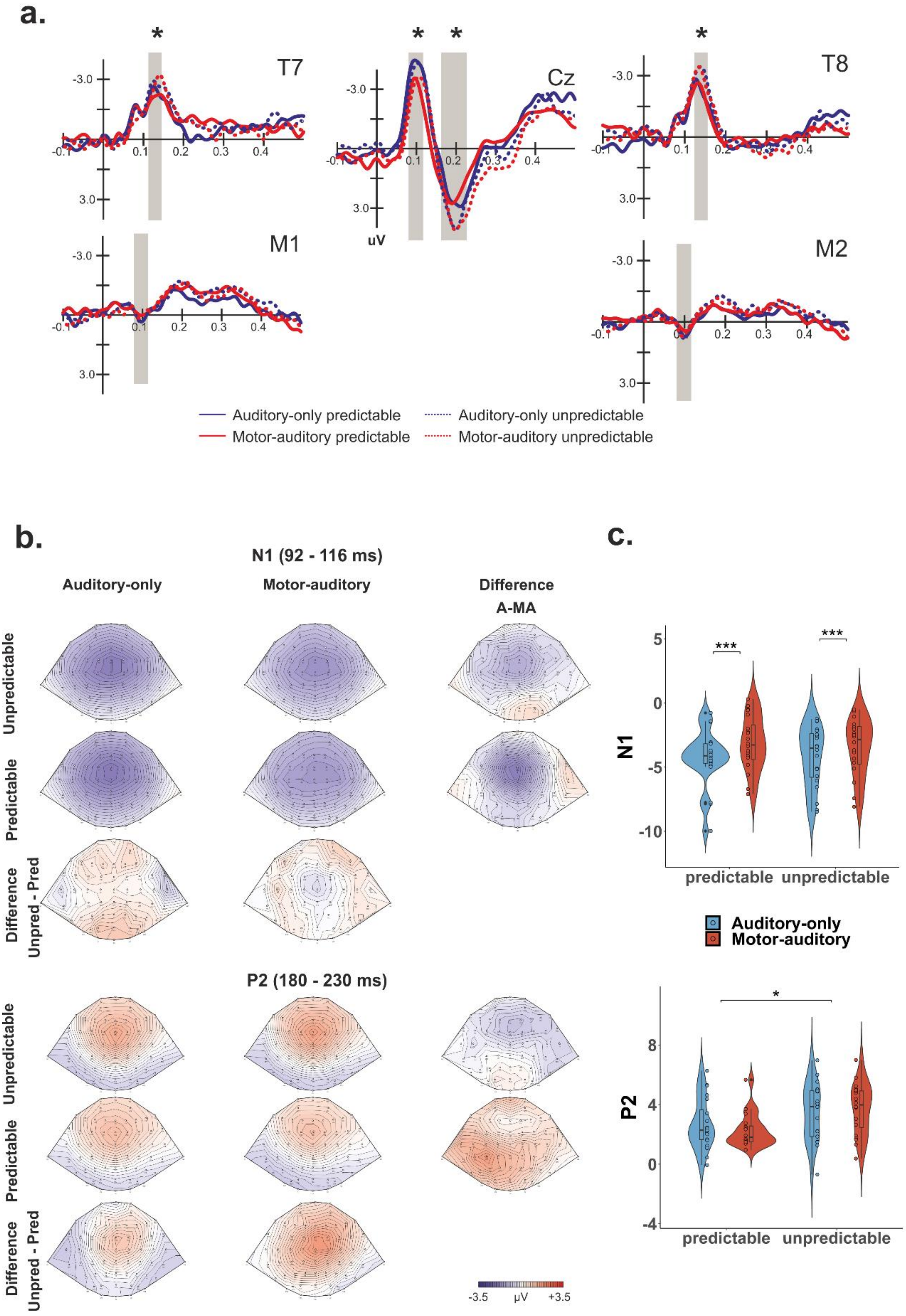

a.
*
T7
-3.0
3.0
-0.1
0.1
0.2
0.3
0.4
*
*
Cz
uV
*
T8
M1
M2
Auditory-only predictable
Motor-auditory predictable
Auditory-only unpredictable
Motor-auditory unpredictable
b.
N1 (92 - 116 ms)
Auditory-only
Motor-auditory
Difference
A-MA
Unpredictable
Predictable
Difference
Unpred - Pred
P2 (180 - 230 ms)
-3.5
μV
+3.5
c.
N1
5
0
-5
-10
***
***
predictable
unpredictable
Auditory-only
Motor-auditory
P2
8
4
0
-4
*
predictable
unpredictable

Figure 5. **a)** Group-average event-related potentials across 20 participants for the motor-auditory (red) and auditory-only (blue), in the predictable and unpredictable sessions (solid and dashed lines, respectively) for sounds presented 750 ms after the button press and/or colour-filling of the cue. Given the longer delay, the data did not require motor and visual correction. Time windows used for the analyses are indicated in gray (Na: 70–90 ms, Tb: 120–150 ms, N1: 92–116 ms, $N1_{mast}$: 88–118 ms, P2: 180–230 ms.). Significant differences in the event-related potentials are indicated by asterisks. **b)** N1 and P2 scalp topographies in the time windows for: the auditory-only condition (left); the motor-auditory condition (middle); and (3) the (A – MA) difference waves, separately for the predictable and the unpredictable session. The third row represents the difference (Unpredictable – Predictable) for the auditory-only (left) and motor-auditory sounds (right). **c)** Violin plots representing the data and differences between conditions for the N1 and P2 at encoding. The width of each curve corresponds with the approximate frequency of data points in each region. In the middle of each density curve is a small box plot, with the rectangle showing the ends of the first and third quartiles and central dot the median. Asterisks (‘*’, ‘***”) indicate p values smaller than .05 and .001, respectively.

As we did for the 150 ms delay data, we tested for further effects during encoding using a cluster-based permutation approach, that is, we tested for a main effect of Source (A – MA), a main effect of Predictability (Unpredictable – Predictable), as well as an interaction between the two factors by comparing the difference waves of the self-generation effects across sessions (i.e., [A – MA in Unpredictable] – [A – MA in Predictable]). We first obtained two significant clusters when comparing A and MA sounds irrespective of predictability: A first positive one, lasting from -250 until -110 ms pre-stimulus (*p* = .01) initially observable over parietal areas and later over central-parietal sensors on the right hemisphere, presumably related to motor activity (since no correction was performed for the long delay data), and a second negative one that was present from 456 until 522 ms post-stimulus (*p* = .04) that was maximal over central-parietal areas. Additionally, when testing for a main effect of Predictability (Unpredictable vs. Predictable), we obtained a significant positive cluster starting 184 ms after sound onset and lasting up to 386 ms post-stimulus (*p* = .01) that was maximal over central sensors, possibly reflecting the effect of predictability on P2 in the ERP analyses. Finally, we also obtained a significant negative cluster when testing for interactions between Source and Predictability in the window -250 until -46 pre-stimulus (*p* < .001). The post-hoc comparisons showed indeed a significant positive cluster in this window for the A vs. MA comparison in the predictable session (*p* < .001) that was observable over parietal and occipital sensors. The comparison between A and MA sounds in the unpredictable session did not yield any significant clusters. We also obtained a significant negative cluster when comparing A unpredictable vs. A predictable sounds that started -250 and lasted until -118 ms pre-stimulus (*p* = .002) and was maximal over central and parietal areas. Conversely, the comparison between MA unpredictable and MA predictable sounds did not yield any significant clusters (*p* > .05).

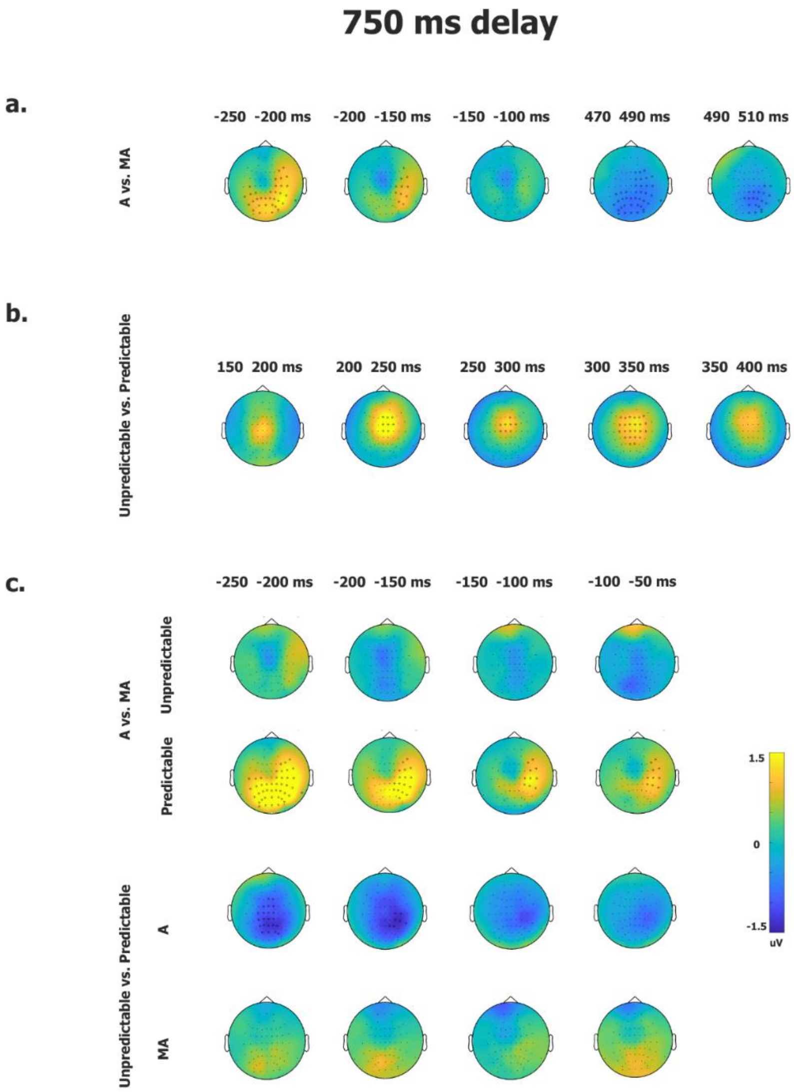


Figure 6. Summary of the main findings derived from the cluster-based permutation analyses for the 750 ms delay. Clusters are averaged within the time interval of 50 ms (time intervals in ms are shown), except for the second cluster in (a). Topographical maps are colour-coded according to the amplitude of the difference. Clusters of electrodes with significant difference between the two conditions are marked in '*' sign ($p < .05$). **a)** Two significant clusters for the main effect of Source (A vs. MA), in the windows -250 until -110 ms pre-stimulus ($p = .01$) and 456 – 522 ms post-stimulus ($p = .04$). **b)** One significant cluster for the main effect of Predictability (Unpredictable vs. Predictable) in the window 184 – 386 ms post-stimulus ($p = .01$). **c)** One significant cluster for the interaction between Source and Predictability in the window -250 until -46 pre-stimulus ($p < .001$). Post-hoc comparisons showed that the significant differences were limited in the comparisons between A vs. MA in the predictable session (in the window -250 until -46 pre-stimulus; $p < .001$) and A unpredictable vs. A predictable (in the window -250 until -118 ms pre-stimulus; $p = .002$).

3.2.2. Comparison of the self-generation and predictability effects across delays

To address the possible three-way interactions between Source, Predictability and Delay, we calculated the self-generation (i.e., Source) effects (A - MA difference waves) and assessed how they were modulated by Predictability and Delay, and we calculated the predictability effects (Unpredictable – Predictable difference waves) and assessed how they were modulated by Source and Delay. To this end, we performed 2x2 ANOVAs (Predictability x Delay on self-generation effects and Source x Delay on predictability effects) on the same components and windows as before: Na: 70–90 ms, Tb: 120–150 ms, N1: 92–116 ms, $N1_{mast}$: 88–118 ms, P2: 180–230 ms). For components identified in more than one electrode, an additional factor of Laterality was introduced in the model. Figure 7 shows the effects of self-generation (a) and predictability (b) represented as the difference A – MA and Unpredictable – Predictable, respectively, in violin plots. A summary of the main effects and interactions obtained in the analyses of this section is reported in Table 3.

Self-generation effects as a function of predictability and delay

Delay and Predictability interactively modulated self-generation effects on N1, Tb, and P2. The post-hoc comparisons for all three components showed that within the 150 ms delay condition, action-driven suppression was significantly more pronounced for the predictable vs. the unpredictable session (N1: $t(19) = -2.72$, $p = .014$, $d = -0.61$, Tb: $t(19) = -3.67$, $p < .001$, $d = -0.58$, P2: $t(19) = 3.80$, $p = .001$, $d = 0.85$) while N1-, P2-, and Tb-suppression for self-generated sounds did not vary as a function of predictability in the 750 ms delay (all *ps* > .05). Additionally, the magnitude of the suppression of these components for self-generated sounds differed between the delays as a function of predictability: For N1 and Tb, motor-driven suppression appeared significantly larger in the 750 ms compared to the 150 ms delay, but only in the unpredictable session ($t(19) = 3.90$, $p < .001$, $d = 0.87$ and $t(19) = 4.04$, $p < .001$, $d = 0.64$) in which for the 150 ms delay we found a significant Tb-enhancement and a trend towards N1-enhancement rather than suppression for self-generated sounds (see above). In contrast, for P2 the suppression effects were larger in the 150 compared to the 750 ms delay only in the predictable session ($t(19) = 3.54$, $p = .002$, $d = 0.79$), but not in the unpredictable one, in which motor-driven P2 attenuation was not found in either delay (see above). As for the $N1_{mast}$ and Na, we did not find significant effects.

In sum, all these interactions confirm that self-generation effects were modulated by predictability at short delays only, with self-generation effects being largest for all components when sounds were predictable and presented at a 150 ms delay, while when sounds were unpredictable and presented at 150 ms delay, there was a general trend towards sensory enhancement rather than suppression.

Predictability effects as a function of source and delay

The source of the predictability (a self-chosen action in MA sounds or visual cue in A sounds) modulated predictability effects on N1, P2 and Tb (see Table 3 and Figure 7). We also found a main effect of Delay on predictability effects for the Tb. However, the effects of Source and Delay on predictability interacted for N1 and Tb, while they did not interact for P2. Finally, we also obtained an interaction between Source and Delay on the predictability effects for N1 and Tb, ($F(1, 19) = 4.52$, $p = .047$, $\eta_p^2 = .19$, and $F(1, 19) = 8.39$, $p = .009$, $\eta_p^2 = .31$), while they did not interact for P2. The main effect of Source on P2 reflected larger P2-suppression for predictable compared to unpredictable sounds when predictability was given by a self-chosen action (i.e., MA sounds), compared to when predictability was afforded by a visual cue (i.e., A sounds). For the N1 and Tb, however, predictability effects were modulated by source in the 150 ms delay only, where predictability afforded by a visual cue enhanced responses, while in comparison predictability afforded by the action had no significant effects (N1: $t(19) = 2.65$, $p = .016$, $d = 0.59$; Tb: $t(39) = 3.53$, $p = .001$, $d = 0.56$). Conversely, in the 750 ms delay, predictability effects on the N1 and Tb were not modulated by Source (both *ps* > .05). Recall that at the 750 ms delay, predictability effects consisted of a significant Tb suppression for predictable sounds (main effect over A and MA of Predictability, see above) while predictability did not show any significant effects on N1 (see Table 2 above). Additionally, on the N1 and Tb components, the effects of predictability afforded by visual cues (A sounds) were significantly different at the short compared to the longer cue-sound delay (N1: *t*(19) = 2.44, *p* = .025, *d* = 0.55; Tb: *t*(39) = 5.87, *p* < .001, *d* = 0.93. For both the N1 and Tb, predictability enhanced responses for A sounds at the 150 ms Delay compared to the responses for A sounds 750 ms Delay. In contrast, predictability effects afforded by motor actions (MA sounds) on the N1 and Tb components (which were in general not significant except for the main effect of predictability on Tb at the 750 ms Delay, see Table 2 above) were not modulated as a function of Delay (both *ps* > .05).

Table 3

Summary of the significant main effects and interactions at encoding for each ERP component when comparing the self-generation and predictability effects across delays.

ERPs	**Self-generation effects as a function of predictability and delay**
N1	Predictability: *F*(1, 19) = 7.12, *p* = .015, η_p^2 = .27 Delay: *F*(1, 19) = 6.61, *p* = .019, η_p^2 = .26 Predictability x Delay: *F*(1, 19) = 5.53, *p* =.030, η_p^2 = .23
P2	Predictability: *F*(1, 19) = 19.11, *p* < .001, η_p^2 =.50 Delay: *F*(1, 19) = 12.66, *p* = .002, η_p^2 = .40 Predictability x Delay: *F*(1, 19) = 6.64, *p* =.018, η_p^2 =.26)
$N1_{mast}$	Non-significant.
Na	
Tb	Delay: *F*(1, 19) = 6.18, *p* = .022, η_p^2 = .25 Predictability x Delay: *F*(1, 19) = 9.00, *p* = .007, η_p^2 =.32
	Predictability effects as a function of Source and Delay
N1	Source: *F*(1, 19) = 6.46, *p* = .020, η_p^2 = .25 Source x Delay: *F*(1, 19) = 4.52, *p* = .047, η_p^2 =.19
P2	Source: *F*(1, 19) = 11.36, *p* = .003, η_p^2 =.37
$N1_{mast}$	Non-significant.
Na	
Tb	Source: *F*(1, 19) = 5.86, *p* = .026, η_p^2 = .24 Delay: *F*(1, 19) = 10.14, *p* = .005, η_p^2 =.35 Source x Delay: *F*(1, 19) = 8.39, *p* = .009, η_p^2 =.31

In sum, these interactions show that predictability effects were modulated by source and delay, but in different ways for the N1 and Tb components on the one hand and the P2 on the other. For the N1 and Tb, again the results confirm that Source and Predictability interact at the 150 ms Delay only, where visually-driven predictability, but not motor-driven predictability enhanced these components. The findings are more inconclusive regarding the P2 component, since the interaction reported in this section (i.e., Source x Delay on predictability effects for P2) indicates that motor-driven predictability led to higher P2-suppression than visually-driven predictability irrespective of the delay; however, a significant interaction between source and predictability was only found in the 150 ms delay and not in the 750 ms delay when analyzed separately (see above). All in all, the data shows a general tendency towards suppressed responses, especially the P2, for more predictable sounds that were self-generated and/or presented at the longer delay; whereas there is a generalized tendency towards enhanced responses, especially the N1 and Tb, for more predictable sounds that were externally-generated and presented at the short delay.

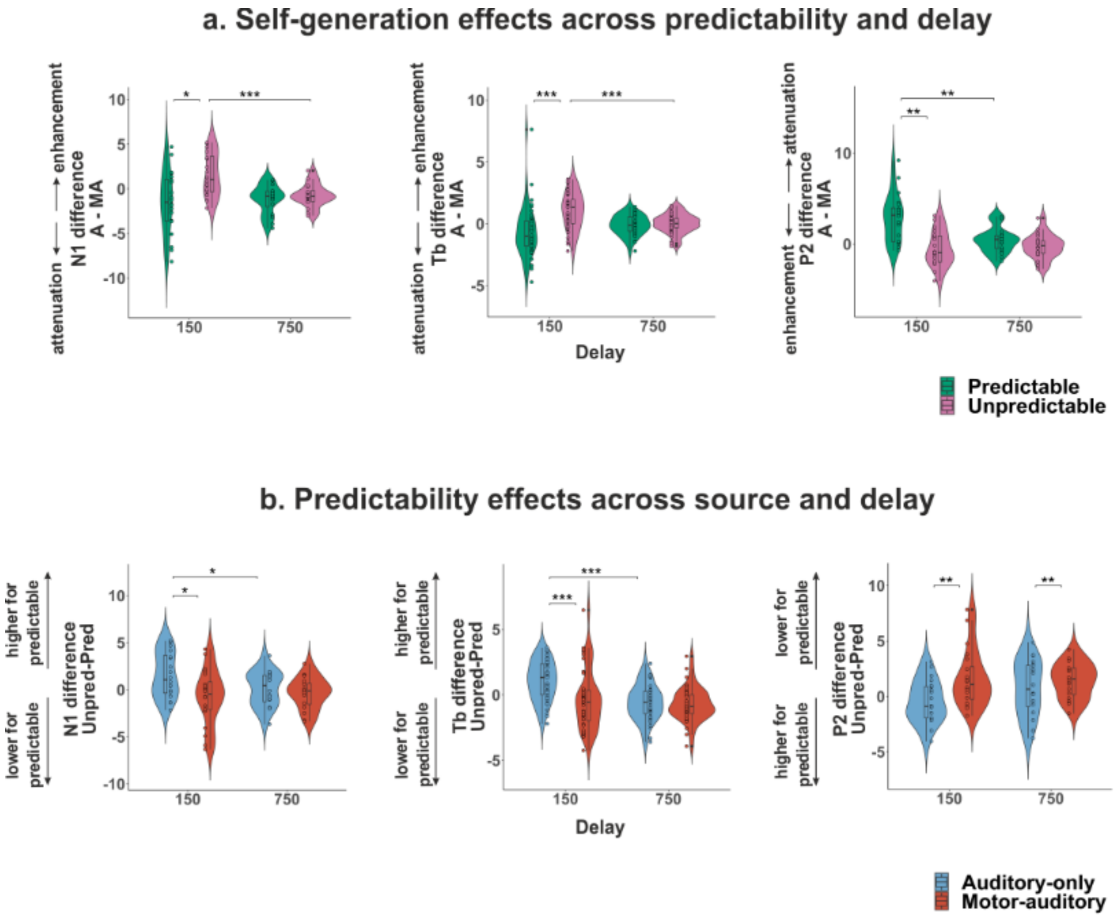


Figure 7. Violin plots representing the magnitude of the self-generation (a) and the predictability (b) effects for the N1, Tb, and P2. The width of each curve corresponds with the approximate frequency of data points in each region. In the middle of each density curve is a small box plot, with the rectangle showing the ends of the first and third quartiles and central dot the median. Asterisks (‘*’, ‘**’, ‘***”) indicate *p* values smaller than .05, .01, and .001, respectively. **a)** Self-generation effects (A – MA) as a function of predictability and delay: Larger attenuation effects correspond to more negative values for the N1 and Tb, and more positive values in the y-axis for the P2. **b)** Predictability effects (Unpredictable – Predictable) as a function of source and delay: Enhancing effects of predictability are represented with more positive values for the N1 and Tb amplitudes, and more negative values for the P2 amplitude in the y-axis.

3.2.3. Visual responses at encoding

In addition to the auditory ERP analyses, we assessed whether self-generation and predictability interact on visual ERPs (Figure 8). These analyses focused on whether the visual responses after the button press are modulated by predictability context (see Methods). This analysis showed significantly larger N1 responses to predictable self-generated stimulation (M = -5.98, SD = 1.93) compared to the unpredictable one, $t(19) = -2.84$, p = .011, d = -0.63 (M = -4.85, SD = 2.44). Note that prediction for the upcoming visual stimulation could only differ between the button presses in predictable and unpredictable sessions (i.e., only in the predictable session the button press could predict which circle would be colour-filled). In contrast, predictions about the visual stimulation could not be formed for the passive A-events in the predictable and unpredictable sessions. Therefore, one would expect no differences when comparing visual N1 responses to A-predictable and A-unpredictable events locked to the filling of the cue. Indeed, a two-sided *t*-test showed no differences in visual N1 between passive A events as a function of predictability session (p > .05).

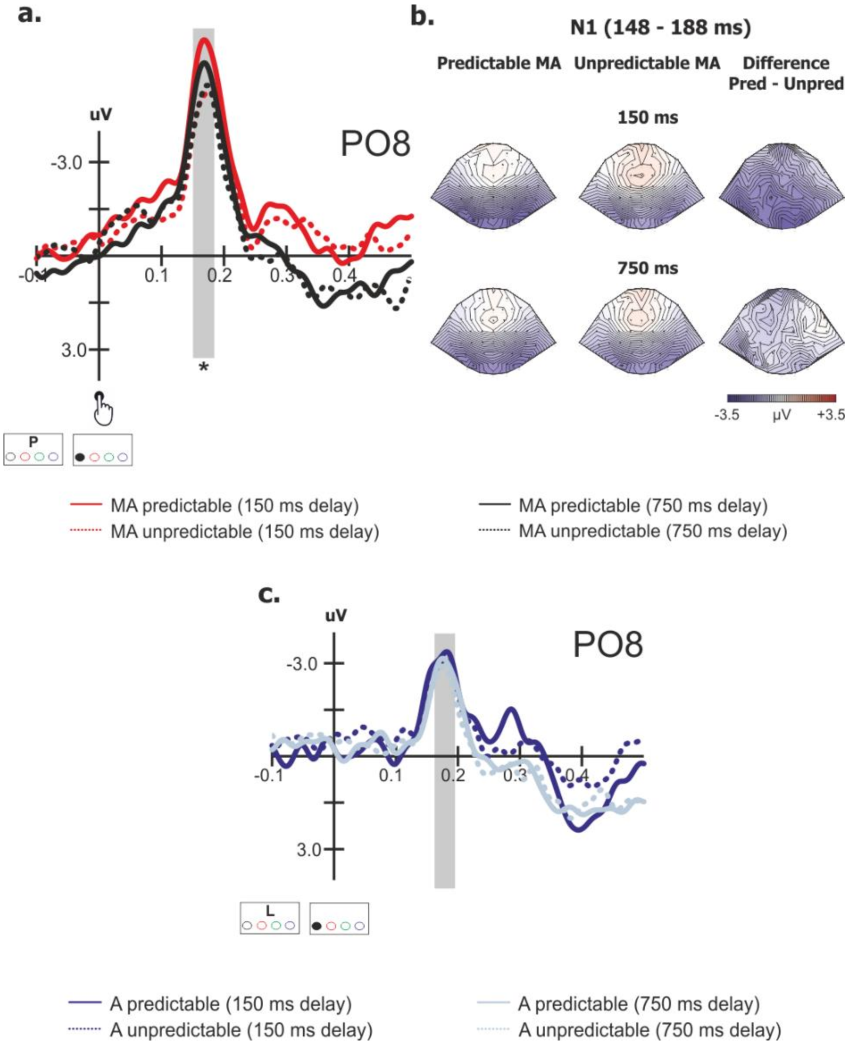


Figure 8. Group-average event-related potentials across 20 participants, focusing on visual responses. The responses are locked to the button press and/or the colour-filling of the cue. Time windows used for the analyses are indicated in gray. The N1 window for the MA events was 148–188 ms and 156–196 for the A events (defined ± 20 ms around the peak for each comparison). The mean N1 amplitudes between the 150 and 750 ms delay conditions for each participant was included in the analyses. Significant findings are indicated by an asterisk. **a)** Visual N1 responses for the MA events following a button press that could predict which circle would be colour-filled (predictable session) or not (unpredictable session), separately for each delay. Responses to predictable self-generated stimulation (i.e., Predictable MA) were amplified compared to the unpredictable one (p = .011; two-sided *t*-test comparing predictable vs. unpredictable MA events irrespective of delay). **b)** Visual N1 scalp topographies for the Predictable MA (left); Unpredictable MA (middle); and the difference waves Predictable MA – Unpredictable MA (right), separately for the 150 and 750 ms delay. **c)** Visual N1 responses for the A events following the passive colour-filling of the cue, separately for each delay. Responses to externally-generated visual stimulation did not differ between the predictable and unpredictable session, since the visual cue could not have been predicted in passive events (p > .05; two-sided *t*-test comparing A events between conditions irrespective of delay).

3.3. *Evoked responses at retrieval*

Figure 9 shows all the studied peaks for the retrieval conditions at the relevant electrodes for each peak. No correction was required here since all the sounds were presented passively at retrieval. Note that at retrieval all sounds were presented without any delay between fixation cross and sound onset. Nevertheless, we decided to analyze separately the sounds that were encoded under the 150- and 750-ms delay condition, respectively, since we did not know whether the delay between action (and/or cue) and sound at encoding would affect the responses to the same sound at retrieval. The time windows defined for each peak were the following: Na 72–92 ms, Tb 120–150 ms, N1 92–112 ms, $N1_{mast}$ 84–114 ms, P2 172–212 ms. Similar to the analyses for the encoding data, we performed within-subjects ANOVAs (Source x Predictability) on N1 and P2 (at Cz), N1 at mastoids ($N1_{mast}$), as well as on Na and Tb responses (at T7 and T8) to assess whether the way the sound was encoded (Encoded as A or Encoded as MA) and the predictability context (Predictable vs. Unpredictable) affected sensory responses when the sound was presented passively at retrieval and was tested for recall. For the Na, Tb, and $N1_{mast}$ an additional factor Laterality was introduced in the ANOVA. A summary of the statistics reported in this section is shown in Table 4. The mean amplitudes for the ERPs at retrieval are presented in Supplementary Material (Tables S4, S5, S6).

150 ms delay

For the 150 ms delay, we did not observe any significant effects on any of the components of interest suggesting that the way the sounds were encoded and their predictability context did not affect sensory responses at retrieval (but see Table 4 for an effect of Laterality on $N1_{mast}$).

750 ms delay

Contrary to the short delay data, analyses of the sounds that were encoded under the 750 ms delay condition showed a series of significant findings related to our effects of interest, but only on N1 and its subcomponents (see Table 4). First, related to the N1, we obtained only a significant main effect of Source, with higher N1 amplitudes for sounds that were encoded as MA compared to those encoded as A. As for the Na, we obtained a significant interaction between Predictability and Laterality, but the post-hoc comparisons (two-sided *t*-tests) showed only a highly significant difference between T7 and T8 only for the predictable session, *t*(39) = -3.90, *p* < .001, *d* = -0.62, while the rest of the post-hoc comparisons did not reach significance (all *ps* > .05). As for the Tb, first we obtained a significant interaction between Source and Predictability, with the post-hoc comparisons (all two-sided *t*-tests) showing only a significant difference between predictable and unpredictable sounds that were encoded as A, *t*(39) = 3.96, *p* < .001, *d* = 0.63, with higher amplitudes for sounds that were encoded as unpredictable. Additionally, we obtained a significant interaction between Source and Laterality, with the post-hoc comparisons (all two-sided *t*-tests) showing significant differences between T7 and T8, but only for sounds encoded as MA, *t*(39) = -2.11, *p* = .041, *d* = -0.33, as well as a significant difference between sounds encoded as A vs. those encoded as MA that was limited only to T7, with larger Tb amplitudes for sounds encoded as MA compared to sound encoded as A, *t*(39) = 2.41, *p* = .021, *d* = 0.38.

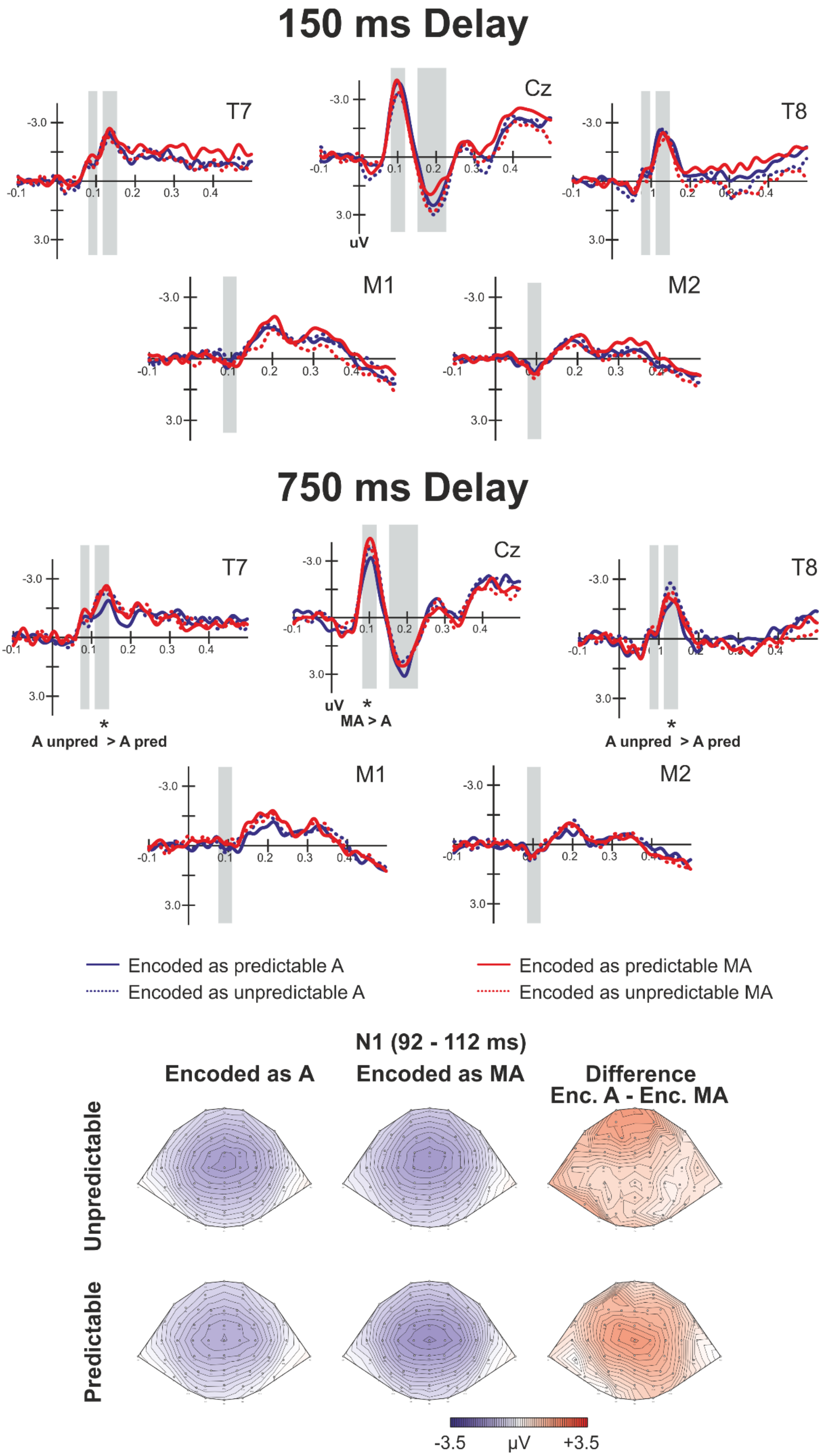

150 ms Delay
T7
Cz
T8
M1
M2
-3.0
3.0
-0.1
0.1
0.2
0.3
0.4
uV
750 ms Delay
*
A unpred > A pred
*
MA > A
Encoded as predictable A
Encoded as unpredictable A
Encoded as predictable MA
Encoded as unpredictable MA
N1 (92 - 112 ms)
Encoded as A
Encoded as MA
Difference
Enc. A - Enc. MA
Unpredictable
Predictable
-3.5
µV
+3.5

Figure 9. Group-average event-related potentials across 20 participants for the Encoded as motor-auditory (MA; red) and auditory-only (A; blue) sounds at retrieval, in the predictable and unpredictable sessions (solid and dashed lines, respectively). The data is shown separately for the responses to sounds presented 150 (top panel) and 750 ms (middle panel) after the button press and/or colour-filling of the cue at encoding. Time windows used for the analyses are indicated in gray (Na 72–92 ms, Tb 120–150 ms, N1 92–112 ms, $N1_{mast}$ 84–114 ms, P2 172–212 ms). Significant findings were obtained only for the sounds that were encoded under the 750 ms delay condition and are indicated by asterisks. Bottom panel depicts the main effect of Source on N1 for the 750 ms delay, showing the N1 scalp topographies for the Encoded as A (left); Encoded as MA (middle); and the difference waves (Encoded as A – Encoded as MA; right), separately for the predictable and the unpredictable session.

Table 4

Summary of the significant main effects (Source, Predictability) and interactions at retrieval separately for each delay and ERP component.

ERPs	**150 ms delay**
N1	Non-significant.
P2	
$N1_{mast}$	Laterality: *F*(1, 19) = 8.01, *p* = .011, η_p^2 = .30
Na	Non-significant.
Tb	
	750 ms delay
N1	Source: *F*(1, 19) = 6.84, *p* = .017, η_p^2 = .26
P2	Non-significant.
$N1_{mast}$	Laterality: *F*(1, 19) = 17.80, *p* < .001, η_p^2 = .48
Na	*F*(1, 19) = 5.56, *p* = .029, η_p^2 = .23 Predictability x Laterality: *F*(1, 19) = 9.29, *p* = .007, η_p^2 =.33
Tb	Predictability: *F*(1, 19) = 6.46, *p* = .02, η_p^2 = .25 Source x Predictability: *F*(1, 19) = 7.09, *p* = .015, η_p^2 =.27 Source x Laterality: *F*(1, 19) = 13.12, *p* = .002, η_p^2 = .41

3.4. Pupil responses at encoding and retrieval

Cluster-based permutation statistics were used to test for possible differences in pupil diameter between the conditions of interest. For the encoding (Figure 10, top panel), we locked the data to the button press and/or the colour-filling of the circle (for the motor-auditory events and for the auditory-only events, respectively) since we were mainly interested in examining the motor-driven modulation of the pupil response, rather than solely the response to the sound. To this end, we tested for possible main effects of Source (A vs. MA) and Predictability (Unpredictable vs. Predictable), as well as for interactions between these two on the pupil responses at encoding, separately for each delay. For the retrieval data (Figure 10, bottom panel), we locked responses to the sound onset since all sounds at retrieval were presented passively and without the presentation of visual cues, other than a constant fixation cross that appeared throughout the retrieval phase.

<u>150 ms delay</u>

At encoding, we obtained a significant main effect of Source starting before the button press and/or colour-filling of the cue (window: -490 to 1,630 ms, *p* < .05), showing a remarkable increase of pupil diameter after button press. We also obtained a significant main effect of Predictability (window: -200 to 760 ms, $p < .05$), with larger pupil diameter for predictable vs unpredictable events. However, this effect seems to be driven by the highly significant interaction (window: -160 to 640 ms, $p < .05$). Specifically, the post-hoc comparisons of the interaction reflected the main effect of Source by showing pupil dilation after button presses compared to the passive events both in the predictable and the unpredictable sessions (both *ps* < .05). These comparisons also showed that while predictability context did not affect pupil responses to A-events (i.e., when merely observing the colour-filling of the cue), it modulated responses following button presses with larger diameter in the predictable vs. the unpredictable session. At retrieval, we did not obtain a significant main effect of Source, however, we obtained a significant main effect of Predictability (window: 840 to 1,810 ms), showing larger pupil responses to sounds encoded in predictable contexts compared to those encoded in the unpredictable ones ($p < .05$). The interaction between Source and Predictability did not reach significance.

<u>750 ms delay</u>

For the 750 ms data at encoding, we obtained similar results as in the 150 ms delay. First, we observed a main effect of Source starting before the button press and/or colour-filling of the cue (window: -490 to 1,830 ms, $p < .05$), showing again increased pupil diameter following a button press. We also observed a significant main effect of Predictability (window: -180 to 1,170 ms, $p < .05$), with larger pupil diameter for predictable compared to the unpredictable events. Similar to the analyses for the 150 ms data, this effect seems to be driven by the significant interaction (window: -150 to 1,170 ms $p < .05$), that showed that button presses elicited larger pupil diameter compared to passive events in both predictability contexts (both *ps* < .05), and that predictability affected pupil diameter but only when an action has been performed (i.e., larger diameter for predictable vs. unpredictable events following a button press, but no differences between passive events as a function of predictability). At retrieval, we obtained a significant main effect of Source (window: 950 to 2,000 ms), showing larger pupil responses to sounds that were previously encoded as MA compared to those encoded as A ($p < .05$). The main effect of Predictability and the interaction between Source and Predictability did not reach significance.

Finally, we also aimed to test whether the effects on pupil were modulated as a function of predictability and delay. To this end, we used the difference A – MA, separately for each predictability session and delay to test for a possible interaction (i.e., ([A – MA in predictable – 150 ms] – [A – MA in predictable – 750 ms]) compared to ([A – MA in unpredictable – 150 ms] – [A – MA in unpredictable – 750 ms]). This analysis did not show a significant interaction between predictability and delay on the pupil effects (A – MA; all *ps* > .05).

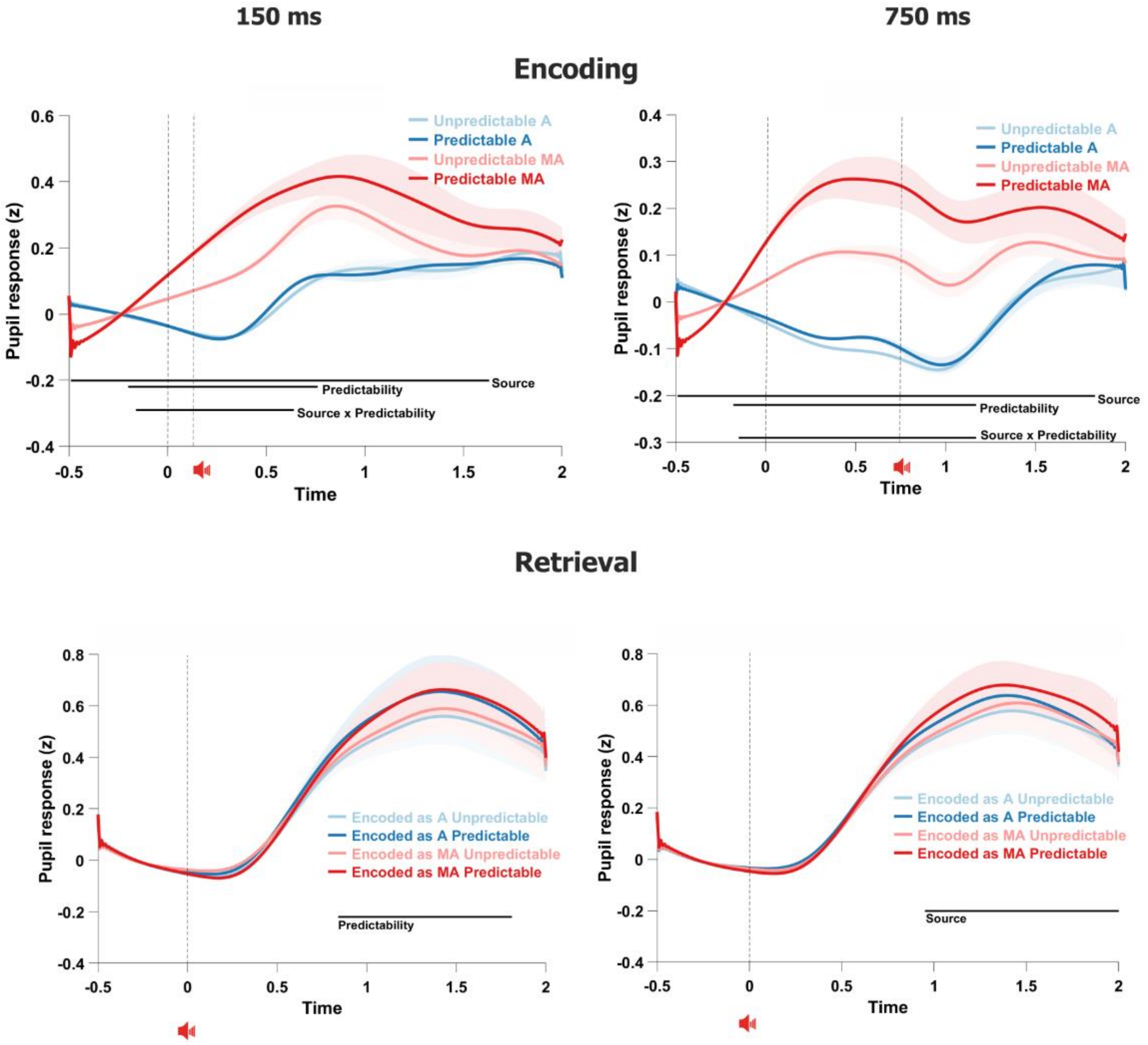


Figure 10. Pupil responses at encoding and retrieval separately for the 150 and 750 ms delay. Black bars indicate significant effects, $p < .05$ (cluster-based permutation tests). Responses at encoding (top panel) were locked to the button press and/or the colour-filling of the cue. Dashed grey lines represent the time of the button press and/or the colour-filling of the cue and the sound onset (150 or 750 ms after the press, respectively). Responses at retrieval (bottom panel) were locked to the sound. Dashed grey line represent the sound onset at retrieval.

3.5. Correlations

Next, we tested for possible correlations between the behavioural performance, pupillometric and electrophysiological data. For the correlation analyses, we focused on the significant neurophysiological effects at encoding (i.e., ERPs and pupil diameter) and the significant behavioural production effect. The effects were introduced in the correlation analyses as the difference between A and MA events separately for each delay condition and predictability session (see Methods). For the significant effects on Tb, we calculated the mean amplitude across the T7 and T8 electrodes. For the pupil data, we calculated the peak of the difference wave (A – MA) within the window of the significant interaction for each delay (i.e., -160 to 640 ms for the 150 ms delay and -150 to 1,170 ms for the 750 ms delay). All the planned correlations are reported in Table 5.

Table 5

Correlations between the self-generation effects, for each combination of predictability and delay. For each one of the ERP components, we specify the direction of the self-generation effect, that is attenuation, enhancement, or no effect (i.e., n.s.). Correlations were performed only for the significant self-generation effects. Pearson's coefficients (r), their corresponding p values, and the Bayes Factors for the alternative hypothesis (BF_{10}) are reported. Significant correlations are highlighted and indicated by an asterisk. Bayes Factors providing evidence (i.e., values > 1) in favour of the alternative hypothesis are underlined. ***a)*** *auditory responses at encoding (N1, P2, and Tb amplitudes) and memory performance,* ***b)*** *neuromodulatory effects at encoding (pupil diameter) and memory performance,* ***c)*** *auditory responses (N1, P2, and Tb amplitudes) and neuromodulatory effects (pupil diameter) at encoding.*

	Correlations between		***r***	***p***	***BF_{10}***
Predictable 150	(a) Memory performance	N1 (attenuation)	-0.094	0.69	0.51
		Tb (attenuation)	-0.047	0.84	0.48
		P2 (attenuation)	0.3	0.2	0.89
	(b) Memory performance	**Pupil dilation**	**0.48**	**0.034***	**2.79**
	(c) Pupil diameter	N1 (attenuation)	-0.38	0.1	1.38
	(c) Pupil diameter	**Tb (attenuation)**	**-0.48**	**0.031***	**2.93**
		P2 (attenuation)	0.011	0.96	0.47
Predictable 750	(a) Memory performance	N1 (attenuation)	0.31	0.18	0.97
		Tb (n.s.)	-	-	-
		P2 (n.s.)	-	-	-
	(b) Memory performance	Pupil dilation	0.19	0.43	0.61
	(c) Pupil diameter	N1 (attenuation)	-0.44	0.052	2.10
		Tb (n.s.)	-	-	-
		P2 (n.s.)	-	-	-
Unpredictable 150	(a) Memory performance	N1 (n.s.)	-	-	-
		Tb (enhancement)	0.25	0.28	0.76
		P2 (n.s.)	-	-	-
	(b) Memory performance	Pupil dilation	-0.044	0.85	0.48
	(c) Pupil diameter	N1 (n.s.)	-	-	-
		Tb (enhancement)	0.2	0.41	0.62
		P2 (n.s.)	-	-	-
Unpredictable 750	(a) Memory performance	N1 (attenuation)	-0.42	0.063	1.85
		Tb (n.s.).	-	-	-
		P2 (n.s.)	-	-	-
	(b) Memory performance	Pupil dilation	0.024	0.92	0.48
	(c) Pupil diameter	N1 (attenuation)	0.044	0.85	0.48
		Tb (n.s.)	-	-	-
		P2 (n.s.)	-	-	-

First, we tested whether the significant self-generation effects at encoding (on N1, P2, and Tb amplitudes) correlated with the significant production effect on memory performance, separately for each delay condition and predictability session (Table 5 (a)). Second, we assessed whether the difference in pupil diameter between auditory-only and motor-auditory events was related to memory performance, separately for each delay condition and predictability session (Table 5 (b)). Third, we tested for possible links between the self-generation effects obtained in the ERP analyses (i.e., N1, P2, and Tb) and the larger pupil diameter for motor-auditory events (within the window of the significant interaction for each delay), separately for each delay condition and predictability session (Table 5 (c)).

As reported in Table 5, the only significant correlations obtained were limited to the predictable session and with short action/cue-sound delays. First, we obtained a significant positive correlation between pupil diameter and memory performance ($r = 0.48$, $p = 0.034$, $BF_{10} = 2.79$), suggesting that the larger the pupil diameter for MA sounds, the worse the memory performance. Second, we obtained a significant negative correlation between Tb attenuation and pupil dilation in the predictable session with short action-sound delays ($r = -0.48$, $p = 0.031$, $BF_{10} = 2.93$), suggesting that the larger the pupil diameter, the stronger the attenuation effects for MA sounds. Both significant correlations are in line with, and extend our previous work (Paraskevoudi & SanMiguel, 2021) and their corresponding Bayes Factors provide supporting, albeit weak, evidence for the alternative hypotheses (i.e., correlation between pupil diameter and memory, and pupil diameter and Tb attenuation). The rest of the correlations did not reach significance (all *ps* > .05).

4. Discussion

In this study, we assessed whether the self-generation effects on sensory processing and memory encoding of sounds are due to stimulus-specific predictions, due to stimulus-unspecific effects of action proximity, or a mixture of both. We employed a paradigm that controlled for all aspects of predictability (*what, when, whether*) and recorded behavioural, electrophysiological, and pupil responses to predictable or unpredictable self- and externally-generated sounds that were either presented shortly after or farther away from the action or the visual cue, respectively. Our aim was to assess the effects of self-generation and predictability, as well as the potential interactions between the two on a) sensory processing and b) memory encoding of sounds, and to investigate the possible role of LC-NE activity in these effects.

4.1. Effects of actions and predictability on sensory processing

Regarding the effects of actions on sensory processing, we examined whether a) sensory attenuation (i.e., measured by auditory evoked potentials) is modulated by predictability and/or action-sound delay, b) actions form a halo of subcortical neuromodulation around them that could be reflected in the pupil diameter, and c) sensory attenuation (i.e., measured by auditory evoked potentials) and subcortical neuromodulation (i.e., measured by pupil diameter) during actions were related. We address the findings related to each one of these aims in the following paragraphs.

4.1.1. Effects of self-generation and predictability on auditory responses

At the electrophysiological level, our findings show effects of the same direction in both short and long action-sound delays; however, when the sound is presented close in time with the action, self-generation interacts with predictability. With short delays, we found the typical sensory attenuation effects in N1, Tb, and P2, even when providing identical what-, when-, and whether-predictability for passive and self-initiated sounds (cf. Klaffehn et al., 2019; Harrison et al., 2021). The attenuation effects were, however, only obtained when the action was fully predictive of the sound, in line with work showing stronger attenuation with more precise match between action-driven sensory predictions and sensory feedback (for a review see Schröger et al., 2015). However, self-generation and predictability no longer interacted when the sound was farther away from the action, affecting N1, Tb, and P2 in differential ways: N1 attenuation was driven by self-generation, while Tb and P2 modulations were driven by predictability.

The first question that arises from these findings is: What can explain the N1 attenuation irrespective of predictability only in the longer delays? One possibility is that with longer action-sound delays, differences in temporal expectations (i.e., expressed as the hazard function) were minimized between predictable and unpredictable sessions, since the conditional probability of an event occurring at a given time given that it has not yet occurred (Nobre et al., 2007) evidently increases as time passes (Janssen & Shadlen, 2005; Yang & Shadlen, 2007). This could possibly have allowed for predicting the temporal onset of the stimulus irrespective of the predictability context, effectively giving rise to a genuine N1 attenuation effect that seems to be driven by both stimulus-specific and unspecific processes (since stimulus-specific effects should occur only when the action is predictive of the stimulus). Therefore, it seems likely that among the three facets of predictions we manipulated here, temporal predictability might have a special role in driving the N1 suppression effects (cf. Schafer & Marcus, 1973). This possibility could be also supported by findings showing that N1 suppression prevails even with 50% action-sound contingency (Paraskevoudi & SanMiguel, 2021), since the action still affords some temporal predictions about the timing of the upcoming sound. Given that equal predictability was provided for both A and MA sounds in the present study, we argue that the prevalence of the suppression effects points to more robust temporal predictions afforded by actions compared to those afforded by sensory cues.

Contrary to N1, P2 was only modulated by predictability (i.e., attenuation for predictable compared to unpredictable sounds; cf. Behroozmand et al., 2016; Chen et al., 2012) with longer action-sound delays (see also Klaffehn et al., 2019 for no differences in P2 between A and MA sounds with longer delays), suggesting that the self-initiated sound was processed as an externally-generated input (Pinheiro et al., 2019). In line with our findings, P2 amplitude is known to increase with increasing action-sound delays (Timm et al., 2016), pointing to a dampening effect of actions on P2 (but not N1), but only when the sound is presented shortly after the action and is predicted by it. A functional dissociation between N1 and P2 has been already proposed, mainly supported by work showing that N1- and P2-suppression are differently modulated by experimental manipulations (e.g., variations in press-to-press intervals, differences in the stimulus generating effector, or action-sound delays; SanMiguel et al., 2013; van Elk et al., 2014; Timm et al., 2016) and cerebellar lesions (Knolle et al., 2012; 2013b), while further evidence reports a link between sense of agency and P2, but not with N1 (Ford et al., 2014; Timm et al., 2016). Interestingly, Tb responses followed a pattern similar to the one observed for P2, (i.e., predictability-driven, rather than action-driven, attenuation) and in fact, recent work has raised the possibility that Tb attenuation – similar to P2 attenuation – might be also dependent on participants' sense of agency (i.e., the feeling of control over actions and their consequences; Han et al., 2021; Gallagher, 2000). Therefore, it is possible that self-generation no longer affects Tb and P2 with decreased sense of agency (cf. Timm et al., 2016), that is, when the sound is far away from the action.

Collectively, these findings replicate work showing larger attenuation effects with more precise match between predictions and sensory feedback (Fu et al., 2006; Houde et al., 2002; Baess et al., 2008), but only when the action immediately triggers a highly predictable sound. Although these findings might provide support to the stimulus-specificity of the effects, it should be noted that we did not find attenuation of the specific N1-component at the mastoids (thought to partly originate from a supratemporal generator; Vaughan & Ritter, 1970). Nevertheless, the lack of polarity reversal does not mean that the specific component of N1 is not modulated; rather, it suggests that attenuation of the supratemporal component of N1 might overlap with the unspecific component that can be simultaneously modulated during self-generation (Horváth et al., 2012).

Additionally, we provide evidence showing opposite effects of predictability as a function of self-generation, namely we observed reduced N1, Tb, and P2 responses for predictable vs. unpredictable sounds in the presence compared to the absence of an action. Specifically, N1 (and Tb) and P2 were differently affected by the interplay of action and predictability, such that with shorter delays, predictability enhanced N1 and Tb responses to predictable sounds in the absence of action (but no differences as a function of predictability in the presence of action), while it suppressed P2 responses for predictable stimuli triggered by an action (but no differences as a function of predictability in the absence of action). In contrast, with longer delays, predictability seems to exert effects only to the Tb and P2 components, suppressing responses to predictable sounds irrespective of the presence of action, supporting previous work suggesting that Tb and P2 might be sensitive to temporal predictions (Behroozmand et al., 2016; Chen et al., 2012; Sowman et al., 2012). These findings suggest that the mere presence of an action determined the direction of the effects and might point to distinct effects of motor- and sensory-based predictions on sensory processing (Darriba et al., 2021; Korka et al., 2019), that modulate differently distinct stages of sensory processing.

In addition to the predictability effects on auditory responses, our design allowed us to assess the effects of predictability on visual responses for stimuli triggered by an action. Note that for the self-generated events in our study, participants had to press a button that would immediately (i.e., with no delay) trigger the colour-filling of a cue (the cue to be colour-filled could have been either fully predicted or not because of the fixed or random button-cue mapping, respectively). Interestingly, and in stark contrast with the findings on the auditory responses, visual responses were enhanced when the visual cue about to be colour-filled could be fully predicted by the action. This discrepancy between the effects of action on auditory vs. visual responses has been already discussed in previous work (Shimaoka et al., 2018; Parker et al., 2020), and points to a critical dependence of the self-generation effects on sensory modality that nevertheless remains out of the scope of the present study.

4.1.2. Effects of actions and predictability on subcortical neuromodulation

The second important finding related to the effects of actions sensory processing is that subcortical neuromodulatory processes (i.e., reflected in pupil diameter; Aston-Jones & Cohen, 2005, Murphy et al., 2014; Joshi et al., 2016) were robustly triggered by actions (cf. Lubinus et al., 2021; Vinck et al., 2015) and started even before action initiation (cf. McGinley et al., 2015; Aston-Jones & Cohen, 2005). More importantly, we show that pupil diameter was interactively modulated by self-generation and predictability, which was mainly driven by a remarkable difference between actions in predictable and unpredictable sessions. While it is possible that the identity prediction related to the cue to be colour-filled or the expectation of hearing a sound (which was only the case in the fully contingent predictable session) might have affected pupil responses, an intriguing possibility is that pupil responses might have been affected by differences between predictable and unpredictable contexts in action preparation and execution that are known to be affected by the action-stimulus contingency of the paradigm (i.e., movement parameters are modulated as a function of whether the action systematically triggers a stimulus or not; Neszmélyi & Horváth, 2017; Neszmélyi & Horváth, 2018).

4.1.3. Links between sensory attenuation and subcortical neuromodulation

Moreover, we also hypothesized that these subcortical neuromodulatory processes – reflected in pupil diameter – might be related to the stimulus-unspecific effects of actions on the auditory evoked responses. Indeed, pupil dilation correlated with the Tb attenuation effects, such that the larger the pupil diameter for self-produced sounds, the larger the suppression of the Tb component. However, this link was only found with fully contingent relationships and short delays between the action and sound (i.e., predictable session with 150-ms delay) where the attenuation effects were strongest. We also observed a non-significant trend between pupil dilation and N1 attenuation effects with fully contingent action-sound relationships, but longer action delays (in line with our previous work; Paraskevoudi & SanMiguel, 2021). Along with the differences in pupil response as a function of predictability, these two findings might suggest that action-induced subcortical processes might relate to the attenuation effects, especially when they are most pronounced (i.e., fully contingent paradigms). In sum, to the best of our knowledge, the present study is the first one to show a link between neuromodulatory processes during actions and sensory attenuation, thereby providing evidence for an alternative stimulus-unspecific mechanism that could at least partly underlie action-induced sensory attenuation, namely subcortical neuromodulatory processes during movement that are possibly mediated by the Locus Coeruleus Norepinephrine system.

4.2. Effects of actions on memory encoding

The second aim of the present study was to examine whether memory encoding of identifiable sounds is affected by semi self-paced actions (and its possible interactions with predictability and/or action-sound delay), and whether any possible effects on memory correlate with sensory suppression and/or subcortical neuromodulation during encoding. We obtained a significant enhancement for sounds that were encoded as self-generated, irrespective of predictability (and delay), providing support to the production effect studies highlighting the beneficial impact of production on memory recall of – typically fully predictable – stimuli (MacDonald & MacLeod, 1998; Brown & Palmer, 2012). This behavioural enhancement for self-generated stimulation seems at odds with predictive coding accounts of memory, but also with findings showing the opposite effect (i.e., behavioural enhancement for the more surprising externally-generated stimuli; Paraskevoudi & SanMiguel, 2021). Differences in the auditory stimuli employed in the two studies might have driven this discrepancy in memory performance (e.g., sound identifiability and duration). However, we believe that the critical factors driving the effect were related to the pure effects of actions on memory when they are not conflated with predictability, and the increased temporal control afforded by the actions in the present study. That is, here, participants knew that after the button press a sound would be generated after either 150 or 750 ms in 100% of the presses for the predictable session and 50% of the presses for the unpredictable session, which might have provided higher temporal control over self-generated stimulation than over the externally-generated inputs, where participants simply waited for the circle to be filled on its own.

Contrary to our previous work that provided only constrained temporal control (since participants were cued to press a specific button at a specific time point, namely, when a horizontal line crossed one of the vertical ones; Paraskevoudi & SanMiguel, 2021), the present design allowed for semi self-paced actions and for choosing between four different buttons. Evidence has already shown that learning and memory might be driven by intentionality (Herwig et al., 2007; Waszak et al., 2005), but also by choice-related processes (i.e., freely choosing which action to perform and when) even when participants' choices do not provide any control over the content of the memoranda (Murty et al., 2015). Another plausible explanation is that the memory weakening reported in our previous work might have been merely driven by differences in temporal predictability (Paraskevoudi & SanMiguel, 2021). However, in the present study, self- and externally-generated sounds were equally predictable (or unpredictable), thereby allowing us to isolate the effect of actions on memory processing without conflating it with temporal predictability. Therefore, we argue that when self-generation affords better predictions, memory will be weakened as proposed by predictive coding accounts (Henson & Gagnepain, 2010), which was supported by the significant correlation between N1 attenuation and memory performance when temporal predictability is not identical between self- and externally-generated inputs (Paraskevoudi & SanMiguel, 2021). In contrast, the present study did not show significant correlations between sensory attenuation and memory enhancement for self-generated stimulation, suggesting that conflating predictability with self-generation might result in effects of opposite direction compared to the memory effects driven by the presence of a self-chosen and semi self-paced action.

Additionally, and in line with previous work in rodents (McGinley et al., 2015) and humans (Paraskevoudi & SanMiguel, 2021), we show that the larger the pupil diameter for self-generated events, the worse the memory performance for these sounds at retrieval, but only in cases with fully contingent action-sound relationships and short action-sound delays (predictable session and 150 ms delay). Although our correlation analyses could only test for linear relationships between memory performance and pupil diameter, this finding could possibly fit with the idea of an inverted U-shaped dependence of behaviour on arousal (Yerkes & Dodson, 1908) – indexed by pupil diameter (McGinley et al., 2015) – suggesting that performance drops during high-arousal states such as movement, while it is optimal at intermediate arousal levels (i.e., states of quiet wakefulness).

5. Conclusion

In sum, the overarching aim of the present work was to control for all aspects of predictability (*what, when, whether*) in order to assess whether the self-generation effects on sensory processing and memory encoding of sounds are due to stimulus-specific predictions, due to the unspecific effects of actions, or a mixture of both. Sensory attenuation for predictable self-generated sounds presented shortly after the action prevailed even when providing identical predictability for the externally-generated ones, albeit with no clear evidence of attenuation in primary auditory areas (i.e., no attenuation on the mastoids), thereby providing only partial support to the specificity of the effects. Meanwhile, N1 attenuation also occurred for both predictable and unpredictable sounds presented farther away from the actions, pointing to some sort of diffused and generalized attenuation mechanism that acts independent of the action-sound contingency. Simultaneously with these effects, subcortical neuromodulatory processes were also triggered by actions that were reflected in increases in pupil diameter during actions. Critically, pupil dilation correlated with the attenuation effects when they were more strongly elicited, providing evidence that unspecific neuromodulatory processes may partly underly the action-induced sensory suppression. Finally, we provide strong evidence that when providing identical predictability for the externally-generated sounds, the presence of self-chosen and semi self-paced actions boosts memory encoding for the self-generated ones, pointing to effects of opposite direction when self-generation is conflated with predictability compared to when predictability is orthogonally controlled. Taken together, the present work shows that actions trigger a cascade of stimulus-specific and unspecific processes – presumably driven by subcortical neuromodulatory areas – that collaboratively shape auditory processing, and that predictability confounds in self-generation paradigms might be critical in determining the direction of the effects of actions on behaviour.

Study III

Supplementary Material

Table S1

Mean N1 and P2 amplitudes at Cz for the encoding phase, separately for each delay, source, and predictability condition. Standard deviation is provided in parentheses.

Factors		150 ms		750 ms	
Source	**Predictability**	**Mean N1 (SD)**	**Mean P2 (SD)**	**Mean N1 (SD)**	**Mean P2 (SD)**
A	predictable	-4.64 (2.73)	3.73 (2.33)	-4.30 (2.20)	2.69 (1.70)
A	unpredictable	-3.07 (2.01)	3.26 (1.90)	-4.19 (2.35)	3.45 (1.94)
MA	predictable	-3.15 (2.54)	0.87 (2.22)	-3.14 (2.12)	2.27 (1.13)
MA	unpredictable	-3.89 (1.62)	2.49 (1.69)	-3.48 (2.22)	3.70 (1.72)

Table S2

Mean Na and Tb amplitudes at T7 and T8 for the encoding phase, separately for each delay, source, and predictability condition. Standard deviation is provided in parentheses.

Factors			150 ms		750 ms	
Source	**Predictability**	**Laterality**	**Mean Na (SD)**	**Mean Tb (SD)**	**Mean Na (SD)**	**Mean Tb (SD)**
A	predictable	T7	-0.90 (1.31)	-3.09 (1.58)	-1.58 (1.20)	-2.34 (1.14)
A	predictable	T8	-0.40 (1.64)	-2.72 (1.92)	-0.89 (1.28)	-2.44 (1.30)
A	unpredictable	T7	-0.93 (1.02)	-2.00 (1.51)	-1.41 (1.10)	-2.73 (1.40)
A	unpredictable	T8	-0.57 (1.46)	-1.67 (1.28)	-0.96 (1.08)	-3.22 (1.58)
MA	predictable	T7	-1.18 (1.63)	-2.35 (1.74)	-1.47 (1.31)	-2.16 (0.80)
MA	predictable	T8	-0.90 (1.71)	-2.12 (2.04)	-0.94 (1.25)	-2.39 (1.15)
MA	unpredictable	T7	-1.36 (1.10)	-2.70 (1.53)	-0.92 (1.19)	-2.87 (1.47)
MA	unpredictable	T8	-0.60 (1.71)	-2.63 (1.60)	-0.56 (1.15)	-3.12 (1.54)

Table S3

Mean N1$_{mast}$ amplitudes at the mastoids (M1, M2) for the encoding phase, separately for each delay, source, and predictability condition. Standard deviation is provided in parentheses.

Factors			150 ms	750 ms
Source	**Predictability**	**Laterality**	**Mean N1$_{mast}$ (SD)**	**Mean N1$_{mast}$ (SD)**
A	predictable	M1	0.13 (1.11)	0.13 (0.78)
A	predictable	M2	0.43 (1.25)	0.54 (0.87)
A	unpredictable	M1	-0.01 (1.80)	0.15 (1.04)
A	unpredictable	M2	0.29 (1.74)	0.51 (0.95)
MA	predictable	M1	0.19 (1.21)	-0.05 (0.78)
MA	predictable	M2	0.28 (1.48)	0.36 (1.01)
MA	unpredictable	M1	-0.17 (1.42)	-0.05 (0.77)
MA	unpredictable	M2	0.15 (1.32)	0.47 (0.89)

Table S4

Mean N1 and P2 amplitudes at Cz for the retrieval phase, separately for each delay, source, and predictability condition. Standard deviation is provided in parentheses.

Factors		**150 ms**		**750 ms**	
Source	**Predictability**	**Mean N1 (SD)**	**Mean P2 (SD)**	**Mean N1 (SD)**	**Mean P2 (SD)**
Encoded as A	predictable	-3.78 (2.44)	2.31 (2.83)	-3.02 (1.96)	2.75 (2.71)
Encoded as A	unpredictable	-3.19 (2.07)	2.69 (2.50)	-3.65 (1.75)	2.18 (2.66)
Encoded as MA	predictable	-3.76 (2.25)	1.72 (2.15)	-4.05 (2.24)	2.41 (2.41)
Encoded as MA	unpredictable	-3.25 (2.36)	2.67 (2.62)	-3.61 (1.80)	2.32 (2.62)

Table S5

Mean Na and Tb amplitudes at T7 and T8 for the retrieval phase, separately for each delay, source, and predictability condition. Standard deviation is provided in parentheses.

Factors			**150 ms**		**750 ms**	
Source	**Predictability**	**Laterality**	**Mean Na (SD)**	**Mean Tb (SD)**	**Mean Na (SD)**	**Mean Tb (SD)**
Encoded as A	predictable	T7	-0.90 (1.53)	-2.27 (1.58)	-1.01 (1.10)	-1.63 (1.35)
Encoded as A	predictable	T8	-0.27 (1.48)	-2.37 (1.60)	-0.13 (1.31)	-1.73 (1.30)
Encoded as A	unpredictable	T7	-0.70 (1.16)	-2.31 (1.44)	-0.68 (1.09)	-2.22 (1.28)
Encoded as A	unpredictable	T8	-0.53 (1.33)	-2.41 (1.37)	-0.40 (1.52)	-2.50 (1.45)
Encoded as MA	predictable	T7	-1.13 (1.11)	-2.43 (1.37)	-1.31 (1.40)	-2.42 (1.20)
Encoded as MA	predictable	T8	-0.47 (1.12)	-2.22 (1.34)	-0.30 (1.66)	-1.89 (1.62)
Encoded as MA	unpredictable	T7	-0.96 (1.03)	-2.13 (1.27)	-0.89 (0.90)	-2.34 (1.24)
Encoded as MA	unpredictable	T8	-0.10 (1.44)	-2.10 (1.56)	-0.48 (1.59)	-2.10 (1.45)

Table S6

Mean $N1_{mast}$ amplitudes at the mastoids (M1, M2) for the retrieval phase, separately for each delay, source, and predictability condition. Standard deviation is provided in parentheses.

Factors			**150 ms**	**750 ms**
Source	**Predictability**	**Laterality**	**Mean $N1_{mast}$ (SD)**	**Mean $N1_{mast}$ (SD)**
Encoded as A	predictable	M1	0.14 (0.90)	0.12 (0.77)
Encoded as A	predictable	M2	0.48 (0.68)	0.42 (0.94)
Encoded as A	unpredictable	M1	-0.02 (1.00)	0.21 (0.96)
Encoded as A	unpredictable	M2	0.49 (0.83)	0.70 (0.97)
Encoded as MA	predictable	M1	0.30 (1.21)	-0.09 (1.18)
Encoded as MA	predictable	M2	0.60 (0.73)	0.47 (0.84)
Encoded as MA	unpredictable	M1	0.22 (1.25)	-0.25 (0.99)
Encoded as MA	unpredictable	M2	0.75 (1.04)	0.57 (0.62)

CHAPTER 5: General discussion

Over the last years, there has been an increasing interest in understanding how the brain uses predictions to optimally interact with the surroundings. Predictions can be formed based on multisensory associations (i.e., audiovisual pairings) or long-term learning (i.e., music or language; for reviews see Press et al., 2020; Bendixen et al., 2012; Schröger et al., 2015; Korka et al., 2021); however, the main focus of the present thesis was to understand how predictions are formed by our own actions. Although substantial work has been done in this domain, several issues remain unknown, often leading to heated debates as for the mechanisms underlying the effects of actions on behaviour and sensory processing (for a review see Schröger et al., 2015; Press et al., 2020; Reznik & Mukamel, 2019). At the core of this debate stand questions related to the direction of the action effects on behaviour (i.e., perceptual suppression or enhancement) and sensory processing (i.e., cancellation or sharpening of sensory responses), the locus and nature of the effects (i.e., stimulus-specific motor-predictions or unspecific mechanisms modulating responses to stimuli falling close in time with an action in an indiscriminate manner), but also the influence of other factors that have been usually confounded with self-generation in previous studies (e.g., stimulus intensity, predictability, and attention).

Specifically, the present thesis constitutes a modest attempt to elucidate the modulatory effects of actions on low-level perceptual processing, high-level memory processes, and neurophysiological responses. Study I focused on the interactive effects of actions and stimulus' intensity on sound detection and loudness discrimination, while carefully assessing which perceptual measure is affected by self-generation and its interaction with sound intensity. Study II sought out to address the effects of motor acts on sensory processing and memory encoding of concurrently presented, but unpredictable, sounds, while exploring for the first time the involvement of neuromodulatory systems in the action-induced modulations of auditory responses. Largely inspired by Study II, Study III assessed whether self-generation and predictability – two factors that have been conflated in previous work (for a review see Hughes et al., 2013a) – interactively shape auditory responses, subcortical neuromodulation, and memory encoding of sounds. Overall, the findings obtained in the present thesis show that actions modulate perception, memory, sensory processing, and subcortical neuromodulation, and interact with other factors in driving these effects (e.g., stimulus intensity, predictability, and action-sound delay), thereby painting a complex picture of how movement shapes the contents of our perceptual and sensory representations. The following sections provide a discussion on the effects of actions separately on perception and memory, sensory responses, and subcortical neuromodulation across the three studies.

5.1. Modulatory effects of self-generation on perception and memory

To date, there is mounting evidence showing that actions modulate perceptual processing in several ways, with the most intuitive observation of such modulations being the fact that although we can be tickled by others, we can never tickle ourselves (e.g., Blakemore, Wolpert et al., 1998; Claxton, 1975; Kilteni et al., 2020). Further work supports that self-production of stimuli might also affect memory encoding, but the direction of the effects remains inconclusive (e.g., Mama & Icht, 2016; Henson & Gagnepain, 2010). The present thesis tapped onto these two research lines related to the effects of actions on low-level perceptual processing (Study I) and memory encoding (Studies II and III). Before moving on to discussing the behavioural findings of the present thesis, it should be noted that to date it remains unclear whether the same mechanisms are responsible for the processing of arbitrary action-stimulus associations (e.g., tones caused by actions) and well-established action-stimulus outcomes that have an innate or overlearned coupling with their associated actions (e.g., our own voice). For example, motor-based forward models might drive the perceptual processing for body-related signals, such as speech or tickling (for a discussion see Dogge et al., 2019), in line with work pointing to prediction-specificity for body-relevant sensory reafference which is thought to be driven by efference copy signals (e.g., Kilteni et al., 2020). However, evidence on specific motor-prediction effects on the perception of arbitrary action-stimulus relationships (e.g., studies using pure tones as the reafferent input; Reznik et al., 2015) is less clear, and does not always point to attenuated perception (e.g., Reznik et al., 2015; Reznik et al., 2014). The stimuli used in the present thesis were an instance of the latter, namely arbitrary tones resulting from participants' button presses.

Related to low-level perceptual processing, the present thesis shows that actions interact with stimulus intensity on perceptual processing, but the effects are limited to perceptual bias measures rather than sensitivity. Specifically, Study I showed that while detection and discrimination sensitivity did not differ as a function of self-generation and intensity, perceived loudness for fully contingent and supra-threshold self-generated sounds (i.e., a cued action always triggered a sound in the discrimination task) was attenuated, in line with previous work showing perceptual attenuation for supra-threshold sensory reafference, that is, auditory stimulation caused *by* our voluntary movement (Sato, 2008; Weiss et al., 2011a, 2011b). Critically, according to dominant cancellation models, this perceptual attenuation is considered as the behavioural manifestation of the attenuated processing in sensory-specific areas (Sato, 2008; Kilteni et al., 2020), implying that the perceived intensity of reafferent input should be always dampened, irrespectively of sound intensity. However, we found the opposite effect at near-threshold intensities, that is, perceptual enhancement for near-threshold sensory reafference, suggesting that other factors might contribute to determining the direction of the effects (e.g., intensity), and raising the possibility of the action effects on loudness perception not being solely driven by cancellation of sensory reafference as suggested by studies employing self-generated touch (Blakemore, Wolpert et al., 1998; Kilteni et al., 2020). Importantly, although the interaction between self-generation and sound intensity has been already described by Reznik and colleagues (2015), here, we extend this work by showing that this interaction is limited to perceptual bias, rather than sensitivity measures, suggesting that sound perception is equally precise in the absence and in the presence of action. This finding is critical considering recent evidence in the somatosensory domain showing that the perceptual bias and sensitivity measures might point to different perceptual phenomena, that is predictive perceptual attenuation for reafferent stimuli caused *by* our movement and gating effects for both reafferent and exafferent information presented *during* movement (but not predicted by it; Kilteni & Ehrsson, 2022), respectively.

Related to high-level memory processes, the present work shows that actions modulate memory performance, but the direction of the effects might depend on factors related to the type of action and the temporal control afforded by it, the confounding factor of predictability, and the type of the sounds to be encoded. Specifically, Study II showed that cued actions triggering non-identifiable, unpredictable sounds weaken memory encoding which also relates to the magnitude of the attenuation effects for these sounds. In contrast, Study III showed the opposite effect, that is memory enhancement for both predictable and unpredictable self-generated identifiable sounds. Importantly, contrary to Study II, this memory enhancement did not relate to the neurophysiological effects of actions on sensory responses. The discrepancy in the findings obtained in Study II and Study III might be driven by the factors mentioned above. First, contrary to Study II, Study III allowed for a variety of action choices (i.e., choosing which action to perform) that is known to yield beneficial effects on memory (Murty et al., 2015). Second, temporal control over the upcoming self-generated stimulus was higher in both predictability sessions of Study III than in Study II: Actions were semi self-paced and allowed participants to know that after the button press a sound would be generated 100% or 50% of the time after either 150 or 750 ms for the predictable and unpredictable session, respectively. This might have provided slightly higher temporal control over self-generated stimulation than over the externally-generated inputs, where participants simply waited for the circle to be filled on its own. Therefore, although temporal predictability and probability of occurrence was equal for both types of sounds, participants only had temporal control over the self-generated stimulation. Third, Study III orthogonally controlled self-generation and predictability, providing equal identity and temporal predictability, but also equal probability of occurrence for both self- and externally-generated sounds. In contrast, temporal predictability was not controlled in Study II (i.e., better temporal predictions for the self-generated sounds). Finally, auditory stimulation differed between the two studies (i.e., non-identifiable noise stimuli in Study II and identifiable human, animal, environmental, and musical sounds in Study III). There is already evidence that different types of sounds (e.g., noise, pure tones, speech, and music) exhibit different patterns of activity (for a recent meta-analysis see Samson et al., 2011), and it is also known that noise stimuli have inherently less structure and therefore afford less predictability than identifiable sounds (Agus et al., 2010). Thus, the employment of non-identifiable stimuli in Study II – along with the confounding factor of temporal predictability in this study – might have triggered saliency-detection mechanisms (Agus et al., 2010; Itti et al., 1998) that boosted the memory of the more surprising externally-generated sounds, as proposed by predictive coding accounts of memory (Henson & Gagnepain, 2010). Lastly, another possibility is that the division of attention in Study II (participants had to be precise in pressing the button when the horizontal line reached one of the vertical ones) might have resulted in memory weakening for the self-generated sounds in Study II (cf. Craik et al., 1996).

In sum, the findings obtained in the three studies of the present thesis raise the question of whether the action effects on perceptual and memory processes can be explained by one unitary framework. Although we cannot exclude the possibility that actions might affect differently low-level perception (Study I) and high-level cognition as in the case of memory (Study II and III), evidence linking the magnitude of sensory responses at encoding with memory performance at retrieval (e.g., Gagnepain et al., 2011) suggests that the differential perceptual processing for self-generated stimulation might define how they will be encoded in memory. However, our findings show that the direction of the effects of actions on both perception and memory is not only defined by the mere presence of an action; rather the effects seem to be highly dependent on factors related to stimulus intensity, action type (cued action vs. several action alternatives), predictability confounds, and the acoustic characteristics of the stimuli (i.e., identifiable vs. non-identifiable sounds), as well.

To date, however, none of the extant models has attempted to provide a complete explanatory framework for the effects of actions on both perception and memory by considering all the factors manipulated in the present thesis (for a schematic representation of the findings and assumptions of previous models see Figure 2a). In the perception domain, cancellation and preactivation models would expect lower intensity percepts (driven by efference copy mechanisms, e.g., Kilteni et al., 2020) and lower detection sensitivity for self-generated sounds (Waszak et al., 2012; Roussel et al., 2013), respectively, whereas sharpening models would expect a representational sharpening for self-generated stimulation which would be limited in sensitivity measures (Yon et al., 2018). Importantly, these accounts cannot explain the interactive effects of self-generation and stimulus intensity. In contrast, the model proposed by Reznik & Mukamel (2019), as well as the opposing process theory by Press and colleagues (2020), highlight that the direction of action effects on perception should depend on stimulus intensity, but they disagree on which aspect of perception is affected (sensitivity or bias, respectively). Additionally, it should be noted that these two accounts have been typically applied to explain findings in studies that have conflated self-generation with predictability (Hughes et al., 2013a) or that differed in the type of action performed (i.e., general active state such as locomotion or button presses; Reznik & Mukamel, 2019) and the comparison of interest to be examined (i.e., predicted vs. mispredicted or predictable vs. unpredictable; Press et al., 2020). Importantly, none of the models presented so far has attempted to link perceptual processing for self-generated stimuli with their encoding in memory.

Conversely, in the memory domain, production effect studies have reported behavioural benefits of actions on memory processing in paradigms with fully predictable stimulation triggered by complex actions (e.g., spoken words or played melodies; MacDonald & MacLeod, 1998; MacLeod et al., 2010; Ekstrand et al., 1966; Hopkins & Edwards, 1972; Conway & Gathercole, 1987; Gathercole & Conway, 1988; Mathias et al., 2015; Brown & Palmer, 2012), but this line of research has been limited to behavioural measures and has not attempted to link the production effect with the immediate sensory processing of self-produced stimulation. In contrast, the predictive coding account of memory suggests that learning and memory are driven by prediction errors, and to the best of our knowledge it is the only framework attempting to link neurophysiological responses for perception with memory performance. However, it should be noted that this account has been typically applied to explain studies manipulating predictions outside of the action domain (Greve et al., 2017; Henson & Gagnepain, 2010; Pine et al., 2018; Gagnepain et al., 2011). In sum, under these two accounts of memory, one would assume memory enhancements or memory decrements for self- compared to externally-generated sounds, respectively, but here we provide evidence for both, suggesting that other factors might be also critical in driving the direction of the action effects on memory processing.

Here, although we cannot exclude the possibility of other factors driving these effects, we propose that the effects of actions on perceptual processing and memory encoding depend on sound intensity, predictability confounds, and the level of temporal control afforded by the action. Specifically, we argue that surprise-driven generalized predictive mechanisms will operate when predictability is confounded with self-generation, and the direction of the effects on perception and memory will depend on sound intensity, since only supra-threshold stimulation can be surprising (Figure 2b, left). However, when predictability is controlled and actions provide temporal control over the self-generated stimulus, memory will be boosted, irrespective of whether the action is fully predictive of the sound or not (Figure 2b, right). We develop this idea in the remaining paragraphs of this section.

We argue that when predictability is conflated with self-generation (as in Study I and Study II where self-generated stimuli afforded better temporal predictability), the effects are driven by surprise, in line with generalized prediction models that are not necessarily dependent on efference copy signals (Friston, 2001, 2005, 2009). According to Bayesian inference models (Friston, 2001, 2005, 2009), surprise is reflected in the distance between two distributions and their associated precision (mathematically expressed as the Kullback-Leibler divergence; Kullback, 1997; Itti & Baldi, 2009): The prior distribution that reflects the probabilistic prediction about the upcoming event, and that once combined with the observed sensory input (i.e., likelihood) computes the posterior probability distribution (Sterzer et al., 2018). Evidently, surprise can only be triggered by clearly perceivable stimulation, that is with loud, supra-threshold inputs (i.e., high signal-to-noise ratio of the sensory input; cf. Press et al., 2020). Therefore, it would be adaptive and computationally plausible to assume that with supra-threshold stimulation, surprise-driven mechanisms would prioritize the processing of the more surprising externally-generated inputs, by attenuating perception to highly predictable and loud sensory reafference. In contrast, under uncertainty (i.e., low-intensity inputs as the near-threshold condition in the discrimination task of Study I), the lower signal-to-noise ratio of the sensory inputs renders the unexpected less surprising, and thus uninformative, shifting the perception towards the expected (Press et al., 2020). This would effectively result in the opposite effect, that is enhanced perceived intensity for the self-generated compared to the externally-generated tones. This explanation fits with the findings obtained in the present thesis showing perceptual attenuation and memory weakening for – temporally – more predictable supra-threshold self-generated sounds (i.e., self-generated sounds in the discrimination task of Study I and in Study II were more predictable in time that the externally-generated ones), pointing to generalized predictive processes, that are not necessarily driven by motor-related efference copy signals (Friston, 2005; but also see Kilteni et al., 2020 for a discussion in the domain of somatosensory attenuation).

However, when the effect of predictability is isolated by providing identical predictability for self- and externally-generated sounds (as in Study III), surprise-driven mechanisms might no longer operate, and other factors, possibly related to the genuine effect of performing an action and the temporal control it affords over the self-initiated stimulus, might dominate in shaping perception and memory. This possibility could account for the findings obtained in Study III that – to the best of our knowledge – was the first to show a highly significant memory enhancement for self-generated sounds irrespective of whether they could be predicted by the action or not, suggesting that predictability cannot be driving this enhancement. It is possible that when predictability is controlled and action is self-chosen and semi self-paced (as in Study III), intentionality and choice-related processes (i.e., freely choosing which action to perform and when) might drive learning and memory (Herwig et al., 2007; Waszak et al., 2005; Murty et al., 2015), possibly by increasing the feeling of control over action outcomes (Wenke et al., 2010) that might allow for better processing of the self-triggered stimulus when an action can be selected among several alternatives compared to when action is limited to one alternative (Barlas et al., 2017).

Collectively, the findings as for the action effects on memory suggest that when self-generation affords better temporal predictions (i.e., as in Study II), memory will be weakened as proposed by predictive coding accounts (Henson & Gagnepain, 2010), which is supported by the significant correlation between N1 attenuation and memory performance (Study II). In contrast, memory performance is enhanced for self-generated and identifiable sounds and does not correlate with the attenuation effects when controlling for predictability (Study III), suggesting that conflating predictability with self-generation might result in effects of opposite direction compared to the memory effects driven by the mere presence of an action. Therefore, self-generation and predictability might act in opposite directions, but in the presence of both, predictability might win over the action effects, even more so when the action is not self-chosen and does not provide temporal control over the upcoming stimulation (as in Study II).

In sum, the present thesis proposes that the direction of the effects on perception and memory might depend on the involvement of predictability in the self-generation effects. Specifically, when actions afford better temporal predictability, perception will be attenuated (but only for supra-threshold stimulation) and memory will be weakened. In contrast, when predictability is controlled and actions provide choice alternatives and temporal control over the self-produced stimulation, memory will be boosted.

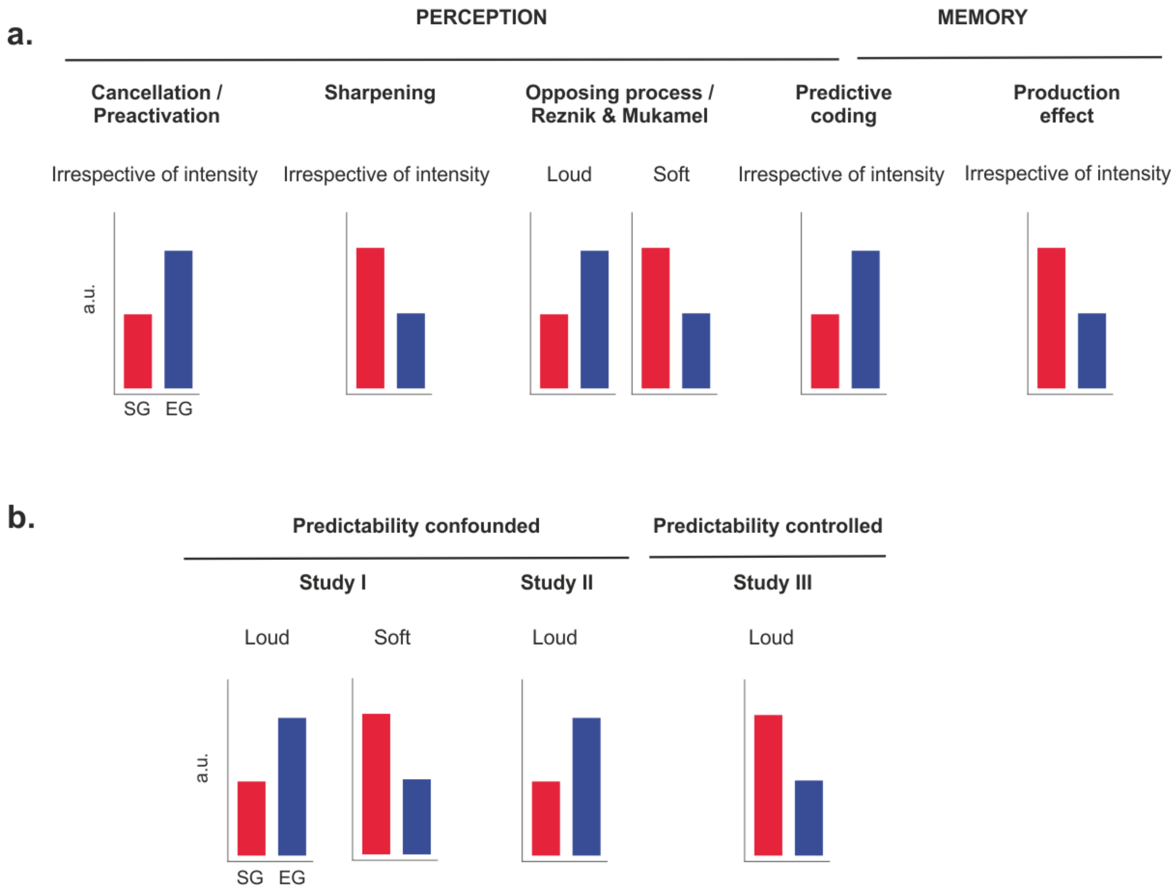


Figure 2. Schematic representation of the assumptions of extant models on how actions affect perception and memory **(a)** and the findings obtained in the present thesis **(b)**. Each barplot represents in arbitrary units (a.u.) the perceived intensity and memory performance in perceptual and memory paradigms, respectively, for self- (SG) and externally-generated sounds (EG). **a)** Assumptions made by previous models on how actions affect perception and memory. Cancellation and preactivation models assume perceptual attenuation, while sharpening models have proposed that perception is enhanced for self-generated and predictable stimulation. These models have not accounted for possible interactions with the stimulus intensity. The opposing process theory by Press et al. (2020) and the model proposed by Reznik and Mukamel (2019) predict an interaction between self-generation and intensity. Predictive coding accounts have been applied both in the perceptual and memory domain, assuming that perception and memory is reduced for predictable stimulation as a result of lower prediction error responses. Finally, production effect studies have reported enhanced memory for self-produced stimulation. **b)** Summary of the findings obtained in the three studies of the present thesis. When predictability was confounded with self-generation, perceived intensity was reduced for loud self-generated sounds, but enhanced for soft self-generated sounds (Study I), and memory performance was reduced (Study II). In contrast, when predictability is controlled, memory is enhanced for self-generated sounds (Study III).

5.2. Modulatory effects of actions on auditory responses

Along with the action effects on perception and memory, one of the core objectives of the present thesis was to elucidate the neurophysiological mechanisms underlying the action effects on sensory processing. This question seems highly relevant given the discrepancy between the assumptions of dominant cancellation models and the findings provided by one part of the sensory attenuation research showing that sensory attenuation might not be purely driven by stimulus-specific motor predictions (Horváth et al., 2012; SanMiguel et al., 2013). On the one hand, the cancellation account assumes that responses to sensory reafference, that is stimuli caused and predicted *by* our actions, will be suppressed due to a forward model that sends an efference copy of the motor command to the corresponding sensory cortex to cancel out predictable, albeit unavoidable, self-generated inputs (Sperry, 1950; von Holst & Mittelstaedt, 1950; Blakemore, Wolpert et al., 1998; Miall & Wolpert, 1996; Wolpert et al., 1995). One important implication of this model is that the suppression effects should be specific to the stimulus (i.e., only predictable self-generated stimuli should be attenuated) and mediated by sensory-specific areas (i.e., the effects should reflect suppressed activity in the sensory-specific areas). On the other hand, there is work showing that attenuation occurs even in the absence of predictive action-stimulus relationships, and there is no clear evidence that the attenuation effects are only stimulus-specific (e.g., SanMiguel et al., 2013; Horváth et al., 2012). Here, we addressed this apparent discrepancy by examining the specificity of the effects, disentangling the effects of predictability and self-generation on sensory responses (Hughes et al., 2013a), and assessing possible interactions between predictability and self-generation that have been largely unexplored (but see Blakemore, Rees et al., 1998; Baess et al., 2008; Klaffehn et al., 2019).

Study II employed self-generated, albeit unpredictable, sounds that coincided with actions, but they were more predictable in time than the externally-generated ones (due to the presence of a cue prompting participants to press a button that would generate a sound half of the time). Despite the absence of a predictive action-sound relationship, we obtained the typical attenuation effects on N1, P2, and Tb (cf. Horváth et al., 2012), but also enhanced P3 for action-sound coincidences. These effects might partly point to unspecific mechanisms, since in this design the identity and the probability of occurrence for the self-generated sounds could not be predicted. Nonetheless, sensory suppression could have been partly driven by the increased temporal predictability for the self-generated sounds (Schafer & Marcus, 1973; Lubinus et al., 2021; Näätänen & Picton, 1987). Indeed, there is some evidence that the magnitude of sensory attenuation effects is reduced when accounting for the effects of temporal predictability (Schafer & Marcus, 1973; Sowman et al., 2012; Weiskrantz et al., 1971), highlighting the need to control for the effects of predictability in self-generation studies.

Study III, therefore, controlled for the confounding factor of predictability by employing a predictable and an unpredictable session where both self- and externally-generated sounds were equally predictable (or unpredictable) in identity, timing, and probability of occurrence and were presented shortly after or farther away from a semi self-paced action or cue, respectively. With short delays, predictability mattered, since attenuation was only obtained in the predictable session in line with work showing stronger attenuation with more precise match between action and its consequences (for a review see Schröger et al., 2015). With longer delays, self-generation and predictability no longer interacted and affected N1, Tb, and P2 in differential ways: N1-attenuation was driven by self-generation only, while Tb- and P2-attenuation were driven by predictability. The question that arises, therefore, is: Where do these findings lead us as for the specificity of the attenuation effects?

Support to the specificity of the effect could be provided by the significant interaction between self-generation and predictability with short action-sound delays in Study III: Attenuation of N1, P2, and Tb, was only observed for predictable self-generated sounds, in line with previous work showing that the suppression effects are specific to the predicted stimulus (McGuire et al., 1996; Hirano et al., 1997; Houde et al., 2002; Hashimoto & Sakai, 2003; Heinks-Maldonado et al., 2005; Fu et al., 2006; Ott & Jäncke 2013), but also with data suggesting that attenuation prevails even when controlling for differences related to temporal predictions and temporal control (Klaffehn et al., 2019; but see also Kaiser & Schütz-Bosbach, 2018). This points to a critical difference between motor-based and sensory-based predictions, in that attenuation occurs for motor-auditory predictions (i.e., self-generated predictable sounds) but not for the visual-auditory ones (i.e., externally-generated predictable sounds that were cued by a visual stimulus). Nevertheless, contrary to previous neuroimaging and magnetoencephalography evidence that the action effects reflect modulations in auditory areas (Blakemore, Rees et al., 1998; Christoffels et al., 2007; Numminen & Curio, 1999; Numminen et al., 1999; Curio et al., 2000; Gunji et al., 2001; Ventura et al., 2009; Niziolek et al., 2013; Tian & Poeppel, 2013; Martikainen et al., 2004; Aliu et al., 2009), we did not find clear evidence that the N1 attenuation reflects modulations of the sensory specific components. This would be reflected in attenuation of the N1 component on the mastoids (thought to partly originate from a supratemporal generator; Vaughan & Ritter, 1970) and the "T complex", that is the Na and Tb (generated by radial sources in the superior temporal gyrus and identified on anterior temporal electrodes; Tonnquist-Uhlen et al., 2003; Wolpaw & Penry, 1975). However, we did not observe attenuation of the specific N1 component at the mastoids, while attenuation for Tb was found for predictable self-generated sounds with short action-delays (Study III), but also for action-sound coincidences that nonetheless afforded better temporal predictability than the passive sounds (Study II), suggesting that increased temporal predictability might be sufficient to elicit suppression of the Tb component.

However, the present thesis also provides strong evidence for the involvement of unspecific processes in the self-generation effects. Attenuation of N1 was observed for sounds that were presented close in time *with* the action but could not be predicted *by* it (Study II and with long action-sound delays in the unpredictable session of Study III), suggesting that N1-suppression is not solely driven by specific motor-predictions (Horváth et al., 2012; SanMiguel et al., 2013). The involvement of unspecific mechanisms in sensory attenuation during actions has been already raised by previous work showing that attenuation prevails for mere action-sound coincidences (Hazemann et al., 1975; Horváth et al., 2012; Makeig et al., 1996; Tapia et al., 1987) and for unrelated auditory inputs during speech (Houde et al., 2002; Kudo et al., 2004), and that it might reflect the modulation of the unspecific N1 component (SanMiguel et al., 2013). Meanwhile, P2 and Tb attenuation for self-produced sounds was observed in unpredictable contexts as well, but only when temporal predictability was not controlled (Study II), suggesting that these two components are sensitive to general predictions afforded by the actions but also to non-motor predictability (reduced amplitude for predictable vs. unpredictable sounds, irrespective of the presence of action, in Study III with longer delays), in line with a more generalized prediction account (Friston, 2005) that does not require motor-specific predictions to attenuate responses.

In sum, these findings suggest that stimulus-specific and unspecific mechanisms co-exist and collaboratively shape auditory responses during actions (cf. Horváth, 2015; Schröger et al., 2015; Flinker et al., 2010). First, sensory attenuation is strongest with better match between prediction and sensory feedback when the latter is presented shortly after the action, and even when controlling for predictability confounds (cf. Klaffehn et al., 2019; Harrison et al., 2021). Second, attenuation effects prevail for sounds that cannot be predicted by the action and these effects are not clearly driven by modulations in primary auditory areas, pointing to some sort of diffused and generalized attenuation mechanism that acts independently of the action-sound contingency and does not necessarily operate based on efferent copy signals. Most importantly, the involvement of such stimulus-unspecific processes in the well-established attenuation effects raises the need to identify the mechanisms driving this unspecific modulation of sensory processing around movement. Here, we hypothesized that subcortical neuromodulation – reflected in pupil diameter – might relate to the unspecific modulation of auditory responses to sounds triggered, but not necessarily predicted by actions.

5.3. Modulatory effects of actions on subcortical neuromodulation

At the core of the present thesis stands the hypothesis that actions might create a halo of neuromodulation around them that could possibly relate to the attenuation effects for self-generated information. This hypothesis was largely inspired by animal work, mainly studies in rodents, that has consistently shown that actions initiate a cascade of neuromodulatory processes (Vinck et al., 2015; Eggerman et al., 2014; McGinley et al., 2015), and that motor and neuromodulatory inputs overlap in auditory areas during movement (Nelson & Mooney, 2016; for a review see Schneider & Mooney, 2018). We, specifically, examined the possibility that the Locus Coeruleus Norepinephrine system (LC-NE) might be activated during movement, effectively creating a halo of neuromodulation that could be possibly related to the unspecific mechanisms that attenuate auditory responses around actions in an indiscriminate manner (i.e., even in the absence of a contingent action-stimulus relationship).

Indeed, in line with previous work, we provide evidence that subcortical neuromodulatory processes, possibly mediated by LC-NE (i.e., reflected in pupil diameter; Aston-Jones & Cohen, 2005, Murphy et al., 2014; Joshi et al., 2016), were robustly triggered by actions (also reported in rodents: Lee & Margolis, 2016; McGinley et al., 2015; Vinck et al., 2015, monkeys: Bornert & Bouret, 2021, and humans: Yebra et al., 2019; Lubinus et al., 2021). Specifically, pupil dilation started already ~200 and ~500 ms before action initiation in Study II and III (cf. Aston-Jones & Cohen, 2005; Reimer et al., 2016) and remained dilated for more than 1 s compared to the passive comparison events. Critically, the pattern of activation was remarkably consistent across the two delay conditions we employed in Study III, in line with evidence suggesting that phasic LC-NE activation is more tightly coupled with behaviour (e.g., motor response) than sensory stimulation (Aston-Jones & Cohen, 2005). The similarity of the pupil response across action-sound delays also suggests that the hypothesized halo of subcortical neuromodulation might last more than the longest delay used in the present thesis. Nevertheless, as we mention in more detail in the next section, the effects of this halo of neuromodulation on sensory responses seem to be strongest when the sound is presented close in time with the action, thereby providing some interim evidence that the effects might fade out with increasing temporal distance between motor act and sensory stimulation.

Importantly, the present thesis extended previous work by showing that movement and predictability interactively modulated pupil diameter, evident by a striking difference between actions generating highly predictable auditory stimulation compared to those that could not predict the identity, timing, and probability of occurrence of a sound. There is some evidence that action-induced LC-NE activity might depend on the type of action, differing between goal-directed actions and general active states representing more automatic behaviours such as grooming (Aston-Jones & Bloom, 1981). Additionally, there is also work from the self-generation domain showing that action preparation and execution are affected by the action-stimulus contingency of the paradigm, being modulated as a function of whether the action systematically triggers a stimulus or not (Neszmélyi & Horváth, 2017; Neszmélyi & Horváth, 2018). Collectively, in line with these findings, the interaction between predictability and movement we obtained here suggests that differences in movement parameters, such as action-stimulus contingency, contribute to the magnitude of the action-induced pupil dilation effects. However, Study III could not disentangle whether one or more aspects of predictability have driven this interaction. Therefore, more work is needed to replicate these interactive effects of actions and predictability on pupil diameter and identify the critical movement parameters that might drive differences in pupil responses during actions.

5.4. Relationships between the action effects on memory, auditory responses, and subcortical neuromodulation

Lastly, one of the main questions of the present work was to understand whether the effects of actions on perception and memory, sensory processing, and subcortical neuromodulation described in the previous sections are related. Although the intensity manipulations in Study I might have also affected arousal level and LC-NE activity, supported by data showing a dependence of LC-NE activity on stimulus intensity (Hayat et al., 2020; Grant et al., 1988; Aston-Jones & Bloom, 1981), subcortical neuromodulation was assessed only in Study II and Study III using pupillometry. Therefore, we will limit our interpretations in the correlations obtained in these two studies, where we have a clear index of neuromodulatory processes thanks to the pupillometry recordings. Study II showed that memory performance independently correlated with sensory attenuation (N1 and Tb) and pupil diameter for self-generated, albeit unpredictable sounds, that is, the larger the N1 and Tb suppression and the greater the pupil diameter, the greater the memory impairment for sounds coinciding with actions. Crucially, contrary to our initial hypothesis, sensory attenuation and pupil dilation did not correlate in Study II. However, Study III revealed a different set of relationships that were limited to fully predictable contexts and short action-sound delays, and that can possibly elucidate a series of questions raised by the present thesis. First, memory performance correlated only with pupil diameter (but not with sensory attenuation as in Study II), pointing to the same direction as the one obtained in Study II: The larger the pupil diameter, the worse the memory performance for self-produced sounds (but note that in Study III, self-generation enhanced, rather than weakened memory). Second, pupil diameter correlated with sensory attenuation (of the Tb component), that is with larger pupil diameter, the attenuation effects are more pronounced. Importantly, contrary to Study II, sensory attenuation did not correlate with memory performance. In what follows, we discuss the links obtained through correlation analyses in an attempt to understand under what circumstances these relationships arise.

First, as for the link between memory performance and pupil diameter, we obtained the same pattern in both studies, namely that increased pupil diameter relates to worse memory performance. However, this relationship was obtained only when the following were fulfilled: Action outcomes should be predictable to some extent (either temporally predictable only, as in Study II, or fully predictable, as in the predictable session of Study III) and should be followed shortly after the action (i.e., zero delay in Study II and 150 ms delay in Study III). In line with this negative correlation between pupil dilation and memory performance, previous work has already shown that later remembered items were accompanied by less dilation during their encoding (Kafkas & Montaldi, 2011). Although at a first glance these findings might suggest a critical role of arousal per se on memory performance, it should be noted that arousal alone is not sufficient to compromise memory. That is, although pupil dilation correlated with worse memory for self-generated sounds in both studies, performance for self-generated sounds was weakened only in Study II. In contrast, Study III showed a highly robust enhancement for sounds encoded as self-initiated. Although our linear correlation approach might not have captured the full nature of the dependence of perception and memory on arousal level, our findings could fit with the idea of an inverted U-shape dependence of behavioural performance on LC-NE activity (Yerkes & Dodson, 1908): High arousal states, such as movement, relate to worse behavioural performance compared to the intermediate arousal states that are thought to yield optimal performance in a range of behavioural tasks (Aston-Jones & Cohen, 2005; McGinley et al., 2015).

Second, memory performance correlated with sensory attenuation, but only when temporal predictability was increased for self- compared to externally-generated sounds (Study II). This link was not obtained when identical predictability was provided for self- and externally-generated sounds. The importance of this finding relies on the conclusions that can be drawn and that are highly relevant for the present thesis: On the one hand, it points to a generalized predictive mechanism – in line with Bayesian inference accounts (Friston, 2001, 2005, 2009) and predictive coding accounts of memory (Henson & Gagnepain, 2010) – that boosts learning and memory for the more surprising events (as in Study II). We speculate that this mechanism would operate for any predictable-unpredictable comparison, that is, it should not be dependent on motor-predictions. On the other hand, the lack of a significant link between sensory attenuation and memory enhancement in Study III suggests that the production effect (MacDonald & MacLeod, 1998; MacLeod et al., 2010; Ekstrand et al., 1966; Hopkins & Edwards, 1972; Conway & Gathercole, 1987; Gathercole & Conway, 1988) is not related to the typical self-generation effects. Instead, it seems that the presence of a self-chosen and semi self-paced action boosts memory performance when predictability confounds are also controlled.

Third, we provide some interim, albeit inconclusive, evidence that these neuromodulatory processes might be behind the action-induced sensory suppression effects. Sensory attenuation of the N1 subcomponent, Tb (generated in the superior temporal gyrus; Näätänen & Picton, 1987), correlated with pupil dilation for self-generated sounds but only when they were presented shortly after the action and could be fully predicted by it (i.e., predictable session with short action-sound delays in Study III). In contrast, when the action-stimulus relationship was not fully contingent, this link is not evident despite the close temporal proximity between the action and the sound (Study II). The significant correlation in Study III fits with previous work pointing to neuromodulatory effects on sensory processing during movement by showing that application of norepinephrine in the barrel cortex reduces overall cortical firing irrespective of cortical layer, but it increases the firing of a small neural population (Castro-Alamancos & Bezdudnaya, 2015). However, if neuromodulatory processes suppress sensory processing in an indiscriminate way, then we should have found this link in Study II and in all the conditions of Study III.

Therefore, we argue that our results have two highly important implications: First, along with the differences in pupil response as a function of predictability in Study III (increased pupil diameter following actions that had predictable sensory consequences), we showed that the more pronounced the attenuation effects, the largest the pupil diameter for self-triggered and predictable inputs, suggesting that the link between action-induced sensory dampening and pupil dilation can be captured when the effects are maximized, namely when the action is fully predictive of the upcoming stimulus. Second, this link between sensory attenuation and pupil dilation was only obtained when the sound was presented shortly after the action, but not with the longer 750 ms delay in Study III. This might provide some interim evidence in favour of the hypothesis that a halo of neuromodulation relates to the attenuation effects when the action is in close temporal proximity with the sensory reafference. Crucially, the effects of such halo on sensory responses seem to fade out with longer action-sound delays.

Finally, we believe that our findings, along with the implications outlined in the previous paragraph, might contribute to elucidating the underlying neurophysiological mechanisms behind the observed link between sensory suppression and subcortical neuromodulation. We propose that the link between sensory attenuation and subcortical neuromodulation might depend on the combination of two processes operating simultaneously. These two processes are outlined in the following paragraphs and depicted in Figure 3.

The first process might result from the LC-NE projections on motor areas. In fact, there is already evidence that LC-NE innervates the primary motor cortex (Chandler et al., 2014) and that norepinephrine release in primary motor areas result in increased firing of motor neurons (Schiemann et al., 2015). Therefore, if LC-NE activity increases motor excitability, this might increase the inhibitory effect of motor engagement on auditory responses (Reznik & Mukamel, 2019), thereby attenuating activity in the auditory cortex. This possibility seems plausible also considering the original proposal by Näätänen and Picton (1987) that the unspecific component of N1 is generated in frontal motor and premotor cortex under the influence of the reticular formation that ultimately results in facilitating motor activity. Combining these lines of evidence, it is likely that the enhancing effects of LC-NE activation on motor excitability further increase the inhibitory effect of motor cortex on auditory neurons, thereby providing a plausible explanation for the links between subcortical neuromodulation and sensory suppression during movement. More importantly, this mechanism would yield stronger attenuation effects under fully predictable contexts, given our data showing larger LC-NE activity under predictable vs. unpredictable actions (see left and middle panel in Figure 3) which would result in higher motor excitability and therefore larger inhibitory effects of the motor cortex on auditory areas.

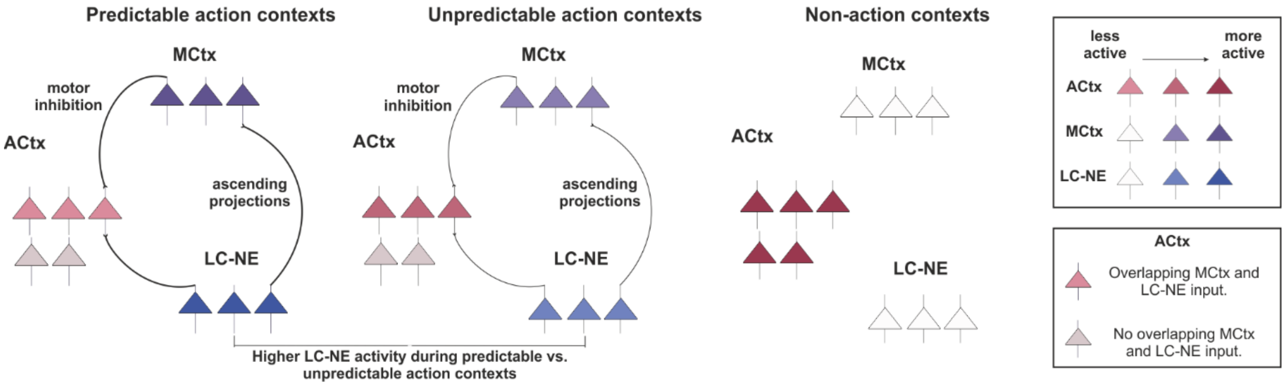


Figure 3. Schematic representation of the proposed mechanism behind the links between sensory attenuation and subcortical neuromodulation during movement. In action contexts (left and middle panel), movement will preferentially modulate activity in neurons that receive overlapping motor and neuromodulatory input. In highly predictable action contexts (left panel), LC-NE activity increases, thereby boosting motor excitability through ascending projections (thick black lines) of LC-NE to motor cortex (i.e., MCtx), which in turn increases the inhibitory effect of MCtx on auditory cortex (i.e., ACtx), resulting in larger attenuation effects. In unpredictable action contexts (middle panel), LC-NE activity increases as well, but to a lesser extent than in predictable action contexts. These increases will boost MCtx activity, but again to a lesser extent than in predictable action contexts (thin black lines). Importantly, although motor engagement will still attenuate ACtx responses, the attenuation effects in this case will be smaller than in predictable action contexts. Finally, in the absence of movement (non-action contexts; right panel), LC-NE and MCtx activity remains silent, and thus, responses in ACtx are merely driven by the sensory input, resulting in stronger responses than responses to stimulation *predicted by* or *presented close in time with* an action (as depicted in the action contexts panels).

The second process we propose is related to the effects of motor and neuromodulatory inputs on auditory areas. There is already evidence showing that motor and neuromodulatory inputs overlap onto the same auditory neurons (Nelson & Mooney, 2016; for a review see Schneider & Mooney, 2018). This might suggest that despite the global effect of neuromodulation in the neocortex, neuromodulatory processes would differentially affect those cells that also receive top-down motor-driven predictions (which would be the case only in conditions that stimulus-specificity mechanisms can operate, that is in highly predictable contexts). These neuromodulatory influences could possibly act as an attentional and orienting filter (Aston-Jones & Cohen, 2005) that facilitates or inhibits sensory responses (depending on the receptor mediating the effects; Vitrac & Benoit-Marand, 2017) for stimulation falling close in time with the action. Given that norepinephrine application in sensory areas has been found to suppress the firing of the vast majority of neurons, but meanwhile increases firing of a small neural population (Castro-Alamancos & Bezdudnaya, 2015), we assume that the net effect on auditory areas would be of inhibitory nature. Therefore, although the entire auditory cortex will receive LC-NE input during movement, only neurons that receive convergent and simultaneous input from motor areas would be the ones modulated. This mechanism would consequently result in greater differences in unspecific modulation between self- and externally-generated sounds, because the latter would lack any neuromodulatory and motor input due to the absence of movement. Thus, we propose that ascending neuromodulatory information might alter auditory cortical responsiveness (Schneider & Mooney, 2018; Reimer et al. 2014), specifically for the neural populations that receive inputs from both motor and neuromodulatory processes.

5.5. Strengths and limitations

The strength of the present thesis lies in addressing auditory processing around actions from the angles of perceptual processing (sound detection and discrimination), high-level cognition (memory), sensory processing (auditory evoked potentials), and subcortical neuromodulation (pupil diameter). Using psychophysics, electroencephalography, and pupillometry, we showed how the action effects and their interactions with other factors (e.g., sound intensity, predictability, action-sound delay) are reflected in behavioural responses and neurophysiological processes. Most importantly, the present work employed a novel experimental design in Study II and Study III, that is a combination of self-generation and a memory recognition task, and was the first – to the best of our knowledge – to attempt to link high-level memory processes, sensory attenuation, and subcortical neuromodulation. We showed that the links between these manifestations of the action effects depend on whether self-generation is conflated with predictability or not, and we provide interim evidence supporting the view that arousal-related unspecific modulatory mechanisms might partly drive the action-induced sensory suppression. Finally, we propose a model that might explain why the links between sensory attenuation and subcortical neuromodulations during actions might be obtained under certain experimental paradigms.

However, we also acknowledge certain limitations of the present work that should be addressed in future studies. First, in Study I, we assessed the effects of actions on detection and discrimination performance, but the action-sound relationship was fully contingent only in the discrimination task. In the detection task, participants had to press a button twice, but the sound was only presented after one of the two button presses. Although the findings of Study I fit well with the idea of the action effects being limited to perceptual bias, rather than sensitivity measures (supported by both tasks of Study I), we cannot exclude the possibility that the 50% action-sound contingency in the detection task might have minimized any potential differences in detection thresholds between active and passive conditions, thereby leading to null findings. Second, the coincidence conditions we employed here (Study II and unpredictable session in Study III) differ from the original coincidence studies by Horváth and colleagues (e.g., Horváth et al., 2012). In the coincidence task by Horváth and colleagues (2012), participants had to press a button several times and concurrently, but independently from the button presses, a sound sequence with random between-sound intervals was presented. This resulted in tones that were presented 0, 250, 500, 750, or 1000+ ms after the button press, and action-sound coincidences were considered the tones that were presented immediately after the button press (i.e., zero delay; Horváth et al., 2012). In contrast, in our coincidence tasks, 50% of the button presses resulted in a sound (after zero or 150/750 ms delay, in Study II and Study III, respectively). Therefore, our coincidence tasks are not directly comparable with the study by Horváth et al. (2012) due to former providing higher probability of occurrence of sounds following an action. Third, we acknowledge that the linear correlation approach we used to examine the possible links between sensory attenuation and pupil dilation might not have fully captured the relationship between the two and therefore we believe that further work is needed to elucidate the links between action-induced subcortical neuromodulation and sensory attenuation.

5.6. Future directions

On a general note, the findings obtained in the present thesis raise the need to further explore the links between the sensory processing and memory encoding of sounds triggered by our own movement. For example, functional neuroimaging work should examine the functional coupling between the areas that are thought to be involved in the self-generation effects (i.e., auditory cortical responses, motor activity, and LC-NE). In a similar fashion, the present thesis raises the need to test for possible links between the action effects on how sounds are processed in sensory-specific and memory-related areas, by examining the links between auditory cortical responses and hippocampal activity during self-generation.

Additionally, understanding the mechanisms underlying sensory attenuation for self-initiated stimulation is highly relevant for clinical work. Sensory attenuation is known to be reduced in schizophrenia (Ford et al., 2001, 2014; Ford & Mathalon, 2005), while evidence has already proposed that schizophrenia might be linked to LC-NE dysfunction (Mäki-Marttunen et al., 2020) and impaired modulation and regulation of neurotransmission (Devor et al., 2017). Future work could, therefore, assess whether the abolished attenuation in schizophrenia relates to abnormalities in neuromodulatory processes.

Meanwhile, the symptoms and abnormalities observed in schizophrenia research might reflect the extreme point of a continuum of schizophrenia that includes similar, albeit less severe and sub-clinical, symptomatology observable in the healthy population (Nelson, Seal et al., 2013; Ettinger et al., 2014). However, to date, only few studies have tested for possible links between schizotypy traits and sensory processing of self-initiated stimulation, showing that reduction in perceptual attenuation is linked to positive schizotypy traits and delusional ideation in non-clinical individuals (Asimakidou et al., 2022; Teufel et al., 2010). This raises the need for future work to examine the inter-individual variability of self-generation effects in the healthy population and address whether sensory and perceptual attenuation, as well as subcortical neuromodulation relate to schizotypy traits that could possibly point to increased liability to schizophrenia and psychosis (Ettinger et al., 2014).

CHAPTER 6: Conclusions

- Actions modulate perceptual and memory processes, but the direction of the effects depends on the predictability afforded by actions and the intensity of the sensory feedback. When actions afford better temporal predictability and generate supra-threshold stimulation, perceived intensity is suppressed and memory is weakened. In contrast, when actions afford better temporal predictability but generate near-threshold stimulation, perceived intensity is enhanced. We propose that these findings point to generalized predictive processes, that are not necessarily driven by motor-related signals. However, when predictability is controlled, such generalized predictive mechanisms might no longer operate, and the presence of an action that affords temporal control over the upcoming stimulus boosts memory encoding.

- Stimulus-specific and unspecific mechanisms co-exist and collaboratively shape auditory responses during actions. In favour of the stimulus-specificity of the effects, sensory attenuation is strongest with fully predictable action-sound relationship and short action-sound delays, even when providing identical predictability for externally-generated stimulation. Meanwhile, unspecific mechanisms operate as well, as shown by the fact that attenuation prevails even in the absence of contingent action-sound relationships and it is not clearly driven by modulations in primary auditory areas.

- Sensory attenuation correlates with weakened memory performance, but only when temporal predictability is higher for self- compared to externally-generated sounds, pointing again to a generalized predictive mechanism that boosts memory for the events eliciting larger prediction errors. However, when predictability is controlled, sensory attenuation no longer relates to the memory enhancements for self-generated sounds, suggesting that the production effect is not related to the typical self-generation effects. Thus, the mere presence of an action boosts memory performance, as long as predictability confounds are controlled and action alternatives are provided.

- Subcortical neuromodulation increases remarkably during actions but relates to the sensory attenuation effects only when both are strongest, that is when actions are fully predictive of the sound and sounds are presented in close temporal proximity with the action. We propose that the underlying neurophysiological mechanism could be conceptualized as a “halo” of neuromodulation around actions, mediated by the Locus Coeruleus Norepinephrine system, that unspecifically modulates auditory processing around movement. This mechanism operates through the ascending projections of the Locus Coeruleus Norepinephrine system to the motor cortex, and through the convergent motor and neuromodulatory inputs on auditory neurons.

- In sum, our results show that actions modulate different facets of auditory processing, namely low-level perception, high-level memory processes, sensory processing, and subcortical neuromodulation, and interact with other factors, such as sound intensity, predictability, action-sound delay, to shape how humans perceive, process, and memorize auditory information generated by their own movement.

CHAPTER 7: References

Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, *66*(4), 610–618. https://doi.org/10.1016/j.neuron.2010.04.014

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

Asimakidou, E., Job, X., & Kilteni, K. (2022). *The positive dimension of schizotypy is associated with a reduced attenuation and precision of self-generated touch* [Preprint]. Neuroscience. https://doi.org/10.1101/2022.01.22.476743

Aston-Jones, G., & Bloom, F. E. (1981). Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *The Journal of Neuroscience*, *1*(8), 876–886. https://doi.org/10.1523/jneurosci.01-08-00876.1981

Aston-Jones, G., & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study: Selective suppression of self-initiated sounds. *Psychophysiology*, *48*(9), 1276–1283. https://doi.org/10.1111/j.1469-8986.2011.01196.x

Baess, 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

Baess, P., Widmann, A., Roye, A., Schröger, E., & Jacobsen, T. (2009). Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. *European Journal of Neuroscience*, *29*(7), 1514–1521. https://doi.org/10.1111/j.1460-9568.2009.06683.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

Baldeweg, T. (2007). ERP Repetition Effects and Mismatch Negativity Generation: A Predictive Coding Perspective. *Journal of Psychophysiology*, *21*(3–4), 204–213. https://doi.org/10.1027/0269-8803.21.34.204

Bang, J. W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific Reports*, *7*(1), 17072. https://doi.org/10.1038/s41598-017-16885-2

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Barlas, Z., Hockley, W. E., & Obhi, S. S. (2017). The effects of freedom of choice in action selection on perceived mental effort and the sense of agency. *Acta Psychologica*, *180*, 122–129. https://doi.org/10.1016/j.actpsy.2017.09.004

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology*, *192*, 101821. https://doi.org/10.1016/j.pneurobio.2020.101821

Bausenhart, K. M., Di Luca, M., & Ulrich, R. (2018). Assessing Duration Discrimination: Psychophysical Methods and Psychometric Function Analysis. In A. Vatakis, F. Balcı, M. Di Luca, & Á. Correa (Eds.), *Timing and Time Perception: Procedures, Measures, & Applications* (pp. 52–78). BRILL. https://doi.org/10.1163/9789004280205_004

Bays, P. M., Flanagan, J. R., & Wolpert, D. M. (2006). Attenuation of Self-Generated Tactile Sensations Is Predictive, not Postdictive. *PLoS Biology*, *4*(2), e28. https://doi.org/10.1371/journal.pbio.0040028

Bays, P. M., Wolpert, D. M., & Flanagan, J. R. (2005). Perception of the Consequences of Self-Action Is Temporally Tuned and Event Driven. *Current Biology*, *15*(12), 1125–1128. https://doi.org/10.1016/j.cub.2005.05.023

Beaton, R., & Miller, J. M. (1975). Single cell activity in the auditory cortex of the unanesthetized, behaving monkey: Correlation with stimulus controlled behavior. *Brain Research*, *100*(3), 543–562. https://doi.org/10.1016/0006-8993(75)90157-2

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., Sangtian, S., Korzyukov, O., & Larson, C. R. (2016). A temporal predictive code for voice motor control: Evidence from ERP and behavioral responses to pitch-shifted auditory feedback. *Brain Research*, *1636*, 1–12. https://doi.org/10.1016/j.brainres.2016.01.040

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. https://doi.org/10.1038/35002078

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Bennett, C., Arroyo, S., & Hestrin, S. (2013). Subthreshold Mechanisms Underlying State-Dependent Modulation of Visual Responses. *Neuron*, *80*(2), 350–357. https://doi.org/10.1016/j.neuron.2013.08.007

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, *42*(1), 33–84. https://doi.org/10.1016/S0165-0173(03)00143-7

Beukema, S., Jennings, B. J., Olson, J. A., & Kingdom, F. A. A. (2019). The Pupillary Response to the Unknown: Novelty Versus Familiarity. *I-Perception*, *10*(5), 204166951987481. https://doi.org/10.1177/2041669519874817

Binda, P., Pereverzeva, M., & Murray, S. O. (2013). Attention to Bright Surfaces Enhances the Pupillary Light Reflex. *Journal of Neuroscience*, *33*(5), 2199–2204. https://doi.org/10.1523/JNEUROSCI.3440-12.2013

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *NeuroReport*, *11*(11), R11–R16. https://doi.org/10.1097/00001756-200008030-00002

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640. https://doi.org/10.1038/2870

Blakemore, S., Rees, G., & Frith, C. D. (1998). How do we predict the consequences of our actions? A functional imaging study. *Neuropsychologia*, *36*(6), 521–529. https://doi.org/10.1016/S0028-3932(97)00145-0

Bolt, N. K., & Loehr, J. D. (2021). Sensory Attenuation of the Auditory P2 Differentiates Self- from Partner-Produced Sounds during Joint Action. *Journal of Cognitive Neuroscience*, *33*(11), 2297–2310. https://doi.org/10.1162/jocn_a_01760

Bornert, P., & Bouret, S. (2021). Locus coeruleus neurons encode the subjective difficulty of triggering and executing actions. *PLOS Biology*, *19*(12), e3001487. https://doi.org/10.1371/journal.pbio.3001487

Bosnyak, D. J., Eaton, R., A., & Roberts, L. E. (2004). Distributed Auditory Cortical Representations Are Modified When Non-musicians Are Trained at Pitch Discrimination with 40 Hz Amplitude Modulated

Tones. *Cerebral Cortex*, *14*(10), 1088–1099. https://doi.org/10.1093/cercor/bhh068

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brooks, J. X., Carriot, J., & Cullen, K. E. (2015). Learning to expect the unexpected: Rapid updating in primate cerebellum during voluntary self-motion. *Nature Neuroscience*, *18*(9), 1310–1317. https://doi.org/10.1038/nn.4077

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

Brown, R. M., & Palmer, C. (2012). Auditory–motor learning influences auditory memory for music. *Memory & Cognition*, *40*(4), 567–578. https://doi.org/10.3758/s13421-011-0177-x

Budinger, E., Laszcz, A., Lison, H., Scheich, H., & Ohl, F. W. (2008). Non-sensory cortical and subcortical connections of the primary auditory cortex in Mongolian gerbils: Bottom-up and top-down processing of neuronal information via field AI. *Brain Research*, *1220*, 2–32. https://doi.org/10.1016/j.brainres.2007.07.084

Buran, B. N., von Trapp, G., & Sanes, D. H. (2014). Behaviorally Gated Reduction of Spontaneous Discharge Can Improve Detection Thresholds in Auditory Cortex. *Journal of Neuroscience*, *34*(11), 4076–4081. https://doi.org/10.1523/JNEUROSCI.4825-13.2014

Burin, D., Battaglini, A., Pia, L., Falvo, G., Palombella, M., & Salatino, A. (2017). Comparing intensities and modalities within the sensory attenuation paradigm: Preliminary evidence. *Journal of Advanced Research*, *8*(6), 649–653. https://doi.org/10.1016/j.jare.2017.08.001

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Cao, L., & Gross, J. (2015a). Attention Wins over Sensory Attenuation in a Sound Detection Task. *PLOS ONE*, *10*(8), e0136585. https://doi.org/10.1371/journal.pone.0136585

Cao, L., & Gross, J. (2015b). Cultural Differences in Perceiving Sounds Generated by Others: Self Matters. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01865

Carcea, I., Insanally, M. N., & Froemke, R. C. (2017). Dynamics of auditory cortical activity during behavioural engagement and auditory perception. *Nature Communications*, *8*(1), 14412. https://doi.org/10.1038/ncomms14412

Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., & Waszak, F. (2010). A New Look at Sensory Attenuation: Action-Effect Anticipation Affects Sensitivity, Not Response Bias. *Psychological Science*, *21*(12), 1740–1745. https://doi.org/10.1177/0956797610389187

Carmel, P. W., & Starr, A. (1963). ACOUSTIC AND NONACOUSTIC FACTORS MODIFYING MIDDLE-EAR MUSCLE ACTIVITY IN WAKING CATS. *Journal of Neurophysiology*, *26*(4), 598–616. https://doi.org/10.1152/jn.1963.26.4.598

Castro-Alamancos, M. A., & Bezdudnaya, T. (2015). Modulation of artificial whisking related signals in barrel cortex. *Journal of Neurophysiology*, *113*(5), 1287–1301. https://doi.org/10.1152/jn.00809.2014

Chagnaud, B. P., Banchi, R., Simmers, J., & Straka, H. (2015). Spinal corollary discharge modulates motion sensing during vertebrate locomotion. *Nature Communications*, *6*(1), 7982. https://doi.org/10.1038/ncomms8982

Chalk, M., Seitz, A. R., & Series, P. (2010). Rapidly learned stimulus expectations alter perception of motion.

Journal of Vision, *10*(8), 2–2. https://doi.org/10.1167/10.8.2

Chandler, D. J., Gao, W.-J., & Waterhouse, B. D. (2014). Heterogeneous organization of the locus coeruleus projections to prefrontal and motor cortices. *Proceedings of the National Academy of Sciences*, *111*(18), 6816–6821. https://doi.org/10.1073/pnas.1320827111

Chapman, C. E., Bushnell, M. C., Miron, D., Duncan, G. H., & Lund, J. P. (1987). Sensory perception during movement in man. *Experimental Brain Research*, *68*(3). https://doi.org/10.1007/BF00249795

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten Percent Electrode System for Topographic Studies of Spontaneous and Evoked EEG Activities. *American Journal of EEG Technology*, *25*(2), 83–92. https://doi.org/10.1080/00029238.1985.11080163

Chen, Z., Chen, X., Liu, P., Huang, D., & Liu, H. (2012). Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. *BMC Neuroscience*, *13*(1), 55. https://doi.org/10.1186/1471-2202-13-55

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping*, *28*(9), 868–879. https://doi.org/10.1002/hbm.20315

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Claxton, G. (1975). Why Can't We Tickle Ourselves? *Perceptual and Motor Skills*, *41*(1), 335–338. https://doi.org/10.2466/pms.1975.41.1.335

Clayton, K. K., Williamson, R. S., Hancock, K. E., Tasaka, G., Mizrahi, A., Hackett, T. A., & Polley, D. B. (2021). Auditory Corticothalamic Neurons Are Recruited by Motor Preparatory Inputs. *Current Biology*, *31*(2), 310-321.e5. https://doi.org/10.1016/j.cub.2020.10.027

Cohen, L. G., & Starr, A. (1987). LOCALIZATION, TIMING AND SPECIFICITY OF GATING OF SOMATOSENSORY EVOKED POTENTIALS DURING ACTIVE MOVEMENT IN MAN. *Brain*, *110*(2), 451–467. https://doi.org/10.1093/brain/110.2.451

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*(3), 341–361. https://doi.org/10.1016/0749-596X(87)90118-5

Costa-Faidella, J., Baldeweg, T., Grimm, S., & Escera, C. (2011). Interactions between "What" and "When" in the Auditory System: Temporal Predictability Enhances Repetition Suppression. *Journal of Neuroscience*, *31*(50), 18590–18597. https://doi.org/10.1523/JNEUROSCI.2599-11.2011

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience*, *9*(8), 587–600. https://doi.org/10.1038/nrn2457

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Cullen, K. E. (2004). Sensory signals during active versus passive movement. *Current Opinion in Neurobiology*, *14*(6), 698–706. https://doi.org/10.1016/j.conb.2004.10.002

Cullen, K. E. (2012). The vestibular system: Multimodal integration and encoding of self-motion for motor control. *Trends in Neurosciences*, *35*(3), 185–196. https://doi.org/10.1016/j.tins.2011.12.001

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, *9*(4), 183–191. 10.1002/(SICI)1097-0193(200004)9:4<183::AID-HBM1>3.0.CO;2-Z

Darriba, Á., Hsu, Y.-F., Van Ommen, S., & Waszak, F. (2021). Intention-based and sensory-based predictions. *Scientific Reports*, *11*(1), 19899. https://doi.org/10.1038/s41598-021-99445-z

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: The role of prediction and context. *Trends in Cognitive Sciences*, *19*(2), 92–99. https://doi.org/10.1016/j.tics.2014.12.004

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

Desantis, A., Waszak, F., & Gorea, A. (2016). Agency alters perceptual decisions about action-outcomes. *Experimental Brain Research*, *234*(10), 2819–2827. https://doi.org/10.1007/s00221-016-4684-7

Devor, A., Andreassen, O. A., Wang, Y., Mäki-Marttunen, T., Smeland, O. B., Fan, C.-C., Schork, A. J., Holland, D., Thompson, W. K., Witoelar, A., Chen, C.-H., Desikan, R. S., McEvoy, L. K., Djurovic, S., Greengard, P., Svenningsson, P., Einevoll, G. T., & Dale, A. M. (2017). Genetic evidence for role of integration of fast and slow neurotransmission in schizophrenia. *Molecular Psychiatry*, *22*(6), 792–801. https://doi.org/10.1038/mp.2017.33

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

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(03), 357. https://doi.org/10.1017/S0140525X00058027

Edeline, J.-M., Manunta, Y., & Hennevin, E. (2011). Induction of selective plasticity in the frequency tuning of auditory cortex and auditory thalamus neurons by locus coeruleus stimulation. *Hearing Research*, *274*(1–2), 75–84. https://doi.org/10.1016/j.heares.2010.08.005

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Eggermann, E., Kremer, Y., Crochet, S., & Petersen, C. C. H. (2014). Cholinergic Signals in Mouse Barrel Cortex during Active Whisker Sensing. *Cell Reports*, *9*(5), 1654–1660. https://doi.org/10.1016/j.celrep.2014.11.005

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. (1966). A frequency theory of verbal-discrimination learning. *Psychological Review*, *73*(6), 566–578. https://doi.org/10.1037/h0023876

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

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 229–240. https://doi.org/10.1037/0096-1523.27.1.229

Endo, N., Ito, T., Watanabe, K., & Nakazawa, K. (2021). Enhancement of loudness discrimination acuity for self-generated sound is independent of musical experience. *PLOS ONE*, *16*(12), e0260859. https://doi.org/10.1371/journal.pone.0260859

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. https://doi.org/10.1162/089892998562997

Escera, C., & Malmierca, M. S. (2014). The auditory novelty system: An attempt to integrate human and animal research: The auditory novelty system. *Psychophysiology*, *51*(2), 111–123. https://doi.org/10.1111/psyp.12156

Ettinger, U., Meyhöfer, I., Steffens, M., Wagner, M., & Koutsouleris, N. (2014). Genetics, Cognition, and

Neurobiology of Schizotypal Personality: A Review of the Overlap with Schizophrenia. *Frontiers in Psychiatry*, *5*. https://doi.org/10.3389/fpsyt.2014.00018

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories—The role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011

Feldman, H., & Friston, K. J. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00215

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. *Journal of Neuroscience*, *30*(49), 16643–16650. https://doi.org/10.1523/JNEUROSCI.1809-10.2010

Fonken, Y. M., Kam, J. W. Y., & Knight, R. T. (2020). A differential role for human hippocampus in novelty and contextual processing: Implications for P300. *Psychophysiology*, *57*(7). https://doi.org/10.1111/psyp.13400

Foote, S. L., Bloom, F. E., & Aston-Jones, G. (1983). Nucleus locus ceruleus: New evidence of anatomical and physiological specificity. *Physiological Reviews*, *63*(3), 844–914. https://doi.org/10.1152/physrev.1983.63.3.844

Foote, S. L., Freedman, R., & Oliver, A. P. (1975). Effects of putative neurotransmitters on neuronal activity in monkey auditory cortex. *Brain Research*, *86*(2), 229–242. https://doi.org/10.1016/0006-8993(75)90699-X

Foote, S. L., & Morrison, J. H. (1987). Extrathalamic Modulation of Cortical Function. *Annual Review of Neuroscience*, *10*(1), 67–95. https://doi.org/10.1146/annurev.ne.10.030187.000435

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

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., 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

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. J. (2001). Book Review: Brain Function, Nonlinear Coupling, and Neuronal Transients. *The Neuroscientist*, *7*(5), 406–418. https://doi.org/10.1177/107385840100700510

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1404), 1771–1788. https://doi.org/10.1098/rstb.2000.0734

Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., Yágüez, L., van Haren, N. E. M., Matsumoto, K., & McGuire, P. K. (2006). An fMRI Study of Verbal Self-monitoring: Neural Correlates of Auditory Verbal Feedback. *Cerebral Cortex*, *16*(7), 969–977. https://doi.org/10.1093/cercor/bhj039

Fukutomi, M., & Carlson, B. A. (2020). A History of Corollary Discharge: Contributions of Mormyrid Weakly Electric Fish. *Frontiers in Integrative Neuroscience*, *14*, 42. https://doi.org/10.3389/fnint.2020.00042

Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods*, *43*(4), 1171–1181. https://doi.org/10.3758/s13428-011-0109-5

Gagnepain, P., Henson, R., Chételat, G., Desgranges, B., Lebreton, K., & Eustache, F. (2011). Is Neocortical–Hippocampal Connectivity a Better Predictor of Subsequent Recollection than Local Increases in Hippocampal Activity? New Insights on the Role of Priming. *Journal of Cognitive Neuroscience*, *23*(2), 391–403. https://doi.org/10.1162/jocn.2010.21454

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*(2), 110–119. https://doi.org/10.3758/BF03213478

Geng, J. J., Blumenfeld, Z., Tyson, T. L., & Minzenberg, M. J. (2015). Pupil diameter reflects uncertainty in attentional selection during visual search. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00435

Gescheider, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed). L. Erlbaum Associates.

Glennon, E., Carcea, I., Martins, A. R. O., Multani, J., Shehu, I., Svirsky, M. A., & Froemke, R. C. (2019). Locus coeruleus activation accelerates perceptual learning. *Brain Research*, *1709*, 39–49. https://doi.org/10.1016/j.brainres.2018.05.048

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Guilford, J. P. (1954). *Psychometric methods.* Mcgraw Hill.

Gunji, A., Hoshiyama, M., & Kakigi, R. (2001). Auditory response following vocalization: A magnetoencephalographic study. *Clinical Neurophysiology*, *112*(3), 514–520. https://doi.org/10.1016/S1388-2457(01)00462-X

Gygi, B., & Shafiro, V. (2010). Development of the Database for Environmental Sound Research and Application (DESRA): Design, Functionality, and Retrieval Considerations. *EURASIP Journal on Audio, Speech, and Music Processing*, *2010*, 1–12. https://doi.org/10.1155/2010/654914

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-Related Brain Potential Correlates of Human Auditory Sensory Memory-Trace Formation. *Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385. https://doi.org/10.1038/nn827

Halgren, E. (1991). Firing of human hippocampal units in relation to voluntary movements. *Hippocampus*, *1*(2), 153–161. https://doi.org/10.1002/hipo.450010204

Han, N., Jack, B. N., Hughes, G., Elijah, R. B., & Whitford, T. J. (2021). Sensory attenuation in the absence of movement: Differentiating motor action from sense of agency. *Cortex*, *141*, 436–448. https://doi.org/10.1016/j.cortex.2021.04.010

Harrison, A. W., Mannion, D. J., Jack, B. N., Griffiths, O., Hughes, G., & Whitford, T. J. (2021). Sensory attenuation is modulated by the contrasting effects of predictability and control. *NeuroImage*, *237*, 118103. https://doi.org/10.1016/j.neuroimage.2021.118103

Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping*, *20*(1), 22–28. https://doi.org/10.1002/hbm.10119

Hayat, H., Regev, N., Matosevich, N., Sales, A., Paredes-Rodriguez, E., Krom, A. J., Bergman, L., Li, Y., Lavigne, M., Kremer, E. J., Yizhar, O., Pickering, A. E., & Nir, Y. (2020). Locus coeruleus norepinephrine activity mediates sensory-evoked awakenings from sleep. *Science Advances*, *6*(15). https://doi.org/10.1126/sciadv.aaz4232

Hazemann, P., Audin, G., & Lille, F. (1975). Effect of voluntary self-paced movements upon auditory and somatosensory evoked potentials in man. *Electroencephalography and Clinical Neurophysiology*, *39*(3), 247–254. https://doi.org/10.1016/0013-4694(75)90146-7

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

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

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857

Herwig, A., Prinz, W., & Waszak, F. (2007). Two Modes of Sensorimotor Integration in Intention-Based and Stimulus-Based Actions. *Quarterly Journal of Experimental Psychology*, *60*(11), 1540–1554. https://doi.org/10.1080/17470210601119134

Hesse, M. D., Nishitani, N., Fink, G. R., Jousmaki, V., & Hari, R. (2010). Attenuation of Somatosensory Responses to Self-Produced Tactile Stimulation. *Cerebral Cortex*, *20*(2), 425–432. https://doi.org/10.1093/cercor/bhp110

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665–667. https://doi.org/10.1038/nn.4284

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback: *NeuroReport*, *8*(9), 2379–2382. https://doi.org/10.1097/00001756-199707070-00055

Hocking, J., Dzafic, I., Kazovsky, M., & Copland, D. A. (2013). NESSTI: Norms for Environmental Sound Stimuli. *PLoS ONE*, *8*(9), e73382. https://doi.org/10.1371/journal.pone.0073382

Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, *25*(1), 16–26. https://doi.org/10.3758/BF03204445

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/S0140525X01000103

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*(4), 534–537. https://doi.org/10.1016/S0022-5371(72)80036-7

Horváth, J. (2013a). Attenuation of auditory ERPs to action-sound coincidences is not explained by voluntary allocation of attention: Action-sound coincidence effect is not attentional. *Psychophysiology*, *50*(3), 266–273. https://doi.org/10.1111/psyp.12009

Horváth, J. (2013b). Action-sound coincidence-related attenuation of auditory ERPs is not modulated by affordance compatibility. *Biological Psychology*, *93*(1), 81–87. https://doi.org/10.1016/j.biopsycho.2012.12.008

Horváth, J. (2014). The role of mechanical impact in action-related auditory attenuation. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(4), 1392–1406. https://doi.org/10.3758/s13415-014-0283-x

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., & Burgyán, A. (2013). No evidence for peripheral mechanism attenuating auditory ERPs to self-induced tones: Auditory ERP attenuation not due to peripheral effect. *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., 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

Hughes, G., Desantis, A., & Waszak, F. (2013a). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, *139*(1), 133–151. https://doi.org/10.1037/a0028566

Hughes, G., Desantis, A., & Waszak, F. (2013b). Attenuation of auditory N1 results from identity-specific action-effect prediction. *European Journal of Neuroscience*, *37*(7), 1152–1158. https://doi.org/10.1111/ejn.12120

Hughes, G., & Waszak, F. (2011). ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *NeuroImage*, *56*(3), 1632–1640. https://doi.org/10.1016/j.neuroimage.2011.02.057

Hupe, J. M., Lamirel, C., & Lorenceau, J. (2009). Pupil dynamics during bistable motion perception. *Journal of Vision*, *9*(7), 10–10. https://doi.org/10.1167/9.7.10

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature Neuroscience*, *8*(2), 234–241. https://doi.org/10.1038/nn1386

Jaramillo, S., & Zador, A. M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature Neuroscience*, *14*(2), 246–251. https://doi.org/10.1038/nn.2688

Jeffreys, H. (1961). *Theory of Probability* (3rd ed., Ser. Oxford Classic Texts in the Physical Sciences.). Oxford Univ. Press.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, *89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology*, *64*(10), 1971–1989. https://doi.org/10.1080/17470218.2011.588335

Kafkas, A., & Montaldi, D. (2015). The pupillary response discriminates between subjective and objective familiarity and novelty. *Psychophysiology*, *52*(10), 1305–1316. https://doi.org/10.1111/psyp.12471

Kaiser, J., & Schütz-Bosbach, S. (2018). Sensory attenuation of self-produced signals does not rely on self-specific motor predictions. *European Journal of Neuroscience*, *47*(11), 1303–1310. https://doi.org/10.1111/ejn.13931

Kelley, D. B., & Bass, A. H. (2010). Neurobiology of vocal communication: Mechanisms for sensorimotor integration and vocal patterning. *Current Opinion in Neurobiology*, *20*(6), 748–753.

https://doi.org/10.1016/j.conb.2010.08.007

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

Kilteni, K., Engeler, P., & Ehrsson, H. H. (2020). Efference Copy Is Necessary for the Attenuation of Self-Generated Touch. *iScience*, *23*(2), 100843. https://doi.org/10.1016/j.isci.2020.100843

Kim, A. J., Fitzgerald, J. K., & Maimon, G. (2015). Cellular evidence for efference copy in Drosophila visuomotor processing. *Nature Neuroscience*, *18*(9), 1247–1255. https://doi.org/10.1038/nn.4083

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences*, *111*(24), 8997–9002. https://doi.org/10.1073/pnas.1319438111

Kimura, M. (2021). Prediction, Suppression of Visual Response, and Modulation of Visual Perception: Insights From Visual Evoked Potentials and Representational Momentum. *Frontiers in Human Neuroscience*, *15*, 730962. https://doi.org/10.3389/fnhum.2021.730962

Kingdom, F. A. A., & Prins, N. (2016). *Psychophysics: A practical introduction* (Second edition). Elsevier/Academic Press.

Kirwan, C. B., & Stark, C. E. L. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory*, *14*(9), 625–633. https://doi.org/10.1101/lm.663507

Klaffehn, A. L., Baess, P., Kunde, W., & Pfister, R. (2019). Sensory attenuation prevails when controlling for temporal predictability of self- and externally generated tones. *Neuropsychologia*, *132*, 107145. https://doi.org/10.1016/j.neuropsychologia.2019.107145

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1-16.

Knapen, T., de Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and Ocular Factors Jointly Determine Pupil Responses under Equiluminance. *PLOS ONE*, *11*(5), e0155574. https://doi.org/10.1371/journal.pone.0155574

Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2012). The Cerebellum Generates Motor-to-Auditory Predictions: ERP Lesion Evidence. *Journal of Cognitive Neuroscience*, *24*(3), 698–706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schröger, E., & Kotz, S. A. (2013a). Prediction errors in self- and externally-generated deviants. *Biological Psychology*, *92*(2), 410–416. https://doi.org/10.1016/j.biopsycho.2012.11.017

Knolle, F., Schröger, E., & Kotz, S. A. (2013b). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, *49*(9), 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577. https://doi.org/10.1017/S0048577201990559

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*, *114*(39), 10473–10478. https://doi.org/10.1073/pnas.1705652114

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cerebral Cortex*, *22*(9), 2197–2206. https://doi.org/10.1093/cercor/bhr310

Korka, B., Schröger, E., & Widmann, A. (2019). Action Intention-based and Stimulus Regularity-based Predictions: Same or Different? *Journal of Cognitive Neuroscience*, *31*(12), 1917–1932. https://doi.org/10.1162/jocn_a_01456

Korka, B., Schröger, E., & Widmann, A. (2020). What *exactly* is missing here? The sensory processing of unpredictable omissions is modulated by the specificity of expected action-effects. *European Journal of Neuroscience*, *52*(12), 4667–4683. https://doi.org/10.1111/ejn.14899

Korka, B., Widmann, A., Waszak, F., Darriba, Á., & Schröger, E. (2021). The auditory brain in action: Intention determines predictive processing in the auditory system—A review of current paradigms and findings. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01992-z

Krawczyk, M. C., Fernández, R. S., Pedreira, M. E., & Boccia, M. M. (2017). Toward a better understanding on the role of prediction error on memory processes: From bench to clinic. *Neurobiology of Learning and Memory*, *142*, 13–20. https://doi.org/10.1016/j.nlm.2016.12.011

Kuchibhotla, K. V., Gill, J. V., Lindsay, G. W., Papadoyannis, E. S., Field, R. E., Sten, T. A. H., Miller, K. D., & Froemke, R. C. (2017). Parallel processing by cortical inhibition enables context-dependent behavior. *Nature Neuroscience*, *20*(1), 62–71. https://doi.org/10.1038/nn.4436

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

Kühn, S., Nenchev, I., Haggard, P., Brass, M., Gallinat, J., & Voss, M. (2011). Whodunnit? Electrophysiological Correlates of Agency Judgements. *PLoS ONE*, *6*(12), e28657. https://doi.org/10.1371/journal.pone.0028657

Kullback, S. (1997). *Information theory and statistics*. Dover Publications.

Kumaran, D., & Maguire, E. A. (2009). Novelty signals: A window into hippocampal information processing. *Trends in Cognitive Sciences*, *13*(2), 47–54. https://doi.org/10.1016/j.tics.2008.11.004

Kunde, W., & Wühr, P. (2004). Actions blind to conceptually overlapping stimuli. *Psychological Research Psychologische Forschung*, *68*(4). https://doi.org/10.1007/s00426-003-0156-3

Lange, K. (2013). The ups and downs of temporal orienting: A review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00263

Lee, C. R., & Margolis, D. J. (2016). Pupil Dynamics Reflect Behavioral Choice and Learning in a Go/NoGo Tactile Decision-Making Task in Mice. *Frontiers in Behavioral Neuroscience*, *10*. https://doi.org/10.3389/fnbeh.2016.00200

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Linares, D., & López-Moliner, J. (2016). quickpsy: An R Package to Fit Psychometric Functions for Multiple Groups. *The R Journal*, *8*(1), 122. https://doi.org/10.32614/RJ-2016-008

Loughlin, S. E., Foote, S. L., & Fallon, J. H. (1982). Locus coeruleus projections to cortex: Topography, morphology and collateralization. *Brain Research Bulletin*, *9*(1–6), 287–294. https://doi.org/10.1016/0361-9230(82)90142-3

Lubinus, C., Einhäuser, W., Schiller, F., Kircher, T., Straube, B., & van Kemenade, B. M. (2021). *Action-based predictions affect visual perception, neural processing, and pupil size, regardless of temporal predictability* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.02.11.430717

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica*, *98*(2–3), 291–310. https://doi.org/10.1016/S0001-6918(97)00047-4

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Majchrowicz, B., & Wierzchoń, M. (2021). Sensory attenuation of action outcomes of varying amplitude and valence. *Consciousness and Cognition*, *87*, 103058. https://doi.org/10.1016/j.concog.2020.103058

Makeig, S., Müller, M. M., & Rockstroh, B. (1996). Effects of voluntary movements on early auditory brain responses. *Experimental Brain Research*, *110*(3). https://doi.org/10.1007/BF00229149

Mäki-Marttunen, V., Andreassen, O. A., & Espeseth, T. (2020). The role of norepinephrine in the pathophysiology of schizophrenia. *Neuroscience & Biobehavioral Reviews*, *118*, 298–314. https://doi.org/10.1016/j.neubiorev.2020.07.038

Mama, Y., & Icht, M. (2016). Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals the superiority of writing over vocalising. *Memory*, *24*(1), 98–113. https://doi.org/10.1080/09658211.2014.986135

Manunta, Y., & Edeline, J.-M. (2004). Noradrenergic Induction of Selective Plasticity in the Frequency Tuning of Auditory Cortex Neurons. *Journal of Neurophysiology*, *92*(3), 1445–1463. https://doi.org/10.1152/jn.00079.2004

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

Martikainen, M. H., Kaneko, K., & Hari, R. (2004). Suppressed Responses to Self-triggered Sounds in the Human Auditory Cortex. *Cerebral Cortex*, *15*(3), 299–302. https://doi.org/10.1093/cercor/bhh131

Martins, A. R. O., & Froemke, R. C. (2015). Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. *Nature Neuroscience*, *18*(10), 1483–1492. https://doi.org/10.1038/nn.4090

Marzecová, A., Schettino, A., Widmann, A., SanMiguel, I., Kotz, S. A., & Schröger, E. (2018). Attentional gain is modulated by probabilistic feature expectations in a spatial cueing task: ERP evidence. *Scientific Reports*, *8*(1), 54. https://doi.org/10.1038/s41598-017-18347-1

Mathias, B., Palmer, C., Perrin, F., & Tillmann, B. (2015). Sensorimotor Learning Enhances Expectations During Auditory Perception. *Cerebral Cortex*, *25*(8), 2238–2254. https://doi.org/10.1093/cercor/bhu030

McCormick, D. A. (1989). Cholinergic and noradrenergic modulation of thalamocortical processing. *Trends in Neurosciences*, *12*(6), 215–221. https://doi.org/10.1016/0166-2236(89)90125-2

McGinley, M. J., David, S. V., & McCormick, D. A. (2015). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, *87*(1), 179–192. https://doi.org/10.1016/j.neuron.2015.05.038

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, *119*(3), 907–917. https://doi.org/10.1093/brain/119.3.907

Melcher, T., Weidema, M., Eenshuistra, R. M., Hommel, B., & Gruber, O. (2008). The neural substrate of the ideomotor principle: An event-related fMRI analysis. *NeuroImage*, *39*(3), 1274–1288. https://doi.org/10.1016/j.neuroimage.2007.09.049

Miall, R. C., & Wolpert, D. M. (1996). Forward Models for Physiological Motor Control. *Neural Networks*, *9*(8), 1265–1279. https://doi.org/10.1016/S0893-6080(96)00035-4

Mifsud, N. G., Beesley, T., Watson, T. L., Elijah, R. B., Sharp, T. S., & Whitford, T. J. (2018). Attenuation of visual evoked responses to hand and saccade-initiated flashes. *Cognition*, *179*, 14–22. https://doi.org/10.1016/j.cognition.2018.06.005

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

Mock, J. R., Foundas, A. L., & Golob, E. J. (2011). Modulation of sensory and motor cortex activity during speech preparation: Modulation of sensory and motor cortex activity. *European Journal of Neuroscience*, *33*(5), 1001–1011. https://doi.org/10.1111/j.1460-9568.2010.07585.x

Mondor, T. A., & Morin, S. R. (2004). Primacy, Recency, and Suffix Effects in Auditory Short-Term Memory for Pure Tones: Evidence From a Probe Recognition Paradigm. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *58*(3), 206–219. https://doi.org/10.1037/h0087445

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.04.2.p061

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology*, *20*(8), 750–756. https://doi.org/10.1016/j.cub.2010.02.045

Müller-Preuss, P., & Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Research*, *215*(1–2), 61–76. https://doi.org/10.1016/0006-8993(81)90491-1

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140–4154. https://doi.org/10.1002/hbm.22466

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The Simple Act of Choosing Influences Declarative Memory. *Journal of Neuroscience*, *35*(16), 6255–6264. https://doi.org/10.1523/JNEUROSCI.4181-14.2015

Myers, J. C., Mock, J. R., & Golob, E. J. (2020). Sensorimotor Integration Can Enhance Auditory Perception. *Scientific Reports*, *10*(1), 1496. https://doi.org/10.1038/s41598-020-58447-z

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Naber, M., Frassle, S., Rutishauser, U., & Einhauser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*(2), 11–11. https://doi.org/10.1167/13.2.11

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/nn.3130

Navarra, R. L., Clark, B. D., Gargiulo, A. T., & Waterhouse, B. D. (2017). Methylphenidate Enhances Early-Stage Sensory Processing and Rodent Performance of a Visual Signal Detection Task. *Neuropsychopharmacology*, *42*(6), 1326–1337. https://doi.org/10.1038/npp.2016.267

Nelson, A., & Mooney, R. (2016). The Basal Forebrain and Motor Cortex Provide Convergent yet Distinct Movement-Related Inputs to the Auditory Cortex. *Neuron*, *90*(3), 635–648. https://doi.org/10.1016/j.neuron.2016.03.031

Nelson, A., Schneider, D. M., Takatoh, J., Sakurai, K., Wang, F., & Mooney, R. (2013). A Circuit for Motor Cortical Modulation of Auditory Cortical Activity. *Journal of Neuroscience*, *33*(36), 14342–14353. https://doi.org/10.1523/JNEUROSCI.2275-13.2013

Nelson, M. T., Seal, M. L., Pantelis, C., & Phillips, L. J. (2013). Evidence of a dimensional relationship between schizotypy and schizophrenia: A systematic review. *Neuroscience & Biobehavioral Reviews*, *37*(3), 317–327. https://doi.org/10.1016/j.neubiorev.2013.01.004

Neske, G. T., Nestvogel, D., Steffan, P. J., & McCormick, D. A. (2019). Distinct Waking States for Strong

Evoked Responses in Primary Visual Cortex and Optimal Visual Detection Performance. *The Journal of Neuroscience*, *39*(50), 10044–10059. https://doi.org/10.1523/JNEUROSCI.1226-18.2019

Neszmélyi, B., & Horváth, J. (2017). Consequences matter: Self-induced tones are used as feedback to optimize tone-eliciting actions: Self-induced tones used as feedback for actions. *Psychophysiology*, *54*(6), 904–915. https://doi.org/10.1111/psyp.12845

Neszmélyi, B., & Horváth, J. (2018). Temporal constraints in the use of auditory action effects for motor optimization. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(11), 1815–1829. https://doi.org/10.1037/xhp0000571

Niell, C. M., & Stryker, M. P. (2010). Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron*, *65*(4), 472–479. https://doi.org/10.1016/j.neuron.2010.01.033

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What Does Motor Efference Copy Represent? Evidence from Speech Production. *Journal of Neuroscience*, *33*(41), 16110–16116. https://doi.org/10.1523/JNEUROSCI.2137-13.2013

Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–470. https://doi.org/10.1016/j.conb.2007.07.006

Norman, K. A. (2010). How hippocampus and cortex contribute to recognition memory: Revisiting the complementary learning systems model. *Hippocampus*, *20*(11), 1217–1227. https://doi.org/10.1002/hipo.20855

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*, *88*(6), 1281–1296. https://doi.org/10.1016/j.neuron.2015.11.035

Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, *272*(1), 29–32. https://doi.org/10.1016/S0304-3940(99)00573-X

Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, *265*(2), 119–122. https://doi.org/10.1016/S0304-3940(99)00218-9

Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences*, *97*(20), 11120–11124. https://doi.org/10.1073/pnas.97.20.11120

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

Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. In *Progress in Brain Research* (Vol. 159, pp. 99–120). Elsevier. https://doi.org/10.1016/S0079-6123(06)59007-7

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

Otazu, G. H., Tai, L.-H., Yang, Y., & Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nature Neuroscience*, *12*(5), 646–654. https://doi.org/10.1038/nn.2306

Ott, C. G. M., & Jäncke, L. (2013). Processing of self-initiated speech-sounds is different in musicians. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00041

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity.

Memory & Cognition, *40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1

Pachitariu, M., Lyamzin, D. R., Sahani, M., & Lesica, N. A. (2015). State-Dependent Population Coding in Primary Auditory Cortex. *Journal of Neuroscience*, *35*(5), 2058–2073. https://doi.org/10.1523/JNEUROSCI.3318-14.2015

Pantev, C., Eulitz, C., Hampson, S., Ross, B., & Roberts, L. E. (1996). The Auditory Evoked “Off” Response: Sources and Comparison with the"On" and the “Sustained” Responses: *Ear and Hearing*, *17*(3), 255–265. https://doi.org/10.1097/00003446-199606000-00008

Paraskevoudi, N., & SanMiguel, I. (2021). *Sensory suppression and increased neuromodulation during actions disrupt memory encoding of unpredictable self-initiated stimuli* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.12.15.472750

Parker, P. R. L., Brown, M. A., Smear, M. C., & Niell, C. M. (2020). Movement-Related Signals in Sensory Areas: Roles in Natural Behavior. *Trends in Neurosciences*, *43*(8), 581–595. https://doi.org/10.1016/j.tins.2020.05.005

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology*, *9*(2), 74–80. https://doi.org/10.5709/acp-0133-x

Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., & Mendelsohn, A. (2018). Knowledge acquisition is governed by striatal prediction errors. *Nature Communications*, *9*(1), 1673. https://doi.org/10.1038/s41467-018-03992-5

Pinheiro, A. P., Schwartze, M., Gutierrez, F., & Kotz, S. A. (2019). When temporal prediction errs: ERP responses to delayed action-feedback onset. *Neuropsychologia*, *134*, 107200. https://doi.org/10.1016/j.neuropsychologia.2019.107200

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 13. https://doi.org/10.1167/15.8.13

Poe, G. R., Foote, S., Eschenko, O., Johansen, J. P., Bouret, S., Aston-Jones, G., Harley, C. W., Manahan-Vaughan, D., Weinshenker, D., Valentino, R., Berridge, C., Chandler, D. J., Waterhouse, B., & Sara, S. J. (2020). Locus coeruleus: A new look at the blue spot. *Nature Reviews Neuroscience*, *21*(11), 644–659. https://doi.org/10.1038/s41583-020-0360-9

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

Poulet, J. F. A., & Hedwig, B. (2003). Corollary Discharge Inhibition of Ascending Auditory Neurons in the Stridulating Cricket. *The Journal of Neuroscience*, *23*(11), 4717–4725. https://doi.org/10.1523/JNEUROSCI.23-11-04717.2003

Poulet, J. F. A., & Hedwig, B. (2006). The Cellular Basis of a Corollary Discharge. *Science*, *311*(5760), 518–522. https://doi.org/10.1126/science.1120847

Press, C., & Cook, R. (2015). Beyond action-specific simulation: Domain-general motor contributions to perception. *Trends in Cognitive Sciences*, *19*(4), 176–178. https://doi.org/10.1016/j.tics.2015.01.006

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Preuschoff, K., t’ Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline’s role in decision making. *Frontiers in Neuroscience*, *5*.

https://doi.org/10.3389/fnins.2011.00115

Pyasik, M., Burin, D., & Pia, L. (2018). On the relation between body ownership and sense of agency: A link at the level of sensory-related signals. *Acta Psychologica*, *185*, 219–228. https://doi.org/10.1016/j.actpsy.2018.03.001

Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., & Tolias, A. S. (2014). Pupil Fluctuations Track Fast Switching of Cortical States during Quiet Wakefulness. *Neuron*, *84*(2), 355–362. https://doi.org/10.1016/j.neuron.2014.09.033

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*(1), 13289. https://doi.org/10.1038/ncomms13289

Requarth, T., & Sawtell, N. B. (2011). Neural mechanisms for filtering self-generated sensory signals in cerebellum-like circuits. *Current Opinion in Neurobiology*, *21*(4), 602–608. https://doi.org/10.1016/j.conb.2011.05.031

Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), Classical Conditioning II: Current Research and Theory (pp. 64-99). New York: Appleton- Century-Crofts.

Reznik, D., Guttman, N., Buaron, B., Zion-Golumbic, E., & Mukamel, R. (2021). Action-locked Neural Responses in Auditory Cortex to Self-generated Sounds. *Cerebral Cortex*, *31*(12), 5560–5569. https://doi.org/10.1093/cercor/bhab179

Reznik, D., Henkin, Y., Levy, O., & Mukamel, R. (2015). Perceived Loudness of Self-Generated Sounds Is Differentially Modified by Expected Sound Intensity. *PLOS ONE*, *10*(5), e0127651. https://doi.org/10.1371/journal.pone.0127651

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

Richer, F., & Beatty, J. (1987). Contrasting Effects of Response Uncertainty on the Task-Evoked Pupillary Response and Reaction Time. *Psychophysiology*, *24*(3), 258–262. https://doi.org/10.1111/j.1469-8986.1987.tb00291.x

Rodenkirch, C., Liu, Y., Schriver, B. J., & Wang, Q. (2019). Locus coeruleus activation enhances thalamic feature selectivity via norepinephrine regulation of intrathalamic circuit dynamics. *Nature Neuroscience*, *22*(1), 120–133. https://doi.org/10.1038/s41593-018-0283-1

Rogawski, M. A., & Aghajanian, G. K. (1980). Modulation of lateral geniculate neurone excitability by noradrenaline microiontophoresis or locus coeruleus stimulation. *Nature*, *287*(5784), 731–734. https://doi.org/10.1038/287731a0

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, *24*(2), 113–121. https://doi.org/10.1016/S0166-2236(00)01685-4

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

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

Roussel, C., Hughes, G., & Waszak, F. (2014). Action prediction modulates both neurophysiological and psychophysical indices of sensory attenuation. *Frontiers in Human Neuroscience*, *8*.

https://doi.org/10.3389/fnhum.2014.00115

Roy, J. E. (2004). Dissociating Self-Generated from Passively Applied Head Motion: Neural Mechanisms in the Vestibular Nuclei. *Journal of Neuroscience*, *24*(9), 2102–2111. https://doi.org/10.1523/JNEUROSCI.3988-03.2004

Roy, J. E., & Cullen, K. E. (2001). Selective Processing of Vestibular Reafference during Self-Generated Head Motion. *The Journal of Neuroscience*, *21*(6), 2131–2142. https://doi.org/10.1523/JNEUROSCI.21-06-02131.2001

Rummell, B. P., Klee, J. L., & Sigurdsson, T. (2016). Attenuation of Responses to Self-Generated Sounds in Auditory Cortical Neurons. *The Journal of Neuroscience*, *36*(47), 12010–12026. https://doi.org/10.1523/JNEUROSCI.1564-16.2016

Salomon, G., & Starr, A. (1963). ELECTROMYOGRAPHY OF MIDDLE EAR MUSCLES IN MAN DURING MOTOR ACTIVITIES. *Acta Neurologica Scandinavica*, *39*(2), 161–168. https://doi.org/10.1111/j.1600-0404.1963.tb05317.x

Samson, F., Zeffiro, T. A., Toussaint, A., & Belin, P. (2011). Stimulus complexity and categorical effects in human auditory cortex: An activation likelihood estimation meta-analysis. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00241

SanMiguel, I., Todd, J., & Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP: Auditory N1 suppression: N1 components. *Psychophysiology*, *50*(4), 334–343. https://doi.org/10.1111/psyp.12024

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), 211–223. https://doi.org/10.1038/nrn2573

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and Cognition*, *17*(4), 1219–1227. https://doi.org/10.1016/j.concog.2008.01.003

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

Sawtell, N. B. (2017). Neural Mechanisms for Predicting the Sensory Consequences of Behavior: Insights from Electrosensory Systems. *Annual Review of Physiology*, *79*(1), 381–399. https://doi.org/10.1146/annurev-physiol-021115-105003

Schafer, E. W. P., & Marcus, M. M. (1973). Self-Stimulation Alters Human Sensory Brain Responses. *Science*, *181*(4095), 175–177. https://doi.org/10.1126/science.181.4095.175

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schiemann, J., Puggioni, P., Dacre, J., Pelko, M., Domanski, A., van Rossum, M. C. W., & Duguid, I. (2015). Cellular mechanisms underlying behavioral state-dependent bidirectional modulation of motor cortex output. *Cell Reports*, *11*(8), 1319–1330. https://doi.org/10.1016/j.celrep.2015.04.042

Schiffer, A.-M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at All the Entropy: Hippocampal, Caudate and Midbrain Contributions to Learning from Prediction Errors. *PLoS ONE*, *7*(5), e36445. https://doi.org/10.1371/journal.pone.0036445

Schneider, D. M. (2020). Reflections of action in sensory cortex. *Current Opinion in Neurobiology*, *64*, 53–59. https://doi.org/10.1016/j.conb.2020.02.004

Schneider, D. M., & Mooney, R. (2018). How Movement Modulates Hearing. *Annual Review of Neuroscience*, *41*(1), 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*****,** 391–395. https://doi.org/10.1038/s41586-018-0520-5

Schneider, D. M., Nelson, A., & Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, *513*(7517), 189–194. https://doi.org/10.1038/nature13724

Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641–664. https://doi.org/10.1111/ejn.12816

Schwarz, K. A., Pfister, R., Kluge, M., Weller, L., & Kunde, W. (2018). Do we see it or not? Sensory attenuation in the visual domain. *Journal of Experimental Psychology: General*, *147*(3), 418–430. https://doi.org/10.1037/xge0000353

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *NeuroReport*, *16*(16), 1781–1785. https://doi.org/10.1097/01.wnr.0000185017.29316.63

Shimaoka, D., Harris, K. D., & Carandini, M. (2018). Effects of Arousal on Mouse Sensory Cortex Depend on Modality. *Cell Reports*, *22*(12), 3160–3167. https://doi.org/10.1016/j.celrep.2018.02.092

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, *136*(6), 943–974. https://doi.org/10.1037/a0020541

Simões-Franklin, C., Whitaker, T. A., & Newell, F. N. (2011). Active and passive touch differentially activate somatosensory cortex in texture perception. *Human Brain Mapping*, *32*(7), 1067–1080. https://doi.org/10.1002/hbm.21091

Simpson, H. M. (1969). EFFECTS OF A TASK-RELEVANT RESPONSE ON PUPIL SIZE. *Psychophysiology*, *6*(2), 115–121. https://doi.org/10.1111/j.1469-8986.1969.tb02890.x

Sinclair, A. H., & Barense, M. D. (2019). Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation. *Trends in Neurosciences*, *42*(10), 727–739. https://doi.org/10.1016/j.tins.2019.08.007

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

Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Experimental Brain Research*, *222*(1–2), 149–157. https://doi.org/10.1007/s00221-012-3204-7

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, *43*(6), 482–489. https://doi.org/10.1037/h0055479

Starr, A. (1964). Influence of motor activity on click-evoked responses in the auditory pathway of waking cats. *Experimental Neurology*, *10*(3), 191–204. https://doi.org/10.1016/0014-4886(64)90062-7

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. MIT Press.

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*(6), 1089–1104. https://doi.org/10.1037/xge0000109

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. *Biological Psychiatry*, *84*(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

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

Suga, N., & Jen, P. H. (1975). Peripheral control of acoustic signals in the auditory system of echolocating bats. *The Journal of Experimental Biology*, *62*(2), 277–311.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Tapia, M. C., Cohen, L. G., & Starr, A. (1987). Attenuation of Auditory-Evoked Potentials during Voluntary Movement in Man. *International Journal of Audiology*, *26*(6), 369–373. https://doi.org/10.3109/00206098709081565

Teufel, C., Kingdon, A., Ingram, J. N., Wolpert, D. M., & Fletcher, P. C. (2010). Deficits in sensory prediction are related to delusional ideation in healthy individuals. *Neuropsychologia*, *48*(14), 4169–4172. https://doi.org/10.1016/j.neuropsychologia.2010.10.024

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

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

Timm, J., Schönwiesner, M., Schröger, E., & SanMiguel, I. (2016). Sensory suppression of brain responses to self-generated sounds is observed with and without the perception of agency. *Cortex*, *80*, 5–20. https://doi.org/10.1016/j.cortex.2016.03.018

Tonnquist-Uhlen, I., Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2003). Maturation of human central auditory system activity: The T-complex. *Clinical Neurophysiology*, *114*(4), 685–701. https://doi.org/10.1016/S1388-2457(03)00005-1

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/ejn.12936

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637. https://doi.org/10.1038/ncomms14637

van Elk, M., Salomon, R., Kannape, O., & Blanke, O. (2014). Suppression of the N1 auditory evoked potential for sounds generated by the upper and lower limbs. *Biological Psychology*, *102*, 108–117. https://doi.org/10.1016/j.biopsycho.2014.06.007

Van Slooten, J. C., Jahfari, S., Knapen, T., & Theeuwes, J. (2019). Correction: How pupil responses track value-based decision-making during and after reinforcement learning. *PLOS Computational Biology*, *15*(5), e1007031. https://doi.org/10.1371/journal.pcbi.1007031

Vaughan, H. G., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology*, *28*(4), 360–367. https://doi.org/10.1016/0013-4694(70)90228-2

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

Vinck, M., Batista-Brito, R., Knoblich, U., & Cardin, J. A. (2015). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron*, *86*(3), 740–754. https://doi.org/10.1016/j.neuron.2015.03.028

Vitrac, C., & Benoit-Marand, M. (2017). Monoaminergic modulation of motor cortex function. *Frontiers in*

Neural Circuits, *11*. https://doi.org/10.3389/fncir.2017.00072

von Holst, E. (1954). Relations between the central Nervous System and the peripheral organs. *The British Journal of Animal Behaviour*, *2*(3), 89–94. https://doi.org/10.1016/S0950-5601(54)80044-X

von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften*, *37*(20), 464–476. https://doi.org/10.1007/BF00622503

von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, *18*(1), 299–342. https://doi.org/10.1007/BF02409636

Waszak, F., Cardoso-Leite, P., & Hughes, G. (2012). Action effect anticipation: Neurophysiological basis and functional consequences. *Neuroscience & Biobehavioral Reviews*, *36*(2), 943–959. https://doi.org/10.1016/j.neubiorev.2011.11.004

Waszak, F., Wascher, E., Keller, P., Koch, I., Aschersleben, G., Rosenbaum, D. A., & Prinz, W. (2005). Intention-based and stimulus-based mechanisms in action selection. *Experimental Brain Research*, *162*(3), 346–356. https://doi.org/10.1007/s00221-004-2183-8

Weiskrantz, L., Elliott, J., & Darlington, C. (1971). Preliminary Observations on Tickling Oneself. *Nature*, *230*(5296), 598–599. https://doi.org/10.1038/230598a0

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011a). The Self in Social Interactions: Sensory Attenuation of Auditory Action Effects Is Stronger in Interactions with Others. *PLoS ONE*, *6*(7), e22723. https://doi.org/10.1371/journal.pone.0022723

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011b). The self in action effects: Selective attenuation of self-generated sounds. *Cognition*, *121*(2), 207–218. https://doi.org/10.1016/j.cognition.2011.06.011

Weller, L., Schwarz, K. A., Kunde, W., & Pfister, R. (2017). Was it me? – Filling the interval between action and effects increases agency but not sensory attenuation. *Biological Psychology*, *123*, 241–249. https://doi.org/10.1016/j.biopsycho.2016.12.015

Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition*, *115*(1), 26–38. https://doi.org/10.1016/j.cognition.2009.10.016

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125–11129. https://doi.org/10.1073/pnas.97.20.11125

Williams, S. R., & Chapman, C. E. (2000). Time Course and Magnitude of Movement-Related Gating of Tactile Detection in Humans. II. Effects of Stimulus Intensity. *Journal of Neurophysiology*, *84*(2), 863–875. https://doi.org/10.1152/jn.2000.84.2.863

Williams, S. R., & Chapman, C. E. (2002). Time Course and Magnitude of Movement-Related Gating of Tactile Detection in Humans. III. Effect of Motor Tasks. *Journal of Neurophysiology*, *88*(4), 1968–1979. https://doi.org/10.1152/jn.2002.88.4.1968

Williams, S. R., Shenasa, J., & Chapman, C. E. (1998). Time Course and Magnitude of Movement-Related Gating of Tactile Detection in Humans. I. Importance of Stimulus Location. *Journal of Neurophysiology*, *79*(2), 947–963. https://doi.org/10.1152/jn.1998.79.2.947

Wolpaw, J. R., & Penry, J. K. (1975). A temporal component of the auditory evoked response. *Electroencephalography and Clinical Neurophysiology*, *39*(6), 609–620. https://doi.org/10.1016/0013-4694(75)90073-5

Wolpert, D., Ghahramani, Z., & Jordan, M. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences*, *109*(9), 3593–3598. https://doi.org/10.1073/pnas.1120118109

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075–1080. https://doi.org/10.1038/nature05852

Yebra, M., Galarza-Vallejo, A., Soto-Leon, V., Gonzalez-Rosa, J. J., de Berker, A. O., Bestmann, S., Oliviero, A., Kroes, M. C. W., & Strange, B. A. (2019). Action boosts episodic memory encoding in humans via engagement of a noradrenergic system. *Nature Communications*, *10*(1), 3534. https://doi.org/10.1038/s41467-019-11358-8

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Yon, D., Gilbert, S. J., de Lange, F. P., & Press, C. (2018). Action sharpens sensory representations of expected outcomes. *Nature Communications*, *9*(1), 4288. https://doi.org/10.1038/s41467-018-06752-7

Yon, D., & Press, C. (2017). Predicted action consequences are perceptually facilitated before cancellation. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(6), 1073–1083. https://doi.org/10.1037/xhp0000385

Yon, D., Zainzinger, V., de Lange, F., Eimer, M., & Press, C. (2019). *Action biases perceptual decisions toward expected outcomes.* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/3zp8n

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002

Zhou, M., Liang, F., Xiong, X. R., Li, L., Li, H., Xiao, Z., Tao, H. W., & Zhang, L. I. (2014). Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. *Nature Neuroscience*, *17*(6), 841–850. https://doi.org/10.1038/nn.3701