Alpha and beta rhythms differentially support the effect of symbols on visual object recognition

Piermatteo Morucci[1*], Francesco Giannelli[2,3], Craig Richter[1], Nicola Molinaro[1,4]

1. Basque Center on Cognition Brain and Language (BCBL), University of the Basque Country UPV/EHU, 20009, Donosti-San Sebastian, Spain

2. Cognition and Brain Plasticity Unit, IDIBELL, 08908, L'Hospitalet de Llobregat, Barcelona, Spain

3. Department of Cognition, Development and Educational Psychology, University of Barcelona, 08035, Barcelona, Spain

4. Ikerbasque, Basque Foundation for Science, 48013, Bilbao, Spain

*Corresponding Author:

Piermatteo Morucci

Basque Center on Cognition Brain and Language

Mikeletegi Pasealekua, 69, 20009 Donostia, Gipuzkoa

Tel: +34 943 309 300 Ext.: 307

p.morucci@bcbl.eu

Abstract

Hearing spoken words can enhance visual object recognition, detection and discrimination. Yet, the mechanism underlying this facilitation is incompletely understood. On one account, words would not bias visual processes at early levels, but rather interact at later decision-making stages. More recent proposals posit that words can alter visual processes at early stages by activating category-specific priors in sensory regions. A prediction of this account is that top-down priors evoke changes in occipital areas before the presentation of visual stimuli. Here, we tested the hypothesis that neural oscillations can serve as a mechanism to activate language-mediated visual priors. Participants performed a cue-picture matching task where cues were either spoken words, in their native or second language, or natural sounds, while EEG and reaction times were recorded. Behaviorally, we replicated the previously reported label-advantage effect, with images cued by words being recognized faster than those cued by natural sounds. A time-frequency analysis of cue-target intervals revealed that this label-advantage was associated with enhanced power in posterior alpha (9-11 Hz) and beta oscillations (17-19 Hz), both of which were larger when the image was preceded by a word compared to a natural sound. Prestimulus alpha and beta rhythms were correlated with reaction time performance, yet they appeared to operate in different ways. Reaction times were faster when alpha power increased, but slowed down with enhancement of beta oscillations. These results suggest that alpha and beta rhythms work in tandem to support language-mediated visual object recognition, while showing an inverse relationship to behavioral performance.

Keywords: electroencephalography, neural oscillations, concepts, categorization, language and perception

Introduction

Hearing certain natural sounds (e.g., the croak of a frog) appears to automatically activate conceptual knowledge, enabling the perceptual system to quickly identify objects in the surroundings (e.g., the presence of a frog). Learning such cross-modal associations represents a crucial prerequisite for mediating interactions with the environment. In humans, conceptual representations can also be activated via language (e.g., "frog"). However, unlike natural sounds, linguistic symbols are categorical, making them uniquely suited to activate semantic information in a format that transcends within-category differences. Whether phylogenetically young systems like language exert similar effects on perception as natural sounds do, and which brain dynamics support such effects is still incompletely understood. In the present study, we test the hypothesis that language boosts visual processes by sharpening categorical priors via the modulation of alpha/beta oscillations.

Conceptual representations activated by auditory cues have been shown to interact with the visual system in different ways. For instance, hearing words and natural sounds can rapidly drive visual attention towards specific entities in a scene (Huettig and Altmann, 2007); facilitate the recognition and discrimination of congruent object categories (Edmiston and Lupyan, 2015; Boutonnet and Lupyan, 2015); lower the detection threshold for ambiguous objects (Lupyan and Ward, 2013; Ostarek and Huettig, 2017); and even cause sensory illusions (Dils and Boroditsky, 2010). While this body of evidence suggests that both linguistic and non-linguistic cues activate content-specific representations, it is currently less clear whether these cues activate the *same* representations. Studies directly targeting this issue have often reported a "label-advantage" effect – that is, a facilitation in the recognition of objects when preceded by words compared to non-linguistic cues (Edmiston and Lupyan, 2015) – suggesting that language represents a more powerful tool to enhance visual processing.

To achieve these facilitatory effects on visual perception, linguistic categories could theoretically follow two possible pathways. On one account, language would not bias perceptual processes at early levels, but rather interacts with later processing

stages such as categorical decision-making. On an alternative account, words can affect visual processing by setting categorical priors with the effect of altering early perceptual processing (Simanova et al., 2016). Support for the latter account comes primarily from EEG studies showing that the better recognition of images preceded by congruent words was associated with modulations of early event related potentials (ERPs) such as the P1 (Boutonnet and Lupyan, 2015, Noorman et al., 2018) – putatively considered an electrophysiological index of low-level visual processes (Spehlmann, 1965). Yet, these experiments targeted the perceptual consequences that language has on visual behavior i.e., they focused on the post-stimulus time interval. The mechanisms that underlie prestimulus effects of language on visual perception are currently poorly understood.

Analysis of oscillatory activity provides an excellent opportunity to study prestimulus language-driven modulations in sensory areas. Based on previous human and animal studies, a candidate mechanism to carry perceptual priors are oscillations in the alpha/beta frequency range. Rhythmic brain activity in these bands has been repeatedly associated with top-down processes (Michalareas et al., 2016, Bressler and Richter, 2017; Arnal and Giraud, 2012).

In the present study, we used a cue-picture matching task to test the hypothesis that language enhances visual object recognition by setting categorical priors via the modulation of alpha/beta oscillations. In contrast to previous studies, we (i) focused on the time interval preceding the onset of the visual object, targeting top-down signaling directly; and (ii) included words in both first (L1) and second language (L2), in order to assess whether the previously reported label-advantage extends to language systems acquired later in development. We hypothesized that, if the label-advantage arises because words provide refined categorical priors to the visual system, then differences in recognizing objects when cued by words vs. natural sounds should be associated with modulations of oscillatory alpha/beta dynamics before the onset of the target picture. Importantly, we should also expect such an oscillatory index to be linked to behavioral performance.

Materials and methods

Participants

We tested a total of twenty-five Basque-Spanish bilingual speakers. Notice that in earlier studies investigating the label advantage in object recognition, a sample size of 15 participants was sufficient to detect the behavioral label-advantage effect (Boutonnet and Lupyan, 2015). Participants were native speakers of Basque who began acquisition of Spanish after three years of age (13 females; age range 18-33, mean: 25.66, SD: 5.45, age of acquisition of Spanish: 4.23 y.o., SD: 1.33). All participants were right-handed, with no history of neurological disorders. Their vision was normal or corrected to normal and received a payment of 10€ per hour for their participation. Before taking part in the experiment, all participants signed an informed consent form. The study was approved by the Basque Center on Cognition, Brain and Language (BCBL) ethics committee in compliance with the Declaration of Helsinki. Participants completed several language proficiency tests in both Spanish and Basque (see Table 1). First, participants were asked to self-rate their language comprehension (on a scale from 1 to 10, where 10 is a native-like level). All participants rated themselves as highly proficient in both Basque and Spanish. Participants also performed "LexTALE", a lexical decision task (Izura et al., 2014; Lemhofer and Broersma, 2012) that tested their vocabulary knowledge. They displayed similarly high scores in both Spanish and Basque. In addition, participants had to name a series of pictures of increasing difficulty in both languages. Here as well, participants achieved native-range scores in both languages. Finally, all participants were interviewed by balanced bilingual linguists who rated them on a scale from 0 to 5: no participants had a score below 4 in either language.

table 1 around here

Stimuli

The visual stimuli included 50 pictures from 10 object categories, referring to both animate (e.g., bird) or inanimate entities (e.g., camera). Each of the 10 categories was

represented by 5 different highly recognizable images (.png extension, white background, 2000x2000 pixels): three color photographs obtained from online image collections, one normed color drawing (Rossion and Pourtois, 2004), and one "cartoon" image (Lupyan and Thompson- Schill, 2012). We selected different instances for each category in order to provide visual heterogeneity.

The audio stimuli included 10 words in Basque (L1), 10 in words Spanish (L2) and 10 natural sounds, each referring to one of the object categories. Words, both in Basque (L1) and in Spanish (L2), were recorded by a Spanish-Basque female speaker. Natural wound stimuli were downloaded from online libraries. Overall, the mean length of the audio stimuli was 0.8 ± 0.05 seconds (Word in L2, mean: 0.81 s, SD: 0.21; Word in L1, mean: 0.77 s, SD: 0.23; Natural Sounds, mean: 0.84 s, SD: 0.2).

In order to test that sounds and images were unequivocally identifiable, a side test was performed. A group of Basque-Spanish bilinguals (N=20) who did not take part in the experiment viewed several images and listened to different sounds. They were told to name the visual and audio stimuli they perceived with the first noun that came to their mind. For the present experiment, we only chose the images and sounds whose names were expressed by all 20 participants. In total, we selected 50 images from 10 categories, 10 words in Basque, 10 words in Spanish, and 10 natural sounds.

Procedure

The EEG experiment was run in a soundproof electrically shielded chamber with dim light. Participants sat on a chair, about sixty centimeters in front of the computer screen. Stimuli were delivered using PsychoPy software (Peirce, 2007). We followed the procedure illustrated by Boutonnet and Lupyan (2015). Participants completed a cued-picture recognition task composed of 300 trials (see Fig.1). On each trial, a fixation point appeared on the center of the screen for one second, then participants heard an auditory cue: either a word in L1, (e.g., igela, "frog), a word in L2 (e.g., rana, "frog") or a natural sound (e.g., a croak).

After 1s from the offset of the cue, a picture appeared on the screen, and participants had to respond "yes" or "no" by pressing one of two buttons on a keyboard

to indicate whether the picture did or did not match the auditory cue at the category level. The picture remained on the screen until the participant's response. In 50% of the trials the picture matched the auditory cue (congruent trials), while the other 50% was a mismatch (incongruent trials). In the case of incongruent trials, a picture belonging to a different category appeared on the screen. Stimuli presentation was randomized for each participant. The entire experiment lasted 40 minutes on average.

figure 1 around here

EEG recording

Electrophysiological activity was recorded from 27 electrodes (Fp1/2, F7/8, F3/4, FC5/6, FC1/2, T7/8, C3/4, CP1/2, CP5/6, P3/4, P7/8, O1/2, F/C/Pz) positioned in an elastic cap (Easycap) according to the extended 10–20 international system. All sites were referenced online to the left mastoid (A1). Additional external electrodes were placed on the right mastoid (A2) and around the eyes (VEOL, VEOR, HEOL, HEOR) to detect blinks and eye movements. Data were amplified (Brain Amp DC) with a filter bandwidth of 0.01-100 Hz, at a sampling rate of 250 Hz. The impedance of the scalp electrodes was kept below 5 kΩ, while eye electrode impedance was kept below 10 kΩ.

EEG preprocessing

All EEG data analysis was performed using Matlab 2014 with the Fieldtrip toolbox (Oostenveld et al., 2011; http: www.fieldtriptoolbox.org) and R (R Core Team, 2015; https: www.r-project.org). For data visualization, we used Matlab or FieldTrip plotting functions, R and the RainCloud plots tool (Allen et al., 2019). The recordings were re-referenced off-line to the average activity of the two mastoids. Epochs of interest were selected based on cue type (word in L1, word in L2, natural sounds) and congruency (match, mismatch), resulting in six different sets of epochs. They were computed from –3 s to 1.5 s with respect to image onset.

Trials in which subjects provided incorrect responses in the behavioral task were removed from the analysis. Spatial-temporal components of the data containing eye and heart artifacts were identified using independent component analysis and subsequently removed. Overall, we removed an average of 2.14 components per subject. We then identified epochs containing additional 'muscle' and 'eye blink' artifacts using an automatic artifact detection procedure (z-value threshold = 12). Trials selected as possibly contaminated by artifacts were visually inspected and removed (~8%). Finally, we removed a few additional trials containing artifacts using a visual inspection procedure (~0.11%). Three participants were excluded from the analysis because more than 25 % of the trials were rejected.

Statistical analysis

Behavior. We used the R environment (version 4.0.0; R Core Team 2020) and lme4 package (Bates et al., 2014) to perform mixed effect regression on reaction time data, following a procedure similar to that illustrated in Boutonnet and Lupyan (2015). Predicted reaction times (calculated from the onset of the target image up to the participant's response) were computed by fitting the model with cue-type (words in L1, words in L2, natural sounds), congruency (match, mismatch) and their interaction as fixed factors, and by adding by subject random slopes for the effect of cue type and congruency. Subsequent pairwise comparisons were performed using estimated marginal means (Bonferroni-corrected for multiple comparisons) with emmeans (Lenth, 2018). Because no reliable interaction was detected, post-hoc comparisons were based on a model with the same syntax as the one presented above but without including the interaction term, in order to facilitate the interpretability of post-hoc analysis. Accuracy was not analyzed statistically because it was near ceiling (98%). For the analysis of behavioral data, we excluded the same three participants that were excluded from the EEG analysis. Moreover, we excluded all incorrect trials (1.88%), as well as a few trials in which participants' responses exceeded 3 s (0.28%). These trials were also excluded from the EEG analysis. Before entering the statistical models, reaction times were log-transformed to improve normality.

Spectral power. A time-frequency analysis of artifact-free EEG trials was performed. Before applying spectral decomposition, the latency of each epoch was reduced to –1.5 s to 0.5 s with respect to image onset. The time-varying power spectrum of single trials was obtained using a Hann sliding window approach (0.5 s window, 0.05 s time steps) for the frequency range between 0 and 30 Hz, zero-padded to 1 s providing a frequency resolution of 1 Hz. Our focus on oscillatory activity up to 30 Hz was motivated by the fact that top-down processes are often associated with oscillations within this frequency band, while higher frequencies have been traditionally linked to bottom-up processing (e.g., Bosman et al., 2012). For the statistical analysis, we computed a single power spectral density estimate for each participant, channel, frequency and epoch by averaging the spectral estimates centered on the -0.75 s to -0.25 s time interval. We selected this time-interval in order to yield more accurate spectral estimates, as activity here is largely uncontaminated by activity evoked by the preceding auditory event or the subsequent visual stimulus.

Grand-average power spectrum. In order to compute the power spectrum, spectral estimates corresponding to congruent and incongruent trials for each cue-type condition were combined, resulting in three different data sets for each cue-type (words in L1, words in L2, natural sounds). Note that subjects were not aware of incongruency in the prestimulus time window, thus the time-frequency representations at this stage should be indistinguishable for the congruent and incongruent conditions. Then, spectral estimates were averaged over trials, participants, channels and cue-type conditions, resulting in a single value for each of the 30 frequency bins (i.e., the grand-average power spectrum). A peak finding algorithm was used to identify spectral peaks as local maxima in the grand-averaged power spectrum. Two peaks, one at 10 Hz and one at 18 Hz emerged from this analysis (Fig. 2A). Based on these peaks, frequencies of interest (FOI) were obtained as the average of the frequency peaks ±1 Hz: that is, 9-11 Hz and 17-19 Hz respectively (Fig. 2B). We refer to these band estimates as the alpha and beta band power.

Figure 2 around here

Prestimulus spectral differences between cues. Spectral estimates for each cue-type (words in L1, words in L2, natural sounds) were averaged over trials. To reduce individual differences in overall EEG power, normalization was applied by converting the time-frequency power for each condition into percent signal change relative to the average power over all three conditions and channels, as performed by Bogaerts et al., (2020). This procedure removes individual differences in signal power, without distorting the relative magnitudes of the conditions, i.e. it functions as a baseline correction, when an appropriate baseline interval is not available. In order to test whether time-frequency representations in the prestimulus time-window differed across cue types, a non-parametric approach was selected (Maris and Oostenveld, 2007). For each FOI, we implemented a cluster-based permutation test based on a dependent sample F-test with the spectral data for each type of cue (words in L1, words in L2, natural sounds) as the dependent variable. This approach is equivalent to a one-way ANOVA but allows to account for the spatial correlation between electrodes (i.e., no a priori region of interest needs to be defined). The minimum number of neighboring electrodes required for a sample to be included in the clustering algorithm was set at 2. The cluster threshold F-value (or t-value) was set at an alpha value at the 85th percentile of their respective distributions. Note that this parameter does not impact the false alarm rate of the test. Rather, it sets a cluster threshold for determining when a sample should be considered as a candidate member of a cluster. Small cluster thresholds usually favor the detection of highly localized clusters with large effect size, while larger cluster thresholds favor clusters with large spatio-temporal extent, and more diffusion of the effect (Maris and Oostenveld, 2007). Because alpha and beta rhythms usually emerge at the network level, we selected a relatively large cluster threshold, i.e. capturing what appears to be quite a globally distributed effect. The number of permutations for the randomization procedure was set at 100000. The critical alpha-level to control the false alarm rate was the standard $\alpha = 0.05$. All resulting p-values were Bonferroni corrected for the number of FOIs. For each FOI, one significant cluster was detected. In order to assess the directionality of the effect, post-hoc non-parametric pairwise comparisons were applied. Specifically, power values for each cue-type condition were averaged over all electrodes belonging to the significant cluster and compared pairwise using paired t-tests. The alpha-level

for the three post-hoc t-tests was Bonferroni corrected for the number of comparisons. This procedure was applied to each FOI separately.

For both the alpha and beta band, post-hoc t-tests revealed that brain data elicited by symbolic cues (words in L1 and L2) come from a similar probability distribution, while both significantly differed from brain activity elicited by natural sounds. This motivated us to pool the data from the spoken word conditions together and contrast this average with that from the natural sound condition using a dependent sample cluster-based t-test (using the same parameters as for the test based on the F-statistic). One significant cluster for each FOI emerged from this comparison: one including 23 (alpha) and 21 (beta) out of 27 electrodes respectively. The power over these clusters served for the analysis of brain-behavior correlation.

Brain-behavior correlation. In order to investigate the link between prestimulus brain rhythms and behavior, correlation analyses were performed over participants. For the correlation analyses reported below, only the trials belonging to the congruent condition were selected; i.e., those trials where the auditory cue matched the object picture. The same analysis pipeline was applied for each FOI.

Trials were averaged, providing an alpha, beta and RT for each participant and condition. To ensure the correlation was not driven by differences between conditions, participants' values were z-scored within conditions. The three conditions were then averaged, providing an alpha-RT and beta-RT pair for each participant. The Spearman correlation was then computed for these pairs for each FOI. To assess the statistical properties of the alpha and beta correlations, we bootstrapped the data over participants. We performed this 100000 times, generating a distribution of bootstrap values. Following Efron and Tibshirani (1986), we computed the percentile bootstrap 95% confidence intervals, and used this distribution to perform statistical tests to determine the difference between the observed correlation coefficients and zero. We finally conducted a two-sample bootstrap test to evaluate the difference between the alpha-RT and beta-RT correlation coefficients (Efron and Tibshirani, 1986).

Results

Effect of cues on visual object recognition. We first analyzed the accuracy. Overall, accuracy was high (98%) and similarly distributed across the three conditions (words in L1 = 98%; words in L2 = 99%, natural sounds = 97%). Participants were clearly at ceiling here; thus, we focused on the analysis of the reaction times. Analysis of reaction time responses showed a main effect of Cue-Type ($\chi^2(2) = 31.9500$, $p < 0.001$) (Fig. 3). This was subsequently unpacked via post-hoc comparisons. Pairwise comparisons using estimated marginal means showed that object images preceded by symbolic cues in both L1 and L2 were identified faster compared to images preceded by natural sounds (words in L1 – natural sounds: $\Delta = -0.08$, SE = 0.01, $p < 0.001$; natural sounds – words in L2: $\Delta = 0.06$, SE = 0.01, $p < 0.001$). On the other hand, the pairwise effect between words in L1 and words in L2 did not reach the significance threshold (words in L1 – words in L2: $\Delta = -0.02$, SE = 0.01, $p = 0.06$). As in previous studies, we also observed a main effect Congruency ($\chi^2(1) = 7.0329$, $p < 0.01$), with matching cue-pictures pairs leading to faster responses compared to mismatching pairs. No reliable Cue-Type by Congruency interaction was detected ($\chi^2(2) = 1.5310$, $p = 0.46$).

Figure 3 around here

Effect of cues on prestimulus alpha rhythms. Differences between spectral power elicited by the three cue-type conditions were assessed using a cluster-based F-test for alpha and beta FOIs separately, focusing on the prestimulus interval. From the analysis of the alpha rhythm, one significant cluster was detected ($p < 0.01$, Bonferroni-corrected for the two FOIs) including several electrodes across the entire scalp (Fig. 4A). In order to assess the directionality of the effect, spectral power for each type of cue was averaged over all the electrodes belonging to the significant cluster and compared pairwise via t-tests. Pairwise comparisons showed that both words in L1 and L2 led to increased alpha power compared to natural sounds ($t(21) = 4.57$, $p < 0.001$ Bonferroni-corrected; $t(21) = 5.48$, $p < 0.001$ Bonferroni-corrected, respectively) (Fig. 4A). No significant difference was detected between words in L1 and L2 ($t(21) = -1.70$, $p = 1$ Bonferroni-corrected).

Figure 4 around here

Effect of cues on prestimulus beta rhythms. Analysis of the beta band revealed a pattern of results similar to the one that emerged from the analysis of the alpha rhythm. The cluster-based F-test detected one cluster ($p < 0.01$, Bonferroni-corrected for number of FOIs) (Fig. 4B). Spectral power in the beta frequency range was averaged over the electrodes of the significant cluster for each type of cue separately and compared pairwise via t-tests. Beta power was larger when images were preceded by words in L1 and L2 compared to natural sounds ($t(21) = 2.68$, $p = 0.04$ Bonferroni-corrected; $t(21) = 4.68$, $p < 0.001$ Bonferroni-corrected, respectively), while no significant difference emerged when comparing words in L1 and L2 ($t(21) = -1.67$, $p = 0.33$ Bonferroni-corrected) (see Fig. 3B).

Relation between prestimulus alpha/beta rhythms and visual object recognition. The results reported so far point to a possible role of alpha and beta rhythms in supporting the label-advantage in object recognition. We further explored the relation between prestimulus alpha/beta oscillations and visual object recognition by correlating prestimulus spectral power and reaction times across participants.

Individual estimates for power and reaction times were correlated using the Spearman rank correlation. This method was selected because reaction time and beta power data significantly deviated from normality, as emerged from a Shapiro-Wilk normality test (reaction time: $W = 0.88$, $p = 0.01$; alpha power: $W = 0.92$, $p = 0.09$; beta power: $W = 0.89$, $p = 0.02$). We observed that both prestimulus alpha and beta power had a relation with reaction time performance (Fig. 5A). Yet, the directionality of the effect was opposite: alpha estimates were negatively correlated with reaction time performance in object recognition (Spearman's rho = -0.34, $p = 0.13$), while the relation between beta power and reaction time was positive (Spearman's rho = 0.33, $p = 0.13$). Though these correlations are not significant, the bootstrap percentile confidence intervals (Fig. 5B) provide marginal evidence that supports that the alpha correlation is less than zero, as shown by the 95% confidence interval only minimally exceeding zero (zero lies at the 93rd-percentile of the alpha bootstrap distribution). This can be

expressed as a p-value using the bootstrap distribution to test the one-tailed hypothesis that the alpha-RT correlation is less than 0 ($p = 0.067$), which again provides marginal evidence that the alpha-RT correlation is negative. Similarly, the bootstrap analysis shows evidence supporting the beta-RT relationship to be positive, with zero appearing at the 4.9th-percentile, which lies close to the border of the 95% confidence interval (Fig. 5B), and showing a significant difference from zero using the one-tailed hypothesis test ($p = 0.049$). Building on this evidence for opposing effects of alpha and beta power on RT, we explicitly tested that these correlations were in fact different. Figure 4B shows that the observed rho values for the alpha-RT and beta-RT correlations fall outside each others 95% confidence intervals. A two-tailed two sample bootstrap test was applied, which revealed that indeed the alpha-RT and beta-RT correlations are significantly different ($p = 0.012$), thus supporting the interpretation that these oscillations affect behavior differentially.

figure 5 around here

Discussion

Spoken words are known to have facilitatory effects on visual object recognition, but the neural mechanism underlying such facilitation is not well established. On one account, language does not influence perception itself but only later processes such as categorical decision making. On another account, language would bias perception at early sensory stages; specifically, via the amplification of category-specific priors in early sensory regions. A prediction of this latter model is that top-down priors evoke changes in occipital areas before the presentation of visual stimuli. In the present study, we tested the hypothesis that neural oscillations can serve as a mechanism to carry language-driven priors about incoming object categories.

To test this hypothesis, we used EEG to measure prestimulus brain activity and characterize the oscillatory dynamics underlying the label-advantage in object recognition. We reasoned that, if objects are recognized faster because spoken words provide more refined categorical priors than natural sounds, then these cues should

differentially modulate prestimulus oscillatory activity in the alpha/beta bands; and we should expect such an oscillatory index to be linked to object recognition performance.

Our results provide evidence that language affects visual perception by biasing prestimulus activity towards the identification of incoming input. We first replicated the previously reported label-advantage and showed that this behavioral effect persisted even when words were presented in a second language, indicating that the label-advantage relies on inherent properties of verbal symbols to deploy precise categorical information. Importantly, the reported behavioral advantage for spoken words was associated with an increase in the posterior alpha and beta rhythms in the time interval between the offset of the cue and the onset of the target object. Correlation analysis revealed that both rhythms contribute to visual object recognition performance. Yet, they appeared to operate in orthogonal directions: object recognition performance improved when alpha power increased, but decreased with the enhancement of beta rhythms. This finding suggests a division of labor between alpha and beta rhythms in orchestrating language-mediated top-down guidance of visual behavior.

Enhancement of alpha oscillations in occipital regions have been largely reported when top-down knowledge is directed by a cue towards a specific feature or direction (Worden et al., 2000; Snyder and Foxe, 2010). Two non-inclusive theoretical interpretations have been advanced to explain this effect. One prominent view is that enhanced alpha power reflects states of inhibition and filtering of task irrelevant information (Jensen and Mazaheri, 2010; Klimesch et al., 2007). More recent proposals however ascribe to neural alpha synchronization a large variety of roles in top-down processing (Palva and Palva 2007; Klimesch 2012; van Kerkoerle et al. 2014). Enhancement of alpha waves in task-relevant regions have been suggested to have excitatory effects reflecting selective amplification of neural representations of object categories (Mo et al., 2011). For instance, M/EEG studies have reported that alpha power increases in grapheme-processing regions as a function of predictability about the identity of letters (Mayer el al., 2016) or in the posterior cortex when meaningful hints precede the discrimination of ambiguous images (Samaha et al., 2018). Our results are in line with this account, and suggests that alpha oscillations can serve as a mechanism to carry language-generated representations about the structure of visual objects. The functional role of prestimulus alpha waves in carrying

object representations is also supported by the negative correlation between individual alpha power and reaction times, showing that participants with higher alpha power were overall faster in recognizing visual objects. This interpretation fits well with the view that alpha waves play a key role in mediating the access to semantic knowledge systems (Klimesch 2012). Under this account, processing verbal symbols may help to form more precise object-representations than natural sounds against which the incoming visual input can be compared, thus resulting in a facilitation in subsequent object recognition.

A novel aspect in our results in contrast to previous similar studies concerns the cue-related differences in the beta rhythm between spoken words and natural sounds. Recent proposals suggest that beta oscillatory activity reflects endogenously driven transitions from latent to active cortical representations of objects categories (Spitzer and Haegens; 2017), as well as the binding of neurocognitive network elements underling a given neural representation (Bressler and Richter, 2016). We speculate that the difference in beta modulations for spoken words vs. natural sounds may reflect a difference in the content of the activated states – and more importantly, in the amount of retrieved conceptual dimensions, e.g. the size of the neurocognitive network state, or load (Bressler and Tognoli, 2006). Behavioral and eye tracking experiments have indeed showed that spoken words activate a rich network of concepts during lexical processing (e.g., Huettig and Altmann, 2005). As a consequence, processing words might lead to the retrieval of knowledge dimensions far beyond purely sensory features of objects, such as conceptual, grammatical and lexical information. This is partially in line with human and monkey studies showing that beta synchronization over parietal and frontal regions carry supramodal information about object categories (Antzoulatos and Miller, 2014, 2016; Wutz et al., 2018). On this account, the positive correlation between beta power and reaction time response may reflect the fact that activating a large space of possibilities is detrimental to successfully performing the current task, as it requires primarily the potentiation of visual features for faster recognition. Overall, these results suggest a division of labor between alpha and beta rhythms in top-down signaling during language-mediated visual object recognition, where alpha rhythms might function to amplify neural

representations of object categories, while beta-frequency synchronization may maintain the neurocognitive network states elicited by the auditory cue.

The present findings inform the broad debate on whether language shapes perception at early or late stages of perceptual processing. Evidence for the former account comes primarily from EEG studies showing that language affect visual processes by modulating early ERP components such as the P1 (Boutonnet and Lupyan 2015; Noorman et al., 2018) and the N170 (Landau et al., 2010). However, studies focusing on post-stimulus activity can also be coherent with a later semantic or decision-making account. Indeed, post-stimulus differences, even if very early, could still emerge from rapid feed-forward integration of visual and linguistic information (Thierry et al., 2009). The reported prestimulus modulations of alpha waves in occipital regions thus provide complementary support for the idea that linguistic influences on perception have a perceptual locus, and shed light on a candidate mechanism underling such influences: the amplification of sensory priors in occipital regions via alpha-band synchronization.

Finally, a novelty in our study compared with previous categorization studies concerns the inclusion of words in L2 as an auditory cue. Our participants were indeed highly proficient Basque-Spanish bilinguals, with comparable levels of proficiency for the two languages, but with the L2 acquired later in development. The effect of top-down processing in bilinguals is currently debated, and largely dependent on factors like proficiency (Kaan, 2014; Hopp, 2013) and age of acquisition (Molinaro et al., 2017). Concerning semantic processing, despite that it is commonly believed that bilinguals access a common semantic system in both languages (e.g., Caramazza and Brones, 1980), recent studies have suggested that top-down processing may be reduced in a second language because of reduced access to perceptual memory resources (e.g., Hayakawa and Keysar, 2018), which are known to play an important role in the generation of visual expectation (Hindy et al., 2016). The reported comparable behavioral and neural responses on the effect of words in L1 and L2 on visual object recognition are in line with the idea that both languages access common conceptual representations, and deploy top-down guidance to the visual system in a similar manner.

On the contrary, the divergent effects of words and natural sounds challenges the hypothesis that these types of cues access a common nonverbal conceptual system. It is relevant to consider why cuing an object with a word results in enhanced visual recognition compared to a natural sound. Symbols have been proposed to be extremely effective in compressing semantic information in a format that transcends within-category differences, thus leading to the amplification of those features which are relevant for distinguishing between exemplars of different categories. On the contrary, natural sounds are inevitably linked to their sources (e.g., the barking of a dog may trigger the representation of a specific exemplar of dog), thus being less effective at cuing a categorical state (Edmiston and Lupyan, 2015). Interestingly, ascribing labels to experience have been shown to also enhance other cognitive functions, such as the retention of items in visual working memory (Souza and Skóra, 2017), learning of novel categories (Lupyan et al., 2007), perceptual categorization across sensory modalities (Miller et al., 2018). These findings indicate that language acts as a powerful tool for compressing information, facilitating different operations important to a multitude of human cognitive processes (Clark 2012).

References

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.

Antzoulatos, E. G., & Miller, E. K. (2014). Increases in functional connectivity between prefrontal cortex and striatum during category learning. Neuron, 83(1), 216-225.

Antzoulatos, E. G., & Miller, E. K. (2016). Synchronous beta rhythms of frontoparietal networks support only behaviorally relevant representations. Elife, 5, e17822.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. Trends in cognitive sciences, 16(7), 390-398.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Bogaerts, L., Richter, C. G., Landau, A. N., & Frost, R. (2020). Beta-band activity is a signature of statistical learning. Journal of Neuroscience, 40(39), 7523-7530.

Bosman, C. A., Schoffelen, J. M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., ... & Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron, 75(5), 875-888.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: a label advantage in object recognition. Journal of Neuroscience, 35(25), 9329-9335.

Bressler, S. L., & Tognoli, E. (2006). Operational principles of neurocognitive networks. International journal of psychophysiology, 60(2), 139-148.

Bressler, S. L., & Richter, C. G. (2015). Interareal oscillatory synchronization in top-down neocortical processing. Current opinion in neurobiology, 31, 62-66.

Clark, A., & Toribio, J. (2012). Magic words: how language augments hum an computation. In Language and Meaning in Cognitive Science (pp. 33-51). Routledge.

Caramazza, A., & Brones, I. (1980). Semantic classification by bilinguals. Canadian Journal of Psychology/Revue canadienne de psychologie, 34(1), 77.

Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. Proceedings of the National Academy of Sciences, 107(37), 16396-16400.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. Cognition, 143, 93-100.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical science, 54-75.

Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. Cognition, 173, 8-15.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nature neuroscience, 19(5), 665-667.

Huettig, F., & Altmann, G. T. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. Visual Cognition, 15(8), 985-1018.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. Second Language Research, 29(1), 33-56.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. Psicológica, 35(1), 49-66.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Frontiers in human neuroscience, 4, 186.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different?. Linguistic Approaches to Bilingualism, 4(2), 257-282.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition–timing hypothesis. Brain research reviews, 53(1), 63-88.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. Trends in cognitive sciences, 16(12), 606-617.

Landau, A. N., Aziz-Zadeh, L., & Ivry, R. B. (2010). The influence of language on perception: listening to sentences about faces affects the perception of faces. Journal of Neuroscience, 30(45), 15254-15261.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. Behavior research methods, 44(2), 325-343.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. R package version, 1(1), 3.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. Psychological science, 18(12), 1077-1083.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. Journal of Experimental Psychology: General, 141(1), 170.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. Proceedings of the National Academy of Sciences, 110(35), 14196-14201.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of neuroscience methods, 164(1), 177-190.

Mayer, A., Schwiedrzik, C. M., Wibral, M., Singer, W., & Melloni, L. (2015). Expecting to see a letter: alpha oscillations as carriers of top-down sensory predictions. Cerebral Cortex, 26(7), 3146-3160.

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J. M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. Neuron, 89(2), 384-397.

Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. Cognition, 171, 172-179.

Mo, J., Schroeder, C. E., & Ding, M. (2011). Attentional modulation of alpha oscillations in macaque inferotemporal cortex. Journal of Neuroscience, 31(3), 878-882.

Molinaro, N., Giannelli, F., Caffarra, S., & Martin, C. (2017). Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals. Cognition, 164, 61-73.

Noorman, S., Neville, D. A., & Simanova, I. (2018). Words affect visual perception by activating object shape representations. Scientific reports, 8(1), 1-10.

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.

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. Journal of Experimental Psychology: Human Perception and Performance, 43(3), 499.

Palva, S., & Palva, J. M. (2007). New vistas for α-frequency band oscillations. Trends in neurosciences, 30(4), 150-158.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. Journal of neuroscience methods, 162(1-2), 8-13.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. Perception, 33(2), 217-236.

Samaha, J., Boutonnet, B., Postle, B. R., & Lupyan, G. (2018). Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. Scientific reports, 8(1), 1-14.

Simanova, I., Francken, J. C., de Lange, F. P., & Bekkering, H. (2016). Linguistic priors shape categorical perception. Language, Cognition and Neuroscience, 31(1), 159-165.

Snyder, A. C., & Foxe, J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. Journal of Neuroscience, 30(11), 4024-4032.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. Cognition, 166, 277-297.

Spehlmann, R. (1965). The averaged electrical responses to diffuse and to patterned light in the human. Electroencephalography and clinical neurophysiology, 19(6), 560-569.

Spitzer, B., & Haegens, S. (2017). Beyond the status quo: a role for beta oscillations in endogenous content (re) activation. eneuro, 4(4).

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. Proceedings of the National Academy of Sciences, 106(11), 4567-4570.

Van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M. A., Poort, J., Van Der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proceedings of the National Academy of Sciences, 111(40), 14332-14341.

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific α-bank electroencephalography increases over occipital cortex. Journal of Neuroscience, 20(6), RC63-RC63.

Wutz, A., Loonis, R., Roy, J. E., Donoghue, J. A., & Miller, E. K. (2018). Different levels of category abstraction by different dynamics in different prefrontal areas. Neuron, 97(3), 716-726.

Figure captions

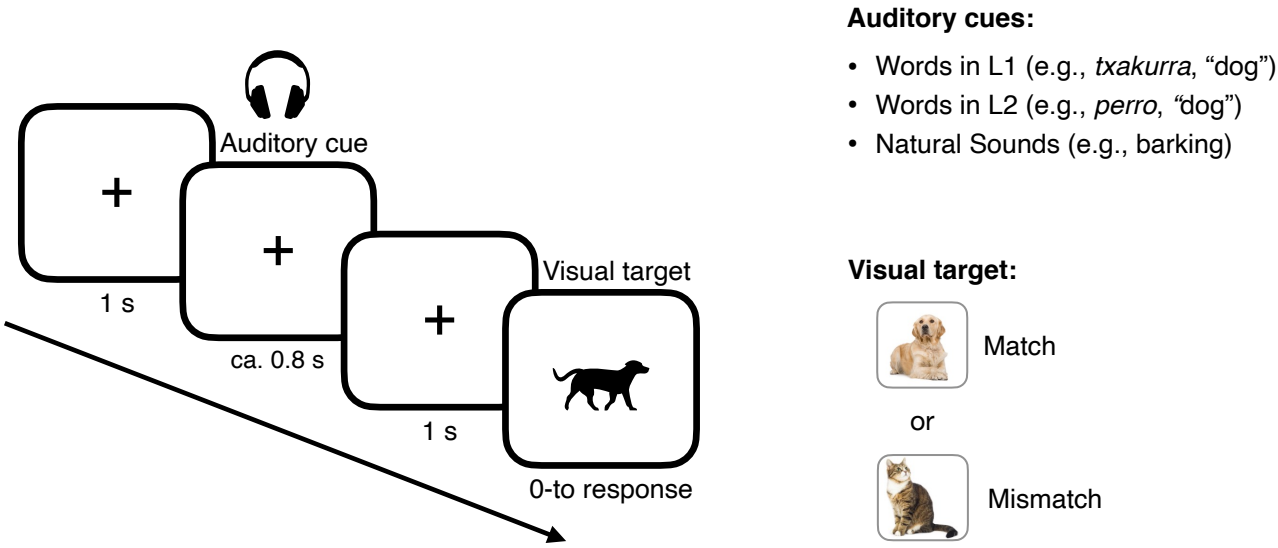

Figure 1. Illustration of the design and trial structure with an example for each possible auditory cue (words in L1, words in L2, natural sounds) and target object condition (Match, Mismatch).

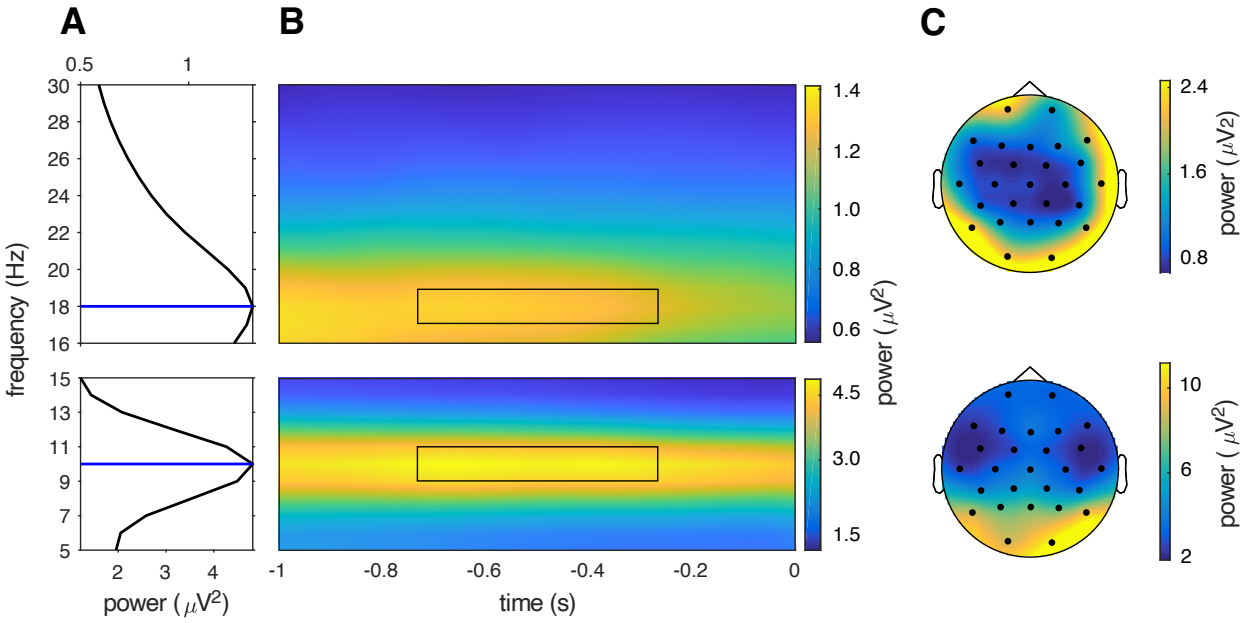

Figure 2. A) Alpha and beta peaks in the grand-average raw power spectrum of all epochs across conditions, during the -0.75 -0.25 pre-target time-interval. The blues lines indicate the power peak as local maxima. B) Time-frequency representation of grand-averaged data for the alpha and beta-band, in the 1 s time window between the offset of the auditory cue and the onset of the image. The black rectangle denotes the time-frequency interval selected for the statistical analysis. C) Topography of the time-frequency interval of interest.

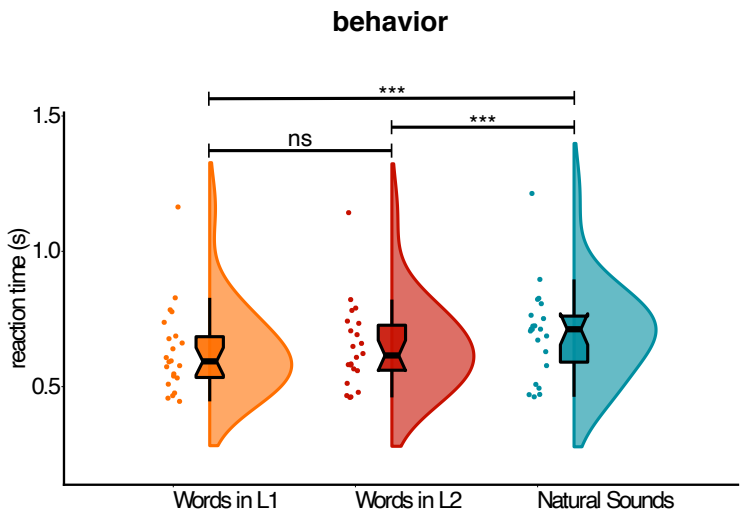

Figure 3. Mean reaction times (correct trials only) showing the main effect of cue-type on visual object recognition performance. Raincloud plots show probability density. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data). The notches indicate the 95% confidence interval around the median. Dots reflect individual subjects.

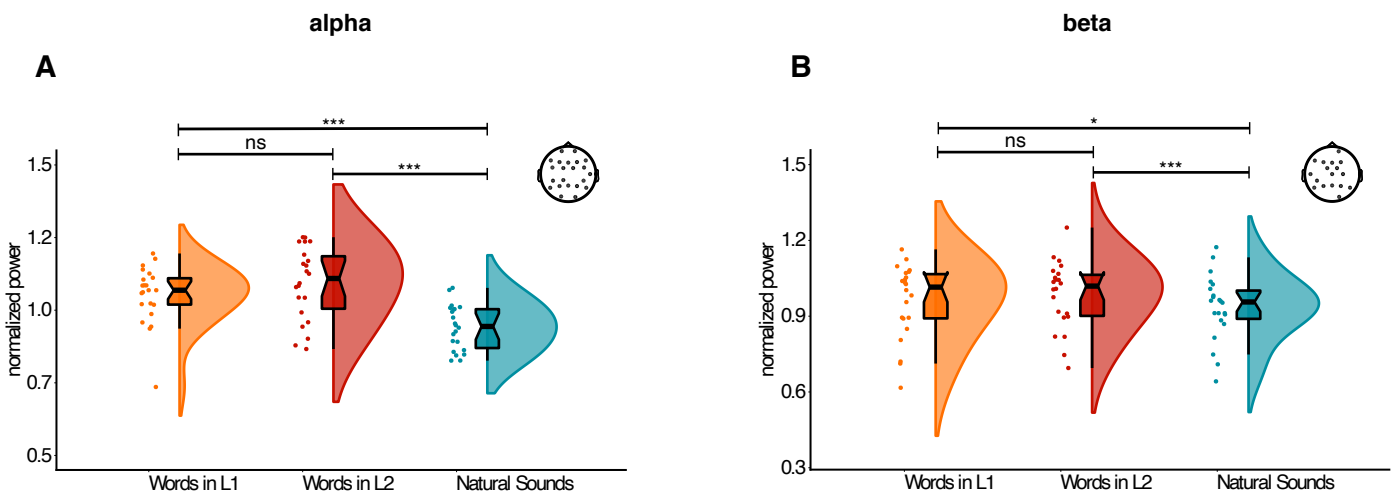

Figure 4. Effect of cues on pre-target alpha (A) and beta power (B) averaged over the electrodes belonging to the significant cluster. Conventions for the plot are the same of Figure 2. The electrodes belonging to the cluster are illustrated on the top-right of each figure.

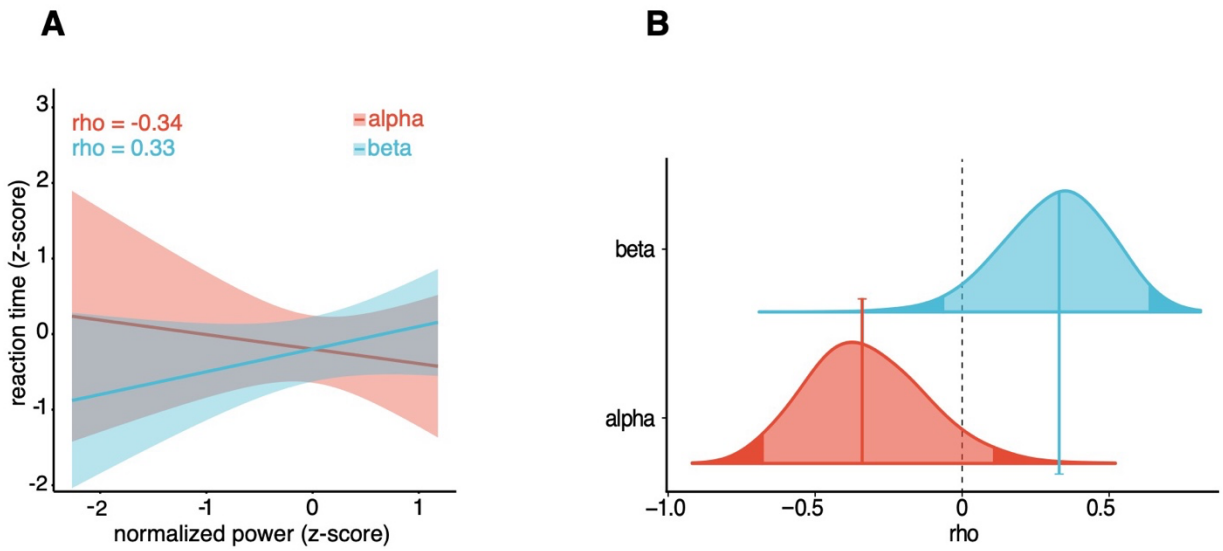

Figure 5. A) Correlation between alpha/beta power and reaction time (RT). Error bars represent 95% confidence interval. B) Bootstrap distributions of alpha/beta Spearman rho values.

Tables

Measure	**Basque**	**Spanish**
Self-evaluation (0-10)	9.04 (0.16)	9.39 (0.24)
LexTALE Basque (0-50); Spanish (0-60)	46.04 (2.67)	54.09 (4.13)
Picture naming (0-65)	64.19 (1.47)	63.38 (1.62)
Interview (0-5)	5 (0)	4.95 (1.33)

Table 1. General proficiency assessment of the participants' linguistic profile.

Conflict of Interests

The authors declare no competing financial interests.

Data sharing and code accessibility

The behavioral and EEG data reported in this paper, experimental materials, and analysis code are available via the Open Science Framework (https://osf.io/ntg9q/). This study has not been preregistered.

Acknowlegments

This research was supported by the Spanish Ministry of Economy and Competitiveness (MINECO), Agencia Estatal de Investigacion (AEI), and Fondo Europeo de Desarrollo Regional (FEDER) (PSI2015–65694-P and RTI2018-096311-B-I00 to NM) and by the Basque government (grant PI_2016_1_0014 to NM). The authors acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, through the “Severo Ochoa” Programme for Centres/Units of Excellence in R&D (SEV-2015-490). Work by Piermatteo Morucci 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. 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 also thanks to Magda Altman and Jose Pérez-Navarro for their helpful comments on the manuscript.

Authors contribution

Piermatteo Morucci: Software, Validation, Formal analysis, Resources, Data Curation, Visualization, Project administration, Writing–original draft, Writing–review &

editing. **Francesco Giannelli**: Conceptualization, Methodology, Software, Investigation, Resources, Data Curation, Project administration, Writing–review & editing. **Craig Richter**: Formal analysis, Supervision, Writing–review & editing. **Nicola Molinaro**: Supervision, Funding acquisition, Project administration, Writing–review and editing.