UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Institut für systemische Neurowissenschaften

Prof. Dr. med. Christian Büchel

Integrating sensory input and prior expectations in structured auditory sequences

Dissertation

zur Erlangung des Doktorgrades PhD
an der Medizinischen Fakultät der Universität Hamburg

vorgelegt von:

Janika Becker
aus Hamburg, Deutschland

Hamburg 2026

Betreuerin: Prof. Dr. Helen Blank

Mitglieder des Thesis-Komitees:
Mitglied: Prof. Dr. Christoph Korn
Mitglied: Prof. Dr. Tobias H. Donner

Gutachtende:
Gutachterin: Prof. Dr. Helen Blank
Gutachter: PD Dr. Patrick Bruns

Mitglieder der Prüfungskommission:
Vorsitz der Prüfungskommission: Prof. Dr. Helen Blank
Mitglied der Prüfungskommission: PD Dr. Patrick Bruns
Mitglied der Prüfungskommission: PD Dr. Regine Klinger

Datum der Disputation: 30.09.2026

Inhaltsverzeichnis / Table of contents

1. Introduction 4
2. Study 1: Sequences of Vowel Pairs in Pupillometry 10
   2.1 Introduction 10
   2.2 Methods 11
   2.3 Results 13
   2.4 Discussion 18
   2.5 Interlude 21
3. Study 2: Sequences of Syllable Triplets in EEG 23
   3.1 Introduction 23
   3.2 Materials & Methods 26
   3.3 Results 46
   3.4 Discussion 71
4. Discussion 87
5. Zusammenfassung / Summary 93
6. Hilfsmittel / Tools 95
7. Abkürzungsverzeichnis / Abbreviations 96
8. Anhang / Supplement 97
   8.1 Supplement ERP 97
   8.2 Supplement: Decoding 101
   8.3 Supplement: Reprint – Study 1 (Pupillometry) 105
9. Literaturverzeichnis / References 106
10. Danksagung 117
11. Eidesstattliche Versicherung 119

1. Introduction

Speech unfolds as a rapid, continuous stream of sound, yet listeners parse it effortlessly into words, phrases, and meaning (Cutler, 2012). Following a conversation in a noisy room, understanding a speaker with an unfamiliar accent, or filling in a word drowned out by background noise all draw on the same underlying ability: extracting statistical regularities from a temporal sequence of sensory input and using them to anticipate what is likely to come next. This capacity, broadly termed statistical learning (Turk-Browne, 2012), operates largely outside awareness (Fiser & Aslin, 2002), requires no explicit instruction or conscious effort (Fiser & Aslin, 2001; Saffran et al., 1996), applies to undifferentiated input using only spatial and temporal co-occurrence statistics (Fiser & Aslin, 2001), and yields knowledge about how specific elements of the environment relate to one another (Perruchet & Pacton, 2006). It applies not only to speech but to sequences of sounds, sights, and events more generally (Perruchet & Pacton, 2006). Because predictable events are processed faster and more efficiently than unpredictable ones (Pinto et al., 2015; Stein & Peelen, 2015), it is generally treated as a mechanism for genuine prediction — prospectively anticipating events before they occur (Turk-Browne, 2012) — rather than only a retrospective description of structure already encountered (Batterink et al., 2025).

That this capacity is already operative early in development was established in a seminal infant study (Saffran et al., 1996): after only two minutes of exposure to a continuous stream of nonsense syllables, 8-month-old infants distinguished three-syllable "words" — sequences with a within-word transitional probability of 1.0 — from equally frequent syllable triplets that merely spanned a word boundary, where the transitional probability dropped to 0.33. This sensitivity to the forward transitional probability between adjacent syllables, later shown to hold even for purely statistical structure stripped of every other cue (Aslin et al., 1998), is the computation that later work in adults builds on — and questions.

Two issues follow from this capacity. First, auditory input must be represented hierarchically: individual phonemes combine into syllables, syllables into words, and words into sequences from which meaning is extracted. Exactly which computational mechanism underlies this process — the pairwise transitional probability between adjacent elements highlighted above, or a more holistic, chunk-based representation of frequently co-occurring units — remains debated (Perruchet, 2019). Second, the world is noisy: spoken speech is frequently ambiguous or acoustically degraded, so listeners must resolve it using their own expectations rather than the acoustic signal alone (Davis & Johnsrude, 2007).

Perception, on this view, is not a passive readout of the sensory signal but the joint product of bottom-up sensory evidence and top-down expectation (de Lange et al., 2018; Summerfield & de Lange, 2014; Teufel & Fletcher, 2020), a claim that holds not only for single stimuli but, as the following chapters investigate, for structured sequences of stimuli over time.

But how exactly does the brain combine acoustic evidence with expectation derived from what has been learned, for instance from sequential regularities?

One influential answer is offered by predictive processing (Clark, 2013), a broad framework holding that the brain is fundamentally predictive: rather than passively registering sensory input, it continuously generates predictions about upcoming input on the basis of its current model of the world, and perception emerges from the interplay between these predictions and the sensory evidence that arrives (Friston, 2010).

Within this framework, predictive coding is the most influential proposal for how predictions and sensory evidence are combined in the brain. On this account, descending connections convey predictions, while ascending connections convey only the residual — the prediction error, or the part of the input the prediction failed to explain (Bastos et al., 2012; Friston, 2005; Rao & Ballard, 1999). Some accounts propose dedicated prediction and prediction-error units that communicate across levels of the neural hierarchy, continuously comparing sensory input against prior expectations (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999; Spratling, 2010). These expectations form an internal model of the world that the brain maintains and updates. Because minimizing prediction error is adaptive — a more accurate model supports better decisions and more efficient processing — the brain is proposed to revise the model whenever predictions and sensory evidence diverge, refining it to better anticipate its environment (Friston, 2010).

To this end, the brain is proposed to actively sample information from its surroundings that the model can use to predict and explain incoming sensory input more precisely (Friston, 2010; Parr et al., 2022). Theories diverge on which information is selected and how the sampling is implemented in the brain. Some — such as Sharpening accounts, often formulated in Bayesian terms (Knill & Pouget, 2004; Kok, Jehee, et al., 2012) — argue that the brain prioritizes the sensory input it expects, in order to arrive at a veridical version of the world despite the ambiguous or noisy signals inherent in sensory processing (Press et al., 2020). This view assumes that our world is, at least to some extent, predictable: we can draw on past experience to anticipate the future and to estimate how likely events are under given circumstances, so it is plausible that the brain would rely on these often-accurate predictions and prioritize predicted sensory events over unpredicted, potentially erroneous or noisy ones. On the neural level, these accounts hypothesize that the prioritization of expected information is achieved by an enhanced response in neurons that encode the expected stimulus, relative to other neurons (Yon et al., 2018), and that this is accompanied by enhanced perception of predicted stimuli or events (Stein & Peelen, 2015). Some studies even suggest that the brain anticipates expected stimuli by pre-activating some stimulus-specific neural template before the stimulus arrives, such that expected input is processed better or faster (Kok et al., 2017).

On the other hand, the events we could not anticipate might be the more important ones, since they can signal that something is not going according to plan and requires

immediate action: a car pulling out unexpectedly in front of you, or a sentence that abruptly turns toward an unexpected meaning. Both demand attention precisely because they were not predicted. This idea is related to cancellation theories, or Dampening accounts (Press et al., 2020): the brain prioritizes unexpected input because it is more informative than signals that were already anticipated and prepared for, as unexpected events may call for immediate action and, at the very least, signal that the internal model needs updating. On the neural level, this might be implemented through neurons encoding expected stimuli being actively suppressed relative to others, such that unexpected events and stimuli are perceived faster and more clearly (Richter et al., 2018).

While these two sets of theories seem to oppose each other, it has been proposed that both may in fact occur — potentially in temporal sequence, or at different levels of the neural hierarchy (Bastos et al., 2012; Press et al., 2020; Rao & Ballard, 1999).

A wide range of research investigates how the brain represents its environment, how it maintains and updates its internal model, and what circumstances affect this process. In perception research, these questions are typically approached by manipulating sensory input to create expectations — for instance through cue-outcome designs (den Ouden et al., 2025), or sequential dependencies (Fiser & Aslin, 2002) — that the brain is expected to pick up on, often automatically and without participants being aware of the process, in line with the idea of implicit statistical learning (Perruchet & Pacton, 2006; Turk-Browne, 2012). Factors such as whether the abstract rules are directly relevant to participants' behaviour — for instance, to fulfil task requirements — or whether they fall within or outside the scope of explicit attention have been extensively studied, yielding diverging evidence, and their precise role is still not fully understood (Alamia et al., 2019; Quirins et al., 2018; Richter & de Lange, 2019). The measure of learning varies, too: direct report by participants, or performance on a task, are the most obvious, intuitive measures. However, for processes that are more subtle and potentially outside participants' awareness, it is reasonable to measure them physiologically instead: via skin conductance, heart rate, gaze or pupil behaviour, or, most directly though at the highest methodological cost, brain activity itself.

As measuring brain activity is not always practical and requires substantial effort and infrastructure, pupillometry — gaze position and pupil dilation — offers a more accessible, if less direct, way of measuring processes that people might not be aware of.

Pupil diameter, for example, is modulated by activity in the noradrenergic locus coeruleus system (Aston-Jones & Cohen, 2005; Joshi & Gold, 2020), and has become a well-established physiological marker of arousal, uncertainty, and surprise-related processing (Joshi & Gold, 2020; Preuschoff et al., 2011), or, conversely, of past experience, familiarity, and expectation (Colizoli et al., 2026; Kafkas & Montaldi, 2015; Naber et al., 2013). Because such processes can operate independently of or beyond people's conscious awareness, and because they index how the brain samples its environment and updates its internal models, researchers have increasingly turned to

pupillometry to study them. In the visual domain, for example, multiple studies have shown that pupil dilation tracks the transition probabilities of a sequence even when observers cannot report the underlying structure, or that pupil and gaze behaviour indicate whether a stimulus or association is recognized as previously encountered, independent of whether that recognition can be explicitly reported (Alamia et al., 2019; Colizoli et al., 2018; Preuschoff et al., 2011; Schwiedrzik & Sudmann, 2020; Zénon, 2019). Whether the same holds in audition, and specifically for sequences of speech-like stimuli, motivated the first study presented in this dissertation.

The first study in this dissertation (Becker, Korn, et al., 2024) translated a visual pair-learning paradigm (Schwiedrzik & Sudmann, 2020) to the auditory domain, across two preregistered experiments. In Experiment 1 (n = 30), eight voice-vowel stimuli were presented in random order and then, after participants were informed of the manipulation, rearranged into fixed pairs. Experiment 2 (n = 31) removed the instruction, aligned the design with the original visual study, and added a familiarity task in which participants judged previously heard ("familiar") against newly recombined ("foil") pairs. Both experiments addressed the same question: does stimulus-evoked pupil dilation track the statistical (pair) structure of an auditory sequence online, as shown previously in vision? And if neither pupil nor behaviour shows such tracking during exposure, does an offline test afterwards nonetheless reveal that the structure was learned — in behaviour, in the pupil, or both?

In both experiments, pupil dilation tracked the presence of a sound, with clear spectral peaks and phase-locking at the presentation rate, but showed no signature of the pair structure: no peak at the pair rate and no condition difference in trial-locked diameter. Explicit judgements in the familiarity task were at chance. Yet pupil diameter in that same task separated the two pair types: dilation was larger to foil than to familiar pairs, following the objective pair category rather than participants' own judgements. The associations were therefore learned and physiologically detectable, even though neither online tracking nor explicit recognition revealed them.

One explanation may lie in the stimulus material: the pair structure was built from spoken vowels alone. Sequential dependencies, however, are extracted primarily from consonants rather than vowels — an asymmetry shown directly for statistical learning of word-like sequences (Bonatti et al., 2005; Nespor et al., 2003). It offers a plausible, though untested, limit on how learnable a vowel-only pair structure could have been.

The pupil thus did not track sequence structure online, unlike in vision and in other auditory statistical-learning paradigms (Aslin et al., 1998; Saffran et al., 1996). The consonant–vowel asymmetry discussed above is one plausible reason. The next study therefore asked: does the brain track sequence structure once it is explicitly learned and carried by syllabic rather than purely vocalic material? And how does it resolve acoustically ambiguous input given those sequence-based expectations?

The second study used a design well established in the statistical-learning literature (Aslin et al., 1998; Franco et al., 2015; Perruchet & Pacton, 2006; Saffran et al., 1996) to install sequential expectations rather than test whether they arise incidentally: 36

participants learned six trisyllabic pseudowords of consonant–vowel syllables, and word knowledge was verified in a battery of tasks. During EEG recording, participants first heard matched sequences without word structure, followed by sequences built from the learned triplets. Triplets rather than pairs make a critical contrast possible, because they separate a position that is merely predictable (the second) from one that also completes the word (the third). A subset of acoustically ambiguous syllables (Morphs) then reveals how much weight the expectation receives. The study is thus organized around two questions: how is auditory input integrated with sequence-based expectations, and what does that integration do to the representation of clear and ambiguous input?

The triplet structure revealed the functionally distinct roles of the syllable positions at both a behavioural and a neural level. A content-specific prior was instantiated only at the medial, second position — predictable from the first syllable, yet not word-final, and thus under the highest sequential constraint. There, clear syllables (Standards) were suppressed below baseline, stimulus decoding was enhanced, and pre-stimulus decoding already carried identity-specific information. Behaviourally, however, the picture was mixed: ambiguous Morphs were categorized as the expected syllable, but clear Standards were identified no better than in unpredictable sequences. The first position, following a word boundary and thus unpredictable, showed none of these neural signatures. Standard syllables there were identified above baseline, while Morph categorization followed the acoustics. The word-final position reversed the direction: responses to both Standard and Morph syllables were enhanced, consistent with completion rather than anticipation, as no pre-stimulus information was present there. Across positions, reliance on the prior grew linearly on ambiguous trials, whereas identification accuracy on Standard trials followed a U-shaped profile with a word-medial dip — the two measures track different things: accuracy where the evidence is clear, and the weight of the prior where it is not.

On the second question, the neural evidence is mixed: Morphs could not be reliably decoded, possibly because responses to two competing syllable identities cancel out. For the decodable Standard stimuli, the pattern argues against Dampening: at the medial position, where suppression was strongest, syllable identity remained decodable — more so than at the first position — and stimulus-specific information was present before onset. This favours Sharpening, though a strict version of that account would also predict a behavioural advantage for expected stimuli and a correlation between suppression and decoding strength. The behavioural advantage was absent; the correlation was not found either, but the measures entering it proved too unreliable for that absent relationship to count as evidence (Section 3.2.9). The findings also remain compatible with a general predictive-coding architecture, while running counter to the specific prediction-error coding accounts that predict the reverse decoding profile (Blank & Davis, 2016; Sohoglu et al., 2024), and add to a literature in which Sharpening and Dampening look less like competing accounts than like descriptions applying at different representational levels or points in time.

Together, these two studies ask a common question — how the brain integrates sequential expectation with incoming sensory evidence, and what that integration does to the underlying representation — using two different measures, paradigms, and degrees of explicitness. Chapter 2 presents the pupillometry study in full; Chapter 3 presents the EEG study; the Discussion returns to what the two studies, read together, suggest about the conditions under which learned structure becomes visible, what happens to the representation of a predictable stimulus — Sharpening or Dampening— why the neural and behavioural measures dissociate, and what follows for how statistical learning is measured.

2. Study 1: Sequences of Vowel Pairs in Pupillometry

Source publication. This chapter presents a condensed account of: Becker, J., Korn, C. W., & Blank, H. (2024). Pupil diameter as an indicator of sound pair familiarity after statistically structured auditory sequence. *Scientific Reports*, 14, 8739. https://doi.org/10.1038/s41598-024-59302-1

All data, experimental designs, statistical results, and interpretations reported in this chapter originate from this publication. The text below has been substantially condensed and independently reworded relative to the published article to avoid textual self-overlap; numerical results are reproduced because they constitute factual findings rather than expressive text. Section 2.3.4 reports additional, previously unpublished analyses conducted specifically for this dissertation to further probe the temporal dynamics of the familiarity-task effect and to empirically evaluate alternative explanations for the sequence-locked findings; Section 2.4 integrates these into a critical discussion of the original study.

2.1 Introduction

Perceptual systems continuously extract statistical regularities from the temporal structure of incoming sensory input, a process broadly termed statistical learning. This extraction appears to occur quickly and largely outside awareness, allowing an observer to form expectations about upcoming events. Violation of such expectations is experienced as surprise and has been linked to noradrenergic signalling from the locus coeruleus (LC-NE system) (Joshi et al., 2016; Murphy et al., 2011). Because pupil diameter is modulated by LC-NE activity, stimulus-evoked pupil dilation is widely used as a peripheral marker of arousal, uncertainty, and surprise-related processing (Bradley et al., 2008; Colizoli et al., 2026; Joshi & Gold, 2020).

In the visual domain, several studies showed that pupil dilation tracks the transition probabilities of sequentially presented stimuli even when observers cannot report the underlying structure (Alamia et al., 2019; Binda et al., 2025; Schwiedrzik & Sudmann, 2020). Alamia et al. (2019) found larger dilations to rare than to frequent letter transitions in a probabilistic sequence, while Schwiedrzik and Sudmann (2020) reported that arranging 18 face identities into fixed sequential pairs produced a smaller pupil response to the first than to the second stimulus of a pair. Because pupil responses to auditory stimuli have been shown to be similarly sensitive to stimulus probability (Becker, Viertler, et al., 2024; Friedman et al., 1973; Qiyuan et al., 1985; Reinhard & Lachnit, 2002) or surprise and prediction-error related effects (Kloosterman et al., 2015; Lavin et al., 2014; Preuschoff et al., 2011; Quirins et al., 2018), our study translated the face-pairing paradigm of Schwiedrzik and Sudmann to the auditory domain, replacing face identity/orientation with recordings of spoken vowels produced by different speakers.

Two preregistered experiments addressed this question. In Experiment 1 (*n* = 30), eight voice–vowel stimuli were presented at 0.5 Hz, first in random order and then — after participants were explicitly informed of the manipulation — rearranged into fixed pairs (0.25 Hz pair rate) (Figure 2.1, Panel A). Experiment 2 (*n* = 31) removed the explicit instruction and scaled the design closer to the original visual study (18 stimuli, 1 Hz presentation rate, 0.5 Hz pair rate), adding a post-sequence familiarity task in which participants judged previously heard ('familiar') versus recombined ('foil') pairs.

The central question in both experiments was whether stimulus-evoked pupil dilation during the structured sequence would track within- versus between-pair transition probability, as in visual studies. In Experiment 2, we additionally asked whether pupil responses during an explicit familiarity judgement would reveal implicit pair learning even without pair-related pupil dilation during sequence presentation.

2.2 Methods

Both experiments recorded right-eye pupil diameter and gaze position at 1000 Hz (EyeLink 1000) while participants, seated with a chin rest at 50 cm from the screen, listened over headphones to sequences of spoken vowels and fixated a central cross. Both experiments were preregistered and approved by the Ethics Committee of the Hamburg Chamber of Physicians. Table 2.1 summarizes the main design parameters that differed between experiments; all other procedures were identical.

Parameter	**Experiment 1 (*n* = 30)**	**Experiment 2 (*n* = 31)**
Stimulus set	8 stimuli per participant, drawn from a pool of 8 voices × 4 vowels (/a e i o/)	18 stimuli per participant, drawn from a pool of 18 voices × 3 vowels (/a i o/)
Presentation rate	0.5 Hz (pair rate 0.25 Hz)	1 Hz (pair rate 0.5 Hz)
Trial / block structure	4 blocks × 76 trials per condition	4 blocks × 81 trials per condition
Concurrent task(s)	Auditory 1-back + visual fixation-change task	Auditory 1-back only
Instructed about pair structure?	Yes, before the structured condition	No
Post-sequence familiarity task	—	18 trials (9 familiar, 9 foil pairs)
High-pass filter cut-off	0.1 Hz	0.25 Hz

Table 2.1. Design parameters that differed between Experiment 1 and Experiment 2.

In both experiments, the random condition always preceded the structured condition, to prevent pair-related expectations from contaminating the random baseline. Final sample sizes (30 and 31, respectively) reflect preregistered exclusion criteria (excessive missing pupil data, poor task performance); in Experiment 1, two further data sets were excluded outside these criteria, because of a technical failure during recording and visibly poor raw-signal quality. The target sample size for Experiment 1 was derived from an a-priori power analysis based on the pair-rate spectral-power

effect reported by Schwiedrzik and Sudmann (2020); the sample size for Experiment 2 followed Experiment 1 and the sample size of that study.

2.2.1 Preprocessing

Pupil traces were cleaned of gaze-contingent artefacts, and gaps from blinks or outlier rejection were closed using shape-preserving (PCHIP) interpolation. A 5 Hz low-pass filter and a high-pass filter (0.1 Hz in Experiment 1, 0.25 Hz in Experiment 2, replacing a preregistered per-block baseline correction that proved infeasible because many participants closed their eyes during inter-block breaks) were applied before down-sampling to 500 Hz and within-participant z-standardization. Preprocessing used the PsPM toolbox (Bach, 2020).

2.2.2 Statistical analysis of the auditory sequences

Three complementary analyses tested whether the pupil response reflected the pair structure of the structured condition.
(1) Welch power spectra (computed per block and averaged within participant and condition) were tested for a spectral peak at the stimulus-presentation rate in both conditions and, additionally, at the pair-presentation rate in the structured condition, by comparing peak power to the mean of four neighbouring frequency bins.
(2) Inter-trial phase clustering (ITPC), computed from 16-s pseudo-trials, quantified phase-locking to the stimulus- and pair-presentation rate.
(3) Trial-locked maximum and minimum pupil diameter (identified from the trial-averaged waveform within condition- and experiment-specific time windows) were submitted to repeated-measures analyses of variance (rmANOVAs) crossing condition (random, structured) with within-trial stimulus or inter-stimulus-interval position, and, for the structured condition only, crossing block (first vs. last) with within-pair position, to test for a change in the pupil response as participants gained exposure to the pairs. Effect sizes are reported as partial η^2 (rmANOVAs) or Cohen's *dz* (*t*-tests) (Cohen, 1988), alongside Bayes Factors BF_{10} quantifying evidence for the alternative relative to the null hypothesis.

2.2.3 Familiarity task (Experiment 2)

After listening to the random and the structured sequences, participants judged 18 pairs (9 previously heard 'familiar' pairs, 9 newly recombined 'foil' pairs) as familiar or unfamiliar via key press. Behavioural accuracy and reaction time were compared between pair types. Pupil diameter was baseline-corrected to the 1-s window preceding the onset of the pair's second stimulus, and maximum diameter in the following 2-s window was compared between familiar and foil pairs (paired *t*-test) and modelled with a linear mixed-effects model predicting trial-wise maximum pupil diameter from the true pair category, the participant's response, and reaction time, with a random intercept for participant.
Section 2.3.4 additionally reports the methods and results of supplementary analyses conducted for this dissertation, including a time-resolved reanalysis of the familiarity-task effect described above.

2.3 Results

2.3.1 Pupil tracks sound presentation, not pair structure

In both experiments, the power spectrum of the pupil signal showed a clear, highly significant peak at the stimulus-presentation rate in both the random and the structured condition. No corresponding peak emerged at the pair-presentation rate in the structured condition, spectral power at the pair rate did not differ between conditions, and the expected condition × frequency interaction was not significant in either experiment (Table 2.2; Figure 2.1, Panels C, E). ITPC was likewise far higher at the stimulus rate than at the pair rate in both experiments, confirming strong phase-locking to individual sound onsets, but ITPC at the pair rate did not differ between the random and structured conditions.

Test	Experiment 1 (n = 30)	Experiment 2 (n = 31)
Spectral peak at stimulus rate (both conditions)	$t(29) = 8.1–10.3$, $p < .001$, $BF_{10} > 100$	$t(30) = 6.5–9.3$, $p < .001$, $BF_{10} > 100$
Spectral peak at pair rate (structured)	$t(29) = 1.19$, $p = .122$, $BF_{10} = 0.65$	$t(30) = 0.16$, $p = .436$, $BF_{10} = 0.22$
Pair-rate power: structured vs. random	$t(29) = −0.12$, $p = .546$, $BF_{10} = 0.18$	$t(30) = −0.39$, $p = .649$, $BF_{10} = 0.14$
rmANOVA power × frequency × condition (interaction)	$F(1,29) = 4.09$, $p = .052$, $\eta p^2 = .124$	$F(1,30) = 0.12$, $p = .733$, $\eta p^2 = .004$
ITPC: stimulus rate vs. pair rate	$t(29) = 11.7$, $p < .001$, $BF_{10} > 100$	$t(30) = 7.23$, $p < .001$, $BF_{10} > 100$
ITPC pair rate: structured vs. random	$t(29) = −0.65$, $p = .521$, $BF_{10} = 0.23$	$t(30) = 0.72$, $p = .477$, $BF_{10} = 0.24$

Table 2.2. Spectral-power and ITPC results, both experiments (as reported in Becker, Korn, et al., 2024). Tests on spectral power were one-tailed, as preregistered; ITPC comparisons were two-tailed.

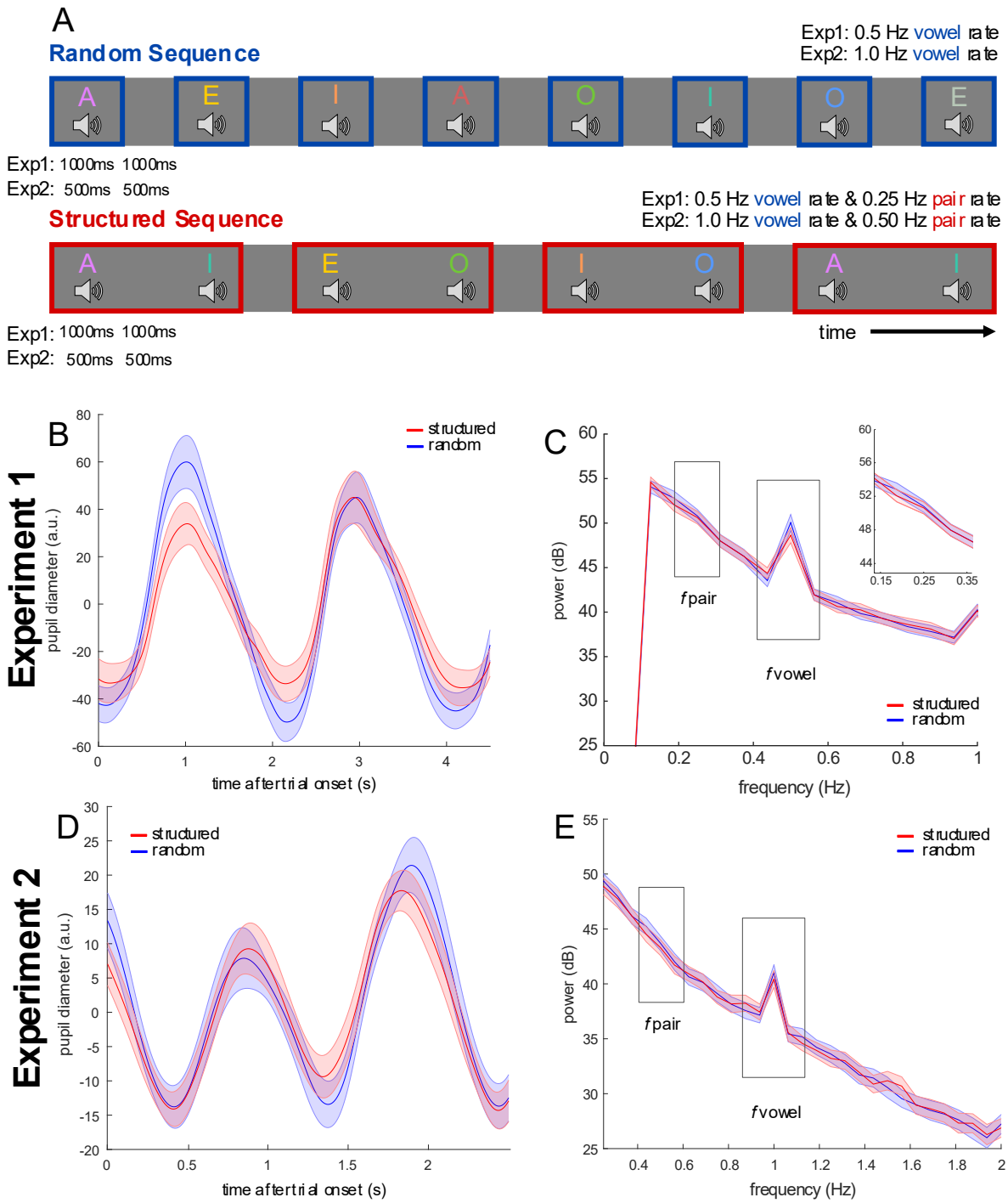


Figure 2.1. (A) Sequence Structure and transition probabilities. Random sequence (top, blue): randomized stream of 8 / 18 sounds for Exp1 / Exp2 (example sequence shown for Exp2), uniform low transition probability (TP = 1/7, 1/17). Structured sequence (bottom, red): the same sounds recombined into fixed pairs (between-pair TP = 1/3, 1/8), with within-pair transitions fully predictable (TP = 1). Experiment 1: (B) Trial-locked pupil diameter and (C) Frequency power spectrum, no pair-frequency peak; Experiment 2: (D) Trial-locked pupil diameter and (E) Frequency power spectrum, no pair-frequency peak. Reproduced from the raw data underlying Figs. 1B–C and 2B–C of Becker, Korn, et al. (2024).

2.3.2 Trial-locked maximum and minimum pupil diameter

Maximum pupil diameter after a stimulus did not depend on condition in either experiment, and did not depend on within-trial stimulus position in Experiment 1; the condition × position interaction was likewise not significant, although it was closest to threshold in Experiment 1 (p = .081; Table 2.3; Figure 2.1, Panel B). In Experiment 2, maximum pupil diameter was consistently larger after the second than the first stimulus of a trial, in both the random and structured condition — an unexpected pattern of pupil cycling at half the stimulus-presentation rate that we interpreted cautiously, since it did not correspond to a peak in the frequency-power spectrum (Figure 2.1, Panel D). This is not a contradiction, but it does bound the interpretation: the trial-locked maxima rest on a single summary value per trial, averaged after exclusion of task trials, whereas the spectrum is computed continuously over the entire block and would capture any genuine periodicity at half the presentation rate. The absence of a corresponding spectral peak therefore argues against a sustained cycling of the pupil signal.

Minimum pupil diameter during the inter-stimulus interval showed no dependence on within-trial position in either experiment, but in Experiment 1 it was smaller (more negative) in the random than in the structured condition, $F(1,29) = 6.45$, p = .017; this effect was not supported by the corresponding Bayes Factor (BF_{10} = 0.76), and no such difference emerged in Experiment 2. Comparing the first and last block of the structured condition revealed no evidence that the pupil response to either stimulus of a pair changed with increasing exposure, in either experiment.

Test	Experiment 1 (n = 30)	Experiment 2 (n = 31)
Max. pupil: condition × stimulus position	$F(1,29) = 3.27$, p = .081, ηp^2 = .101	$F(1,30) = 0.30$, p = .589, ηp^2 = .010
Max. pupil: main effect of stimulus position	$F(1,29) = 0.45$, p = .509, ηp^2 = .015	$F(1,30) = 9.10$, p = **.005**, ηp^2 = .233
Min. pupil (ISI): main effect of condition	$F(1,29) = 6.45$, p = .017, ηp^2 = .182	$F(1,30) = 0.33$, p = .568, ηp^2 = .011
Learning-over-time: block × within-pair position	$F(1,20) = 2.15$, p = .158, ηp^2 = .097 (n=21)	$F(1,23) = 1.90$, p = .181, ηp^2 = .076 (n=24)

Table 2.3. Selected trial-locked pupil-diameter results (full ANOVA tables in the original publication).

2.3.3 Familiarity task: pupil diameter distinguishes familiar from foil pairs

Behavioural performance in the familiarity task gave no indication that participants explicitly learned the pairs: categorization accuracy (M = 52 %) did not differ from chance, $t(30) = 1.45$, p = .157, and reaction times for familiar and foil pairs did not differ, $t(30) = -0.93$, p = .362. Pupil diameter, however, distinguished the two pair types: maximum pupil diameter after the second stimulus of a pair was larger for foil than for familiar pairs, $t(30) = 2.74$, p = .010, dz = 0.49, BF_{10} = 4.32 (foil M = 0.958 vs. familiar M = 0.728 a.u.; Figure 2.2). A linear mixed-effects model confirmed that the true pair category, but not the participant's response or reaction time, significantly predicted trial-wise maximum pupil diameter (true category β = 0.198, $t(554) = 2.19$, p = .029),

indicating that the pupil response tracked the objective rather than the subjectively reported pair category. The magnitude of this pupil dissociation did not correlate with behavioural accuracy across participants, $r(29) = .29$, $p = .11$.

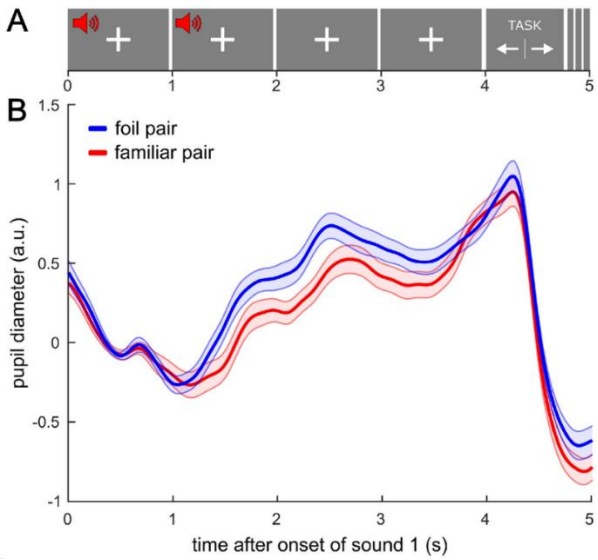


Figure 2.2. Familiarity Task: Maximum baseline-corrected pupil diameter, familiar vs. foil pairs (mean ± SEM, n = 31). Original Fig. 3 from (Becker, Korn, et al., 2024).

2.3.4 Supplementary analyses

The analyses below were not part of the original publication. They were conducted for this dissertation using the raw pupillometry, onset and behavioural data underlying Becker, Korn, et al. (2024) and the analysis pipeline described in Section 2.2.1, in order to, firstly, resolve the familiarity-task effect at finer temporal resolution, and, secondly, empirically evaluate alternative explanations discussed in Section 2.4 rather than leaving them as unquantified possibilities.

2.3.4.1 Familiarity effect: cluster-based permutation test

Method. To examine the time-course of the familiarity-task pupil effect at finer resolution than the single time-window comparison in Section 2.3.3, we applied a nonparametric cluster-based permutation test (Maris & Oostenveld, 2007) to the same 0–2000-ms post-stimulus-2 window used for the original maximum-pupil analysis. To this end, for each of the 31 participants, baseline-corrected pupil waveforms were averaged across the nine trials of each pair type, and a paired *t*-test (foil minus familiar) was computed at every sample of the resulting 500-Hz time series. Contiguous samples exceeding a two-tailed cluster-forming threshold of $|t| > 2.042$ ($\alpha = .05$, $df = 30$) were grouped into clusters, and the summed *t*-value within each cluster ('cluster mass') served as the test statistic. The null distribution of the maximum cluster mass was estimated from 10000 sign-flip permutations (randomly inverting the sign of each participant's foil-minus-familiar difference wave), and a cluster-level *p*-value was computed as the proportion of permutations whose maximum cluster mass equalled or exceeded the mass of each observed cluster.

Result. One significant cluster emerged: pupil diameter was larger for foil than familiar pairs from 1017 to 1649 ms after the onset of the second stimulus (summed cluster *t*-mass = 863.0, $p_{cluster}$ = .016). Two further candidate clusters, from 465–695 ms (p = .068) and 1949–1998 ms (p = .123), did not reach significance. The significant window falls in the second half of the pupil dilation response rather than at its earliest rising phase, suggesting a comparatively late, sustained process rather than an immediate low-level acoustic response to the second stimulus (Hoeks & Levelt, 1993; Wierda et al., 2012). This localizes the originally reported familiarity-task effect in time.

2.3.4.2 Balance of trial count across conditions

Method. The trial-locked analyses in Section 2.3.2 excluded any trial containing, or immediately following, a task event (on average ~36 % of trials in Experiment 1, ~22 % in Experiment 2), after additional exclusion of on average 1.4 (Exp. 1) / 1.0 (Exp. 2) data blocks per participant due to artefact quality. Because exclusion is tied to the position of task events within the sequence, a potential imbalance in the number of retained trials between the random and structured condition could introduce unequal measurement noise that mimics or masks a genuine condition effect. We tested this directly by comparing, per participant, the number of trials retained for the trial-locked analyses between conditions.

Result. The number of retained trials did not differ significantly between conditions in either experiment (Experiment 1: M = 165.0 vs. 156.9 trials for random vs. structured, $t(29) = -0.80$, p = .433, BF_{10} = 0.26; Experiment 2: M = 227.8 vs. 216.1, $t(30) = -1.23$, p = .229, BF_{10} = 0.38). This makes an unequal-trial-count / unequal-noise account of the condition comparisons in Section 2.3.2 unlikely.

2.3.4.3 Voice salience variability and individual sensitivity to pair structure

Method. Naturalistic, multi-speaker voice stimuli can differ in low-level acoustic salience independently of the pair structure (Liao et al., 2016). If such salience differences introduce noise into the pupil signal, and if participants vary in their sensitivity to this salience, the resulting pupil responses may be differentially modulated by noise across individuals. To account for this, we computed a salience-variability index: for each participant, we extracted the phasic pupil amplitude for every individual occurrence of every voice in the random condition only (where transition probability is uniform across stimuli, so any systematic between-voice differences reflect stimulus-specific salience rather than sequence structure) and then calculated the standard deviation of the resulting per-voice mean amplitudes.

This index was correlated, across participants, with the absolute magnitude of the condition × stimulus-position interaction contrast on maximum pupil diameter (i.e., the per-participant scalar equivalent of the interaction term reported in Table 2.3).

Result. This index did not correlate significantly with the interaction-contrast magnitude in either experiment (Experiment 1: $r(28)$ = .26, p = .158; Experiment 2: $r(29)$ = .13, p = .491). Altogether, the data do not indicate that naturalistic voice-salience heterogeneity is a major driver of individual variability in sensitivity to the pair structure.

2.4 Discussion

Across two independent, preregistered experiments, stimulus-evoked pupil dilation during an ongoing auditory sequence tracked the presence of a sound (as indicated by strong, reliable spectral peaks and phase-locking at the stimulus-presentation rate) but not the statistical pair structure of that sequence: neither spectral power, ITPC, nor trial-locked maximum/minimum pupil diameter distinguished the random from the structured condition in the way we predicted from the equivalent visual paradigm (Schwiedrzik & Sudmann, 2020). This null result contrasts with prior demonstrations of probability-sensitive pupil responses in audition using tones or musical sequences (Basgol et al., 2026; Bianco et al., 2020; Milne et al., 2021), so sensory modality per se is unlikely to be the decisive factor.

Two stimulus-related factors were considered as contributors to the null result; only the first remains plausible.
Firstly, the stimuli varied only in vowel identity superimposed on naturalistic, multi-speaker voice recordings. The psycholinguistic literature on the 'consonant–vowel asymmetry' suggests that sequential statistical dependencies are extracted primarily from consonants rather than vowels (Bonatti et al., 2005; Toro et al., 2008), which may have limited how strongly the pair structure was tracked online. The familiarity task shows that the pairs were learned, so the asymmetry cannot explain a failure of learning as such, but only a weaker moment-to-moment signal; and even that reading remains untested directly (e.g., against a matched consonant–vowel version of the paradigm), so this is only a plausible, not established, explanation. A second candidate, namely that acoustic salience differences across the naturalistic multi-speaker stimuli added noise to the pupil signal, was addressed in Section 2.3.4.3 and did not receive support: individual differences in voice-salience variability were unrelated to individual sensitivity to the pair structure. This weakens, without fully excluding, the salience-noise account and shifts explanatory weight back towards the consonant–vowel and task-relevance factors discussed below.

Another important factor to consider for the absent effect is that condition order (random vs. structured) is fully confounded with time-on-task in both experiments: the random condition always preceded the structured condition, by design, to prevent pair-related expectations from contaminating the random baseline. This fixed order could have affected our results in multiple ways. Firstly, we cannot rule out that participants experienced fatigue, habituation, or other slow physiological drifts in the second half of the experiment which would affect pupil dilation (Aston-Jones & Cohen, 2005; Mathôt, 2018). Following data exclusion based on noisiness, this could have resulted in an unequal count of retained trials contributing to the statistical conditions, which Section 2.3.4.2 renders unlikely. Alternatively, potential changes of mental or physical state over time could have masked the pupillary pair-tracking effect we were looking for or could have contributed to other condition differences that were observed, such as the smaller minimum pupil diameter in the random than the structured condition in Experiment 1. Another candidate for an explanation is offered by Iorio et al. (2026) who

found reduced statistical learning during a structured sequence when their participants were exposed to a random sequence first, relatively to the other way around. This is addressed in more detail in the discussion in chapter 4.
Overall, because order and condition are perfectly collinear by design, we cannot resolve ambiguity with the present dataset. This would require a counterbalanced or order-randomized follow-up including the trade-off that participants might then form pair-related expectations before the random block. Alternatively, to avoid or at least attenuate structure-specific expectations in such a design, one could consider using different stimulus sets across order conditions, depending on the specific research question.

Finally, the concurrent 1-back repetition-detection task — present in both experiments and paired in Experiment 1 with an additional visual task — kept participants engaged with the basic alternation structure of the sequence but may have drawn attention away from the higher-order pair structure, which was task-irrelevant throughout both experiments; this remains a plausible, if untested, contributor alongside the factors above.

Despite the absence of a sequence-locked pupil signature of pair structure, the post-sequence familiarity task revealed that participants had, in fact, implicitly learned the pair associations: pupil diameter was reliably larger for newly recombined foil pairs than for previously heard familiar pairs, and a linear mixed-effects model showed that this response tracked the objective pair category rather than participants' (at-chance) explicit judgements. This dissociation between null behavioural discrimination and a reliable, moderately strong physiological signature is consistent with statistical learning proceeding, at least partly, outside conscious access, and adds to the literature in which pupillometry outperforms explicit report as a measure of implicit association learning (Alamia et al., 2019; Gomes et al., 2015; Kafkas & Montaldi, 2015; Schwiedrzik & Sudmann, 2020). At the same time, it is worth noting explicitly that the 2-s analysis window used for this effect overlaps with the period during which participants know a familiarity judgement is imminent (the response prompt appeared 3 s after the second stimulus). Thus, the present design cannot fully separate a perceptual-surprise account of the foil-vs-familiar difference from a decision-uncertainty account, since foil pairs are plausibly both more surprising and harder to judge with confidence, and both accounts predict the same direction of effect (Colizoli et al., 2018; Nassar et al., 2012; Preuschoff et al., 2011; Urai et al., 2017).
The cluster-based permutation analysis (Section 2.3.4.1) narrows the space of explanations, although it cannot adjudicate between these two accounts, since both predict a late effect: the significant familiar-vs-foil difference was confined to a roughly 600-ms cluster centred just after the peak of the dilation response (1017–1649 ms post-stimulus), rather than spanning the entire 2-s window or concentrating in its earliest rising phase, where a purely bottom-up acoustic-novelty response would be expected to dominate. A difference concentrated in this later, more sustained portion of the response therefore speaks rather against a purely bottom-up acoustic-novelty account and locates the effect among the post-perceptual processes, whether

recognition- or decision-related (Colizoli et al., 2018; de Gee et al., 2014). Separating stimulus-driven from decision-related pupil dilation could be achieved by jittering the response-prompt onset, or by combining pupillometry with EEG to track the two processes independently.

Taken together, the original findings, extended by the supplementary analyses in Section 2.3.4, support a moderately cautious conclusion: pupil diameter clearly indexed the behavioural relevance of a learned auditory association once probed explicitly (the familiarity task), but did not detectably respond to transition-probability structure that was task-irrelevant during the exposure sequence itself. The evidence for that absence, while consistent across analyses and supported by Bayes Factors that range from inconclusive to moderate evidence for the null, remains partially confounded with time-on-task. This boundary condition is important for the broader interpretation of pupillometric statistical-learning effects across sensory modalities and should inform the design of future auditory sequence-learning studies, for instance by counterbalancing condition order where feasible, considering stimulus type (tones vs. naturalistic speech, vowels vs. consonants), and using time-resolved analysis approaches, such as the cluster-based method demonstrated here, that are sensitive to effects that are temporally localized rather than uniformly distributed across an analysis window.

2.5 Interlude

The pupil tracked the presence of a sound but not the sequence's statistical structure while that structure was merely being listened to — a finding that contrasts both with the visual study that inspired the design (Schwiedrzik & Sudmann, 2020) and with other auditory statistical-learning findings (Aslin et al., 1998; Saffran et al., 1996).
Two candidate explanations for the absence of pupillary pair-structure tracking seemed most plausible, alongside the order confound between the random and structured condition described in Section 2.4. One is the consonant–vowel asymmetry noted earlier: a sequence built from vowels alone may simply carry a weaker statistical signal than one that also incorporates consonants (Bonatti et al., 2005; Nespor et al., 2003). The other is that the pair structure was entirely task-irrelevant during exposure: participants performed a concurrent 1-back task that kept them engaged with individual sounds but gave them no reason to attend to how those sounds were grouped, which may have limited how much the sequence's structure was actually used, regardless of whether or not it could in principle have been learned.

This raises the questions that the second study of this dissertation was designed to answer: does the brain track sequence structure once predictability is explicitly established, with stimuli suited better to carrying that structure and a task that encouraged participants to actually use it? And, if so, how does the brain resolve input that is acoustically ambiguous, given the expectations that a sequence structure has built up?
Triplet-based artificial-word paradigms as used in the second study have a long history in the statistical-learning literature (Peña et al., 2002; Saffran et al., 1996), and a growing number of neuroscientific studies have used multivariate pattern analysis or time-resolved decoding to ask what happens to the neural representation of a stimulus once it becomes predictable — in audition (Blank & Davis, 2016; Demarchi et al., 2019; Schneider & Blank, 2026; Sohoglu & Davis, 2020; Ufer & Blank, 2024) as well as in vision (Alink et al., 2026; Garlichs & Blank, 2024; Kok et al., 2017; McDermott et al., 2026; Rittershofer et al., 2026). The high temporal resolution of EEG and MEG allows the pre- and post-stimulus dynamics to be tracked separately, which pupillometry, by design, averages over. This is also where the pupil study's own discussion pointed: separating a stimulus-driven pupil response from a later, decision-related contribution requires a method that can resolve the two in time. The second study takes up that suggestion, though with a more specific question in mind.

Taking both candidate explanations for the lack of pupillary tracking into account as mentioned above, the second study made two corresponding design changes. First, it replaced vowel-only pairs with consonant–vowel syllable triplets, which is a considerably better established stimulus format from the statistical-learning literature (Aslin et al., 1998; Perruchet & Pacton, 2006; Saffran et al., 1996). Second, rather than leaving sequence structure learning incidental combined with an unrelated task, word knowledge was made explicit from the outset and independently verified through a battery of learning tasks: 36 participants were directly taught six trisyllabic

pseudowords built from consonant–vowel syllables, so that the sequential expectations under investigation were guaranteed to actually be in place, rather than a matter of whether incidental exposure alone had been sufficient to establish them. The learned syllable triplets were then presented, alongside randomized and structurally identical but unlearned syllable sequences, during EEG recording.

Relative to the pairwise pupil design, the trisyllabic structure was a further, deliberate change. With three syllables rather than two, a learned word separates two properties that a pair of stimuli necessarily confounds, because the second element is at once both the first predictable one and the last one of that unit. Additionally, the new design featured stimulus ambiguity as a means to investigate expectation effects from yet another angle: if listeners could identify every syllable from the acoustic signal alone, they would have little reason to draw on their sequence knowledge at all, and the weight they give to their expectation would rather be a theoretical possibility than a likely observation. A subset of syllables was therefore constructed as an acoustic blend between two learned alternatives ("Morphs"), so that the effect of expectations could be read out trial by trial. The following chapter introduces both features, among others, in the terms used throughout.

3. Study 2: Sequences of Syllable Triplets in EEG

3.1 Introduction

Perception is not a readout of the sensory signal alone. What listeners hear depends jointly on the input and on what the context makes likely (Brodbeck et al., 2022; Clayards et al., 2008; Norris et al., 2003; Schneider & Blank, 2026). Speech is the paradigmatic case: syllables arrive in a rapid, partly redundant stream, and the statistics of a language make each element more or less predictable from those that precede it. The brain has been shown to track the regularity of such an auditory stream closely enough to approximate even an ideal observer under certain conditions (Barascud et al., 2016). The general introduction set out why this matters for perception; the present chapter asks a narrower question: given that listeners have learned a sequence, and that predictions are instantiated within it, how does this affect the neural representation of sounds — predicted and unpredicted alike?

Theories of predictive processing (Clark, 2013; Friston, 2005, 2010) would posit that the identity of each sound is estimated by combining the acoustic evidence with an expectation derived from the learned sequence, each weighted by how informative it is. This leads to two hypotheses: Firstly, predictions should matter most where the sensory evidence is least informative, that is, for acoustically ambiguous input. And, secondly, the influence of expectations should follow the sequential constraint on the current element. It should therefore be strongest where the context narrowly determines the next element, and absent where the context leaves it open. Because the second and third syllable of a word are equally determined by the preceding context, both were expected to show expectation-related effects; the third position, however, is additionally word-final, so any divergence between the two would isolate the contribution of word completion from that of sequential constraint alone.

Predictive coding is a highly influential proposal for how the combination of predictions and sensory input is carried out on the neural level (Bastos et al., 2012; Friston, 2005; Rao & Ballard, 1999; Srinivasan et al., 1982). Canonical hierarchical formulations posit a division of labour between two directions of message passing: descending connections convey predictions about the expected input, while ascending connections convey only the residual, the part of the input that the prediction failed to explain. These two directions may correspond to two distinct neuronal populations, one transmitting predictions and one signalling the deviation from them; direct evidence for such an implementation, however, remains limited (Walsh et al., 2020). If the neural system works according to predictive coding, feedforward signals are not a copy of the sensory stimulus but an error term, which should shrink as predictions about that stimulus improve.

A signature often cited in support of predictive coding is expectation suppression: a reduced sensory response to expected relative to unexpected stimuli, reported across modalities and recording techniques (de Lange et al., 2018; Feuerriegel et al., 2021; Richter et al., 2018; Summerfield et al., 2008; Todorovic & de Lange, 2012).

Two caveats limit what a reduced response to expected stimuli can show. First, comparing expected against unexpected stimuli cannot say which of the two moved: the expected response may be suppressed, or the unexpected one enhanced by surprise. Deciding between them requires a baseline condition that differs from the other conditions in predictability alone, and constructing one is a recognized difficulty in this literature (Feuerriegel et al., 2021). Second, a smaller amplitude has explanations that do not involve prediction error at all: stimulus-specific adaptation and reduced attentional gain make the same prediction (Feuerriegel et al., 2021; Walsh et al., 2020), and predictive coding is in any case only one of several architectures that could implement the computation described above (Aitchison & Lengyel, 2017). A design meant to test it therefore has to hold repetition and physical similarity constant across conditions.

What does this amplitude reduction imply for population coding? A univariate event-related potential (ERP) sums across distinct populations, so an amplitude change alone cannot reveal which population changed. The informative measure is the representation itself, and this is where two accounts diverge: Sharpening holds that expectation suppresses units tuned away from the expected feature, leaving a smaller but more selective, more decodable representation (Kok, Jehee, et al., 2012). Dampening holds the reverse: units encoding the expected component are themselves suppressed, so unexpected signals stand out but the overall representation becomes less informative and less decodable (Press et al., 2020; Richter et al., 2018).
At the computational level the same contrast is framed as Sharpened Signals versus Prediction Errors, that is, whether what is passed forward is the sensory input weighted by the prior or the residual that remains once the prediction has been subtracted (Blank & Davis, 2016). For clear and correctly predicted input the two framings converge, since prediction-error coding and Dampening both predict a smaller and less informative representation. They come apart once the input is ambiguous or the prediction is violated: only prediction-error coding predicts a reversal, with representations of expected input becoming less informative as that input grows clearer, and more informative when the input mismatches a strong prediction — a signature reported for degraded spoken words (Blank & Davis, 2016) and for syllable predictions within spoken words (Sohoglu et al., 2024). Ambiguous input is therefore the condition under which the two framings could in principle be told apart, and the Morph condition of the present study was designed to provide it.
What distinguishes these accounts empirically overall is therefore not the amplitude but the joint pattern of amplitude and decodability, which is also in the focus of the present study.

Both Sharpening and Dampening are functionally defensible, which is part of why the debate persists. Prioritizing expected input as in Sharpening is a way of arriving at a veridical percept despite noise, on the assumption that the world is largely predictable; prioritizing unexpected input as in Dampening is a way of spending limited capacity where it is informative, since only a violation signals that the internal model needs revision. The Opposing Process Theory (OPT) reconciles the two by placing them in

sequence: an early bias toward the expected percept, followed by a reversal that emphasizes the unexpected (Press et al., 2020). Because the theory specifies the order of the two processes but not their time windows, it accommodates a wide range of findings. As a result, studies reporting support for OPT locate the effects very differently, for instance, before stimulus onset in one case (Rittershofer et al., 2026) or between roughly 100 and 200 ms after onset in another (McDermott et al., 2026).

Testing whether predictions shape perceptual content requires input that the acoustics do not settle on their own. With a clearly intelligible stimulus, an expectation can only add to evidence that is already sufficient, and the weight the system assigns to it is not readable from the response. With an ambiguous stimulus that weight becomes observable: what the listener reports is a direct read-out of how strongly the prediction is being used. Morphing a stimulus between two learned alternatives is an established way of achieving this, for instance, in faces (Garlichs & Blank, 2024) as well as in speech (Schneider & Blank, 2026; Xu & Jiang, 2026).

Whether a prediction is genuinely anticipatory, rather than reconstructed after the fact, is decided before the stimulus arrives. If the system instantiates a content-specific prediction, information about the upcoming stimulus should be recoverable from pre-stimulus activity. Such pre-stimulus templates have been reported in vision (Kok et al., 2017) and in audition (Demarchi et al., 2019), although the auditory finding is under active methodological dispute, because a systematic relation between a classifier's confusions and a sequence's transition structure can mimic the same result (Abdoun et al., 2026). Converging evidence that is not exposed to this particular confound comes from omission designs, in which the response elicited when a predicted sound fails to arrive differs depending on which sound was predicted (SanMiguel et al., 2013). A claim of this kind therefore stands or falls with the controls for that confound, which the Discussion (Section 3.4.5) addresses directly for the present design.

The design follows the requirements set out above: participants learned six trisyllabic pseudowords built from consonant–vowel syllables and then heard them embedded in continuous syllable sequences during EEG recording, alongside sequences carrying no word structure. Trisyllabic words result in three within-word positions: The first (P1) follows a word boundary and carries no within-word transition constraint. The second (P2) is fully determined by its predecessor but does not complete the word. The third (P3) is both determined and word-final. A subset of syllables was acoustically ambiguous between two learned alternatives (Morphs), so that the weight given to a prior expectation could be read out trial by trial while the acoustic evidence was held ambiguous by construction.

The study therefore tested whether expectations strengthen the neural representation of expected or of unexpected information in syllables, particularly in ambiguous Morphs, and whether the resulting pattern favours Sharpening or Dampening.

3.2 Materials & Methods

In this study, participants learned a set of six trisyllabic pseudowords (hereafter "words") and listened to sequences. In the sequences, mostly intelligible syllables (Standards) were presented, but sometimes Standards were replaced by acoustically ambiguous syllables (Morphs). Each Morph was an acoustic blend, combining the sound of the expected Standard syllable with another, unexpected Standard syllable. During the sequences, participants were occasionally required to indicate via keypress which syllable they heard last.

This design allows us to examine how sequential expectations shape the neural representation of ambiguous speech sounds (Fig. 3.1).

This study is preregistered (Becker et al., 2025).

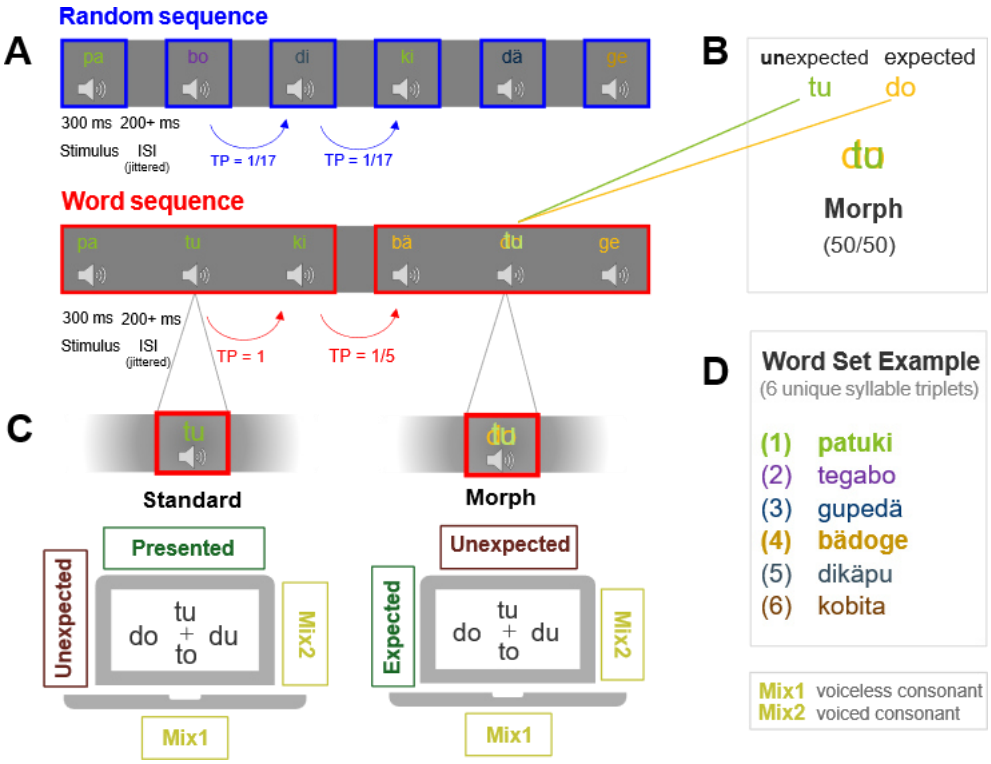


Figure 3.1. Experimental paradigm: sequences and response categorization. **(A)** Sequence Structure and transition probabilities. Random sequence (top, blue): 18 Standard syllables in randomized order (no self-repetition), uniform low transition probability (TP = 1/17). Word sequence (bottom, red): same syllables recombined into six unique trisyllabic pseudowords, balanced between-word transitions (TP = 1/5), within-word transitions predictable (TP = 1). **(B)** On a subset of trials, the Standard Syllable is replaced by an ambiguous Morph Syllable (50/50 blending of the expected and an unexpected Standard Syllable). **(C)** Sequence Task response categorization, two examples for within-word P2. Left task example: On Standard trials (presented syllable “tu”), the four options are the *presented* syllable plus three foils built from its Morph counterpart (*unexpected*: “do”) and two voicing-based confusion groups (*mix1* = voiceless consonant: “to”; *mix2* = voiced consonant: “du”). Right task example: On Morph trials (expected syllable “do”, presented: ambiguous blend of “do” and “tu”), the same four options are shown, relabelled to *expected* (“do”), *unexpected* (“tu”), and the same *mix1*/*mix2* groups. **(D)** Example word set consisting of six unique trisyllabic pseudowords.

3.2.1 Participants

Thirty-eight healthy volunteers were invited, two were excluded due to technical difficulties, leaving 36 adults (19 female, age: 27.39 ± 5.77 (20-40) years) who participated in the EEG experiment and are reported here. All participants were right-handed, native German speakers, and reported normal or corrected-to-normal vision, no history of neurological or psychiatric disorders, no hearing- or reading-related impairments, and no current or regular use of drugs or medication. Participants gave written informed consent prior to participation and were compensated at €13/hour. The study was approved by the Ethics Committee of the Hamburg Chamber of Physicians and conducted in accordance with the Declaration of Helsinki.
The sample size was fixed a priori by the counterbalancing scheme: the individualized word sets are constructed such that, across exactly 36 participants, every Standard syllable and every Morph occurs equally often at each within-word position.

3.2.2 Stimuli

Standard and Morph syllable corpus. Thirty-six German consonant-vowel (CV) syllables were recorded from a female native speaker (Neumann TLM 103 microphone), truncated to 300 ms, and loudness-normalized to 75 dB using a custom Praat script (Boersma & Weenink, 2020). To equate the recordings in fundamental frequency, so that F0 could not become a syllable-identifying cue in itself (Shannon et al., 1995), all recordings were pitch-corrected to ~181 Hz using Praat (Boersma & Weenink, 2020) and noise-vocoded (16 channels, a degree of vocoding that preserves full intelligibility (Blank & Davis, 2016)) using a custom MATLAB script (The MathWorks Inc., 2020).

From these 36 Standard syllables, 18 acoustically ambiguous Morph syllables were created by blending pairs of Standard syllables (both consonant and vowel) using TandemSTRAIGHT (Kawahara et al., 2008) in MATLAB. Syllable pairs were chosen to be phonetically similar while sharing no consonant or vowel (e.g., "tu"/"do"); each Standard syllable contributed to only one Morph. Standards and Morphs were evaluated in prior online studies to be intelligible (Standards) or perceived ambiguously (Morphs) with regard to their constituent consonant and vowel.

Word sets. Each participant was assigned an individualized set of six syllable triplets ("words"), drawn from 18 of the 36 Standard syllables and 9 of the 18 Morphs. Each Morph could replace either of its two constituent Standard syllables, at the position (1st, 2nd, or 3rd syllable) that syllable occupied in one of two different words. Assignment was balanced across the 36 participants such that, cumulatively, every Standard syllable and every Morph occurred equally often at each word position. Because syllable-position assignment, and thus Morph-position assignment, was fully counterbalanced across participants, the acoustic content entering each within-word position was identical at the group level. Position effects therefore cannot be attributed to stimulus acoustics.

Trial design. On each trial in a sequence, participants listened to a 300 ms syllable followed by an interstimulus interval (ISI) of ≥ 200 ms, ensuring that only the target sound occurred within a 500-ms analysis window. To avoid neural entrainment to a predictable stimulus rhythm and generic, time-locked rather than content-locked anticipatory activity, (Nobre & van Ede, 2018; Schroeder & Lakatos, 2009), we jittered the ISI, yielding a mean ISI of *M* = 268 ms, SD = 22 ms (mean stimulus-onset asynchrony, SOA, 568 ms).

Sequence design. Two sequence conditions were used: a Random condition, comprising two unstructured blocks and one block that already carried the word structure without disclosing it, and a word condition with syllables ordered into the learned triplets. Throughout, we refer to the two unstructured blocks as R1 and R2, to the third block of the Random condition, which carried the word structure without disclosing it, as the hidden-structure block R3, and to the five structured blocks as W1-W5. The ordinal slot a syllable occupies within a word is referred to as its within-word position 1-3, abbreviated P1-P3.

Random sequence condition. The two unstructured blocks (R1-R2) presented 306 Standard-syllable trials, realizing 305 of the 306 possible ordered syllable-to-syllable transitions exactly once, with each syllable occurring 17 times and a minimum lag of six syllables between repetitions of the same syllable to prevent accidental clustering of identical stimuli within local portions of the sequences. All syllables were followed by a Sequence Task probe (as described in the next section) approximately equally often; probes occurred, on average, every 13.3 ± 2.79 syllables (range 9–18), yielding 21.57 ± 1.10 (range 19–25) probes per block. Note that the third block in the Random condition, R3, followed the rules (i.e. trial number and presentation of Morphs) of the word sequence condition, unbeknownst to the participants.

Word sequence condition. Each of the five blocks presented 360 syllable trials (300 Standards, 60 Morphs): all six words occurred 20 times with maximally balanced word-to-word transitions. Blocks contained both Standard syllables (16–17 presentations/syllable/block) and Morphs (6–8 presentations/Morph/block, since each Morph replaced two different Standard syllables); on average, every second word contained a Morph. Two-thirds of the Sequence Task probes followed Morph syllables, to obtain a reliable behavioural estimate of the sequence-expectation effect. Probes occurred, on average, every 13.45 ± 2.73 syllables (range 9–18), yielding 25.38 ± 1.05 (range 23–28) probes per block.

3.2.3 Sequence Task and Response Categories

During Sequence Task probe trials (maximum duration 4 s, onset following the same SOA as for syllables), participants identified the most recently heard syllable by selecting one of four randomly arranged options via arrow keys.
For Morph probes, the options were derived from the two Standard syllables comprising the Morph. These always differed in both their plosive consonant (voiced: b, d, g; voiceless: p, t, k) and their vowel (a, ä, e, i, o, u):

- *Expected* / *Presented*: The syllable predicted by the learned word structure. (For Standards: the *presented*, correctly identified syllable; for Morphs: the *expected* component in an ambiguous Morph)
- *Unexpected*: Syllable from the same within-word position of another word. (For Standards: not presented. For Morphs: the other syllable used to create the Morph.)
- *Mix1*: A voiceless consonant (combined with either vowel to create a new, non-redundant option).
- *Mix2*: A voiced consonant (combined with the remaining vowel to create a new, non-redundant option).

For example, for a Morph composed of "do" (voiced) and "tu" (voiceless): *expected* and *unexpected* are "do" and "tu", respectively, *mix1*= "to", and *mix2* = "du". For Standard-syllable probes, alternatives were defined analogously, with the *expected* option being the actually presented syllable (see Figure 3.1).

3.2.4 Procedure and Paradigm

EEG recordings took place in a single session; stimuli were presented using PsychoPy (Peirce et al., 2019). The main experiment lasted 60–75 minutes, excluding briefing, EEG setup, and debriefing.

Setup and instructions. Following EEG setup and task instructions, participants were familiarized with all 36 Standard syllables to acclimate to the noise-vocoded sounds and completed a short practice block of the Sequence Task.

Part 1 – Random sequence (3 blocks). Participants listened to pseudo-randomized Standard-syllable sequences while fixating a central cross; syllables were displayed only as response options during Sequence Task probes. Blocks were separated by short breaks with visual feedback on task performance (% probes caught, mean RT). Unbeknownst to participants, the third block already followed the triplet structure of the (not-yet-learned) words and included Morphs, allowing us to test for incidental (implicit/statistical) learning of the word structure prior to explicit training.

Part 2 – Word learning. Participants were instructed to learn six trisyllabic words and informed that recall would be tested at the end of the session. Words were presented repeatedly, both auditorily and as on-screen text and participants were given the opportunity to speak them out loud and take a moment to internalize them. This was followed by a Continuation Task in which participants heard the first one or two syllables of a word and selected the correct continuation from the displayed options.

Part 3 – Word sequence (5 blocks). Participants listened to sequences structured according to their six learned words; this differed from Part 1 only in syllable order and in the presence of Morphs, which replaced one syllable in every other word. Participants again responded to Sequence Task probes. Between blocks, participants

were reminded of their words (repeated audiovisual presentation) and completed a brief Continuation Task, to prevent forgetting and to limit interference from Morph exposure on the learned word representations. After the third block, participants had a self-paced break.

Part 4 – Word-learning test. Learning was assessed with three tasks: (i) Free Recall (typing the six words), (ii) a Familiarity Task (12 trials; participants heard their learned word and a foil in random order, and indicated which was familiar), and (iii) a Certainty Rating (6-point scale, 0 = "not learned" to 5 = "very well learned", for each of the six words).

3.2.5 Statistical Analyses

Difference tests. Wherever Mauchly's test indicated a violation of sphericity in rmANOVAs, we report Greenhouse-Geisser-corrected degrees of freedom and *p*-values as p_{GG}. As effect sizes, we report generalized eta squared (η^2_G) for rmANOVA main effects and interactions, Cohen's *d* for one-sample and paired t-tests, and Hedges' g together with BF_{10} for post-hoc comparisons. All post-hoc comparisons following a significant effect were Holm-corrected within each family of tests addressing one research question and are reported as p_{holm}. Bonferroni correction was used to control the family-wise error rate across sets of a priori, non-overlapping analyses addressing related but distinct research questions (e.g., separate correlational analyses) and is reported as p_{bonf}.

Correlation tests. For all cross-subject correlations, we report both Pearson product-moment and Spearman rank correlation coefficients, r and ρ, rather than selecting a single method post hoc, since choosing parametric and non-parametric statistics after inspecting the data introduces researcher degrees of freedom and can affect significance in small samples (de Winter et al., 2016). Pearson correlations were computed after excluding outliers identified via Cook's distance (threshold $D_i > 4/n$) (Cook, 1977), and the 1.5 × interquartile range (IQR), each applied separately to the two variables entering a given correlation. Spearman correlations, being inherently robust to extreme values through rank transformation, were computed on the full, unexcluded sample.

Convergence between parametric and non-parametric methods (comparable direction, magnitude, and significance of effects) was taken as evidence of a robust association, whereas divergence was treated as diagnostic of outlier-driven or non-linear relationships. 95 % confidence intervals for correlation and paired-comparison effect sizes were obtained via percentile bootstrapping (3,000 resamples, participant-level paired resampling).

Modelling. Binary, trial- or item-level outcomes were modelled in two complementary ways. A mixed-effects logistic regression (GLMM) with by-participant random intercepts — and, where indicated, a crossed random intercept for word or syllable identity — estimates the effect for a typical participant, allowing for stable differences

in how readily participants give a given response. A generalized estimating equation (GEE) with an exchangeable working correlation structure and cluster-robust ("sandwich") standard errors (Zeger & Liang, 1986) estimates the corresponding population-averaged effect without assuming a distribution for those between-participant differences. The two approaches target different estimands, so their coefficients differ systematically in magnitude, population-averaged effects being attenuated relative to subject-specific ones (Fitzmaurice et al., 2011; Neuhaus et al., 1991); we therefore compared the two fits in terms of the direction and statistical significance of an effect rather than the size of its coefficient. As in the correlation analyses above, convergence across the two specifications was taken as evidence that a conclusion does not rest on the assumptions of either — in particular, on the distributional assumption placed on the random effects.
Which model was treated as primary was determined by the data structure of each analysis: the GEE where response categories are rare and the GLMM does not converge stably, the mixed model otherwise. Continuous outcomes were analysed with linear mixed models throughout. The specific model is stated for each analysis below and any case in which the two specifications would support different conclusions is reported at the relevant point in the Results in Section 3.3.1.

Software. All analyses were performed in Python (3.11.9), using toolboxes pingouin (0.5.5) and statsmodels (0.14.4) for statistical modelling, as well as matplotlib (3.8.2) and seaborn (0.13.2) for visualization.

3.2.6 Behavioural Data Analysis

We analysed behaviour for different task groups: Sequence Tasks, Retrieval Tasks, and Continuation Task.

In the Sequence Task participants indicated during the sequences which syllable they heard last. The Sequence Task is therefore our best indicator of online perception during the sequences. The Retrieval Tasks were administered at the end of the experiment and tested how well participants learned and remembered their words (Free Recall, Familiarity Task, Certainty rating). The Continuation task preceded every word sequence block and served primarily word training but at the same time is an indicator of word learning over time. Lastly, we analysed how performance correlated between tasks.

Sequence Task

For each participant, we computed the proportion of trials assigned to each response category and mean reaction time (RT), separately for Standard and Morph syllables and, in the word condition, for each within-word position (positions P1, P2, P3).

Sequence Task response categorization
For each response category separately, we fitted a trial-level binomial mixed-effects model (category chosen [1] vs. not [0] on every trial; logit link) with within-word position

(3-level categorical factor), stimulus type (Standard/Morph), and their interaction as fixed effects and a random intercept per participant, alongside a GEE. Where the Position × Stimulus Type interaction was significant, we further decomposed the categorical position effect into orthogonal linear and quadratic contrasts as appropriate for 3 equally-spaced levels, computed separately for each stimulus type, to test whether the position effect within a given stimulus type was adequately summarized by a single linear slope (non-significant quadratic contrast) or instead required the full categorical/non-monotonic description (significant quadratic contrast). A linear slope is reported and interpreted only in the former case.

To visualize and describe pairwise differences between response categories within a given position × stimulus-type combination, we additionally computed Wilcoxon signed-rank tests between all category pairs, Holm-corrected within each position × stimulus-type combination (see Results, Figure 3.2).

Reaction times per response category (Morphs only)
RTs were analysed at the trial level with a linear mixed-effects model, fitted on individual Morph trials (log(RT) ~ within-word position × response category, random intercept per participant). RT was log-transformed to account for its right-skewed distribution. Standard trials were not analysed this way due to insufficient trial counts in non-expected response categories (pooled across all 36 participants: 7–10 unexpected trials and 15–35 mix2 trials per position, with 26–32 and 16–27 of 36 participants, respectively, contributing zero trials to a given position × category cell; see Figure 3.2 and Section 3.3.1.1).

Significance of the position main effect, category main effect, and their interaction was assessed with likelihood-ratio tests (LRT; fitted by maximum likelihood), comparing the full model against a reduced model without that term. Because the outcome is log-transformed, a coefficient b corresponds to a multiplicative effect on RT of $\exp(b)$, e.g. $b = -0.05$ means RTs were shorter by a factor of ≈ 0.95, i.e. about 5 % faster.

As a robustness check, the same model was refitted with an additional crossed random intercept for syllable identity, to account for syllable-level variability in addition to participant-level variability. To characterize the shape of the position effect within *expected* responses, we computed orthogonal linear and quadratic contrasts (as described above for response categorization) from this same model's coefficients.

Reaction times per stimulus type and within-word position
Per participant we computed the mean log-RT in each stimulus type × position cell and tested stimulus type, within-word position and their interaction with a 2 × 3 rmANOVA, followed by Holm-corrected paired contrasts. A trial-level linear mixed model, log(RT) ~ stimulus type × position + (1 + position | participant), was fitted alongside as a cross-check. It carries a per-participant random slope for position so that the within-participant position effect is not evaluated against pooled trial-level variance. As a further robustness check this model was refitted with an additional crossed random intercept for the identity of the expected syllable.

Retrieval Tasks

Free Recall

After completing the experimental session, participants' explicit memory for the six learned pseudowords was assessed with a Free Recall task, in which they typed as many of their six words as they could remember into six blank text fields, without time pressure. Once entered, a word could not be changed. Each participant's response was scored as the number of correctly reproduced words (0-6).

Familiarity Task

In the Familiarity Task, participants completed 12 trials in which they heard their previously learned (familiar) word and a foil word presented sequentially in randomized order and indicated via key press whether the first or second word was the one they had learned. Foil words were created by replacing one syllable with a similar-sounding syllable occurring at the same position in another word (i.e., the *unexpected* syllable in Sequence Task probes) or by re-ordering the syllables. This yielded, per participant, the percentage of correct 2AFC judgements across the 12 trials (chance level = 50 %). Group-level performance was tested against chance with a one-sample *t*-test and, due to the ceiling effect of accuracy, additionally checked for a robustness using a non-parametric Wilcoxon signed-rank test against chance.

Certainty Rating

In the Certainty Rating task, participants rated each of their six learned words, presented individually, on a 0 ("not learned") to 5 ("very well learned") scale, indicating their subjective confidence in having learned that specific word. For each participant we computed the minimum, mean, and maximum rating across their six words, as well as the full distribution of individual word-level ratings across the sample.

Relating retrieval task performance

To test whether participants performance in the retrieval tasks was related, we computed Pearson correlations for each pair of the retrieval tasks Free Recall (*n* words recalled), Familiarity Task (% correct trials) and Certainty Rating (mean rating) across participants. We additionally computed Spearman correlation coefficients ρ as a rank-based robustness check given the discrete and bounded nature of the Free Recall and Familiarity Task.

Continuation Task

The Continuation Task was administered before each of the five word-sequence blocks to assess prediction performance and to support word learning through active prediction. On each trial, participants heard the onset of one of their six words (first syllable or first two syllables) and selected the correct continuation from two displayed syllables (correct and false, corresponding to the two options expected and unexpected from the Sequence Task probe) within a 4-s window; misses triggered a warning. After the response, feedback was provided for the chosen syllable (green highlight for correct, red for false), followed by a single audiovisual presentation of the full word.

Accuracy and RT were analysed at the trial level so that each trial contributes in proportion to the data provided by participant and word, and participant- and word-level variability are estimated separately (each participant's six words being unique to them). Asking whether prediction performance in the Continuation Task increased over time and whether it differed between positions, the hypotheses concern main effects of Block (Block 1 reference) and within-word Position (P2 reference). So the primary models were additive: a population-averaged binomial GEE for accuracy (logit link, participant clusters, cluster-robust SEs) and a log-RT linear mixed model (REML; participant random intercept plus a word-level variance component). The Block × Position interaction was tested by adding the term (joint Wald test / likelihood-ratio test).
A crossed participant × word binomial GLMM gave the same fixed-effect pattern and, like the RT model, showed substantial word-level variance. This shows that words differed reliably in how well they could be predicted, motivating the word-level analyses in the following section.

Relating behavioural tasks: syllable- and word-level analyses

Relating Continuation Task accuracy to retrieval tasks

We wanted to quantify whether the regular training as induced by the Continuation Task between word sequence blocks facilitated word learning. To this end, we used the mean Continuation Task accuracy per word, averaged over all blocks and both syllable positions (yielding a Continuation Task word-level score), for each of the 216 (participant × word) observations to predict each words' Certainty rating and Familiarity accuracy (linear mixed effects model, LME, random intercept per participant) and whether it was freely recalled (GEE, binomial family, exchangeable working correlation, participant as cluster).

Associating word learning strength and Continuation Task Performance with expected responses in Sequence Task Probes

To test whether categorization in the Sequence Task during the word sequence was related to word learning strength, we fitted GEEs predicting the trial-level binary *expected* response from that word's Free Recall, Familiarity Task, and Certainty Rating scores. We fitted models separately for Standard and Morph syllables as Standard trials indicate correct identification of the expected (i.e. actually presented) syllable, whereas Morph trials rather capture whether participants perceived the deviating or expected component in the Morph.

In addition to the three retrieval measures, we entered each word's Continuation Task word-level score as defined above as a predictor of the *expected* response, once on its own and once jointly with the Certainty rating, to test whether the more objective training measure accounted for Sequence Task behaviour over and above the subjective rating measure. Because Continuation Task accuracy is defined only for the second and third syllable of a word (the first syllable serves rather as the retrieval cue), models including it were restricted to Sequence Task probes at P2 and P3, with syllable

position as a covariate; Morph probes at P1 were analysed separately, using the word-level Continuation Task score, as a generalization test.
For the binary *expected* response, the Continuation Task predictors were decomposed into within- and between-participant components, and the adequacy of the GEE specification was verified in two ways. First, to confirm that participant-level clustering captured the non-independence of repeated probes of the same word, all GEEs were refitted with clustering at the item (participant × word) level; this left the direction and significance of every effect unchanged. Second, a binomial GLMM with random intercepts for participant and word was fitted as a random-effects sensitivity check, which confirmed the direction and significance of the GEE effects.
Sequence Task probe response times (log-transformed) were analysed with linear mixed models (random intercept for participant, word variance component), fitted separately for Standard and Morph probes and predicted by the word's Continuation Task accuracy; for Standard probes we additionally used the word's mean Continuation Task RT as a same-modality convergent-validity check.

Assessing phoneme-specific response behaviour in the Sequence Task
To evaluate biases or asymmetries in responses to the Sequence Task probes based on speech features, we decomposed task responses along two dimensions. *Expected* and *unexpected* responses were defined by whole-syllable identity relative to the presented/nominally-expected syllable and its full phonetic opposite. Separately, each response was scored for whether the chosen consonant and the chosen vowel each matched the expected syllable, computed directly from the selected response syllable. Trials were grouped by the voicing of the expected consonant (voiced: /b/, /d/, /g/; voiceless: /p/, /t/, /k/) and, within the voiced-expected subset, by place-of-articulation pair (labial, alveolar, velar); vowels were grouped by pair (a/ä, e/i, o/u).

All analyses were conducted at the participant level. The primary dependent measure was, per participant, the difference *unexpected* − *expected* in percent across the four response categories (or, for the voicing contrast, the proportion of "unexpected" responses among trials answered "expected" or "unexpected"). For single-factor within-participant contrasts, a linear mixed-effects model with a random intercept by participant is equivalent to a paired / one-sample *t*-test, which is what we report. A population-averaged logistic model (GEE, binomial family, exchangeable working correlation, clustered on participant) is reported alongside as a cross-check. Trial-level mixed models were also fitted but gave anticonservative *p*-values for sparse cells (2–5 trials/participant/cell) and therefore are not relied upon. The same measures were computed at within-word P2 and P3. Reaction times were compared with paired *t*-tests on per-participant cell means. Descriptive proportions use the mean of per-participant proportions.

3.2.7 EEG Data Acquisition

EEG acquisition

EEG was recorded from 64 Ag/AgCl active electrodes (ActiCap64, Brain Products, Germany) using BrainAmp amplifiers, positioned according to the extended 10–20 system, with reference and ground at FCz and Iz, respectively. Four additional electrodes recorded vertical and horizontal electrooculography. Impedances were kept ≤10 kΩ; data were sampled at 1000 Hz.

EEG preprocessing

EEG preprocessing was performed using MNE-Python (1.9.0) (Gramfort et al., 2013). The preprocessing followed a preregistered pipeline, consistent with previous experiments in our lab (Schneider & Blank, 2026). Line noise was removed with a notch filter at 50 Hz and its harmonics; data were then demeaned and visually inspected for excessive noise. Noisy channels were identified and interpolated in 5 of 36 participants (1-8 channels per participant, M = 0.44, SD = 1.48 channels across the full sample).

To detect brief but pronounced muscle artefacts, a separate copy of the data was band-pass filtered (110–140 Hz, IIR) and segmented into −500 to 2500 ms epochs. For each channel, data were z-scored jointly across all epochs and all time points; the maximum absolute z-score across time was then taken for each channel and epoch, and these per-channel values were averaged across channels to obtain a single artefact score per trial. Trials were flagged as bad using a participant-specific threshold and confirmed by visual inspection. Continuous data were band-pass filtered (0.5–50 Hz, IIR) and epoched from −500 to 2500 ms relative to audio onset. Independent component analysis was used to remove ocular, cardiac, and muscle-related components (Median = 7.0, Mean = 13.72, SD = 14.37 components removed per participant). The mean exceeds the median because a small number of participants required extensive ocular and muscle correction; every removal was confirmed by visual inspection of the data. All data were finally downsampled to 200 Hz to reduce computational load for subsequent analyses. Data per trial were baseline-corrected by subtracting the mean pre-stimulus signal (-100 to 0 ms).

For ERP analyses, epochs were re-referenced to the average of all channels and bad trials were excluded as artefacts can distort trial-averaged waveforms (excluded trials in word sequence condition: Standards: 13.24 ± 6.59 % per participant (M ± SD), Morphs: 13.44 ± 6.47 %; unstructured blocks R1-R2: Standards: 6.45 ± 5.16 %). For decoding analyses, no trials were excluded, as even artefact-contaminated trials can carry information relevant to decoding, and retaining all trials preserves the full representational structure of the balanced stimulus presentation (Section 3.2.9). Trial data used for decoding were smoothed with a moving average of nine time points (45 ms).

3.2.8 Univariate ERP analysis

For univariate EEG analyses, we computed ERP trial data for each condition and per participant. Statistical analyses were run across participants if not stated otherwise.

ERPs: Spatiotemporal Cluster Tests

Statistical Testing ERPs

Event-related potentials (ERPs) were analysed in a data-driven approach using spatiotemporal cluster-based permutation tests (Maris & Oostenveld, 2007) in MNE-Python to identify condition-related differences while controlling for multiple comparisons across electrodes and time. For each participant, condition-specific ERPs and analyses were computed, and differences between conditions were tested against zero over the 0–500 ms post-stimulus window. Spatial adjacency was defined via Delaunay triangulation of the digitized channel coordinates to enforce spatial contiguity. Clusters were defined as sets of adjacent electrodes and time points exceeding a data-driven threshold derived from the *t*-distribution (df = 35, $|t| = 2.03$), and cluster significance was assessed using 10,000 permutations (two-tailed, $\alpha = .05$), using a fixed random seed (seed = 42) for reproducibility. To rule out that effects were due to the specified fixed threshold, we re-ran the complete confirmatory ERP analysis battery with threshold-free cluster enhancement (TFCE, Smith & Nichols (2009)); as implemented in MNE-Python using default threshold settings) as a pre-specified robustness check holding all other test settings identical. Corroborating the reported findings with minimum *p*-values being smaller under TFCE for most contrasts, every significant cluster was recovered, and no parametric null became significant apart from a single exception explicitly noted (Position effect: P1 vs. P3, pooled across Stimulus Type).

We note that, as cluster-based permutation tests control the family-wise error (FWE) rate across the entire search space, the temporal and spatial extent of a significant cluster is not an estimate of the true onset, offset, or topography of an effect (Sassenhagen & Draschkow, 2019). Cluster extents are therefore reported descriptively, to characterize where in the search space evidence against the null hypothesis accumulated, as one cannot make inferential claims about effect latency or location. Each cluster was summarized by its cluster mass, that is, the sum of the *t*-values it contains.

Where multiple contrasts addressed one research question (e.g., pairwise position or stimulus-type comparisons), Bonferroni correction was applied within that family of tests, reported as p_{bonf}.

Main effects of stimulus type and position

To start with, we tested for univariate differences in ERPs of stimulus type (Standard vs. Morph stimuli) and position (1, 2, 3), following the theoretical consideration that prediction effects should not only be tested against each other but also against a neutral baseline (Feuerriegel et al., 2021). We chose our Random-sequence condition as neutral baseline as it presented the same stimuli in an unpredictable order before explicit word learning. We note, however, that the Random condition is not an optimal neutral baseline to compare against (Feuerriegel et al., 2021) because of the fixed

order of conditions and the change in stimulus inventory (random: Standards only, word sequence: Standards and Morphs). This is further addressed in the Discussion.

For stimulus type, we tested Standards against Morphs, and each against the Random sequence condition, as one family of three comparisons (Standard vs. Morph, Standard vs. Random, Morph vs. Random; $\alpha = .05/3 \approx .0167$).

For position, we tested the three pairwise position comparisons (P1 vs. P2, P1 vs. P3, P2 vs. P3) together with each position against the Random condition, as one family of six comparisons ($\alpha = .05/6 \approx .0083$).

Interaction of stimulus type and position

For the interaction of stimulus type and position, we contrasted the Morph-minus-Standard difference wave against itself across the same three position pairs (difference-of-differences; $\alpha = .05/3 \approx .0167$). This contrast does not involve the Random sequence condition.

Position-by-stimulus-type effects

As a follow-up to the different effects for both stimulus type and position, we investigated each position individually, testing Standards against Morphs directly together with each against the Random condition, as one family of three comparisons per position (P1, P2, P3, each: Standard vs. Morph, Standard vs. Random, Morph vs. Random; $\alpha = .05/3 \approx .0167$ per position).

Visualization

Where cluster tests covered several pairwise contrasts within one family (e.g., Standard/Morph/Random), we plot all conditions in one shared ERP panel with a single FWE-corrected significance strip beneath, rather than one figure per contrast. Most effects appeared as two topographically mirrored clusters, frontal and posterior, in closely matching time windows. This is expected under an average reference, which by construction sums to zero across the montage at each time point (Luck, 2014), and is further consistent with the classic antero-posterior polarity inversion reported for the auditory N1–P2 complex and the mismatch negativity, commonly attributed to a dominant, tangentially oriented generator in or near auditory cortex (note that P2 here denotes an ERP component, not within-word position 2) (Näätänen et al., 2007; Näätänen & Picton, 1987).

Each panel's topography and ERP difference wave are anchored to one pre-specified reference cluster (typically its posterior member, for topographic clarity), at that cluster's peak latency (maximum |t|). The significance-bar strip beneath each panel shows every tested pairwise comparison, distinguishing clusters that survive FWE correction (black) from those that do not (grey, labelled "n.s., corrected"), so that no borderline effect is omitted. Reported time windows and *p*-values always reflect the original cluster-permutation tests, never values recomputed from the visualization, to avoid double-dipping. All individual clusters, including non-reference and non-significant ones, are shown separately in the Supplement.

Control Analyses

Control Analysis for the "Hidden-Structure Block". To test whether the absence of position main effects in the hidden-structure block was driven by reduced statistical power (~112 trials/position vs. ~520 in the pooled word sequence dataset), we conducted an exploratory control analysis. We re-ran the spatiotemporal cluster permutation tests across all five single, N-matched word sequence blocks (~101-107 trials/position) to directly evaluate position effects at comparable sample sizes.

Morph-Task Trial Bias. Sequence Task probes followed Morph syllables at a higher rate than Standard syllables (28.4 % vs. 2.8 %; 66.9 % of all task probes followed a Morph trial). To test whether the Morph > Standard main effect reflected task relevance or response preparation rather than genuine prediction-related processing, we re-ran the spatio-temporal cluster permutation test restricted to trials that were never followed by a probe.

Note that probe rates were near-identical across within-word positions, both pooled (7.0-7.2 %) and separately by stimulus type (Morph 28.1-28.7 %, Standard 2.7-2.9 % at each of the three positions), so this probe imbalance is no confound for the position main effect. Probe onset relative to the preceding syllable onset was at minimum 512 ms across all 7,036 probed trials, so a probe screen can never fall inside the 0-500 ms analysis window.

Random Sequence Trial-Count Matching. The Random sequence as a control condition, pooled across positions, comprises more trials per participant (~573–600) than Standard-P2 alone (~435). To test whether the Standard-P2 suppression relative to the Random sequence (as a central finding, see Section 3.3.2) was influenced by the trial-count imbalance between conditions, we conducted a trial-count-matched subsampling control: across 200 iterations, each participant's Random sequence trial pool was randomly subsampled without replacement to match that participant's own Standard-P2 trial count, and the spatio-temporal cluster permutation test was re-run against the resulting matched Random sequence ERP.

Baseline Correction Control. ERPs were baseline-corrected using a -100–0 ms pre-stimulus window. Given the design's mean SOA of 568 ms (300 ms syllable duration + 268 ms mean, jittered inter-stimulus interval), this window falls within the preceding syllable's own late response, and since the preceding syllable is confounded with within-word position (P2 always follows P1, P3 always follows P2), any position-dependent offset in this window could in principle contaminate the reported position effects. To test this, we re-ran the full confirmatory cluster-permutation battery (stimulus-type main effect, position main effect, position-within-stimulus-type, the stimulus-type × position interaction, and all Random-sequence/hidden-word-block corroboration contrasts) on ERP epochs processed identically to the main pipeline but without baseline correction, and additionally tested for position- and stimulus-type-dependent differences within the nominal baseline window itself on these uncorrected epochs.

3.2.9 Multivariate Decoding analyses

One-vs-rest time-generalized decoding

Decoding analyses were performed using the MVPy toolbox for Python (Schneider, n.d.) and conducted separately for each participant. Classifiers were trained independently at each time point (5 ms steps) on voltage patterns across all EEG electrodes.

Critically, each trained classifier was then tested not only at the timepoint it was trained on but at every timepoint, yielding a full train-time × test-time generalization (TG) matrix for every decoding measure. The matrix diagonal (train time = test time) corresponds to standard time-resolved decoding. The off-diagonal entries indicate whether the neural pattern discriminating syllables at one moment generalized to, or reactivates at, other moments (temporal generalization, King & Dehaene (2014)).

A regularized linear classifier (ridge regression with a one-vs-rest multiclass strategy, combined with Ledoit–Wolf covariance estimation) was fit at each training timepoint using repeated stratified k-fold cross-validation (5 folds × 50 repeats = 250 folds) on Random-sequence trials (i.e., the first two blocks of the Random sequence condition). Within each fold, the classifier was trained on 80 % of the Random-sequence trials and tested on the held-out 20 % to estimate within-condition decoding accuracy. Word-sequence trials never entered training and were therefore fully out-of-sample; predictions for these trials were averaged across the 250 fold-specific classifiers.

Per participant, this procedure produced TG decoding matrices
(i) within-condition validation on Random sequences,
(ii) generalization to Standard syllables in word sequences, and
(iii) generalization to Morph syllables in word sequences.

Expected and unexpected components

Note that for the word sequence all four decoding measures, as described below, were computed twice per trial: for the expected syllable (following the learned word structure, which equals the true, presented syllables on Standard trials and the expected component in Morph trials) and for its designated unexpected Morph component (see section Sequence Task options; the *unexpected* stimulus information both in Standard and Morph trials). This yields parallel expected- and unexpected component measures for both Standard and Morph trials.

Measures of Decoding Performance

Recorded decoding measures include:
(i) decision values
(ii) MRR (mean reciprocal rank)
(iii) accuracy
(iv) ROC AUC (area under the receiver-operating characteristic curve; hereafter AUC)

Decision values. As the classifier's goal is to distinguish which syllable was presented, it produces one continuous decision value per candidate syllable, per trial, per timepoint, reflecting how strongly the multichannel EEG pattern on that trial resembles

the pattern the model learned to associate with that syllable. Every trial therefore yields a full vector of decision values, one per candidate syllable, at each decoding time point. The decision value vector gets evaluated either for all 18 candidate syllables, or for only the selection of syllables that are valid candidates in the experimental design, e.g., only 6 syllables can occur at a given within-word position in the word sequence, producing one of the following decoding measures.

MRR. For each trial, the mean reciprocal rank captures the reciprocal of the rank the target syllable (i.e., the expected or presented syllable) received when all candidate syllables were ordered by decision value (1 = ranked first), averaged across trials. Chance level is the corresponding harmonic-number expectation for the number of candidates, i.e., chance level for the full 18-syllable inventory: 0.194, and when restricted to the 6 syllables valid at a given within-word position: 0.408. In contrast to accuracy, which is based on the top-ranked syllable only, and AUC, which compares all syllable pairs, MRR offers a graded perspective on the classifiers' rating of specific syllables (e.g. the expected or unexpected syllable) relative to all others. In other words, a high MRR value indicates that a target syllable is ranked near the top classifier's choices even if it is not the classifier's top guess.

*Accuracy***.** Decoding accuracy is defined as the proportion of trials on which the syllable with the highest decision value matches the syllable presented (or expected in a Morph trial), evaluated against chance (chance level for 18-syllable set: 1/18, for 6-class position restricted set: 1/6).

AUC. For each syllable, the AUC quantifies how well its graded decision value discriminates trials on which it was presented from trials on which it was not, independent of any specific classification threshold. It indicates whether the signal contains decodable information about the target syllable, instead of just indicating whether it was the MRR-top-rated syllable as accuracy does. Also, AUC chance level is always 0.5 and thus easily comparable across 18-class and 6-class position-restricted analyses. For these reasons, we chose AUC as primary decoding measure.

Statistical Testing of Decoding Performance

We statistically evaluated the time-generalized decoding performance for the available combinations of conditions (metrics: accuracy, AUC, MRR; for P1-P3 and aggregate; for Random sequence, Standard and Morph trials in the word sequence; for expected and unexpected syllables) in the time window spanning a trial and its preceding interstimulus interval, i.e., 0 to 500 ms after stimulus onset on the train axis and -200 ms to 500 ms on the test axis of the matrix, across participants. Group-level significance was assessed using cluster-based sign-flip permutation tests (10,000 permutations, fixed random seed, cluster-forming threshold derived from the t-distribution at $p < .05$, two-tailed).

At each timepoint (diagonal decoding) or each train-time × test-time cell of the decoding matrix, a one-sample *t*-test compared each participant's decoding value against the appropriate baseline. This baseline was chance-level for AUC (0.5), for accuracy (1/18 for full syllable set or 1/6 for the position-restricted set), or MRR (the harmonic-number chance-level for the corresponding number of classes), and 0 for condition-based contrast measures (Condition A minus Condition B).

Contiguous timepoints (1D, for diagonal decoding) or contiguous cells (2D, using a matrix-based 8-connected neighbourhood spanning both train and test time) exceeding the cluster-forming threshold were grouped into candidate clusters. Each cluster was summarized by its cluster mass (the sum of its *t*-values). On each permutation, participants' signs were randomly flipped, and the maximum resulting cluster mass was recorded to construct an empirical null distribution. A candidate cluster's *p*-value was defined as the proportion of permutations yielding a maximum cluster mass equal or greater than the observed mass.

Notably, the 1D diagonal decoding and the 2D decoding matrix can yield differing cluster results. While timepoints in the 1D diagonal only form clusters with temporal neighbours, cells in the 2D matrix must additionally cohere with adjacent training rows. Consequently, effects in the 2D matrix are less likely to survive correction. We therefore treat the diagonal test as our primary, more sensitive significance check, and the full 2D matrix as an additional, more conservative criterion, except for off-diagonal effects, which by construction cannot be captured by the diagonal test and for which the two-dimensional analysis is the only available test.

One-vs-rest trial-resolved decoding

Next to the primary dataset consisting of the TG decoding matrices, a second dataset was generated via a fully separate, independently executed analysis run, using the same underlying training data and classifier recipe (ridge one-vs-rest, Ledoit–Wolf covariance, 5-fold × 50-repeat stratified cross-validation) but its own newly-fit classifiers, restricted to the diagonal (train time = test time only). The second run additionally retained per-trial, per-class decision values and probabilities at every time point rather than only fold-averaged summary metrics. This trial-resolved dataset enabled the single-trial analyses reported in this study, such as control analyses and linking decoding evidence to individual trials' behavioural responses and phoneme identity, which are not possible from the fold-averaged TG matrices alone. Despite being independently re-fit, the two datasets were confirmed to agree closely: the decoding diagonal of the trial-resolved dataset correlated with the diagonal of the TG dataset at $r > .9999$ (essentially identical means, peak magnitudes, and peak latencies), and independent cluster-permutation tests conducted on the trial-resolved dataset reproduced the significance pattern established with the TG dataset (e.g., significant decoding for Standard word-sequence trials in the same time window, and null decoding for Morph trials). This shows that the decoding pipeline itself, not merely a fixed set of model outputs, is highly reproducible.

Further analyses on decoding performance

Correlation of pre-stimulus decoding (ROC/MRR) cluster and expected responses in Sequence Task

For Standard syllables at within-word P2, the decoding matrix of both ROC and MRR values for the expected syllable identified two significant, distinct sub-clusters: firstly, a compact, on-diagonal block (both train and test time post-stimulus, ROC: train 145–500 ms × test 150–500 ms; MRR: train 265–475 ms × test 230 to 500 ms;) consistent

with the 1D decoding along the diagonal, and secondly, off-diagonal sub-cluster whose entire test-time extent falls before stimulus onset (ROC: train 150–500 ms × test -200 to -65 ms; MRR: train 135–500 ms × test -195 to -35 ms), in the following referred to as “pre-stimulus decoding(/ROC/MRR) cluster”.

We wanted to test for a relationship of the pre-stimulus decoding clusters and a behavioural indicator of syllable perception during the sequences, i.e., the Sequence Task probes. As participants mostly responded with the *expected* syllable in Sequence Tasks of Standard trials, leaving little response variance to base a correlation on, we computed the correlation for Morph trials where task responses varied more between response categories. Importantly, during the pre-stimulus time window of Morph trials participants are not aware that they will hear a Morph next. Given the 1:5 ratio of Morph and Standards, participants likely expect to hear a Standard syllable.

To compute the correlation, the exact cell coordinates of the pre-stimulus component of the Standard Position-2 decoding cluster were transferred onto the Morph matrices for *expected* syllables (i.e., the aggregate matrix and each of Positions 1–3 separately). Each participant's mean decoding value (ROC AUC or MRR) within this transferred set of cells was correlated with that participant's proportion of *expected* responses on Morph trials in the Sequence Task (the overall proportion for the aggregate matrix, the position-matched proportion for each position). Pearson correlations were computed after removing bivariate outliers (Cook's distance > 4/n and/or 1.5 × IQR on either variable); Spearman correlations used the full sample. Estimates were corroborated with percentile bootstrap 95 % confidence intervals (3,000 resamples); for the Pearson intervals the outliers were identified once on the full sample and held out of every resample. Bonferroni correction was applied across the four tested matrices ($\alpha = .0125$; see Figure 3.7).

As a sanity check on whether this neural–behaviour relationship depends on the exact cell selection of the pre-stimulus cluster, we repeated the correlation after averaging decoding over the entire pre-stimulus window rather than the cluster coordinates transferred from Standard P2: the full rectangle spanning the whole post-stimulus training axis (0–500 ms) and the whole 200 ms pre-stimulus test window (−200–0 ms), applied to the same 2D Morph ROC AUC (and MRR) matrices, with all other analysis choices (behavioural measure, outlier handling, bootstrap CIs, and Bonferroni FWE correction) held identical.

Position contrast on the decoding time course

To test whether syllable-identity decoding differs between within-word positions, we contrasted the on-diagonal decoding time course (train time = test time; Standards in W1-W5, expected label; ROC AUC as the primary metric, MRR secondary; computed per position over the six syllables occurring at that position) between positions, paired across participants. For each participant we computed a difference wave of the on-diagonal decoding time courses over 0–500 ms and submitted it to a cluster-based permutation test against zero (10,000 sign-flip permutations, cluster-forming threshold

$p < .05$ two-tailed, cluster mass = Σt). The theory-driven, pre-specified confirmatory contrast (see Section 3.3.3) was P2 vs P1. P1, rather than the Random condition, served as the unpredictable comparison as it was drawn from the same session and trial structure, and is the only within-word position not constrained by a preceding within-word transition. P2 vs P3 and P1 vs P3 were computed as secondary, descriptive contrasts. Because the P1 vs P2 contrast was confirmatory and pre-specified, no multiple-comparison correction was applied. The two secondary contrasts are reported uncorrected for completeness. As a complementary summary we additionally compared the mean diagonal AUC across a fixed a priori 100–500 ms window between positions (paired *t*-test, with a Bayes Factor BF_{01}, for the confirmatory contrast).

Correlation of ERP effects and Decoding

As the literature differentiates Sharpening and Dampening mechanisms based on the relationship of the univariate EEG activity, i.e., typically expectation suppression, and the multivariate decoding performance, we wanted to investigate potential correlations in our data speaking for neural Sharpening (positive correlation) or Dampening (negative correlation).

To this end, we correlated the suppression magnitude of the established P2 ERP suppression cluster with, firstly, the pre-stimulus decoding cluster itself and, secondly, mean decoding performance (AUC, MRR) in that position's own, independently established decoding time window, using Pearson (with Cook's-distance/IQR outlier exclusion) and Spearman correlations. Suppression magnitude at P2 was quantified as the signed Standard-minus-Random amplitude difference in P2's own cluster, so that more positive values indicate stronger suppression (a smaller-magnitude deflection for Standard relative to Random).

To investigate whether these participant-level correlations are informative in the first place, we estimated the split-half reliability of each per-participant decoding metric (400 random trial splits, Spearman-Brown step-up, $n = 36$).

The position-restricted MRR/AUC measures entering the reported tests had low reliability (P2 ≈ 0.2; P1 and the 18-class aggregate ≈ 0). We therefore interpret null correlations in this section as evidence against a *strong* Sharpening- or Dampening-type coupling only, not as evidence for its absence, and we do not treat them as adjudicating between the two accounts.

Beyond this primary test, we exploratorily also ran more analyses, such as linking single-trial ERP amplitude to single-trial decoding evidence or using alternative a priori and data-driven time-window/channel definitions. However, single-trial ERP–decoding analyses were not interpreted, as the evidence had near-zero split-half reliability, and difference-based measures with low reliability cannot support individual-differences inference, so a null association is uninformative rather than evidential. This reflects the *reliability paradox*: paradigms that yield robust group-level effects often show poor between-unit reliability precisely because a small true between-unit variance is what makes the group effect clean (Hedge et al., 2018), and, for ERP measures specifically,

internal consistency depends strongly on the number of trials per participant per cell (Clayson & Miller, 2017).

We emphasize that this reliability-based caution applies specifically to the null correlations reported in this section. It does not carry over to the significant, independently replicated association between pre-stimulus decoding and behavioural expectation strength reported in Section 3.3.3. Measurement error attenuates observed correlations toward zero rather than inflating them, so a correlation that survives outlier-robust estimation and label-permutation testing is a conservative lower bound on a true effect, not an artefact of noise. In other words, low reliability disarms the null suppression–decoding correlations, whereas the positive pre-stimulus finding stands on its own and is, if anything, underestimated.

Control Analysis: Resampling Standard trials to test Morphs low trial number

To determine whether the absence of significant 1D Morph decoding was driven purely by statistical underpowering (due to a 5:1 Standard-to-Morph trial ratio), we implemented a trial-count-matched control analysis. For each participant, 100 independent iterations drew a subset of Standard trials matched exactly in trial count (~300) and position distribution to that participant's own Morph trials. AUC metrics were recomputed from the trial-level decision values and submitted to 1D cluster-based sign-flip permutation tests (0–500 ms, 10,000 permutations, cluster-forming $p < .05$) against chance level (0.5), logging the complete cluster distribution across iterations. Full-N Standard ($n \approx 1500$) and un-subsampled Morph ($n \approx 300$) reference tests were included as benchmarks.

Control Analysis: Accounting for phoneme-specific behavioural asymmetries

To test whether phoneme identity-specific behavioural asymmetries in the Sequence Task response could confound, or be confounded with, the neural decoding results, we re-analysed the existing trial-level decoding output for word-sequence Morph trials at the first syllable position (N=3,578 trials across 36 participants). For each trial, we extracted the decision values corresponding to the trial's expected syllable and its designated unexpected counterpart, averaged each over the 0–500 ms post-stimulus window, and computed their difference (expected minus unexpected) as a trial-level neural evidence measure. Trials were grouped into the same phoneme subgroups used in the behavioural analysis (i.e. voiced- versus voiceless-expected consonant, each of the three consonant pairs and their two members, and each of the three vowel pairs and their two members) and the neural evidence measure was tested against zero within each subgroup using one-sample *t*-tests, with between-subgroup comparisons conducted via participant-clustered GEEs.

Phoneme composition as a possible confound for the pre-stimulus decoding–behaviour correlation is ruled out by construction and a direct consequence of the fully counterbalanced stimulus design: the proportion of voiced consonants (50 %) and of each specific consonant and vowel identity (16.7 % each) among the trained syllable set was identical for all 36 participants.

Control analysis: trial timing, baseline correction, and pre-stimulus carry-over
Due to the design's mean SOA of 568 ms (300 ms syllable duration + 268 ms mean, jittered ISI), the neural response to each syllable partially overlaps with the pre-stimulus interval of the next, and within-word position is perfectly confounded with the identity of the preceding syllable. Two control analyses tested whether this overlap could contribute to the reported neural position effects.
First, the full confirmatory ERP cluster-permutation battery was re-run on data processed identically on the main pipeline but without baseline correction, and a separate diagnostic cluster-permutation test was applied to the nominal −100 to 0 ms baseline window itself (on uncorrected data) to test whether position or stimulus type already differ there.
Second, in the decoding domain, the Position-2 pre-stimulus cluster mask (train 135–500 ms × test −195 to −35 ms for MRR; train 150–500 ms × test −200 to −65 ms for ROC-AUC) was transferred cell-for-cell onto Position 1's own time-generalized decoding matrix, with the training-time axis unchanged and the test-time axis advanced by the mean SOA (114 samples of 5 ms), so that each transferred cell indexes the same absolute post-onset latency relative to Position 1's own onset. This region was treated as a single a priori region of interest, so no new cluster search was run within it. Per-subject mean decoding in the transferred region was compared against chance and against Position 2's own value in the original cluster, using one-sample and paired *t*-tests, Wilcoxon signed-rank tests, Cohen's *dz*, and Bayes factors, for both MRR and ROC-AUC.

3.3 Results

Results are organized in three blocks addressing different levels of measurement: firstly, what listeners reported (behavioural results), secondly, how neural amplitude responded to stimuli (ERPs), and, thirdly, what could be decoded from the population pattern (decoding analyses). Within each block we first report the structured word sequence, then the contrasts against the random-sequence baseline, and finally the control analyses that bound the interpretation.

3.3.1 Behavioural Analyses

3.3.1.1 Sequence Tasks

Word Sequence: Sequence Task
During the Random sequences and the triplet sequences, participants had to indicate via keypress from time to time which syllable they heard last. The results show different response patterns and reaction times for Morphs and Standards, indicating different underlying perceptual or decisional processes.

Response categorization: a word-boundary advantage for Standards, a linear rise for Morphs

Presented responses for Standards, indicating correct identification of the *expected* syllable, dominated at every within-word position (P1 87.9 %, P2 81.2 %, P3 87.6 %; more frequent than every other category at every position, Wilcoxon signed-rank, Holm-corrected within position, all $p_{holm} < .001$). On Morph trials, which presented ambiguous input and thus had no correct identification, the *expected* proportion instead rose across position (25.9 % → 39.9 % → 54.5 %) and was the most frequent category at P2 and P3. However, at P1 it fell below *unexpected* (32.7 %, $p_{holm} < .001$). Per-position pairwise contrasts are shown in Figure 3.2, Panel A.

A trial-level binomial model (*expected* vs. rest ~ position × stimulus type; GEE with participant as the clustering unit and robust SEs) yielded a Position × Stimulus Type interaction (Wald $\chi^2(2) = 39.06$, $p < .001$).
Orthogonal polynomial contrasts within each stimulus type separated the two profiles: Standard trials showed a significant quadratic curvature ($b = 1.05$, $z = 4.11$, $p < .001$), but no linear trend ($z = -0.24$, $p = .81$), indicating a U-shaped profile with a word-medial dip for *expected* responses, whereas Morph trials showed a significant linear rise ($b = 0.62$, $z = 9.86$, $p < .001$; no quadratic component: $z = -0.21$, $p = .83$) of *expected* responses across positions. No other response category showed a significant Position × Stimulus Type interaction (all $\chi^2(2) < 4.8$, $p > .09$), so the effect is specific to *expected* responses. The parallel mixed-effects logistic model reproduced the same pattern for *expected* responses; it did not estimate stably for the three rarer response categories, so we report the GEE throughout this analysis.
The response-category profile of Standard P2 in the word-sequence, creating the word-medial dip for *expected* responses, matched the Random condition almost exactly (P2: 81.2 % vs. 81.2 %; $t(35) = -0.02$, $p = .99$, $BF_{01} = 5.6$, see Figure 3.2), whereas identification at the two word-boundary positions, P1 and P3, exceeded it (P1: $t(35) = 3.08$, $p = .004$; P3: $t(35) = 3.21$, $p = .003$). The reduced proportion of *expected* responses at Standard P2 in the word sequence therefore reflects an advantage at the boundary positions rather than a cost at the word-medial position.

Overall, the proportion of *expected* responses in Standard and Morph trials index different quantities and take different shapes: For Standards, the proportion indicates identification accuracy, which at the word-medial position did not exceed the Random-sequence level obtained without any sequence structure but was elevated above it at the two word-boundary positions. Responses not falling into the *expected* category fall into the mixed categories rather than the *unexpected*, suggesting reduced report precision rather than a shift toward the expectation-relevant competitor in Standard trials. For Morphs, *expected* responses may index the weight given to prior knowledge from word learning or triplet structure, relative to the ambiguous acoustics, and this proportion increases monotonically across the word.

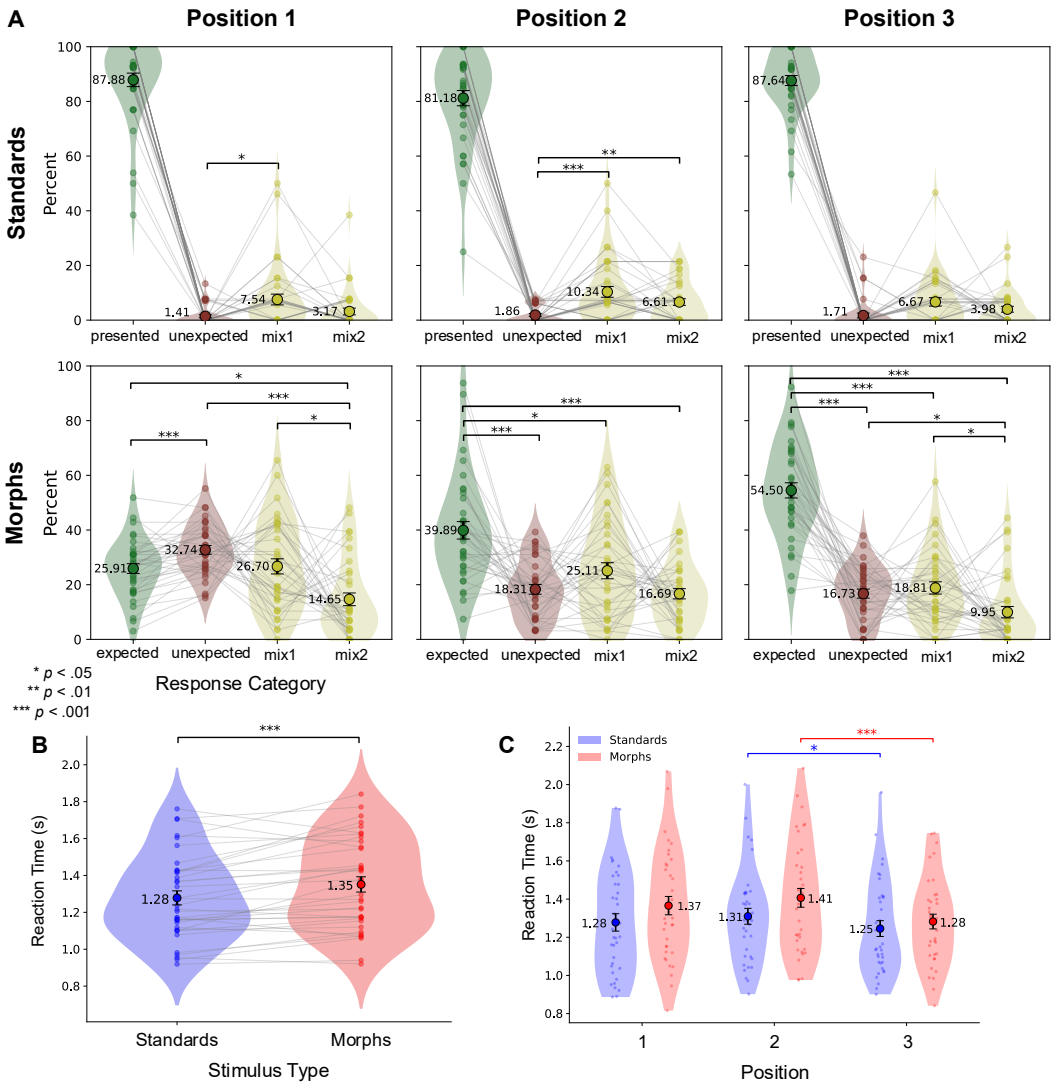


Figure 3.2. Sequence Task categorization and reaction times in the word sequence. Asterisks: * $p < .05$, ** $p < .01$, *** $p < .001$. Dots: participants, markers: group mean ± SEM. **(A)** Response-category proportions by within-word position for Standard (top) and Morph (bottom) trials. Brackets: pairwise Wilcoxon signed-rank tests between response categories within each position, Holm-corrected within panel. For Standard trials the *presented* (being also the expected and correct) response was more frequent than every other category (all $p < .001$; brackets omitted for legibility) and follows a non-monotonic P1 ≈ P3 > P2 pattern (87.9 % → 81.2 % → 87.6 %; P1 vs P2: $p = .008$, P2 vs P3: $p = .034$, P1 vs P3: *n.s.*), paralleling the ERP position effect. For Morph trials (no objectively correct response for an ambiguous stimulus) the *expected*-consistent rate rises monotonically across position (25.9 % → 39.9 % → 54.5 %; all pairwise $p < .001$), tracking the words' transition-probability structure; at P1 it lies even below the *unexpected* rate (32.7 %; $p < .001$). **(B)** Mean RT (s) per participant by stimulus type, collapsed across position. Morph (red, 1.35 s) responses are slower than Standard (blue, 1.28 s) in a LME on log-transformed single-trial RT (log(RT) ~ stimulus type × position + (1|participant)): $\chi^2(1) = 22.87$, $p < .001$; $t(35) = -3.93$, $p < .001$, $g = -0.31$. **(C)** Mean RT (s) per participant by within-word position and stimulus type. Brackets: within-stimulus-type Holm-corrected paired contrasts (P2 vs. P3: Standards: $p = .027$, Morphs: $p < .001$). P1 did not differ from P2 in either stimulus type; P1 vs. P3 was marginal for Morphs ($p = .071$) and null for Standards. The stimulus type × position interaction was not significant ($F(2, 70) = 1.15$, $p = .32$).

Morphs: faster expected responses at the word-final position

Standard-trial RTs were too sparse in the non-*expected* categories for trial-level modelling (7–10 *unexpected* and 15–35 *mix2* trials per position across all 36 participants combined; 26–32 and 16–27 participants, respectively, with no trials in a given position × category cell).

Therefore, RT analyses were restricted to Morph trials (a total of 3,057 trials for 36 participants). A linear mixed-effects model on log-RT (log(RT) ~ position × response category + participant random intercept) showed main effects of position ($\chi^2(2) = 34.39$, $p < .001$) and category ($\chi^2(3) = 24.94$, $p < .001$) and a Position × Category interaction ($\chi^2(6) = 15.66$, $p = .016$).

The interaction was localized to P3: relative to *expected* responses and relative to P1, *mix1*, *mix2* and *unexpected* were disproportionately slow at P3 (interaction b ≈ 0.11–0.13, all $p < .02$). Per-participant contrasts at P3 confirmed that *expected* responses were faster than *mix1* ($t(32) = -2.76$, $p_{holm} = .019$, $g = -0.50$) and *mix2* ($t(25) = -3.12$, $p_{holm} = .013$, $g = -0.73$), and non-significantly faster than *unexpected* ($t(32) = -1.89$, $p_{holm} = .068$). Within *expected* responses, RT did not fall gradually across position: P1 and P2 did not differ, but both were slower than P3 (orthogonal contrasts: linear $z = -4.69$, $p < .001$; quadratic $z = -3.11$, $p = .002$), which stands in contrast to the gradual rise in *expected*-response proportions.

Slower responses to Morphs, and a word-final reaction-time advantage

Reaction times, analysed at the participant level (2 × 3 repeated-measures ANOVA on mean log-RT), showed main effects of stimulus type ($F(1, 35) = 19.45$, $p < .001$) and within-word position ($F(2, 70) = 8.15$, $p_{GG} < .001$), but no interaction ($F(2, 70) = 1.15$, $p = .32$). Standard probes (1.28 ± 0.04 s) were responded to about 5 % faster than Morph probes (1.35 ± 0.04 s; $t(35) = -4.08$, $p < .001$, $dz = -0.68$).

The position effect showed a slowing at the medial relative to the word-final position rather than a monotonic gradient across positions. For both Standards and Morphs, responses were slower at P2 than at P3 (Standard: +5 %, $t(35) = 2.76$, $p_{holm} = .027$, $g = 0.29$; Morph: +10 %, $t(35) = 5.81$, $p_{holm} < .001$, $g = 0.49$); P1 versus P3 showed a non-significant trend (Standard: +3 %, $p_{holm} = .55$; Morph: +6 %, $p_{holm} = .071$), and P1 and P2 did not differ in either stimulus type (Standard: $p_{holm} = .55$; Morph: $p_{holm} = .12$; on *expected* Morphs trials P2 was numerically faster than P1). The trial-level mixed model with a per-participant random slope for position reproduced this pattern (position $\chi^2(2) = 27.5$, $p < .001$; P2 vs. P3: $p < .001$ in both types; P2 vs. P1: $p = .29$ for Standards and $p = .08$ for Morphs).

The absence of a stimulus type × position interaction indicates that the Morph slowing did not differ reliably across positions. Given the diverging position and stimulus-type effects reported so far, we nonetheless examined the Standard–Morph difference at each position (paired contrasts on participant cell means): Standards were faster than Morphs at P1 and P2 (P1: −5.4 %, $t(35) = -2.84$, $p = .008$, $BF_{10} = 5.4$; P2: −6.4 %, $t(35)$

= −3.19, p = .003, BF_{10} = 11.9) but not at P3 (−2.7 %, $t(35)$ = −1.63, p = .11, BF_{01} = 1.7). Processing at the word-final position thus converged for the two stimulus types (see Figure 3.2, Panel C).

Random Sequence: Response-Category Baseline and a Test for Implicit Learning

Sequence Task: Response-Category Proportions

In the Random blocks (R1-R2, no triplet structure, Standards only, no Morphs), *presented* responses dominated (M ± SEM: 81.2 ± 2.1 %), followed by *mix1* (10.9 ± 1.6 %), *mix2* (6.8 ± 0.8 %) and *unexpected* (1.1 ± 0.3 %). All six pairwise category comparisons were significant (Wilcoxon signed-rank; p_{holm} < .001 for every pair except *mix1* vs. *mix2*, p_{holm} = .010; see Fig. 3.3, Panel A). This mirrors the response profile for Standard syllables in the word sequence.

Hidden-Structure Block: Response Categories

The third block in the Random condition (R3) embedded the same triplet structure and Morphs as the word sequence, but without participants being instructed about it or any prior explicit word learning. Probe trials were sparse (~2.9 Standard and ~5.6 Morph trials per position and participant), so proportions are reported descriptively only.

The response profile broadly resembled the word sequence (Fig. 3.2): for Standards, presented responses dominated at every position (81–87 %), with unexpected ≤ 3.5 %; for Morphs, responses were distributed across all four categories, with *expected* showing a positive trend across position (23.7 % → 28.7 % → 30.6 %, Δ 6.9 pp) while *unexpected* stayed flat (~24–26 %). This trend is shallower than in the first word-sequence block (Δ 12.6 pp), which however followed explicit word training. Given the trial counts we can neither confirm nor exclude a position effect indicative of implicit statistical learning.

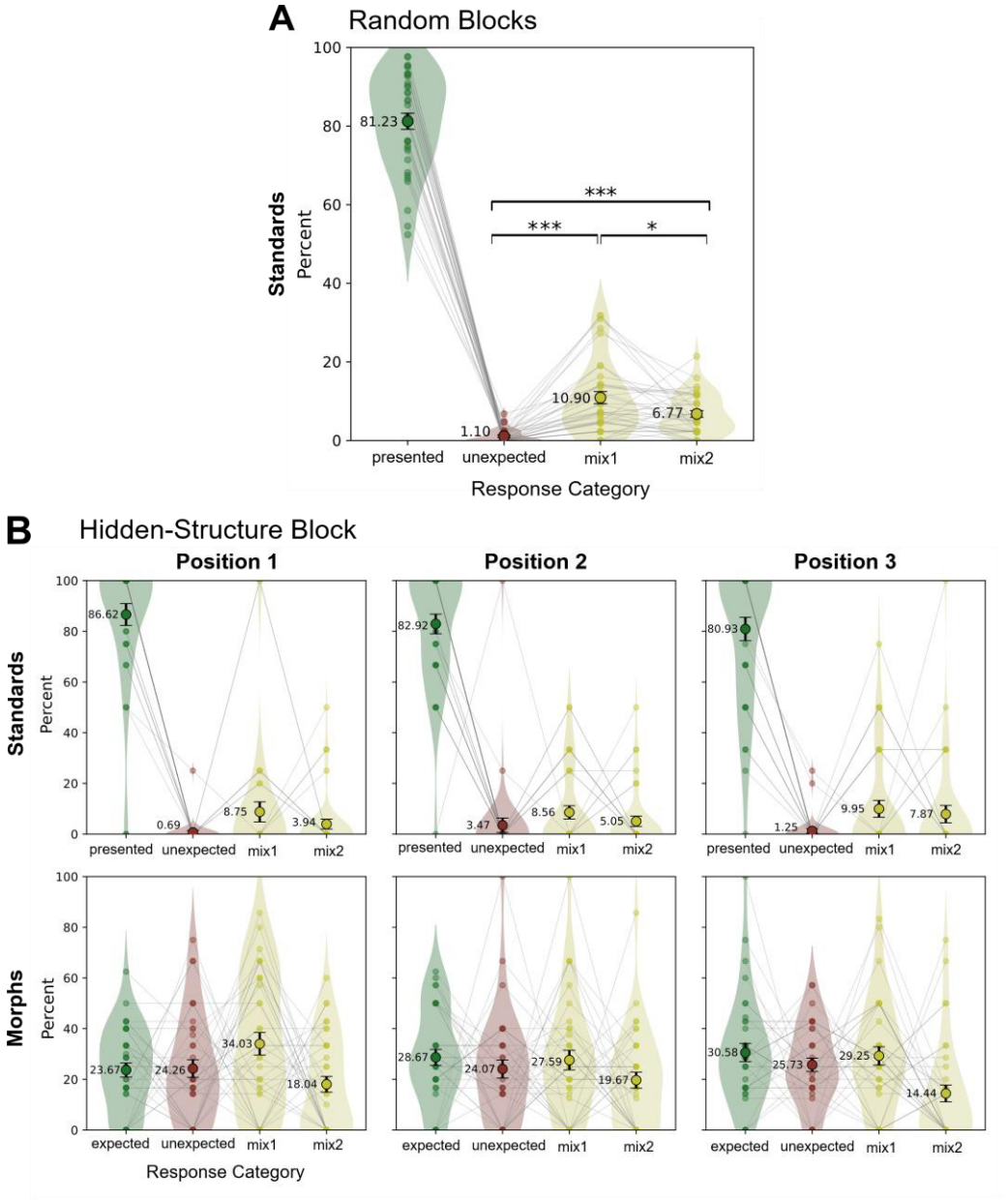


Figure 3.3. Random condition: Sequence Task responses. (A) Random-baseline blocks (R1-2) (no triplet structure, no Morphs). Sequence-task probe responses in the Random condition, Standards only; violins/dots show per-participant proportions, markers group mean ± SEM. Brackets: pairwise Wilcoxon signed-rank tests, Holm-corrected across all six category pairs (*** $p <$ .001, * $p <$ .05). *Presented* responses were more frequent than every other category (all $p_{holm} <$ .001; brackets omitted for legibility). **(B) Hidden-structure block (R3).** Sequence-task probe responses in the random-sequence block that was covertly word-structured but not preceded by any word learning or training, for Standard (top) and Morph (bottom) trials, by within-word position. Violins/dots: per-participant proportions; markers: group mean ± SEM. No significance testing (on average 2.9 / 5.6 Standard / Morph probe trials per position and participant). Descriptively, Standard *presented* rates decrease slightly across position (86.6 → 80.9 %) and Morph *expected* rates rise (23.7 → 30.6 %), paralleling the direction of the word-sequence position gradient (Figure 3.2, Panel A) but more weakly (Δ 6.9 pp vs. 12.6 pp in word-sequence Block 1). Axis labels as elsewhere: presented (Standard) / expected (Morph) = the syllable consistent with the triplet structure.

3.3.1.2 Word Learning: Retrieval Tasks and Continuation Task

Successful word learning

All neural and behavioural effects reported below presuppose that participants actually learned their six pseudowords. Three retrieval tasks administered at the end of the session, Free Recall, a two-alternative Familiarity Task, and a per-word Certainty rating, establish that they did, and the Continuation Task administered before each structured block provides a record of how that word knowledge developed over the course of the word sequence blocks.

Free Recall. Participants recalled on average 3.78 ± 1.74 (M ± SD, median = 4) of their six words. 19.4 % recalled all six, 2.8 % none, and 69.4 % at least half.

Familiarity Task. Recognition was near ceiling (M ± SD: 93.29 ± 7.41 %, 95 % CI [90.8, 95.8]) and, as a manipulation check, well above chance ($t(35) = 35.07$, $p < .001$, $d = 5.85$, $BF_{10} > 1000$; Wilcoxon: $W = 0$, $p < .001$); 41.7 % of participants scored 100 % and none below 75 %.

Certainty Ratings. The mean rating was 3.30 ± 0.63 (M ± SD, participant means 2.0–4.67). Across all 216 word ratings the full scale was used (0: $n = 7$; 1: $n = 26$; 2: $n = 38$; 3: $n = 32$; 4: $n = 51$; 5: $n = 62$), and participants differentiated between their own six words (mean within-participant range 3.4 points) rather than rating them uniformly.

Cross-task convergence. Free Recall performance correlated with mean Certainty ratings ($r = .53$, $p_{bonf} = .003$; $\rho = .51$, $p_{bonf} = .004$) but not with Familiarity Task accuracy ($r = .27$, $p_{bonf} = .339$; $\rho = .25$, $p_{bonf} = .431$). Likewise, Certainty ratings did not correlate with Familiarity Task accuracy ($r = .28$, $p_{bonf} = .306$; $\rho = .215$, $p_{bonf} = .622$) which might be due to the near-ceiling scores in the Familiarity Task.

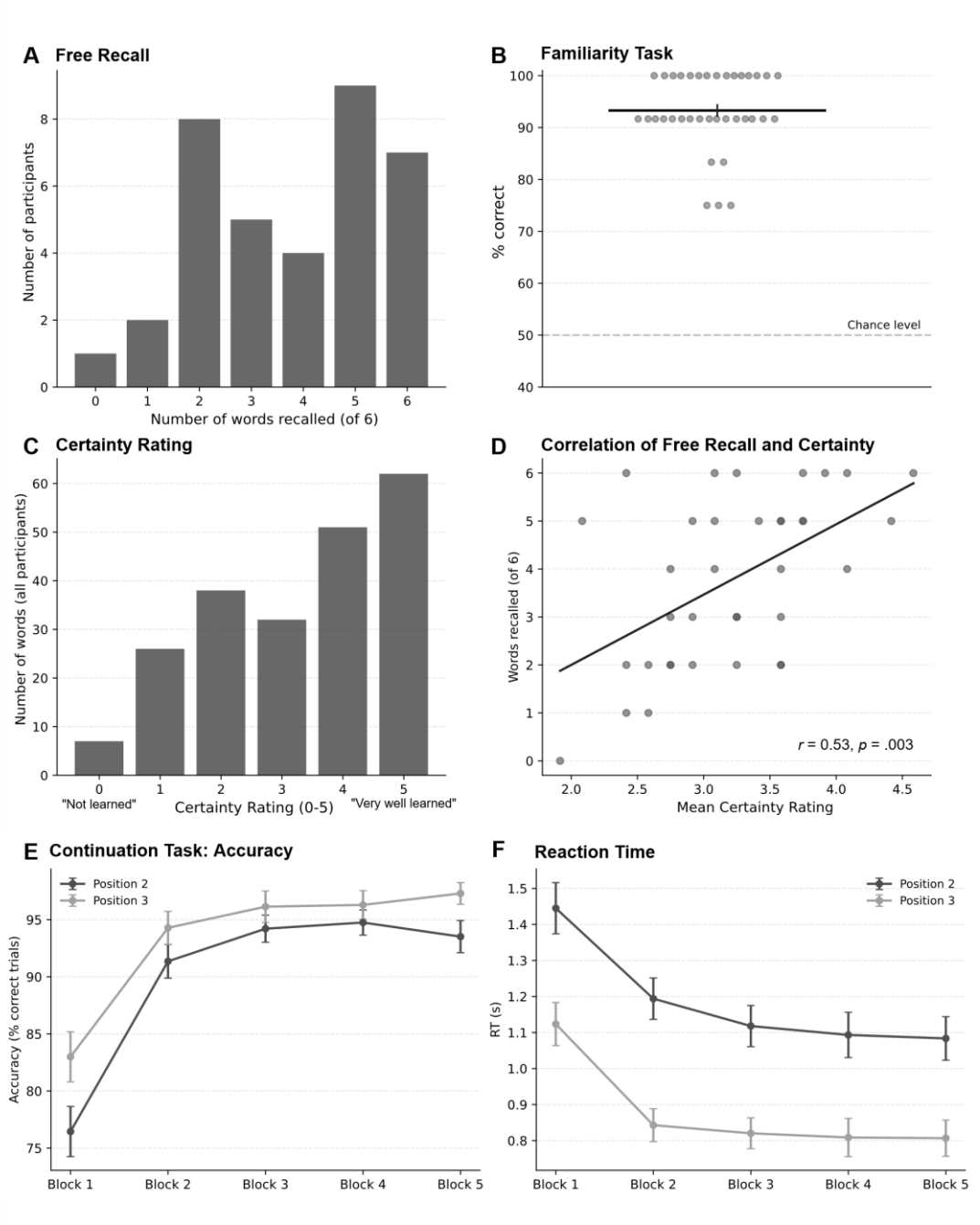


Figure 3.4. Behavioural evidence for successful word learning.
Retrieval tasks: (A) Free Recall: distribution of words (of 6) freely recalled per participant (*M* = 3.78, *SD* = 1.74; 19.4 % recalled all six). **(B)** Familiarity Task (2AFC recognition): per-participant accuracy, dashed line = chance (50 %); near ceiling performance (mean ± SEM, 93.3 ± 1.2 %), significantly above chance ($t(35) = 35.07$, $p < .001$, $d = 5.85$). **(C)** Certainty Rating: distribution of word-level ratings (0–5, all 216 participant × word observations), concentrated toward high ratings (word-level *M* = 3.30, *SD* = 1.50) but spanning the full scale. **(D)** Cross-task convergence: mean Certainty rating correlates with words recalled ($r = .53$, $p_{bonf} = .003$).
Continuation Task: (E) Accuracy and **(F)** reaction time (mean ± SEM) across the five training blocks, for P2 (predicting syllable 2 from syllable 1) vs. P3 (predicting syllable 3 from syllables 1–2). Both measures improved sharply from Block 1 to Block 2 and then plateaued (Block main effects: accuracy $p < .001$, RT $p < .001$); P3 was predicted more accurately and faster than P2 throughout, an advantage already present in Block 1 and thus not a product of learning (Position × Block interaction non-significant).

Continuation Task: better and faster prediction of syllable 3 than syllable 2

Prediction performance improved rapidly over word training in the Continuation Task which preceded each Word Sequence block (so Continuation Task blocks are numbered by the word block they precede, see Fig. 3.4).

In additive models, accuracy was higher in every block after Block 1 (b = 0.65–1.12, all $z \geq 5.2$, $p < .001$) and RT correspondingly shorter (b = −0.21 to −0.29, all $|z| \geq 9.7$, $p < .001$). Observed accuracy rose from 76–83 % (Block 1) to 91–97 % (Block 2 on), while observed RT fell from 1.12–1.44 s to 0.81–1.19 s. Across all blocks, participants were more accurate and faster at P3 than P2 (accuracy: b = 0.43, z = 3.26, p = .001; RT: b = −0.30, z = −21.5, $p < .001$). The Block × Position interaction was not significant (accuracy: $\chi^2(4) = 3.31$, p = .51; RT: $\chi^2(4) = 0.96$, p = .92), indicating that the P3 advantage was present from Block 1 and stable, instead of a learning product over time.

Word-specific learning strength predicts Morph categorization

Prediction training is linked to word-learning strength

A word's Continuation Task accuracy predicted its Certainty rating (b = 2.83, $p < .001$) and Familiarity accuracy (b = 0.29, p = .006); the Free Recall association was directionally consistent but weaker (b = 1.72, p = .096). The measures thus index a common word-specific learning signal.

Better-learned words had their Morphs categorized as expected more often

When the trial-level *expected* response was predicted from a word's Free Recall, Familiarity and Certainty scores, no measure predicted Standard-syllable categorization (all $p > .35$). Certainty showed a weak version of the effect for Morphs (b = 0.10, p = .024), suggesting that words rated more certain had their Morph syllables categorized more often as *expected*.

However, a more objective measure captured it far more strongly: On Morph trials, Continuation Task accuracy predicted the *expected* response (GEE: b = 0.40 per SD, z = 4.30, $p < .001$; b = 2.66 raw), with the mirror effect on *unexpected* responses (GEE b = −0.31 per SD, p = .028). Entered together with Continuation accuracy, Certainty no longer contributed (p = .45).

The effect of Continuation Task was a word-level effect within participants (within: b = 0.29 per SD, between: b = 0.41 per SD, both $p < .001$). In other words, among a participant's own six words, the ones they predicted better in the Continuation Task were the ones whose Morphs they more often heard as *expected* (and less as *unexpected*). This effect held whether predictability was measured overall or at the matching syllable position, and after removing the shared increase of both measures across training blocks. Simply put, this coupling of Continuation Task performance and Morph categorization tracks which words a participant knows better, independent of overall practice or the shared rise of both measures across training.

The coupling seemed specific to predictable positions

The word-level predictor, i.e. a word's Continuation Task accuracy averaged over its P2 and P3 probes, predicted Morph categorization as *expected* at P2 and P3 but not at P1 (b = 0.16 per SD, p = .28). The accuracy × position interaction was not significant (P2 vs P3: $p \approx 1.0$; P1 vs P2/P3: p = .065 / .074), but it was also low-powered (~4–5 Morph probes per participant × word × position). However, the direction nonetheless follows the design: at P1, the upcoming syllable is not predictable from context as word order is randomized, and is never probed by the Continuation Task, which only tests syllables at P2 and P3. That is, better word knowledge shifts Morph categorization where it supports and allows a specific prediction; where none is possible, better-learned words behave no differently. The effect thus acts through prediction of a particular upcoming syllable, not through a position-independent tendency tied to merely knowing the word.

For Standards, both cross-task links were between participants, not between words

On Standard probes (P2, P3), accuracy in the Continuation Task predicted a participant's overall accuracy (i.e. rate of *expected* responses) (between-participant component: b = 0.39 per SD, p = .001) but not the relative standing of their own words (within: b = 0.07 per SD, p = .43): participants who were more accurate in the Continuation Task were also more accurate on the Standard probes.

The RT-RT link (Continuation-RT → probe-RT) was likewise entirely between-participant on both Standard and Morph probes (between: b = 0.08 per SD, $p < .001$; within: b = 0.011, p = .45/.41), while Continuation Task accuracy predicted probe RT on neither (p = .11/.77), indicating a shared response-speed trait across two timed forced-choice tasks, resolving no individual words.

Consistent with the near-ceiling Standard *expected* rate, the only word-level cross-task effect in this analysis is the one on Morph categorization.

Ruling out a general engagement confound

A participant who is more engaged throughout the session could plausibly show higher Continuation Task accuracy, higher scores in all retrieval tasks, and a greater tendency to give the *expected* response in the Sequence Task; or in short, producing the associations reported above without any word-specific mechanism.

Two features of the data argue against this: First, if participant-level engagement is assumed constant within a session, it cannot drive the within-participant component of the Morph accuracy effect (i.e., that better predicted words were more often categorized as *expected*, b = 0.29 per SD, $p < .001$). This contrasts each participant's better- vs. worse-learned words against their own mean and thereby partials out anything constant within a participant, such as motivation, general task aptitude, response bias, or rating-scale use.

Second, we exploratorily added four explicit participant-level engagement proxies to the model of the *expected* proportion on Morph probes (Continuation Task miss rate, mean RT on the Word Sequence Task probes, Sequence Task accuracy in the pre-learning Random blocks R1–R2, and mean Certainty rating), and found that they left

the pooled Continuation Task-accuracy coefficient essentially unchanged (b = 2.66 → 2.62). None of the four proxies predicted the *expected* proportion on its own, although two were themselves correlated with overall Continuation Task accuracy (pre-learning Random-block accuracy r = .51; mean Certainty rating r = .58).
Together these indicate that the behavioural effect on Morph categorization is word-specific and not an artefact of between-participant engagement differences.

These analyses do not, however, localize the word-specific bias. It could be perceptual, i.e., knowing a word better makes its ambiguous Morph sound more like the expected syllable, or decisional, i.e., the Morph is heard as ambiguous, but participants weigh their expectation more heavily when choosing a response for words they have learned better.
The within-participant decomposition excludes only a participant-constant response tendency; a word-by-word decisional bias of this kind is not separable from a perceptual effect on behavioural data alone (see Section 3.4.8).

Phoneme-specific response analysis: tendency towards voiceless consonants
To quantify potential perceptual biases related to syllable constituents, we analysed Sequence Task performance separately for consonants and vowels.

On Morph trials the *unexpected* response was chosen more often than the *expected* response at P1 (32.7 % vs. 25.9 %; +6.8 pp; $t(35) = 4.56$, $p < .001$), which is the reverse of the ordering at P2 and P3. Decomposing each response into its consonant and vowel, both dimensions leaned toward the unexpected phoneme by a similar amount (consonant: 46.3 % "expected", $t(35) = -2.96$, $p = .005$; vowel: 46.9 %, $t(35) = -3.57$, $p = .001$).
For consonants, the lean was toward the voiceless member of each consonant pair, chosen on 59.7 % of trials ($t(35) = 2.55$, $p = .015$) and more so when the expected consonant was voiced, i.e. when the voiceless choice opposed the expectation (63.1 % vs. 56.4 %; $t(35) = 2.75$, $p = .009$). Accordingly, the reversal reached significance only in trials with a voiced expected consonant (+14.2 pp, $t(35) = 2.40$, $p = .022$; voiceless-expected: −0.9 pp, p = .880), although the direct contrast was not significant ($p = .160$). Descriptively the effect was graded across pairs (/d/–/t/ +19.9, /b/–/p/ +12.1, /g/–/k/ +11.0 pp) but only /d/–/t/ was individually reliable ($p = .030$), while the pairs did not differ significantly (Friedman $p = .300$). The same voiceless preference appeared at P2 and P3, where it aligned with the position-related expectation and produced a larger shift toward *expected* for voiceless- than voiced-expected trials (P2: −35.3 vs. −8.3 pp, $p = .003$; P3: −48.6 vs. −24.2 pp, $p = .028$).
For vowels, two pairs carried a large expectation-independent favourite: "e" over "i" (85.2 %, $p < .001$) and "a" over "ä" (66.1 %, $p = .016$), which, being orthogonal to the expectation, cancel when pooled. Only o/u showed a genuine net lean toward the unexpected vowel ($t(35) = -2.81$, $p = .008$), which cannot be attributed to the Morph specific idiosyncrasies as it runs in both directions, and thus accounts for the entire vowel-level contribution to the reversal.

The consonant and vowel biases were not independent: on a subject-clustered cross-check the /d/–/t/ effect was stronger in the e/i than the a/ä vowel context (p = .019). However, we cannot conclude whether these phoneme-related findings reflects a perceptual interaction between segments or rather acoustic idiosyncrasies of the particular Morph sounds.
Reaction times were consistent with a fast, low-effort bias rather than an effortful override: within voiced-expected trials, *unexpected* responses were faster than *expected* ones (1.32 vs. 1.62 s, $t(27) = -2.93$, p = .007, for n = 28 participants with trials in both cells).

Overall, the Morph P1 response reversal is robust and reflects small biases on both phoneme dimensions. For consonants it is a combination of a stable, largely position-independent preference for the voiceless member of each pair (chosen ~60 % of the time at every position) and an expectation effect that is weak at P1, so the fixed preference dominates and produces the *unexpected* surplus, but strong at P2 and P3, where the response instead tracks which consonant is predicted. While we cannot show that these biases fully account for the reversal, given the available power, or localize the consonant bias to a specific pair, we report these findings as a primary candidate contributing to the effect observed at Morph P1.

3.3.2 Univariate EEG Analyses (ERPs)

ERP effects of Stimulus Type and Within-Word Position

Figures summarize ERPs per contrast family in one panel (see Methods for cluster-selection and visualization conventions); all clusters are shown individually in the Supplement. For better differentiation of clusters, we report p-values with four decimal places.

ERPs per Stimulus Type

Standard < Morph. Morph stimuli elicited a larger response than Standard stimuli (Morph > Standard) in two topographically mirrored fronto-posterior cluster pairs spanning most of the epoch: an early pair (105–290 ms; frontal: p = .0024; posterior: p = .0038) and a late pair (290–500 ms; frontal: p = .0008; posterior: p = .0008). This difference is likely driven by an increased processing load for ambiguous relative to intelligible syllables.
Standard < Random. Standard ERPs were suppressed relative to Random (Random > Standard) in an early cluster (40–280 ms, p = .0250) which did not survive FWE correction, followed by a significant late mirrored cluster pair (290–485 ms; frontal: p = .0044; posterior: p = .0010). This indicates sustained expectation suppression of predictable Standard stimuli relative to unpredictable Standards.
Morph > Random. Morph ERPs were significantly enhanced relative to Random (Morph > Random) in an early cluster (115–275 ms, p = .0257) which did not survive FWE correction, followed by a significant late mirrored cluster pair (340–500 ms; posterior: p = .0059; frontal: p = .0021). This suggests that the heightened processing

load required to resolve ambiguous syllables exceeds the baseline processing of unpredictable but intelligible syllables.

Taken together, the three contrasts are jointly consistent with an amplitude ordering of Morph > Random > Standard, with all differences being significant after FWE from approximately 300 ms onward.

ERPs per Position

P1 > P2. P1 differed from P2 in a mirrored cluster pair extending over most of the epoch (80–500 ms; frontal: $p = .0019$; posterior: $p = .0007$; P1 > P2). This is consistent with the absence of transition-based expectations at word onset, relative to the expectation suppression engaged at the second position.

P2 < P3. P2 differed from P3 predominantly in the late window (215–500 ms; posterior: $p = .0032$; frontal: $p = .0016$; P3 > P2), underscoring the functional shift from active expectation suppression at P2 to word completion and triplet closure processes at P3.

P1 ≥ P3. The P1-P3 contrast produced an early, mirrored structure (95–250 ms; frontal: $p = .0389$; posterior: $p = .0352$) for P1 > P3 that did not survive FWE correction under a fixed cluster-forming threshold. TFCE, added as a sensitivity check, showed a significant early window (~115–205 ms, corrected $p = .009$). On either analysis, any P1-P3 difference is restricted to an early, plausibly stimulus-locked window and is weaker than the P1-P2 and P2-P3 difference. The positional pattern is thus P1 ≈ P3 in the late window. The absent (or early-only) P1-P3 difference does not imply the two positions are equivalent: as the contrasts against the Random baseline show below, P1 and P3 differ in their relationship to that baseline.

P1 ≈ Random. P1 showed no clusters at any threshold relative to Random. We therefore found no evidence for predictive modulation of the response to word-initial syllables, as would be expected if these are not constrained by a preceding within-word transition.

P2 ≤ Random. P2 was suppressed below Random in an early time window during stimulus presentation (40–240 ms, posterior, $p = .0233$) that did not survive FWE correction, followed by a significant time window in the second half of the trial (255–500 ms; frontal: $p = .0024$; posterior $p = .0014$). This topographically mirrored the Morph vs. Standard difference, reflecting an expectation suppression effect for the predictable second syllable.

P3 > Random. P3 was enhanced above Random in a single broad central cluster (250–500 ms, $p = .0010$). Unlike the P2 suppression, a fronto-posterior dipolar pattern that mirrors the Morph-minus-Standard effect (see above), the P3 enhancement formed a single, broadly distributed central positivity with no mirrored frontal counterpart (Figure S1).

Stimulus Type × Position Interaction. When directly contrasting the Morph-minus-Standard difference waves across position pairs, a candidate cluster emerged for the contrast between the two positions at which the separate analyses diverge most, namely P2 vs. P3 (130–280 ms, 22 central channels, $p = .031$), in the expected direction (Morph > Standard at P2 reversing to Standard > Morph at P3) and with a

topography resembling that of the position effect. Given its directional and topographic consistency with the established position effect, but its failure to survive FWE correction (p_{bonf} = .092, α = .0167), we treat the interaction as unresolved rather than as absent.

When interpreting this cluster, two observations should be kept in mind: Because the test operates on the difference of difference waves, the direction (Standard > Morph at P3) means that the Morph-minus-Standard difference is smaller at P3 than at P2 in this early window, not that Standard amplitudes exceed Morph amplitudes at P3; no contrast in this dataset shows the latter. Also, unlike the mirrored fronto-posterior cluster pairs seen for the stimulus-type main effect and for the pairwise position contrasts, the interaction produced a single central cluster, the same configuration as the independently obtained P3-versus-Random effect (see Figure S1); 15 of its 22 channels (68 %) also fall within that 28-channel cluster, although the time windows overlap only partially (130-280 ms for the interaction here against 250-500 ms for P3 vs Random there), so the interaction precedes rather than coincides with the established P3 effect.

We also note that the six cell-versus-Random contrasts reported below do not depend on this interaction test: each cell is tested against the Random baseline as an external unpredictable floor rather than against the other cells, so they are not simple-effects follow-ups of this interaction and are neither licensed nor withheld by it.

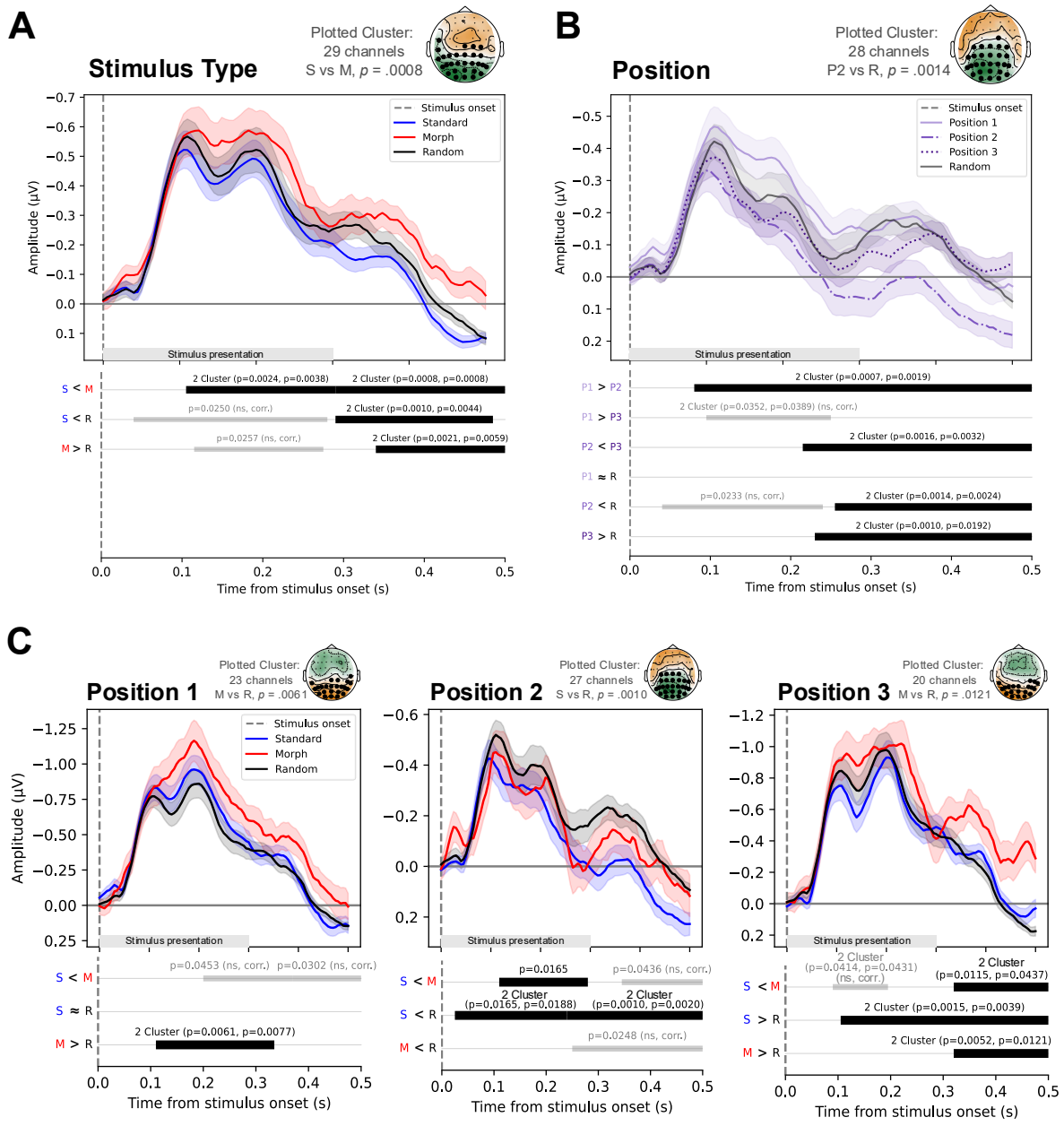


Figure 3.5. Stimulus-type, position, and position-within-stimulus-type ERP effects.
(A) Grand-average ERPs (posterior channels, *n* = 29) for Standard (blue), Morph (red), and Random (black), mean ± SEM. Topography: difference wave at the reference cluster's peak (Standard vs. Morph, 445 ms, *p* = .0008). Below: pairwise cluster tests (α = .0167). Operators (<, >, ≈) per row were added to clarify directions of effects, but cluster tests were two-tailed. Morph exceeded both Standard and Random from ~100 ms onward (early and late clusters, both *p* ≤ .0038); Standard and Random converged early, diverged only late (*p* ≤ .0044). **(B)** Same conventions for P1/2/3 (light/medium/dark purple) vs. Random, posterior channels (*n* = 28). Topography: P2-vs-Random reference cluster, 320 ms, *p* = .0014. P1 exceeded P2 throughout (p ≤ .0019); P2 was suppressed below Random from ~250 ms (*p* = .0014); P3 was enhanced above Random from ~250 ms (*p* = .0010); P1 did not differ from Random at any latency. **(C)** Standard vs. Morph vs. Random per within-word position (posterior channels; *n* = 23/27/20; each position: α = .0167). P1: Morph is enhanced above Random (110–335 ms, *p* = .0061), Standard unaffected. P2: Standard is suppressed below Random (35–240 ms, *p* = .0165; 240–500 ms, *p* = .0010) while Morph does not differ from Random but is enhanced above Standard (110–280 ms, *p* = .0165). P3: both Standard and Morph are enhanced above Random (Standard: 105–500 ms, *p* = .0015; Morph: 320–500 ms, *p* = .0121), diverging from each other only late (*p* = .0115, Morph > Standard).

ERPs per Stimulus Type and Position

P1 *(Figure 3.5, left panel; reference topography: Morph-vs-Random, posterior cluster, p = .0061).*

P1: Standard ≈ Random. Standard ERPs at P1 showed no difference compared to the Random baseline. We thus found no evidence that intelligible, contextually unconstrained word-initial syllables were processed differently in the structured and unstructured sequence contexts.

P1: Morph > Random. Morph ERPs were significantly enhanced compared to the Random baseline during stimulus presentation (110–335 ms) across a topographically mirrored pair of clusters (posterior, 23 channels, p = .0061; frontal, 31 channels, p = .0077). This likely reflects an increased processing effort for ambiguous sensory input at the unpredictable initial within-word position.

P1: Standard ≤ Morph. The contrast between Morph and Standard ERPs revealed two posterior-only clusters in a later window (200–370 ms, p = .0453; 370–500 ms, p = .0302). Although neither cluster survived FWE correction, this numerical trend aligns with the main effect of Morphs exceeding Standards (shown in panel A) and likely reflects heightened ambiguity-driven processing load.

P2 *(Figure 3.5, middle panel; reference topography: Standard-vs-Random, posterior cluster, p = .0010).*

P2: Standard < Morph. Morph ERPs significantly exceeded **Standard** ERPs during stimulus presentation (110–280 ms, p = .0165, fronto-central topography), characteristic of a typical auditory mismatch response. Although the second cluster after stimulus offset did not survive family-wise correction (345–500 ms, p = .0436), Morphs and Standards were distinguishable at P2, at least in an early window, consistent with ambiguity-related responses superimposed on the expectation suppression seen for Standard stimuli.

P2: Standard < Random. Standard ERPs were suppressed relative to the Random baseline across two merged time windows: an early cluster extending over 25–240 ms combining a fronto-central cluster (p = .0188, non-significant after FWE correction) and a posterior-temporal cluster (35–240 ms, p = .0165), and a second, more robust window spanning the latter half of the trial (240–500 ms), combining a frontal (p = .0020) and a posterior (p = .0010) cluster. While the early cluster should be interpreted with caution given its threshold significance, it aligns with an expectation suppression effect, wherein predictable Standard stimuli elicit reduced neural activation. This suppression effect becomes substantially more pronounced in the second window.

P2: Morph ≤ Random. ERPs exhibited a trend toward a more negative amplitude relative to the Random baseline over a single left frontotemporal cluster during the second half of the trial (250–500 ms, 14 channels, p = .0248, non-significant after FWE correction). This deflection may reflect a late mismatch response or heightened processing load associated with resolving ambiguity in speech-relevant areas. During stimulus presentation, Morph ERPs did not differ reliably from the Random baseline, unlike at P1 and in the hidden-structure block. While the corresponding interaction did

not survive correction, this can be interpreted as suggestive evidence that structural expectation attenuates the baseline-relative elevation of ambiguous stimuli.

P3 *(Figure 3.5, right panel; reference topography: Morph-vs-Random, posterior cluster, p = .0121).*

P3: Standard < Morph. When contrasting Standard and Morph ERPs, early amplitude differences did not survive FWE correction (90–195 ms, parietal p = .0414, frontal p = .0431). We therefore found no reliable early difference between Morphs and Standards at this position. This would be consistent with structural expectation attenuating early sensory differences across stimulus types. However, following stimulus offset, Morph amplitudes significantly exceeded Standard ERPs (320–500 ms, merging a significant frontal p = .0115 cluster with a non-significant posterior p = .0437 cluster), suggesting that ambiguity-related processing load is superimposed on top of the triplet-closure effect.

P3: Standard > Random. At P3, Standard ERPs exceeded the Random baseline in a sustained window across two complementary clusters: an early-onset right fronto-temporal cluster (105–500 ms, merging a p = .0039 and a p = .0015 cluster) and a later parieto-central cluster in the second half of the trial (250–500 ms, 27 channels, p = .0015, see Figure S1 for a separate visualization). Because P3 represents the deterministic boundary of a triplet — where sequence predictability reaches its maximum before resetting — this late positive component likely reflects triplet completion and word closure.

P3: Morph > Random. Morph ERPs were likewise significantly enhanced compared to the Random condition following stimulus offset (320–500 ms; posterior, 20 channels, p = .0121; frontal, 26 channels, p = .0052). Paralleling the Standard condition at P3, this supports the notion that the late trial effect is position-specific and tied to triplet closure.

ERP Hidden-Structure Block

The last block of the Random condition followed the triplet structure and included Morphs – unbeknownst to the participants.

Standard < Morph. The direct contrast between Standard and Morph syllables exhibited a two-window, topographically mirrored pattern similar to the main structured sequence effect: Morphs elicited a larger response than Standard syllables in an early pair of clusters (~95–200 ms; frontal: p = .0369; posterior: p = .0355) that did not survive FWE correction, and a late pair of clusters (300–500 ms; frontal: p = .0095; posterior: p = .002). This pattern indicates a robust late ambiguity response even in the absence of explicit sequence awareness.

Standard ≈ Random. Standard ERPs showed no clusters relative to the Random baseline, whether corrected or uncorrected. We therefore found no evidence for expectation suppression of Standard syllables in this block.

Morph > Random. Morph ERPs differed significantly from the Random baseline in a mirrored early/late pattern parallel to the direct Morph vs Standard contrast. This included a significant early window (100–285 ms, comprising three temporally overlapping sub-clusters; posterior: p = .0019; frontal: p = .0214 and p = .0282) and a

significant late window (295–500 ms; posterior: $p = .0019$). Ambiguous stimuli therefore elicited an enhanced response relative to unambiguous ones in this block as well, that is, without explicit knowledge of the sequence structure.

Positions. We found no evidence for differences between the three positions, or between any position and the Random condition.

Control Analyses

Hidden word block and trial counts. The absence of positional effects in the hidden-structure block was not determined purely by trial count. N-matched single structured blocks (W1–W4) exhibited consistent nominal trends in three of the four blocks for each contrast, even though none survived FWE correction: P2 vs. P3 in W1–W3 ($p = .031$–$.040$; W4: $p = .423$) and P1 vs. P2 in W1, W3 and W4 ($p = .051$–$.066$; W2: $p = .409$). In contrast, nominal trends in the hidden-structure block were substantially flatter (P2 vs. P3: $p = .583$; P1 vs. P2: $p = .084$).

Despite an identical sample size (~102 trials/position), W5 yielded highly robust, significant position contrasts across all pairwise comparisons ($p \leq .0043$). While this is likely due to accumulated learning, we cannot distinguish it from time-on-task effects, since block number in this design confounds both.

Morph-Task Trial Bias. The Morph > Standard effect was not attributable to task-relevant probe bias: restricted to the ~72 % of Morph trials never followed by a probe (M = 62 vs. 86 trials/position for the full sample), the effect remained significant across all four original clusters, with closely matching topography and timing (early: 95–280 ms, $p = .0026$, 32 channels; late: 320–500 ms, $p = .0019$, 25 channels; original: ~105–290 ms, $p \leq .004$ and ~290–500 ms, $p \leq .0008$, respectively).

Random Sequence Trial-Count Matching. The Standard-P2 suppression was likewise not attributable to the trial-count imbalance relative to the Random condition: the effect remained significant in every one of 200 trial-count-matched iterations (200/200, 100 %; median cluster $p = .0005$), closely matching, or even slightly exceeding, the original, full-sample result ($p = .0010$).

Baseline Correction Control. Restricting the analysis to the nominal baseline window itself (on uncorrected epochs) revealed a significant position-dependent difference (P1 vs. P3: $p = .002$; P2 vs. P3: $p < .001$; P1 vs. P2: $p = .093$, non-significant after FWE) and a significant stimulus-type difference (Morph vs. Standard: $p < .001$), confirming that some ERP overlap from the preceding syllable is present at this SOA even within the window used for baseline correction. Critically, however, re-running the complete cluster-permutation battery without baseline correction left every originally significant effect intact. The stimulus-type main effect and the P1-vs-2 and 2-vs-3 contrasts all remained significant without baseline correction, and in several cases with numerically stronger statistics (e.g., P1 vs. P2: $p = .0007$ with baseline correction vs. $p = .0001$ without). This indicates that these effects do not depend on the specific baseline choice.

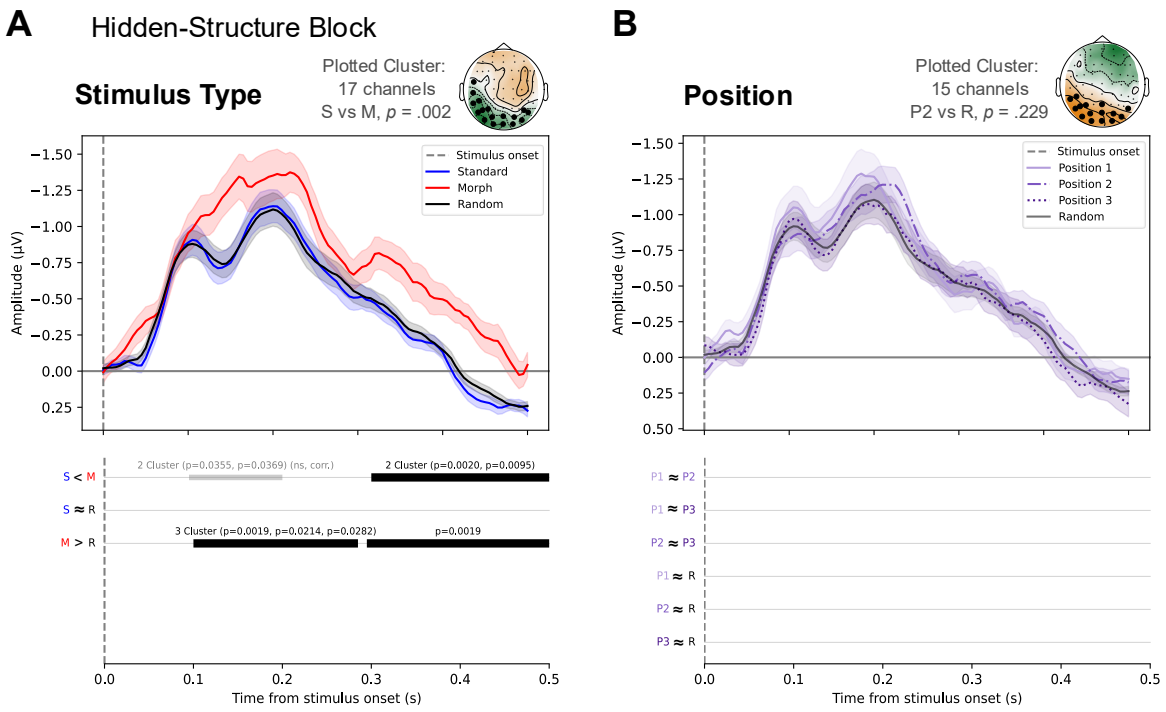


Figure 3.6. ERPs per Stimulus-type and within-word position in Hidden-Structure Block. **(A)** Grand-average ERPs (posterior channels, *n* = 17) for Standard (blue), Morph (red), and Random (black) syllables, mean ± SEM. Topography shows the difference wave at the peak latency (445 ms) of the reference cluster (Standard vs. Morph, posterior, *p* = .0020). Below: cluster-based permutation results for all three pairwise comparisons (3-test family, α = .0167); black bars = corrected-significant clusters, grey = *n.s.* after correction, each labelled with its own *p*-value. Morph elicited a larger response than both Standard and Random from ~300 ms onward (both p ≤ .0020); Standard and Random did not differ.
(B) Same conventions for P1/2/3 (light/medium/dark purple) vs. Random, posterior channels (*n* = 15). No comparison reached significance, even uncorrected. Topography shows the P2-vs-Random contrast (from the word sequence figure's reference comparison, retained here for cross-block comparability) at its peak (235 ms), non-significant here (*p* = .229): the P2 suppression effect does not generalize to the Hidden-Structure Block.

3.3.3 Multivariate EEG Analyses (Decoding)

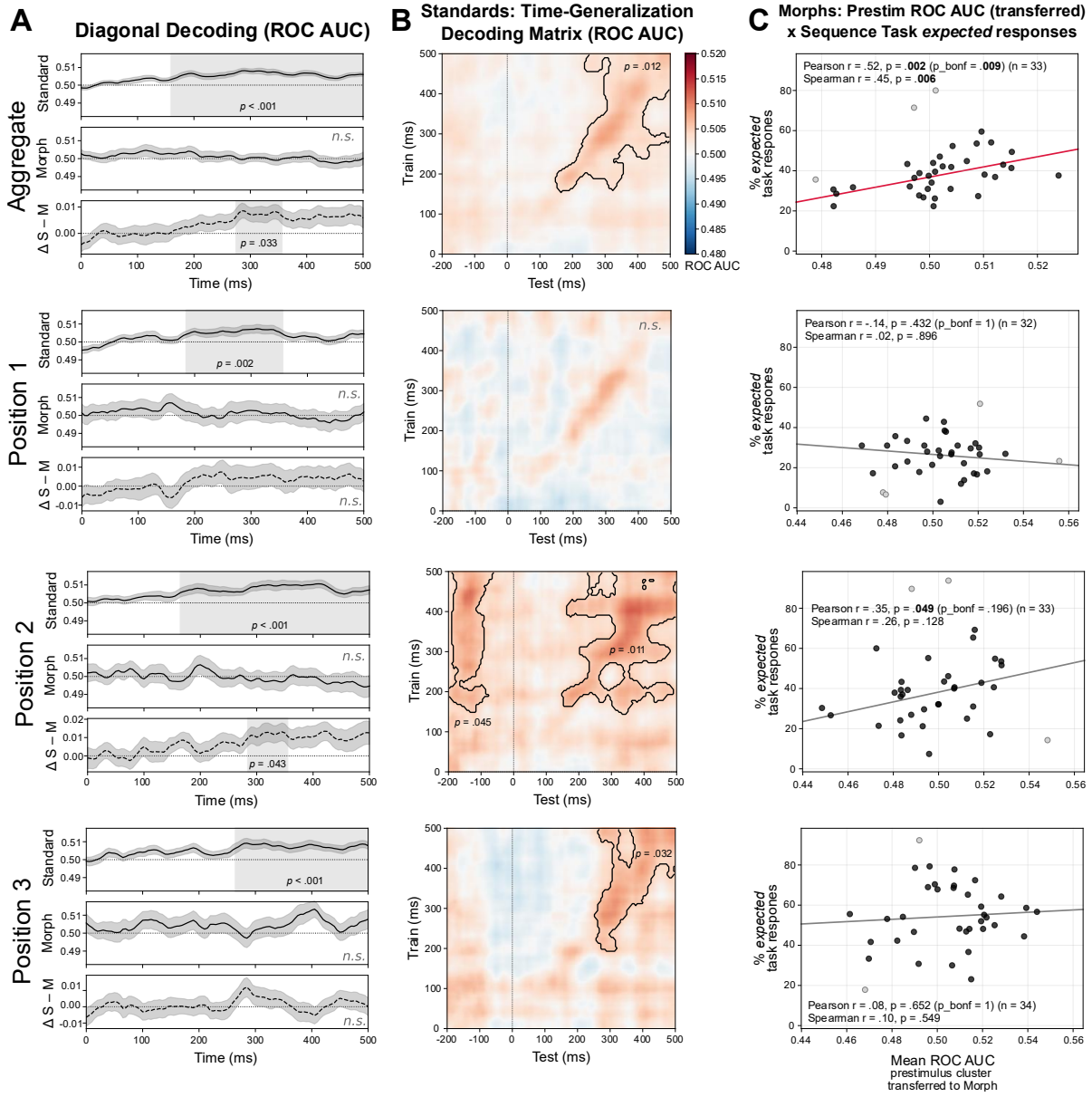


Figure 3.7. Windowed time-generalized (TG) decoding of Standard syllables and its relationship to Morph-trial behaviour (AUC).

Rows: aggregate and within-word P1–3. **(A) Left column:** diagonal (train = test) decoding time course for Standard (top), Morph (middle) trials and their difference wave (bottom), mean ± SEM, dotted line = chance (.5); cluster-permutation-significant windows shaded grey. Standard decoding is significant at every position (all $p \leq .002$); Morph decoding never differs from chance; the Standard−Morph difference reaches significance in aggregate ($p = .033$) and at P2 ($p = .043$) only. **(B) Middle column:** Time-generalization matrix of Standards (train 0–500 ms × test −200–500 ms), colour = AUC, black contour = cluster-permutation-significant region(s). A significant on-diagonal (post-stimulus) cluster is present in aggregate, at P2, and at P3 (not P1); P2 uniquely shows a second, pre-stimulus-generalizing cluster (test < 0 ms, $p = .0453$). **(C) Right column:** coordinates of the pre-stimulus cluster at Standard P2 (from the middle column) were applied to the Morph matrix (not shown) of each position/aggregate (per row); mean AUC of Morph cluster correlated per participant with their proportion of *expected* responses in the Sequence Task per position/aggregate (grey points = outliers excluded). Significant correlation after FWE correction (red line) only in aggregate (Pearson $r = .52$, $p_{bonf} = .009$); P2 shows a weaker, uncorrected trend ($r = .35$, $p = .049$); no relationships in P1 and P3.

Decoding of Standard Stimuli: better for later positions and a pre-stimulus cluster at the medial position

Using the 1D diagonal of the AUC matrices, we could decode the presented syllable in Standard trials of the word sequence both in aggregate and split by position (overall: 160–500 ms, $p < .001$; P1: 185–355 ms, $p = .002$; P2: 165–500 ms, $p < .001$; P3: 265–500 ms, $p < .001$), whereby the cluster identified for P1 spanned a shorter interval than those identified for the other positions. For Morph trials, the 1D decoding time courses of the expected component did not reveal any clusters (see Fig. S3 for decoding of unexpected components). Indicated by difference waves, the AUC values of Standard trials in aggregate ($p = .033$, 278–358 ms) and for P2 ($p = .043$, 288–353 ms) were significantly higher than those of Morph trials.

In 2D time-generalization AUC matrices, we identified a significant above-chance cluster in aggregate (cluster $p = .0118$, mean AUC = .504), spanning train times of 155–500 ms and test times of 150–500 ms. Decoding was successful at the level of positions: Mean decoding performance was comparable at P2 and P3 (both AUC = .507) while the clusters differed in extent (P2: $p = .0114$, train 145–500 ms, test 150–500 ms; P3: $p = .0321$, train 195–500 ms, test 255–500 ms). P1 did not reach significance for AUC in this window.

At P2, a second, off-diagonal cluster was also significant ($p = .045$), in which classifiers trained on post-stimulus activity (train time 150–500 ms) generalized to pre-stimulus test data (test time -200 to -65 ms), indicating that pre-stimulus neural activity already carried a signal resembling the code the classifier later relies on to decode syllable identity post-stimulus. Corroborating this finding, the same analyses run for time-generalized MRR matrices revealed a similar cluster ($p = .027$).

Linking neural information to behaviour: Pre-stimulus stimulus information correlates with expected responses

We wanted to test whether decoding performance of this pre-stimulus cluster was associated with behavioural measures of perception, such as performance in the Sequence Task during the word sequence. As the response variance of Sequence Task performance in Standard trials was low (>80 % *expected* responses), we transferred the exact cell coordinates to Morph 2D AUC Decoding matrices to then compute a mean pre-stimulus AUC value for Morph trials. To test whether an effect would not only show at P2 but possibly for other positions as well, we transferred the cluster from Standard P2 to all Morph Positions and aggregate. Importantly, participants did not know before stimulus onset that they would hear a Morph, so expectations before stimulus onset should correspond between Standard and Morph trials.

Aggregated across positions, mean pre-stimulus AUC values were positively correlated with the proportion of expected responses in Morph trials (Pearson $r = .515$, $p = .002$, $p_{bonf} = .009$, $n = 33$ [3 outliers excluded], 95 % bootstrap CI [.306, .695]; Spearman $\rho = .449$, $p = .006$, $n = 36$, 95 % CI [.165, .659]). P2 showed a weaker

relationship which did not survive FWE correction (Pearson r = .345, p = .049, p_{bonf} = .196, n = 33 [3 excluded], 95 % bootstrap CI [.016, .609]; Spearman ρ = .258, p = .128, n = 36, 95 % CI [−.094, .582]). P1 and P3 did not indicate a relationship (P1: Pearson r = −.144, p = .432, n = 32 [4 excluded], 95 % bootstrap CI [−.409, .118]; Spearman ρ = .023, p = .896, n = 36, 95 % CI [−.327, .357]. P3: Pearson r = .080, p = .652, n = 34 [2 excluded], 95 % bootstrap CI [−.197, .347]; Spearman ρ = .103, p = .549, n = 36, 95 % CI [−.233, .410]).
Averaging over the entire pre-stimulus AUC window rather than the transferred cluster from Standard P2 reproduced this pattern almost exactly: the aggregate correlation remained significant and survived FWE correction (Pearson r = .588, p < .001, *pbonf* = .001, n = 33 [3 excluded], 95 % bootstrap CI [.379, .751]; Spearman ρ = .413, p = .012, n = 36, 95 % CI [.135, .648]), Position 2 was again nominal but did not survive correction (Pearson r = .376, p = .031, p_{bonf} = .123, 95 % bootstrap CI [.054, .634]; Spearman ρ = .324, p = .054, 95 % CI [−.021, .627]), and Positions 1 and 3 showed no relationship (P1: Pearson r = .035, 95 % bootstrap CI [−.293, .361]; P3: Pearson r = .151, 95 % bootstrap CI [−.105, .385]). The decoding–behaviour correlation is therefore not an artefact of the specific cells identified by the cluster-permutation test.

Overall, the relationship between pre-stimulus AUC and task performance was significant when aggregating across positions. The relationship at P2 did not survive FWE but the bootstrap interval excluded zero. As aggregation averages over roughly three times as many trials, this could be either explained by a general, position-unspecific relationship or mainly by an effect of P2 which is too noisy to be detected at the single-position level. However, we cannot clearly distinguish between these possibilities or a combination of them.

To corroborate this series of decoding results, we computed all decoding analyses as reported so far using the more nuanced MRR values instead of AUC. Reassuringly, we identified matching 1D and 2D decoding clusters, comparable expected-unexpected differences, as well as even stronger neural-behavioural correlation effects surviving FWE correction at P2 (figure and statistics in Figure S2).

Is syllable decodability enhanced at the predictable position?
Given that theories such as Dampening or Sharpening would predict differences of decoding performance between predictable and unpredictable stimuli, we compared decoding performance between P2 and P1.
While Sharpening predicts stronger syllable decoding at a predictable position (such as P2) than at an unpredictable control (P1), Dampening predicts the reverse. All three positions carried reliable syllable information on the diagonal (P1: 185–355 ms, p = .002; P2: 165–500 ms, p < .001; P3: 265–500 ms, p < .001), so P1 is not at floor and the contrast is a genuine test of relative representational strength.
The confirmatory P2 − P1 difference-wave cluster test on the on-diagonal time course returned a late, post-stimulus significant cluster in both metrics (ROC AUC: 365–435 ms, p = .045; MRR: 405–450 ms, p = .041), with better decoding at P2. Neither

secondary contrast (P2 vs. P3, P1 vs. P3) yielded a cluster; for P2 vs. P3 the evidence favoured no difference (Δ = +0.001, p = .60, BF_{01} = 4.88).
The full time-generalization matrix showed only a non-significant cluster for that late difference (P2 > P1: train 355–445 ms × test 300 to 500 ms, MRR: p = .082, ROC AUC: p = .139), which is plausibly the more conservative measure. But it also revealed an entirely pre-stimulus cluster (significant for MRR, train 130–500 ms × test −180 to −75 ms, p = .040, ROC AUC same region, p = .089), reflecting P2 being decodable before stimulus onset while P1 was at chance. Note that this difference reaches significance in MRR but not in the pre-specified primary metric (ROC AUC, p = .089); the region-of-interest comparison reported in Section 3.3.4 (addressing pre-stimulus decoding carry-over), however, confirms the P2-over-P1 advantage in both metrics.

Does univariate suppression correlate with multivariate decodability at the predictable position?
As both theories, Sharpening and Dampening, predict a relationship between univariate expectation suppression and multivariate decoding performance (a positive relationship for Sharpening and a negative for Dampening), we tested whether the Standard P2-specific expectation suppression seen in the ERPs correlated with the corresponding decoding performance. We found no evidence for such a correlation at either predictable within-word position (P2 or P3), for either decoding metric (AUC or MRR), and regardless of whether the pre-stimulus cluster or the 1D diagonal decoding time course was used (see Table S1 in the Supplement). Bayes factors (BF_{01}) consistently fell in the 3.6-4.6 range across all six tests.

This test has limited sensitivity in the present data, however. The per-participant decoding estimates for this analysis have low split-half reliability (r_{SB} ≈ 0.2 for the position-restricted measures, ≈ 0 when collapsed across position; an instance of the reliability paradox described in the Methods Section), compounded by the modest trial count per participant per position. A genuine suppression–decoding coupling of small-to-moderate size would be expected to appear negligible here. We therefore report this analysis to document that the Sharpening/Dampening prediction was tested directly, but we do not treat its null outcome or the accompanying Bayes factors as adjudicating between the two accounts.

3.3.4 Control analyses

Low trial count does not explain the absence of one-dimensional Morph decoding. To test whether the absence of significant one-dimensional Morph ROC decoding was attributable to the lower Morph trial count, we subsampled Standard trials to match the Morph trial count and repeated the decoding analysis. The resampled Full-N Standard Trial 1D Decoding yielded a robust, extended cluster (162–498 ms, p < .0001). When Standard Trial counts were reduced to match Morph numbers (N ≈ 300), the group-level detection rate dropped to 66 % (i.e. significant clusters in 66/100 iterations; median best p = .011, IQR [.0015, .078]). Crucially, 100 % of iterations produced at least one candidate cluster surviving the initial uncorrected threshold. Real Morph trials at the same sample size yielded zero clusters at any

threshold. Overall, trial count alone cannot fully account for the complete absence of Morph decoding. Instead, the qualitative divergence suggests that Morph stimuli elicit a genuinely weaker or different neural representation, maybe due to ambiguity-resolving or parallel processes, which renders representational patterns less distinct than those of unambiguous Standard syllables.
Read together with the univariate results, this null finding also adds to another perspective: Morph syllables elicit the largest amplitudes of any condition while yielding no decodable identity, whereas predictable Standard syllables at P2 elicit the smallest amplitudes while their identity remains decodable. A larger response carrying less information is not what accounts of processing-load would predict, since more extensive processing of a stimulus should render its identity more, not less, recoverable. One reading is that a substantial part of the response to an unresolvable stimulus reflects a residual rather than a representation; a second is that a 50/50 blend of two trained classes is inherently hard for a categorical classifier to assign to either of them. The two are addressed in Section 3.4.

Phoneme-specific behavioural asymmetries do not account for the decoding results. To assess whether the phoneme-identity-specific behavioural asymmetries discovered in the Sequence Task performance could account for, or otherwise confound, the neural decoding results, we re-examined the existing trial-level decoding evidence separately for each phoneme subgroup identified in the behavioural analysis. Neural decoding evidence did not differ from zero in any subgroup, including the two subgroups carrying the largest behavioural asymmetries (expected consonant = “d”: t = -0.06, p = .950; expected vowel = “e”: t = -0.48, p = .631), nor did it differ significantly between voiced- and voiceless-expected trials (b = -0.00004, p = .777) or across vowel-pair contexts (all p > .22). This indicates that the behavioural phoneme-identity asymmetries are not accompanied by a corresponding difference in decodable whole-scalp neural response amplitude within this time window, and that the previously reported absence of reliable Morph-trial decoding is not an artefact of pooling across phoneme subgroups with opposing behavioural effects.

Trial timing and baseline correction do not change ERP results. Re-running the confirmatory ERP battery without baseline correction left every reported positive effect intact: the stimulus-type main effect and the P1-vs-P2 and P2-vs-P3 contrasts all remained significant, several with numerically stronger cluster statistics. The diagnostic test on the −100 to 0 ms window did detect a position- and stimulus-type-dependent offset there (P1 vs. P3: p = .002; P2 vs. P3: p < .001; Morph vs. Standard: p < .001), confirming genuine ERP overlap at this SOA. These pre-stimulus offsets are an expected consequence of the SOA rather than the source of the reported effects: what licenses the latter is the re-run itself, in which each of them survived the removal of baseline correction (Section 3.3.2).

No evidence for pre-stimulus decoding carry-over. Transferring the P2 pre-stimulus decoding cluster onto P1’s own decoding matrix (SOA-shifted) showed no above-chance decodability of the preceding syllable in that window for either metric

(MRR: M = .4084 vs. .4083 chance, $t(35)$ = 0.05, p = .962, Wilcoxon p = .810, dz = 0.01, BF_{01} = 5.6; ROC-AUC: M = .5007 vs. .5000, $t(35)$ = 0.39, p = .701, Wilcoxon p = .625, dz = 0.06, BF_{01} = 5.2).
In direct contrast to that null finding, in the original cells at P2, syllable identity was decoded well above chance, corresponding to a medium effect (MRR dz = 0.73; ROC-AUC dz = 0.67), whereas the preceding-syllable estimate is within sampling noise of zero ($|dz| \leq 0.06$), with BF_{01}s of ≈ 5 favouring its absence. P1 also decoded significantly worse than P2 in a direct paired comparison of the same cells (MRR: $t(35) = -3.24$, p = .003, $dz = -0.54$, BF_{10} = 13.5; ROC-AUC: $t(35) = -2.47$, p = .019, $dz = -0.41$, BF_{10} = 2.5).
This converges with a logical argument independent of any single test: the preceding-syllable confound is symmetric across positions (Position 3 is exactly as deterministically predicted by Position 2 as Position 2 is by Position 1), so a carry-over account predicts an analogous pre-stimulus cluster at Position 3, however, none is observed. Together, these results argue against a carry-over contribution to the pre-stimulus decoding finding.

Decoding of unexpected components. An exploratory decoding analysis of the unexpected (non-presented for Standards) component for Standards and Morphs complements the null result reported above for the expected component in Morph trials: Using the 1D diagonal, we found one nominally significant cluster in Morph trials at P2, where the AUC time course dipped below chance early in the trial (37–97 ms, p = .048). However, this cluster did not appear in the corresponding MRR time course, nor in the Standard-versus-Morph difference wave for this window. Comparable marginal clusters occurred elsewhere in the same analysis (Standard trials, P3, AUC: below chance, 47–132 ms, p = .038; Standard trials, aggregate, MRR: above chance, 242–298 ms, p = .022). Given their short and inconsistent windows, marginal p-values, and especially the absence of convergence across metrics and conditions, we regard these clusters as rather unreliable. For transparency, we report the exploratory 1D decoding results for the unexpected component in the Supplement (Figure S3). In the corresponding 2D time-generalization matrices, no cluster reached significance for the unexpected component, in either Standard or Morph trials.

Random sequence decodability. Cluster-based permutation tests also identified significant clusters for the Random sequence on which the classifiers were trained (Accuracy/AUC/MRR, all $p \leq .0013$). These clusters were significantly below chance, which we attribute to the Random design's minimum six-syllable repetition lag (see Supplement, Figure S4); Random-sequence decoding is therefore not a positive validation of the classifier.

3.4 Discussion

Participants listened to two kinds of syllable stream: pseudo-randomized sequences, and structured sequences in which the same syllables were ordered into six trisyllabic pseudowords they had explicitly learned. In the structured stream, syllables were occasionally replaced by acoustically ambiguous Morphs, blending the expected syllable with an unexpected one, and participants intermittently reported which syllable they had just heard.

In this study, three levels of measurement gave three answers. Behaviourally, listeners reported the expected component of a Morph increasingly often at later within-word positions, while identification of unambiguous Standard syllables was elevated over the random-sequence baseline only at the two word-boundary positions, not at the medial one. Electrophysiologically, responses to predictable Standards were suppressed below baseline at the medial position and rose above it at the final position, while Morphs exceeded Standards throughout. Representationally, Standard syllable identity was decodable at all three positions, accompanied — at the medial position only — by syllable-specific information before stimulus onset that tracked how far listeners' reports shifted toward the expectation.

Taken together, these results show that a content-specific sequential prior is instantiated at the first predictable position within a word and governs what is heard when the input is ambiguous. A behavioural advantage at the two word edges of intelligible input qualifies this picture further, asking how neural predictive processes align with behavioural patterns rooted in speech perception in a sequence context.

3.4.1 What happened at each position?

At the first position, which follows a word boundary and is therefore unpredictable, behaviour and physiology converge on bottom-up processing: Providing ambiguous input that cannot be shaped by expectations, Morph responses were dominated by properties of the stimulus itself while showing a pronounced electrophysiological response. Similarly, Standard amplitudes were indistinguishable from the random-sequence baseline, and no pre-stimulus information was detectable. Clear sensory input was correctly identified – better even than in the Random condition despite being unpredictable.

At the second position, fully determined by its predecessor but leaving the word incomplete, four things converge. Firstly, Standard amplitudes were suppressed below the random-sequence baseline, uniquely at this position. Secondly, the suppression also extended to Morphs, pulling their otherwise elevated early response down to baseline. Thirdly, pre-stimulus activity carried identity-specific information about the upcoming syllable. And lastly, behavioural categorization of Morphs had shifted substantially, though not maximally, toward the expected syllable, while Standard identification, at this position only, did not exhibit any categorization advantage of the sequence structure (81.18 % vs the 81.23 % in the random sequence, $BF_{01} = 5.6$).

At the third position, both predictable and word-final, the electrophysiological direction reversed: both Standards and Morphs exceeded the baseline, especially in a late window, suggesting a word-closure response. No pre-stimulus information was detectable, while Morph responses reached their highest proportion of *expected* categorizations and Standards identification accuracy increased to the same level as for the word-initial position, (87.6 % vs. 87.9 %, both elevated relative to the 81.2 % baseline and medial position).

Table 3.1 summarizes the pattern across the three positions.

	Position 1 (onset)	**Position 2 (medial)**	**Position 3 (offset)**
ERP vs. Random baseline	≈ baseline	suppressed (255–500 ms)	enhanced (250–500 ms)
Syllable decodability	1D: brief (185–355 ms); 2D: absent	1D: sustained (165–500 ms), late P2 > P1; 2D: present	1D: late (265–500 ms); 2D: present
Pre-stimulus identity decoding	absent	present	absent
Standard-syllable identification	87.9 %, above baseline	81.2 %, at baseline	87.6 %, above baseline, faster RT
Morph-syllable categorization	acoustically biased	partially shifted toward expected (~40 %)	most shifted toward expected (~55 %), faster RT
Reading	weak sequential constraint; onset segmentation	maximal sequential constraint; prediction live	strong sequential constraint; word-closure signal

Table 3.1. Position summary table.

3.4.2 Ambiguity and expectation effects – an interplay

The acoustic ambiguity of stimuli produced an effect largely independent of position: Morphs elicited larger responses than either Standards or the unpredictable baseline and did so even in the block that covertly carried the word structure (the hidden-structure block). This might indicate that ambiguity resolution recruits additional processing resources (Xu & Jiang, 2026), regardless of explicit structural knowledge.

While this main effect of ambiguity is visible at every position, with Morphs eliciting a higher amplitude than Standards, it also interacts with position-related expectation. The interaction lies in the comparison against baseline: at the first position, Morphs are reliably elevated above it; at the second, the amplitude of both Morphs and Standards is pulled down to or below it; at the final position, ambiguity and closure combine additively so that both rise above baseline. Taken together, expectation effects do not abolish the ambiguity response but rather offset it, and the offset is specific to the position.

We find suppression at the second position, which has the highest sequential constraint, with both the preceding and the following syllable fixed and fully predictable. A reduced amplitude at the most predictable position is compatible with the concept of expectation suppression (Feuerriegel et al., 2021). Expectation suppression effects can be confounded with repetition suppression or low-level stimulus adaptation, in which reduced amplitude reflects merely repeated presentation of physically similar stimuli rather than expectation as such (Tang et al., 2018). Our design counters this by construction: syllable identity, condition, and position were balanced and pseudo-randomized, and syllables composing a word were selected to be phonetically different. So, the word-medial suppression we report is unlikely to result from a syllable's raw presentation frequency or physical properties.

A reduced amplitude alone is only half the picture; the more interesting part concerns the changes of the underlying population code. Sharpening and Dampening differ in their assumption what information the response still carries, so with regard to our multivariate decoding results, we turn to this question next.

3.4.3 Amplitude and decodability: Sharpening and Prediction Errors?

As introduced earlier in this dissertation, Sharpening and Dampening make different predictions about the underlying population code, differing in whether decodability should increase (Kok, Jehee, et al., 2012) or decrease (Richter et al., 2018) alongside a reduced amplitude.

Following that distinction, Dampening is the account our data are least compatible with. We find a suppressed amplitude at the medial position, while syllable identity of Standards remains reliably decodable (165–500 ms, $p < .001$), even higher than at the unpredictable first position, and accompanied by a pre-stimulus cluster carrying identity-specific information. A representation of expected stimuli under Dampening would look the opposite: less, not more, decodable information for expected stimuli.

Our neural data are instead better compatible with predictions of Sharpening accounts: syllables at the medial position are suppressed in amplitude yet decodable in both the more sensitive diagonal analysis and the more conservative time-generalized matrix. A confirmatory contrast showed that decoding performance at this predictable position exceeds the unpredictable first position in two measures (ROC AUC: 365–435 ms, $p = .045$; MRR: 405–450 ms, $p = .041$), overlapping the window (165–1000 ms) in which Sharpening effects have been reported in other auditory studies (Schneider & Blank, 2026), noting that our epoch ends at 500 ms and cannot speak to the later part of that range.

Our pattern is also broadly compatible with the opposing process theory (Press et al., 2020), though our data cannot confirm its central claim directly. OPT proposes an early bias toward the expected stimulus, followed by a later, reactive enhancement of stimuli that turn out surprising. We find something widely in line with the first half — pre-stimulus, content-specific activity resembling the expected syllable — but cannot

evaluate the second half, since the decoding patterns of our Morphs do not promote sustained or reliable identity information for either expected or unexpected stimuli. The only exception is a short, early cluster (37–97 ms, p = .048) in Morph decoding at the medial position, indicating reduced information for the unexpected component. This would fit the direction OPT's early bias would predict, but the clusters' borderline significance and lack of cross-metric consistency mean we do not treat it as direct evidence (Supplement, Figure S3). A recent EEG study using observed action outcomes reports a two-phase pattern that matches the OPT better (Rittershofer et al., 2026): higher decoding of the expected outcome before stimulus onset, then decoding of the unexpected outcome from ~155 ms post-stimulus. While we share the pre-stimulus activity, our Morphs do not qualify as a valid comparison to their unexpected condition, since Morphs do not fully mismatch the prediction in our design. Thus, we cannot report evidence for or against the second half of OPT's claim and remain with OPT as a plausible frame for the pre-stimulus result only.

Although it was not tested here as an a priori hypothesis, an alternative reading of our amplitude and decoding pattern allows for an interpretation in terms of the canonical predictive-coding architecture (specifically, a verbal rather than a simulated version of this account), which posits separate prediction (representation) and error units (Bastos et al., 2012; Friston, 2005): if expected sensory evidence is largely explained away by a strong prediction, little error signal remains to drive amplitude, producing the low amplitude we observe at the medial position of Standards. Decodable identity information, meanwhile, would instead be carried by the still-active representation of the predicted syllable in prediction units — consistent with the pre-stimulus decodable pattern we report at this position. Following King and Dehaene (2014), the broader, time-generalized decoding clusters at the medial and final positions are compatible with a maintained representation, in this case of the expected syllable, with classifiers finding syllable-identity information present over an extended window — consistent with a prediction that could already be active before syllable onset. Conversely, the narrow, diagonal-only pattern at the unpredictable first position is compatible with the signature of a single feedforward pass of sensory input, in the absence of a syllable-specific expectation.

However, the decoding pattern of Morphs is not what specific prediction-error signalling accounts would lead one to expect. These accounts rest on an interaction between prior knowledge and sensory evidence: representations of predicted input are suppressed when prediction and input match, and enhanced when they mismatch (Blank & Davis, 2016; Sohoglu et al., 2024). Our design holds sensory evidence constant and therefore samples only part of that interaction, but extrapolating from it, whatever information a Morph trial carries should represent its unexpected component, since Morphs combine expected and unexpected information and it is the mismatch that is forwarded as error. In our study, the acoustically ambiguous Morphs generate a large, unresolved error, producing the large amplitude one would expect under a predictive-coding architecture with separate prediction and error units — the more general framing we adopt from here on, as distinct from the specific prediction-error

coding accounts just addressed. Instead, Morph trials carry no reliable decodable identity information for either component, and we do not have a confident resolution for this. Two considerations are worth noting.

Firstly, one principled possibility is that error units, under this architecture, primarily encode the magnitude of a mismatch rather than a stimulus-specific code (Walsh et al., 2020), which would explain why the identity of the unexpected component is not recoverable from this signal, independent of any interpretation of Standards.

Secondly, a further, purely speculative but theoretically plausible extension of this account concerns the short, early below-chance cluster (~37–97 ms) we observe for the unexpected component of Morphs at the word-medial position (Section 3.3.4, Supplement, Figure S3). Under Sharpening, expectation is thought to actively suppress sensory units tuned away from the expected feature, potentially already during the initial feedforward sweep of stimulus processing. For Morphs, which combine expected and unexpected acoustic information, this logic is consistent with exactly the kind of early, transient suppression of the unexpected component we observe. While the specific time window precedes the point at which expectation-related suppression has been reported to emerge in early auditory evoked fields more generally (100–200 ms; Todorovic & de Lange (2012)), this divergence might be explained by the differences in measure (univariate amplitude vs. multivariate decoding) and in kind (general suppression vs. a Sharpening-consistent, direction-specific effect), so an identical time course would not be strictly predicted. In this context, another speculative possibility is that in our paradigm, expectation is already active well before stimulus onset — as evidenced by the pre-stimulus decodable cluster we report — which could allow it to act on the earliest part of the incoming signal rather than building up gradually over the post-stimulus window.

As Morph processing proceeds beyond this initial window and recurrent, error-related computations come to dominate, carrying the unresolved mismatch forward as amplitude rather than as identity-specific information, any early Sharpening-consistent signal would be expected to be progressively overwritten by non-specific error-related activity — compatible with the disappearance of this cluster later in the trial. We stress that this temporal handoff is not established by our data and remains purely speculative: what our data show is only that the early cluster exists (though with limited robustness), which is directionally consistent with this account, and is not contradicted by the absence of later effects.

Overall, we would read Sharpening and predictive coding in its general, architectural sense not as competing accounts but as compatible descriptions at different levels of one architecture: a scalp-recorded amplitude sums over both an error-related and a representational population, while a classifier reads mainly the structure of the pattern — so a reduced amplitude alongside preserved decodability is consistent with such an architecture.

We are aware that this reading departs from an assumption commonly made in this literature: on the canonical microcircuit, predictions are conveyed by deep and prediction errors by superficial cortical layers, so that non-invasive measures are assumed to be dominated by error-related activity (Bastos et al., 2012; Kok et al., 2016; Walsh et al., 2020) — an assumption under which the representations reported for

speech are read as error signals (Blank & Davis, 2016; Sohoglu et al., 2024). Our pre-stimulus result suggests that scalp EEG is not exclusively sensitive to error-related activity, since content-specific information about the upcoming syllable is recoverable from the scalp EEG before any input has arrived from which an error could be computed (Section 3.4.5). The suggested compatibility with a predictive-coding architecture should not be read as support for prediction-error coding of syllable identity: the first is a claim about how signals are passed between levels, the second about what a sensory representation contains.
A structurally comparable division between the two mechanisms, Sharpening and prediction error processing, operating at different levels or stages of processing within the same data, has recently been reported for speech more generally (Schneider & Blank, 2026) and across the face-processing hierarchy (Garlichs & Blank, 2024).
Taken together, our neural results fit within a broader body of work suggesting that expectation shapes neural representation through more than one mechanism at once. Our own data do not adjudicate between these mechanisms directly, since the comparison that would do so — decoding of ambiguous Morphs — was uninformative; thus what we claim for the present dataset is compatibility, not positive evidence for their co-occurrence.

Finally, we note that our behavioural findings are, at least in part, hard to reconcile with either a Sharpening account or the architectural reading sketched above: identification accuracy for Standards is not elevated at the medial position, despite this position carrying the strongest neural signature of prediction. This runs counter to what both theoretical accounts would predict, and to other work showing facilitated, faster, and clearer perception of predicted stimuli more generally (Pinto et al., 2015; Sohoglu et al., 2014; Stein & Peelen, 2015; Teufel et al., 2018). This is a puzzle we take up directly in the next section.

3.4.4 Reconciling behaviour and electrophysiology

The neural and behavioural results point in different directions at the medial position. ERP suppression and above-chance pre-stimulus decoding are both compatible with Sharpening and, as discussed above, with a predictive-coding architecture. Both are readings that, on a straightforward extension, would predict a behavioural advantage for a confirmed prediction. But we do not find one: identification accuracy for Standard syllables exceeds the Random baseline at the two word-boundary positions (87.9 % at Position 1, 87.6 % at Position 3) but not at the medial position (81.2 %, indistinguishable from an 81.2 % baseline, BF_{01} = 5.6) — exactly where the neural evidence for a content-specific representation is strongest. The advantage we do find instead sits at the two positions our neural measures are less informative about: above-chance pre-stimulus decoding was found only at the medial position, but at neither edge.

The reaction-time data make this picture clearer: responses were reliably slower at Position 2 than at Position 3, in both Standard and Morph trials. Positions 1 and 2, by contrast, did not differ reliably. The RT advantage of Position 3 is in line with a broader

literature on syllable-triplet processing, where responses to the second and third syllables of a learned word are typically faster than to word-initial syllables (Batterink, 2017; Batterink & Paller, 2017; Franco et al., 2015). But in the same way, the plateau across Positions 1 and 2 is less consistent with this literature, which typically reports a three-step gradient rather than a two-step pattern. This discrepancy matches the overall picture in our data, where Position 2 diverges from both neighbouring positions across several measures.

Batterink et al. (2025), using also a trisyllabic word-learning design, add a relevant nuance: recognition accuracy for a given word predicted faster RTs specifically to that word's third syllable, with no equivalent relationship at the second syllable or for general RT facilitation, which held even for words participants could not explicitly recognize. That their data tie specifically a final-position RT advantage to explicit word knowledge, not a medial one, supports our P3-specific findings: higher accuracy and lower RT both in the Continuation and the Sequence task, and no RT difference between Standards and Morphs (in contrast to P1 and P2). While keeping in mind that our designs differ– their learning was implicit and variable across items, while ours was uniformly explicit – this provides independent support for considering the medial and final positions asymmetrically rather than as one class of "predictable" positions.

However, we note that our RT results generally warrant some caution in interpretation. Our task design differs crucially from common trisyllabic designs. Those paradigms, e.g., in rapid serial auditory presentation (Franco et al., 2015) often use target-detection tasks (Akça et al., 2023; Batterink, 2017), requiring a single keypress to one pre-specified target syllable or other signal. In contrast, our categorization task required participants to identify the last heard item and locate their response among multiple on-screen options at unpredictable points in the sequence. RT here plausibly reflects visual search and decision time as much as auditory processing speed, so we treat the RT as informative but not as strong independent support in the context of this literature.

Nonetheless, the accuracy and RT data converge on the same central point: we do not find a Sharpening-like behavioural signature at the position with the clearest neural evidence for prediction.
What we do find is a behavioural advantage located at the two word edges. Literature on artificial-language rule learning offers a candidate explanation: listeners preferentially extract and generalize regularities located at the edges of a sequence, i.e., the first and last elements, while regularities defined over medial elements are learned far more slowly, or not at all, even when the statistical evidence for them is equally strong. For example, listeners readily generalize a rule linking the first and last syllables of an artificial-language sequence to new items, but struggle to do so for an equivalent rule defined over medial syllables, even after extensive exposure (Endress et al., 2009; Endress & Bonatti, 2007; Peña et al., 2002). Notably, this edge-restriction appeared specific to generalization: statistical tracking of medial regularities remained intact even when their generalization failed (Endress & Mehler, 2009). This suggests that edge-anchored and statistical, and thus potentially predictive, computations are

dissociable, consistent with our observation of an edge advantage that does not simply track predictability. If word edges draw attentional resources that the medial position does not, independent of predictability, this could account for a behavioural advantage restricted to the two boundary positions without implying that medial predictions have no behavioural impact in general. On this view, the prediction may still be present, but its behavioural expression would be weakest where, we assume, attention is least engaged.

How attention shapes Sharpening and Dampening is itself an active research question (Alink & Blank, 2021; Jiang et al., 2013; Kok, Rahnev, et al., 2012; Rungratsameetaweemana & Serences, 2019; Summerfield & Egner, 2016), and our paradigm sits in interesting contrast to it. Sabio-Albert et al. (2026), manipulating attention directly in a cross-modal prediction paradigm, found that attention is required for a predicted stimulus to produce a Sharpening-like neural representation. Our results, as we interpret them, stand in contrast to that finding: the strongest neural evidence for an anticipatory representation appears at the medial position, which is maximally constrained by the sequence and therefore informationally redundant, and where – suggested by our behavioural data and in line with findings from literature on language-rule learning (Endress et al., 2009; Endress & Bonatti, 2007; Endress & Mehler, 2009; Peña et al., 2002) – attention is lower.

By contrast, the positions that seem to receive more attention based on identification accuracy show weaker neural signatures but clearer behavioural benefits. If identification accuracy is taken as a proxy for attention, our findings would sit uneasily with the idea that Sharpening depends on attention; however, we lack an independent measure of attention to test this directly. They also reveal a broader mismatch between what predictive processing accounts would lead us to expect for behaviour in a sequence containing both clear and ambiguous stimuli, and what the behavioural data actually show. Instead, the behavioural pattern aligns with the speech and word processing literature, where perceptual enhancement typically occurs at the edges of a unit. As our paradigm is grounded in both traditions, this offers the chance to integrate insights from both.

On this view, we do not think the neural and behavioural patterns need to be read as two versions of the same underlying effect. The most straightforward way to reconcile them is to treat them as answers to different questions. The neural measures ask where in the word a content-specific prediction is represented, and the answer is clearly the medial position. The behavioural measures ask where in the word a response benefits from any source of facilitation – of which neural prediction is only one candidate among others, such as, e.g., attention, lexical closure, and decision-level processes related to word learning. Here, the answer is the two edges which separate and thus identify a word as a single unit in a sequence.

Under this interpretation, the focus shifts: rather than asking why Sharpening does not yield a behavioural advantage, we recognize that the two patterns can coexist, given that they likely arise from different and only partly overlapping sources.

Whatever the exact division of labour between prediction, attention, and decision-level factors turns out to be, the neural evidence that a content-specific prediction is instantiated ahead of time is itself the more foundational claim, and the one our design speaks to most directly — we turn to it next.

3.4.5 Identity-specific pre-stimulus information – is it real?

The pre-stimulus finding is the most direct evidence in this study that a prediction is instantiated ahead of time, rather than inferred after stimulus presentation. A classifier trained on post-stimulus activity generalized to a pre-stimulus test window and recovered the correct syllable, hinting that information about the upcoming syllable's identity was present before stimulus onset, exclusively at the medial position, across two decoding measures. This signal is not merely present but functionally graded: across participants, its strength predicts how far listeners' percept of an ambiguous Morph shifted toward the expected syllable, suggesting that the prediction we detect neurally might be the same one that shapes what listeners consciously report hearing.

However, we treat this finding with some caution: positive reports of very early or even pre-stimulus sensory information of expected stimuli, in line with early Sharpening or sensory pre-activation, exist in vision (Blom et al., 2020; Kok et al., 2017; Myers et al., 2015; Rittershofer et al., 2026; Tirou et al., 2026) and in audition (Demarchi et al., 2019). However, other studies investigating predictive processing did not find pre-stimulus decoding of expected stimuli, e.g. (Alink et al., 2026; Hu et al., 2026).
Even the auditory finding is now the subject of an active methodological dispute: a specific bias between a classifier's confusion matrix and the sequence's transition matrix can produce the same pre-stimulus decoding pattern from data with no predictive content at all (Abdoun et al., 2026), a critique the original authors have partly rebutted with newly controlled data (Demarchi et al., 2026). We take this less as a verdict than as a benchmark: pre-stimulus decoding claims of this kind are only as credible as their control for this specific confound, which is what the next paragraph addresses for our own design.

For our data, we addressed two candidate artefacts. First, a carry-over effect of the preceding syllable is excluded by a control analysis: applying the same cluster coordinates to the preceding syllable's own decoding matrix, at the time window corresponding to when the current syllable is presented, yielded no reliable decoding. In other words, decoding of syllable n-1 was not still elevated at the moment syllable n is played. Moreover, a pure carry-over account would also predict an analogous cluster at the third position, which is equally determined by the second syllable, but none was observed.
Second, the specific bias Abdoun et al. (2026) identify, i.e., a correlation between the classifier's confusion matrix and the sequence's transition matrix, which can mimic genuine pre-stimulus decoding when physically similar stimuli are also more likely to follow one another, is unlikely to apply to our design for three reasons: firstly, the training sequence's transition structure was fully first-order counterbalanced, so every syllable was equally likely to follow every other, and the pairs a classifier might confuse

are therefore no more likely to be sequential neighbours than any other pair. Secondly, the assignment of syllables to positions was balanced across participants, so any residual bias would vary in direction between participants and cancel at the group level — unlike designs in which the same transition structure is fixed for everyone, as in Demarchi et al. (2019). And, thirdly, the within-word transitions we use are between phonetically non-overlapping syllables by design, which limits how much a classifier's confusions could align with them in the first place.

How does our finding fit with other positive literature reporting pre-stimulus decoding of expected stimuli? Previous studies derived a template from a contrast between conditions: Kok et al. (2017) explicitly built theirs from 'the difference in decoding between validly and invalidly predicted gratings,' and Rittershofer et al. (2026) follow the same logic for action outcomes. Our medial-position cluster, by contrast, shows above-chance pre-stimulus decoding within the Standard condition alone, across two measures and with classifiers decoding actual syllable identity, rather than a binary validity contrast. Given the differences between designs (feature space, modality, number of classes), our finding might rather be considered a different kind of demonstration of the same underlying effect. Still, where that work has related pre-stimulus decoding strength to behaviour, our result sits in a broadly similar place: Kok et al. (2017) found a correlation between pre-stimulus decoding and behavioural benefit only when the expected feature was task-relevant (r = .44, p = .035) and not when it was task-irrelevant (r = −.13, p = .55). Similarly, our own pre-stimulus decoding strength at the medial position predicts how often listeners resolved an ambiguous Morph toward the expected syllable (r = .515, p_{bonf} = .009 aggregated across positions; r = .46, p_{bonf} = .027 at the medial position specifically, MRR metric). We read this as a genuine parallel worth drawing, while noting our behavioural measure, i.e., the proportion of an ambiguous stimulus resolved toward the prediction, could arguably be understood as even a more direct read of a prediction filling in a percept, than a speed or accuracy benefit on a fully visible stimulus.

One further alternative deserves explicit consideration. Because the response elicited when a predicted sound fails to arrive is itself sound-specific (SanMiguel et al., 2013), a signal preceding stimulus onset could in principle reflect a negative prediction error — the error registered while an expected stimulus is overdue — rather than the prediction itself. Our timing parameters make this unlikely here. The stimulus-onset asynchrony was jittered only narrowly (ISI M = 268 ms, SD = 22 ms), so a syllable is on average not late at all and at most a few tens of milliseconds late on extreme trials, whereas the pre-stimulus clusters begin 180 to 200 ms and end 65 to 75 ms before onset. Almost the entire window therefore precedes the earliest point at which a syllable could be registered as overdue. A negative prediction error would moreover be expected at the third position as well, where the temporal expectation is at least as precise, and no cluster was found there.

Finally, we have not identified an alternative account of this pre-stimulus effect beyond genuine sequential predictive effects, and the control analyses above address the candidates we could think of. We say this with the field's recent history in mind rather

than as reassurance: studies investigating similar predictive processes in this literature have reported null findings for stimulus-specific pre-onset information, and even the positive auditory finding discussed above is still being actively contested. We therefore report our own effect while deliberately situating it within this contested space, rather than treating it as a definitive contribution to the positive evidence.

3.4.6 Why the medial position, and not the final one?

That expectation effects appear at the medial rather than the final position is initially counterintuitive: the final syllable is more predictable in terms of accumulated context and better learned according to the Continuation Task. We want to share some considerations that might make the pattern intelligible, one of them a positive contribution of our design itself.

First, the initial position and the medial position stand in a cue-confirmation relationship, distinct from the closure-related point we make below about the final position: the first syllable generates a prediction about the syllable to come, and the medial position is where that prediction is confirmed or violated — a step still required for word identification, and therefore behaviourally live, even though the medial syllable's identity is already fully determined at a purely sequential-statistical level. This is not necessarily in tension with calling the same position "informationally redundant" elsewhere in this chapter: statistically the identity of the medial syllable is redundant given the fixed preceding and following syllable in this design, but the confirmation that closes the prediction initiated at the first position has not yet happened, which leads to the next point.

Second, a prediction being available is not the same as it being used with maximal efficiency — a distinction that fits the two-process picture we draw throughout this chapter, in which a graded, accumulating use of the prediction in Morph trials sits alongside the categorical baseline-versus-structured comparison for Standards. At the medial position, where the word is in principle fully determined by its first syllable, identification is well below both edge positions, and the Morphs' categorization as expected reaches only around 40 % (instead of ~55 % at the final position). An ideal observer making use of the learned transition structure in an optimal fashion would do better than that. That the effect of expectation instead grows across the word, as evident in the Morph trials' *expected* ratings rising linearly from the first to the last position, suggests that listeners accumulate evidence toward a word identity rather than committing at the earliest point that would theoretically allow it. As discussed in Section 3.4.4, independent behavioural evidence from a different statistical-learning paradigm converges on the same separation: Batterink et al. (2025) tie a final-, rather than medial-, position reaction-time advantage specifically to explicit recognition of the learned item, consistent with the medial and final positions drawing on separable sources of facilitation rather than forming one undifferentiated 'predictable' class.

Third, this dissociation of the two predictable positions is also evident in the configuration of the neural effects: the medial suppression is a fronto-posterior dipolar

pattern, whereas the final enhancement is a single broad central positivity with no such mirrored counterpart (Results 3.3.2). This difference suggests not two phases of one process, but two different ones. While the medial position might be categorized as expectation suppression, the final position's structure-level profile (present for Standard and Morph alike) instead resembles more an endogenous closure or boundary response at the completion of a learned unit.
A component candidate for this evoked potential at the final position could be the closure positive shift observed at the boundaries of learned linguistic units in the absence of acoustic cues, over central and centro-parietal electrodes (Steinhauer et al., 1999; Steinhauer & Friederici, 2001). Because the analysable window in our study is restricted to ~0–500 ms, we cannot determine whether this positivity is sustained, and we therefore offer this correspondence as framing rather than component identification. Instead – or alongside – a dedicated closure component, we would also think an attentional contribution plausible: the upcoming word boundary is itself unpredictable, and we already treat the first and the last position as the ones where attention is comparatively higher. In that line of thought, the same shift in attention toward an approaching, unpredictable onset could contribute to this late enhancement without requiring a separate, closure-specific mechanism.

Lastly, the dissociation between 'first predictable' and 'sequence-final' as we report it is hard to separate in two-element paradigms (Kok, Jehee, et al., 2012; McDermott et al., 2026; Richter et al., 2022; Tang et al., 2018), including our own pupil study (Becker, Korn, et al., 2024) and some of the studies reporting pre-stimulus decoding as well (e.g., Kok et al., 2017; Rittershofer et al., 2026). If the second element is simultaneously the first predictable one and the sequence-final one, then anticipatory and closure-related components are confounded by design. A trisyllabic structure separates them: here, the pre-stimulus, content-specific signature belongs to the medial, predictable-but-not-final position, while the closure response belongs to the last one. And this is where our findings extend existing literature on predictive processing.

3.4.7 Does incidental exposure suffice for statistical learning?

In the block that carried the word structure without disclosing it, no significant position effects were present in any measure. In electrophysiological data, Standard amplitudes were indistinguishable from the Random baseline throughout the trial. Ambiguity effects, by contrast, were fully preserved, with Morphs eliciting larger potentials than Standards. The straightforward reading of this finding is that the position effects require explicit knowledge, or at least the amount of exposure and training our explicit condition provided, whereas ambiguity resolution does not.

Behavioural results resembled the categorization pattern from the word sequence, with Standards showing high identification accuracy, while Morph responses were more distributed and, numerically, followed a positive trend across positions. However, a reliable analysis was not feasible due to low trial count, so we refrain from concluding that implicit statistical learning has taken place.

A further constraint on the implicit-statistical-learning question is structural: the hidden-structure block was preceded by two unstructured blocks. Since the order of random sequences preceding structured conditions has been shown to affect statistical learning negatively (Iorio et al., 2026), the data might look different had our participants been exposed to the hidden-structure block before the random sequences.

3.4.8 Limitations

Some aspects of our design and analysis choices deserve caution when interpreting the results above and thus are described below.

Expectation Suppression

As elsewhere in the literature, we treated expectation suppression here as the empirical phenomenon that accounts such as Dampening and Sharpening are proposed to explain; establishing its presence would thus be a prerequisite for adjudicating between them (Press et al., 2020; Richter et al., 2018; Summerfield & Egner, 2009). In our data, we identified a reduced amplitude at the medial position for Standards (posterior-temporal: 35–240 ms, p = .0165; fronto-central: 25–240 ms, p = .0188, the latter non-significant after FWE correction) as well as a reduced amplitude when pooling across stimulus type (40–240 ms, posterior, p = .0233, also non-significant after FWE correction) (see Section 3.3.2). These clusters span the time window where we would expect expectation suppression to occur based on previous literature (~100-200 ms, Todorovic & de Lange (2012)), but borderline significance indicates that this effect may not be robust, unlike the post-offset clusters, which survive FWE correction.

However, there are other studies reporting only a weak or no expectation suppression (den Ouden et al., 2023), or even enhanced activity for expected stimuli (Amado et al., 2016; Larsson & Smith, 2012). Sabio-Albert et al. (2026), for instance, could not confirm their own preregistered dampening hypothesis for unattended predicted stimuli. In this context, Feuerriegel et al. (2021) review four confounds that recur across expectation-suppression studies, among them attention. Alink & Blank (2021) go further, proposing that at least part of what is reported as expectation suppression across this literature could instead (or additionally) reflect reduced saliency-driven attention to predictable stimuli rather than reduced prediction error as such. We do consider attention a credible contributing factor for our overall findings (see Section 3.4.4), and it could in principle account for the reduced amplitude at the medial position. However, we attribute that particular effect more to endogenous word-edge effects rooted in speech processing than to exogenous stimulus saliency.

As Feuerriegel et al. (2021) point out, expectation suppression is not always clearly distinguished from enhancement of the surprising alternative, and a fully neutral baseline is difficult to construct. Following their criteria, our Random condition is also not an ideal neutral baseline: it contains no ambiguous syllables, so it differs from the structured condition in stimulus-set composition and task context, not only in sequential predictability. Considering this, comparisons against it are better read as comparisons against a condition without sequential structure, not against a genuinely neutral perceptual state.

Stimuli

The ambiguity effect of Morphs was established by measuring against the Random baseline and showed an overall higher amplitude than the baseline and Standards. An elevated amplitude elicited by the ambiguity of morphed speech (Xu & Jiang, 2026), can, however, be confounded by saliency or surprise. In our study, Morphs had a lower frequency relative to Standard trials (1:5), and infrequent stimuli can elicit larger responses for attentional or oddball-like reasons independent of predictability (Jacobsen & Schröger, 2001; May & Tiitinen, 2010). We cannot fully rule this out, though the specificity of the position-by-condition pattern reported in this study, showing that Morphs are not uniformly elevated regardless of position, is not what a pure frequency account would predict on its own.

All stimuli, Morphs, Standard and Random alike, were spectrally vocoded (see Methods). This was a deliberate design choice, intended to keep the identification task non-trivial, and therefore keep sequence structure behaviourally relevant, even for acoustically clear input. It is presumably also why Standard-syllable identification sits at 87–88 % at the two boundary positions rather than at ceiling.

In the same line, response patterns for Morphs showed a spread across categories, demonstrating that they were indeed more ambiguous than Standard stimuli. Yet, the final set of Morph stimuli still varied in their perceptual ambiguity despite prior online validation. This is a common challenge of morphing stimuli; studies quantifying perceptual ambiguity typically permit a certain range of ambiguity scores (Schneider & Blank, 2026) into which our stimulus set falls when benchmarked against the Sequence Task categorization patterns. However, physical residual differences in ambiguity across Morphs likely contributed to the categorization patterns observed for Morph trials in the initial position, where *unexpected* responses exceeded *expected* ones (see Section 3.3.1.1). This was further compounded by a voicing-based saliency asymmetry in our stimuli, biasing perception towards the voiceless consonant (Chen et al., 2017; Liberman et al., 1957), which resulted in *mix1* responses (defined by a voiceless consonant) being chosen numerically (and in part significantly) more often than the *mix2* options (starting with a voiced consonant).

Other studies facing the challenge of optimizing the ambiguity of morphed stimuli chose to calibrate their stimuli individually per participant (Garlichs & Blank, 2024). However, this was not feasible in our case since calibrating perceived ambiguity per listener would have meant to let each participant adjust their own acoustic material. This would have likely undermined our intention of Morphs being perceived as ambiguous versions of previously encountered Standard stimuli, instead of establishing them as a separate stimulus set on their own with participants being consciously aware of them.

Behavioural confounds

The behavioural expectation effect we observed in our categorization task, showing that Morphs are increasingly perceived as their expected component at later positions, could in principle be task-induced rather than directly reflecting perception. One expression of this concern follows the idea that motivated participants simply learn better and select the expected option more often, potentially because they believe they are supposed to correctly reproduce the trained words in the Sequence Task, rather

than report their actual percept. Participants were, after all, trained in the Continuation Task to do exactly that, and could in principle carry the same strategy into the Sequence Task.
The data argue against this: the Morph categorization effect holds within participants across their own six words (see Section 3.3.1.2), and adding four explicit engagement proxies leaves the pooled coefficient essentially unchanged. In practical terms, listeners' choices still varied within their own set of six words in a way that a stable, participant-level motivation difference cannot explain, and this pattern is unchanged when explicit engagement indicators are added to the model.
A second version of the initial concern is harder to rule out: listeners may weight their expectation more heavily for well-known words without the Morph itself sounding different to them, which would point to a genuine top-down effect at the decision rather than the perceptual stage (Ganong, 1980), connected to a broader debate about whether expectation effects are better characterized as cognitive than sensory (Gabhart et al., 2025).
Two observations speak to this point: first, the pre-stimulus measure that predicts the behavioural shift towards the *expected* response (see Section 3.3.3) is recorded before task onset and therefore cannot be affected by decisional processes. Second, *expected* responses at the final position are no faster than *unexpected* ones, whereas a low-effort completion strategy, e.g., neglecting the sensory evidence and simply completing the word as learned, would be expected to produce faster answers specifically for the *expected* option. Neither argument fully settles whether a genuinely perceptual bias or a decision-criterion shift underlies the Morph effect, but together they make it less likely to be an artefact of participant motivation or deliberate strategy.

Correlating univariate and multivariate data
One aim of this project was to identify neural implementations of predictive processes, such as Sharpening or Dampening accounts, in our data. These accounts usually assume a correlation between the central phenomena of expectation suppression and decoding performance (positive under Sharpening, negative under Dampening). We computed such correlations across participants, but it was null across six such tests (BF_{01} = 3.6–4.6, favouring the absence of a relationship). However, we do not read much into this: individual-differences correlations built on cluster-based, group-level measures with low split-half reliability are typically underpowered for exactly this kind of test, independent of whether a true effect exists (Hedge et al., 2018). This is a known limitation of this approach generally, not a specific weakness of our design, so we treat the null finding as uninformative rather than as evidence against either account.

Morph decoding
Finally, as follows from the discussion in Section 3.4.3, the absence of Morph decoding is difficult to interpret from a single Morph-component mixing ratio. In future studies, a graded morph series could potentially help to separate two accounts that a 50/50 blend conflates: if ambiguity degrades identity information as such, decodability should decline monotonically with mixing. If the null instead arises because two representations are equally active and cancel in the pattern, decodability should re-

emerge at intermediate ratios and do so in the direction of the expected syllable. Crossing such a series with within-word position would additionally yield a per-position estimate of how much acoustic evidence an expectation can override, rather than the single point the present design provides.

3.4.9 Conclusion

In this EEG study, we show that listeners who have learned short, trisyllabic pseudowords instantiate a content-specific prediction for a particular syllable within the word before that syllable arrives, and that this prediction shapes which syllable they hear when the acoustic input is ambiguous. The prediction is most strongly anchored at the medial syllable — predictable from the word's first syllable, but not yet the word's last — and it is electrophysiologically and topographically distinct from the response that accompanies completion of the word. Where the prediction is accurate, the evoked response to that syllable is attenuated and its identity is clearly represented; where the syllable is acoustically ambiguous and cannot be resolved from bottom-up input alone, the neural response is amplified and carries no clearly recoverable syllable identity.

How the attenuation is realized in the underlying population code is constrained, but not finally settled, by our data. A dampened representation is the account our data fit least: identity remains decodable exactly where amplitude is smallest. A sharpened one is better supported: decoding at the predictable position exceeds the unpredictable control. Behaviour, however, shows none of the facilitation this account would predict.

Our behavioural findings, then, point to something the prediction account alone does not cover: identification of clear, unambiguous syllables was best at the two word edges, not at the medial position where the neural evidence for prediction is strongest. The two measures therefore order the three positions by different properties of the sequence — one by how tightly a syllable is constrained by its predecessor, the other by its place relative to the word boundary. What makes boundary positions advantageous for identification is not settled by these data; a shift of attention toward word onsets and offsets is one candidate, though the present design offers no independent measure of attention to test it. The broader implication of this divergence, and what it suggests about the relation between predictive processing and word-level speech processing, is taken up in the following Discussion.

4. Discussion

Both studies in this dissertation asked a version of the same question: once a listener has picked up the statistical structure of a sound sequence, how are the predictable stimuli processed? The pupillometry study asked whether an autonomic signal tracks a vowel-pair structure incidentally while that structure is merely being listened to, and whether the pair association can be recovered afterwards. The EEG study asked where inside a learned trisyllabic word prediction is instantiated, what happens to the neural representation of a syllable once it becomes predictable, and what listeners report hearing when that prediction meets ambiguous acoustic input. Taken together, the two studies support a claim narrower than "statistical learning happens" — and more useful: structure was learned in both studies, but whether it became visible depended on what was measured, when it was measured, and under which conditions the structure was encountered.

Summary

In the pupillometry study (Chapter 2), stimulus-evoked pupil dilation tracked the presence of a sound reliably but carried no signature of the pair structure itself: no spectral peak at the pair rate, no difference in phase locking, and no reliable difference in trial-locked pupil diameter between the structured and the random condition. With Bayes factors in the anecdotal-to-moderate range, we would rather speak of a structure not detected than a structure shown to be absent. However, the more informative half of the result came afterwards: in the post-sequence familiarity task, explicit judgements were at chance (≈ 52 %), yet pupil diameter was reliably larger for recombined foil pairs than for previously heard familiar pairs ($t(30) = 2.74$, $p = .010$, $dz = 0.49$), and this response followed the objective pair category rather than what participants themselves reported. The associations had been learned; only one of the two available readouts could see them.

The EEG study (Chapter 3) replaced vowel pairs with consonant–vowel syllables, made word knowledge explicit and verified it independently, and added a third within-word position, so that a merely predictable position could be distinguished from one that also completes the word. Its central result is that predictability within a learned word is not one thing. At the first position, which follows a word boundary and cannot be predicted, no neural signature of expectation was present. At the medial position — determined by its predecessor but not yet completing the word — responses to predictable syllables were suppressed below the random-sequence baseline, syllable identity was decoded more strongly than at the unpredictable first position, and identity-specific information was already present before the syllable arrived. At the word-final position the direction reversed into an enhancement without any pre-stimulus information, which fits completion of the learned unit better than anticipation of its content. Behaviourally, ambiguous Morphs were heard as the syllable that fits the learned word increasingly often as the word unfolded (25.9 %, 39.9 % and 54.5 % expected responses across the three positions). A prediction is therefore instantiated

as early as the sequence allows, and it demonstrably shapes what listeners hear when the acoustic evidence does not settle the matter on its own.

At the level of the underlying representation, these results constrain the accounts introduced at the start of this dissertation without settling the matter. Dampening fits worst: at the medial position, amplitude was smallest exactly where syllable identity remained most decodable, whereas a dampened representation of an expected stimulus should be less decodable, not more. Sharpening fits the neural pattern better, although not in its narrowest form: that version would also predict a behavioural advantage for expected stimuli and a coupling between suppression and decoding strength. Neither was observed, but the measures entering the correlation were too unreliable for its absence to count against the account. At the level of representational content, the results run counter to specific prediction-error coding accounts, which predict the reverse decoding profile: ambiguous Morphs decoded best from their informative residuals, Standards worst because they should largely have been explained away (Blank & Davis, 2016; Sohoglu et al., 2024). At the level of architecture, however, they remain compatible with canonical predictive coding, in which separate populations encode predictions and prediction errors — on the assumption that EEG-based multivariate pattern analysis is sensitive to both. These are claims of different kinds, and on the reading we consider most defensible, Sharpening accounts and prediction-error processing can occur on different levels of one architecture: a scalp-recorded amplitude sums across error-related and representational populations, while a classifier reads the structure of the pattern.

What this dissertation adds to that debate is a distinction the debate usually collapses. Sharpening and Dampening are typically tested by contrasting a predictable stimulus with an unpredictable one, and in two-element sequences the predictable element is necessarily also the final one, so anticipation and closure cannot be told apart. A three-syllable word separates them, and once separated they behave differently: the anticipatory, content-specific signature belongs to the first predictable position, the closure-like enhancement to the last, and the two differ in scalp configuration, as in our case the medial effect forms a fronto-posterior dipolar pattern and the final one a single broad central positivity. Whether an expectation effect looks Sharpening-like therefore depends not only on whether a stimulus is predictable, but on where inside a learned unit the prediction is read out.

The most instructive result of this dissertation is one that does not resolve. The clearest neural evidence for a content-specific prediction and the clearest behavioural benefit occur at different positions. Identification of unambiguous syllables exceeded the random-sequence baseline at the two word edges (87.9 % and 87.6 % against 81.2 %), but was indistinguishable from it at the medial position (81.2 %, $BF_{01} = 5.6$) — exactly where suppression, enhanced decoding and pre-stimulus information converge. A coherent account of this divergence is that the two measures answer different questions: the neural measures ask where a content-specific prediction is represented, the behavioural measures ask where a response benefits from any source of facilitation, of which prediction is only one.

Luo et al. (2025) make a structurally similar case from a different angle. Manipulating word-length uniformity leaves online target detection unaffected while markedly impairing offline segmentation, which they attribute to different computational demands: detection requires no more than stored syllable transitions, whereas segmentation additionally requires that unit boundaries are located. Our data show a similar divergence within a single design, in the same participants and trials — the neural measures track transition constraint and peak word-medially, while identification accuracy tracks unit boundaries and peaks at the word edges. Notably, Luo et al. attribute boundary sensitivity to the rhythm that uniform-length streams afford, in line with Wang et al. (2026). However, our SOA was jittered precisely to preclude entrainment, so the boundary advantage observed here is unlikely to result from rhythmic cueing.

Beyond the laboratory, this positional pattern has a plausible everyday counterpart. A learned prior filling in an ambiguous syllable is the ordinary situation of listening to degraded speech — in noise, in an unfamiliar accent, with a hearing impairment, or in a second language — and our results make a reasonably specific prediction about it: such filling-in should be strongest not at the edges of a word, but at the point where the word first becomes determined.

Two sequence studies: lessons about tasks, conditions and stimuli

Set side by side, the pupil study and the EEG study may look as though they disagree: one found no trace of structure in an ongoing sequence, the other found a great deal. Two considerations resolve that apparent contrast. The first is what each measure can see: pupil dilation is a single, slow, averaged signal. It can indicate that something was learned (Nassar et al., 2012; Preuschoff et al., 2011), but not how it is represented. Only time-resolved multivariate EEG allowed the second study to address the representational question at all. The second is that the two studies differ in ways that plausibly determined how much there was to find: the pupil sequence was built from vowels alone, although sequential dependencies are extracted primarily from consonants (Nespor et al., 2003), and its sequence structure remained incidental to an unrelated task throughout exposure, whereas word knowledge in the EEG study was explicit and verified before the sequence of interest began. Despite the differences, both studies share a dissociation in their results: in both, a sensitive measure and an overt behavioural measure came apart — pupil against report in one, neural representation against identification accuracy in the other — and they did so in opposite directions.

That recurring dissociation carries the clearest methodological lesson of this work: it concerns how learning is tested. Both studies used a two-alternative familiarity task, and in both the behavioural data proved uninformative — but for opposite reasons: a null effect against a reliable physiological effect in the pupil study, and ceiling performance (93.3 %) in the EEG study. Soares et al. (2025) report a comparable dissociation, with reliable differences in EEG components between stream conditions — though only in the explicitly instructed tasks — alongside forced-choice performance that, while above chance throughout, did not differentiate those conditions at all. They

read this as evidence that online and offline measures index distinct processes and call for measures sensitive to the time course of processing as well as alternative post-learning tasks. Forced-choice measures also show limited reliability elsewhere. For instance, in 5- to 8-year-olds, serial recall showed substantially higher test–retest reliability than a two-alternative forced choice familiarity task and revealed learning even in children who scored below chance on the forced-choice task (Kidd et al., 2020). What limits the familiarity task is its granularity: it yields a binary verdict — above chance or not — and little beyond it. Alternative formats worth considering include serial recall (Isbilen et al., 2020) and, where feasible, tracking learning online via neural measures (Batterink & Paller, 2017). Either way, we would follow Conway et al. (2025), who argue that robustness is improved when a phenomenon is observed across multiple methods, and that individual studies should therefore incorporate several measures of statistical learning rather than rely on a single task, with indirect measures valuable precisely because they track learning as it unfolds. Importantly, as they add, a diversity of measures only helps if it is clear what each of them measures. For future work on implicit statistical learning, we would recommend building an item-level or time-resolved measure into the design from the outset, alongside rather than instead of a familiarity task.

This recommendation reaches beyond the laboratory, since statistical learning is not only a laboratory phenomenon: it is the mechanism most often invoked for how infants segment their first words, and individual differences in it have been related to language and reading outcomes and to developmental language disorder (Isbilen & Christiansen, 2022; Lammertink et al., 2017). Much of that literature rests on the same offline forced-choice measures that proved least informative in both studies presented here, and whose psychometric properties have also been questioned by others, as discussed earlier. Where the visibility of learning depends on a measure of that kind, a null forced-choice result in a developmental or clinical group supports rather less than it might appear to, and the online or physiological readouts discussed here would seem the more informative complement.

Another design lesson concerns the order in which structure is encountered. Iorio et al. (2026) show that initial exposure to random input measurably reduces subsequent statistical learning during a structured sequence of pseudowords, behaviourally and neurally. Both studies here are affected: in the pupil study, the random condition necessarily preceded the structured one, so condition and time on task are perfectly collinear by design. In the EEG study, the block that carried the word structure without disclosing it followed two full random blocks, which closely matches the manipulation of Iorio et al.; that covertly structured block in our study produced only a shallow behavioural trend (23.7 %, 28.7 % and 30.6 % *expected* Morph responses) and no reliable neural effect. Whether incidental exposure alone would have sufficed therefore remains open rather than answered in the negative. For future designs, we would seriously consider counterbalancing the condition order. Bruns et al. (2025) make a related point from perceptual learning: whether a second learning episode disrupts or

facilitates an earlier one depends on whether the learner is aware of what is being learned - precisely the dimension on which our two studies differ.

Task relevance and attention run through both studies and are the hardest of these conditions to pin down. In the pupil study the structure was irrelevant to the concurrent task throughout exposure, which is a plausible part of why online tracking failed — though it does not explain the offline result, and informing participants about the structure in advance (Experiment 1) did not help either: knowing that an information source exists and using it while listening are not the same thing. In the EEG study, task relevance and explicit training are confounded by design. The paradigm nonetheless produced a result that the simplest attentional account does not predict: the strongest anticipatory signature appeared at the medial position, the most redundant position in the word and, judging by the behavioural data, the one that attracts the least attention. Whether expectation effects require attention at all is unresolved in the wider literature, and this dissertation adds a case that does not fit the straightforward version of the claim.

The stimulus material is the last of these conditions, and logically prior to all the others: a measure can only detect learning if the input carried a learnable signal in the first place. The consonant–vowel asymmetry is the strongest lesson at this level, and the design change that most plausibly separates the two studies. Unit length is a second such factor, and it changed two things at once. It plausibly strengthened the signal available online, alongside the shift to consonant–vowel syllables and explicit training (though these three changes are confounded across the two studies, so their contributions cannot be separated here). And in addition to that, it made a distinction visible that a pair cannot express: in a two-element unit the second element is at once the first predictable one and the last, so anticipation and closure are necessarily conflated, and the pupil study could not have found the positional pattern the EEG study did even in principle.

Several limitations bound these conclusions beyond those discussed in the individual chapters. The sample size of the EEG study was fixed by the counterbalancing scheme rather than by a power analysis. This was a design-driven choice, but it left individual-differences correlations underpowered, which may limit how reliably a core assumption of the accounts discussed here can be quantified (Hedge et al., 2018). The Random condition provides a comparison without sequential structure rather than a neutral perceptual state, since it also differs in stimulus-set composition. In the pupil study, the confound between condition and order cannot be resolved with the present data. And both studies used pseudowords stripped of prosody, coarticulation and meaning: this is what makes their statistical structure interpretable, and at the same time what makes the step to natural speech an assumption rather than a demonstration. Whether the positional pattern reported here survives in speech that carries those cues is an open question, and a tractable one.

A first step would be to lengthen the unit beyond three syllables. With four syllables there are two medial positions that are equally constrained by their predecessor but

differ in distance from the word boundary, which separates anticipation from closure more sharply than three syllables allow. The account sketched above (Luo et al., 2025) makes a testable prediction for such a design: if the neural measures track transition constraint while identification tracks unit boundaries, the anticipatory signature should be comparable across both medial positions, whereas the behavioural advantage should remain confined to the edges. The same design could vary the degree of ambiguity across positions rather than holding it at a single ratio, which would show whether the point at which expectation overrides the acoustics shifts with sequential constraint — a graded, item-level readout on both measures of exactly the kind recommended above. Combined with an item-level or time-resolved learning measure alongside a familiarity task, such a study would put both strands of this dissertation to a single joint test: where prediction is instantiated within a learned unit, and what each class of measure can see of it.

This dissertation opened with the observation that listeners parse a continuous speech stream into words effortlessly and without noticing. Neither study contradicts that impression, but both suggest that the effortlessness is conditional rather than automatic: it holds in everyday listening because task relevance, exposure order and rich stimulus material usually align. Pulling them apart, as both studies here did in different ways, is what made the machinery visible — and it is where looking remains most informative.

5. Zusammenfassung / Summary

Zusammenfassung

Sprache erreicht das Ohr als kontinuierlicher Lautstrom, aus dem Hörerinnen und Hörer mühelos Wörter herauslösen. Diese Fähigkeit beruht wesentlich auf statistischem Lernen: der Extraktion von Übergangswahrscheinlichkeiten zwischen aufeinanderfolgenden Elementen und ihrer Nutzung für Erwartungen an das jeweils nächste Element. Wie das Gehirn eingehende akustische Evidenz mit solchen sequenziellen Erwartungen verrechnet, und was diese Verrechnung mit der neuronalen Repräsentation eines vorhersagbaren Lautes macht, ist Gegenstand der vorliegenden Arbeit. Zwei Studien nähern sich dieser Frage mit unterschiedlichen Messverfahren und Graden der Explizitheit.

Die erste Studie demonstriert, dass eine gelernte Assoziation physiologisch nachweisbar sein kann, ohne sich während des Lernens oder im Verhalten zu zeigen. In zwei präregistrierten Pupillometrie-Experimenten (n = 30, n = 31) wurden gesprochene Vokale zunächst in zufälliger Reihenfolge und anschließend in festen Paaren dargeboten. Die Pupille folgte zuverlässig der Darbietung einzelner Laute, zeigte aber in keinem der erhobenen Maße eine verlässliche Signatur der Paarstruktur. In der abschließenden Wiedererkennungsaufgabe blieb das Verhaltensmaß auf Zufallsniveau, während der Pupillendurchmesser zuverlässig zwischen zuvor gehörten und neu kombinierten Paaren unterschied – unabhängig von der subjektiven Einschätzung der Teilnehmenden.

Die zweite Studie trennt zwei Effekte, die sich in Paar-Sequenzen kaum auseinanderhalten lassen. 36 Personen lernten explizit sechs Pseudowörter aus je drei Konsonant-Vokal-Silben, die anschließend im EEG in kontinuierlichen Sequenzen dargeboten wurden, neben Sequenzen ohne Wortstruktur. Ein Teil der Silben war eine akustisch mehrdeutige Mischung zweier gelernter Alternativen. Bei diesen stieg der Anteil erwartungskonformer Kategorisierungen über die drei Positionen linear an – der Einfluss von Erwartung nahm über das Wort hinweg zu. Die Identifikation klarer Silben hingegen folgte der Wortkontur. Sie war an den beiden Wortgrenzen besser als in der unstrukturierten Sequenz, nicht an der mittleren Position – genau dort aber waren die neuronalen Erwartungseffekte am stärksten: die Amplituden waren reduziert und die Silbenidentität bereits vor Stimulusbeginn nachweisbar. An der finalen Wortposition kehrte sich die Amplitudenrichtung um, passend zum Abschluss der Einheit. Positionsabhängige Vorhersagbarkeit ist damit einer von mehreren Faktoren, die die Wortverarbeitung im Sequenzkontext prägen.

Beide Studien zeigen ein ähnliches Muster: in der ersten divergieren gelerntes Wissen und sein Ausdruck im Verhalten; in der zweiten zeigen sich neuronale und behaviorale Erwartungseffekte in der Wortverarbeitung unterschiedlich über Positionen. Gut sichtbar wird diese Trennung erst bei Einheiten mit mehr als zwei Elementen. Für die Untersuchung statistischen Lernens heißt das: die Sequenzstruktur entscheidet mit, welche Effekte auftreten können, und ein einzelnes Verhaltensmaß sagt über Strukturwissen wenig aus.

Summary

Speech reaches the ear as a continuous stream of sound, from which listeners extract words without apparent effort. This ability rests substantially on statistical learning: on extracting the transitional probabilities between successive elements and using them to form expectations about the element to come. How the brain combines incoming acoustic evidence with such sequential expectations, and what that combination does to the neural representation of a predictable sound, is the subject of this dissertation. Two studies approach the question with different measures and different degrees of explicitness.

The first study shows that a learned association can be physiologically detectable without becoming apparent during learning or in behaviour. In two preregistered pupillometry experiments (n = 30, n = 31), spoken vowels were presented first in random order and then in fixed pairs. The pupil reliably followed the presentation of individual sounds but showed no reliable signature of the pair structure in any of the measures obtained. In a final recognition task, participants' judgements remained at chance, while pupil diameter reliably distinguished previously heard from newly recombined pairs — independently of what participants themselves reported.

The second study separates two effects that are hard to tell apart in two-element sequences. Thirty-six participants explicitly learned six pseudowords, each made up of three consonant–vowel syllables, which were then presented in continuous sequences during EEG recording, alongside sequences without word structure. A subset of syllables was an acoustically ambiguous blend of two learned alternatives. For these, the proportion of expectation-consistent categorizations rose linearly across the three positions – the influence of expectation increased as the word unfolded. Identification of clear syllables, by contrast, followed the contour of the word. It was better at the two word boundaries relative to the unstructured sequence, but not at the medial position – which is precisely where the neural expectation effects were strongest: amplitudes were reduced and syllable identity was already detectable before stimulus onset. At the word-final position the direction of the amplitude effect reversed, consistent with completion of the unit. Position-dependent predictability is therefore one of several factors shaping word processing in a sequence context.

Both studies show a similar pattern: in the first study, learned knowledge and its expression in behaviour diverge; in the second, neural and behavioural expectation effects are distributed differently across positions in word processing. This separation becomes clearly visible only in units of more than two elements. For research on statistical learning this means that the structure of the sequence co-determines which effects can appear at all, and that a single behavioural measure says little about what has been learned.

6. Hilfsmittel / Tools

Beim Verfassen der vorliegenden Dissertation wurden KI-basierte Tools („Chatbots“) verwendet. Ihre Nutzung sowie die daraus generierten Anteile werden hiermit kenntlich gemacht.

Anthropic (*Claude*, Versionen: Claude Opus 5, Claude Sonnet 5, Abrufzeitraum: 2026). Verfügbar unter: https://claude.ai/

- Verwendungszweck:
    - Korrektur von Grammatik und Rechtschreibung des Manuskripts
    - Verbesserung von Lesbarkeit, Satzfluss und akademischem Schreibstil
    - Überprüfung von Programmcode und Analysedaten
    - Überprüfung von logischer Konsistenz und Strukturierung von Argumentationslinien

Google (*Gemini*, Versionen: 3.6 Flash, 3.1 Pro, Abrufzeitraum: 2026). Verfügbar unter: https://gemini.google.com/

- Verwendungszweck:
    - Unterstützung bei der Suche relevanter Literatur (z. B. Generierung von Schlagworten und Suchstrategien)

7. Abkürzungsverzeichnis / Abbreviations

Abbreviation	Meaning
2AFC	two-alternative forced choice
AgCl	silver/silver-chloride (electrode material)
ANOVA	analysis of variance
AUC	area under the curve (here: under the ROC curve)
BF_{01}	Bayes factor in favour of the null hypothesis
BF_{10}	Bayes factor in favour of the alternative hypothesis
CI	confidence interval
CV	consonant–vowel (syllable structure)
DFG	Deutsche Forschungsgemeinschaft
EEG	electroencephalography
ERP	event-related potential
FCz	fronto-central midline electrode (10–20 system)
FWE	family-wise error
GEE	generalized estimating equation
GLMM	generalized linear mixed model
IIR	infinite impulse response (filter type)
IQR	interquartile range
ISI	inter-stimulus interval
ISN	Institut für Systemische Neurowissenschaften
ITPC	inter-trial phase clustering
LC-NE	locus coeruleus–norepinephrine system
LME	linear mixed-effects model
LRT	likelihood-ratio test
MEG	magnetoencephalography
MRR	mean reciprocal rank
N1–P2	auditory N1–P2 event-related potential complex
OPT	Opposing Process Theory
P1–P3	within-word position 1 to 3
PCHIP	piecewise cubic Hermite interpolating polynomial
PsPM	PsychoPhysiological Modelling toolbox
R1–R2	random-sequence blocks 1 and 2
R3	hidden-structure block
REML	restricted maximum likelihood
rmANOVA	repeated-measures analysis of variance
ROC	receiver operating characteristic
RT	reaction time
SD	standard deviation
SEM	standard error of the mean
SOA	stimulus onset asynchrony
TFCE	threshold-free cluster enhancement
TG	time-generalized (decoding)
TP	transition probability
W1–W5	structured (word-sequence) blocks 1 to 5

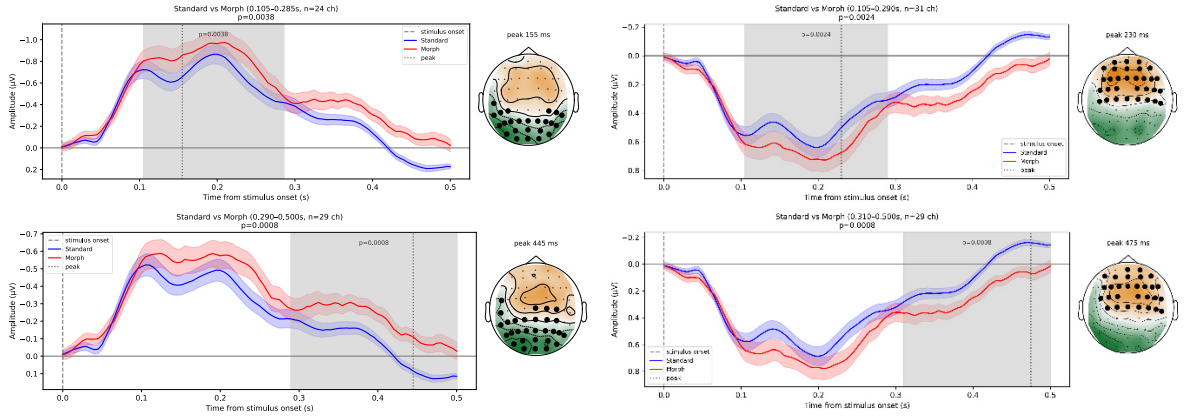
8. Anhang / Supplement

8.1 Supplement ERP

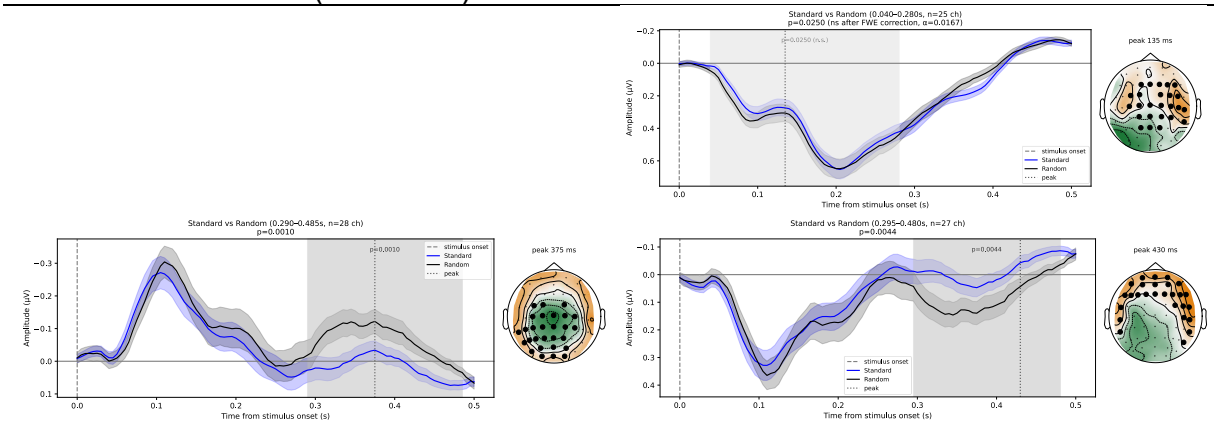
Figure S1. Collection of original electrode clusters and significant ($p < .05$) time windows as returned by the spatiotemporal cluster-based permutation tests (time windows: dark grey = survived FWE correction, light grey = $p < .05$ but did not survive FWE correction).

ERPs per Stimulus Type

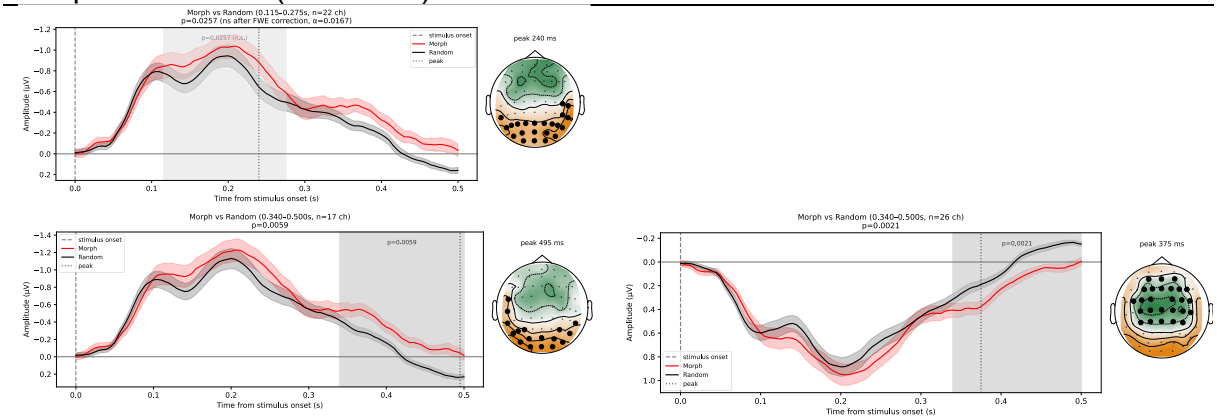
Morph > Standard (4 clusters)



Standard < Random (3 clusters)

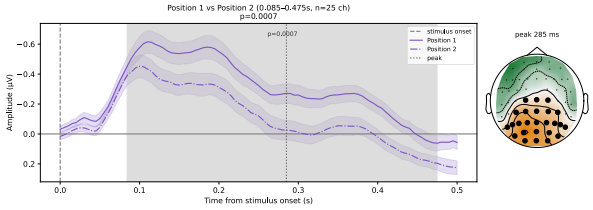


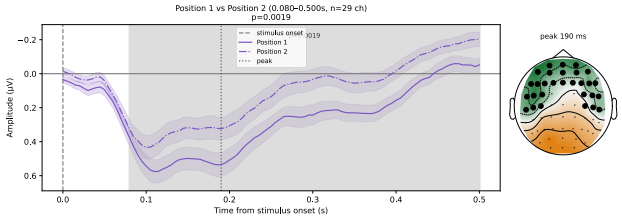
Morph > Random (3 clusters)



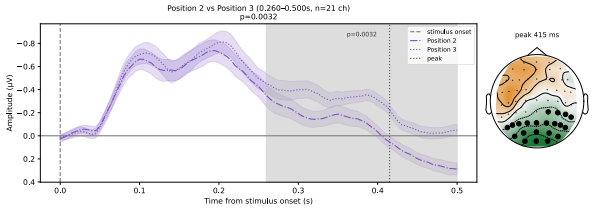
ERPs per Position

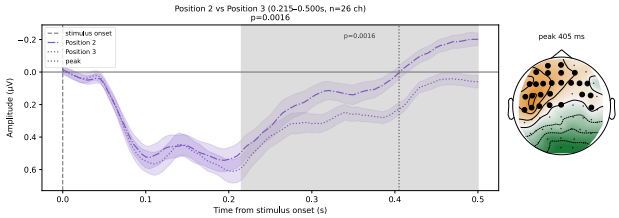
P1 > P2 (2 clusters)



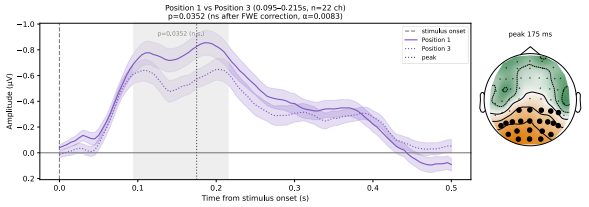


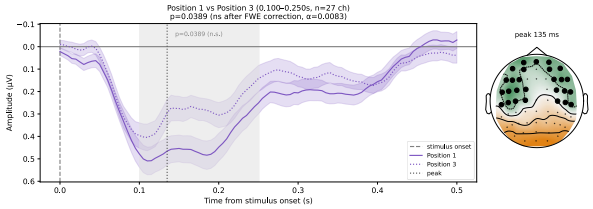
P2 < P3 (2 clusters)





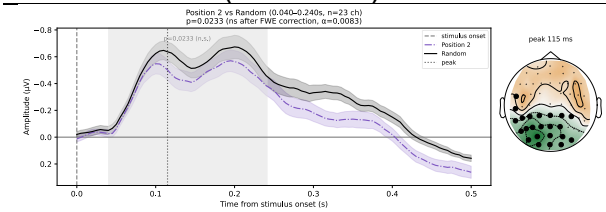
P1 ≈ P3 (fixed threshold) or P1 < P3 (TFCE) (2 clusters)

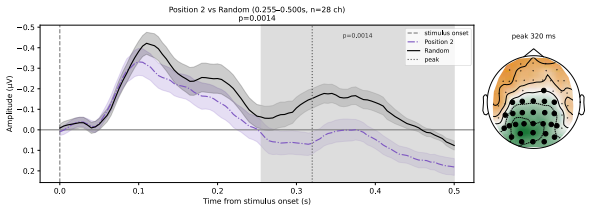


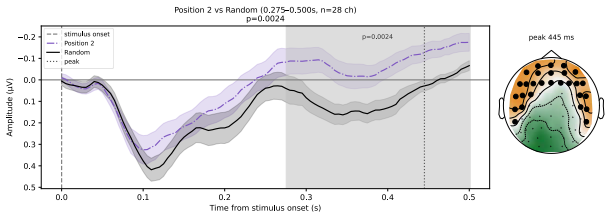


P1 ≈ Random (no clusters)

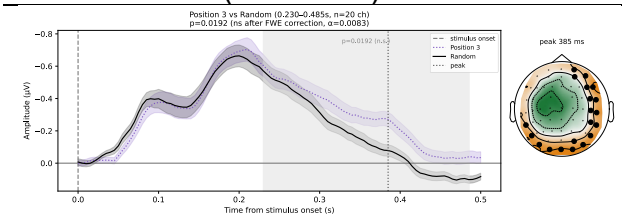
P2 < Random (3 clusters)

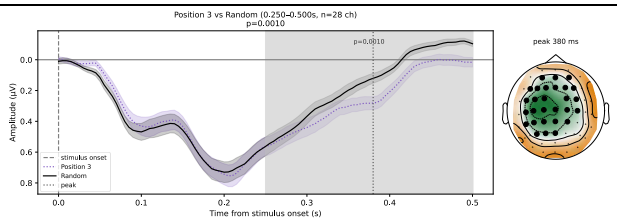






P3 > Random (2 clusters)

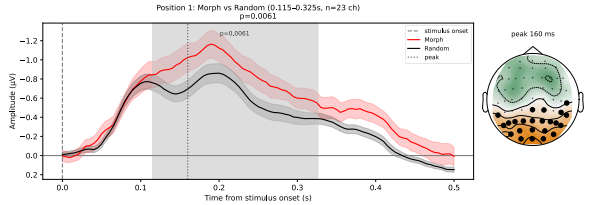


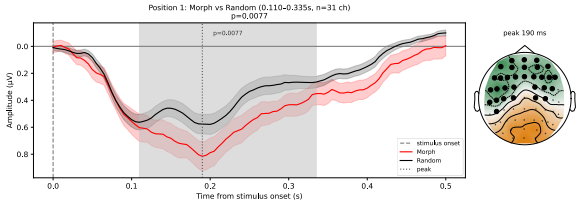


ERPs per Stimulus Type and Position

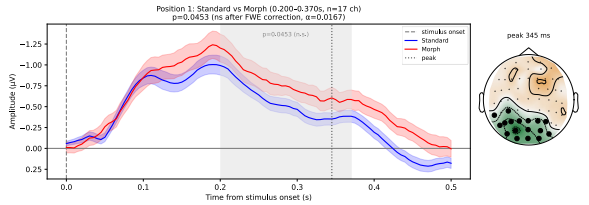
P1: Standard ≈ Random (no clusters)

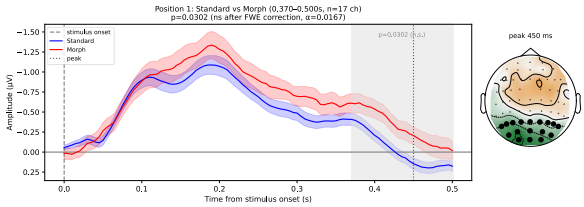
P1: Morph > Random (2 clusters)



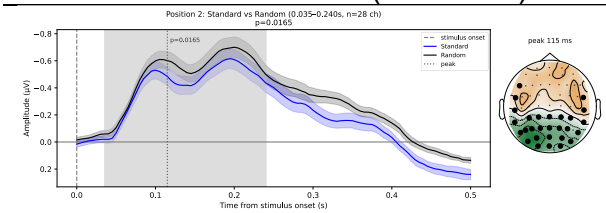


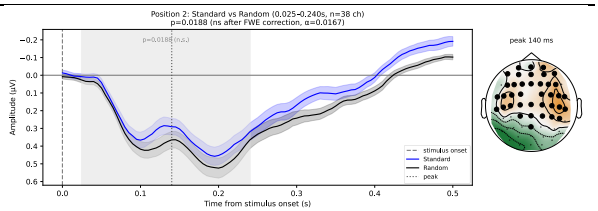
P1: Morph ≈ Standard (2 clusters)

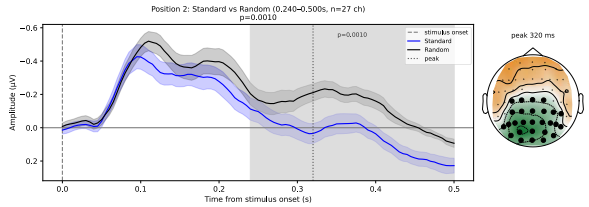


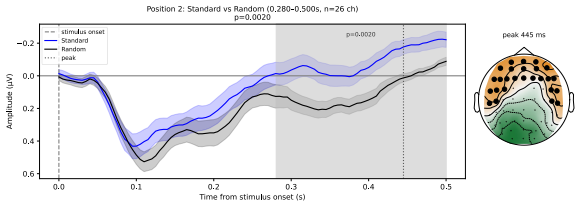


P2: Standard < Random (4 clusters)

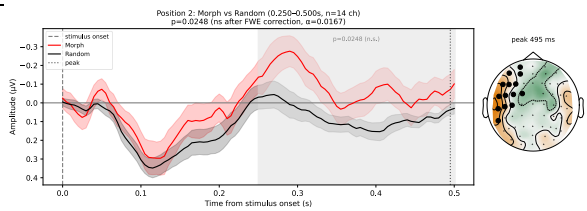




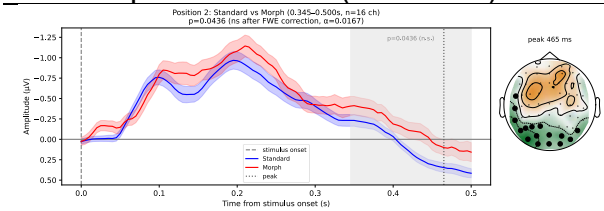


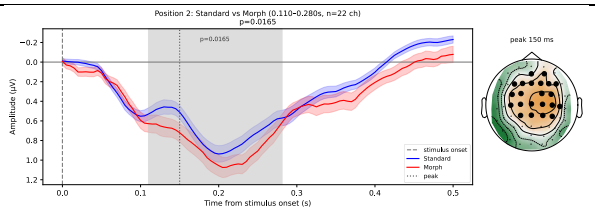


P2: Morph < Random (1 cluster)

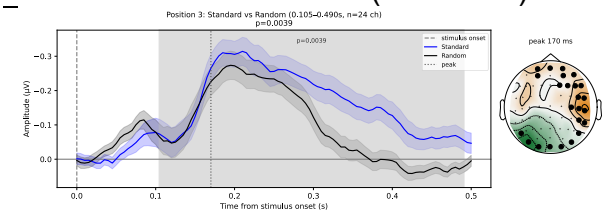


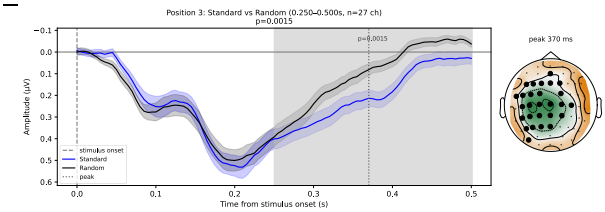
P2: Morph > Standard (2 clusters)



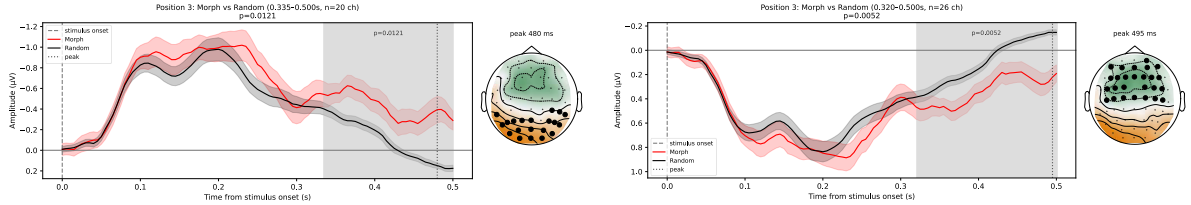


P3: Standard > Random (2 clusters)

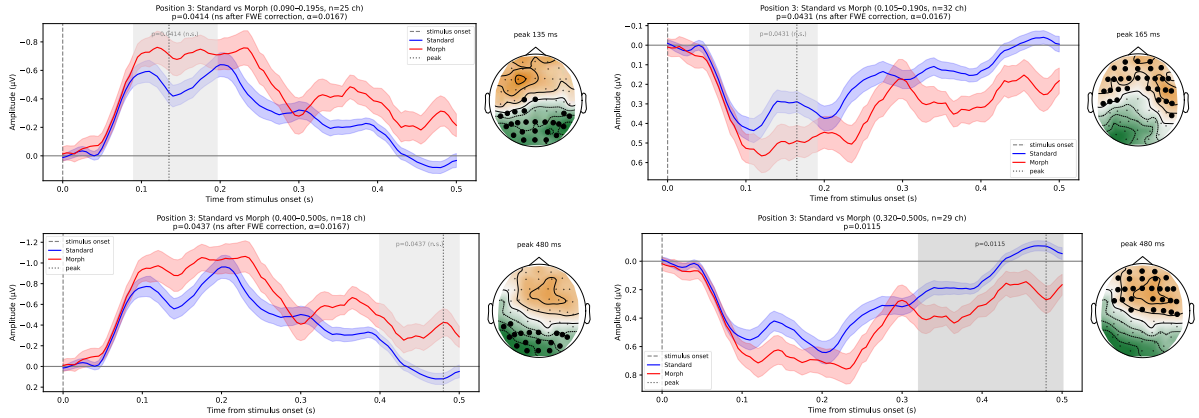




P3: Morph > Random (2 clusters)

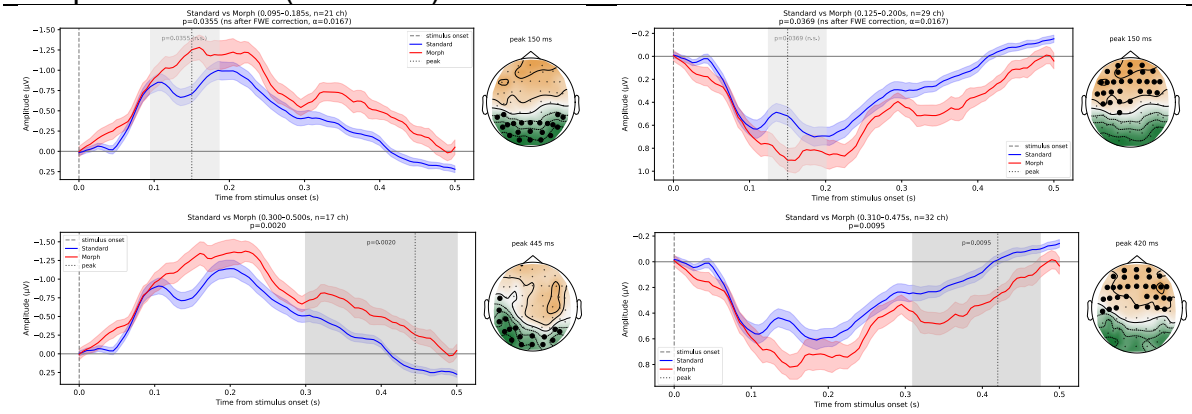


P3: Morph > Standard (4 clusters)



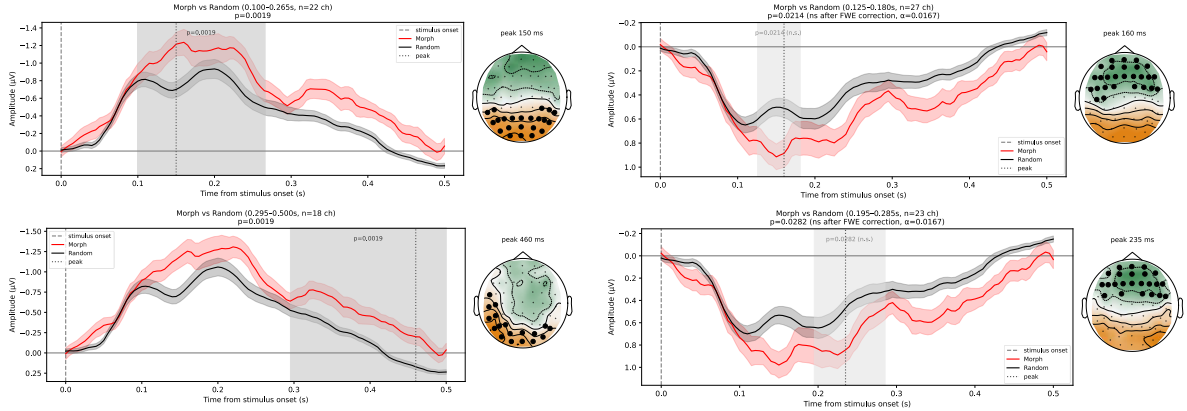
ERPs in Hidden-Structure Block

Morph > Standard (4 clusters)



Standard ≈ Random (none)

Morph > Random



8.2 Supplement: Decoding

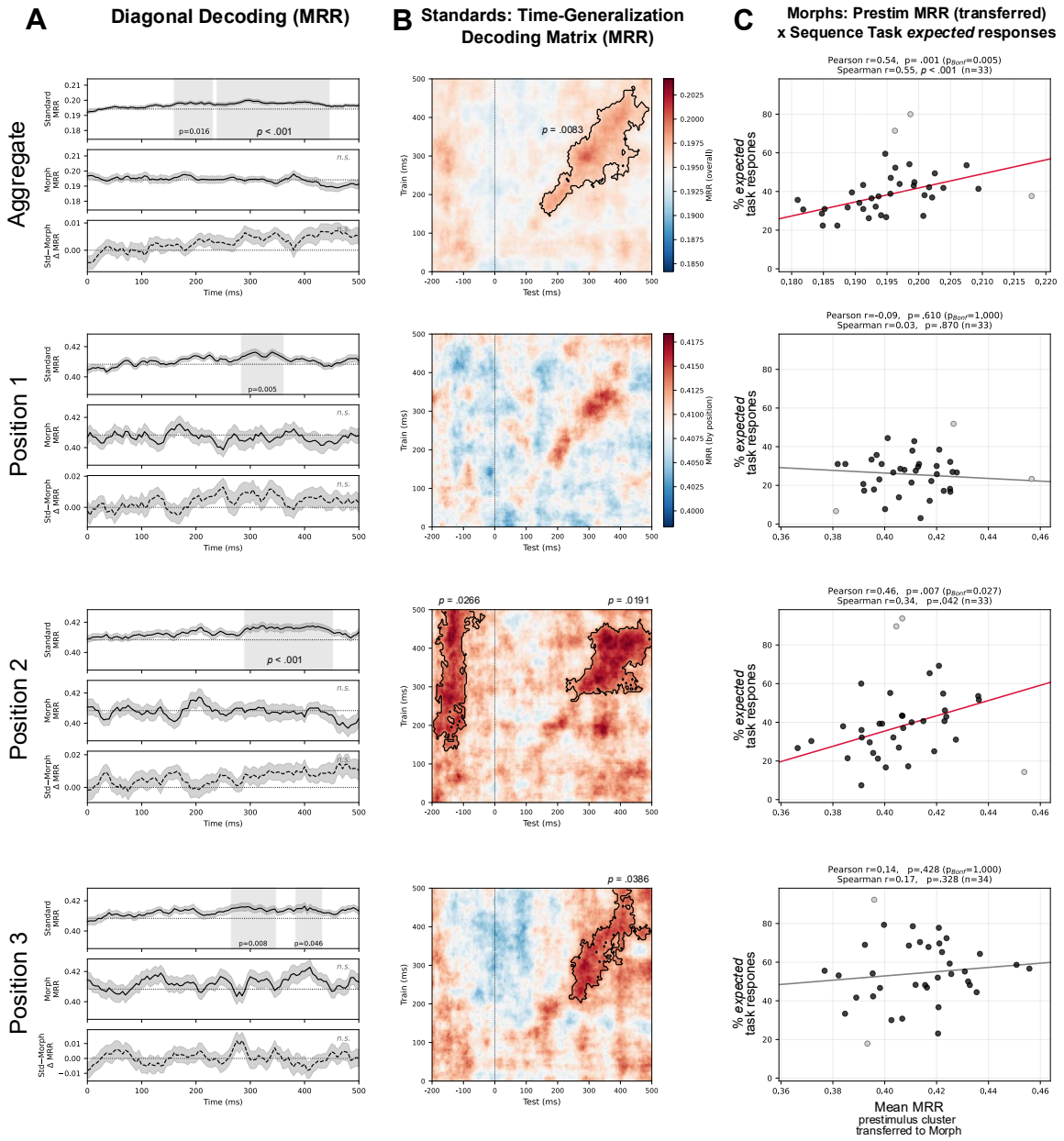


Figure S2. Windowed time-generalized (TG) decoding of Standard syllables and its link to Morph-trial behaviour — MRR (corroborating metric). Direct MRR analogue of the AUC figure (see Figure 3.7), same row/column layout (aggregate, P1–3 × diagonal decoding / TG matrix / behaviour correlation). **(A) Left column**: Standard MRR decoding is significant in every row (all $p \leq .016$); Morph MRR never differs from chance. The Standard−Morph difference wave does not reach significance in any row, but trends in the same direction as AUC throughout (Standard > Morph). **(B) Middle column**: Standard MRR time-generalization matrix; a significant on-diagonal cluster is present in aggregate, at P2, and at P3 (not P1). P2 again uniquely shows an additional pre-stimulus cluster (test < 0 ms, $p = .0266$), alongside its on-diagonal cluster ($p = .0191$). **(C) Right column**: the pre-stimulus cluster coordinates from Standard P2 (middle column), applied to the matching Morph-trial matrices per row (not shown); one point per participant's mean MRR decoding strength in pre-stimulus Morph cluster plotted against how often that participant judged the Morph as *expected* in the Sequence Task. As with AUC, the correlation is significant and Bonferroni-corrected in aggregate ($r = .54$, $p_{bonf} = .005$) and, here more robustly, also at P2 ($r = .46$, $p_{bonf} = .027$); P1 and P3 show no relationship.

Table S1. *ERP suppression/enhancement × decoding correlations (all n.s.).*

Position	Decoding measure	Metric	Pearson r	*p*	BF_{01}	Spearman r	*p*
2	Pre-stimulus cluster	MRR	−.06	.762	4.08	.04	.839
2	Pre-stimulus cluster	AUC	.10	.586	3.95	.03	.863
2	Diagonal (1D)	AUC	−.13	.455	3.58	−.08	.647
2	Diagonal (1D)	MRR	−.05	.785	4.64	−.05	.762
3	Diagonal (1D)	AUC	−.09	.618	4.14	−.13	.461
3	Diagonal (1D)	MRR	−.10	.606	4.17	−.05	.767

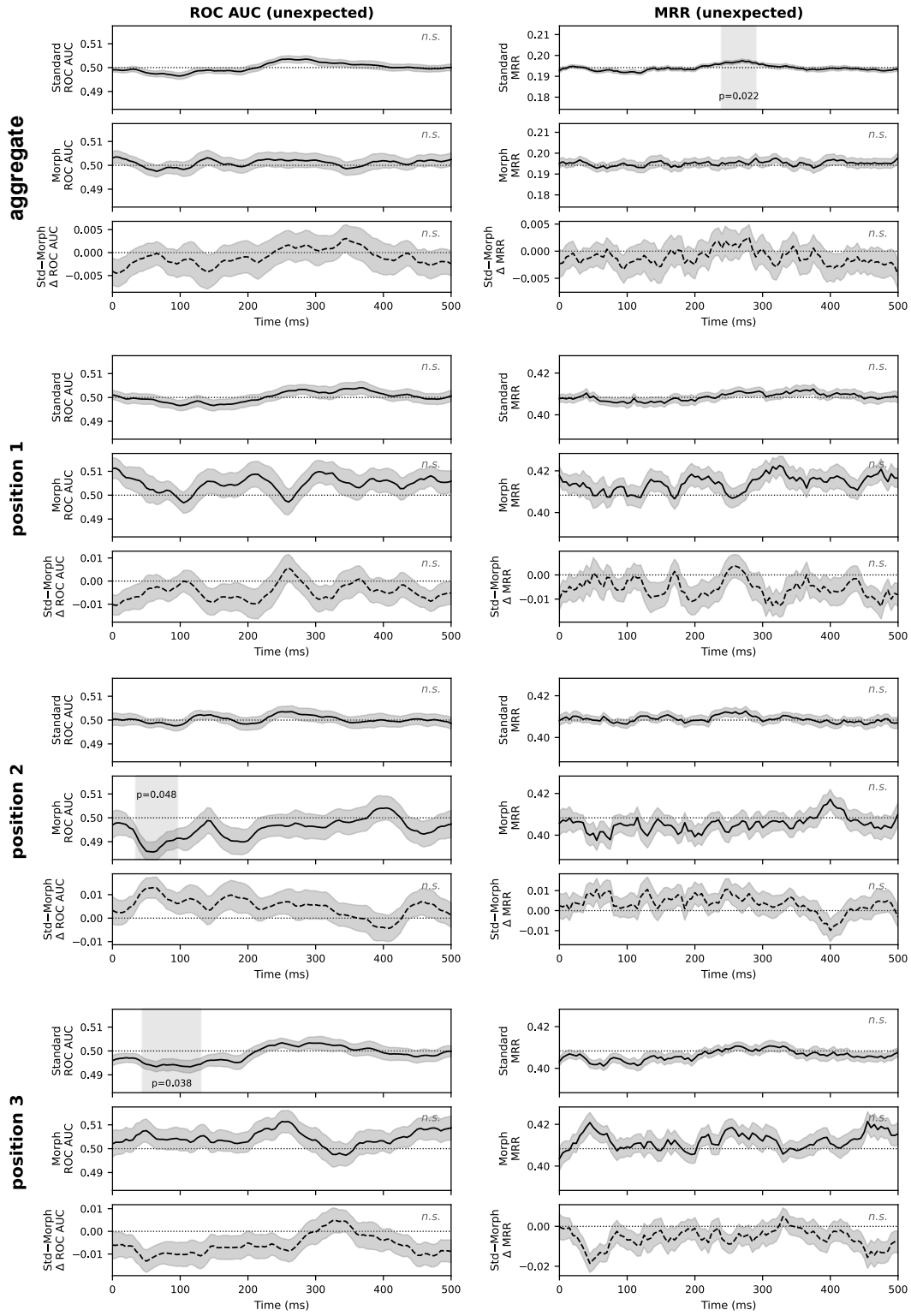


Figure S3. Time-resolved decoding of the unexpected (non-presented) component by position. Rows: aggregate, P1–P3. Columns: AUC (left), MRR (right); each panel shows Standard and Morph 1D diagonal decoding (0–500 ms) plus the Standard−Morph difference wave, with grey bands marking cluster-corrected $p < .05$. Three nominally significant clusters emerge (Morph, P2, AUC: 37–97 ms, $p = .048$; Standard, P3, AUC: 47–132 ms, $p = .038$; Standard, aggregate, MRR: 242–298 ms, $p = .022$); none replicate across metric or condition, and the corresponding 2D time-generalization analysis (not shown) yields no significant cluster for the unexpected component for either Standards or Morphs.

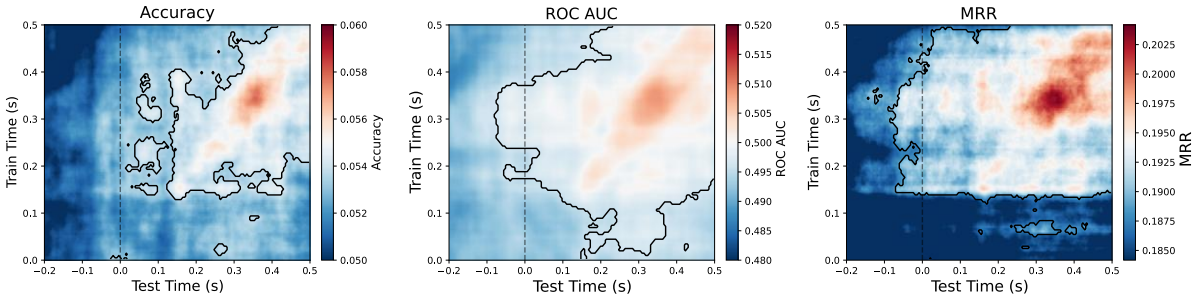


Figure S4. Random-sequence (localizer) time-generalized decoding, all three metrics. Time-generalization matrices (train 0–500 ms × test −200–500 ms, *n* = 36) for the Random-sequence condition, shown for Accuracy (left), AUC (middle), and MRR (right). Black contour = significant cluster per metric (cluster-permutation test); each is significantly below chance, covering 70 % (Accuracy, $p < .0001$), 45 % (AUC, $p = .0013$), and 44 % (MRR, $p = .0006$) of the tested time space. We attribute this to the Random-sequence design's minimum six-syllable repetition lag: residual activity resembling a recently-heard syllable is, by design, evidence against that syllable recurring. A separate, visually prominent above-chance region (~300–400 ms train × 300–500 ms test) is present in AUC and MRR but does not reach significance ($p = .224$ and $p = .192$, respectively). Random-sequence decoding should therefore not be read as straightforward positive-decoding validation; word-sequence generalization (see main figures) is the more direct evidence for the decoding approach.

8.3 Supplement: Reprint – Study 1 (Pupillometry)

Becker, J., Korn, C. W., & Blank, H. (2024). Pupil diameter as an indicator of sound pair familiarity after statistically structured auditory sequence. *Scientific Reports*, 14, 8739. https://doi.org/10.1038/s41598-024-59302-1

scientific reports

Check for updates

OPEN

Pupil diameter as an indicator of sound pair familiarity after statistically structured auditory sequence

Janika Becker[1✉], Christoph W. Korn[1,2] & Helen Blank[1]

Inspired by recent findings in the visual domain, we investigated whether the stimulus-evoked pupil dilation reflects temporal statistical regularities in sequences of auditory stimuli. We conducted two preregistered pupillometry experiments (experiment 1, *n* = 30, 21 females; experiment 2, *n* = 31, 22 females). In both experiments, human participants listened to sequences of spoken vowels in two conditions. In the first condition, the stimuli were presented in a random order and, in the second condition, the same stimuli were presented in a sequence structured in pairs. The second experiment replicated the first experiment with a modified timing and number of stimuli presented and without participants being informed about any sequence structure. The sound-evoked pupil dilation during a subsequent familiarity task indicated that participants learned the auditory vowel pairs of the structured condition. However, pupil diameter during the structured sequence did not differ according to the statistical regularity of the pair structure. This contrasts with similar visual studies, emphasizing the susceptibility of pupil effects during statistically structured sequences to experimental design settings in the auditory domain. In sum, our findings suggest that pupil diameter may serve as an indicator of sound pair familiarity but does not invariably respond to task-irrelevant transition probabilities of auditory sequences.

We constantly sample sensory information from our surroundings and update our inner model of the world based on the perceived relationships in our environment. The extraction of regularities such as the temporal co-occurrence of stimuli is thought to happen unconsciously and quickly, after only a few repetitions of the regularity[1,2]. Based on the perceived regularities, we can generate expectations about our environment. The violation of these expectations, potentially signalling a need for expectation updating, is experienced as surprise[3] and linked to the neuromodulator norepinephrine (NE)[4–6] which is generated in the Locus Coeruleus (LC). The LC-NE system is associated with the regulation of global arousal and projects to neural circuits modulating pupil diameter[7]. Consequently, pupillometry is often used as a proxy for global arousal[8–11] and surprise-driven stimulus responses[3,10,12–16].

Recent studies found stimulus-evoked pupil dilation to respond to transition probabilities in statistically structured visual sequences[12,17], without participants being aware of any rule driving the stimulus presentation or being able to reproduce the statistical regularities. In these studies, participants were presented with sequences of visual stimuli in which the stimulus transitions were governed by different probabilities. In the study of Alamia et al.[12], transitions between stimuli (four letters presented at 1 Hz) followed a probabilistic Markovian process, with some letter transitions occurring more frequently than others. The authors found a larger pupil dilation in response to rare letter transitions in comparison to more frequent transitions in the letter stream. Schwiedrzik and Sudmann[17] presented participants initially with a randomly ordered sequence of 18 faces at a rate of 2 Hz, followed by the same faces presented in a sequence with a pair structure. While they showed pupil cycling at the stimulus presentation rate in both conditions, pupil diameter additionally was smaller in response to the first stimulus than to the second stimulus of a pair in the structured sequence.

As pupil dilation was also shown to respond to stimulus probability in the auditory domain[16,18–21], we conducted two experiments to investigate whether pupil diameter indicates statistical regularities, as defined by stimulus transition probabilities, in auditory sequences. For this purpose, we transferred the visual study design

[1]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany. [2]Section Social Neuroscience, Department of General Psychiatry, University of Heidelberg, 69115 Heidelberg, Germany. ✉email: jan.becker@uke.de

by Schwiedrzik and Sudmann[17] to the auditory domain. We opted for recordings of voices uttering different vowels to align our auditory stimuli with their visual stimuli consisting of face images with different head orientations. Matching their experiment structure, our participants were exposed to sequences of auditory stimuli in two conditions. In the first condition, the stimuli were presented in a randomized order. In the second condition, the same stimuli were temporally structured in pairs (Fig. 1A). In experiment 1, we presented eight voices with a stimulus presentation rate of 0.5 Hz and informed participants about the upcoming pair regularity in the structured sequence. As we could not reproduce the results from the visual domain using these settings, we conducted a second experiment that mirrored the design of the visual study (i.e., 18 face images presented at 2 Hz) more closely[17]. To this end, in experiment 2, we increased the number of voices to 18 and the stimulus presentation rate to 1 Hz (note that 2 Hz was not feasible due to the auditory stimulus duration), we did not inform participants about any statistical rules driving the stimulus sequences, and we conducted a familiarity task after the sequences to determine whether participants learned the pairs presented in the structured condition.

For the familiarity task, we found that pupil diameter following a stimulus pair differed depending on whether participants were exposed to that pair in the structured sequence or not, indicating that pupil diameter can serve as an indicator of learned auditory regularities. However, in both experiments, we did not find the expected pupil response to the transition probabilities during the presentation of the sequences. This observation points out that pupil effects in response to statistically structured sequences are susceptible to the specific experimental settings, potentially particularly in the auditory domain.

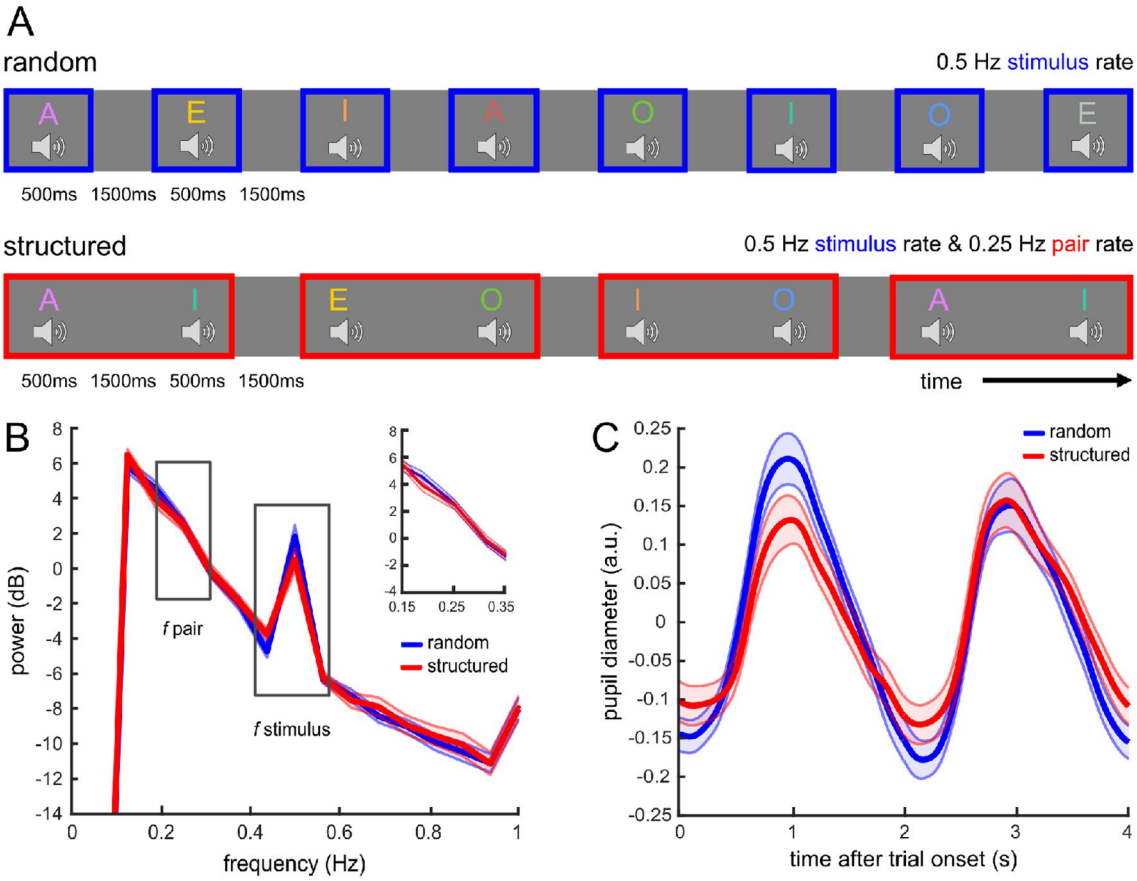


Figure 1. Design and main results of experiment 1. (**A**) Design of experiment 1. Participants were presented with auditory stimuli in two conditions: in the random condition, eight spoken vowels were presented in a random order with a stimulus presentation rate of 0.5 Hz, and afterwards, in the structured condition, the same stimuli were arranged in pairs, resulting in the pair presentation rate of 0.25 Hz. In this illustration, the colours of the vowels represent different voices. A white fixation cross was presented on the screen throughout the experiment. (**B**) Frequency power spectrum. The peak in spectral power at the stimulus presentation rate (0.5 Hz) is significant compared to the mean of the four surrounding frequency bins (*t*-tests, $p < 0.001$) in both the random and the structured condition, but there is no significant peak at the pair presentation rate (0.25 Hz, $p > 0.05$). (**C**) Mean pupil diameter for two consecutively presented stimuli (i.e., a 'trial') averaged over participants in the random (blue) and structured (red) condition (excluding trials with a task event and the following trial). The pupil dilates after sound onset (t = 0 s and t = 2 s) and peaks about one second later. Shading represents standard error of the mean (SEM) across participants.

Results

In two preregistered experiments, we analysed the pupil diameter of 30 (experiment 1) and 31 participants (experiment 2) who listened to auditory sequences of vowels spoken by different speakers. In both experiments, participants were first presented with the stimuli in a random order (random condition), and afterwards, with the same stimuli structured in pairs (structured condition). In the random condition, stimulus transition probabilities were equal between all stimuli, while in the structured condition, the stimulus transition probabilities depended on the pair structure of the sequence. Specifically, the stimulus transition probabilities within pairs were at maximum (i.e., voice 'X' was always followed by voice 'Y') and the stimulus transition probabilities between pairs were at minimum (i.e., stimulus pair XY could be followed by the pairs AB, CD, EF, or XY with the same probability).

In experiment 1, participants listened to sequences consisting of eight stimuli (8 voices, 4 vowels) presented at 0.5 Hz and had to respond to an infrequent visual task (i.e., a luminance change of the fixation cross) and auditory task (i.e., auditory 1-back repetition detection task) (Fig. 1A). Also, they were informed before the structured condition that auditory stimuli would be presented in pairs. In experiment 2, a new sample of participants listened to sequences consisting of 18 stimuli (18 voices, 3 vowels) presented at 1 Hz while responding to an infrequent auditory task (i.e., auditory 1-back repetition detection task) (Fig. 2A). They were not informed about any structure in the stimulus presentation. After listening to the stimulus sequences in experiment 2, participants completed a familiarity task to test whether they learned the stimulus pairs in the structured condition.

In both experiments, behavioural performance in the auditory 1-back task was high across the analysed data blocks (Exp. 1: $M = 89.8\%$, $SD = 6.5$; Exp. 2: $M = 78.7\%$, $SD = 9.8$). Yet we did not find the hypothesized pupil effects of the statistical pair structure during the auditory sequences. However, pupil diameter during the familiarity task after experiment 2 differed between familiar pairs, which participants were previously exposed to in the structured condition, and foil pairs, indicating that pupil diameter reflected the learned association of auditory stimuli.

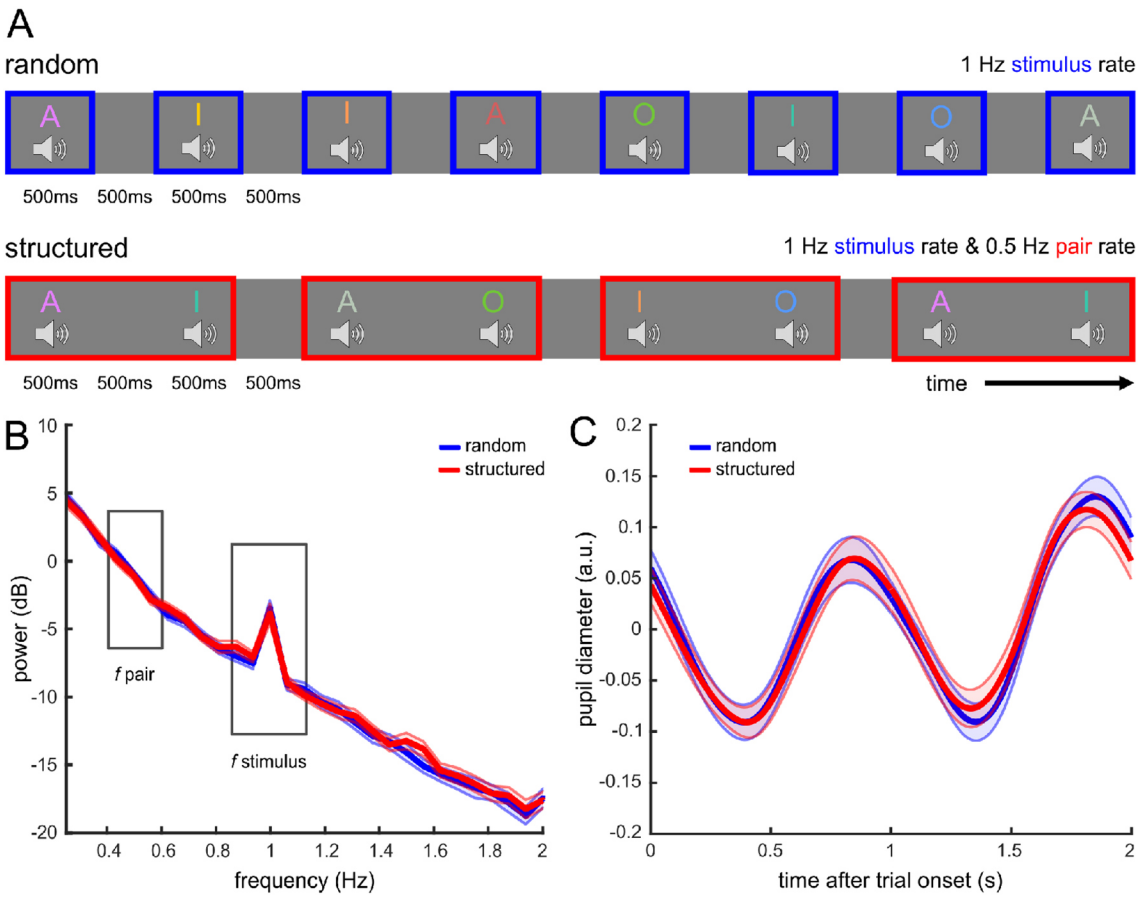


Figure 2. Design and main results of experiment 2. (**A**) Design of experiment 2. Participants listened to 18 different voice stimuli in the random and structured condition, presented at a stimulus presentation rate of 1 Hz. In the structured condition, the pair presentation rate was 0.5 Hz. (**B**) Frequency power spectrum. The power spectrum shows a significant peak in both conditions at the stimulus presentation rate (1 Hz, $ps < 0.001$), but no significant peak at the pair presentation rate (0.5 Hz, $p > 0.05$). (**C**) Pupil diameter locked to the onset of a trial for the random (blue) and the structured (red) condition. Shading represents SEM across participants.

Pupil dilation reflects sound presentation but not statistical pair structure

We investigated whether the pupil dilation responses during the auditory sequences reflect, firstly, the stimulus presentation rate (i.e., dilation in response to every sound), and secondly, in the structured condition, also reflect the pair presentation rate (i.e., dilation in response to every sound pair). Consequently, we tested for two peaks in a frequency power spectrum by comparing the peak value to the mean of the four surrounding frequency bins: one peak at the stimulus presentation rate (Exp. 1: 0.5 Hz; Exp. 2: 1 Hz) in both conditions (random and structured) and an additional peak at the pair presentation rate in the structured condition only (Exp. 1: 0.25 Hz; Exp. 2: 0.5 Hz).

As expected, the frequency power spectra of both experiments showed a distinct peak at the stimulus presentation rate in both conditions (Exp. 1: random condition: $t(29) = 10.318$, $p < 0.001$, $d_z = 1.88$, $BF_{10} > 100$; structured condition, $t(29) = 8.101$, $p < 0.001$, $d_z = 1.48$, $BF_{10} > 100$; Exp. 2: random condition, $t(30) = 9.349$, $p < 0.001$, $d_z = 1.679$, $BF_{10} > 100$; structured condition, $t(30) = 6.544$, $p < 0.001$, $d_z = 1.175$, $BF_{10} > 100$, see Figs. 1B, 2B). However, there was no peak at the pair presentation rate in the structured conditions (Exp. 1: $t(29) = 1.188$, $p = 0.122$, $d_z = 0.217$, $BF_{10} = 0.647$; Exp. 2: $t(30) = 0.163$, $p = 0.436$, $d_z = 0.029$, $BF_{10} = 0.219$). Consequently, we neither found the hypothesized interaction of condition and frequency in a repeated-measures ANOVA (rmANOVA, Exp. 1: $F(1, 29) = 4.093$, $p = 0.052$, $\eta_p^2 = 0.124$, $BF = 0.44$; Exp. 2: $F(1, 30) = 0.119$, $p = 0.733$, $\eta_p^2 = 0.004$, $BF = 0.166$) nor a significant difference in power between the random and structured conditions at the pair presentation rate (Exp. 1: $t(29) = -0.118$, $p = 0.546$, $d_z = -0.022$, $BF_{10} = 0.178$; Exp. 2: $t(30) = -0.387$, $p = 0.649$, $d_z = -0.07$, $BF_{10} = 0.144$).

Additionally, we calculated the intertrial phase clustering (ITPC) to test pupil entrainment to the auditory stimulus and pair presentation. The ITPC across both conditions was higher at the stimulus presentation rate than at the pair presentation rate (Exp. 1: mean difference = 0.34, $t(29) = 11.7$, $p < 0.001$, $d_z = 2.14$, $BF_{10} > 100$; Exp. 2: mean difference = 0.26, $t(30) = 7.23$, $p < 0.001$, $d_z = 1.3$, $BF_{10} > 100$), showing that the pupil dilation tracked the auditory stimuli. However, there was no difference in ITPC at the pair presentation rate between the random and the structured condition (Exp. 1: mean difference = 0.01, $t(29) = -0.65$, $p = 0.52$, $d_z = -0.12$, $BF_{10} = 0.23$; Exp. 2: mean difference = 0.01, $t(30) = 0.72$, $p = 0.48$, $d_z = 0.13$, $BF_{10} = 0.24$), indicating that the pupil dilation did not entrain to the statistical structure of the sound pairs in the structured condition.

Pupil diameter responds to sound presentation but not statistical pair structure

For complementary tests to the frequency power analyses, we extracted the pupil diameter at the onset of a trial. We defined a trial as spanning two consecutively presented stimuli including their following inter-stimulus interval (ISI), respectively (i.e., trial duration: Exp. 1: 4000 ms; Exp. 2: 2000 ms). In the structured condition, a trial consisted of the two stimuli composing a pair. In both experiments, pupil diameter peaked in the time window after each sound onset (Figs. 1C, 2C), corresponding to the peaks at the stimulus presentation rate in the frequency power spectra (Figs. 1B, 2B). In accordance with the frequency power analyses, maximum pupil diameter after sound onset (time window: Exp. 1: 0–2000 ms, Exp. 2: 200–1000 ms) did not show a significant interaction of condition and stimulus position within a trial (rmANOVA: Exp. 1: $F(1, 29) = 3.269$, $p = 0.081$, $\eta_p^2 = 0.101$, $BF = 0.942$; Exp. 2: $F(1, 30) = 0.298$, $p = 0.589$, $\eta_p^2 = 0.01$, $BF = 0.187$). Pupil diameter did not significantly differ between the random and structured condition (no main effect of condition, Exp. 1: $F(1, 29) = 3.232$, $p = 0.083$, $\eta_p^2 = 0.1$, $BF = 0.61$; Exp. 2: $F(1, 30) = 0.724$, $p = 0.402$, $\eta_p^2 = 0.024$, $BF = 0.138$). Also, pupil diameter did not significantly differ between the stimulus positions within a trial in experiment 1 (no main effect of stimulus position: $F(1, 29) = 0.447$, $p = 0.509$, $\eta_p^2 = 0.015$, $BF = 0.155$). In experiment 2, maximum pupil diameter was larger in response to every second stimulus in both conditions (main effect of stimulus position: $F(1, 30) = 9.096$, $p = 0.005$, $\eta_p^2 = 0.233$, $BF = 112.036$; pupil diameter after stimulus 1: $M = 0.128$ a.u., $SD = 0.097$, after stimulus 2: $M = 0.176$, $SD = 0.08$).

Next, we tested for an effect of the sequence structure on minimum pupil diameter during the ISI. In both experiments, minimum pupil diameter during the ISI (time window after sound onset: Exp. 1: 1000–3000 ms, Exp. 2: 700–1200 ms) did neither show an interaction of condition and ISI position within a trial (rmANOVA, Exp. 1: $F(1, 29) = 0.599$, $p = 0.445$, $\eta_p^2 = 0.02$, $BF = 0.2$; Exp. 2: $F(1, 30) = 0.051$, $p = 0.822$, $\eta_p^2 = 0.002$, $BF = 0.184$) nor a main effect of ISI position within a trial (Exp. 1: $F(1, 29) = 0.453$, $p = 0.506$, $\eta_p^2 = 0.015$, $BF = 0.184$; Exp. 2: $F(1, 30) = 1.017$, $p = 0.321$, $\eta_p^2 = 0.033$, $BF = 0.234$). In experiment 1, a significant main effect of condition ($F(1, 29) = 6.446$, $p = 0.017$, $\eta_p^2 = 0.182$, $BF = 0.76$) indicated that minimum pupil diameter was smaller in the random condition ($M = -0.221$ a.u., $SD = 0.114$) than in the structured condition ($M = -0.194$, $SD = 0.114$). There was no main effect of condition in experiment 2 ($F(1, 30) = 0.333$, $p = 0.568$, $\eta_p^2 = 0.011$, $BF = 0.132$).

Finally, we tested whether a differential pupil dilation response to the pair structure developed over time in the structured condition, i.e., whether the pupil dilation response to the stimuli of a pair differed between the beginning (first block) and the end (last block) of the structured condition. For both experiments, maximum pupil diameter showed no significant main effects or interaction, so there was neither a difference in pupil diameter between the first and the last block of the structured condition (rmANOVA, no main effect of block, Exp. 1: $F(1, 20) = 3.394$, $p = 0.08$, $\eta_p^2 = 0.145$, $BF = 0.355$; Exp. 2: $F(1, 23) = 0.004$, $p = 0.952$, $\eta_p^2 < 0.01$, $BF = 0.152$), nor in response to the first and second stimulus within a pair (no main effect of stimulus position, Exp. 1: $F(1, 20) = 0.02$, $p = 0.89$, $\eta_p^2 < 0.001$, $BF = 0.147$; Exp. 2: $F(1, 23) = 2.286$, $p = 0.144$, $\eta_p^2 = 0.09$, $BF = 0.878$), and no significant interaction of block and stimulus position within a pair (Exp. 1: $F(1, 20) = 2.147$, $p = 0.158$, $\eta_p^2 = 0.097$, $BF = 0.386$; Exp. 2: $F(1, 23) = 1.9$, $p = 0.181$, $\eta_p^2 = 0.076$, $BF = 0.467$).

Familiarity task: pupil diameter differs between familiar and foil pairs

After experiment 2, participants completed a familiarity task in which they were presented with 'familiar' stimulus pairs, which were part of the previously presented structured condition, and 'foil' pairs (i.e., pairs consisting

of the first stimulus of one pair and the second stimulus of another pair), which were not presented in the structured condition. After the pair presentation, participants had to indicate via keypress whether the pair was familiar or unfamiliar (Fig. 3A). Behavioural measures did not indicate that participants learned the pairs, as accuracy was not significantly above chance ($M = 52\%$ rated correctly, $SD = 9$, range = 28–67; $t(30) = 1.452$, $p = 0.157$, $d_z = 0.261$, $BF_{10} = 0.495$) and reaction time did not differ between familiar ($M = 1.051$ s, $SD = 0.583$) and foil pairs ($M = 1.135$ s, $SD = 0.522$, $t(30) = -0.926$, $p = 0.362$, $d_z = -0.166$, $BF_{10} = 0.284$). However, maximum pupil diameter after the pair presentation differed significantly between familiar and foil pairs ($t(30) = 2.735$, $p = 0.01$, $d_z = 0.491$, $BF_{10} = 4.32$), with a larger pupil diameter after foil pairs ($M = 0.958$, $SD = 0.366$) than after familiar pairs ($M = 0.728$, $SD = 0.377$, Fig. 3B). We explored whether participants who performed better in the familiarity task also showed a stronger differential pupil response and found no correlation of the mean difference in maximum pupil diameter between foil and familiar pairs with the participants' task accuracy ($r = 0.29$, $p = 0.11$).

Additionally, a linear mixed-effects model showed that maximum pupil diameter was best explained by the true pair category ($\beta = 0.198$, $t(554) = 2.19$, $p = 0.029$), while the estimated pair category as indicated by the participants' response ($\beta = -0.126$, $t(554) = -1.38$, $p = 0.167$) and the reaction time ($\beta = 0.016$, $t(554) = 0.4$, $p = 0.684$) did not significantly contribute to the maximum pupil diameter. The positive estimate of the predictor 'true category' supports the previous finding of a larger pupil diameter in response to foil pairs, whereby the pupil diameter reflects the actual rather than the reported pair category.

In sum, we did not observe an effect of the statistical pair structure on the pupil response during the presentation of the auditory sequences. However, we found that maximum pupil diameter in a subsequent familiarity task distinguished between the stimulus pairs of the structured condition and foil pairs, indicating that participants learned the pairs.

Discussion

In this study, we tested whether the stimulus-evoked pupil dilation during auditory sequences differs depending on the statistical structure of the sequence. In contrast to previous observations of pupil diameter reflecting differences in the statistical structure of visual sequences of letters or faces[12,17], we did not find corresponding pupil effects during auditory sequences of spoken vowels. However, in a subsequent familiarity task, pupil diameter differed between familiar and foil stimulus pairs. We speculate that the extent to which the stimulus-evoked pupil dilation response in the auditory domain reflects learned statistical regularities, based on stimulus transition probabilities, depends on factors such as the sequence complexity, stimulus distinctiveness, and relevance of the regularities to the task at hand.

We conducted two preregistered pupillometry experiments in which participants were exposed to sequences of auditory stimuli (vowels spoken by different speakers). In the 'random' condition, the stimuli were presented in a random order. Afterwards, in the 'structured' condition, the same stimuli were temporally grouped in pairs, such that every second stimulus was predictable based on its predecessor. Participants performed an infrequently and randomly occurring 1-back auditory repetition detection task. In experiment 1, participants listened to 8

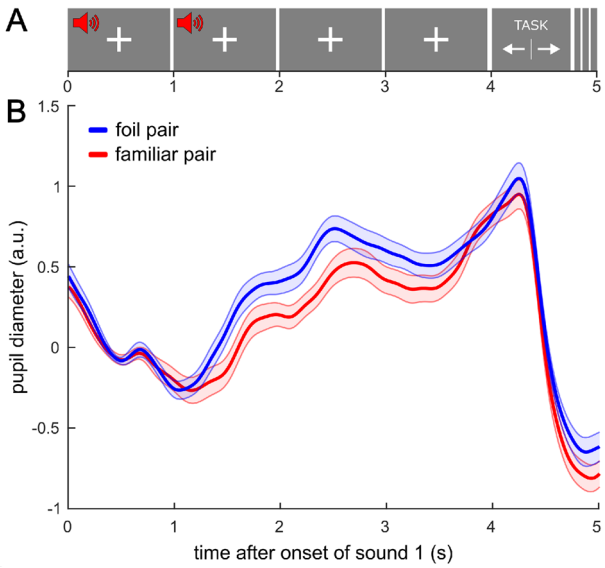


Figure 3. Pupil diameter during the familiarity task (familiar vs. foil pairs). (**A**) Participants heard either a stimulus pair they were exposed to during the structured sequence ('familiar pair') or a new combination of stimuli ('foil pair'). Stimulus 1 was presented at t = 0 s and stimulus 2 at t = 1 s. At t = 4 s, the choice options were presented onscreen and participants had to indicate via button press whether the presented pair was familiar or unfamiliar to them. (**B**) Maximum pupil diameter in the two-second window after the onset of stimulus 2 was larger for foil pairs, indicating that pupil diameter reflected implicit learning of the pairs in the structured condition. Shading represents SEM across participants.

stimuli (i.e., 8 different speakers uttering one of 4 vowels) at a presentation rate of 0.5 Hz. As a recent study[16] found that some participants only exhibited pupil dilation effects in response to violations of global regularity after they were explicitly made aware of the regularity, we informed participants about the pair structure before the start of the structured sequence. For experiment 2, we adapted our experimental design to increase the similarity to the main experiment of Schwiedrzik and Sudmann[17]. For this purpose, we increased the number of auditory stimuli from 8 to 18 (i.e., 18 different speaker-vowel combinations) as well as the stimulus presentation rate (from 0.5 to 1 Hz) and we did not inform participants about the sequence structure. To assess learning of the stimulus pairs in experiment 2, participants also completed a familiarity task after being exposed to the sequences.

During the familiarity task, the pupil diameter was smaller following familiar pairs, i.e., sound pairs which participants were previously exposed to in the structured condition, in comparison to foil pairs, i.e., new pair combinations of the stimuli. This finding, supported by the results of a linear mixed-effects model, suggests that pupil diameter provides a valuable measure to assess implicitly learned stimulus associations, provided that the stimuli are made behaviourally relevant to the listeners. In contrast to the differential pupil response, the behavioural results, i.e., reaction time and accuracy in rating pairs as familiar or unfamiliar, did not indicate that participants explicitly learned the auditory stimulus pairs. This finding is in line with previous studies, where behavioural measures of rule awareness similarly did not indicate that participants had learned the statistical patterns underlying the sequences. (i.e., frequent letter transitions or face pairs, see Alamia et al.[12], experiment 2, familiarity and generative task, and Schwiedrzik and Sudmann[17], experiment 2, card sorting task). In our experimental design, however, the usefulness of reaction time as a measure of learning is constrained due to the delayed response option, allowing participants to prepare their answers in advance.

Our study supports previous studies challenging the broad applicability of pupil diameter as an indicator of statistical learning, emphasizing the need for cautious interpretation[22]. However, we cannot conclude from the absence of pupil dilation in response to statistical regularities in our study that the pupil does not reflect auditory statistical regularities in general. More sensitive measures, such as electroencephalography[23–25], or additional indicators of ocular activity[26], such as electrooculogram, saccades or blinks, may be needed to detect an auditory-regularity effect in both the exposure phase and the familiarity task. In the following sections, we discuss potential reasons for the absence of the expected pupil effect of the pair structure during the auditory sequences.

As the effects of previous visual studies did not easily translate to the auditory domain, the sensory modality might be considered a critical factor for pupil effects of statistical structures during stimulus sequences. The visual and the auditory domain are based on inherently different environmental statistics, such that visual stimuli are more stable and hence changes are more surprising than in the constantly dynamic auditory domain, so there might be modality-specific mechanisms of processing temporal statistical regularities[27,28]. However, there is a body of existing research using pupillometry and auditory stimuli to investigate the processing of probabilities and predictability. For example, sound-evoked (phasic) pupil dilation differed depending on the stimulus occurrence probability[18–21]. Also, sustained (tonic) pupil diameter differed depending on the statistical structure of an auditory sequence[1]. Milne et al.[1] found larger sustained pupil dilation during random auditory sequences than during sequences with a probabilistic or deterministic structure, possibly indicating that predictability facilitated sequence processing in the structured sequences. Generally, auditory processing is thought to have a higher temporal sensitivity than visual processing[27,29,30] and an advantage in sequential processing[28,31]. Therefore, as previous studies showed pupil effects of predictability with auditory stimuli, the modality alone does not appear to account for the lack of pupil effects during the auditory sequences in our study.

A second distinction between our study and previous visual research was the presentation rate. Due to the auditory stimulus length of 0.5 s, the stimulus presentation rate in our experiment 2 (1 Hz) was lower than the presentation rate of the visual stimuli (2 Hz) used by Schwiedrzik and Sudmann[17], but the same as used by Alamia et al.[12] who also reported pupil effects of stimulus transition probabilities in a visual sequence. Furthermore, previous auditory studies with stimulus presentation rates comparable to ours found predictability-related differences in the pupil response to sound sequences. For example, Qiyuan et al.[19] presented auditory sequences of 50-ms tones at 1.5-s intervals (i.e., 0.666 Hz stimulus presentation rate) and reported that the amplitude of the sound-evoked pupil dilation was modulated by tone probability. Predictability-related pupil effects were also observed in musical research, e.g., by Bianco et al.[32] who used 5-s sequences of 20 tones (i.e., 4 Hz stimulus presentation rate) and found a difference in the pupil dilation response between expected and unexpected tones. Although their paradigm included deviants and, thus, is not directly comparable to our paradigm, it shows that predictability-related effects are observable in the pupil response to auditory stimuli with an even higher presentation rate. Therefore, it is unlikely that the absence of probability-related pupil effects in our study is due to the chosen stimulus presentation rates in the sequences.

A dissimilarity between our study and the majority of previous auditory studies is the specific choice of stimuli, as previous studies predominantly used tones[1,19,20,32,33], whereas we used vowels spoken by different speakers. However, some studies have successfully used verbal auditory stimuli to investigate statistical learning based on transition probabilities in sequences[34], especially in the field of language acquisition[35]. In a seminal study, Saffran et al.[36] showed that 8-month-old infants learned the transition probabilities of auditory syllables (i.e., consonants followed by a vowel) within a short period of time. In contrast to their experiment and the large body of studies replicating this effect, our stimulus material consisted of multiple speakers and contained only vowels. According to the consonant–vowel hypothesis[37,38], humans extract words in a continuous artificial stream of speech primarily based on the statistical dependencies of consonants, not vowels. Therefore, the usage of only vowels in our study might have hampered statistical learning of the transition probabilities, even though the familiarity task indicated that participants were able to learn the vowel pairs. Consequently, our paradigm is not fully comparable to the mentioned studies in the domain of language acquisition.

Furthermore, our stimulus material consisted of voice recordings from multiple speakers who naturally differed in their voice characteristics and, thus, potentially in saliency. The stimuli were adjusted to have equal

duration and loudness, while intentionally varying in frequency-related features, such as pitch and formant frequencies, as these are crucial for defining voice and vowel identity[39,40]. Our stimuli were derived from the natural spectrum of frequency variations rather than artificially generated, so any potential differences in saliency are inherently embedded within the defining features of our stimuli. As saliency draws attention[41] and the pupil reflects temporal attention[26], a variation in stimulus-related saliency could have added noise to the observed stimulus-evoked pupil dilation in our study, potentially obscuring a regularity-related effect. Yet, given the balanced stimulus sets over participants and the randomized order of stimuli and pairs in the sequences, any saliency-related pupil responses to single stimuli should be cancelled out during the calculation of intertrial phase clustering or frequency power over the entire sequences. Overall, this supports the conclusion that pupil diameter can reflect statistical regularities but might not be a broadly applicable indicator of regularities in sequences with diverse stimuli.

As fatigue or habituation effects are known to be reflected in the pupil dilation response[42], these might also affect the pupil responses in our experimental design. For example, in experiment 1, minimum pupil diameter was smaller between stimuli in the random condition than in the structured condition, although the effect was not supported by the related Bayes Factor. As the random condition was presented first, it was probably less affected by fatigue or habituation effects, so the pupil diameter amplitude was larger. Nevertheless, it is unlikely that potential regularity-related pupil effects were completely masked by fatigue as the duration of our experiments and the length of blocks are comparable to other sequence experiments that found regularity-related pupil effects in the visual[12,17] or auditory domain[1].

Moreover, although research indicates that statistical learning can happen automatically, rapidly, and unconsciously[2], studies suggesting that statistical learning is task-independent also observed that pupil effects of statistical learning disappeared when attention was diverted from the statistically structured stimuli[12,43]. We included the auditory 1-back repetition task in both experiments to increase participants' engagement and for comparability to the study by Schwiedrzik and Sudmann[17]. As the temporal statistical structure of the sequence was critical for the 1-back repetition task, we assumed that participants paid attention to the temporal structure of the sequence. To prevent any unintended shifts of attention by assigning tasks in two distinct modalities in our first experiment, we removed the visual task in the second experiment. It is unclear whether the 1-back task affected learning of the statistical structure in our experiments: the task could have facilitated the processing of the pair structure or could have disrupted the emerging pair structure by requiring participants to direct their attention to the stimulus repetitions. Unexpected auditory repetitions, i.e., infrequent repetitions of the same stimulus in a sequence of alternating stimuli[44], have been found to draw attention even when to be ignored during a visual memory task[45] and elicit a pupil dilation at first occurrence[6,46]. Hence in our design, participants just needed to react to the attention-drawing violation of the basic regularity of stimulus alternation and could have neglected the pair structure. However, the difference in the pupil response to familiar and foil pairs in the familiarity task indicates that participants processed the pair regularity of the structured sequence at least implicitly.

As a final side note, in experiment 2, we observed a larger pupil diameter in response to every second stimulus in both the random and the structured condition. As there was no statistical structure in the random condition, which the pupil might have responded to, we report this unexpected finding and cautiously interpret this as a coincidental additional occurrence of pupil cycling at 0.5 Hz to the auditory stimuli presented at 1 Hz. Further studies are needed to find out whether this observation is a false positive result, as it is not evident in the corresponding frequency power spectrum, or whether it indeed reflects some kind of automatic pupil cycling at 0.5 Hz during random auditory sequences (potentially in relation to fatigue effects[47] or bottom-up effects of peripheral autonomic function[48]).

In summary, there are several differences between our study design and previous studies, which showed differential pupil responses to regularities in visual sequences as well as to auditory stimuli, that may account for the absent pupil response to the pair structure in our auditory sequences. The pupil effect in our subsequent familiarity task suggests that the behavioural relevance of statistical regularities might be crucial to observe a transition probability-related stimulus-evoked pupil dilation in the auditory domain. Previous visual studies found pupillometry to be a promising method for monitoring the acquisition of sensory statistical regularities in sequences online[12,17], potentially able to replace explicit questioning or subsequent testing of learning success in the form of familiarity or generative tasks. However, our study questions the generalizability of this effect, indicating that this might not readily apply to any sensory sequence and may depend on the interplay of the specific experimental settings, such as the modality, stimulus characteristics, timing or task relevance and difficulty.

The pupil dilation response during the explicit familiarity task indicated that participants learned auditory pairs in the structured condition, but pupil diameter during the structured sequence did not vary with the statistical regularity in our study. In conclusion, pupil diameter may indicate sound pair familiarity but may not universally respond to task-irrelevant transition probabilities in auditory sequences.

Materials and methods

General procedure

In two preregistered pupillometry experiments, participants listened to auditory sequences. The pupillometry measurements took place in a darkened room (experiment 1) and with a dim light behind the participants (experiment 2). Participants placed their heads on a chin support to ensure head stability and a constant distance of 50 cm from the monitor. Using an EyeLink 1000 eye tracker, the gaze and diameter of the participants' right pupils were recorded at a sampling rate of 1000 Hz. The experiment was implemented in Matlab (version

2013b[49]). Auditory stimuli were presented via headphones (Sennheiser HD 201) and button presses were recorded with a conventional PC keyboard.

The auditory sequences consisted of single vowels spoken by different speakers. The sequences were presented in two conditions which differed only in the statistical structure of the stimulus presentation[17]: first, in the random condition, the stimuli were presented in a randomized order. Afterwards, in the structured condition, the same stimuli were temporally arranged in fixed pairs, such that the first stimulus of a pair predicted the subsequently presented second stimulus of that pair. The stimulus pairs were presented in a randomized order. Participants always took part in the random condition first and in the structured condition afterwards. This was to ensure that participants did not have any pair-related expectations about the stimulus presentation structure in the random condition which could have affected the pupil response.

Ethics Statement

Both experiments were approved by the Ethics Committee of the Chamber of Physicians in Hamburg and were performed in accordance with the ethical guidelines and regulations. All participants gave written informed consent to privacy terms and conditions before they participated in the study. Both experiments were preregistered before any data analyses were conducted (Exp. 1: https://osf.io/d8utw, Exp. 2: https://osf.io/8gfn5).

Experiment 1

Participants

A total of $N = 38$ participants were measured for experiment 1. Eight data sets were excluded from the analysis (six sets met the predefined exclusion criteria related to task performance and missing pupil data proportion, two others were excluded due to technical failure during the measurement and the raw visualized data exhibiting large artefacts). Thus, the data sets of a final $n = 30$ participants (21 females and 9 males; $M = 26.2$ years, $SD = 4.0$, range 18–33) were used for further analyses.

Using the software G*Power[50], the preregistered sample size was estimated with a power analysis based on the effect size reported by Schwiedrzik and Sudmann[17] for the difference in spectral power at the pair presentation rate between the structured and random condition: $t(29) = 3.451$; $p = 0.002$; *Hedges* $g = 0.449$, which was transformed to d_z using the formula Cohen's $d_z = \frac{t}{\sqrt{n}}$[51]. Aiming for 0.90 power to achieve an effect size of $d_z = 0.63$ with an alpha error probability of 0.05, the power analysis yielded a target sample size of $N = 29$. As we invited more participants to prospectively compensate for data set exclusion, we ended with full data sets of 30 participants. Participants were recruited via the university job portal and received monetary compensation for their participation. Participants were required to have very good German language skills, be between 18 and 35 years old and physically and mentally healthy. Wearing glasses or contact lenses as well as regular use of medication or other intoxicants were considered exclusion criteria. The participants were asked to refrain from wearing heavy eye makeup and from consuming caffeine before the measurement.

Stimuli, apparatus, and procedure

For our study, we aimed at auditory stimuli that resembled the visual stimuli used in the study by Schwiedrzik and Sudmann[17]. They presented 18 images of different faces with one of three head orientations. To maintain the human character of the stimuli, we decided on voices as stimuli, and to account for the varying head orientation, we let the voices speak different vowels. The extent of similarity between our auditory and their visual stimuli was not quantified.

For our experiment 1, we wanted to facilitate learning of the stimuli and thus reduced the complexity of the sequence by presenting $n = 8$ voices. To balance the mapping of the vowels to the voices, we decided on $n = 4$ different spoken vowels (/a/, /e/, /i/, and /o/). We recorded eight different speakers (4 females, 4 males) of different heights (167–193 cm) and ages (20–55 years) with the microphone TLM 103 by Neumann. The recordings were standardized in volume (75 dB) and length (duration: 0.5 s) with version 2.4.2 of Audacity recording and editing software[52]. The voice characteristics, i.e., formant shift ratio and median pitch, of some speakers were slightly altered using the program Praat[53] to create comparable but distinct voice stimuli. We created a pool of 32 stimuli, with eight voices speaking four vowels, from which we drew a pseudorandomized set of eight stimuli for every participant. It was predefined that every set contained all eight voices (i.e., every voice once) and all four vowels (i.e., every vowel twice). Thus, each participant heard eight different voices, with two voices speaking the same vowel. The specific voice-vowel combinations were pseudorandomly selected for every participant.

Experimental design

In experiment 1, each of the two conditions (random and structured) was about 20 min long and consisted of four blocks with 152 stimulus presentations each. An auditory stimulus was presented every 2 s, i.e., the stimulus presentation rate was 0.5 Hz. Consequently, the pair presentation rate in the structured condition was 0.25 Hz[17] (Fig. 1B). We chose this timing for the stimulus presentation because we found the sound-evoked pupil dilation response to be slow in a previous experiment[21], and similar observations were reported elsewhere[54,55].

For the statistical analyses, we artificially segmented the stimulus sequences in 'trials', whereby we defined a trial as two consecutively presented stimuli and their following ISI. In the random condition, the stimuli were presented in a randomized order, thus, the transition probabilities p for the two stimuli of a trial were the same ($p = 0.14$). In the structured condition, the pair structure of the sequence established different transition probabilities within a trial: while the stimulus transition probability within a pair was maximal ($p = 1$), the transition probability between pairs was minimal ($p = 0.25$). It was predetermined that no stimulus would be presented twice in a row (unless it was a task trial, see below) or three times in a row.

One trial (i.e., two stimuli of 500 ms duration, followed by an ISI of 1500 ms each) had a total duration of 4 s. One block consisted of 76 trials with 152 auditory stimulus presentations and a total duration of 304 s. There was a break of 20 s between each of the four blocks within each condition and a 5-min break between the two conditions. Before the main experiment, participants completed a two-minute training session to get used to the task and the experimental conditions.

Participants had one visual and one auditory task during the experiment to ensure that they paid attention to the auditory stimuli while keeping their eyes open for eye tracking. The auditory task was a 1-back repetition detection task, so participants were asked to press a button whenever the same stimulus (i.e., voice-vowel combination) was presented twice in a row. For the visual task, participants were instructed to press a button in response to a subtle luminance change of the fixation cross (1000 ms) on the screen. Both tasks occurred infrequently and randomly. About 20% of the 76 trials per block contained a task, with half of them being auditory task trials and the other half being visual task trials. The fixation-cross change (i.e., visual task) occurred only during the ISI and, in the structured condition, occurred equally within and between stimulus pairs. Each of the eight auditory stimuli was used for the 1-back repetition task at least once per block. There was a maximum of one task per trial. As explicit knowledge supports the prediction of upcoming stimuli[56], we informed participants in experiment 1 before the start of the structured condition that the sequence would be structured in pairs.

Preprocessing

All preprocessing and data analyses were conducted in Matlab (version 2019b and 2021b[49]) using the publicly available Matlab toolbox PsPM (version 5.0.0[57]).

First, the raw data transmitted by the eye tracker were split into data per block. In a second step, outliers were excluded, which were defined based on gaze. As a change in gaze direction may affect the measurement of pupil diameter[58,59], pupil diameter data points were excluded whenever the respective data points of gaze exceeded the range of four degrees visual angle around the centre of the fixation cross.

Data blocks were excluded when more than 33% of the pupil diameter data were missing after outlier exclusion or when participants missed more than 30% of the auditory task. Based on these criteria, we excluded an average of $M = 1.4$ blocks ($SD = 1.57$, median = 1, range: 0–4) per participant in experiment 1. If more than two blocks per condition were excluded, the entire data set was excluded and replaced with the data set of a new participant. Furthermore, all raw data were visualized and inspected to check for poor data quality. This would be indicated by an abnormal distribution of the power spectra, by large remaining artefacts in the preprocessed data or by other striking anomalies in the visualized data. As stated previously, bad data quality was clearly visible in one data set, which was excluded from further analyses. Gaps in pupil diameter data, due to blinks or outlier exclusion, were closed using a shape-preserving interpolation method (Piecewise Cubic Hermite Interpolating Polynomial, Pchip[60]). Then, filters were applied to remove slow drifts and smooth the data. First, we constructed a low-pass filter at 5 Hz similar to the one used by Schwiedrzik and Sudmann[17], as a one-pass/zero-phase, Kaiser-windowed sinc finite impulse response filter (filter order = 2677; transition width = 2.0 Hz; pass band = 0–4.0 Hz; stop band = 6–500 Hz; maximal pass band deviation = 0.0010 (0.10%); stop band attenuation = − 60 dB[61]).

Initially, as noted in the preregistration, a baseline subtraction per block as well as detrending of the data were planned. However, after data collection, the subtraction of a valid baseline turned out to be unfeasible, because a lot of participants kept their eyes closed in the breaks between the blocks which left no data to draw a baseline from. So instead a two-pass/zero-phase, Butterworth infinite impulse response high-pass filter at 0.1 Hz was applied (filter order = 11; transition width = 0.05 Hz; pass band = 0.1–500 Hz; stop band attenuation = − 60 dB[61]). In a last preprocessing step, the sampling rate was reduced to 500 Hz. As preregistered, we z-standardized all pupil data to account for inter-individual differences in pupil amplitude.

Statistical analyses

The analysis steps are for the most part adopted from the methods described by Schwiedrzik and Sudmann[17]. We expected the pupil to visibly dilate after the presentation of an auditory stimulus in both conditions. Crucially, in the structured condition, we expected the pupil to react differently to each of the two stimuli composing a pair due to the different transition probabilities.

To substantiate the assumption of pupil dilation following auditory stimuli, we tested whether the pupil dilates at the stimulus presentation rate of 0.5 Hz in both conditions. For this purpose, we performed a spectral power analysis. A discrete Fourier transform (DFT) of the pupil data was calculated per block, applying Welch's method[62] as implemented in the Matlab function *pwelch.m*. In order to obtain a spectral resolution of 0.0625 Hz, a Hanning window of 50,820 points (one-third of a block length), 16,940 overlapping points (one-third of the window length), and 8000 DFT points were used. Then, the power spectra were converted to decibels by taking the decadic logarithm and multiplying by 10. After averaging the remaining power spectra over blocks per participant, we computed a mean power spectrum over all participants per condition.

We expected a spectral peak at the stimulus presentation rate (i.e., 0.5 Hz) in the power spectra of both conditions. To statistically test the existence of such, we compared spectral power at 0.5 Hz to the mean of the four surrounding frequency bins (i.e., two above and two below), by means of a one-tailed paired *t*-test for both conditions, respectively. In the structured condition, we additionally expected a spectral peak at the pair presentation rate (i.e., 0.25 Hz), which was also tested by comparing spectral power at 0.25 Hz to the mean of the four surrounding frequency bins (two above, two below) in the structured condition only, using a one-tailed paired *t*-test. Then, we tested whether spectral power at the pair rate was significantly higher in the structured condition than in the random conditions with a one-tailed paired *t*-test. As preregistered and based on the hypothesized results of the former analyses, we conducted a two-way repeated measures ANOVA (rmANOVA)

of spectral power with frequency (2 levels: 0.25 Hz, 0.5 Hz) and condition (2 levels: structured, random) as within-subject factors which we expected to show an interaction of frequency and condition.

For all rmANOVA and paired *t*-tests of pupil diameter, we defined the pupil response to a stimulus as the maximum pupil diameter in the 2000-ms time window after stimulus onset. To calculate maximum and minimum pupil diameter, we averaged pupil diameter during a four-second trial per participant over all included trials per condition (for the analyses with factor condition) or block (for the analyses with factor block). For the analyses of pupil diameter, we included only trials that did not contain a task and neither were preceded by a task trial, as rare events cause massive and several seconds long-lasting pupil dilations[12]. Excluding every task trial and their subsequent trial resulted in the exclusion of an average of $M = 27.46$ trials (range: 22–32) out of 76 trials per block per participant in experiment 1. Considering the relatively high percentage of excluded trials, we reduced the proportion of task trials in experiment 2 by increasing the overall amount of trials and removing the visual task.

To test whether maximum pupil diameter following the stimuli was different between conditions and if this difference was specific to the stimulus position within a trial (which defines the stimulus probability), a two-way rmANOVA of maximum pupil diameter was conducted with the within-subject factors condition (2 levels: structured, random) and stimulus position (2 levels: first stimulus, second stimulus).

Based on the results of Schwiedrzik and Sudmann[17], we tested whether minimum pupil diameter following the ISI within pairs differed from minimum pupil diameter following the ISI between pairs in the structured condition (Fig. 1C). For this purpose, we conducted a two-way rmANOVA of pupil diameter following an ISI with condition (2 levels: structured, random) and position of the ISI within a trial (2 levels: first ISI, second ISI) as within-subject factors. Pupil diameter following an ISI was defined as the minimum pupil diameter in the time window from 1000 to 3000 ms after stimulus onset.

Furthermore, we wanted to determine whether pupil diameter following the two stimuli of a pair in the structured condition changed over time, as participants learned the pairs and thereby the associated stimulus probabilities. To test this, pupil dimeter following the first and second stimulus of a pair was compared between the first and last block of the structured condition. For this purpose, a two-way rmANOVA of pupil diameter following a stimulus was performed with block (2 levels: first block, last block) and stimulus position within a pair (2 levels: first position, second position) as within-subject factors.

To test whether the pupil dilation entrained to the auditory stimuli presented at 0.5 Hz and to the pairs presented at 0.25 Hz in the structured condition, we calculated intertrial phase clustering (ITPC)[17,63]. ITPC indicates the consistency of an oscillation's phase angle across trials. Possible ITPC values range from a minimum of 0, when all phase angles are uniformly distributed, to a maximum of 1, when phase angles are identical across trials. To calculate ITPC, the preprocessed pupil data were divided into pseudo trials of 16 s, starting with the onset of the first stimulus of a pair (the timing was chosen based on the procedure in the visual study[17]). Using the Matlab toolbox Fieldtrip (version 20240214[64]), we performed a Fourier transform for each pseudo trial per block and participant (tapers: discrete prolate spheroidal sequences (DPSSs); spectral resolution: 0.0625 Hz; minimal spectral smoothing), resulting in a range of complex numbers. ITPC was computed with these complex numbers and the formula:

$$ITPC = \left| n^{-1} \sum_{r=1}^{n} e^{ik_{fr}} \right|$$

for n trials with the phase angle k on trial r at the frequency f. We calculated the mean ITPC per participant and condition across included blocks at the stimulus and pair presentation rate and conducted statistical comparisons between conditions and rates using paired *t*-tests.

The significance of analyses was assumed based on the standard $p < 0.05$ criterion. Conducted tests of difference were two-tailed unless stated otherwise. As effect sizes, we report partial eta squared for rmANOVA and Cohen's d_z for paired *t*-tests[51]. Also, we report Bayes Factors for all analyses. Bayes Factors of main factors and interactions in rmANOVA are computed by dividing the Bayes Factor of the full model by the restricted model not including the respective main effect or interaction.

Experiment 2

We conducted a second experiment with adjusted settings to make the design more similar to the one of Schwiedrzik and Sudmann[17]. The methods of experiment 2 correspond to experiment 1 if not stated otherwise in the following.

Participants

A new sample of $N = 35$ participants was tested for experiment 2. Five data sets were excluded from the analysis based on the predefined exclusion criteria relating to task performance and missing pupil data proportion. Thus, the data sets of a final $n = 31$ participants (22 females and 9 males; $M = 27.7$ years, $SD = 4.9$, range 20–38) were used for further analyses. The sample size was based on experiment 1 and the sample size reported by Schwiedrzik and Sudmann[17]. For experiment 2, we included healthy participants with corrected-to-normal vision using soft contact lenses.

Stimuli, apparatus, and procedure

For experiment 2, we adjusted and extended the stimulus pool from experiment 1. To more closely relate to the original study by Schwiedrzik and Sudmann[17] using images of 18 face identities with three head orientations, we used 18 voices and three vowels (i.e., /a/, /i/, and /o/). To this end, we recorded 10 more voices (5 females, 5 males) of different height (160–199 cm) and age (25–59 years) with the same microphone as in experiment 1. A total of 18 voices speaking 3 vowels resulted in a pool of 54 stimuli. Each participant listened to a pseudorandomized

set of 18 stimuli. We predetermined that all 18 voices were included in a set and that the number of vowels was balanced. As we had three different vowels, every vowel appeared six times in a set. The mapping of vowels and the speakers' gender was also balanced, thus, every vowel was spoken by three male and three female voices.

Experimental design

To make the experimental design more similar to the timing of the main experiment by Schwiedrzik and Sudmann[17], we increased the stimulus presentation rate for our experiment 2, i.e., the stimulus presentation rate was 1 Hz and the pair presentation rate in the structured condition was 0.5 Hz (Fig. 2B). Both conditions were about 11 min long and consisted of four blocks with 162 stimulus presentations each. An auditory stimulus was presented every second, so one trial (i.e., two stimuli of 500 ms duration, followed by an ISI of 500 ms each) had a total duration of 2 s. One block consisted of 81 trials with a total duration of 162 s. There was a break of 10 s between each of the four blocks of each condition and about 5 min break between the two conditions. Before the experiment, participants completed a two-minute training session to get used to the task and the experimental conditions.

To make experiment 2 more similar to Schwiedrzik and Sudmann[17], participants were only assigned with the auditory 1-back task, equivalent to their visual 1-back task. About 11% of the 81 trials per block were task trials. Per condition, each of the 18 auditory stimuli comprised the task trial at least twice. Task trials occurred randomly in the sequence but task trials did not appear consecutively and neither within the first 3 trials of a block. Also, unlike experiment 1, we did not inform the participants about any structures of the sequences.

Familiarity task

After listening to the sequences in both conditions in experiment 2, participants completed a familiarity test with 18 trials. In each trial, they were presented with a pair of two auditory stimuli. Their task was to categorize the pair via button press as familiar or unfamiliar. Throughout a trial, a white fixation cross was presented onscreen. The trial started with the first auditory stimulus, followed by the second auditory stimulus after one second (Fig. 3A). Four seconds after trial onset, the familiarity task instruction was displayed on the screen, asking participants to press either the left or right arrow key depending on whether the order of voices seemed familiar or unfamiliar to them. The response via key press initiated the white fixation cross display and a silent period of three seconds until the next trial started. Participants were presented with the 9 familiar stimulus pairs, i.e., the ones that participants heard in the structured condition, as well as 9 foil pairs, i.e., a random new combination of the first stimulus of each pair with the second stimulus of a different pair. The trials were presented in a randomized order.

Preprocessing

The preprocessing of the pupil data was performed similarly to experiment 1. As the frequencies of interest changed, we adjusted the high-pass filter to 0.25 Hz. (i.e., pass band = 0.25–5 Hz). Also, as the length of blocks changed, the Hanning window for the Welch's method was adjusted fit one-third of a block length and the overlapping points to fit one-third of the window length. For experiment 2, data blocks were excluded when more than 33% of the pupil diameter data were missing after outlier exclusion or when participants missed more than 50% of the auditory task trials. If less than two blocks for either condition were available for analyses, the entire data set was excluded and replaced with the data set of a new participant. We excluded an average of $M = 0.97$ blocks ($SD = 1.17$, median = 0, range: 0–4) per participant in experiment 2. Excluding task trials and the trials following task trials from analyses of pupil diameter resulted in the exclusion of an average of $M = 17.87$ trials (range: 16–20) out of 81 trials per block per participant.

Statistical analyses

Main experiment. We conducted the same statistical analyses of pupil diameter for experiment 2 as for experiment 1. Due to the different timing, the analyses of spectral peaks and ITPC are based on the stimulus presentation rate at 1 Hz and pair presentation rate at 0.5 Hz. Pupil diameter maxima were identified in the time window of 200–1000 ms after stimulus onset and minima were calculated in the time window from 700 to 1500 ms after stimulus onset. We chose these time windows to prevent identifying a maximum or minimum value from the response to the preceding stimulus or ISI, as the pupil response is slightly delayed as visible in Fig. 2C.

Familiarity task. To check whether participants learned the pairs in the structured condition, we analysed performance and pupil diameter in the familiarity task. Performance was assessed by comparing reaction times to familiar and foil pairs by means of paired *t*-tests and by calculating whether performance accuracy was above chance. Reaction time was measured as the time in seconds of the keypress since the onscreen display of the task. We also tested whether the pupil responded differently to the familiar and foil pairs presented in the familiarity task and, thus, may serve as an indirect measure of familiarity. For this end, we preprocessed the pupil data in the same way as for the main experiment and computed the average pupil diameter per participant over all 9 trials per condition (familiar and foil pairs). Corresponding to the analyses in the experiment 1, we identified maximum pupil diameter in the 2000 ms window after onset of the second stimulus and compared maximum pupil diameter between familiar and foil pairs using a paired *t*-test. As participants could not judge the familiarity of the pair based on the first presented stimulus, but only based on the second one, we defined a 1 s-baseline before onset of the second stimulus. Additionally, to test whether participants who performed better in the familiarity task exhibited a stronger differential pupil response, we used Pearson's correlation coefficient to measure the relationship of the mean difference in pupil diameter between foil and familiar pairs per participant with their percentage of correct categorization of the pairs.

Furthermore, we tested to which extent the pupil response reflects the actual auditory pair category and the behavioural measures of the participants (i.e., the estimated pair category as indicated by the keypress or the reaction time). To this end, we fitted a linear mixed-effects model in which we predicted the maximum pupil diameter in the familiarity task with the three fixed factors 'true category' (categorical; familiar = 0, foil = 1), 'indicated category' (categorical; familiar = 0, foil = 1) and 'reaction time' (continuous) as well as a random intercept for participants to account for the repeated-measures design. Note that the three fixed predictors are not correlated, as the correlation values between two predictor variables across participants do not differ from zero (true category and indicated category: mean correlation = 0.05, $t(29) = 1.44$, $p = 0.16$, $d_z = 0.26$, $BF_{10} = 0.49$; true category and reaction time: mean correlation = 0.08, $t(30) = 1.83$, $p = 0.08$, $d_z = 0.33$, $BF_{10} = 0.84$; indicated category and reaction time: mean correlation = 0.01, $t(29) = 0.22$, $p = 0.83$, $d_z = 0.04$, $BF_{10} = 0.2$).

Data availability

The datasets analysed in the current study are available in the OSF repository https://osf.io/t465x/.

Received: 11 December 2023; Accepted: 9 April 2024
Published online: 16 April 2024

References

1. Milne, A. E., Zhao, S., Tampakaki, C., Bury, G. & Chait, M. Sustained pupil responses are modulated by predictability of auditory sequences. *J. Neurosci.* **41**, 6116–6127 (2021).
2. Turk-Browne, N. B. Statistical learning in perception. in *Encyclopedia of the Sciences of Learning* (ed. Seel, N. M.) 3182–3185 (Springer, 2012). https://doi.org/10.1007/978-1-4419-1428-6_1707.
3. Preuschoff, K., 't Hart, B. & Einhauser, W. Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Front. Neurosci.* **5**, 115 (2011).
4. Payzan-LeNestour, E. & Bossaerts, P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput. Biol.* **7**, e1001048 (2011).
5. Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–692 (2005).
6. Zhao, S. *et al.* Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. *Nat. Commun.* **10**, 4030 (2019).
7. Joshi, S. & Gold, J. I. Pupil size as a window on neural substrates of cognition. *Trends Cogn. Sci.* **24**, 466–480 (2020).
8. Krishnamurthy, K., Nassar, M. R., Sarode, S. & Gold, J. I. Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nat. Hum. Behav.* **1**, 1–11 (2017).
9. Murphy, P. R., Robertson, I. H., Balsters, J. H. & O'connell, R. G. Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology* **48**, 1532–1543 (2011).
10. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15**, 1040–1046 (2012).
11. Urai, A. E., Braun, A. & Donner, T. H. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nat. Commun.* **8**, 14637 (2017).
12. Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A. & Zenon, A. Pupil-linked arousal responds to unconscious surprisal. *J. Neurosci.* **39**, 5369–5376 (2019).
13. de Berker, A. O. *et al.* Computations of uncertainty mediate acute stress responses in humans. *Nat. Commun.* **7**, 10996 (2016).
14. Kloosterman, N. A. *et al.* Pupil size tracks perceptual content and surprise. *Eur. J. Neurosci.* **41**, 1068–1078 (2015).
15. Lavin, C., San Martín, R. & Rosales Jubal, E. Pupil dilation signals uncertainty and surprise in a learning gambling task. *Front. Behav. Neurosci.* **7**, 218 (2014).
16. Quirins, M. *et al.* Conscious processing of auditory regularities induces a pupil dilation. *Sci. Rep.* **8**, 14819 (2018).
17. Schwiedrzik, C. M. & Sudmann, S. S. Pupil diameter tracks statistical structure in the environment to increase visual sensitivity. *J. Neurosci.* **40**, 4565–4575 (2020).
18. Friedman, D., Hakerem, G., Sutton, S. & Fleiss, J. L. Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. *Electroencephalogr. Clin. Neurophysiol.* **34**, 475–484 (1973).
19. Qiyuan, J., Richer, F., Wagoner, B. L. & Beatty, J. The pupil and stimulus probability. *Psychophysiology* **22**, 530–534 (1985).
20. Silvestrin, F., Penny, W. D. & FitzGerald, T. H. B. Pupil dilation indexes automatic and dynamic inference about the precision of stimulus distributions. *J. Math. Psychol.* **101**, 102503 (2021).
21. Becker, J., Viertler, M., Korn, C. W. & Blank, H. The pupil dilation response as an indicator of visual cue uncertainty and auditory outcome surprise. *Eur. J. Neurosci.* https://doi.org/10.1111/ejn.16306 (2024).
22. Zhang, F. & Emberson, L. L. Opposing timing constraints severely limit the use of pupillometry to investigate visual statistical learning. *Front. Psychol.* **10**, 465752 (2019).
23. Benjamin, L. *et al.* Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *Dev. Sci.* **26**, e13300 (2023).
24. Batterink, L. J., Mulgrew, J. & Gibbings, A. Rhythmically modulating neural entrainment during exposure to regularities influences statistical learning. *J. Cogn. Neurosci.* **36**, 107–127 (2024).
25. Aller, M., Økland, H. S., MacGregor, L. J., Blank, H. & Davis, M. H. Differential auditory and visual phase-locking are observed during audio-visual benefit and silent lip-reading for speech perception. *J. Neurosci.* **42**, 6108–6120 (2022).
26. Jin, P., Zou, J., Zhou, T. & Ding, N. Eye activity tracks task-relevant structures during speech and auditory sequence perception. *Nat. Commun.* **9**, 5374 (2018).
27. Frost, R., Armstrong, B. C., Siegelman, N. & Christiansen, M. H. Domain generality versus modality specificity: The paradox of statistical learning. *Trends Cogn. Sci.* **19**, 117–125 (2015).
28. Conway, C. M. & Christiansen, M. H. Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol.: Learn. Mem. Cogn.* **31**, 24–39 (2005).
29. Chen, L. & Vroomen, J. Intersensory binding across space and time: A tutorial review. *Atten. Percept. Psychophys.* **75**, 790–811 (2013).
30. Emberson, L. L., Conway, C. M. & Christiansen, M. H. Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Q. J. Exp. Psychol.* **64**, 1021–1040 (2011).
31. Kemény, F. & Lukács, Á. Sequence in a sequence: Learning of auditory but not visual patterns within a multimodal sequence. *Acta Psychol.* **199**, 102905 (2019).
32. Bianco, R., Ptaszynski, L. E. & Omigie, D. Pupil responses to pitch deviants reflect predictability of melodic sequences. *Brain Cogn.* **138**, 103621 (2020).
33. Southwell, R. *et al.* Is predictability salient? A study of attentional capture by auditory patterns. *Philos. Trans. R. Soc. B* **372**, 20160105 (2017).

34. Hochmann, J.-R. & Papeo, L. The invariance problem in infancy: A pupillometry study. *Psychol. Sci.* **25**, 2038–2046 (2014).
35. Romberg, A. R. & Saffran, J. R. Statistical learning and language acquisition. *WIREs Cogn. Sci.* **1**, 906–914 (2010).
36. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
37. Bonatti, L. L., Peña, M., Nespor, M. & Mehler, J. Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychol. Sci.* **16**, 451–459 (2005).
38. Toro, J. M., Nespor, M., Mehler, J. & Bonatti, L. L. Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychol. Sci.* **19**, 137–144 (2008).
39. Krumbiegel, J., Ufer, C. & Blank, H. Influence of voice properties on vowel perception depends on speaker context. *J. Acoust. Soc. Am.* **152**, 820–834 (2022).
40. Kreitewolf, J., Gaudrain, E. & von Kriegstein, K. A neural mechanism for recognizing speech spoken by different speakers. *NeuroImage* **91**, 375–385 (2014).
41. Alink, A. & Blank, H. Can expectation suppression be explained by reduced attention to predictable stimuli?. *NeuroImage* **231**, 117824 (2021).
42. Zekveld, A. A., Koelewijn, T. & Kramer, S. E. The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends Hear.* **22**, 2331216518777174 (2018).
43. Richter, D. & de Lange, F. P. Statistical learning attenuates visual activity only for attended stimuli. *eLife* **8**, e47869 (2019).
44. Heilbron, M. & Chait, M. Great expectations: Is there evidence for predictive coding in auditory cortex?. *Neuroscience* **389**, 54–73 (2018).
45. Hughes, R. W., Vachon, F. & Jones, D. M. Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *J. Exp. Psychol. Learn. Mem. Cogn.* **33**, 1050–1061 (2007).
46. Marois, A., Pozzi, A. & Vachon, F. Assessing the role of stimulus novelty in the elicitation of the pupillary dilation response to irrelevant sound. *Audit. Percept. Cogn.* **3**, 1–17 (2020).
47. Warga, M., Lüdtke, H., Wilhelm, H. & Wilhelm, B. How do spontaneous pupillary oscillations in light relate to light intensity?. *Vis. Res.* **49**, 295–300 (2009).
48. Madsen, J. & Parra, L. C. Brain-body interaction during natural story listening. Preprint at https://doi.org/10.1101/2023.01.31.526511 (2023).
49. The MathWorks Inc. MATLAB [Computer software]. Versions 2013b, 2019b, 2021b (2020).
50. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
51. Lakens, D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* **4**, 62627 (2013).
52. Audacity Team. Audacity(R): Free Audio Editor and Recorder [Computer software]. Version 2.4.2 (2022).
53. Boersma, P. & Weenink, D. Praat: Doing phonetics by computer [Computer software]. Version 6.1.16 (2020).
54. Beatty, J. Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology* **19**, 167–172 (1982).
55. Hoeks, B. & Levelt, W. J. M. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behav. Res. Methods Instrum. Comput.* **25**, 16–26 (1993).
56. Willingham, D. B., Nissen, M. J. & Bullemer, P. On the development of procedural knowledge. *J. Exp. Psychol.: Learn. Mem. Cogn.* **15**, 1047–1060 (1989).
57. Bach, D. R. PsPM—Psycho-physiological modelling [Computer software]. Version 5.0.0 (2020).
58. Brisson, J. *et al.* Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behav. Res.* **45**, 1322–1331 (2013).
59. Hayes, T. R. & Petrov, A. A. Mapping and correcting the influence of gaze position on pupil size measurements. *Behav. Res.* **48**, 510–527 (2016).
60. Dan, E. L., Dînşoreanu, M. & Mureşan, R. C. Accuracy of Six Interpolation Methods Applied on Pupil Diameter Data. in *2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)* 1–5 (2020). https://doi.org/10.1109/AQTR49680.2020.9129915.
61. Widmann, A., Schröger, E. & Maess, B. Digital filter design for electrophysiological data—A practical approach. *J. Neurosci. Methods* **250**, 34–46 (2015).
62. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15**, 70–73 (1967).
63. Cohen, M. X. *Analyzing Neural Time Series Data: Theory and Practice* (The MIT Press, 2014). https://doi.org/10.7551/mitpress/9609.001.0001.
64. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Intell. Neurosci.* **2011**, 1:1-1:9 (2011).

Acknowledgements

This work was funded by the Emmy Noether program of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG; Grant No DFG BL 1736/1-1 to HB). CWK is supported by the DFG Emmy Noether Research Group [392443797]. We acknowledge financial support from the Open Access Publication Fund of UKE - Universitätsklinikum Hamburg-Eppendorf. We thank Mark Lustig for support with data collection.

Author contributions

JB planned the study, collected and analysed the data, and wrote the paper; CK was involved in planning the study and reading the final version of the paper; HB planned the study and wrote the paper.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024

9. Literaturverzeichnis / References

1. Abdoun, O., Todorov, D., Poublan-couzardot, A., Tirou, C., Lutz, A., Vernet, M., & Quentin, R. (2026). No evidence of neural feature-specific pre-activation during the prediction of an upcoming stimulus. *Nature Communications*, *17*(1), 4639. https://doi.org/10.1038/s41467-026-73568-1
2. Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology, Computational Neuroscience*, *46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010
3. Akça, M., Vuoskoski, J. K., Laeng, B., & Bishop, L. (2023). Recognition of brief sounds in rapid serial auditory presentation. *PLOS ONE*, *18*(4), e0284396. https://doi.org/10.1371/journal.pone.0284396
4. Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A., & Zenon, A. (2019). Pupil-Linked Arousal Responds to Unconscious Surprisal. *The Journal of Neuroscience*, *39*(27), Article 27. https://doi.org/10.1523/JNEUROSCI.3010-18.2019
5. Alink, A., Becker, J., & Blank, H. (2026). *Sequential prioritization of expected and unexpected information in visual processing* (p. 2026.03.23.713176). bioRxiv. https://doi.org/10.64898/2026.03.23.713176
6. Alink, A., & Blank, H. (2021). Can expectation suppression be explained by reduced attention to predictable stimuli? *NeuroImage*, *231*, 117824. https://doi.org/10.1016/j.neuroimage.2021.117824
7. Amado, C., Hermann, P., Kovács, P., Grotheer, M., Vidnyánszky, Z., & Kovács, G. (2016). The contribution of surprise to the prediction based modulation of fMRI responses. *Neuropsychologia*, *84*, 105–112. https://doi.org/10.1016/j.neuropsychologia.2016.02.003
8. Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063
9. 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*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709
10. Bach, D. R. (2020). *PsPM - Psycho-Physiological Modelling* (Version 5.0.0) [Computer software]. pspm.sourceforge.net
11. Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, *113*(5), Article 5.
12. Basgol, H., Raab, F., Dayan, P., & Franz, V. H. (2026). Modality-general sensitivity of pupil responses to regularity violations. *Cognitive, Affective, & Behavioral Neuroscience*, *26*(4), 1684–1705. https://doi.org/10.3758/s13415-026-01423-3
13. Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038
14. Batterink, L. J. (2017). Rapid Statistical Learning Supporting Word Extraction From Continuous Speech. *Psychological Science*, *28*(7), 921–928. https://doi.org/10.1177/0956797617698226

15. Batterink, L. J., Hsiung, S., Herrera-Chaves, D., & Köhler, S. (2025). Implicit prediction as a consequence of statistical learning. *Cognition*, *258*, 106088. https://doi.org/10.1016/j.cognition.2025.106088
16. Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004
17. Becker, J., Korn, C. W., & Blank, H. (2024). Pupil diameter as an indicator of sound pair familiarity after statistically structured auditory sequence. *Scientific Reports*, *14*(1), 8739. https://doi.org/10.1038/s41598-024-59302-1
18. Becker, J., Schneider, F., & Blank, H. (2025). *Processing of expected and unexpected sensory input during auditory syllable sequences [Preregistration]*. OSF. https://osf.io/xhmb6/overview?view_only=458f931404a24bed84d037be5996ac79
19. Becker, J., Viertler, M., Korn, C. W., & Blank, H. (2024). The pupil dilation response as an indicator of visual cue uncertainty and auditory outcome surprise. *European Journal of Neuroscience*, *59*(10), 2686–2701. https://doi.org/10.1111/ejn.16306
20. Bianco, R., Ptasczynski, L. E., & Omigie, D. (2020). Pupil responses to pitch deviants reflect predictability of melodic sequences. *Brain and Cognition*, *138*, 103621. https://doi.org/10.1016/j.bandc.2019.103621
21. Binda, P., Terzo, C., Turi, M., & Burr, D. C. (2025). Pupillometric signature of implicit learning of statistical regularities. *Current Biology*, *35*(6), 1431-1435.e2. https://doi.org/10.1016/j.cub.2025.02.011
22. Blank, H., & Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLOS Biology*, *14*(11), Article 11. https://doi.org/10.1371/journal.pbio.1002577
23. Blom, T., Feuerriegel, D., Johnson, P., Bode, S., & Hogendoorn, H. (2020). Predictions drive neural representations of visual events ahead of incoming sensory information. *Proceedings of the National Academy of Sciences*, *117*(13), 7510–7515. https://doi.org/10.1073/pnas.1917777117
24. Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.16) [Computer software]. https://www.praat.org
25. Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science*, *16*(6), 451–459. https://doi.org/10.1111/j.0956-7976.2005.01556.x
26. Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), Article 4. https://doi.org/10.1111/j.1469-8986.2008.00654.x
27. Brodbeck, C., Bhattasali, S., Cruz Heredia, A. A., Resnik, P., Simon, J. Z., & Lau, E. (2022). Parallel processing in speech perception with local and global representations of linguistic context. *eLife*, *11*, e72056. https://doi.org/10.7554/eLife.72056
28. Bruns, P., Shibata, K., & Watanabe, T. (2025). Conscious awareness determines whether a second learning session disrupts or facilitates earlier learning. *Current Biology*, *35*(19), 4803-4809.e2. https://doi.org/10.1016/j.cub.2025.08.009
29. Chen, S., Zhu, Y., & Wayland, R. (2017). Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PLOS ONE*, *12*(7), e0180656. https://doi.org/10.1371/journal.pone.0180656

30. 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
31. Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. https://doi.org/10.1016/j.cognition.2008.04.004
32. Clayson, P. E., & Miller, G. A. (2017). Psychometric considerations in the measurement of event-related brain potentials: Guidelines for measurement and reporting. *International Journal of Psychophysiology, Rigor and Replication: Towards Improved Best Practices in Psychophysiological Research*, *111*, 57–67. https://doi.org/10.1016/j.ijpsycho.2016.09.005
33. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9780203771587/statistical-power-analysis-behavioral-sciences-jacob-cohen
34. Colizoli, O., de Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-31985-3
35. Colizoli, O., van Leeuwen, T. M., Rutar, D., & Bekkering, H. (2026). Pupil dilation offers a time-window on prediction error. *eLife*, *14*, RP105287. https://doi.org/10.7554/eLife.105287
36. Conway, C. M., Jenkins, H. E., Milne, A. E., Singh, S., & Wilson, B. (2025). Addressing the theory crisis in statistical learning research. *Npj Science of Learning*, *10*(1), 68. https://doi.org/10.1038/s41539-025-00359-6
37. Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics*, *19*(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493
38. Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. The MIT Press. https://doi.org/10.7551/mitpress/9012.001.0001
39. Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, Auditory Cortex 2006 - The Listening Brain*, *229*(1), 132–147. https://doi.org/10.1016/j.heares.2007.01.014
40. de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences*, *111*(5), E618–E625. https://doi.org/10.1073/pnas.1317557111
41. de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), Article 9. https://doi.org/10.1016/j.tics.2018.06.002
42. de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, *21*(3), 273–290. https://doi.org/10.1037/met0000079
43. Demarchi, G., Hartmann, T., Hauswald, A., Schubert, J., Reisinger, L., Topalidis, P., Benz, K., Schmidt, F., Gehmacher, Q., & Weisz, N. (2026). Reply to: “No evidence of neural feature-specific pre-activation during the prediction of an upcoming stimulus.” *Nature Communications*, *17*(1), 4638. https://doi.org/10.1038/s41467-026-73567-2

44. Demarchi, G., Sanchez, G., & Weisz, N. (2019). Automatic and feature-specific prediction-related neural activity in the human auditory system. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-11440-1
45. den Ouden, C., Kashyap, M., Kikkawa, M., & Feuerriegel, D. (2025). Limited Evidence for Probabilistic Cueing Effects on Grating-Evoked Event-Related Potentials and Orientation Decoding Performance. *Psychophysiology*, *62*(5), e70076. https://doi.org/10.1111/psyp.70076
46. den Ouden, C., Zhou, A., Mepani, V., Kovács, G., Vogels, R., & Feuerriegel, D. (2023). Stimulus expectations do not modulate visual event-related potentials in probabilistic cueing designs. *NeuroImage*, *280*, 120347. https://doi.org/10.1016/j.neuroimage.2023.120347
47. Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299. https://doi.org/10.1016/j.cognition.2006.09.010
48. Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *Quarterly Journal of Experimental Psychology*, *62*(11), 2187–2209. https://doi.org/10.1080/17470210902783646
49. Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, *13*(8), 348–353. https://doi.org/10.1016/j.tics.2009.05.005
50. Feuerriegel, D., Vogels, R., & Kovács, G. (2021). Evaluating the evidence for expectation suppression in the visual system. *Neuroscience & Biobehavioral Reviews*, *126*, 368–381. https://doi.org/10.1016/j.neubiorev.2021.04.002
51. Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392
52. Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458
53. Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied Longitudinal Analysis* (1st ed.). Wiley. https://doi.org/10.1002/9781119513469
54. Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid Serial Auditory Presentation: A New Measure of Statistical Learning in Speech Segmentation. *Experimental Psychology*, *62*(5), 346–351. https://doi.org/10.1027/1618-3169/a000295
55. Friedman, D., Hakerem, G., Sutton, S., & Fleiss, J. L. (1973). Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. *Electroencephalography and Clinical Neurophysiology*, *34*(5), Article 5. https://doi.org/10.1016/0013-4694(73)90065-5
56. Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), Article 1456. https://doi.org/10.1098/rstb.2005.1622
57. Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), Article 2. https://doi.org/10.1038/nrn2787
58. Gabhart, K. M., Xiong, Y. (Sophy), & Bastos, A. M. (2025). Predictive coding: A more cognitive process than we thought? *Trends in Cognitive Sciences*, *29*(7), 627–640. https://doi.org/10.1016/j.tics.2025.01.012
59. Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

60. Garlichs, A., & Blank, H. (2024). Prediction error processing and sharpening of expected information across the face-processing hierarchy. *Nature Communications*, *15*(1), 3407. https://doi.org/10.1038/s41467-024-47749-9
61. Gomes, C. A., Montaldi, D., & Mayes, A. (2015). The pupil as an indicator of unconscious memory: Introducing the pupil priming effect. *Psychophysiology*, *52*(6), Article 6. https://doi.org/https://doi.org/10.1111/psyp.12412
62. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267
63. Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1
64. 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), Article 1. https://doi.org/10.3758/BF03204445
65. Hu, Z., Tran, D. M. D., & Rideaux, R. (2026). Faster But Less Precise: Expectation Enhances Response Speed While Reducing Sensory Fidelity. *The Journal of Neuroscience*, *46*(28), e0154262026. https://doi.org/10.1523/JNEUROSCI.0154-26.2026
66. Iorio, C., Edalati, M., Moghimi, S., & Tillmann, B. (2026). Initial exposure to random input decreases subsequent statistical learning: Behavioral and neural evidence. *Brain and Cognition*, *198*, 106460. https://doi.org/10.1016/j.bandc.2026.106460
67. Isbilen, E. S., & Christiansen, M. H. (2022). Statistical Learning of Language: A Meta-Analysis Into 25 Years of Research. *Cognitive Science*, *46*(9), e13198. https://doi.org/10.1111/cogs.13198
68. Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically Induced Chunking Recall: A Memory-Based Approach to Statistical Learning. *Cognitive Science*, *44*(7), e12848. https://doi.org/10.1111/cogs.12848
69. Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, *38*(4), 723–727. https://doi.org/10.1111/1469-8986.3840723
70. Jiang, J., Summerfield, C., & Egner, T. (2013). Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain. *The Journal of Neuroscience*, *33*(47), 18438–18447. https://doi.org/10.1523/JNEUROSCI.3308-13.2013
71. Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition. *Trends in Cognitive Sciences*, *24*(6), Article 6. https://doi.org/10.1016/j.tics.2020.03.005
72. 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), Article 1. https://doi.org/10.1016/j.neuron.2015.11.028
73. Kafkas, A., & Montaldi, D. (2015). The pupillary response discriminates between subjective and objective familiarity and novelty. *Psychophysiology*, *52*(10), Article 10. https://doi.org/10.1111/psyp.12471

74. Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3933–3936. https://doi.org/10.1109/ICASSP.2008.4518514
75. Kidd, E., Arciuli, J., Christiansen, M. H., Isbilen, E. S., Revius, K., & Smithson, M. (2020). Measuring children's auditory statistical learning via serial recall. *Journal of Experimental Child Psychology*, *200*, 104964. https://doi.org/10.1016/j.jecp.2020.104964
76. King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002
77. Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, *41*(8), Article 8. https://doi.org/10.1111/ejn.12859
78. Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007
79. Kok, P., Bains, L. J., Mourik, T., Norris, D. G., & Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology*, *26*(3), 371–376. (ISI:000369502900029). https://doi.org/10.1016/j.cub.2015.12.038
80. Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), Article 2. https://doi.org/10.1016/j.neuron.2012.04.034
81. 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
82. 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
83. Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical Learning in Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3474–3486. https://doi.org/10.1044/2017_JSLHR-L-16-0439
84. Larsson, J., & Smith, A. T. (2012). fMRI Repetition Suppression: Neuronal Adaptation or Stimulus Expectation? *Cerebral Cortex*, *22*(3), 567–576. https://doi.org/10.1093/cercor/bhr119
85. Lavin, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience*, *7*, 218. https://doi.org/10.3389/fnbeh.2013.00218
86. Liao, H.-I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2016). Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychonomic Bulletin & Review*, *23*(2), 412–425. https://doi.org/10.3758/s13423-015-0898-0
87. Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

88. Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique, second edition*. MIT Press.
89. Luo, M., Cao, R., & Wang, F. H. (2025). Statistical learning of uniform- and mixed-length artificial languages: Different computational mechanisms support different task demands. *eLife*, *13*. https://doi.org/10.7554/eLife.95761.2
90. 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
91. Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.18
92. May, P. J. C., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x
93. McDermott, H. H., de Martino, F., Schwiedrzik, C. M., & Auksztulewicz, R. (2026). Dissociable dynamic effects of expectation during statistical learning. *eLife*, *13*, RP103689. https://doi.org/10.7554/eLife.103689
94. Milne, A. E., Zhao, S., Tampakaki, C., Bury, G., & Chait, M. (2021). Sustained Pupil Responses Are Modulated by Predictability of Auditory Sequences. *The Journal of Neuroscience*, *41*(28), Article 28. https://doi.org/10.1523/JNEUROSCI.2879-20.2021
95. Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477
96. Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), Article 11. https://doi.org/10.1111/j.1469-8986.2011.01226.x
97. Myers, N. E., Rohenkohl, G., Wyart, V., Woolrich, M. W., Nobre, A. C., & Stokes, M. G. (2015). Testing sensory evidence against mnemonic templates. *eLife*, *4*, e09000. https://doi.org/10.7554/eLife.09000
98. 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
99. 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
100. Naber, M., Alvarez, G., & Nakayama, K. (2013). Tracking the allocation of attention using human pupillary oscillations. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00919
101. 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
102. Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue E Linguaggio*, *2*, 203–229.
103. Neuhaus, J. M., Kalbfleisch, J. D., & Hauck, W. W. (1991). A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data. *International Statistical Review / Revue Internationale de Statistique*, *59*(1), 25. https://doi.org/10.2307/1403572

104. Nobre, A. C., & van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, *19*(1), 34–48. https://doi.org/10.1038/nrn.2017.141
105. Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9
106. Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.
107. Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y
108. Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-Driven Computations in Speech Processing. *Science*, *298*(5593), 604–607. https://doi.org/10.1126/science.1072901
109. Perruchet, P. (2019). What Mechanisms Underlie Implicit Statistical Learning? Transitional Probabilities Versus Chunks in Language Learning. *Topics in Cognitive Science*, *11*(3), 520–535. https://doi.org/10.1111/tops.12403
110. Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006
111. 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
112. 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
113. Preuschoff, K., 't Hart, B., & Einhauser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Frontiers in Neuroscience*, *5*. https://www.frontiersin.org/article/10.3389/fnins.2011.00115
114. Qiyuan, J., Richer, F., Wagoner, B. L., & Beatty, J. (1985). The pupil and stimulus probability. *Psychophysiology*, *22*(5), Article 5. https://doi.org/10.1111/j.1469-8986.1985.tb01645.x
115. Quirins, M., Marois, C., Valente, M., Seassau, M., Weiss, N., El Karoui, I., Hochmann, J.-R., & Naccache, L. (2018). Conscious processing of auditory regularities induces a pupil dilation. *Scientific Reports*, *8*(1), 14819. https://doi.org/10.1038/s41598-018-33202-7
116. Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), Article 1. https://doi.org/10.1038/4580
117. Reinhard, G., & Lachnit, H. (2002). The effect of stimulus probability on pupillary response as an indicator of cognitive processing in human learning and categorization. *Biological Psychology*, *60*(2–3), Article 2–3. https://doi.org/10.1016/S0301-0511(02)00031-5
118. Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. *eLife*, *8*, e47869. https://doi.org/10.7554/eLife.47869
119. Richter, D., Ekman, M., & Lange, F. P. de. (2018). Suppressed Sensory Response to Predictable Object Stimuli throughout the Ventral Visual Stream.

Journal of Neuroscience, *38*(34), 7452–7461. https://doi.org/10.1523/JNEUROSCI.3421-17.2018
120. Richter, D., Heilbron, M., & De Lange, F. P. (2022). Dampened sensory representations for expected input across the ventral visual stream. *Oxford Open Neuroscience*, *1*, kvac013. https://doi.org/10.1093/oons/kvac013
121. Rittershofer, K., Wang, Y., Eimer, M., Kok, P., Yon, D., & Press, C. (2026). *Paradoxical influences of prediction are resolved across time*. Neuroscience. https://doi.org/10.64898/2026.03.05.709935
122. Rungratsameetaweemana, N., & Serences, J. T. (2019). Dissociating the impact of attention and expectation on early sensory processing. *Current Opinion in Psychology, Attention & Perception*, *29*, 181–186. https://doi.org/10.1016/j.copsyc.2019.03.014
123. Sabio-Albert, M., Richter, D., Fuentemilla Garriga, L., & Pérez-Bellido, A. (2026). Attention and Explicit Knowledge Drive Predictive Sharpening in Early Visual Cortex. *Journal of Cognitive Neuroscience*, 1–15. https://doi.org/10.1162/JOCN.a.2689
124. Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), Article 4. https://doi.org/10.1006/jmla.1996.0032
125. SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2013). Hearing Silences: Human Auditory Processing Relies on Preactivation of Sound-Specific Brain Activity Patterns. *Journal of Neuroscience*, *33*(20), 8633–8639. https://doi.org/10.1523/JNEUROSCI.5821-12.2013
126. Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335
127. Schneider, F. (n.d.). *MVPy* [Computer software]. Retrieved https://github.com/FabulousFabs/MVPy
128. Schneider, F., & Blank, H. (2026). Sensory sharpening and semantic prediction errors unify competing models of predictive processing in human speech comprehension. *PLOS Biology*, *24*(1), e3003588. https://doi.org/10.1371/journal.pbio.3003588
129. Schroeder, C. E., & Lakatos, P. (2009). The Gamma Oscillation: Master or Slave? *Brain Topography*, *22*(1), 24–26. https://doi.org/10.1007/s10548-009-0080-y
130. Schwiedrzik, C. M., & Sudmann, S. S. (2020). Pupil Diameter Tracks Statistical Structure in the Environment to Increase Visual Sensitivity. *Journal of Neuroscience*, *40*(23), Article 23. https://doi.org/10.1523/JNEUROSCI.0216-20.2020
131. Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, *270*(5234), Article 5234. https://doi.org/10.1126/science.270.5234.303
132. Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061
133. Soares, A. P., Paiva, D., Lema, A., Pereira, D. R., Rodrigues, A. C., & Oliveira, H. M. (2025). Speech Stream Composition Affects Statistical Learning: Behavioral and Neural Evidence. *Brain Sciences*, *15*(2), 198. https://doi.org/10.3390/brainsci15020198

134. Sohoglu, E., Beckers, L., & Davis, M. H. (2024). Convergent neural signatures of speech prediction error are a biological marker for spoken word recognition. *Nature Communications*, *15*(1), 9984. https://doi.org/10.1038/s41467-024-53782-5
135. Sohoglu, E., & Davis, M. H. (2020). Rapid computations of spectrotemporal prediction error support perception of degraded speech. *eLife*, *9*, e58077. https://doi.org/10.7554/eLife.58077
136. Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 186–199. https://doi.org/10.1037/a0033206
137. Spratling, M. W. (2010). Predictive Coding as a Model of Response Properties in Cortical Area V1. *Journal of Neuroscience*, *30*(9), 3531–3543. https://doi.org/10.1523/JNEUROSCI.4911-09.2010
138. Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *216*(1205), 427–459. https://doi.org/10.1098/rspb.1982.0085
139. 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
140. Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191–196. https://doi.org/10.1038/5757
141. Steinhauer, K., & Friederici, A. D. (2001). Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *Journal of Psycholinguistic Research*, *30*(3), 267–295. https://doi.org/10.1023/A:1010443001646
142. Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nat Rev Neurosci*, *15*, 745–756. (25315388). https://doi.org/10.1038/nrn3838
143. Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. (19716752). https://doi.org/10.1016/j.tics.2009.06.003
144. Summerfield, C., & Egner, T. (2016). Feature-Based Attention and Feature-Based Expectation. *Trends in Cognitive Sciences*, *20*(6), 401–404. https://doi.org/10.1016/j.tics.2016.03.008
145. Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. (ISI:000258720000009). https://doi.org/10.1038/Nn.2163
146. Tang, M. F., Smout, C. A., Arabzadeh, E., & Mattingley, J. B. (2018). Prediction error and repetition suppression have distinct effects on neural representations of visual information. *eLife*, *7*, e33123. https://doi.org/10.7554/eLife.33123
147. Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, *8*(1), 10853. https://doi.org/10.1038/s41598-018-28845-5

148. Teufel, C., & Fletcher, P. C. (2020). Forms of prediction in the nervous system. *Nature Reviews Neuroscience*, *21*(4), 231–242. https://doi.org/10.1038/s41583-020-0275-5
149. The MathWorks Inc. (2020). *MATLAB* (Version 9.7.0.1471314 (R2019b)) [Computer software].
150. Tirou, C., Abdoun, O., Vékony, T., Tosatto, L., Brovelli, A., Vernet, M., Németh, D., & Quentin, R. (2026). Learning regularities in noise engages both neural predictive activity and representational changes. *Nature Communications*, *17*(1), 8372. https://doi.org/10.1038/s41467-026-74824-0
151. Todorovic, A., & de Lange, F. P. (2012). Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. *Journal of Neuroscience*, *32*(39), Article 39. https://doi.org/10.1523/JNEUROSCI.2227-12.2012
152. Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding Words and Rules in a Speech Stream: Functional Differences Between Vowels and Consonants. *Psychological Science*, *19*(2), 137–144. https://doi.org/10.1111/j.1467-9280.2008.02059.x
153. Turk-Browne, N. B. (2012). Statistical Learning in Perception. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 3182–3185). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_1707
154. Ufer, C., & Blank, H. (2024). Multivariate analysis of brain activity patterns as a tool to understand predictive processes in speech perception. *Language, Cognition and Neuroscience*, *39*(9), 1117–1133. https://doi.org/10.1080/23273798.2023.2166679
155. 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), Article 1. https://doi.org/10.1038/ncomms14637
156. Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321
157. Wang, F. H., Zevin, J., Trueswell, J., & Mintz, T. H. (2026). Repetition induced rhythm as an alternative account to statistical word segmentation: A model and meta-analysis. *Cognition*, *276*, 106648. https://doi.org/10.1016/j.cognition.2026.106648
158. Wierda, S. M., Rijn, H. van, Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), Article 22. https://doi.org/10.1073/pnas.1201858109
159. Xu, R., & Jiang, X. (2026). Ambiguity in vocal expression perception: Evidence from voice morphing and its electrophysiological correlates. *NeuroImage*, *336*, 121987. https://doi.org/10.1016/j.neuroimage.2026.121987
160. 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
161. Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, *42*(1), 121. https://doi.org/10.2307/2531248
162. Zénon, A. (2019). Eye pupil signals information gain. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1911), Article 1911. https://doi.org/10.1098/rspb.2019.1593

10. Danksagung

Hiermit möchte ich mich bei all den Menschen bedanken, die mich in den vergangenen Jahren fachlich und persönlich begleitet und damit zum Gelingen dieser Arbeit beigetragen haben.

Zunächst gilt der größte Dank meiner Doktormutter, Prof. Dr. Helen Blank. Liebe Helen, vor sechs Jahren habe ich als studentische Hilfskraft in deiner frisch gegründeten Arbeitsgruppe angefangen, einen ersten Einblick in die Neurowissenschaften und in deine Arbeitsweise erhalten – und bin dann, wie man sieht, geblieben: erst als Hilfskraft und Praktikantin, dann als Masterandin, und nun schließe ich meine Promotion bei dir ab, während du inzwischen zur Professorin berufen wurdest. Ich kann daher ohne jede Untertreibung sagen, dass ich dir den größten Teil meiner praktischen wissenschaftlichen Ausbildung verdanke. Danke für diese hervorragende Ausbildung und für die gemeinsame Arbeit! Danke außerdem dafür, dass du jederzeit erreichbar und ansprechbar warst, wenn ich Rat oder Hilfe gebraucht habe, und dass dabei immer auch Raum für menschliches Verständnis füreinander blieb – das ist ein Glück, das längst nicht alle Promovierenden haben. Dein Einsatz, deine Zielstrebigkeit und deine Hingabe an die Wissenschaft sind mir über all die Jahre eine große Inspiration gewesen und werden mir auch zukünftig ein Vorbild sein.

Begleitet wurde diese Arbeit außerdem von den Mitgliedern meines Thesis-Komitees. Prof. Dr. Christoph Korn danke ich für seinen konstruktiven Rat auf Augenhöhe, seine verlässliche Hilfsbereitschaft und die Zusammenarbeit als Co-Autor – und dafür, dass er die gemeinsame wissenschaftliche Arbeit mit einer ausgesprochen wertschätzenden Art verbindet. Prof. Dr. Tobias Donner verdanke ich eine pragmatische und flexible Herangehensweise, durch die sich Fragen und Termine stets unkompliziert klären ließen. PD Dr. Patrick Bruns als Zweitgutachter und PD Dr. Regine Klinger danke ich für ihr Interesse an dieser Arbeit und ihre Bereitschaft, Teil meines Prüfungskomitees zu sein.

Ein besonders herzlicher Dank gilt meinen Kolleg:innen am ISN, die über die Jahre eine verlässliche Quelle der Inspiration und der schönen Momente waren. Annika Garlichs, als das soziale Herz unserer Gruppe, lebt vor, wie sich fachliche Stärke mit echter Herzlichkeit verbinden lässt. Carina Ufer verbindet eine außergewöhnliche Auffassungsgabe mit einer beeindruckend authentischen Art – in fachlicher Begeisterung ebenso wie im täglichen Miteinander. Fabian Schneider, unser Statistik- und Programmier-Pro, vereint technische Souveränität mit großer Hilfsbereitschaft – seine Python-Toolbox hat die Analysen dieser Dissertation entscheidend vereinfacht, und seine stets entspannte, offene Art macht die Zusammenarbeit umso schöner. Philipp Schumann und Mai-Carmen Requena-Komuro, die in der zweiten Hälfte meiner Promotion zur Gruppe gestoßen sind, danke ich für den fachlichen Austausch,

die gemeinsamen Group-Lunches, das offene Ohr in schwierigen Phasen – und besonders für die Ermutigung, für mich selbst einzustehen.
Julia Ruge war die zuverlässigste Büronachbarin, die man sich wünschen kann: bei ihr konnte man sich beschweren, die geteilte Leidenschaft für Maiswaffeln schamlos ausleben, und ihre Arbeitsmoral war mir immer ein Vorbild. Sepideh Khoneiveh danke ich für eine Freundschaft, die über das Büro hinausgeht und meinen menschlichen Horizont erweitert hat. Nicht zuletzt waren die Jahre am Institut dank Carina Jaap, Jan Ostrowski, Nina Lüthi, Maria Bruntsch, Ivana Tanasic und Oleg Solopchuk eine unvergesslich schöne Zeit – mit unzähligen Spieleabenden, gemeinsamen Brunchs und vielen weiteren Unternehmungen
Ohne unsere studentischen Hilfskräfte wäre vieles langsamer gegangen: Mark Lustig und Antonia Gelpke haben mich bei der Datenerhebung tatkräftig unterstützt, Eylül Karabiber als Praktikantin bei der Literaturrecherche.

In der Schlussphase dieser Promotion haben sich meine Freundinnen und Freunde zu Recht vernachlässigt gefühlt und sind mir trotzdem mit Verständnis begegnet. Danke für eure Nachsicht!

Calina Augustin danke ich für ihr Verständnis, ihre Herzlichkeit und einen Rückhalt, auf den ich mich bedingungslos verlassen kann. Sarah Al Nussairi ist unvergleichlich in ihrer entspannten, optimistischen Art, die mich immer wieder ansteckt. Zu dritt mit euch ist die Zeit einfach am schönsten – danke für eure Bestärkung in dieser Zeit und für diese wunderbare Freundschaft!

Maximilian Ochs danke ich insbesondere für seine Ermutigung und seinen Zuspruch in den Phasen, in denen mir das Vertrauen in die eigenen Fähigkeiten abhandengekommen war. Seine Worte haben mir gerade in schwierigen Momenten immer wieder geholfen, an mich selbst zu glauben und einen Sinn in dem zu sehen, was ich tue.

Gerrit von Wissel hat mich mit Geduld, Verständnis und Fürsorge durch die letzte Phase dieser Promotion begleitet. Seine Bereitschaft, sich auf die besonderen Anforderungen dieser Zeit einzustellen, und sein Rückhalt waren mir eine große Unterstützung auf dem Weg zum Abschluss.

Der letzte und persönlichste Dank gilt meiner Familie: meinen Eltern, meinen Brüdern und Hanni. Ich weiß, dass ich mich jederzeit auf euch verlassen kann und dass ihr hinter mir steht. Dieses Wissen trägt durch alle Zeiten und ist schlicht unbezahlbar.

Vielen Dank!

11. Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe, insbesondere ohne entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten, verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Das gilt insbesondere auch für alle Informationen aus Internetquellen.
Soweit beim Verfassen der Dissertation KI-basierte Tools („Chatbots“) verwendet wurden, versichere ich ausdrücklich, den daraus generierten Anteil deutlich kenntlich gemacht zu haben. Die „Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG“ aus September 2023 wurde dabei beachtet.
Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.
Ich erkläre mich damit einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Datum Unterschrift