

Language experience shapes predictive coding of rhythmic sound sequences

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint version 1

December 22, 2023 (this version)

Sent for peer review

August 21, 2023

Posted to preprint server

August 3, 2023

Piermatteo Morucci , Sanjeev Nara, Mikel Lizarazu, Clara Martin, Nicola Molinaro

Department of Fundamental Neurosciences, University of Geneva, Switzerland • Basque Center on Cognition Brain and Language (BCBL), 20009, Donostia-San Sebastian, Spain • Mathematical Institute, Department of Mathematics and Computer Science, Physics, Geography, Justus-Liebig-Universität Gießen, Germany • Ikerbasque, Basque Foundation for Science, 48013, Bilbao, Spain

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Perceptual systems heavily rely on prior knowledge and predictions to make sense of the environment. Predictions can originate from multiple sources of information, including contextual short-term priors, based on isolated temporal situations, and contextindependent long-term priors, arising from extended exposure to statistical regularities. While the effects of short-term predictions on auditory perception have been welldocumented, how long-term predictions shape early auditory processing is poorly understood. To address this, we recorded magnetoencephalography data from native speakers of two languages with different word orders (Spanish: functor-initial versus Basque: functor-final) listening to simple sequences of binary sounds alternating in duration with occasional omissions. We hypothesized that, together with contextual transition probabilities, the auditory system uses the characteristic prosodic cues (duration) associated with the native language's word order as an internal model to generate long-term predictions about incoming non-linguistic sounds. Consistent with our hypothesis, we found that the amplitude of the mismatch negativity elicited by sound omissions varied orthogonally depending on the speaker's linguistic background and was most pronounced in the left auditory cortex. Importantly, listening to binary sounds alternating in pitch instead of duration did not yield group differences, confirming that the above results were driven by the hypothesized long-term "duration" prior. These findings show that experience with a given language can shape a fundamental aspect of human perception - the neural processing of rhythmic sounds - and provides direct evidence for a long-term predictive coding system in the auditory cortex that uses auditory schemes learned over a lifetime to process incoming sound sequences.

eLife assessment

This study presents **useful** observations about how the human brain uses long-term priors (acquired during our lifetime of listening) to make predictions about expected sounds - an open question in the field of predictive processing. However, the evidence as currently presented is **incomplete**. Both the theoretical background and analysis approach should be strengthened.

Introduction

According to predictive coding theories of perception, sensory processes and perceptual decisions are described as a process of inference, which is strongly shaped by prior knowledge and predictions (Clark 2013 [↗](#); Friston 2005 [↗](#); Rao & Ballard 1999 [↗](#)). Predictions can be derived from different sources of information, forming a hierarchical predictive system (De Lange et al., 2018 [↗](#)). Each level of the predictive hierarchy houses an internal model encoding prior information about the structure of the external environment. When a prediction is violated, the prediction error is computed and used to adjust the corresponding prior and internal model. This results in a constantly evolving system that generates and refines predictions based on incoming sensory input and prior experience.

In the auditory domain, great progress in the understanding of the predictive capabilities of the auditory system has been made using the oddball design and its variations (see Heilbron and Chait, 2018 [↗](#), for a review). In these designs, participants are usually presented with sequences of tones encoding a certain rule that is then violated by a ‘deviant’ event. Such deviants elicit a sharp evoked response in the EEG signal which has been defined as “mismatch negativity” (MMN). The MMN peaks at about 0.100–0.250 seconds from stimulus onset and exhibits enhanced intensity over secondary temporal, central, and frontal areas of topographic scalp maps (Sams et al., 1985 [↗](#), Garrido et al., 2009 [↗](#)). Within the predictive coding framework, the MMN is putatively considered an index of cortical prediction error.

Functionally, the mechanism underlying the MMN operates over both conscious and preconscious memory representations. MMN responses to auditory violations are observable when the participant is not paying attention to the auditory task and have been reported even in states of sleep (Sallinen et al., 1994 [↗](#), Sculthorpe et al., 2009 [↗](#), Strauss et al. 2015 [↗](#)) and coma (Fischer et al., 2000 [↗](#)). Given its automatic nature, the anticipatory mechanism underlying the MMN has been suggested to reflect a form of ‘*primitive intelligence*’ in the auditory cortex (Näätänen et al., 2001 [↗](#)). In the present study, we show that life-long experience with a spoken language can shape this automatic anticipatory mechanism.

Experimental studies using the oddball design and its derivations have been important to unveil the sensitivity of the auditory predictive system to local statistical regularities and transition probabilities (Heilbron and Chait, 2018 [↗](#)). However, these studies have primarily examined so called contextual (or short-term) predictive signals. These predictions are usually based on rules acquired in the context of an experimental task – that is, rules linked to short-term memory – and have a short-lived impact on sensory processing. Yet, one core assumption of current predictive coding models is that the brain also deploys predictions based on long-term memory representations (Serès & Seitz, 2013 [↗](#); Yon & De Lange, 2018; Teufel & Fletcher, 2020 [↗](#)). Such long-term predictions may emerge via learning of regularities and co-occurring patterns that are relatively stable throughout the lifespan of an organism. Because arising over long timescales, these experiential priors become encoded into the tuning properties of sensory cortices, forming a computational constraint on bottom-up sensory processing (Teufel & Fletcher, 2020 [↗](#)).

Long-term priors may have long-lasting effects on perception. One example from the visual domain is the systematic bias in humans toward the perception of cardinal orientations (Girshick et al., 2011 [↗](#)). This bias has been linked to the presence of a long-term prior that mirrors the statistics of the visual environment, that is, the preponderance of cardinal orientations in visual input (Girshick et al., 2011 [↗](#)). Monkey studies on visual processing have shown that the visual system employs long-term priors to generate long-term predictions of incoming data (Meyer & Olson, 2011 [↗](#)). Yet, whether similar predictive coding schemes subserve cortical computation in the human auditory system remains unsettled.

Here, we take a cross-linguistic approach to test whether the auditory system generates long-term predictions based on life-long exposure to auditory regularities, using rules that extend beyond those acquired in the recent past. Currently, one critical behavioral example of the effect of long-term experience on auditory perception is the influence of language on rhythmic grouping (Iversen et al., 2008 [DOI](#), Molnar et al., 2016 [DOI](#)): Sequences of two tones alternating in duration are usually perceived by speakers of functor-initial languages (e.g., Spanish, English) as repetition of short-long groups separated by a pause, while speakers of functor-final languages (e.g., Basque, Japanese) report a bias for the opposite long-short grouping pattern. This perceptual effect has been linked to the co-occurrence statistics underlying the word order properties of these languages. Specifically, the effect has been proposed to depend on the quasi-periodic alternation of short and long auditory events in the speech signal, which reflect the linearization of function words (e.g., articles, prepositions) and content words (e.g., nouns, adjectives, verbs). In functor-initial languages, like English or Spanish, short events (i.e., function words; e.g., *un*, *a*) normally combine with long ones (i.e., content words; e.g., *ordenador*, computer) to form “short-long” phrasal chunks (Fig. 1, A [DOI](#)). By contrast, in functor-final languages like Japanese and Basque, short events (i.e., function words; e.g., *bat*, *a*) normally follow long ones (i.e., content words; e.g., *ordenagailu*, computer), resulting in “long-short” phrasal units (Fig. 1, B [DOI](#)). Regular exposure to such language-specific phrasal structures has been proposed to underlie the automatic grouping biases of non-linguistic sounds (Iversen et al., 2008 [DOI](#), Molnar et al., 2016 [DOI](#)), suggesting the presence of an auditory “duration prior” that mirrors the word-order and prosodic properties of a given language.

We hypothesize that the auditory system uses the proposed “duration prior” as an internal model to generate long-term predictions about incoming sound sequences. In predictive coding terms, our hypothesis posits that the human auditory system upweights neural activity towards the onset of certain high-level events, based on the statistics of a given language.

To test this hypothesis, two groups of Basque ($n = 20$) and Spanish ($n = 20$) dominant participants were presented with 30 second rhythmic sequences of two tones alternating in duration at fixed intervals (Fig. 1, C [DOI](#)), while magnetoencephalography (MEG) was monitoring their cortical activity. To measure prediction error, random omissions of long and short tones were introduced in each sequence. Omission responses allow to examine the presence of putative error signals decoupled from bottom-up sensory input, offering a critical test for predictive coding (Walsh et al 2020 [DOI](#), Heilbron and Chait, 2018 [DOI](#)).

If, in line with our hypothesis, the human auditory system uses long-term linguistic priors as an internal model to predict incoming sounds, the following predictions ensue. The omission of a long tone should represent the violation of two predictions in the Basque, but not in the Spanish group: a short-term prediction based on the statistics of the previous stimuli (i.e., a prediction about a new tone), and a long-term prediction based on the statistics of the Basque’s phrasal structure (i.e., a prediction about a new phrasal chunk). Consequently, such an omission response should lead to a larger prediction error in the Basque compared to the Spanish group (Fig. 1, E [DOI](#)). An orthogonally opposite pattern is expected when the deviant event is reflected in the omission of a short tone (Fig. 1, E [DOI](#)). We tested these predictions against a control condition having the same alternation design as the experimental condition, but with the two tones alternating in pitch instead of duration (Fig. 1, D [DOI](#)). Here, no difference between groups is expected, as both groups should rely on short-term, but not long-term priors (Fig. 1, F [DOI](#)). Finally, we performed reconstruction of cortical sources to identify the regions supporting long-term auditory priors.

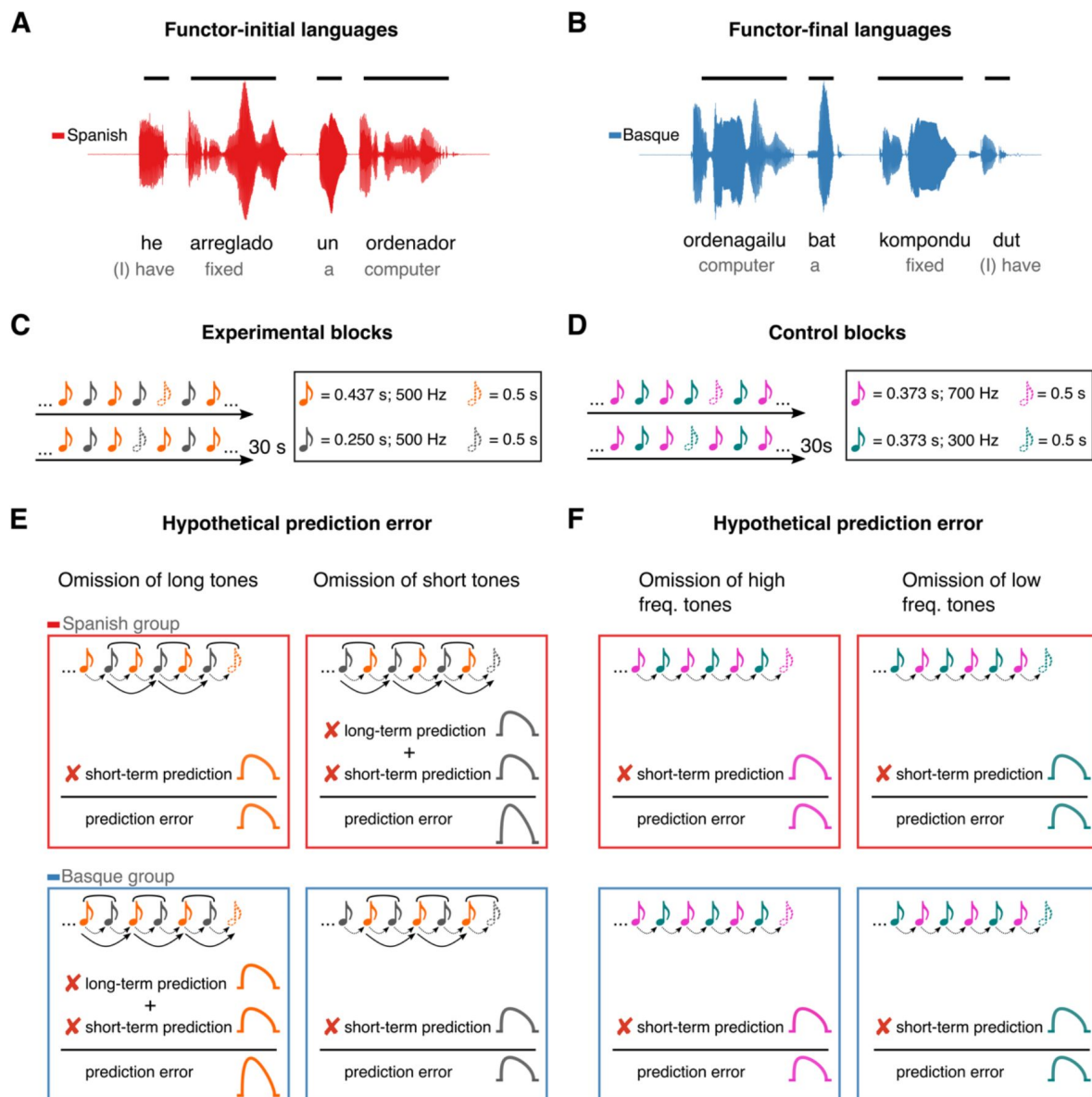


Figure 1.

Experimental design and rationale of the study.

Panels A and B illustrate the contrast between funtor-initial and funtor-final word order in Spanish and Basque, as well as its consequences on their prosodic structure. Panels C and D show the design of the experimental and control conditions, respectively. The structure of the design is the same in both conditions (*ababab*), with 30 s sequences of two tones alternating at fixed intervals and occasional omissions. In the experimental condition, tones alternate in duration but not frequency, whereas in the control condition tones alternate in frequency but not duration. Panels E and F depict the hypothesized error responses associated with the different types of omissions for experimental and control conditions, respectively. Round brackets above the tones reflect the grouping bias of the two languages, based on their word order constraints. Dotted lines reflect short-term predictions based on the transition probabilities of the previous stimuli. Solid lines reflect long-term predictions based on the phrasal chunking scheme of the two languages.

Results

We first examined the responses evoked by the omissions of tones. MEG responses time-locked to the onset of the tones and omissions from the Basque and Spanish dominant groups were pulled together and compared via cluster-based permutation test analysis (Maris & Oostenveld, 2007). Cluster analysis was performed over several time-points to identify spatiotemporal clusters of neighboring sensors where the two conditions differ (*Methods*). This analysis revealed an early effect of omission responses arising around 0.100 s from deviance onset ($P < 0.0001$), including several channels over the entire scalp (**Fig. 2 A, B**). The latency and topographical distribution of the effect resemble the one elicited by a classical mismatch response, with strong activations over left and right temporal regions (**Fig. 2 A, B**). This finding aligned with previous reports showing that the omission of an expected tone in a regular sequence of sounds generates larger event-related fields (ERF) than an actual tone (e.g., Yabe et al., 1997; Raji et al., 1997).

Our main question was on the presence of long-term predictions induced by the linguistic background of the participants during the processing of simple binary auditory sequences. To assess this, we tested for the presence of an interaction effect between the linguistic background of the participants (Basque, Spanish) and the type of tone omitted (long, short) in modulating the amplitude of the MMN. Omission-MMN responses in this analysis were calculated by subtracting the ERF elicited by a given tone from its omission (Garrido et al., 2009). A cluster-based permutation test was used to test the interaction effect between omission type (long vs short) and the linguistic background of the participants (Basque vs Spanish). As the cluster-based permutation test is designed to compare two conditions at a time, we tested for an interaction effect by subtracting the MMN elicited by the omission of a long tone from the MMN elicited by the omission of a short tone for each participant, and then compared the resulting differences between groups.

For this and the following contrasts, we selected a predefined time window of interest between 0.100 and 0.250 s, which covers the typical latency of the MMN (Näätänen et al., 2007; Garrido et al., 2009). Cluster analysis revealed a significant interaction ($P = 0.03$), which was particularly pronounced over left frontotemporal channels (see **figure 2, C**). To unpack the interaction, we averaged data samples for each participant and condition over the channels belonging to the significant cluster and time points of interest. This resulted in two ERFs for each participant, one for each type of omission-MMN (long, short). We then compared ERFs for each omission type between the two groups using a one-sided independent sample t-test, testing the hypothesis that participants deploy long-term expectations about the onset of abstract language-like grouping units (**fig. 1, E**). Specifically, we compared (i) the omission-MMN responses generated by the omission of long tones in the Basque vs the Spanish group, and (ii) the MMN responses generated by the omission of short tones in the Basque vs the Spanish group. Consistent with the hypothesis, we found that omissions of long tones generated a larger MMN response in the Basque compared Spanish group ($t(38) = 2.22$; $P = 0.03$ FDR-corrected; $d = 0.70$), while the omission of short tones generated a larger omission-MMN in the Spanish compared to Basque group ($t(38) = -2$; $P = 0.03$ FDR-corrected; $d = 0.63$) (**fig. 2, D, E**). Notice that, given the structure of our design, a *between-group* comparison (e.g., comparing the ERF between the Basque and Spanish groups) is more suited to test our hypothesis than a *within-group* comparison (e.g., comparing the ERF evoked by the omission of long vs short tones within the Basque group), as the pre-stimulus baseline activity is virtually identical across conditions only in the *between-group* contrast. To further assess that the interaction was driven by the hypothesized long-term “duration prior”, the same analysis pipeline was applied to the data from the control condition. Here, no significant omission-type x language background interaction was detected (no cluster with $P < 0.05$). To further check that no interaction was present in the control study, we averaged the data samples over the channels and time points in which we detected a significant interaction in the test condition and ran an independent sample t-test by comparing MMN responses elicited by the omission of high and low-frequency tones in both groups (*Methods*). Even within this subset of channels, no between-group

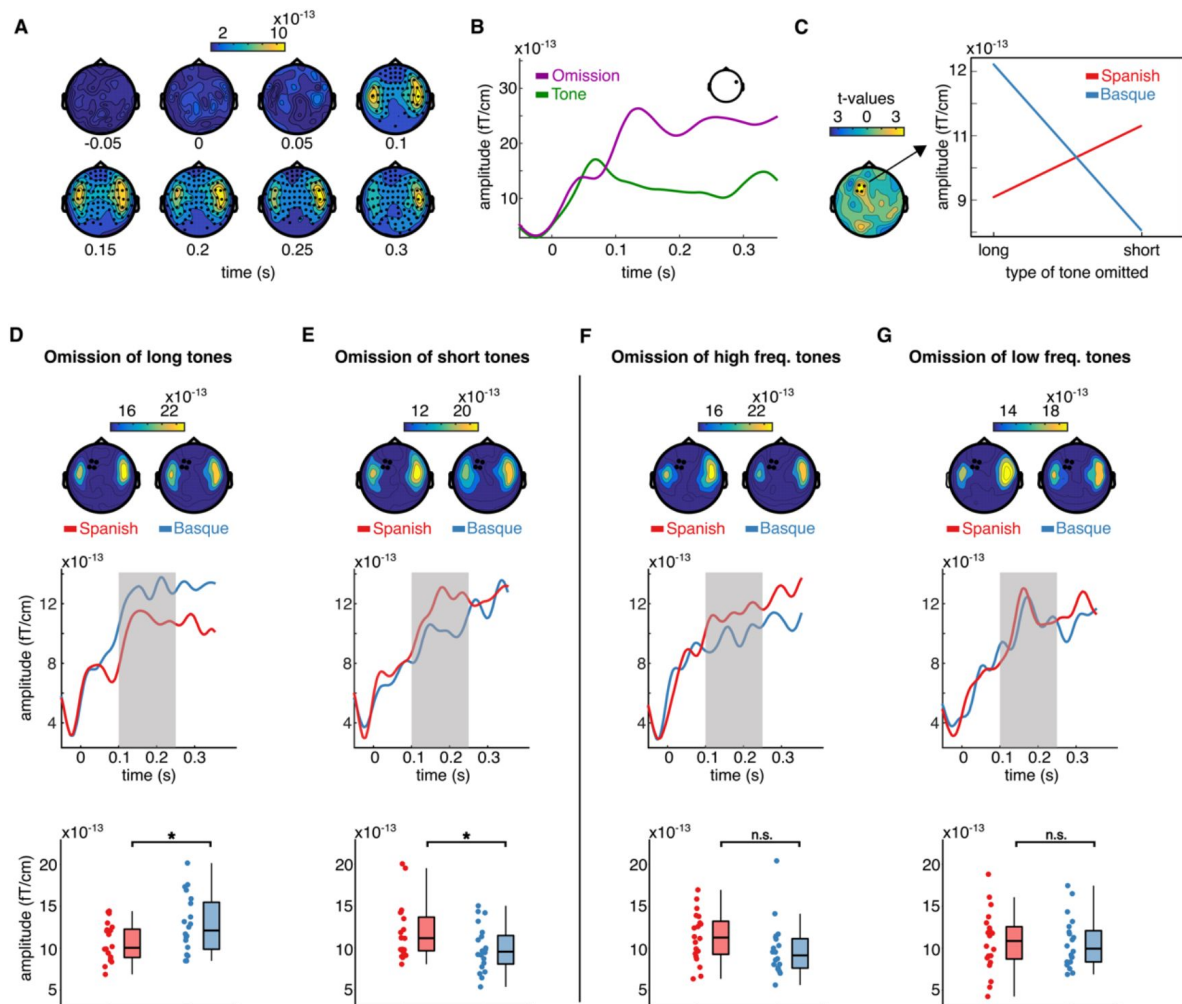


Figure 2.

Sensor-level topography and time course of neural responses to omitted sounds across groups and conditions.

Panel A shows the temporal unfolding and topographical distribution of the overall effect of omission (omissions minus tones). Channels belonging to the significant cluster are highlighted. Panel B shows the ERF generated by omissions and tones in a representative channel. Panel C (left) shows the topography of the t distribution of the interaction effect between the language background of the participants (Spanish, Basque) and the type of omission MMN (short, long). Channels belonging to the significant interaction cluster are highlighted. The interaction effect was present only in the experimental condition. Panel C (right) shows the averaged MEG activity over the 0.100 – 0.250 s time window and channels belonging to the significant cluster for each group and condition separately. Panels D, E, F, G show the effect of language experience in modulating the amplitude of the omission MMN associated with each experimental and control contrast. Topographies (top) show the scalp distribution of the averaged activity over the 0.100 – 0.250 s time window. Channels belonging to the significant interaction cluster are highlighted. ERFs (middle) show the temporal unfolding of brain activity averaged over the channels belonging to the significant interaction cluster for each contrast and group. The shaded area indicates the time window of interest for the statistical analysis. Box-plots (down) show the mean MEG activity for each participant over the 0.100 – 0.250 s time window and the channels belonging to the significant interaction cluster. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data). The notches indicate the 95% confidence interval around the median. Dots reflect individual subjects. In D–G, asterisks indicate statistical significance for each contrast using a one-sided, independent sample t -test with FDR correction for multiple comparisons [statistical significance: ** $p < 0.01$, * $p < 0.05$, and *ns* $p > 0.05$, respectively].

difference was detected between MMN responses evoked by omissions of high ($t(38) = -1.6$; $P = 0.12$ FDR-corrected; $d = -0.51$), and low-frequency tones ($t(38) = -1.1$; $P = 0.55$ FDR-corrected; $d = -0.04$) (**fig. 2, F, G** [↗](#)).

A linearly constrained minimum variance (LCMV) beamformer approach (Van Veen et al., 1997 [↗](#)) was used to reconstruct the cortical sources of the MEG signal. We first focused on the source activity underlying the effect of omission (**Fig. 2 A, B** [↗](#)). Source activity was calculated for the ERF averaged in the 0.100 – 0.250 s interval for both tones and omissions. In order to isolate regions underlying the effect of omission, whole-brain maps for the ratio of the source activity associated with omission responses and tones were created (**Fig. 3 A** [↗](#)). A large network of regions showed a stronger response to omissions compared to auditory tones, including the bilateral inferior frontal gyri, premotor cortices, angular gyri, as well as the right superior temporal gyrus (**Fig. 3 A** [↗](#)).

We then examined the cortical origin of the interaction that emerged at the sensor level (**Fig. 2 C** [↗](#)). In line with the MMN analysis at the sensor level, we focused on the cortical activity of the difference between omissions and tone for each condition (short, long) and group (Basque, Spanish). This analysis was aimed at identifying the cortical origin of the hypothesized long-term predictions. We hypothesized that such long-term priors might be linked to long-term experience with the rhythmic properties of the two languages. As such, they were expected to arise around early auditory cortices. An alternative hypothesis is that these priors are linked to the abstract syntactic structure of the two languages. Under this account, long-term predictions would be generated via long-range feedback from regions associated with syntactic processing, such as the left inferior frontal gyrus. To disentangle these possibilities, we focused on three regions of interest, which included the inferior frontal gyrus (*pars triangularis* and *pars opercularis*) and the early left and right auditory cortices (*Heschl's gyrus* and *superior temporal gyrus*). As shown in **Fig. 3 B, C, D** [↗](#), the only region that showed an interaction trend consistent with the sensor-level data is the left auditory cortex (**Fig. 3 B** [↗](#)), in which omission of long events generated a stronger response in the Basque compared to the Spanish group, while the opposite pattern was observed when a short tone was omitted. Besides documenting an interaction, which was the analysis of interest of our study, we also searched for a main effect of omission type. A cluster-based permutation test was used to compare the MMN responses elicited by omissions of long tones and omissions of short tones averaged across the two groups. The results showed that long-tone omissions generated a larger omission MMN response than short-tone omissions ($P = 0.03$), with the cluster including several frontal channels (**fig. 4 A** [↗](#)). This effect was consistent in the Basque ($P = 0.003$), but not in the Spanish group (no clusters with $P < 0.05$). No main effect of language background was detected (no clusters with $P < 0.05$) (**fig. 4 B** [↗](#)). In the control condition, neither a main effect of omission type nor an effect of language background was detected (no clusters with $P < 0.05$).

Discussion

By comparing MEG data from native speakers of functor-initial (i.e., Spanish) and functor-final languages (i.e., Basque) listening to simple binary sequences of tones with occasional violations, we show that experience with a given language can shape a very simple aspect of human perception, such as the neural processing of a binary rhythmic sound. This finding suggests that the human auditory system uses structural patterns of their native language to generate predictive models of non-linguistic sound sequences. This result highlights the presence of two distinct systems for the generation of auditory predictive models, one relying on the transition probabilities governing the recent past, and another relying on natural sound statistics learned over a lifetime.

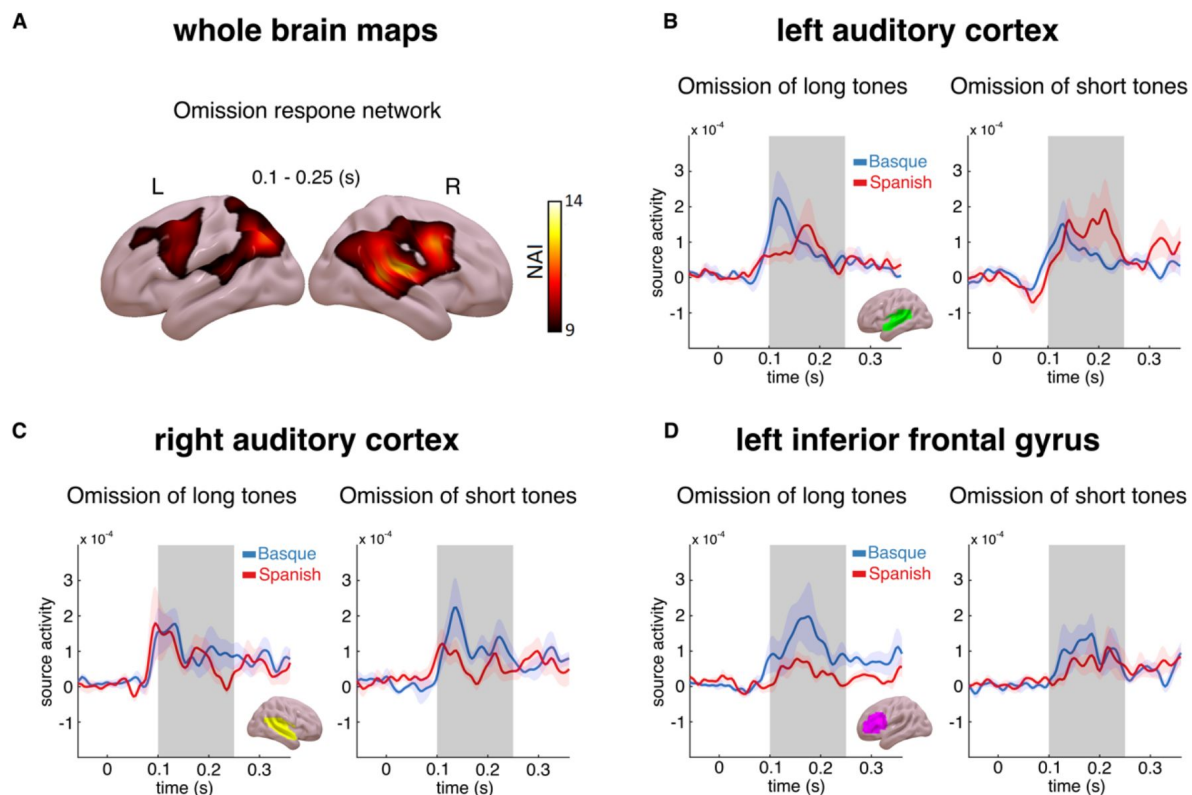


Figure 3.

Source activity underlying the omission network and long-term predictions.

Panel A shows brain maps representing source activity underlying the omission response network over the 0.100 – 0.250 s time window. Panels B, C and D show source activity associated with the omission MMN responses in the left auditory cortex (left *Hesch's gyrus* and left *superior temporal gyrus*), right auditory cortex (right *Hesch's gyrus* and right *superior temporal gyrus*), and the left inferior frontal gyrus (*pars triangularis* and *pars opercularis*), respectively. The shaded grey area shows the 0.100 – 0.250 s time window, which was selected for the statistical analysis of the sensor-level data.

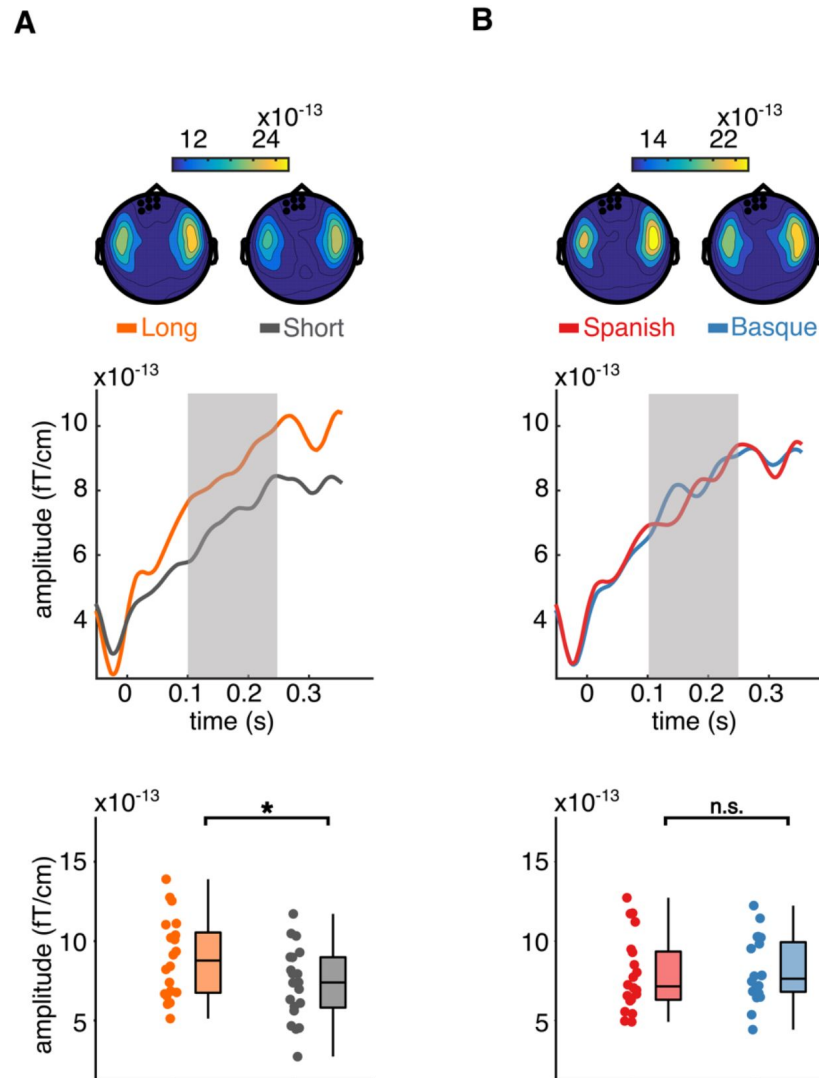


Figure 4.

Main effects of omission type and language background.

Panel A shows the ERFs and topographies reflecting the main effect of omission type, with omissions of long auditory events generating larger omission MMN than short events between groups. Panel B shows the ERFs and topographies reflecting the (lack of) main effect of language background, with overall no group differences in the amplitude of the omission MMN. Because no main effect of language background was detected, Panel B uses the same channels of Panel A as representative channels for plotting the ERF. The remaining conventions for the plot are the same as in [Figure 2](#).

We first looked at the responses generated by sound omissions. In line with previous reports, we found that omissions of expected tones in an auditory sequence generate a sharp response in the event-related field. Such omission responses have been suggested to reflect pure error signals decoupled from bottom-up sensory input (Hughes et al., 2001 [↗](#); Wacogne et al., 2011). On the other hand, other studies have proposed that omission responses could reflect pure predictions (Bendixen et al., 2009 [↗](#), San Miguel et al., 2013 [↗](#)). While the exact nature of such responses is currently debated and likely dependent on factors such as task and relevance, the latency and topography of the omission response in our data resemble those evoked by a classical mismatch response (Fig. 2 A, B [↗](#)). Analysis of cortical sources also supports this interpretation. Indeed, source activity associated with omissions leads to stronger responses compared to tones over a distributed network of regions, including the bilateral inferior frontal gyri, premotor areas, angular gyri, and right superior temporal gyrus. Overall, these findings are consistent with the idea that omission responses reflect, at least in part, prediction error signals.

Importantly, we showed that when an unexpected omission disrupts a binary sequence of sounds, the amplitude of the omission MMN varies orthogonally depending on the speaker's linguistic background. Omissions of long auditory events generate a larger omission MMN in the Basque compared to the Spanish group, while omissions of short sounds lead to a larger omission MMN responses in the Spanish compared to the Basque group. We hypothesized that this effect is linked to a long-term “duration prior” originating from the syntactic function-content word order of language, and specifically, from its acoustic consequences on the prosodic structure. Importantly, no difference between groups was detected in a control task in which tones alternate in frequency instead of duration, suggesting that the reported effect was driven by the hypothesized long-term linguistic priors instead of uncontrolled group differences.

It is important to note that both Spanish and Basque speakers are part of the same cultural community in Northern Spain. These languages share almost the same phonology and orthography. However, Basque is a non-Indo-European language (an isolated language) with no typological relationship with Spanish. It is thus very unlikely that the current findings are driven by cultural factors that are not language-specific (e.g., exposure to different musical traditions or educational and writing systems).

How would such long-term priors arise? One possible interpretation is that long-term statistical learning of the duration prosodic pattern of native language shapes the tuning properties of early auditory regions, affecting predictive coding at early stages. Such language-driven tuning is arguably important for reducing the prediction error during the segmentation of speech material into phrasal units, as it allows the auditory system to generate a functional coding scheme, or auditory template, against which the incoming speech input can be parsed. Such an auditory template is likely to be recycled by the auditory system to build top-down predictive models of non-linguistic auditory sequences.

The idea that the auditory system implements long-term predictions based on the prosodic structure of the native language could explain the previously reported behavioral influence of language experience on rhythmic grouping (Iversen et al., 2008 [↗](#), Molnar et al., 2016 [↗](#)): when listening to sequences of two tones alternating in duration, like those used in the present study, speakers of functor-initial languages report to perceive the rhythmic sequences as a repetition of “short-long” units, while speakers of functor-final languages have the opposite “long-short” grouping bias (Iversen et al., 2008 [↗](#), Molnar et al., 2016 [↗](#)). Despite lacking a direct behavioral assessment (but see Molnar et al., 2016 [↗](#), for related behavioral evidence), our results indicate that this perceptual grouping effect can be explained within a predictive coding framework that incorporates long-term prior knowledge into perceptual decisions. Under such an account, the auditory system internalizes the statistics underlying the prosodic structure of language and uses

this knowledge to make long-term predictions of incoming sound sequences. Such long-term predictions would bias auditory processing at early, rather than later decision-making stages, affecting how rhythmic sounds are experienced.

Our work capitalized on a specific aspect of natural sound acoustic – the duration pattern in Basque and Spanish prosody – as a testbed to assess the presence of long-term priors in the auditory system. Despite our work being restricted to this specific feature, it is likely that the auditory system forms several other types of long-term priors using the spectrotemporal features that dominate the auditory environment. Support for this claim comes from (i) studies showing that the human auditory system uses the statistics underlying the acoustic structure of speech and music to form perceptual grouping decisions (Młynarski & McDermott, 2019 [↗](#)); and (ii) behavioral experiments reporting off-line effects of language experience on auditory perception based on different acoustic features (Liu et al., 2023). For instance, native speakers of languages in which pitch carries phonemically meaningful information (i.e., tone languages, e.g., Mandarin Chinese) benefit from a behavioral advantage in non-linguistic pitch discrimination tasks as compared to speakers of non-tone languages like English (Bidelman et al., 2013 [↗](#)). Similarly, speakers of languages that use duration to differentiate between phonemes (e.g., Finnish, Japanese) manifest an enhanced ability to discriminate the duration of non-linguistic sounds (Tervaniemi et al., 2006 [↗](#)). Our results, in conjunction with these studies, suggest that the auditory system forms long-term priors and predictions over development, using the co-occurrences that dominate the natural stimulus statistics. Yet, our results leave open the question of whether these long-term priors can be updated during adulthood, following extensive exposure to new statistical dependencies. This can be tested by exposing adult speakers to natural sounds encoding rules that “violate” the long-term prior (e.g., a language with opposite prosodic structure) and exploring the effects of such short-term exposure to behavioral and neural performance.

One potential alternative to the conjecture that the “duration prior” is linked to the spectrotemporal features of a language is that the prior depends on abstract syntactic/word-order rules. This latter account would predict that violations of long-term predictions in our study would lead to larger error responses in regions sensitive to syntactic variables, such as the left inferior frontal gyrus. Instead, the former account would predict that violations of long-term predictions elicit stronger responses in early left-lateralized auditory regions, which are putatively associated with early speech processing. The reconstruction of cortical sources associated with the omission of short and long tones in the two groups seems to suggest the “duration prior” might indeed depend more on the acoustic properties of a language than its syntactic configurations. Indeed, an interaction trend mirroring the one at the sensor level was only present in early left auditory regions, but not in the left inferior frontal gyrus or the right auditory regions (**fig. 3, B, C, D** [↗](#)).

Our results are in line with predictive coding models stating that predictions are organized hierarchically. When two predictive signals, one short-term and one long-term are violated, the amplitude of the prediction error is larger compared to a scenario in which only one short-term prediction is violated. This result complements previous studies using the local-global design showing that the same deviancy presented in different contexts gives rise to different error signals, such as the MMN and the P3 (Bekinschtein et al., 2009 [↗](#); Wacongne et al., 2011 [↗](#)). These studies provide empirical evidence that predictive coding of auditory sequences is organized at different functional levels, with early sensory regions using transition probabilities to generate expectations about the present, and frontal and associative regions inferring the global structure of an auditory event. Our results extend this work by providing direct evidence for the presence of a system in the auditory cortex that uses long-term natural sound statistics to generate long-term predictions. This interpretation is also supported by the reconstruction of cortical sources. Indeed, while the overall omission effect is larger in the right hemisphere (**Fig. 3 A** [↗](#)), the interaction is evident in the left hemisphere (**Fig. 3 B** [↗](#)). This finding further suggests that distinct cortical systems, supporting different predictive models, underlie the generation of the omission MMN.

Our findings are also consistent with more recent predictive coding models incorporating the idea of “stubborn” predictive signals – that is, predictions resilient to model updates. Unlike short-term expectations, long-term predictions are usually implemented as a computational constraint on input data, thus being largely unaffected by short-term experience (Teufel & Fletcher, 2020 [DOI](#)). In our study, the deployment of long-term predictions does not represent an effective coding strategy to perform the task. Yet, listeners still seem to assign different weights to incoming data, using a “default” predictive coding scheme that resembles the segmentation strategy used to parse speech material. Why should a neural system rely on such stubborn priors even when irrelevant to solving a given perceptual task? One possibility is that implementing stable priors as a constraint on perception is computationally less expensive in terms of metabolic costs than recalibrating cortical internal models anew based on any type of novel experience. Another possibility is that relying on unchanging predictive schemes helps the system to form coherent models in front of environmental contingencies, thus reflecting an effective computational strategy for the reduction of the long-term prediction error. Defining how stubborn predictions emerge during learning and what their computational role is represents an important challenge to understanding the role of prior experience in perceptual inference.

We also reported a main effect of omission type, indicating that the MMN generated by the omission of a long tone was generally larger compared to that generated by the omission of a short one. Because such group effect was consistent only in the Basque group, it is possible that it merely reflects a larger sensitivity of the auditory system of this group to the omission of long events, in line with the interaction reported above. Alternatively, this effect could be driven by the fact that, during language processing, major predictive resources are invested in predicting the onset of long events, compared to short ones, as the formers usually refer to content words i.e., semantically relevant events. Consequently, the auditory system may apply a similar predictive scheme also during the processing of non-linguistic sound sequences, independently of language background. Independently on the interpretation, the lack of a main effect of omission type in the control condition suggests that the long omission effect is driven by experience with the native language.

Our results also refine previous studies showing modulatory effects of (long-term) musical expertise on the MMN (e.g., Vuust et al., 2005 [DOI](#); 2009 [DOI](#)). These studies indicate that responses to violation during auditory rhythm perception are larger when the listener is an expert musician compared to a non-musician, pointing to the role of long-term auditory experience in shaping early predictive mechanisms. In our study, we manipulated long-term prediction orthogonally, with clear-cut predictions about the effect of language experience on early auditory predictive processing. Our results thus provide direct evidence for the presence of an active system in the auditory cortex that uses long-term priors to constrain information processing of incoming auditory stimuli.

Materials and Methods

Participants

In total, 20 native speakers of Spanish (mean age: 25.6 years, range: 20–33, 13 females) and 20 native speakers of Basque (mean age: 27.11 years, range: 22–40, 17 females) took part in the experiment. Members of the two groups were selected based on self-reported scores for exposure (percentage of time exposed to a given language at the time of testing) and speaking (percentage of time speaking a given language at the time of testing). Participants from the Basque group were living in a Basque-speaking region of the Basque Country. They all reported having learned Basque as a first language, being primarily exposed to Basque during daily life (mean exposure: 69%; SD: 13.28, range: 50–90%) and using it as a main language for communication (mean speaking: 77%; SD: 10.56; range: 60–90%). All native speakers of Basque reported having learned

Spanish as a second language. However, they had overall low exposure (mean exposure: 22%; SD: 10.31, range: 10-40%) and speaking scores for Spanish (mean speaking: 17%; SD: 7.33; range: 10-30%). In this respect, it is important to notice that previous behavioral studies on perceptual grouping in Basque bilinguals showed that language dominance is the main factor driving non-linguistic rhythmic grouping (Molnar et al., 2016). Therefore, despite limited exposure to the Spanish language, the formation of the hypothesized “duration prior” in the Basque group should be primarily linked to experience with the dominant language (i.e., Basque), with no or only minimal influence from Spanish. Participants from the Spanish dominant group were coming from different regions of Spain. All of them learned Spanish as their first language, and had high self-reported scores for Spanish exposure (mean exposure: 79%; SD: 9.67, range: 60-100%) and speaking (mean speaking: 88.5%; SD: 6.7; range: 80-100%). Spanish participants reported having learned a second or third language after childhood (e.g., Basque, English, Italian, and Catalan).

Stimuli and experimental design

Stimuli were created using Matlab Psychtoolbox and presented binaurally via MEG-compatible headphones. Experimental stimuli consisted of 60 sequences of two tones alternating in duration (short tones: 0.250 s; long tones: 0.437 s) with fixed intervals (0.02 s). Both long and short tones had a frequency of 500 Hz. The beginning and end of each tone were faded in and out of 0.015 s. Overall, each sequence consisted of 40 short-long tone pairs, for a total of 80 unique tones per sequence, and lasted around 30 s. Half of the sequences started with a long tone and half with a short tone. The beginning and the end of each sequence were faded in and faded out of 2.5 s to mask possible grouping biases. In each sequence, 2 to 6 tones were omitted and substituted with a 0.6 s silence gap. The larger gap was introduced to avoid that activity related to the onset of the tone following the omission overlaps with the activity generated by the omitted tone. Tone omissions occurred pseudorandomly, for a total of 240 omissions (120 short and 120 long). The pseudorandomization of the omissions consisted in separating the omissions within each sequence of at least 7 tones. In the control condition, sequences consisted of tones alternating in frequency at fixed intervals (0.02 s). High-frequency tones had a frequency of 700 Hz, while low-frequency tones had a frequency of 300 Hz. Both high and low-frequency tones had an overall duration of 0.343 s. This duration was selected to keep the overall length of the sequences equal to that of the test condition, by keeping the total number of 80 tones per sequence. As in the test condition, tones and sequences were faded in and out of 0.015 s and 2.5 s respectively. In each sequence, 2 to 6 tones were omitted and substituted with a 0.6 s silence gap.

Overall, the experiment was divided into two main blocks: test and control. The order in which the blocks were presented was counterbalanced across participants. Each block consisted of 60 sequences and lasted around 35 minutes. Each sequence was separated by an 8 s silence gap. Every twenty sequences, a short pause was introduced. The end of each block was followed by a longer pause.

Participants were requested to minimize movement throughout the experiment, except during pauses. Subjects were asked to keep their eyes open, to avoid eyes movements by fixating on a cross on the screen. Similarly to previous studies, the only task that was asked to subjects was to count how many omissions were present in each sequence (e.g., Bekinschtein et al., 2009) - and report it at the end of the sequence during the 8 s silence gap. Participants only received instructions at the very beginning of the task, and no verbal or written instructions were introduced during the task.

MEG Recordings

Measurements were carried out with the Elekta Neuromag VectorView system (Elekta Neuromag), which comprises 204 planar gradiometers and 102 magnetometers in a helmet-shaped array. ECG and electrooculogram (EOG) (horizontal and vertical) were recorded simultaneously as auxiliary channels. MEG and auxiliary channels were low-pass filtered at 330 Hz, high-pass filtered at 0.03

Hz, and sampled at 1 KHz. The head position with respect to the sensor array was determined by five head-position indicator coils attached to the scalp. The locations of the coils were digitized with respect to three anatomical landmarks (nasion and preauricular points) with a 3D digitizer (Polhemus Isotrak system). Then, the head position with respect to the device origin was acquired before each block.

Preprocessing

Signal space separation correction, head movement compensation, and bad channels correction were applied using the MaxFilter Software 2.2 (Elekta Neuromag). After that, data were analyzed using the FieldTrip toolbox (Oostenveld et al., 2011 [DOI](#)) in Matlab (MathWorks). Trials were initially epoched from 1.2 s before to 1.2 s after the onset of each tone or omitted tone. Epochs time-locked to the onset of short and long tones were undersampled to match approximately the number of their corresponding omissions. Trials containing muscle artifacts and jumps in the MEG signal were detected using an automatic procedure and removed after visual inspection. Subsequently, independent component analysis (Bell and Sejnowski, 1995 [DOI](#)) was performed to partially remove artifacts attributable to eye blinks and heartbeat artifacts (Jung et al., 2000 [DOI](#)). To facilitate the detection of components reflecting eye blinks and heartbeat artifacts, the coherence between all components and the ECG/EOG electrodes was computed. Components were inspected visually before rejection. On average, we removed 14.28% (SD = 5.71) of the trials and 2.45 (SD = 0.55) components per subject. After artifact rejection, trials were low-pass filtered at 40 Hz and averaged per condition and per subject. ERFs were baseline corrected using the 0.05 s preceding trial onset and resampled to 256 Hz. The latitudinal and longitudinal gradiometers were combined by computing the root mean square of the signals at each sensor position in order to facilitate the interpretation of the sensor-level data.

ERF analysis

Statistical analyses were performed using FieldTrip (Oostenveld et al., 2011 [DOI](#)) in Matlab 2014 (MathWorks) and R studio for post-hoc analysis. For data visualization, we used Matlab or FieldTrip plotting functions, R studio and the RainCloud plots tool (Allen et al., 2019 [DOI](#)). Plots were then arranged as cohesive images using Inkscape (<https://inkscape.org/> [DOI](#)). All comparisons were performed on combined gradiometer data. For statistical analyses, we used a univariate approach in combination with cluster-based permutations (Maris & Oostenveld, 2007 [DOI](#)) for family-wise error correction. This type of test controls the type I error rate in the context of multiple comparisons by identifying clusters of significant differences over space and time, instead of performing a separate test on each sensor and sample pair. Two-sided paired- and independent-samples t-tests were used for within- and between-subjects contrasts, respectively. The minimum number of neighboring channels required for a sample to be included in the clustering algorithm was set at 3. The cluster-forming alpha level was set at 0.05. The cluster-level statistic was the maximum sum of t-values (maxsum) and the number of permutations was set to 100000. To control for the false alarm rate, we selected the standard $\alpha = 0.05$. For the first analysis only, in which we compared ERF generated by pure tones vs omitted tones, we used a time-window between -0.05 and 0.350 s and considered both the spatial and temporal dimensions in the cluster-based permutation test. This explorative analysis was performed to assess the effect of unexpected omission, as well as its temporal unfolding. In all the remaining analyses, MMN responses were calculated by subtracting the ERFs of the tone from the ERFs of the omission. Moreover, all the contrasts were conducted using the average activity in the latency range between 0.100 and 0.250 s, which covers the typical latency of the MMN (Näätänen et al., 2007 [DOI](#); Garrido et al., 2009 [DOI](#)). This approach uses spatial clusters to test for the difference between conditions. When multiple clusters emerged from a comparison, only the most significant cluster was reported. When an interaction effect was detected, post hoc analyses were performed to assess its directionality. This was done by averaging data for each participant and condition over the preselected latency and channels belonging to the significant clusters (see *Results*). The ERFs generated by this averaging

were compared using one-sided independent sample t-tests. All p-values resulting from the post hoc t-test comparisons were FDR-corrected for multiple comparisons. All effect sizes reported are Cohen's d (Cohen, 1988).

Source reconstruction

Source reconstruction mainly focused on the statistically significant effects observed at the sensor-level ERF analysis. Individual T1-weighted MRI images were first segmented into scalp, skull, and brain components using the segmentation algorithm implemented in Fieldtrip (Oostenveld et al., 2011). A standard Montreal Neurological Institute (MNI) brain template available in SPM toolbox was used for the participants whose MRI was not available. Co-registration of the anatomical MRI images with MEG signal was performed using the manual co-registration tool available in Fieldtrip. The source space was defined as a regular 3D grid with a 5 mm resolution, and the lead fields were computed using a single-sphere head model for three orthogonal source orientations. The whole brain cortical sources of the MEG signals were estimated using a linearly constrained minimum variance (LCMV) beamformer approach (Van Veen et al., 1997). Only planar gradiometers were used for modelling source activity. First, we compared the average source activity associated with pure omissions and standard tones in the time range of 0.1 – 0.25 s after the stimulus onset. The covariance matrix used to derive LCMV beamformer weights was calculated from the 0.2 s preceding and 0.4 s following the onset of events (i.e., tones and omissions). Neural activity index (NAI) was computed as the ratio of source activity of tone to the omission ($NAI = S_{\text{Omission}}/S_{\text{Tone}}$). A non-linear transformation using the spatial-normalization algorithm (implemented in SPM8; Friston et al., 1994) was employed to transform individual MRIs to the standard Montreal Neurological Institute (MNI) brain. The source maps were plotted using the Surf Ice tool (<https://www.nitrc.org/projects/surfice/>)

Further, we selected predefined regions of interest for the subsequent analysis using the AAL atlas (Tzourio-Mazoyer et al., 2002). The regions of interest included the left and right auditory cortex (*Heschl's gyrus* and *superior temporal gyrus*) and the inferior frontal gyrus (*pars triangularis* and *pars opercularis*). We created LCMV filter using the same parameters used for whole brain source reconstruction. The virtual electrode in the source space corresponding to each ROI was generated. Singular vector decomposition (SVD) was applied to select the component with maximum variance. Later, the differential omission mismatch negativity was computed by subtracting the power of the long tones from the power of the long omissions and the power of the short tones from the power of the short omissions. This procedure was applied to each group separately (i.e., Spanish and Basque natives).

Acknowledgements

This research was supported by the Basque Government through the BERC 2022-2025 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S. Work by PM received support from “la Caixa” Foundation (ID 100010434) through the fellowship LCF/BQ/IN17/11620019, and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 713673. CM received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No: 819093), the Spanish Ministry of Economy and Competitiveness (PID2020-113926GB-I00), and the Basque Government (PIBA18_29). NM was supported by the Spanish Ministry of Economy and Competitiveness (PSI2015-65694-P, RTI2018-096311-B-I00, PDC2022-133917-I00). Work by ML received support from Juan de la Cierva IJC2020-042886-I. SN acknowledges the support from “The Adaptive Mind,” funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. We wish to express our gratitude to the BCBL lab staff and the research assistants who helped to recruit the participants and collect the data. We thank Ram Frost for providing helpful comments on the manuscript.

Author Contributions

PM, CM and NM designed the research; PM collected and analyzed the data; SN and ML helped with the analysis of the source-level data; PM created the figures; PM wrote the original draft. NM and CM reviewed and edited the original draft.

Data and materials availability

Data and codes for reproducing the analysis and figures are publicly available via the Open Science Framework (OSF): <https://osf.io/6jep8/> 

References

1. Clark A. (2013) **Whatever next? Predictive brains, situated agents, and the future of cognitive science** *Behavioral and brain sciences* **36**:181–204
2. Friston K. (2005) **A theory of cortical responses** *Philosophical transactions of the Royal Society B: Biological sciences* **360**:815–836
3. Rao R. P., Ballard D. H. (1999) **Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects** *Nature neuroscience* **2**:79–87
4. De Lange F. P., Heilbron M., Kok P. (2018) **How do expectations shape perception?** *Trends in cognitive sciences* **22**:764–779
5. Heilbron M., Chait M. (2018) **Great expectations: is there evidence for predictive coding in auditory cortex?** *Neuroscience* **389**:54–73
6. Sams M., Paavilainen P., Alho K., Näätänen R. (1985) **Auditory frequency discrimination and event-related potentials** *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* **62**:437–448
7. Garrido M. I., Kilner J. M., Stephan K. E., Friston K. J. (2009) **The mismatch negativity: a review of underlying mechanisms** *Clinical neurophysiology* **120**:453–463
8. Sallinen M., Kaartinen J., Lyytinen H. (1994) **Is the appearance of mismatch negativity during stage 2 sleep related to the elicitation of K-complex?** *Electroencephalography and Clinical Neurophysiology* **91**:140–148
9. Sculthorpe L. D., Ouellet D. R., Campbell K. B. (2009) **MMN elicitation during natural sleep to violations of an auditory pattern** *Brain research* **1290**:52–62
10. Strauss M., Sitt J. D., King J. R., Elbaz M., Azizi L., Buiatti M., Dehaene S. (2015) **Disruption of hierarchical predictive coding during sleep** *Proceedings of the National Academy of Sciences* **112**:E1353–E1362
11. Fischer C., Morlet D., Giard M. H. (2000) **Mismatch negativity and N100 in comatose patients** *Audiology and Neurotology* **5**:192–197
12. Näätänen R., Tervaniemi M., Sussman E., Paavilainen P., Winkler I. (2001) **‘Primitive intelligence’ in the auditory cortex** *Trends in neurosciences* **24**:283–288
13. Seriès P., Seitz A. R. (2013) **Learning what to expect (in visual perception)** *Frontiers in human neuroscience* **7**
14. Yon D., de Lange F. P., Press C. (2019) **The predictive brain as a stubborn scientist** *Trends in cognitive sciences* **23**:6–8
15. Teufel C., Fletcher P. C. (2020) **Forms of prediction in the nervous system** *Nature Reviews Neuroscience* **21**:231–242

16. Girshick A. R., Landy M. S., Simoncelli E. P. (2011) **Cardinal rules: visual orientation perception reflects knowledge of environmental statistics** *Nature neuroscience* **14**:926–932
17. Meyer T., Olson C. R. (2011) **Statistical learning of visual transitions in monkey inferotemporal cortex** *Proceedings of the National Academy of Sciences* **108**:19401–19406
18. Iversen J. R., Patel A. D., Ohgushi K. (2008) **Perception of rhythmic grouping depends on auditory experience** *The Journal of the Acoustical Society of America* **124**:2263–2271
19. Molnar M., Carreiras M., Gervain J. (2016) **Language dominance shapes non-linguistic rhythmic grouping in bilinguals** *Cognition* **152**:150–159
20. 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**:242–268
21. Maris E., Oostenveld R. (2007) **Nonparametric statistical testing of EEG-and MEG-data** *Journal of neuroscience methods* **164**:177–190
22. Yabe H., Tervaniemi M., Reinikainen K., Näätänen R. (1997) **Temporal window of integration revealed by MMN to sound omission** *Neuroreport* **8**:1971–1974
23. Rajj T., McEvoy L., Mäkelä J. P., Hari R. (1997) **Human auditory cortex is activated by omissions of auditory stimuli** *Brain research* **745**:134–143
24. 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**:2544–2590
25. Van Veen B. D., Van Drongelen W., Yuchtman M., Suzuki A. (1997) **Localization of brain electrical activity via linearly constrained minimum variance spatial filtering** *IEEE Transactions on biomedical engineering* **44**:867–880
26. Hughes H. C., Darcey T. M., Barkan H. I., Williamson P. D., Roberts D. W., Aslin C. H. (2001) **Responses of human auditory association cortex to the omission of an expected acoustic event** *Neuroimage* **13**:1073–1089
27. Wacongne C., Labyt E., Van Wassenhove V., Bekinschtein T., Naccache L., Dehaene S. (2011) **Evidence for a hierarchy of predictions and prediction errors in human cortex** *Proceedings of the National Academy of Sciences* **108**:20754–20759
28. Bendixen A., Schröger E., Winkler I. (2009) **I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system** *Journal of Neuroscience* **29**:8447–8451
29. 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**:8633–8639
30. Mlynarski W., McDermott J. H. (2019) **Ecological origins of perceptual grouping principles in the auditory system** *Proceedings of the National Academy of Sciences* **116**:25355–25364
31. Liu J., Hilton C. B., Bergelson E., Mehr S. A. (2023) **Language experience predicts music processing in a half-million speakers of fifty-four languages** *Current Biology* **33**:1916–1925

32. Bidelman G. M., Hutka S., Moreno S. (2013) **Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music** *PloS one* **8**
33. Tervaniemi M., Jacobsen T., Rottger S., Kujala T., Widmann A., Vainio M., Schroger E. (2006) **Selective tuning of cortical sound feature processing by language experience** *European Journal of Neuroscience* **23**:2538–2541
34. Bekinschtein T. A., Dehaene S., Rohaut B., Tadel F., Cohen L., Naccache L. (2009) **Neural signature of the conscious processing of auditory regularities** *Proceedings of the National Academy of Sciences* **106**:1672–1677
35. Vuust P., Pallesen K. J., Bailey C., Van Zuijen T. L., Gjedde A., Roepstorff A., Østergaard L. (2005) **To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians** *Neuroimage* **24**:560–564
36. Vuust P., Pallesen K. J., Bailey C., Van Zuijen T. L., Gjedde A., Roepstorff A., Østergaard L. (2005) **To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are leftlateralized in musicians** *Neuroimage* **24**:560–564
37. Oostenveld R., Fries P., Maris E., Schoffelen J. M. (2011) **FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data** *Computational intelligence and neuroscience* **2011**:1–9
38. Makeig S., Bell A., Jung T. P., Sejnowski T. J. (1995) **Independent component analysis of electroencephalographic data** *Advances in neural information processing systems* **8**
39. Jung T. P., Makeig S., Westerfield M., Townsend J., Courchesne E., Sejnowski T. J. (2000) **Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects** *Clinical neurophysiology* **111**:1745–1758
40. Allen M., Poggiali D., Whitaker K., Marshall T. R., Kievit R. A. (2019) **Raincloud plots: a multi-platform tool for robust data visualization** *Wellcome open research* **4**
41. Cohen J. (2013) **Statistical power analysis for the behavioral sciences**
42. Friston K. J., Holmes A. P., Worsley K. J., Poline J. P., Frith C. D., Frackowiak R. S. (1994) **Statistical parametric maps in functional imaging: a general linear approach** *Human brain mapping* **2**:189–210
43. Tzourio-Mazoyer N., Landeau B., Papathanassiou D., Crivello F., Etard O., Delcroix N., Joliot M. (2002) **Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain** *Neuroimage* **15**:273–289

Article and author information

Piermatteo Morucci

Department of Fundamental Neurosciences, University of Geneva, Switzerland, Basque Center on Cognition Brain and Language (BCBL), 20009, Donostia-San Sebastian, Spain

For correspondence: pmorucci@bcbl.eu

Sanjeev Nara

Basque Center on Cognition Brain and Language (BCBL), 20009, Donostia-San Sebastian, Spain,
Mathematical Institute, Department of Mathematics and Computer Science, Physics,
Geography, Justus-Liebig-Universität Gießen, Germany

Mikel Lizarazu

Basque Center on Cognition Brain and Language (BCBL), 20009, Donostia-San Sebastian, Spain

Clara Martin

Basque Center on Cognition Brain and Language (BCBL), 20009, Donostia-San Sebastian, Spain,
Ikerbasque, Basque Foundation for Science, 48013, Bilbao, Spain

Nicola Molinaro

Basque Center on Cognition Brain and Language (BCBL), 20009, Donostia-San Sebastian, Spain,
Ikerbasque, Basque Foundation for Science, 48013, Bilbao, Spain

Copyright

© 2023, Morucci et al.

This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editors

Reviewing Editor

Maria Chait

University College London, London, United Kingdom

Senior Editor

Barbara Shinn-Cunningham

Carnegie Mellon University, Pittsburgh, United States of America

Reviewer #1 (Public Review):

Summary:

In this work, the authors study whether the human brain uses long-term priors (acquired during our lifetime) regarding the statistics of auditory stimuli to make predictions respecting auditory stimuli. This is an important open question in the field of predictive processing.

To address this question, the authors cleverly profit from the naturally existing differences between two linguistic groups. While speakers of Spanish use phrases in which function words (short words like articles and prepositions) are followed by content words (longer words like nouns, adjectives, and verbs), speakers of Basque use phrases in the opposite order. Because of this, speakers of Spanish usually hear phrases in which short words are followed by longer words, and speakers of Basque experience the opposite. This difference in the order of short and longer words is hypothesized to result in a long-term duration prior that is used to make predictions regarding the likely durations of incoming sounds, even if they are not linguistic in nature.

To test this, the authors used MEG to measure the mismatch responses (MMN) elicited by the omission of short and long tones that were presented in alternation. The authors report an

interaction between the language background of the participants (Spanish, Basque) and the type of omission MMN (short, long), which goes in line with their predictions. They supplement these results with a source-level analysis.

Unfortunately, serious concerns regarding the predictions put forward by the authors, and the interaction effect found, make the interpretation of these results difficult.

Strengths:

This work has many strengths. To test the main question, the authors profit from naturally occurring differences in the everyday auditory experiences of two linguistic groups, which allows them to test the effect of putative auditory priors consolidated over years. This is a direct way of testing the effect of long-term priors.

The fact that the priors in question are linguistic, and that the experiment was conducted using non-linguistic stimuli (i.e. simple tones), allows for testing of whether these long-term priors generalize across auditory domains.

The experimental design is elegant and the analysis pipeline is appropriate. This work is very well written. In particular, the introduction and discussion sections are clear and engaging. The literature review is complete.

Weaknesses:

There are two main issues in this work. The first one pertains to the predictions put forward by the researchers, and the second with the interaction effect reported.

1. With respect to the predictions, the authors propose that the subjects, depending on their linguistic background and the length of the tone in a trial, can put forward one or two predictions. The first is a short-term prediction based on the statistics of the previous stimuli and identical for both groups (i.e. short tones are expected after long tones and vice versa). The second is a long-term prediction based on their linguistic background. According to the authors, after a short tone, Basque speakers will predict the beginning of a new phrasal chunk, and Spanish speakers will predict it after a long tone.

In this way, when a short tone is omitted, Basque speakers would experience the violation of only one prediction (i.e. the short-term prediction), but Spanish speakers will experience the violation of two predictions (i.e. the short-term and long-term predictions), resulting in a higher amplitude MMN. The opposite would occur when a long tone is omitted. So, to recap, the authors propose that subjects will predict the alternation of tone durations (short-term predictions) and the beginning of new phrasal chunks (long-term predictions).

The problem with this is that subjects are also likely to predict the completion of the current phrasal chunk. In speech, phrases are seldom left incomplete. In Spanish is very unlikely to hear a function-word that is not followed by a content-word (and the opposite happens in Basque). On the contrary, after the completion of a phrasal chunk, a speaker might stop talking and a silence might follow, instead of the beginning of a new phrasal chunk.

Considering that the completion of a phrasal chunk is more likely than the beginning of a new one, the prior endowed to the participants by their linguistic background should make us expect a pattern of results actually opposite to the one reported here.

1. The authors report an interaction effect that modulates the amplitude of the omission response, but caveats make the interpretation of this effect somewhat uncertain. The authors report a widespread omission response, which resembles the classical mismatch response (in MEG) with strong activations in sensors over temporal regions. Instead, the interaction found is circumscribed to four sensors that do not overlap with the peaks of activation of the omission response. Furthermore, the boxplot in Figure 2E suggests that part of the interaction effect might be due to the presence of two outliers (if removed, the effect is no longer significant). Overall, it is possible that the reported interaction is driven by a main effect of omission type which the authors report, and find consistently only in the Basque group (showing a higher amplitude omission response for long tones than for short tones).

Because of these points, it is difficult to interpret this interaction as a modulation of the omission response. It should also be noted that in the source analysis, the interaction only showed a trend in the left auditory cortex, but in its current version the manuscript does not report the statistics of such a trend.

- <https://doi.org/10.7554/eLife.91636.1.sa2>

Reviewer #2 (Public Review):

Summary:

Morucci et al. tested the influence of linguistic prosody long-term priors in forming predictions about simple acoustic rhythmic tone sequences composed of alternating tone duration, by violating context-dependent short-term priors formed during sequence listening. Spanish and Basque participants were selected due to the different rhythmic prosody of the two languages (functor-initial vs. Functor final, respectively), despite a common cultural background. The authors found that neuromagnetic responses to casual tone omissions reflected the linguistic prosody pattern of the participant's dominant language: in Spanish speakers, omission responses were larger to short tones, whereas in Basque speakers, omission responses were larger to long tones. Source localization of these responses revealed this interaction pattern in the left auditory cortex, which the authors interpret as reflecting a perceptual bias due to acoustic cues (inherent linguistic rhythms, rather than linguistic content). Importantly, this pattern was not found when the rhythmic sequence entailed pitch, rather than duration, cues. To my knowledge, this is the first study providing neural signatures of a known behavioral effect linking ambiguous rhythmic tone sequence perceptual organization to linguistic experience.

The conclusions of the study are well supported by the data, albeit weakly by the source analysis, but I have the impression that the rationale of the study and the analyses performed may be missing an important aspect of rhythmic sequence perception, namely the involvement of entrained oscillatory activity to the perceived rhythm, particularly phase alignment to pattern onsets. This view would not change the impact of the results but add depth to their interpretation.

Strengths:

1. The choice of participants. The bilingual population of the Basque country is perfect for performing studies that need to control for cultural and socio-economic background while having profound linguistic differences. In this sense, having dominant Basque speakers as a sample equates that in Molnar et al. (2016), and thus overcomes the lack of direct behavioral evidence for a difference in rhythmic grouping across linguistic groups. Molnar et al. (2016)'s evidence on the behavioral effect is compelling, and the evidence on neural signatures provided by the present study aligns with it.
2. The experimental paradigm. It is a well-designed acoustic sequence, that considers aspects such as gap length insertion, to be able to analyze omission responses free from subsequent stimulus-driven responses, and which includes a control sequence that uses pitch instead of duration as a cue to rhythmic grouping, which provides a stronger case for the differences found between groups to be due to prosodic duration cues.
3. Data analyses. Sound, state-of-the-art methodology in the event-related field analyses at the sensor level.

Weaknesses:

1. Despite the evidence provided on neural responses, the main conclusion of the study reflects a known behavioral effect on rhythmic sequence perceptual organization driven by linguistic background (Molnar et al. 2016, particularly). Also, the authors themselves provide a good review of the literature that evidences the influence of long-term priors in neural responses related to predictive activity. Thus, in my opinion, the strength of the statements the authors make on the novelty of the findings may be a bit far-fetched in some instances.
2. Albeit the paradigm is well designed, I fail to see the grounding of the hypotheses laid by the authors as framed under the predictive coding perspective. The study assumes that responses to an omission at the beginning of a perceptual rhythmic pattern will be stronger than at the end. I feel this is unjustified. If anything, omission responses should be larger when the gap occurs at the end of the pattern, as that would be where stronger expectations are placed: if in my language a short sound occurs after a long one, and I perceptually group tone sequences of alternating tone duration accordingly, when I hear a short sound I will expect a long one following; but after a long one, I don't necessarily need to expect a short one, as something else might occur.
3. In this regard, it is my opinion that what is reflected in the data may be better accounted for (or at least, additionally) by a different neural response to an omission depending on the phase of an underlying attentional rhythm (in terms of Large and Jones rhythmic attention theory, for instance) and putative underlying entrained oscillatory neural activity (in terms of Lakatos' studies, for instance). Certainly, the fact that the aligned phase may differ depending on linguistic background is very interesting and would reflect the known behavioral effect.
4. Source localization is performed on sensor-level significant data. The lack of source-level statistics weakens the conclusions that can be extracted. Furthermore, only the source reflecting the interaction pattern is taken into account in detail as supporting their hypotheses, overlooking other sources. Also, the right IFG source activity is not depicted, but looking at whole brain maps seems even stronger than the left. To sum up, source localization data, as informative as it could be, does not strongly support the author's claims in its current state.

- <https://doi.org/10.7554/eLife.91636.1.sa1>

Reviewer #3 (Public Review):

Summary:

The paper investigates the effects of long-term linguistic experience on early auditory processing, a subject that has been relatively less studied compared to short-term influences. Using MEG, the study examines brain responses to auditory stimuli in speakers of Spanish and Basque, whose syntactic rules provide different degrees of exposure to durational patterns (long-short vs short-long). The findings suggest that both long-term language experience, as well as short-term transitional probabilities, can shape auditory predictive coding for non-linguistic sound sequences, evidenced by differences in mismatch negativity amplitudes localised to the left auditory cortex.

Strengths:

The study integrates linguistics and auditory neuroscience in an interesting interdisciplinary way that may interest linguists as well as neuroscientists. The fact that long-term language experience affects early auditory predictive coding is important for understanding group and individual differences in domain-general auditory perception. It has importance for neurocognitive models of auditory perception (e.g. inclusion of long-term priors), and will be of interest to researchers in linguistics, auditory neuroscience, and the relationship between language and perception. The inclusion of a control condition based on pitch is also a strength.

Weaknesses:

The main weaknesses are the strength of the effects and generalisability. The sample size is also relatively small by today's standards, with $N=20$ in each group. Furthermore, the crucial effects are all mostly in the $.01 > P < .05$ range, such as the crucial interaction $P=.03$. It would be nice to see it replicated in the future, with more participants and other languages. It would also have been nice to see behavioural data that could be correlated with neural data to better understand the real-world consequences of the effect.

- <https://doi.org/10.7554/eLife.91636.1.sa0>

Author Response

Reviewer 1 (Public Review):

1. With respect to the predictions, the authors propose that the subjects, depending on their linguistic background and the length of the tone in a trial, can put forward one or two predictions. The first is a short-term prediction based on the statistics of the previous stimuli and identical for both groups (i.e. short tones are expected after long tones and vice versa). The second is a long-term prediction based on their linguistic background. According to the authors, after a short tone, Basque speakers will predict the beginning of a new phrasal chunk, and Spanish speakers will predict it after a long tone.

In this way, when a short tone is omitted, Basque speakers would experience the violation of only one prediction (i.e. the short-term prediction), but Spanish speakers will experience the violation of two predictions (i.e. the short-term and long-term predictions), resulting in a higher amplitude MMN. The opposite would occur when a long tone is omitted. So, to recap, the authors propose that subjects will predict the alternation of tone durations (short-term predictions) and the beginning of new phrasal chunks (long-term predictions).

The problem with this is that subjects are also likely to predict the completion of the current phrasal chunk. In speech, phrases are seldom left incomplete. In Spanish is very unlikely to hear a function-word that is not followed by a content-word (and the opposite happens in Basque). On the contrary, after the completion of a phrasal chunk, a speaker might stop talking and a silence might follow, instead of the beginning of a new phrasal chunk.

Considering that the completion of a phrasal chunk is more likely than the beginning of a new one, the prior endowed to the participants by their linguistic background should make us expect a pattern of results actually opposite to the one reported here.

Response: We acknowledge the plausibility of the hypothesis advanced by Reviewer #1. We would like to further clarify the rationale that led us to predict that the hypothesized long-term predictions should manifest at the onset of (and not within) a “phrasal chunk”. The hypothesis does not directly concern the probability of a short event to follow a long one (or the other way around), which to our knowledge has not been systematically quantified in previous cross-linguistic studies. Rather, it concerns how the auditory system forms higher-level auditory chunks based on the rhythmic properties of the native language, which is what the previous behavioral studies on perceptual grouping have addressed (e.g., Iversen 2008; Molnar et al. 2014; Molnar et al. 2016). When presented with sequences of two tones alternating in duration, Spanish speakers typically report perceiving the auditory stream as a repetition of short-long chunks separated by a pause, while speakers of Basque usually report the opposite long-short grouping bias. These results suggest that the auditory system performs a chunking operation by grouping pairs of tones into compressed, higher-level auditory units (often perceived as a single event). The way two constituent tones are combined depends on linguistic experience. Based on this background, we hypothesized the presence of (i) a short-term system that merely encodes a repetition of alternations rule and predicts transitions from one constituent tone to the other ($a \rightarrow b \rightarrow a \rightarrow b$, etc.); (ii) a long-term system that encodes a repetition of concatenated alternations rule and predicts transitions from one high-level unit to the other ($ab \rightarrow ab$, etc.). Under this view, we expect predictions based on the long-term system to be stronger at the onset of (rather than within) high-level units and therefore omissions of the first constituent tone to elicit larger responses than omissions of the second constituent tone.

In other words, the omission of the onset tone would reflect the omission of the whole chunk. On the other hand, the omission of the internal tone would be better handled by the short-term system, involved in processing the low-level structure of our sequences.

A similar concern was also raised by Reviewer #2. We will include the view proposed by Reviewer #1 and Reviewer #2 in the updated version of the manuscript.

1. The authors report an interaction effect that modulates the amplitude of the omission response, but caveats make the interpretation of this effect somewhat uncertain. The authors report a widespread omission response, which resembles the classical mismatch response (in MEG) with strong activations in sensors over temporal regions. Instead, the interaction found is circumscribed to four sensors that do not overlap with the peaks of activation of the omission response.

Response: We appreciate that all three reviewers agreed on the robustness of the data analysis pipeline. The approach employed to identify the presence of an interaction effect was indeed conservative, using a non-parametric test on combined gradiometers data, no a priori assumptions regarding the location of the effect, and small cluster thresholds ($\text{cfg.clusteralpha} = 0.05$) to enhance the likelihood of detecting highly localized clusters with large effect sizes. This approach led to the identification of the cluster illustrated in Figure 2c,

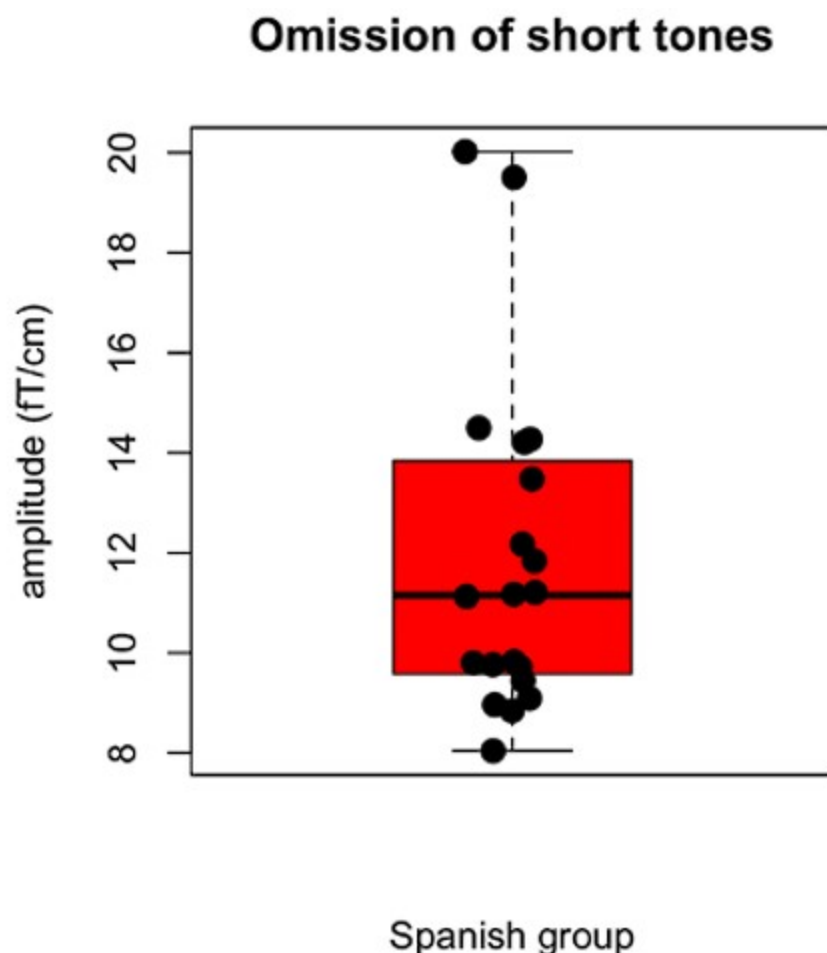
where the interaction effect is evident. The fact that this interaction effect arises in a relatively small cluster of sensors does not alter its statistical robustness. The only partial overlap of the cluster with the activation peaks might simply reflect the fact that distinct sources contribute to the generation of the omission-MMN, which has been demonstrated in numerous prior studies (e.g., Zhang et al., 2018; Ross & Hamm, 2020).

Furthermore, the boxplot in Figure 2E suggests that part of the interaction effect might be due to the presence of two outliers (if removed, the effect is no longer significant). Overall, it is possible that the reported interaction is driven by a main effect of omission type which the authors report, and find consistently only in the Basque group (showing a higher amplitude omission response for long tones than for short tones). Because of these points, it is difficult to interpret this interaction as a modulation of the omission response.

Response: The two participants mentioned by Reviewer #1, despite being somewhat distant from the rest of the group, are not outliers according to the standard Tukey's rule. As shown in Author response image 1 below, no participant fell outside the upper ($Q3+1.5 \times IQR$) and lower whiskers ($Q1-1.5 \times IQR$) of the boxplot.

Author response image 1.

The presence of a main effect of omission type does not impact the interpretation of the interaction, especially considering that these effects emerge over distinct clusters of channels.



The code to generate Author response image 1 and the corresponding statistics have been added to the script “analysis_interaction_data.R” in the OSF folder (<https://osf.io/6jep8/>).

It should also be noted that in the source analysis, the interaction only showed a trend in the left auditory cortex, but in its current version the manuscript does not report the statistics of such a trend.

Response: Our interpretation of the results for the present study is mainly driven by the effect observed on sensor-level data, which is statistically robust. The source modeling analyses (in non-invasive electrophysiology) provide a possible model of the candidate brain sources driving the effect observed at the sensor level. The source showing the interactive effect in our study is the left auditory cortex. More details and statistics will be provided in the reviewed version of the manuscript.

Reviewer #2 (Public Review):

1. Despite the evidence provided on neural responses, the main conclusion of the study reflects a known behavioral effect on rhythmic sequence perceptual organization driven by linguistic background (Molnar et al. 2016, particularly). Also, the authors themselves provide a good review of the literature that evidences the influence of long-term priors in neural responses related to predictive activity. Thus, in my opinion, the strength of the statements the authors make on the novelty of the findings may be a bit far-fetched in some instances.

Response: We will consider the suggestion of reviewer #2 for the new version of the manuscript. Overall, we believe that the novelty of the current study lies in bridging together findings from two research fields - basic auditory neuroscience and cross-linguistic research - to provide evidence for a predictive coding model in the auditory that uses long-term priors to make perceptual inferences.

1. Albeit the paradigm is well designed, I fail to see the grounding of the hypotheses laid by the authors as framed under the predictive coding perspective. The study assumes that responses to an omission at the beginning of a perceptual rhythmic pattern will be stronger than at the end. I feel this is unjustified. If anything, omission responses should be larger when the gap occurs at the end of the pattern, as that would be where stronger expectations are placed: if in my language a short sound occurs after a long one, and I perceptually group tone sequences of alternating tone duration accordingly, when I hear a short sound I will expect a long one following; but after a long one, I don't necessarily need to expect a short one, as something else might occur.

Response: A similar point was advanced by Reviewer #1. We tried to clarify our hypothesis (see above). We will consider including this interpretation in the updated version of the manuscript.

1. In this regard, it is my opinion that what is reflected in the data may be better accounted for (or at least, additionally) by a different neural response to an omission depending on the phase of an underlying attentional rhythm (in terms of Large and Jones rhythmic attention theory, for instance) and putative underlying entrained oscillatory neural activity (in terms of Lakatos' studies, for instance). Certainly, the fact that the aligned phase may differ depending on linguistic background is very interesting and would reflect the known behavioral effect.

Response: We thank the reviewer for this comment, which is indeed very pertinent. Below are some comments highlighting our thoughts on this.

1. We will explore in more detail the possibility that the aligned phase may differ depending on linguistic background, which is indeed very interesting. However, we believe that even if a phase modulation by language experience is found, it would not negate the possibility that the group differences in the MMN are driven by different long-term predictions. Rather, since the hypothesized phase differences would be driven by long-term linguistic experience, phase entrainment may reflect a mechanism through which long-term predictions are carried. On this point, we agree with the Reviewer when says that “this view would not change the impact of the results but add depth to their interpretation”.
2. Related to the point above: Despite evoked responses and oscillations are often considered distinct electrophysiological phenomena, current evidence suggests that these phenomena are interconnected (e.g., Studenova et al., 2023). In our view, the hypotheses that the MMN reflects differences in phase alignment and long-term prediction errors are not mutually exclusive.
3. Despite the plausibility of the view proposed by reviewer #2, many studies in the auditory neuroscience literature putatively consider the MMN as an index of prediction error (e.g., Bendixen et al., 2012; Heilbron and Chait, 2018). There are good reasons to believe that also in our study the MMN reflects, at least in part, an error response.

In the updated version of the manuscript, we will include a paragraph discussing the possibility that the reported group differences in the omission MMN might be partially accounted for by differences in neural entrainment to the rhythmic sound sequences.

Reviewer #3 (Public Review):

The main weaknesses are the strength of the effects and generalisability. The sample size is also relatively small by today's standards, with N=20 in each group. Furthermore, the crucial effects are all mostly in the .01>P<.05 range, such as the crucial interaction P=.03. It would be nice to see it replicated in the future, with more participants and other languages. It would also have been nice to see behavioural data that could be correlated with neural data to better understand the real-world consequences of the effect.

Response: We appreciate the positive feedback from Reviewer #3. Concerning this weakness highlighted: we agree with Reviewer #3 that it would be nice to see this study replicated in the future with larger sample sizes and a behavioral counterpart. Overall, we hope this work will lead to more studies using cross-linguistic/cultural comparisons to assess the effect of experience on neural processing. In the context of the present study, we believe that the lack of behavioral data does not undermine the main findings of this study, given the careful selection of the participants and the well-known robustness of the perceptual grouping effect (e.g., Iversen 2008; Yoshida et al., 2010; Molnar et al. 2014; Molnar et al. 2016). As highlighted by Reviewer #2, having Spanish and Basque dominant “speakers as a sample equates that in Molnar et al. (2016), and thus overcomes the lack of direct behavioral evidence for a difference in rhythmic grouping across linguistic groups. Molnar et al. (2016)’s evidence on the behavioral effect is compelling, and the evidence on neural signatures provided by the present study aligns with it.”

References

1. Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology*, 83(2), 120-131.
2. Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex?. *Neuroscience*, 389, 54-73.
3. Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America*, 124(4), 2263-2271.
4. Molnar, M., Lallier, M., & Carreiras, M. (2014). The amount of language exposure determines nonlinguistic tone grouping biases in infants from a bilingual environment. *Language Learning*, 64(s2), 45-64.
5. Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition*, 152, 150-159.
6. Ross, J. M., & Hamm, J. P. (2020). Cortical microcircuit mechanisms of mismatch negativity and its underlying subcomponents. *Frontiers in Neural Circuits*, 14, 13.
7. Simon, J., Balla, V., & Winkler, I. (2019). Temporal boundary of auditory event formation: An electrophysiological marker. *International Journal of Psychophysiology*, 140, 53-61.
8. Studenova, A. A., Forster, C., Engemann, D. A., Hensch, T., Sander, C., Mauche, N., ... & Nikulin, V. V. (2023). Event-related modulation of alpha rhythm explains the auditory P300 evoked response in EEG. *bioRxiv*, 2023-02.
9. Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*, 115(2), 356-361.
10. Zhang, Y., Yan, F., Wang, L., Wang, Y., Wang, C., Wang, Q., & Huang, L. (2018). Cortical areas associated with mismatch negativity: A connectivity study using propofol anesthesia. *Frontiers in Human Neuroscience*, 12, 392.