



EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations

João M. Correia*, Bernadette Jansma, Lars Hausfeld, Sanne Kikkert and Milene Bonte

Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht Brain Imaging Center (M-BIC), Maastricht University, Maastricht, Netherlands

Edited by:

Peter Indefrey, University of Dusseldorf, Germany

Reviewed by:

Jonas Obleser, Max Planck Institute for Human Cognitive and Brain Sciences, Germany
Dirk Koester, Bielefeld University, Germany

*Correspondence:

João M. Correia, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht Brain Imaging Center (M-BIC), Maastricht University, Oxfordlaan 55, 2nd floor, Room 014, 6229 EV Maastricht, Netherlands
e-mail: joao.correia@maastrichtuniversity.nl

Spoken word recognition and production require fast transformations between acoustic, phonological, and conceptual neural representations. Bilinguals perform these transformations in native and non-native languages, deriving unified semantic concepts from equivalent, but acoustically different words. Here we exploit this capacity of bilinguals to investigate input invariant semantic representations in the brain. We acquired EEG data while Dutch subjects, highly proficient in English listened to four monosyllabic and acoustically distinct animal words in both languages (e.g., “paard”–“horse”). Multivariate pattern analysis (MVPA) was applied to identify EEG response patterns that discriminate between individual words within one language (within-language discrimination) and generalize meaning across two languages (across-language generalization). Furthermore, employing two EEG feature selection approaches, we assessed the contribution of temporal and oscillatory EEG features to our classification results. MVPA revealed that within-language discrimination was possible in a broad time-window (~50–620 ms) after word onset probably reflecting acoustic-phonetic and semantic-conceptual differences between the words. Most interestingly, significant across-language generalization was possible around 550–600 ms, suggesting the activation of common semantic-conceptual representations from the Dutch and English nouns. Both types of classification, showed a strong contribution of oscillations below 12 Hz, indicating the importance of low frequency oscillations in the neural representation of individual words and concepts. This study demonstrates the feasibility of MVPA to decode individual spoken words from EEG responses and to assess the spectro-temporal dynamics of their language invariant semantic-conceptual representations. We discuss how this method and results could be relevant to track the neural mechanisms underlying conceptual encoding in comprehension and production.

Keywords: EEG decoding, EEG oscillations, speech perception, spoken word recognition, bilinguals, semantic representations, conceptual representation

INTRODUCTION

Speech processing is a surprisingly flexible and accurate cognitive ability that allows humans to comprehend spoken language in real-time. At the individual word level, speech processing requires a continuous mapping of complex and variable auditory input signals to words and their semantic-conceptual representations. In turn, when we speak, we start from ideas and concepts and convert these into articulatory motor programs. In multilingual environments, these transformations involve the extraction of unified semantic concepts from variable acoustic/phonological word forms in native and non-native languages. When and how the bilingual brain performs these language-invariant conceptual transformations remains essentially unknown and is a focus of the present electroencephalography (EEG) study.

EEG allows studying non-invasively and with high temporal resolution the neural dynamics of speech processing. The

temporal dynamics of EEG signals are informative of temporal order effects during speech processing. ERP (event-related potential) components at early time intervals, 100–200 ms after word onset, have been associated with phonetic/phonological processing (Dumay et al., 2001; Sanders and Neville, 2003; Bonte and Blomert, 2004; Uusvuori et al., 2008). Intermediate time intervals (200–300 ms) have been suggested to reflect early aspects of lexical access (Van den Brink et al., 2001; Hagoort et al., 2004; Salmelin, 2007; Bonte et al., 2009), followed by lexical/semantic processing in the 300–600 ms window, as indicated by ERP modulations dependent on semantic attributes of words, semantic priming and semantic context (Kutas and Hillyard, 1980; Hagoort, 2008). Spatially, this temporal signature of speech processing may reflect a spread of information from primary auditory areas to anterior temporal and frontal regions, mid-inferior and posterior temporal regions (Marinkovic et al., 2003) corresponding to the

network of brain areas observed in functional magnetic resonance imaging (fMRI) studies of speech processing (Binder et al., 2000; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). Complementary to ERP modulations, the oscillatory dynamics of EEG signals measured extracranially (Hagoort et al., 2004; Shahin et al., 2009; Doelling et al., 2014; Strauß et al., 2014) and intracranially (Luo and Poeppel, 2007; Giraud and Poeppel, 2012) have provided important insights regarding the function of underlying neural oscillations. Namely, an entrainment of theta band oscillations to the phoneme/syllable rate of speech signals, and the entrainment of gamma band oscillations to the phase of such theta band oscillations are suggested to reflect synchronization mechanisms that optimize the parsing of the speech signal into its relevant units (Lakatos et al., 2005; Giraud and Poeppel, 2012; Obleser et al., 2012; Peelle and Davis, 2012).

A challenge is to investigate how these temporal and oscillatory EEG dynamics encode the representation of specific speech units, such as individual words and concepts. Recently, methods based on machine learning comprising multivariate statistics (MVPA, multivariate pattern analysis, Formisano et al., 2008a; Haxby et al., 2011) have shown their potential to solve this challenge. MVPA of EEG signals extends traditional univariate methods by exploiting the interaction between multiple signal features (e.g., spectro-temporal features across multiple electrodes and/or time points) using classification algorithms (Chan et al., 2011b; Hausfeld et al., 2012; Herrmann et al., 2012; Brandmeyer et al., 2013). The higher sensitivity of MVPA to find information content within brain imaging signals has significantly contributed to our understanding of the brain's responses to speech and language. In fMRI studies, multi-voxel patterns across early and higher-order auditory cortex have been shown to successfully predict the (perceptual) identity of individual speech sounds and speaker's voices (Formisano et al., 2008b; Kilian-Hütten et al., 2011; Bonte et al., 2014). Furthermore, fMRI responses in inferior parietal areas have been shown to differentiate words across different semantic categories [e.g., tools and dwellings, Shinkareva et al. (2011)]. At a more fine-grained within-category level, MVPA was recently shown to accurately predict which spoken noun a bilingual listener was listening to in one language (e.g., "horse" in English) based on the fMRI response patterns to equivalent nouns in the other language (e.g., "paard" in Dutch; Correia et al., 2014). This generalization of the meaning of words across languages specifically relied on focal regions, including the left anterior temporal lobe (left-ATL), suggesting the existence of "hub" regions organizing semantic-conceptual knowledge in abstract form (Damasio et al., 1996; Scott et al., 2000; Patterson et al., 2007; Visser and Lambon Ralph, 2011; Correia et al., 2014). Although more challenging in terms of the robustness of single trial estimates, also spatially/temporally distributed EEG/MEG patterns have been observed to discriminate individual speech sounds (Hausfeld et al., 2012), and words from different perceptual and semantic categories (Simanova et al., 2010; Chan et al., 2011b; Sudre et al., 2012). Classification performances in EEG-MVPA studies on speech processing are typically low [e.g., below 0.55 in binary classification of spoken vowels, Hausfeld et al. (2012); or below 0.60 in binary classification of spoken words, Simanova et al. (2010)]. Besides the low signal-to-noise

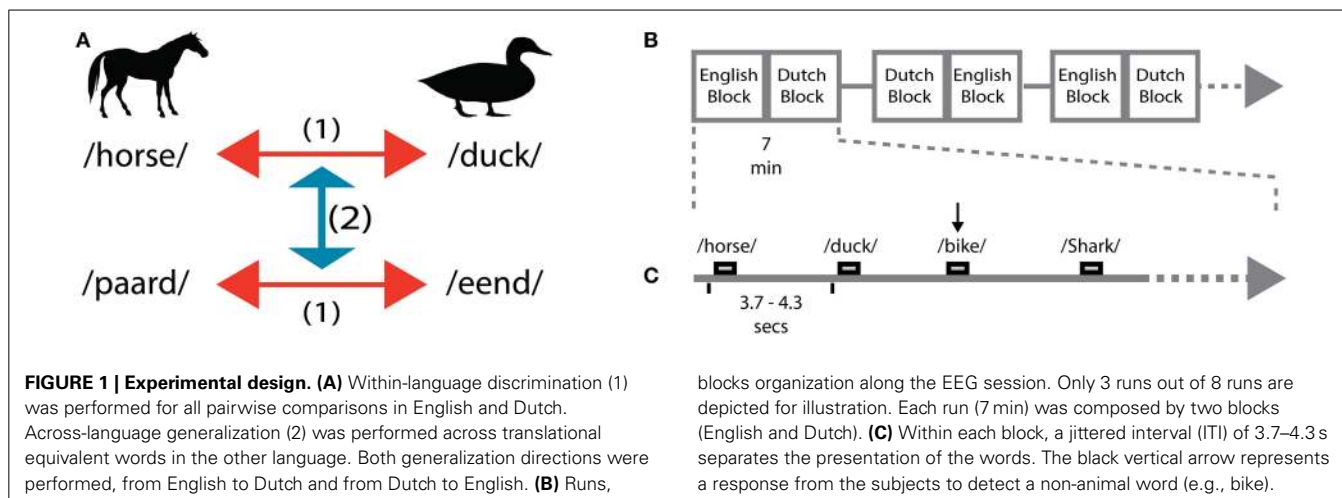
ratio of single trial EEG signals, EEG-based classification of individual words may be limited by the continuous and temporally variable processing of their phonological and semantic features (Van Petten et al., 1999). Importantly, however, multivariate approaches in EEG allow unraveling subtle differences in the neural processing of individual speech sounds that remain obscured in univariate approaches relying on average activation differences between experimental conditions.

Here, we employ MVPA to investigate spectro-temporal EEG response patterns capable of discriminating semantic-conceptual representations of words at the fine-grained level of within-category distinctions (animal nouns). To this end, we exploit the unique capacity of bilingual subjects to access semantic-conceptual information of spoken words from two languages. In separate Dutch and English blocks, we asked bilingual participants to listen to individual animal nouns (66.6% trials) and to detect non-animal target nouns (33.3% trials). The non-animal target nouns were presented as control task to ensure speech comprehension at every word presentation, but were not included in the analysis. Following supervised machine learning approaches, we trained multivariate classifiers (linear-SVM) to predict the identity of the perceived animal noun from new (untrained) samples of EEG activity (**Figure 1A**). In a first analysis we aimed to identify the EEG correlates involved in within-language word discrimination. To this end we trained classifiers to discriminate EEG responses to English (e.g., "horse" vs. "duck") and Dutch (e.g., "paard" vs. "eend") nouns. Importantly, stimuli included three exemplars of each noun, pronounced by three different female speakers, allowing for speaker-invariant word discrimination ("*within-language*"). In a second analysis we aimed to assess the EEG correlates involved in language-independent decoding of the animal nouns ("*across-language*"). Here we trained classifiers to discriminate EEG responses to words in one language (e.g., in English, "horse" vs. "duck") and tested whether this training generalizes and allows discrimination of EEG responses to the corresponding nouns in the other language (e.g., in Dutch, "paard" vs. "eend"). Importantly, all words were acoustically-phonetically distinct both within and across languages. Based on this approach, we aimed to investigate whether language-independent representations are detectable in the EEG responses to individual spoken words. In particular, this approach allowed us to extract critical time windows and frequency ranges within the EEG relevant to semantic-conceptual encoding.

METHODS

PARTICIPANTS

Sixteen native Dutch (L1) participants proficient in English (L2) took part in the study (8 males and 8 females, right-handed, mean age = 28.9 SD = 3.4). The participants were undergraduate or post-graduate students of Maastricht University studying or working in an English speaking environment. All participants reported normal hearing abilities and were neurologically healthy. English proficiency was assessed with the LexTALE test, a vocabulary test including 40 frequent English words and 20 non-words (Lemhöfer and Broersma, 2012). The mean test score was 89.6% correct (SD = 11.2%). This score is well above the average score (70.7%) of a large group of Dutch and Korean advanced learners



of English performing the same test (Lemhöfer and Broersma, 2012). For comparison reasons, participants also conducted the Dutch version of the vocabulary test. The mean Dutch proficiency score was 94.1% (SD = 3.3). The study was approved by the Ethical Committee of the Faculty of Psychology and Neuroscience at the University of Maastricht, The Netherlands.

STIMULI

Stimuli consisted of Dutch and English spoken words representing four different animals (English: “Bull,” “Duck,” “Horse,” and “Shark,” and the Dutch equivalents: “Stier,” “Eend,” “Paard,” and “Haai”) and six inanimate object words (English: “Bike,” “Coat,” “Dress,” “Road,” “Suit,” and “Town”; and the Dutch equivalents: “Fiets,” “Jas,” “Jurk,” “Weg,” “Pak,” and “Stad”). All animal nouns were monosyllabic and acoustically/phonetically distinct from each other both within and across languages. Phonetic distance between word pairs was quantified using the Levenshtein distance, which gives the number of phoneme insertions, deletions and/or substitutions required to change one word into the other, divided by the number of phonemes of the longest word (Levenshtein, 1965). On a scale from 0 (no changes) to 1 (maximum number of changes), the mean (SD) Levenshtein distances corresponded to 0.83 (0.15) for Dutch word pairs, 0.93 (0.12) for English word pairs and 1.00 (0.00) for English-Dutch word pairs. Furthermore, all animal nouns had an early age of acquisition in Dutch (mean = 5.28 years SD = 0.98; De Moor et al., 2000) and a medium-high frequency of use expressed on a logarithmic scale in counts per million tokens in Dutch (mean = 1.29 SD = 0.71) and in English [mean = 1.50 SD = 0.42; Celex database, Baayen et al. (1995)]. To add acoustic variability and allow for speaker-invariant MVPA analysis, the words were spoken by three female native Dutch speakers with good English pronunciation. Stimuli were recorded in a sound proof chamber at a sampling rate of 44.1 kHz (16 bit resolution). Post-processing of the recorded stimuli was performed in PRAAT software (Boersma and Weenink, 2013) and included band-pass filtering (80–10,500 Hz), manual removal of acoustic transients (clicks), length equalization, removal of sharp onsets and offsets using 30 ms ramp envelopes, and amplitude equalization (average

RMS). Stimulus length was equated to 600 ms (original range: 560–640 ms) using PSOLA (75–400 Hz as extrema of the F0 contour). We carefully checked the stimuli for possible alterations in F0 after length equation and did not find any detectable changes. We assured that the produced stimuli were unambiguously comprehended by the participants during the stimuli familiarization phase prior to the experiment.

EXPERIMENTAL PROCEDURES

The experimental session was organized in 8 runs, each run containing 2 blocks (one Dutch and one English). Each block included 36 nouns: 24 animal nouns and 12 (33.3%) non-animal nouns. The order of English and Dutch blocks was counter-balanced across runs: odd runs started with an English block followed by a Dutch block; even runs started with a Dutch block followed by an English block (Figure 1B). Participants were instructed to actively listen to the stimuli and to press a button (with the left index finger) whenever they heard a non-animal word. The goal of the task was to help maintaining a constant attention level throughout the experiment and to promote speech comprehension at every word presentation. All participants paid attention to the words as indicated by a mean (SD) detection accuracy of 98.3 (1.4) %. Data from non-animal trials were excluded from further analysis. The 24 animal nouns in each block corresponded to 6 repetitions of each of the 4 animal nouns. Because nouns were pronounced by 3 different speakers, each physical stimulus was repeated twice in each block. Stimulus presentation was pseudo-randomized within each block, avoiding consecutive presentations of the same words or sequences of words. Throughout the experiment, each animal noun was presented 48 times per language.

EEG ACQUISITION AND PREPROCESSING

Data were recorded with a sampling rate of 250 Hz in an electrically shielded and sound-proof room from 62 electrode positions (Easycap, Montage Number 10, 10–20 system) relative to a left mastoid reference signal. The ground electrode was placed on the Fz electrode. Impedance levels were kept below 5 k Ω . During the EEG measurement, stimuli were presented binaurally at a comfortable intensity level. According to an event-related design

(Figure 1C), the averaged inter-trial-interval between two stimuli was 4 s (jittered randomly between 3.7 s and 4.3 s). Each run took 7 min, resulting in a total EEG measurement time of 56 min. A gray fixation cross against a black background was used to keep the visual stimulation constant during the whole duration of a block. Block and run transitions were marked with written instructions. Participants were instructed to minimize eye-movements during the auditory presentation and fixate on the fixation cross.

Data preprocessing was performed using EEGLab (Delorme and Makeig, 2004) and included band-pass filtering (0.1–100 Hz) followed by epoch extraction locked to the onset of the animal nouns (−1000 to 1000 ms) and baseline correction (−1000 to 0 ms).

Removal of signal artifacts was performed in two steps. First, the data were visually inspected and epochs containing non-stereotypical artifacts including high-amplitude, high-frequency muscle noise, swallowing, and electrode cable movements, were rejected (mean 4.31 trials per subject, SD 2.36). Second, stereotypical artifacts related to eye movements, eye-blinks and heart beat artifacts were corrected with extended INFOMAX ICA (Lee et al., 1999) as implemented in EEGLAB. Because data were recorded at 62 channels, runica decomposed the data in 62 component activations per subject. These component activations were categorized as EEG activity or non-brain artifacts by visual inspection of their scalp topographies, time courses, and frequency spectra. Criteria for categorizing component activations as EEG activity included (1) a scalp topography consistent with an underlying dipolar source, (2) spectral peak(s) at typical EEG frequencies, and (3) regular responses across single trials, i.e., an EEG response should not occur in only a few trials (Delorme and Makeig, 2004). Based on these criteria, component activations representing non-brain artifacts were removed, and EEG data were reconstructed from the remaining component activations representing brain activity. The resulting ICA-pruned data sets were baseline corrected (−1000 to 0 ms) and used for further analysis.

ERP AND ERSF ANALYSIS

First, in order to validate typical EEG responses to spoken words reported in the literature, we performed univariate analyses. These were conducted in EEGLab (Delorme and Makeig, 2004) and included: (1) an ERP analysis based on the average amplitude of signal change over time with respect to baseline (−1000 to 0 ms) and (2) an ERSF (event-related spectral perturbation) analysis based on averaged power changes of all words over frequency and time with respect to baseline (−1000 to 0 ms). For the ERSF analysis we employed a Hanning taper fast fourier transform (FFT) filter from 1 to 60 Hz on a linear frequency scale with steps of 2 Hz, producing 30 filtered signals. Group statistics for the ERP and ERSF were performed at random-effects using two-sided Wilcoxon tests for each time-point vs. zero baseline and corrected for multiple comparisons using FDR (alpha = 5%).

MULTIVARIATE CLASSIFICATION ANALYSIS

Multivariate classification was employed to investigate whether specific temporal or spectrotemporal EEG signatures enable the

discrimination of words within and across languages. To this end we used a supervised machine learning algorithm (linear support vector machines, linear-SVM; Cortes and Vapnik, 1995) as implemented by the Bioinformatics Matlab toolbox (maximum number of learning iterations = 15,000). Classifications were performed to evaluate whether patterns of EEG data pertained relevant information encoding the representations of spoken words (within-language discrimination) as well as their language invariant semantic-conceptual representations (across-language generalization). All classifications were binary (i.e., chance-level is 0.5) and involved discrimination and generalization between two words. The results of these binary predictions were then averaged across all possible pair-wise classifications. Additional methodological steps encompassing the computational strategy to validate the classification results (cross-validation) and to select the EEG features used for classification (feature selection) are described below.

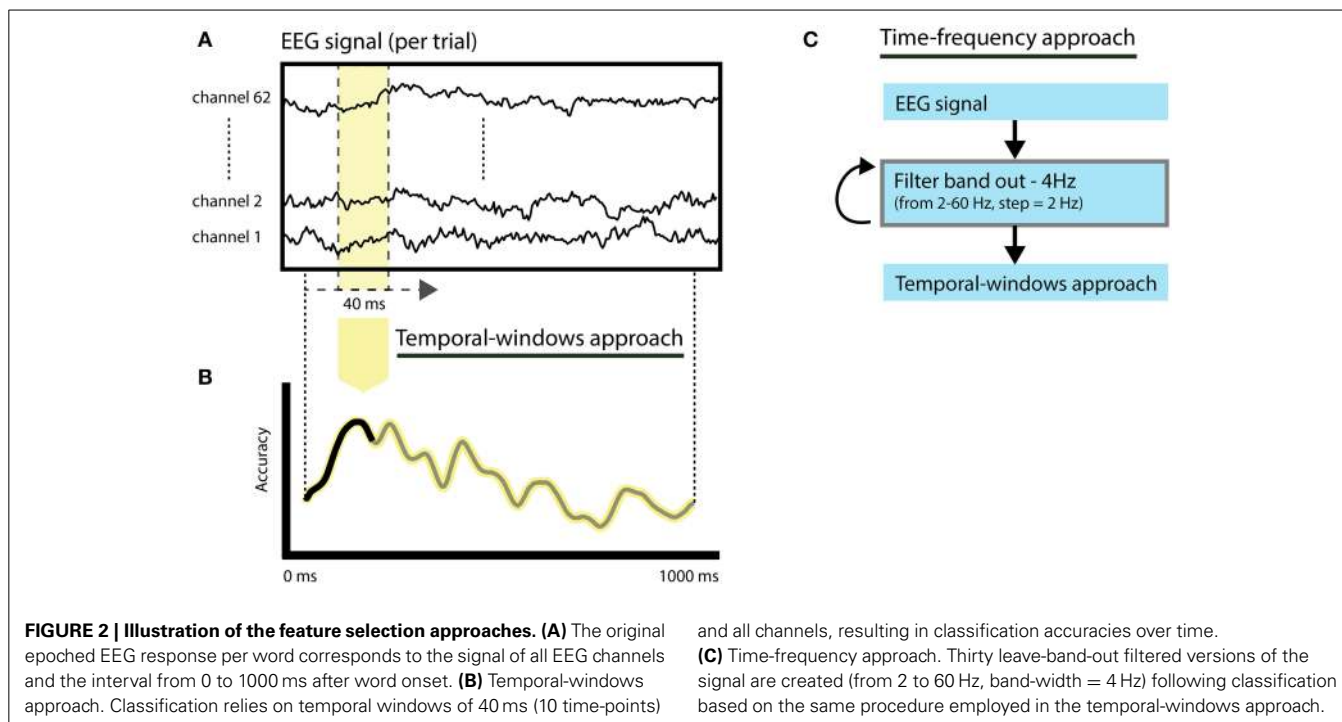
CROSS-VALIDATION APPROACHES

Cross-validation of the multivariate classification analysis served two purposes: (1) to obtain robust estimates of the discrimination accuracies; (2) to allow generalization of classes by using distinct class groupings during the training and testing phases of classification. Cross-validation for within-language word discrimination relied on speaker identity. Here, we trained a classifier to discriminate words based on samples recorded from two out of the three speakers that pronounced the words (32 trials per word) and tested whether this training was able to generalize the left-out speaker pronouncing the same words (16 trials per word). This cross-validation procedure assured word discrimination invariant to neural activations specific to acoustic-phonetic characteristics of the speakers. Cross-validation for across-language generalization of semantic concepts relied on language independent information of the words. Here, we trained a classifier to discriminate words within one language (48 trials per word) and tested whether this training generalized to the other language (48 trials per word). Hence, in across-language generalization, we aimed to isolate semantic conceptual properties of the words that were language invariant.

FEATURE SELECTION APPROACHES

Temporal-windows approach (shifting-windows + all channels)

To investigate the temporal evolution of spoken word decoding, we selected EEG response features (Figure 2A) using shifting-windows (width = 40 ms—10 time points) across all channels (Figure 2B). Restricting the EEG signal features to specific time windows permits the calculation of changes in classification accuracies over time informative of spoken word processing. Because the temporal-windows approach reduces the number of EEG features used for classification, it increases the temporal sensitivity of the classifiers to speaker and language invariant information of the spoken words due to a potentially better match between the training and testing patterns (Hausfeld et al., 2012). Additionally, it reduces the high dimensionality of the feature space, thus avoiding degraded classification performances (model overfitting; for a description, see Norman et al., 2006). The empirical null distribution was computed per subject using 200 label permutations.



Individual statistical significance ($p < 0.05$) was calculated based on deviance from permuted accuracies. Group level statistics were calculated based on the overlap of significant subjects across time intervals using a binomial test with $n = 16$ (number of subjects) and $p = 0.05$ (Darlington and Hayes, 2000; Hausfeld et al., 2012) and corrected for multiple comparisons (time windows) using FDR correction ($\alpha = 5\%$).

Time-frequency approach (filtered-band-out + shifting-windows + all channels)

To assess the importance of brain oscillations in specific frequency bands to the performance of the classifiers we employed a feature selection approach combining temporal shifting windows and filter-band-out (Figure 2C). The original epoched EEG responses (-1000 to 1000 ms) were filtered prior to classification using an FIR (finite impulse response) filter as implemented in EEGLab (Delorme and Makeig, 2004). The width of the filtered-out frequency band was set to 4 Hz, centered on frequencies from 2 up to 60 Hz and in frequency steps of 2 Hz, producing 30 filtered signals. For each of the filtered signal versions, we subsequently performed the *temporal-windows* approach to assess the importance of each frequency band over time. The importance of the left-out frequency band was quantified in terms of a change in classification performance with respect to the non-filtered signal. To prevent a modulation of time-frequency importance due to differences in the original classification accuracy, a normalization of the importance of each time-frequency bin with respect to the accuracy limits (0–1) was performed using “odds-ratio” normalization (Szumilas, 2010). Odds-ratio values above 1 indicate a reduction of classification accuracy after a specific frequency band is filtered out. This approach allowed us to investigate the contribution of each frequency band over time

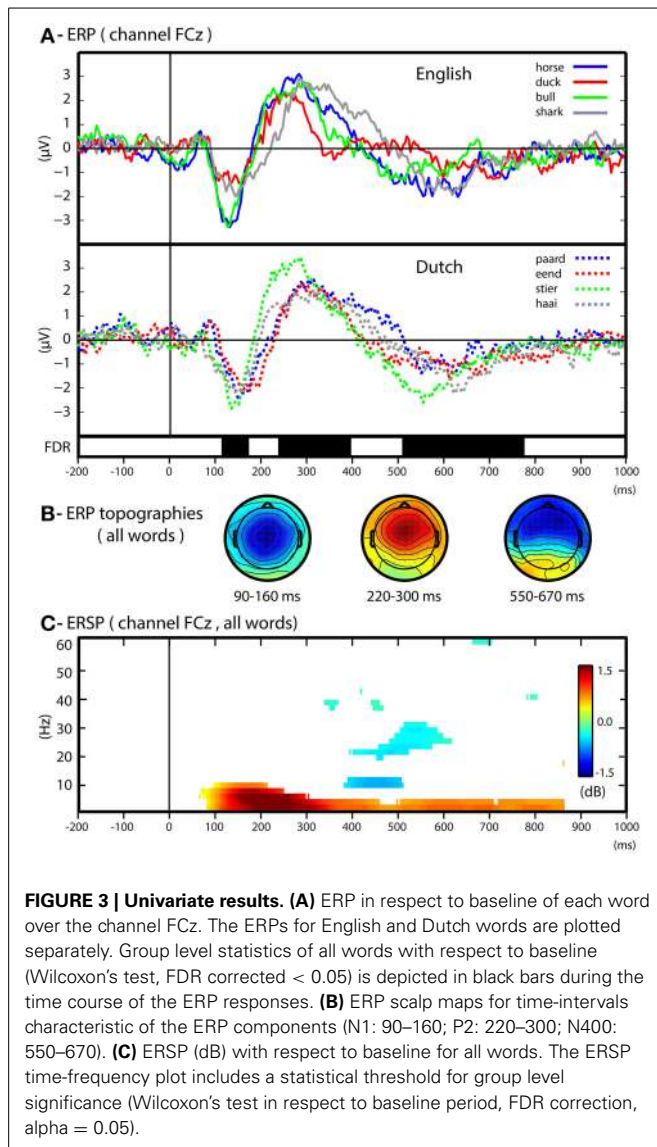
without disrupting EEG spectral interactions that may be crucial in many cognitive processes, including speech processing (Giraud and Poeppel, 2012; Henry and Obleser, 2012; Peelle and Davis, 2012). Group statistics were performed in random-effects (two-sided Wilcoxon’s test) and corrected for multiple comparisons using FDR correction ($\alpha = 5\%$).

RESULTS

ERPs AND TIME-FREQUENCY ANALYSIS

We first conducted univariate analyses of ERP and time-frequency changes relatively to stimulus baseline in order to assess the overall spectro-temporal characteristics of EEG responses evoked by the animal words. Figure 3 illustrates the averaged ERP responses elicited by the different animal words, including the expected ERP peaks (channel Fcz, Figure 3A) and their corresponding topographies (Figure 3B), in the N1 window (120–160 ms), the P2 window (230–390 ms) and the N400 window (550–800 ms). To assess univariate differences between the ERP responses we conducted all possible word-to-word contrasts within the same language (e.g., horse vs. duck), as well as all possible concept-to-concept contrasts (e.g., horse + paard vs. duck + eend). None of the possible contrasts yielded significant differences within or across participants.

The analysis of averaged power changes in different frequency bands (Figure 3C) shows an average power increase (ERS, event-related synchronization) of slow oscillations (1–10 Hz) starting 100 ms after stimulus onset, followed by a steep reduction in alpha power (ERD, event-related desynchronization) between 400 and 500 ms. At later time intervals, the ERS of slow oscillations (1–8 Hz) was maintained. These differences did not allow the systematic discrimination of individual words nor of language-independent concepts.



MULTIVARIATE ANALYSIS (MVPA)

The multivariate analysis consisted of assessing the ability of multivariate classifiers to discriminate words within the same language and across first and second language in bilingual subjects. To assess the contribution of specific EEG features used for classification we used two feature selection approaches: a *temporal-windows approach*, relying on restricted time intervals (40 ms) shifted over time and all EEG channels; and a *time-frequency approach*, relying on a combined selection of features using the *temporal-windows approach* and a moving filter-band-out procedure (4 Hz bands with an step of 2 Hz).

The *temporal-windows* feature selection approach enabled identifying specific time-intervals related to word decoding. Within-language discrimination (**Figure 4A**) was significantly possible throughout most of the time-course from ~ 50 until 620 ms after word onset. Within this broad time window, salient local maxima of accuracies were identified for the temporal windows (40 ms) around 160 ms (accuracy = 0.535), 225 ms

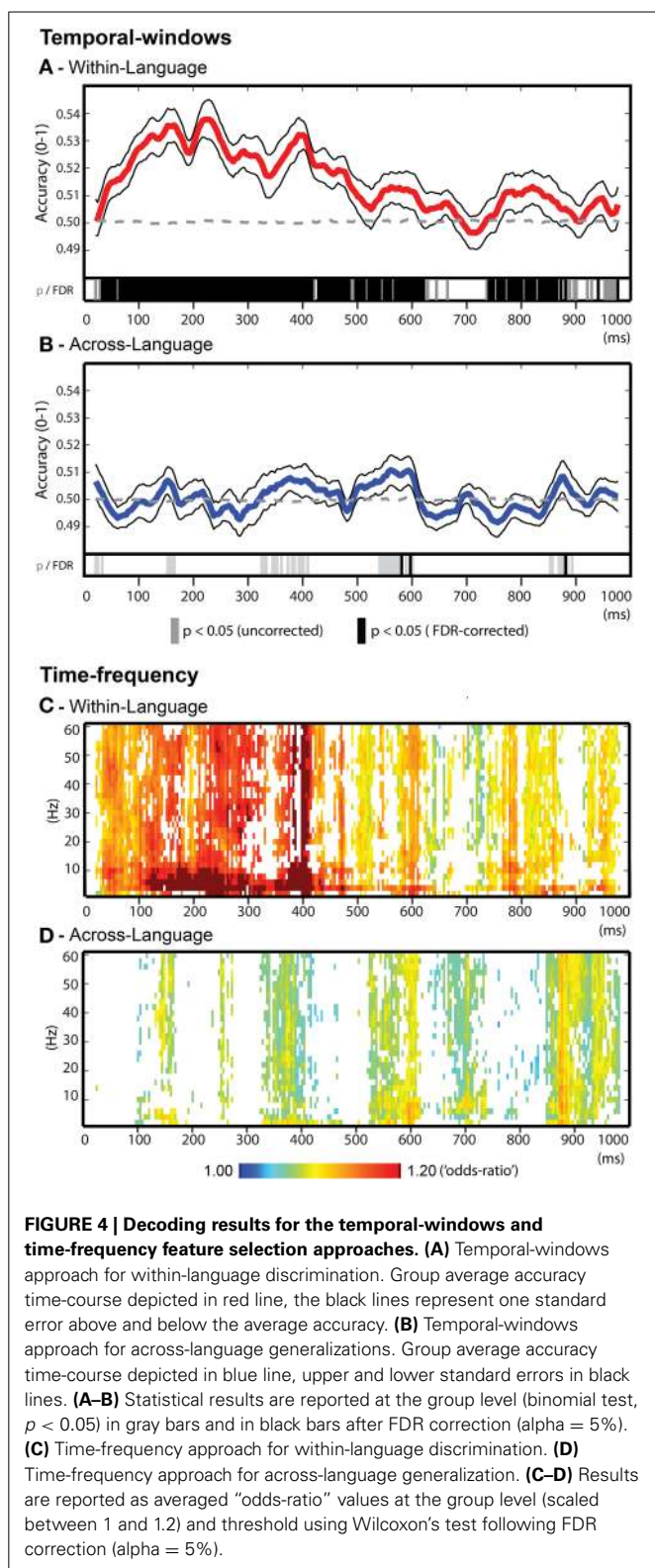
(accuracy = 0.537), 390 ms (accuracy = 0.533), 570 ms (accuracy = 0.513), and 820 ms (accuracy = 0.512). Interestingly, cross-language generalization (**Figure 4B**) led to significant classification in more restricted temporal windows with significant results between 550 and 600 ms (maximum accuracy = 0.511) and 850–900 ms (maximum accuracy = 0.508). A further time-interval showing a trend (uncorrected $p < 0.05$) for cross-language generalization capacity was observed around 400 ms (maximum accuracy = 0.507).

The *time-frequency* feature selection approach assessed the contribution of oscillatory activity in specific frequency bands to word decoding across the different time windows. For this purpose, “odds-ratio” values were computed, group averaged and thresholded for statistical significance (random-effects, FDR = 5%). Overall, the temporal profiles of the *time-frequency approach* match consistently with that of the *temporal-windows approach*, confirming that reductions in classification accuracy due to the omission of specific frequency bands occurred in time windows relevant for word decoding (**Figure 4C**). For within-language discrimination of words, reductions in classification accuracy especially occurred when omitting slow oscillations (below 12 Hz, delta, theta and alpha). For cross-language generalization (**Figure 4D**), the period around 600 ms that showed significant generalization capacity, was characterized by accuracy reductions when filtering out frequencies up to 10 Hz (delta-theta-alpha). In other time windows a contribution of slow oscillations was also observed for this analysis, although involving slower oscillations (delta/ low theta, below 6 Hz). Visual inspection of **Figures 4C–D** further suggested that besides the sensitivities for oscillations below 12 Hz, for both types of analysis smaller classification drops occurred across gamma band (above 30 Hz) as well as across broad-band oscillation profiles.

DISCUSSION

By combining EEG MVPA and an experimental design that exploits the unique capacities of bilingual listeners we identified specific time windows and oscillations enabling within-category discrimination of individual spoken words. We demonstrated within-language word decoding in a broad time-window from ~ 50 to 620 ms after word onset with a strong contribution of slow oscillations (below 12 Hz). Most importantly, we were able to isolate specific time windows, including the 550–600 ms window, in which EEG features enabled the generalization of the meaning of the words across their Dutch and English word forms. Our results demonstrate the feasibility of using MVPA to identify individual word representations based on speech evoked EEG signals. Furthermore, they indicate the advantage of feature selection approaches in assessing temporal and temporal-oscillatory EEG response features in classification.

The univariate analyses illustrate ERP and oscillatory responses typically elicited by individual spoken words (Kutas and Federmeier, 2000; Hagoort et al., 2004; Bastiaansen et al., 2008; Bonte et al., 2009; Strauß et al., 2014) indicating a progression from acoustic-phonetic to lexical-semantic processing. The ERPs to the individual words show variability as a consequence of acoustic-phonetic differences and other word-specific properties. However, these differences did not



allow the systematic discrimination of individual words nor of language-independent concepts. The prevalence of slow oscillatory activity (below 12 Hz) while subjects listened to the words indicates the crucial role of these frequencies in the

processing and comprehension of speech (Hagoort et al., 2004; Giraud and Poeppel, 2012; Strauß et al., 2014). The analysis also showed that the univariate frequency power changes were not suitable for distinguishing individual words or across-language generalization of semantic concepts.

Importantly, the multivariate analyses allowed finding neural time-course correlates of the individual words that were invariant to the acoustic-phonetic characteristics of the speakers (within-language discrimination) as well as to the language in which the meaning was presented (across-language generalization). Within-language word discrimination relied on acoustic-phonetic and semantic-conceptual differences between the nouns, but also on possible other differences reflecting their individual properties. Accordingly, within-language discrimination was possible for both approaches of feature selections employed. In the *temporal-windows* approach (**Figure 4A**), investigating the temporal evolution of classification across consecutive short time-intervals of 40 ms, classification performance was significant from ~50 until 620 ms after word onset. In accordance with the ERP literature, decoding in this broad time window may reflect a progression from phonetic-phonological processing (100–200 ms; Dumay et al., 2001; Sanders and Neville, 2003; Bonte and Blomert, 2004; Uusvuori et al., 2008) to initial lexical access (200–300 ms; Van den Brink et al., 2001; Hagoort et al., 2004; Salmelin, 2007; Bonte et al., 2009), and lexical semantic processing (300–600 ms; Kutas and Hillyard, 1980; Hagoort, 2008). These results are also consistent with previous single trial auditory word classification (Simanova et al., 2010) that showed initial prominent classification capability centered around 240 ms followed by a second less prominent capability around 480 ms after word onset.

The second multivariate analysis - across-language generalization - relied uniquely on language invariant semantic-conceptual properties of the nouns. This analysis, and especially the temporal-window approach (**Figure 4B**), revealed language invariant EEG features coding for the animal words in much more restricted time-windows including the 550–600 ms window and the 850–900 ms window at the end of the EEG epoch. ERP research has commonly associated similar time intervals with lexical-semantic processing of words across different task and sentence contexts (Kutas and Federmeier, 2000; Hagoort, 2008). Here, we indicate the potential of EEG signals to represent semantic-conceptual information of individual words independent of their acoustic-phonetic implementation or word-form. In order to isolate these input-invariant lexical-semantic representations we used animal nouns that were acoustically-phonetically distinct both within and across languages and were presented together with non-animal nouns that served as targets. In everyday speech processing, it is more difficult to disentangle input-driven vs. input-independent processes as initial lexical-semantic access is influenced by both acoustic-phonetic word form information (McClelland and Elman, 1986; Marslen-Wilson, 1987) and semantic or task context (Bonte, 2004; Obleser et al., 2004; Çukur et al., 2013), leading to early lexical and/or semantic ERP modulations around 200–300 ms (e.g., Van den Brink et al., 2001; Bonte et al., 2006; Travis et al., 2013; Strauß et al., 2014). Our approach presents a way to disentangle these aspects of comprehension. Importantly, by using words belonging to the

same semantic category—animals—we reduced the influence of larger scale semantic category differences that can also drive the decoding of individual nouns (Simanova et al., 2010; Chan et al., 2011b; Shinkareva et al., 2011).

In later time-windows, significant classification for within-language discrimination (750–900 ms) and across-language generalization (850–900 ms) may reflect effects specific to our paradigm. That is, the slow presentation of words and/or the use of a target detection task, may have led to e.g., subvocal rehearsal in working memory (Kutas and Federmeier, 2000; Baddeley, 2003; Buchsbaum et al., 2011) and/or response monitoring toward the end of the trial (Wang et al., 2012).

In bilinguals, the active translation of written words during speech production tasks has been shown to elicit ERP differences for translation direction around 400 ms after word presentation (Christoffels et al., 2013). In the current study the effect of direct translations was minimized in several ways. First, we avoided active translations from second to native language and vice-versa by separately presenting words in Dutch and English blocks and using catch trials consisting of Dutch and English non-animal words, respectively. Furthermore, we used a selection of words with relatively early age of acquisition and of medium-high frequency of use in both languages.

To further understand the EEG temporal patterns allowing classification, we employed a time-frequency feature selection approach that assessed the relative contribution of oscillatory bands. We observed a significant contribution of slow EEG oscillations (below 12 Hz) for within-language and across-language classification, which links to the synchronization of oscillatory bands observed in the ERSP analysis. Furthermore, in the time windows during which the slower oscillations most strongly influenced classification performance, results also indicated a contribution from higher, gamma band oscillations (above 30 Hz). It would be interesting to replicate this possible co-occurrence of slower and gamma band modulations in future studies with bilinguals, and, in particular to test how they relate to suggested processing of (phonemes, syllables and semantic information (Lakatos et al., 2005; Giraud and Poeppel, 2012; Peelle and Davis, 2012; Peña and Melloni, 2012).

We may hypothesize that the neural processing underlying the EEG-based translations of animal nouns occurs in a brain network that was recently identified in an fMRI study using a comparable bilingual paradigm (Correia et al., 2014). In particular, in this previous study, language-invariant classification of animal words was found to rely on focal brain regions, including the left anterior temporal lobe (left-ATL), corroborating the existence of “hub” regions organizing semantic-conceptual knowledge in abstract form. Correspondingly, recent models of conceptual knowledge (Patterson et al., 2007), brain lesion studies (Damasio et al., 1996) and neuroimaging evidence (Visser et al., 2012; Correia et al., 2014) locate a possible semantic hub within the left-ATL, integrating distributed semantic-conceptual information throughout the cortex. Furthermore, distributed neural representations of semantic information may also connect to modality specific brain regions subserving perception and action (Martin, 2007; Meyer and Damasio, 2009). Interestingly, magnetoencephalography (MEG) studies have related time windows

starting at 400 ms after spoken word onset to semantic processing in bilateral anterior temporal areas (Marinkovic et al., 2003; Chan et al., 2011a), suggesting a putative link between the present finding of language-independent word decoding in the 550–600 ms time window and processing in these brain regions. At present, this spatial-temporal association remains speculative, but similar classification paradigms using simultaneous fMRI and EEG recordings (De Martino et al., 2011) may allow investigating the joint spatio-temporal representation of spoken words. Furthermore, earlier indications of semantic/conceptual representations of our words are observed in a spread time window between 320 and 420 ms after word onset (uncorrected $p < 0.05$). These and possibly even earlier semantic activations elicited by the individual animal words may be more difficult to detect due to variability in the exact timing of these initial activations.

Overall, our results show the benefit of EEG-based MPVA to investigate the representation of semantic concepts independently of the input language and more generally of individual spoken words independently of the speaker. Although the obtained accuracies are relatively low, they demonstrate the sensitivity of multivariate classification to distinguish subtle representations extracted from single-trial EEG responses that may not be present in the averaged EEG signal across multiple trials (Makeig et al., 2002; Hausfeld et al., 2012). Furthermore, our results show the potential of feature selection approaches based on moving temporal windows to highlight time windows associated with the neural processing of specific characteristics of speech and language (e.g., language independent semantic processing, see also Simanova et al., 2010; Chan et al., 2011b; Hausfeld et al., 2012). Future studies including different sets of words, languages or feature selection approaches may help confirming the generalization of our results. Beyond decoding language-invariant semantic-concepts during listening, EEG-based MVPA may also be used to investigate whether semantic-concepts share a similar neural representation during reading and speaking (Hickok et al., 2011; Pickering and Garrod, 2013). When we speak, we start from ideas and concepts and convert these into articulatory motor programs. ERP studies on speech production (e.g., picture naming), relate early windows, 100–200 ms after stimulus onset to interactive processing of visual encoding and accessing concepts for language use (Rahman and Sommer, 2003; Redmann et al., 2014). Like in speech comprehension, this interaction between input-dependent and abstract semantic-conceptual representations in speech production, together with their strong context and task-dependency (e.g., Jescheniak et al., 2002; Aristei et al., 2011), makes it difficult to isolate abstract semantic conceptual representations using univariate analysis methods. Because our EEG-based MVPA approach may disentangle these processes, it would thus be of interest to employ this same approach in speech production studies (e.g., and Schmitt et al., 2000; Koester and Schiller, 2008). In particular, a similar bilingual paradigm involving word naming in bilingual speakers would allow investigating the timing of language-independent semantic-conceptual representations. Furthermore, the classification of spoken words across and within languages in bilingual speakers and across and within speech modality (perception and production) may allow to investigate neural representations crucial for the initiation of

speech production (Levelt, 1989; Rahman and Sommer, 2003; Indefrey and Levelt, 2004; Indefrey, 2011), as well as, for the monitoring of speech output (Hickok et al., 2011).

REFERENCES

- Aristei, S., Melinger, A., and Rahman, A. R. (2011). Electrophysiological chronometry of semantic context effects in language production. *J. Cogn. Neurosci.* 23, 1567–1586. doi: 10.1162/jocn.2010.21474
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Baddeley, A. (2003). Working memory and language: an overview. *J. Commun. Disord.* 36, 189–208. doi: 10.1016/S0021-9924(03)00019-4
- Bastiaansen, M. C. M., Oostenveld, R., Jensen, O., and Hagoort, P. (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain Lang.* 106, 15–28. doi: 10.1016/j.bandl.2007.10.006
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528. doi: 10.1093/cercor/10.5.512
- Boersma, P., and Weenink, D. (2013). Praat: doing phonetics by computer. *Glott Int.* 5, 341–345. Available online at: <http://www.praat.org/>
- Bonte, M. (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clin. Neurophysiol.* 115, 409–423. doi: 10.1016/S1388-2457(03)00361-4
- Bonte, M., and Blomert, L. (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clin. Neurophysiol.* 115, 409–423. doi: 10.1016/S1388-2457(03)00361-4
- Bonte, M., Hausfeld, L., Scharke, W., Valente, G., and Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J. Neurosci.* 34, 4548–4557. doi: 10.1523/JNEUROSCI.4339-13.2014
- Bonte, M., Parviainen, T., Hytönen, K., and Salmelin, R. (2006). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cereb. Cortex* 16, 115–123. doi: 10.1093/cercor/bhi091
- Bonte, M., Valente, G., and Formisano, E. (2009). Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *J. Neurosci.* 29, 1699–1706. doi: 10.1523/JNEUROSCI.3694-08.2009
- Brandmeyer, A., Farquhar, J. D. R., McQueen, J. M., and Desain, P. W. M. (2013). Decoding speech perception by native and non-native speakers using single-trial electrophysiological data. *PLoS ONE* 8:e68261. doi: 10.1371/journal.pone.0068261
- Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., et al. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain Lang.* 119, 119–128. doi: 10.1016/j.bandl.2010.12.001
- Chan, A. M., Baker, J. M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., et al. (2011a). First-pass selectivity for semantic categories in human anteroventral temporal lobe. *J. Neurosci.* 31, 18119–18129. doi: 10.1523/JNEUROSCI.3122-11.2011
- Chan, A. M., Halgren, E., Marinkovic, K., and Cash, S. S. (2011b). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage* 54, 3028–3039. doi: 10.1016/j.neuroimage.2010.10.073
- Christoffels, I. K., Ganushchak, L., and Koester, D. (2013). Language conflict in translation: an ERP study of translation production. *J. Cogn. Psychol.* 25, 646–664. doi: 10.1080/20445911.2013.821127
- Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., and Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *J. Neurosci.* 34, 332–338. doi: 10.1523/JNEUROSCI.1302-13.2014
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Çukur, T., Nishimoto, S., Huth, A. G., and Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nat. Neurosci.* 16, 763–770. doi: 10.1038/nn.3381
- Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., and Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature* 380, 499–505. doi: 10.1038/380499a0
- Darlington, R. B., and Hayes, A. F. (2000). Combining independent p values: extensions of the Stouffer and binomial methods. *Psychol. Methods* 5, 496–515. doi: 10.1037/1082-989X.5.4.496
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- De Martino, F., de Borst, A. W., Valente, G., Goebel, R., and Formisano, E. (2011). Predicting EEG single trial responses with simultaneous fMRI and relevance vector machine regression. *Neuroimage* 56, 826–836. doi: 10.1016/j.neuroimage.2010.07.068
- De Moor, W., Ghyselinck, M., and Brysbaert, M. (2000). A validation study of the age-of-acquisition norms collected by Ghyselinck, De Moor, and Brysbaert. *Psychol. Belg.* 40, 99–114.
- Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85 Pt 2, 761–768. doi: 10.1016/j.neuroimage.2013.06.035
- Dumay, N., Benraïss, A., Barriol, B., Colin, C., Radeau, M., and Besson, M. (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *J. Cogn. Neurosci.* 13, 121–143. doi: 10.1162/089892901564117
- Formisano, E., De Martino, F., Bonte, M., and Goebel, R. (2008b). “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* 322, 970–973. doi: 10.1126/science.1164318
- Formisano, E., De Martino, F., and Valente, G. (2008a). Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magn. Reson. Imaging* 26, 921–934. doi: 10.1016/j.mri.2008.01.052
- Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1055–1069. doi: 10.1098/rstb.2007.2159
- Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441. doi: 10.1126/science.1095455
- Hausfeld, L., De Martino, F., Bonte, M., and Formisano, E. (2012). Pattern analysis of EEG responses to speech and voice: influence of feature grouping. *Neuroimage* 59, 3641–3651. doi: 10.1016/j.neuroimage.2011.11.056
- Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., et al. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416. doi: 10.1016/j.neuron.2011.08.026
- Henry, M. J., and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20095–20100. doi: 10.1073/pnas.1213390109
- Herrmann, B., Maess, B., Kalberlah, C., Haynes, J.-D., and Friederici, A. D. (2012). Auditory perception and syntactic cognition: brain activity-based decoding within and across subjects. *Eur. J. Neurosci.* 35, 1488–1496. doi: 10.1111/j.1460-9568.2012.08053.x
- Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422. doi: 10.1016/j.neuron.2011.01.019
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:255. doi: 10.3389/fpsyg.2011.00255
- Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001
- Jescheniak, J. D., Schriefers, H., Garrett, M. F., and Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *J. Cogn. Neurosci.* 14, 951–964. doi: 10.1162/089892902760191162
- Kilian-Hütten, N., Valente, G., Vroomen, J., and Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J. Neurosci.* 31, 1715–1720. doi: 10.1523/JNEUROSCI.4572-10.2011
- Koester, D., and Schiller, N. O. (2008). Morphological priming in overt language production: electrophysiological evidence from Dutch. *Neuroimage* 42, 1622–1630. doi: 10.1016/j.neuroimage.2008.06.043

- Kutas, M., and Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/S1364-6613(00)01560-6
- Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.
- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi: 10.1152/jn.00263.2005
- Lee, T. W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/089976699300016719
- Lemhöfer, K., and Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44, 325–343. doi: 10.3758/s13428-011-0146-0
- Levelt, W. J. M. (1989). *Speaking: from Intention to Articulation*. Cambridge, MA: MIT Press.
- Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Dokl. Akad. Nauk SSSR* 163, 845–848.
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic brain sources of visual evoked responses. *Science* 295, 690–694. doi: 10.1126/science.1066168
- Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., and Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* 38, 487–497. doi: 10.1016/S0896-6273(03)00197-1
- Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition* 25, 71–102. doi: 10.1016/0010-0277(87)90005-9
- Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45. doi: 10.1146/annurev.psych.57.102904.190143
- McClelland, J. L., and Elman, J. L. (1986). The TRACE Model of Speech Perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- Meyer, K., and Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci.* 32, 376–382. doi: 10.1016/j.tins.2009.04.002
- Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005
- Obleser, J., Elbert, T., and Eulitz, C. (2004). Attentional influences on functional mapping of speech sounds in human auditory cortex. *BMC Neurosci.* 5:24. doi: 10.1186/1471-2202-5-24
- Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. *Front. Hum. Neurosci.* 6:250. doi: 10.3389/fnhum.2012.00250
- Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987. doi: 10.1038/nrn2277
- Peelle, J. E., and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3:320. doi: 10.3389/fpsyg.2012.00320
- Peña, M., and Melloni, L. (2012). Brain oscillations during spoken sentence processing. *J. Cogn. Neurosci.* 24, 1149–1164. doi: 10.1162/jocn_a_00144
- Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495
- Rahman, R. A., and Sommer, W. (2003). Does phonological encoding in speech production always follow the retrieval of semantic knowledge? Electrophysiological evidence for parallel processing. *Brain Res. Cogn. Brain Res.* 16, 372–382. doi: 10.1016/S0926-6410(02)00305-1
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331
- Redman, A., Fitzpatrick, I., Hellwig, F., and Indefrey, P. (2014). The use of conceptual components in language production: an ERP study. *Front. Psychol.* 5:363. doi: 10.3389/fpsyg.2014.00363
- Salmelin, R. (2007). Clinical neurophysiology of language: the MEG approach. *Clin. Neurophysiol.* 118, 237–254. doi: 10.1016/j.clinph.2006.07.316
- Sanders, L. D., and Neville, H. J. (2003). An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native speakers. *Brain Res. Cogn. Brain Res.* 15, 228–240. doi: 10.1016/S0926-6410(02)00195-7
- Schmitt, B. M., Münte, T. F., and Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* 37, 473–484. doi: 10.1111/1469-8986.3740473
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123 Pt 12, 2400–2406. doi: 10.1093/brain/123.12.2400
- Shahin, A. J., Picton, T. W., and Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain Cogn.* 70, 259–266. doi: 10.1016/j.bandc.2009.02.008
- Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., and Just, M. A. (2011). Commonality of neural representations of words and pictures. *Neuroimage* 54, 2418–2425. doi: 10.1016/j.neuroimage.2010.10.042
- Simanova, I., van Gerven, M., Oostenveld, R., and Hagoort, P. (2010). Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PLoS ONE* 5:e14465. doi: 10.1371/journal.pone.0014465
- Strauß, A., Kotz, S. A., Scharinger, M., and Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *Neuroimage* 97C, 387–395. doi: 10.1016/j.neuroimage.2014.04.005
- Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., et al. (2012). Tracking neural coding of perceptual and semantic features of concrete nouns. *Neuroimage* 62, 451–463. doi: 10.1016/j.neuroimage.2012.04.048
- Szumilas, M. (2010). Explaining odds ratios. *J. Can. Acad. Child Adolesc. Psychiatry* 19, 227–229.
- Travis, K. E., Leonard, M. K., Chan, A. M., Torres, C., Sizemore, M. L., Qu, Z., et al. (2013). Independence of early speech processing from word meaning. *Cereb. Cortex* 23, 2370–2379. doi: 10.1093/cercor/bhs228
- Uusvuori, J., Parviainen, T., Inkinen, M., and Salmelin, R. (2008). Spatiotemporal interaction between sound form and meaning during spoken word perception. *Cereb. Cortex* 18, 456–466. doi: 10.1093/cercor/bhm076
- Van den Brink, D., Brown, C. M., and Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *J. Cogn. Neurosci.* 13, 967–985. doi: 10.1162/089892901753165872
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., and Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 394–417. doi: 10.1037/0278-7393.25.2.394
- Visser, M., Jefferies, E., Embleton, K. V., and Lambon Ralph, M. A. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J. Cogn. Neurosci.* 24, 1766–1778. doi: 10.1162/jocn_a_00244
- Visser, M., and Lambon Ralph, M. A. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *J. Cogn. Neurosci.* 23, 3121–3131. doi: 10.1162/jocn_a_00007
- Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J.-M., Magyari, L., et al. (2012). Beta oscillations relate to the N400m during language comprehension. *Hum. Brain Mapp.* 33, 2898–2912. doi: 10.1002/hbm.21410

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 June 2014; accepted: 13 January 2015; published online: 06 February 2015.

Citation: Correia JM, Jansma B, Hausfeld L, Kikkert S and Bonte M (2015) EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Front. Psychol.* 6:71. doi: 10.3389/fpsyg.2015.00071

This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Correia, Jansma, Hausfeld, Kikkert and Bonte. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.