Supplementary Material: Neural manifolds carry reactivation of phonetic representations during semantic processing

Pavo Orepic[1], Wilson Truccolo[2,3], Eric Halgren[4], Sydney S. Cash[5], Anne-Lise Giraud[1,6], Timothée Proix[1*]

[1]Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland.
[2]Department of Neuroscience, Brown University, Providence, Rhode Island, United States of America.
[3]Carney Institute for Brain Science, Brown University, Providence, Rhode Island, United States of America.
[4]Department of Neuroscience & Radiology, University of California San Diego, La Jolla, California, United States of America.
[5]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America.
[6]Institut Pasteur, Université Paris Cité, Hearing Institute, Paris, France.

*Corresponding author(s). E-mail(s): timothee.proix@unige.ch;

Phonetic and semantic mTRF features

Supplementary Table 1 | Vowels used in the mTRF models. Duration is computed as the mean in milliseconds across all occurrences of the same phoneme.

			First formant		*Second formant*		
Vowel	**Repetitions**	**Duration**	**Low**	**High**	**Front**	**Middle**	**Back**
/ae/	64	87.8	1	0	1	0	0
/ʌ/	103	47.0	1	0	0	0	1
/ɪ/	61	56.1	0	1	1	0	0
/ɜ/	43	89.4	1	0	0	1	0
/uː/	31	81.8	0	1	0	0	1
/ɒ/	49	73.4	1	0	0	0	1
/ɛ/	44	66.4	1	0	1	0	0
/ɔ/	29	91.5	1	0	0	0	1
/i/	55	89.6	0	1	1	0	0
/u/	7	57.1	0	1	0	0	1
/ʊ/	8	33.8	0	1	0	0	1

Supplementary Table 2 | Consonants used in the mTRF models. Duration is computed as the mean in milliseconds across all occurrences of the same phoneme.

			Articulation place						*Articulation manner*				
Consonant	**Repetitions**	**Duration**	**Bilabial**	**Labiodental**	**Alveolar**	**Velar**	**Uvular**	**Glottal**	**Plosive**	**Nasal**	**Fricative**	**Approximant**	**Lateral approximant**
/ŋ/	19	52.8	0	0	0	1	0	0	0	1	0	0	0
/k/	135	41.0	0	0	0	1	0	0	1	0	0	0	0
/l/	123	47.9	0	0	1	0	0	0	0	0	0	0	1
/t/	120	53.0	0	0	1	0	0	0	1	0	0	0	0
/n/	89	46.2	0	0	1	0	0	0	0	1	0	0	0
/p/	75	36.7	1	0	0	0	0	0	1	0	0	0	0
/f/	45	36.7	0	1	0	0	0	0	0	0	1	0	0
/d/	51	34.0	0	0	1	0	0	0	1	0	0	0	0
/ʁ/	100	38.3	0	0	0	0	1	0	0	0	1	0	0
/ʃ/	22	70.7	0	0	1	0	0	0	0	0	1	0	0
/s/	104	65.0	0	0	1	0	0	0	0	0	1	0	0
/b/	79	22.0	1	0	0	0	0	0	1	0	0	0	0
/dʑ/	18	59.0	0	0	1	0	0	0	0	0	1	0	0
/m/	60	46.6	1	0	0	0	0	0	0	1	0	0	0
/g/	36	35.0	0	0	0	1	0	0	1	0	0	0	0
/v/	13	28.5	0	1	0	0	0	0	0	0	1	0	0
/h/	25	37.9	0	0	0	0	0	1	0	0	1	0	0
/z/	13	50.0	0	0	1	0	0	0	0	0	1	0	0
/tɕ/	31	67.8	0	0	1	0	0	0	0	0	1	0	0
/w/	22	35.8	0	0	0	1	0	0	0	0	0	1	0
/θ/	10	29.0	0	0	1	0	0	0	0	0	1	0	0

Supplementary Table 3 | Semantic features for the semantic task.

	Semantic features	**Repetitions**
Conceptual category	Object	200
	Animal	200
Perceptual category	Bigger than a foot	200
	Smaller than a foot	200
Semantic decision	Bigger than a foot	139
	Smaller than a foot	261

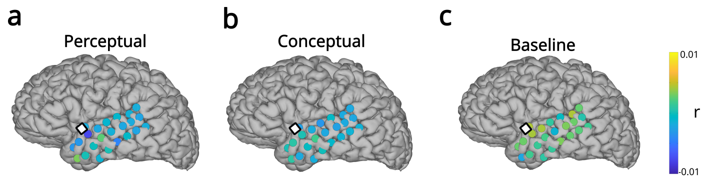
Supplementary Table 4 | Word class features for natural speech perception.

Word class	**Repetitions**	**Unique repetitions**
Noun	147	97
Verb	105	49
Adjective	26	17
Adverb	31	13
Article	49	5
Auxiliary	15	8
Demonstrative	17	4
Quantifier	7	3
Preposition	88	25
Pronoun	78	18
Conjunction	32	7
Interjection	41	12
Number	28	14

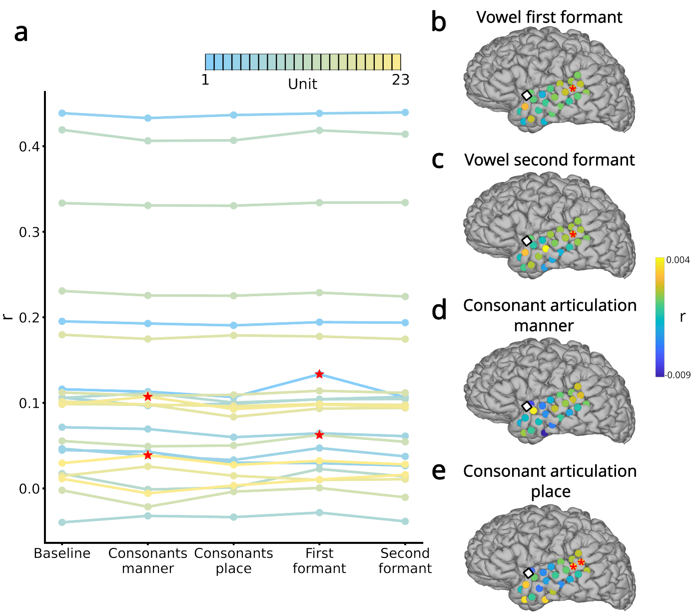
Supplementary Table 5 | Lexical semantics feature for natural speech perception.

Lancaster sensorimotor norm	**Strength**
Auditory	1.83
Gustatory	0.36
Haptic	1.13
Interoceptive	1.2
Olfactory	0.44
Visual	2.73
Foot, leg	1.03
Hand, arm	1.49
Head	2.37
Mouth	1.4
Torso	0.91

Additional mTRF models



Supplementary Fig. 1 | Non-significant and baseline ECoG models a. Encoding of perceptual category across ECoG channels. Colors indicate differences in r values compared to the baseline model. No channel was significantly higher ($p < 0.05$) than the chance level r values of a surrogate distribution. **b.** Encoding of conceptual category across ECoG channels. Colors and significance as in a. **c.** Encoding of the baseline model (word onset and acoustic edges) across ECoG channels.



Supplementary Fig. 2 | Phonetic encoding at the single-unit and LFP level in the aSTG. a. Pearson correlation coefficient (r values) for models including each phonetic feature group fitted to each of the 23 single units with the highest firing rates in the ensemble. The lines for different units are color-coded based on firing rate, from lower (blue) to higher (yellow). Red stars indicate units for which r values of the fitted models are significantly higher ($p < 0.05$) than the chance level r values of a surrogate distribution. **b-e** Encoding of vowel first formant, vowel second formant, consonant manner of articulation, and consonant place of articulation across ECoG channels. Colors indicate differences in r values compared to the baseline model. Red stars indicate significance as above.

Clustering control analyses for the semantic task

To confirm our clustering results, we conducted several control analyses. We first confirmed that our results did not change when using PC spaces with more than two PC dimensions. Thus, we computed the clustering index for all four phonetic features (vowel first formant, vowel second formant, consonant manner of articulation, consonant place of articulation) up to the sixth PC dimension (Supplementary Fig. 3). The results observed for two dimensions largely generalized when using more dimensions for the PC space.

For the vowel second formant group, we additionally ascertained that the lack of clustering was not caused by the fact that it contained three phonetic features (front, middle, and back position of the tongue) instead of two as for the first formant group (high, low). We replicated the lack of clustering effects by using two instead of three phonetic features (front and back position of the tongue, Supplementary Fig. 4).

To confirm the observed clustering of phonemes along vowel first formant and consonant manner groups (and the lack of clustering along vowel second formant and consonant place groups), we performed three alternative analyses: (i) linear discriminant analysis (LDA) classifier (Supplementary Fig. 5); (ii) for vowels, rank regression analysis (rank regression was performed to test whether vowels not only separate, but are also ordered by their first formant value in the PC space, Supplementary Fig. 6); (iii) k-means clustering (Supplementary Fig. 7).

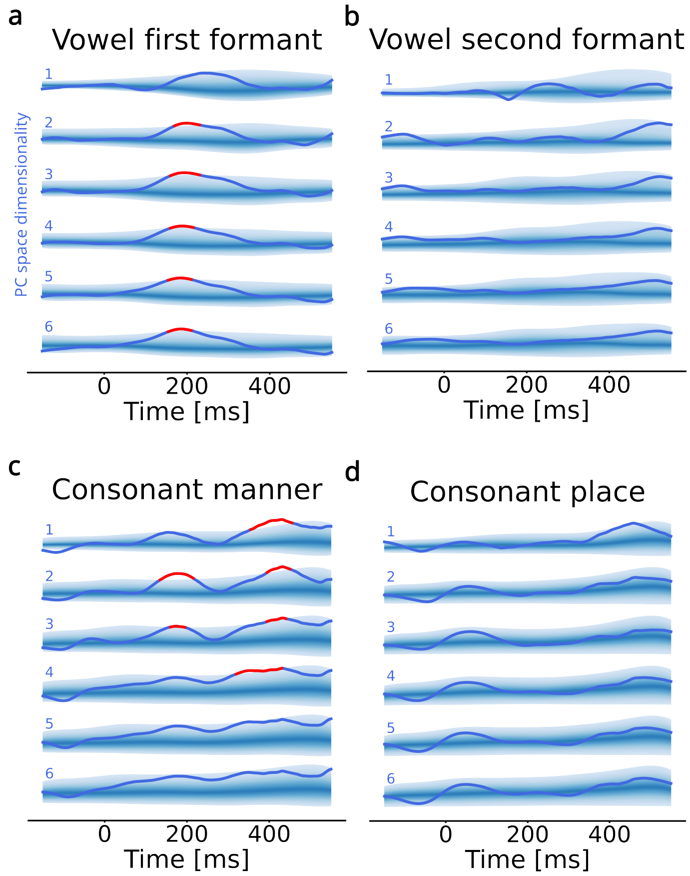
Linear discriminant analysis (LDA) classifier

We ran the LDA classifier for each of the 4 phonetic groups (vowel first formant, vowel second formant, consonant manner, consonant place). For each time point, we first ran an LDA classifier which is a Gaussian mixture model to compute the means of the multivariate normal distributions for each class. Then, we computed the average Euclidean distance between all class means and compared it against the distribution of 1000 surrogates. This is similar to the between-cluster distances from our clustering algorithm, with the difference that here the class mean (cluster centroid) is a parameter of the estimated multivariate distribution, and not computed directly by averaging class elements.

Replicating the clustering index analyses, at 200 ms we observed a significant separation of vowels into first formant categories (Supplementary Fig. 5a). Linear separators estimated at 0 and 200 ms are shown in Supplementary Fig. 5b. No peaks was observed for the vowel second formant (Supplementary Fig. 5c). LDA also indicated the two previously observed peaks for consonant manner (200 and 400 ms), however, here, they did not reach significance (Supplementary Fig. 5d). Finally, LDA revealed a significant peak for consonant place at 50 ms (Supplementary Fig. 5e) that was not observed before.

Rank regression

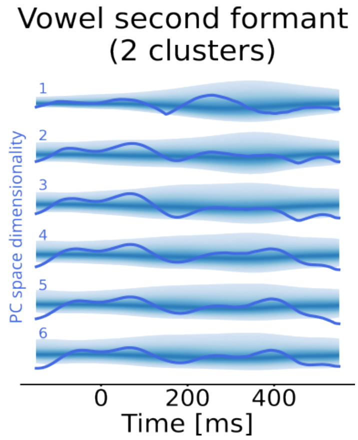
Since formants are actually ordered by their frequency values along the first and second formant axis, we additionally explored whether the actual ordering of formant frequency values was encoded in the low-dimensional space. To that aim, we ran a



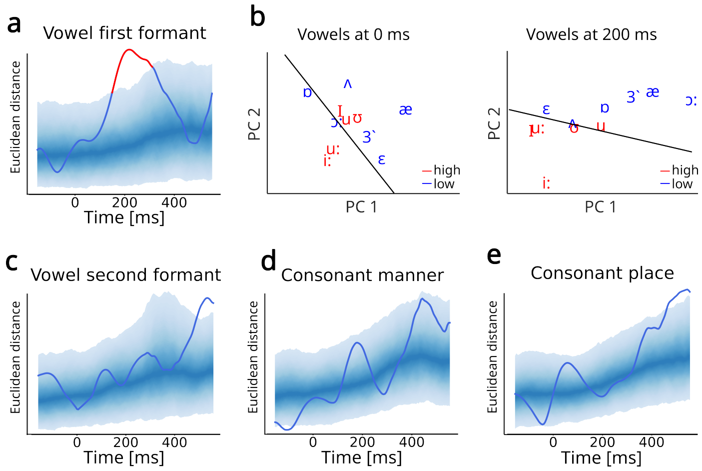
Supplementary Fig. 3 | Clustering index across dimensions during semantic categorization. a. Clustering index for vowels grouped by first formant (high and low tongue position) up to six PC dimensions. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Red segments indicate significant periods after multiple comparison correction (cluster-based test). Numbers indicate the dimensionality of the PC space. **b.** Clustering index for vowels grouped by second formant up to six PC dimensions. Coloring and numbering as in **a**. **c.** Clustering index for consonants grouped by manner of articulation up to six PC dimensions. Coloring and numbering as in **a**. **d.** Clustering index for consonants grouped by place of articulation up to six PC dimensions. Coloring and numbering as in **a**.

rank regression analysis, where a rank value (1-7) was assigned to each vowel, based on the standard IPA table. At each time point, the ranked order of vowels was correlated with their coordinates on the first three axes (PC1, PC2, and PC3), and compared against a distribution of 1000 surrogates.

We observed a significant ordering of vowels based on their first formant ranks on all three PCs (Fig. 6a). We performed the same regression for the second formant values

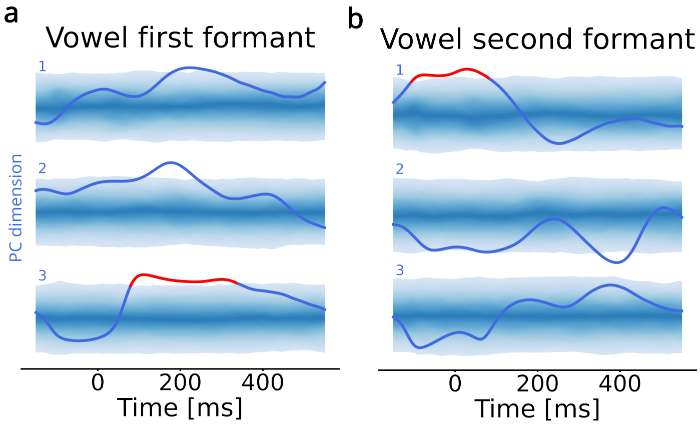


Supplementary Fig. 4 | Clustering for the vowel second formant group using only two phonetic features. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles.

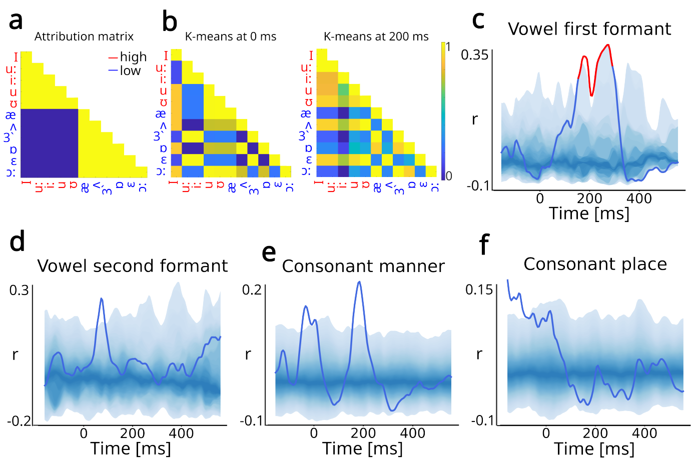


Supplementary Fig. 5 | Linear discriminant analysis during semantic categorization. a. Vowel first formant. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Red segments indicate significant periods after multiple comparison correction (cluster-based test). **b.** Estimated linear separators for vowel first formant feature at 0 and 200 ms. **c** Vowel second formant. Shading and red segments are as in **a**. **d.** Consonant manner of articulation. Coloring as in **a**. **e.** Consonant place of articulation. Coloring as in **a**.

and observed a significant correlation at about 0 ms, but only on PC1 (Supplementary Fig. 6b). As the two consonant groups are based on categorical and not continuous variables, rank regression analysis is not applicable.



Supplementary Fig. 6 | Rank regression during semantic categorization. a. Correlation between vowel first formant ranks and coordinates on PC axes. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Red segments indicate significant periods after multiple comparison correction (cluster-based test). Numbers indicate PC dimensions. **b.** Correlation between vowel second formant ranks and coordinates on PC axes. Coloring and numbering as in **a**.



Supplementary Fig. 7 | K-means clustering during semantic categorization. a. Attribution matrix for the vowel first formant feature. **b.** K-means matrices for the vowel first formant feature at 0 and 200 ms. **c.** Correlation between vowel first formant attribution matrix (**a**) and corresponding k-means matrices (**b**) at each time point. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Red segments indicate significant periods after multiple comparison correction (cluster-based test). **d.** Correlation between vowel second formant and corresponding k-means attribution matrices at each time point. Coloring as in **c**. **e.** Correlation between consonant manner and corresponding k-means matrices at each time point. Coloring as in **c**. **f.** Correlation between consonant place and corresponding k-means matrices at each time point. Coloring as in **c**.

K-means clustering

We used k-means clustering to investigate whether the same clustering of phonemes would emerge in a data-driven fashion. Our clustering results of the main text necessitate assigning a priori each of the N phonemes to a cluster (e.g., for the vowel first formant feature see Supplementary Fig. 7a), and the clustering index is then quantified based on the Euclidean distances of those a priori chosen clusters. On the contrary, k-means clustering is an unsupervised method that partitions all N phonemes into k clusters based on their proximity in PC space. Thus, for each time point, we first ran k-means clustering 1000 times, as the clustering results might change based on the algorithm's random initialization. Then, we computed an average N-by-N attribution matrix that indicated how often each of the N phonemes was clustered together (Supplementary Fig. 7b). Finally, we correlated the resulting k-means attribution matrix with the attribution matrix of the actual, linguistically-based clusters, and compared the correlation value against the distribution of correlation values for the 1000 surrogates at each time point (Supplementary Fig. 7c-f).

Replicating previous clustering results for the vowel first formant group, this analysis revealed a significant period at about 200 ms during which data-driven clusters overlapped with the linguistic ones (Supplementary Fig. 7c). This analysis further revealed a significant peak at 75 ms for vowel second formant that was not present before (Supplementary Fig. 7d). The 200-ms peak was also observed again for the consonant manner group, with an additional peak at about 0 ms, and no significant peak at 400 ms (Supplementary Fig. 7e). No significant peaks were observed for consonant place (Supplementary Fig. 7f).

Despite its advantageous data-driven perspective, there are also some important drawbacks of this method, that can be observed in the shape of the correlation curve for the vowel first formant group (Supplementary Fig. 7c). Namely, at 200 ms there is a decrease in the correlation value (although still significant), because at that time point vowels /ʌ/ and /ɛ/, which belong to the 'low' group, are closer in space to the 'high' group (compare the 200-ms insets on Fig. 2e and Supplementary Fig. 7b). However, it is apparent that, despite their spatial proximity to the 'high' group, vowels /ʌ/ and /ɛ/ can be attributed to the 'low' group because the two clusters can be linearly separated right next to the positions of these two vowels in space (Supplementary Fig. 5b).

Summary of control analyses

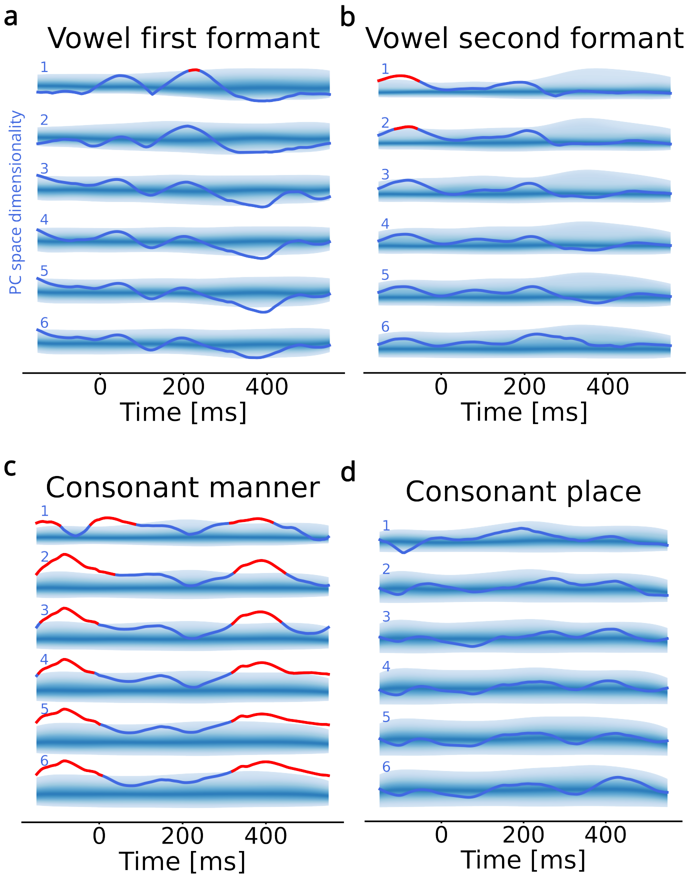
Supplementary Table 6 summarizes all peaks observed across different analyses for all 4 groups. Although some analyses sporadically revealed significant peaks for vowel second formant (k-means) and consonant place (LDA) groups, the peaks that were consistent across all analyses are the same peaks as observed in our initial clustering analysis: 200 ms for vowel first formant and consonant manner, and 400 ms for consonant manner alone.

Supplementary Table 6 | Summary of the significant window locations during semantic categorization identified through control analyses. Bold entries indicate peaks that were consistent across analyses. Entries in brackets indicate existing peaks that did not cross the significance threshold. Entries in red indicate peaks surviving multiple comparison correction (cluster-based test). All values are in ms.

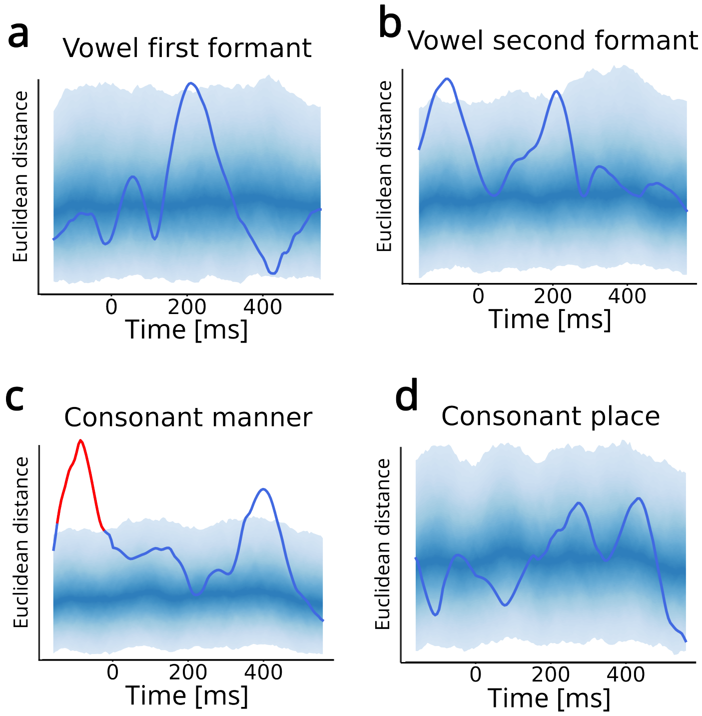
Phonetic feature group	**Clustering**	**LDA**	**Rank regression**	**K-means**
Vowel first formant	**200**	**200**	**200**	**200**
Vowel second formant			0	75
Consonant manner	**200, 400**	**(200, 400)**	not applicable	0, **200**
Consonant place		450	not applicable	

Clustering control analyses for natural speech perception

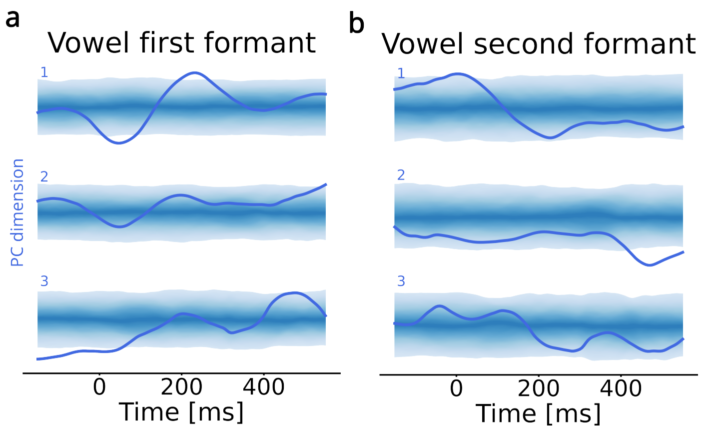
We performed the same control analyses for natural speech perception for each of the four feature groups (vowel first formant, vowel second formant, consonants manner, consonant place): clustering up to six dimensions of the PC space (Supplementary Fig. 8), LDA (Supplementary Fig. 9), rank regression (Supplementary Fig. 10), and correlation with K-means attribution (Supplementary Fig. 11). We then summarized the significant peaks observed during natural speech perception across all analyses. (Supplementary Table 7).



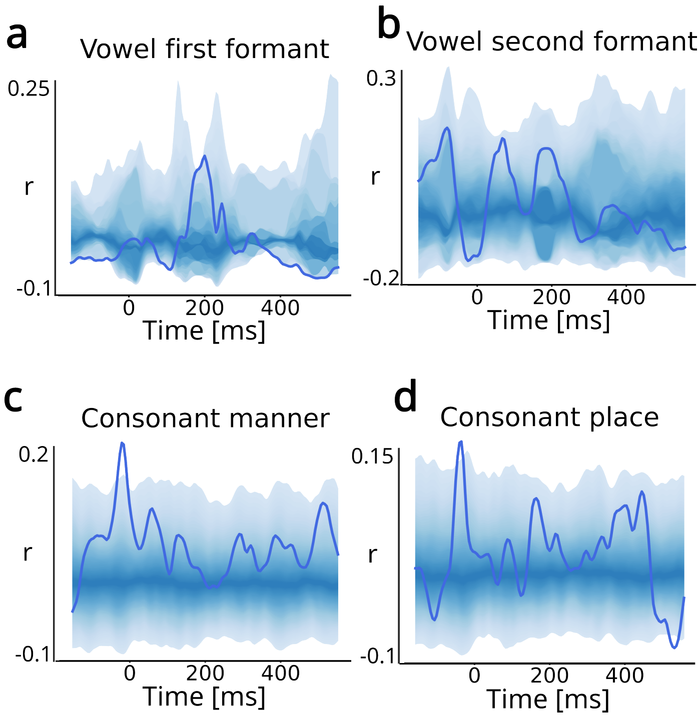
Supplementary Fig. 8 | Clustering index across dimensions during natural speech perception. a. Clustering index for vowels grouped by the first formant (high and low tongue position) up to 6 PC dimensions. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Numbers indicate the dimensionality of the PC space. **b.** Clustering index for vowels grouped by the second formant up to six PC dimensions. Coloring and numbering as in **a**. **c.** Clustering index for consonants grouped by the manner of articulation up to six PC dimensions. Shading and numbering as in **a**. **d.** Clustering index for consonants grouped by the place of articulation up to six PC dimensions. Coloring and numbering as in **a**.



Supplementary Fig. 9 | Linear discriminant analysis during natural speech perception. a. Vowel first formant. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Red segments indicate significant periods after multiple comparison correction (cluster-based test). **b.** Vowel second formant. Coloring as in **a**. **c.** Consonant manner of articulation. Coloring as in **a**. **d.** Consonant place of articulation. Coloring as in **a**.



Supplementary Fig. 10 | Rank regression during natural speech perception. a. Correlation between vowel first formant ranks and coordinates on PC axes. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Numbers indicate PC dimensions. **b.** Correlation between vowel second formant ranks and coordinates on PC axes. Coloring and numbering as in **a**.



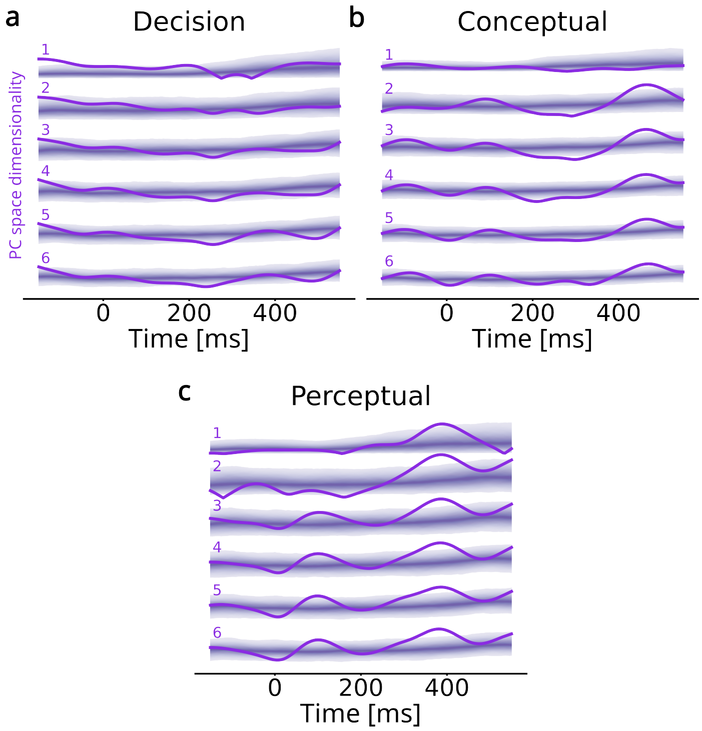
Supplementary Fig. 11 | K-means clustering during natural speech perception. a. Correlation between vowel first formant and corresponding k-means attribution matrices at each time point. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. **b.** Correlation between vowel second formant and corresponding k-means attribution matrices at each time point. Coloring as in **a**. **c.** Correlation between consonant manner and corresponding k-means attribution matrices at each time point. Coloring as in **a**. **d.** Correlation between consonant place and corresponding k-means attribution matrices at each time point. Coloring as in **a**.

Supplementary Table 7 | Summary of the significant window locations during natural speech perception identified through control analyses. Bold entries indicate peaks that were consistent across analyses. Entries in brackets indicate existing peaks that did not cross the significance threshold. Entries in red indicate peaks surviving multiple comparison correction (cluster-based test). All values are in ms.

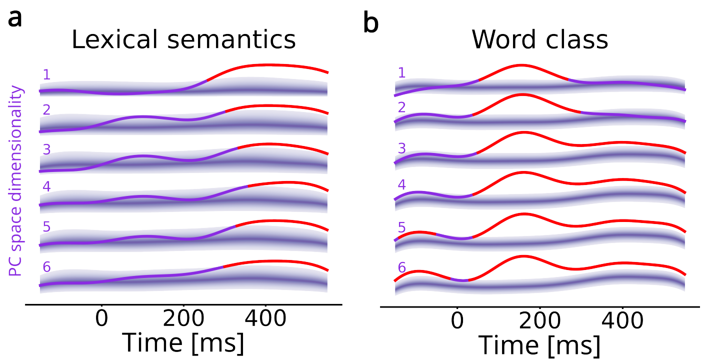
Phonetic feature group	Clustering	LDA	Rank regression	K-means
Vowel first formant	(**200**)	**200**	**200**	**200**
Vowel second formant	**-100**, (**200**)	**-100**, **200**	0	
Consonant manner	**-100**, **400**	**-100**, **400**	not applicable	-25
Consonant place			not applicable	-25

Control analyses for semantic encoding

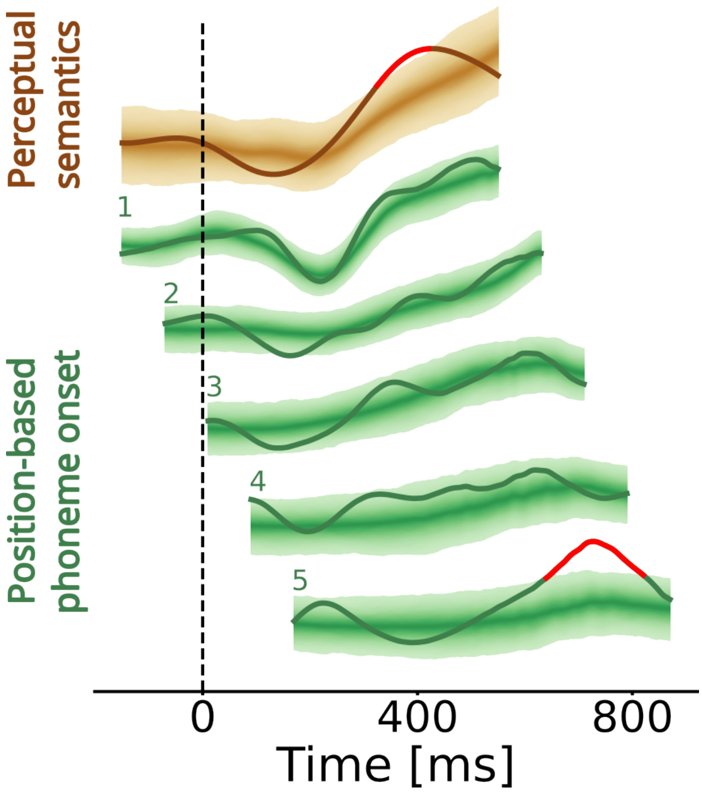
We also performed control analyses for the semantic encoding results. We confirmed that our results hold when using a different number of PC dimensions to compute the Euclidean distances between semantic features, both for the semantic task (Supplementary Fig. 12) and natural speech perception (Supplementary Fig. 13).



Supplementary Fig. 12 | **Encoding of semantic features across dimensions during semantic categorization. a.** Decision kernels. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. **b.** Conceptual kernels. Coloring and numbering as in **a**. **c.** Perceptual kernels. Coloring and numbering as in **a**.



Supplementary Fig. 13 | Encoding of semantic features across dimensions during natural speech perception. a. Lexical semantic kernels. The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Red segments indicate significant periods after multiple comparison correction (cluster-based test). Numbers indicate the dimensionality of the PC space. **b.** Word class kernels. Coloring and numbering as in **a**.



Supplementary Fig. 14 | Concurrent encoding of phonetic and semantic features during semantic task Kernels for the position-based phoneme onsets (green) within a word shifted by average phoneme duration (80 ms) and aligned with the perceptual semantic kernel at word onset (brown). The 95% confidence region of the surrogate (chance-level) distribution is shown with brighter shading for increasingly peripheral percentiles. Red segments indicate significant periods after multiple comparison correction (cluster-based test).