

Shared and language-specific phonological processing in the human temporal lobe

<https://doi.org/10.1038/s41586-025-09748-8>

Received: 9 January 2025

Accepted: 13 October 2025

Published online: 19 November 2025

Open access

 Check for updates

Ilina Bhaya-Grossman^{1,2,3}, Matthew K. Leonard^{1,2,3}, Yizhen Zhang^{1,2}, Laura Gwilliams^{1,2}, Keith Johnson⁴, Junfeng Lu⁵ & Edward F. Chang^{1,2,3}✉

All spoken languages are produced by the human vocal tract, which defines the limited set of possible speech sounds. Despite this constraint, however, there exists incredible diversity in the world's 7,000 spoken languages, each of which is learned through extensive experience hearing speech in language-specific contexts¹. It remains unknown which elements of speech processing in the brain depend on daily language experience and which do not. In this study, we recorded high-density cortical activity from adult participants with diverse language backgrounds as they listened to speech in their native language and an unfamiliar foreign language. We found that, regardless of language experience, both native and foreign languages elicited similar cortical responses in the superior temporal gyrus (STG), associated with shared acoustic–phonetic processing of foundational speech sound features^{2,3}, such as vowels and consonants. However, only during native language listening did we observe enhanced neural encoding in the STG for word boundaries, word frequency and language-specific sound sequence statistics. In a separate cohort of bilingual participants, this encoding of word- and sequence-level information appeared for both familiar languages in the same individual and in the same STG neural populations. These results indicate that experience-dependent language processing involves dynamic integration of both shared acoustic–phonetic and language-specific sequence- and word-level information in the STG.

When listening to speech in a foreign language, we hear a fast, continuous and uninterpretable stream of speech sounds^{4,5}. However, when we listen to speech in a language we know, we hear these sounds as words. This specialized perception of speech in one's native languages develops over the course of language acquisition^{4,6–11}. A central question in speech neuroscience is how and to what extent the adult brain's processing of speech is specific to one's native languages.

Previous work has identified the human STG, a non-primary auditory area, as crucial for speech perception^{3,12–14}. Both native and foreign or unfamiliar speech elicit a broadly similar pattern of activation in the bilateral middle STG^{15–23}. As a result, models of speech processing have hypothesized that the middle STG performs similar complex acoustic processing across native and foreign speech^{13,24}. However, it is unclear to what extent acoustic–phonetic and higher-level speech representations in STG are affected by language experience, particularly given that language-specific features such as phonetic categories^{2,25,26} and lexical tone^{27,28} are encoded there. Furthermore, STG neural populations have been shown to encode aspects of whole words^{3,29}, which are highly specific to individual languages and must be learned through extensive exposure. In this study, we ask which features are affected by language experience in the human temporal cortex during natural speech processing.

To address this question, we report on a rare dataset collected over the span of 10 years in which we leveraged high-density

electrocorticography (ECoG) recordings from a cohort of Spanish, English and Mandarin monolingual speakers passively listening to naturally spoken sentences in their native language and a foreign language (Fig. 1a). The high spatial and temporal resolution of direct cortical surface ECoG enabled us to ask not only about similarities and differences in neural populations activated by native and unfamiliar foreign languages but also much more specifically about how transient speech features are encoded depending on language experience.

We found that the overall magnitude of neural activity in the bilateral STG is similar for native and foreign languages, driven by the largely conserved encoding of acoustic–phonetic speech features. Given this result, we hypothesized that neural signatures present in ECoG recordings of known language processing might be realized at a different representational level: namely, in the enhanced encoding of sequences of acoustic–phonetic features that form larger, language-specific compositional structures, such as syllables and words. Indeed, we found that neural populations more robustly encoded the sequential structure of listeners' native language, including identifying where individual words begin and end in natural speech. In a separate cohort of Spanish–English bilingual participants, we found that neural populations encoded phoneme sequence and word-level information for both familiar languages, demonstrating experience-dependent mechanisms for speech processing across multiple languages. Finally, in participants

¹Department of Neurological Surgery, University of California, San Francisco, CA, USA. ²Weill Institute for Neuroscience, University of California, San Francisco, CA, USA. ³Joint Program in Bioengineering, University of California, Berkeley and University of California, San Francisco, Berkeley, CA, USA. ⁴Department of Linguistics, University of California, Berkeley, Berkeley, CA, USA. ⁵Department of Neurosurgery, Huashan Hospital, Fudan University, Shanghai, China. ✉e-mail: edward.chang@ucsf.edu

with varying degrees of proficiency with English, we found that proficiency modulated the extent to which word-level information in speech could be neurally decoded. We propose that these results support a neurobiological model of human speech processing wherein STG neural populations perform dynamic integration of shared acoustic–phonetic and language-specific phoneme–sequence and word-level information during native language listening.

Shared phonetic encoding in native and foreign speech

To understand how language knowledge affects neural responses to natural speech, we recorded high-density ECoG from participants with diverse language backgrounds undergoing intracranial monitoring for epilepsy while they listened to spoken sentences. Each participant heard sentences in both their known native language and an unknown foreign language (Fig. 1a and Extended Data Fig. 1), allowing us to directly compare activity in the same neural populations.

Because our research question focuses on cross-linguistic phonology rather than morphology, syntax or semantics, we used as stimuli read sentences that were designed to comprehensively span each language's phonemic inventory^{30–32}. We collected ECoG recordings from 20 monolingual participants who spoke Spanish, English or Mandarin (14 left hemisphere, 11 female, 9 Spanish speakers, 8 English speakers and 3 Mandarin speakers), which are three of the most common languages in the world, each spoken by more than half a billion people and originating from two different language families (Indo-European and Sino-Tibetan). Although some participants had received exposure to other languages, they all reported little to no proficiency in the foreign language (Extended Data Table 1) and that they were unable to comprehend when presented with foreign speech. All participants listened to English speech (mean across participants = 20.44 min), Spanish speakers also listened to Spanish (mean = 20.80 min), Mandarin speakers also listened to Mandarin (mean = 23.68 min), and English speakers also listened to Spanish and/or Mandarin (Fig. 1a).

First, we asked whether the same neural populations respond to both native and foreign speech. We identified 883 electrodes with significantly greater activity during speech than silence in any language condition (high-frequency activity (HFA), 70–150 Hz; see Methods for details and Extended Data Fig. 2 for coverage). An example from a Spanish listener illustrates that the same electrodes were responsive to both native and foreign speech, particularly in the STG (Fig. 1b). Furthermore, the average response dynamics of these electrodes were highly correlated between language conditions (Pearson $r(250) = 0.98$, $P < 0.001$; Fig. 1c). These examples demonstrate that in an individual participant, the same neural populations are active in response to speech regardless of language knowledge.

Across all participants, most speech-responsive electrodes were located in the STG, with further responses in the middle temporal, postcentral, precentral and supramarginal gyrus (Fig. 1e). Electrodes with the strongest speech activity were active in response to both native and foreign speech (Extended Data Fig. 3), and 79.84% of electrodes (82.20% median, 10.70% s.d. across participants) showed significant responses to both languages. Single-electrode HFA peaks were not consistently higher, nor was the number of active electrodes significantly more when participants listened to native speech (Extended Data Fig. 3). These results demonstrate a highly overlapping area of cortex activated in response to native and foreign speech¹⁹.

We next asked whether these neural populations, which were largely active in response to both languages, also encoded the same information, independent of whether the presented language was comprehensible to the listener. All spoken languages consist of the same broad classes of acoustic–phonetic features, which describe short segments such as vowels or consonants^{2,3,33}, and therefore neural tuning to these broad classes may be conserved across native and foreign speech. However, Spanish, English and Mandarin phonology differ systematically

(for example, in how phonetic features are used to distinguish segments^{34–36}), which may give rise to language-specific or selective neural tuning to certain phonetic features.

We evaluated whether language experience affects encoding of acoustic–phonetic features in natural speech. Specifically, we asked, in cases where an acoustic–phonetic feature is present across languages, is it neurally encoded in the same way? To address this question, we fit a temporal receptive field (TRF; with acoustic–phonetic features)³⁷ to each language separately and asked the extent to which acoustic–phonetic feature tuning was the same across native and foreign speech. We compared electrode feature tuning across languages by correlating the weights from TRF models that had been fit on each language separately.

We identified electrodes with significant model fits and tuning to specific acoustic–phonetic features and observed that their response patterns were highly correlated across languages. For three example electrodes, tuning to vowel formants, changes in the speech amplitude envelope (peakRate³⁸) and fricative manner of articulation showed similar feature weights in native and foreign speech (Fig. 1f). Across electrodes with significant TRF model fits in both languages ($R^2 > 0.1$), feature weights across languages were highly correlated (Pearson $r(1992) = 0.86$, $P < 0.001$; Fig. 1g).

Overall, we found that 155 of 166 electrodes (93.37%) had TRF feature weight profiles correlated above chance ($P < 0.05$ compared to uncorrected non-parametric permuted distribution generated from shuffling TRF weight labels; Fig. 1h). We further tested this by training a TRF model in the native speech condition and testing in the foreign speech condition or vice versa. We found that this cross-training still produced significantly correlated model predictions (Extended Data Fig. 4). These results demonstrate that in addition to the same brain regions being active in response to native and foreign languages, acoustic–phonetic speech feature tuning in local neural populations is largely conserved, indicating that these populations encode the shared auditory content relevant for general speech processing.

Finally, we considered the fact that languages systematically differ from one another in their specific inventory of speech sounds and that language-specific inventories influence speech sound perception, such as categorical perception^{39–41}. It is therefore possible that despite the high resolution of the neural data and the sensitivity of TRF-encoding models, neural populations could encode more subtle, categorical differences in phonology between languages⁴². Although natural speech is not ideal for testing these differences because of its inherent multivariate, complex nature, we examined two specific examples of speech sound differences across languages to test whether they are prevalent in STG activity: voice onset time (VOT) and vowel formants.

Spanish and English differ systematically in how the acoustic cue of VOT is used to determine phoneme category membership for plosive sounds^{36,39}. We found a small but significant effect of native language on the distribution of electrodes tuned to VOT in the predicted direction (more electrodes tuned to voiced plosives in Spanish speakers) but no clear difference in how single electrodes encode plosives (Extended Data Fig. 5). In addition to VOT, Spanish and English differ in the number and acoustic realization of language-specific vowel categories⁴³. In line with our VOT results, we found no consistent, significant differences between the neural encoding of vowels in English and Spanish speakers (Extended Data Fig. 6). These results indicate that although categorical differences between phonemes can be affected by language experience (perhaps best studied with stimuli that parametrically dissociate acoustic cues, such as in ref. 35), the dominant pattern in STG when presented with natural speech is one of shared acoustic–phonetic representation across languages where categorical effects are overridden by the availability of other cues.

Together, these results indicate that the same STG neural populations respond to speech regardless of whether the listener is familiar with the language presented. Particularly in high-level auditory areas like the STG, neural populations encode largely the same acoustic–phonetic

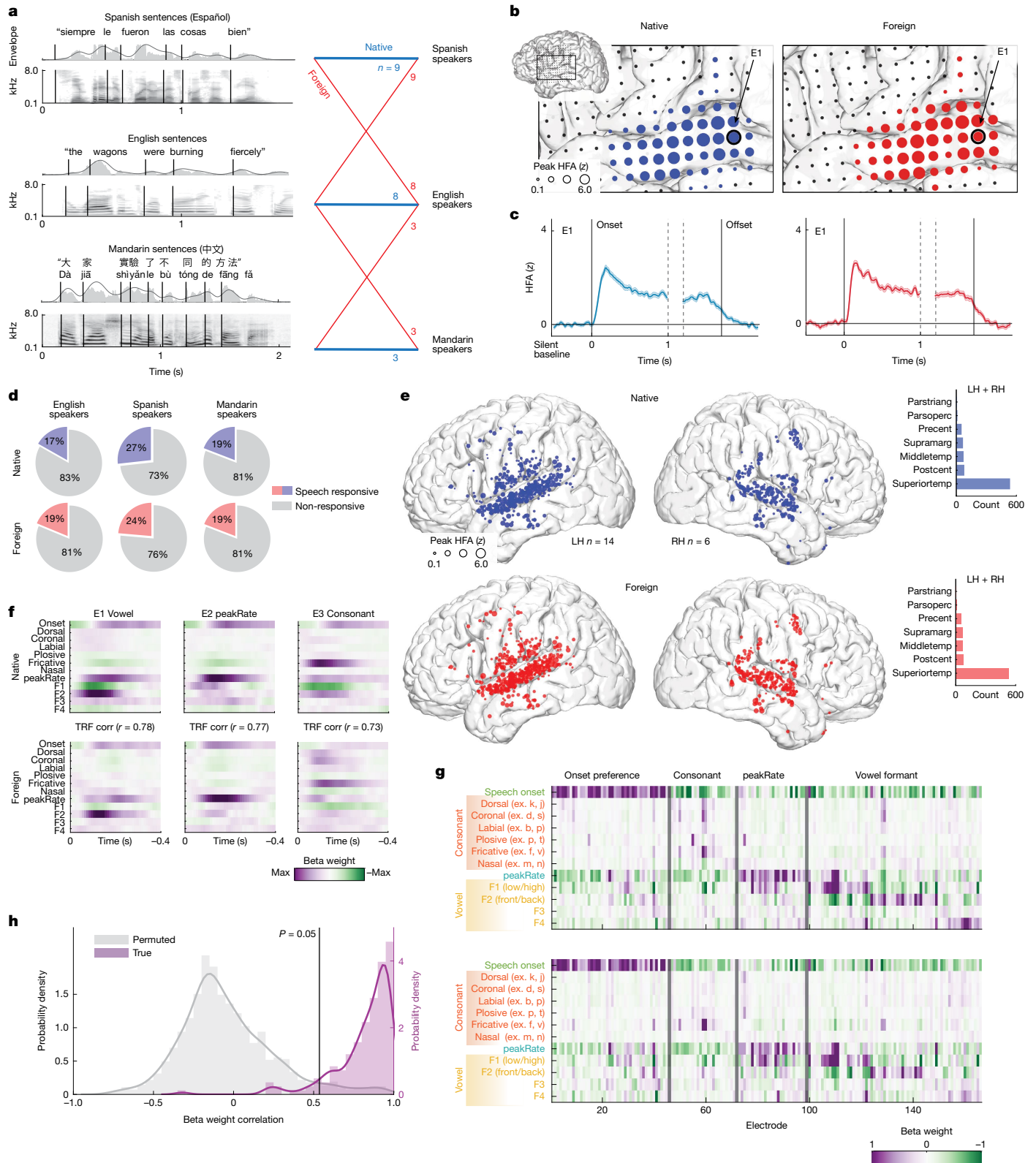


Fig. 1 | See next page for caption.

speech content, reflecting common vocal tract articulator movements and the resulting shared phonological structure of different languages.

Language experience enhances word encoding

Given that the brain represents individual speech sounds similarly regardless of language experience (Fig. 1), we hypothesized that

experience is reflected more strongly in listeners' ability to compose these sounds into phonological sequences and words. This hypothesis can be intuited by the fact that when we hear someone speaking an unfamiliar language, we may be able to tell broadly which consonants and vowels we hear, but we find it substantially more challenging to identify where one word ends and another begins. This is because natural speech does not have reliable cues to word boundaries⁴⁴⁻⁴⁶ and different

Fig. 1 | Shared acoustic–phonetic processing in STG across native and foreign speech. **a**, Spanish-, English- and Mandarin-speaking ECoG participants passively listened to native and foreign speech. Left, sound envelope and spectrogram of example stimuli. Right, coloured lines indicate speech condition, labelled with the number of participants in each group. **b**, In an example Spanish listener, the same neural populations are active in response to native (left) and foreign (right) speech. Electrode size indicates magnitude of the peak HFA averaged sentence response; colour indicates speech condition. **c**, Average HFA sentence response for an example electrode is highly correlated across native and foreign speech (Pearson $r(250) = 0.986, P < 0.001$). Neural responses time-aligned to sentence onset and offset; colour indicates speech condition. Shaded patches indicate standard error of the mean across sentences. **d**, Proportions of speech-responsive electrodes for Spanish, English and Mandarin speakers are similar across native (top) and foreign (bottom) speech, ranging from 17% to 27% across groups. **e**, Speech-responsive electrodes across 20 participants primarily localized to the STG in native (top) and foreign (bottom) speech conditions across hemispheres. Electrode size indicates

magnitude of the peak HFA averaged sentence response. Inset histograms show electrode counts across native and foreign speech in each anatomical region. **f**, Acoustic–phonetic TRF weights across three example electrodes show strong correlations between native (top) and foreign (bottom) speech conditions. Different electrodes show tuning to distinct acoustic–phonetic features common to both languages. **g**, TRF weights across speech-responsive electrodes for native (top) and foreign (bottom) speech are highly correlated (Pearson $r(1,992) = 0.86, P < 0.001$). Top, columns (electrodes) ordered by the acoustic–phonetic feature showing the maximal weight. Bottom, columns ordered identical to the top for visual comparison. **h**, Distribution of TRF native–foreign weight correlations across electrodes (purple) is higher compared to the non-parametric permuted distribution (grey). Black vertical line indicates the 95th percentile in the permuted distribution. Corr, correlation; LH, left hemisphere; middletemp, middle temporal gyrus; parsoperc, pars opercularis; parstriang, pars triangularis; postcent, postcentral gyrus; precent, precentral gyrus; RH, right hemisphere; superiotemp, superior temporal gyrus; supramarg, supramarginal gyrus; surp, surprisal.

languages such as Spanish, English and Mandarin are highly distinct at the level of speech sound sequences and words, specifically with respect to phonotactic and timing-related differences^{34,47}. As a result, we were able to specifically test the hypothesis that when listening to one's native language, compared to an unfamiliar foreign language, language experience supports the neural encoding of phonological structure⁴⁸, which is in turn used to segment the continuous speech stream into a series of auditory objects like words.

To test this hypothesis, we fit TRF models with the same acoustic–phonetic features as in Fig. 1 and other features that we predicted would be relevant for identifying the phonological structure of words in continuous speech: word boundary, word length^{49,50}, word frequency^{51,52} and phoneme surprisal (negative log probability⁵³; for a full predictor list, see Methods; Fig. 2a). We limited the following analyses to the Spanish- and English-speaking participants to perform a fully crossed comparison in which all participants listened to the same two languages (English and Spanish).

To evaluate the specific encoding of word-level features and phoneme surprisal independent of acoustic–phonetic tuning, we used unique R^2 as a measure of explained variance^{12,54}. Across speech-responsive electrodes, word boundary, length and frequency explained significantly greater unique variance for native compared to foreign speech (linear mixed effects (LME) word boundary $t(243) = 4.00$, standard error (s.e.) = 0.0008, $P < 0.001$; length $t(123) = 3.65$, s.e. = 0.0006, $P < 0.001$; frequency $t(253) = 5.18$, s.e. = 0.0006, $P < 0.001$; Fig. 2b). We also found that phoneme surprisal explained significantly greater unique variance in native compared to foreign speech ($t(29) = 3.58$, s.e. = 0.0017, $P = 0.0012$; Fig. 2b), indicating that response to sublexical sequence information is also dependent on language experience. These language-experience-dependent effects were in contrast to no differences for acoustic–phonetic feature encoding (Fig. 2c; all $P > 0.05$), consistent with the analysis of acoustic–phonetic feature weights (Fig. 1e–g). As a further control, we tested differences in unique variance explained by pitch (absolute pitch and relative pitch⁵⁵) and the continuous speech envelope, both of which showed no significant differences (pitch $t(175) = 0.55$, s.e. = 0.0027, $P = 0.58$; envelope $t(473) = 1.191$, s.e. = 0.0009, $P = 0.23$). This result establishes that a key neural difference between hearing one's native versus a foreign language is in how the phonological structure of speech is encoded towards the goal of perceiving words.

Next, we asked where on the lateral brain surface language-experience-dependent information was predominantly encoded (we combined the word-level and phoneme-surprisal features because encoding of each is strongly correlated; Extended Data Fig. 7). We found that electrodes showing significant unique variance for these features were primarily located in the bilateral mid-STG (Fig. 2d,e). These electrodes made up 34% of the speech-responsive electrodes identified in

Fig. 1 in the native speech condition and 12% of the speech-responsive electrodes in the foreign speech condition (Fig. 2f). Across single participants, the percentage of electrodes significantly encoding word-level and phoneme-surprisal features was significantly higher in the native language (Extended Data Fig. 8; signed-rank test $n = 17, P = 0.0012$).

If encoding of phonological structure is largely dependent on language experience, it is perhaps surprising that we find any significant encoding of this information in foreign speech. To understand the extent to which the encoding of phonological structure in foreign speech reflects the same processes as in the native language, we compared the unique variance for word-level and phoneme-surprisal features across languages. A minority of electrodes encoded phonological structure only in the foreign language (lower-right quadrant), potentially implicating a learning signal or tuning to new sound structure⁵⁶. Overall, encoding of phonological structure was significantly stronger in the native language (word level and phoneme surprisal $t(179) = 4.77$, s.e. = 0.0018, $P < 0.001$), including for electrodes that had significant effects in both native and foreign languages (Fig. 2g). Furthermore, the pattern of unique variance was not significantly correlated between languages (Fig. 2g; Spearman $r(98) = 0.16, P = 0.12$). Together, this indicates that although the brain may leverage native-like processing to attempt to parse foreign speech, the ability to encode phonological sequence information and words is significantly stronger in one's native language.

STG segments words in native speech

When we hear speech in a foreign language, a salient challenge for listeners is identifying when individual words begin and end (word segmentation)^{44,57}. This is in part due to the fact that there are no reliable acoustic cues to word boundaries. Even prominent acoustic cues to word boundaries in English, like sharp increases in the speech envelope at word onset^{58–60}, are unreliable because they can also occur at syllable boundaries (Fig. 3a,b). Indeed, a decoder trained to classify word boundaries using 0.2 s of the spectrogram around the boundary performs with a mean accuracy of 0.72 (median area under the receiving operating curve (AUC) = 0.76 logistic regression, 20-fold cross-validated), meaning that more than one in four words are incorrectly segmented on the basis of acoustic cues alone (Fig. 3c).

Given that listeners show enhanced representation of linguistic features relevant for words in their native language (word boundary, length and phoneme-surprisal features; Fig. 2b), we hypothesized that neural activity would also be better able to discriminate between word and syllable boundaries in one's native compared to a foreign language. We first compared neural responses aligned to boundary events in individual electrodes. We found electrodes in STG that showed differential responses to word and syllable boundaries only for native speech for both Spanish and English speakers (Fig. 3d,e). Specifically,

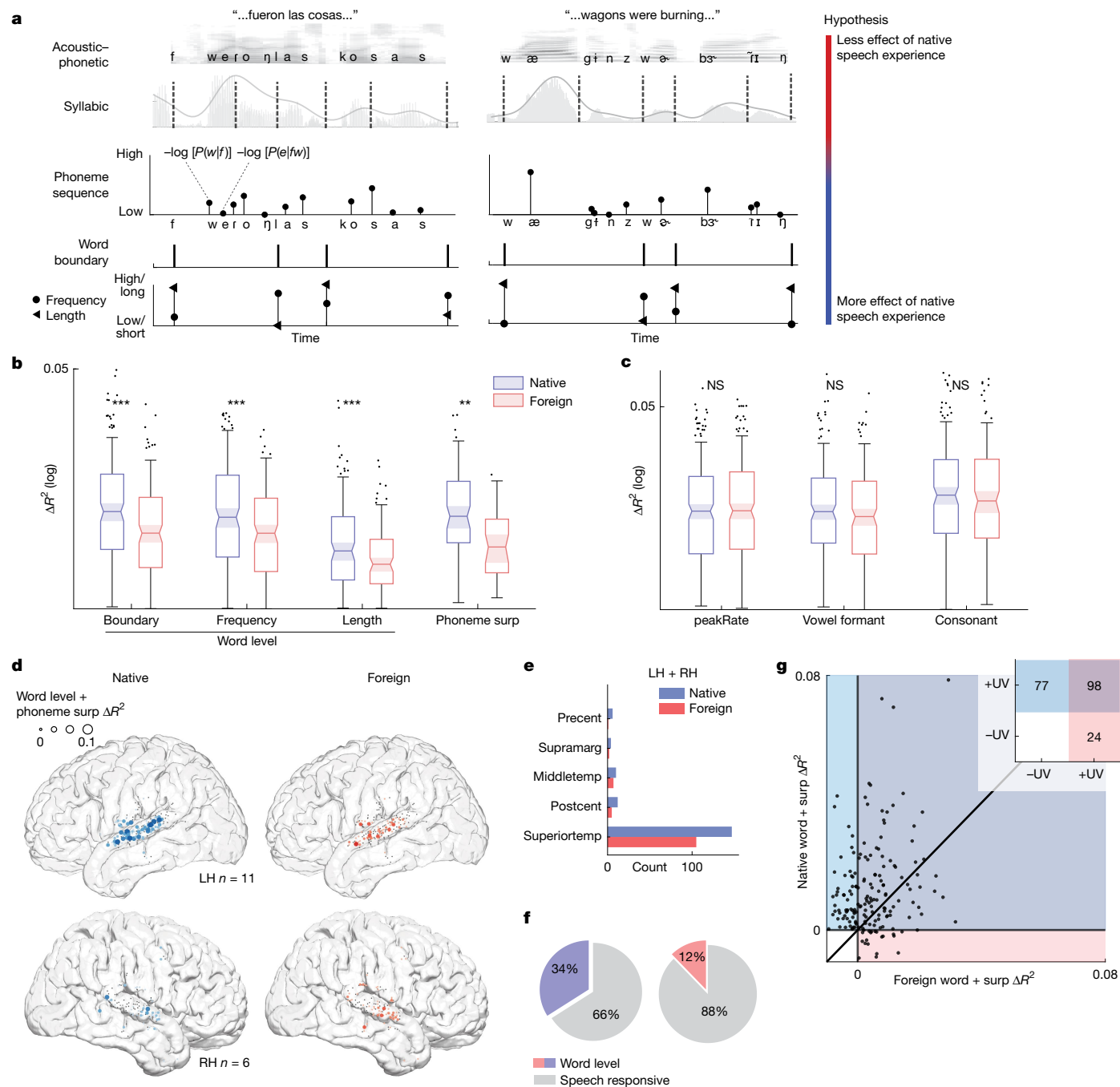


Fig. 2 | Enhanced encoding of word-level and phoneme-surprisal features in native speech. **a**, Spanish (left) and English (right) sentences annotated with phonemes, syllable boundaries, phoneme surprisal, word boundaries, frequencies and lengths. We hypothesized that neural encoding of speech representations requires increasing language experience (colour bar at right). **b**, Across electrodes, the variance in the neural response uniquely explained by word-level features is significantly higher in native versus foreign speech (LME model boundary $***P < 0.001$, frequency $***P < 0.001$, length $***P < 0.001$; see Methods for formula) and phoneme surprisal (LME $**P = 0.0012$). Box plots show non-outlier maximum and minimum values (whiskers), median (centre line) and lower and upper quartiles (box limits) corresponding to electrodes surpassing a significance threshold for model $R^2 > 0.05$ and $\Delta R^2 > 0.001$ (boundary $n = 246$, frequency $n = 256$, length $n = 126$). **c**, Unique variance for acoustic-phonetic features does not significantly differ in native versus foreign speech (LME peakRate $P = 0.41$, formant $P = 0.60$, consonant $P = 0.31$; see Methods for formula). Box plots show non-outlier maximum and minimum values (whiskers), median (centre line) and lower and upper quartiles (box limits)

corresponding to electrodes surpassing a significance threshold for model $R^2 > 0.05$ and $\Delta R^2 > 0.001$ (peakRate $n = 502$; formant $n = 316$; consonant $n = 270$). **d**, Electrodes with positive unique variance for word and phoneme-surprisal features primarily located in mid-STG; native (left) and foreign (right) speech. Electrode size indicates unique variance magnitude. Black scatters correspond to speech-responsive electrodes. **e**, Electrode counts with significant word and phoneme-surprisal feature unique variance (permutation test versus shuffled distribution) across native and foreign speech in each anatomically defined brain region. **f**, Proportion of speech-responsive electrodes with significant word and phoneme-surprisal feature unique variance (permutation test versus shuffled distribution) for speech conditions across hemispheres; native (left) and foreign (right) speech. **g**, Ninety-eight electrodes show positive unique variance for word and phoneme surprisal in both native and foreign languages; unique variance for word and phoneme surprisal is not significantly correlated across languages (Spearman $r(98) = 0.16$, $P = 0.12$). Inset plot shows maximum count in each quadrant of the scatter plot. NS, not significant.

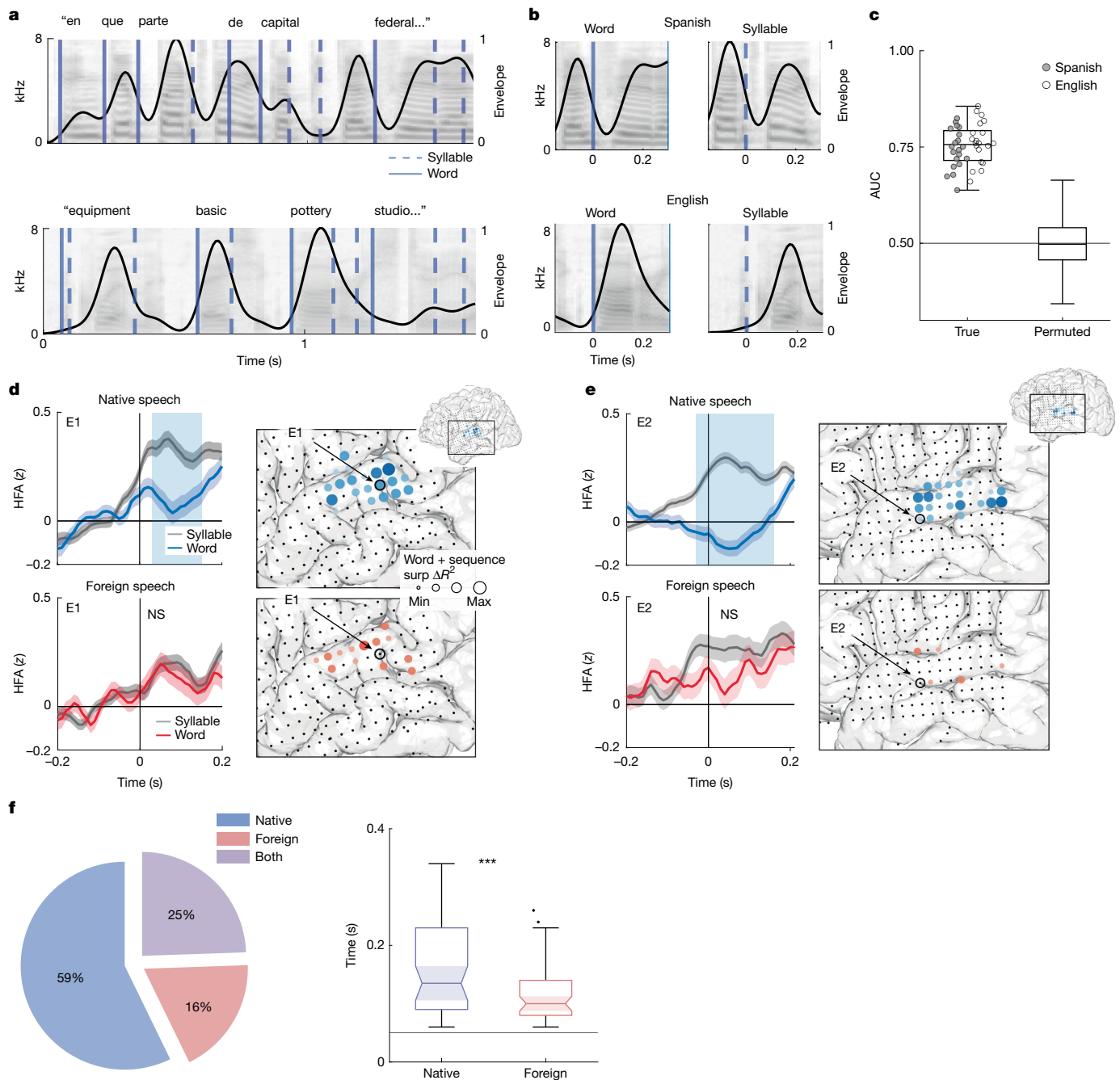


Fig. 3 | Enhanced discrimination between word and syllable boundaries in native speech. **a**, Spectrogram and speech envelope for example Spanish (top) and English (bottom) sentence stimuli. Word and syllable boundaries marked in solid and dashed lines, respectively. **b**, Spectrogram and speech envelope for example word and syllable boundaries are closely matched to one another in Spanish (top) and English (bottom) but differ across languages. Example word and syllable boundaries marked as in **a**, **c**. Acoustic classification of word boundaries using the spectrogram across English and Spanish speech. Each scatter point represents 1/20 folds. Permuted distribution median AUC of 0.50 (25 permutations with 20 folds each for each language). Box plots show maximum and minimum values (whiskers), median (centre line) and 25th to 75th percentiles (box limits). **d**, Left, example electrode (E1) from an English listener differentiates word and syllable boundaries in native (top) but not in foreign (bottom) speech. Rectangular patches indicate time points with a

significant difference (per-time-point two-sided ANOVA $P < 0.01$ with Bonferroni correction). Continuous shaded patches indicate standard error of the mean HFA response across boundaries. Right, anatomical location of E1 and the unique variance for word and phoneme-surprisal features in the example participant. **e**, Left, example electrode (E2) from a Spanish listener differentiates word and syllable boundaries in native (top) but not in foreign (bottom) speech. Rectangular patches indicate time points with a significant difference (per-time-point two-sided ANOVA $P < 0.01$ with Bonferroni correction). Continuous shaded patches indicate standard error of the mean HFA response across boundaries. Right, anatomical location of E2 and unique variance for word and phoneme-surprisal features in the example participant. **f**, Word versus syllable boundary discrimination is significant in more electrodes (left) and for a longer duration of the neural response (right) in native versus foreign speech (LME $***P < 0.001$; see Methods for formula).

in native speech, both example electrodes show a negative deflection in amplitude immediately after a word boundary as compared to the positive peak in amplitude at the syllable boundary⁶¹. Of the

electrodes that discriminated between boundary events (50 ms of contiguous significant difference, $P < 0.01$ with Bonferroni correction), most did so only in native speech (Fig. 3f; 59%). Further, for electrodes

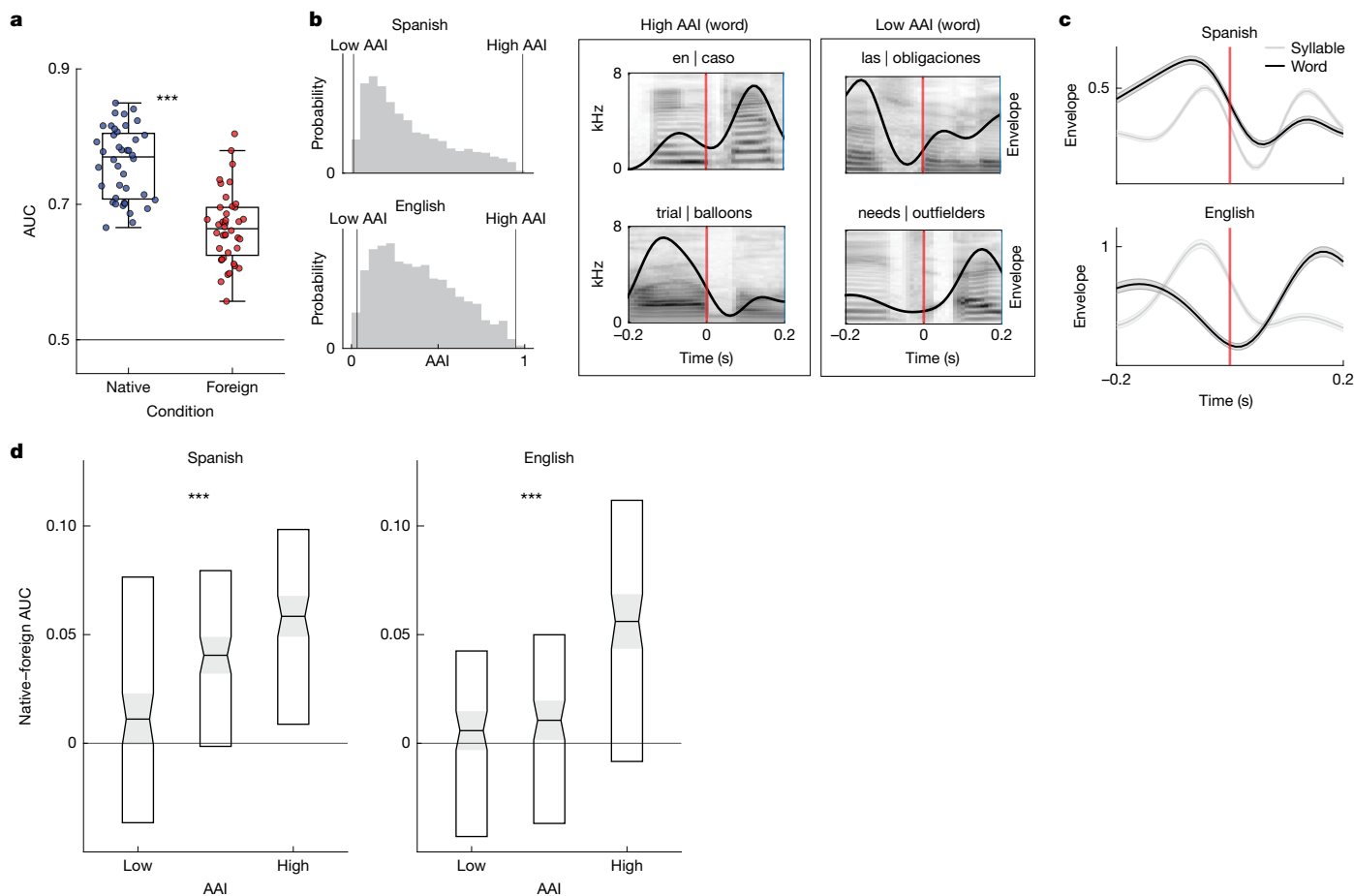


Fig. 4 | Enhanced neural word-boundary decoding in native speech.

a, Neural classification of word boundaries is significantly better in native versus foreign speech (each scatter point corresponds to the AUC of 1/20 folds combined across English and Spanish; one-tailed two-sample *t*-test $***P < 0.001$). Box plots show the maximum and minimum values (whiskers), median (centre line) and 25th to 75th percentiles (box limits). **b**, Left, AAIs span a similar range in Spanish (top) and English (bottom). AAI values were derived from an acoustic word-boundary decoder trained separately on Spanish and English speech stimuli. Right, spectrogram and envelope of high- and low-AAI examples, such as ‘trial | balloons’ and ‘needs | outfielders’, which reflect the acoustics of highly unlikely and likely transitions at a word boundary in American English, respectively. **c**, Low-AAI examples reflect the language-specific

acoustic properties of word and syllable boundaries in Spanish (top) and English (bottom) as demonstrated by the average speech envelope. This resemblance can be seen by comparing low-AAI examples in **b** to the average envelope aligned to the word boundary (black trace). Shaded patches around each trace are the standard error of the envelope across boundary events in the given language. **d**, Native language sensitivity to word boundaries is enhanced for ambiguous word boundaries; Spanish (left) and English (right). The difference between native and foreign neural word-boundary decoding (AUC) increases as AAI increases. Box plots show median (centre line) and 25th to 75th percentiles (box limits), with each distribution corresponding to 225 pairwise AUC difference values generated by random sampling. AUC differences show a significant effect of bin (one-way ANOVA $***P < 0.001$ in both languages).

that discriminated between boundary events in both languages (Fig. 3f; 25%), discrimination occurred for significantly longer in native speech (Fig. 3f; $t(485) = 9.987$, *s.e.* = 0.6955, $P < 0.001$). These results indicate that language experience enhances the brain’s ability to discriminate between word and syllable boundaries in continuous speech, leveraging knowledge about the phonological structure of words to overcome ambiguous acoustic cues.

To quantify the extent to which neural word segmentation was dependent on language experience, we trained a decoder to use ECoG recordings to classify word and syllable boundaries in each participant group and compared the performance (AUC) across cross-validated folds (0.2 s before to 0.4 s after each boundary; 20-fold cross-validated logistic regression). We found that word-boundary decoding performance was significantly higher in native compared to foreign speech (native AUC = 0.77 ± 0.052 ; foreign AUC = 0.66 ± 0.053 ; Fig. 4a). We also found that this effect was consistent in each speech corpus separately and in single participants (Extended Data Fig. 9).

Although word segmentation is challenging without language experience, it is still possible for naive listeners to identify some

boundaries on the basis of salient acoustic cues^{29,62} (Fig. 4b,c and Extended Data Fig. 10). Therefore, we hypothesized that language experience particularly enhances word-boundary decoding for words with unreliable acoustic cues. We characterized each word boundary according to an acoustic ambiguity index (AAI) that measures the extent to which acoustic cues could be used to classify each instance as a word or within-word syllable boundary (see Methods for details; Fig. 4b). We computed AAIs for Spanish and English separately because acoustic and prosodic cues to word boundaries are language specific^{29,63,64} (Fig. 4c and Extended Data Fig. 10). Critically, in both languages, the envelope of high-AAI word-boundary examples (Fig. 4b) strongly resembles the average envelope aligned to within-word syllable boundaries rather than word boundaries (Fig. 4c). This indicates that if unfamiliar listeners were to use average envelope patterns alone to identify word boundaries, they would incorrectly classify the high-AAI boundary trials.

Next, we binned the continuous AAI metric into three discrete bins and compared native versus foreign neural word-boundary decoder performance (decoders from Fig. 4a). To do this, we randomly sampled

trials to generate a distribution of AUC values in each bin and computed the pairwise difference between AUC values across speech conditions (native–foreign). We found that as acoustic cues became more ambiguous (increasing AAI), the difference in AUC between the native and foreign neural decoders increased as well, which we found to be significant using a one-way analysis of variance (ANOVA) (Fig. 4e; Spanish $F(2, 672) = 12.88, P < 0.001$; English $F(2, 672) = 24.91, P < 0.001$). This result demonstrates that enhanced neural decoding of word boundaries in native speech is probably driven not only by sensitivity to the acoustic–phonetic cues to word boundaries but also by cues that draw on language experience and are not explicit in the acoustic signal.

STG identifies words across languages

The vast majority of the world's population is proficient in more than one language. Although it is known that in bilingual individuals, both familiar languages evoke neural activity in the same core set of brain regions^{65,66}, it remains unclear what this activity reflects. Here we took advantage of rare opportunities to record from bilingual participants while they listened to continuous speech in both of their familiar languages. We specifically asked the extent to which encoding of both acoustic–phonetic and word-level information is the same for both languages and how these representations differ as a function of second-language proficiency.

First, we focused on recordings from eight Spanish–English bilingual participants (five left hemisphere, three female) who were highly proficient in both languages (see Extended Data Table 3 for language profiles and Extended Data Fig. 2 for coverage). Consistent with the results from participants familiar with only one language (Fig. 1), we found that similar anatomical regions were active regardless of whether bilingual participants listened to English or Spanish, with the highest proportion of electrodes located in the STG (Fig. 5a). Activity in this region could be explained by encoding of acoustic–phonetic speech features (Fig. 5b), which was significantly correlated across languages (64 of 67 electrodes, 95.2%) and had TRF feature weight profiles that were correlated above chance ($P < 0.05$ compared to uncorrected non-parametric permuted distribution generated from shuffling TRF weight labels; Fig. 5c). Thus, in the brain of a bilingual participant, encoding of spectrotemporal speech content in the STG is substantially similar in multiple languages (Fig. 5b,c).

Above, we showed that language knowledge most prominently affects representations of word-level and phoneme-sequence speech information. Because Spanish–English bilingual participants are familiar with the statistics of Spanish and English, we hypothesized that neural activity would reflect encoding of this information in both languages, unlike in the monolingual participants (Fig. 2). Using a LME model, we found that word-boundary, frequency, length and phoneme-surprisal unique variance was not significantly different between the two languages in bilingual participants (Fig. 5d) (boundary $t(163) = -0.05, \text{s.e.} = 0.0009, P = 0.96$; frequency $t(187) = -1.93, \text{s.e.} = 0.0008, P = 0.055$; length $t(177) = -1.20, \text{s.e.} = 0.0008, P = 0.23$; surprisal $t(73) = 0.76, \text{s.e.} = 0.0011, P = 0.45$). Furthermore, encoding of word-level and phoneme-surprisal features was not significantly different in bilingual participants as compared to the native language condition for monolingual participants (Fig. 4e) (boundary $t(321) = -1.28, \text{s.e.} = 0.0010, P = 0.20$; frequency $t(346) = 0.23, \text{s.e.} = 0.0010, P = 0.82$; length $t(318) = 1.37, \text{s.e.} = 0.0072, P = 0.17$; surprisal $t(144) = -1.09, \text{s.e.} = 0.0010, P = 0.28$). Together, these results demonstrate that language experience—even for multiple languages in a single brain—modulates the extent to which word-level and phoneme-surprisal features affect neural representations in the temporal lobe.

Bilingual participants provide a unique opportunity to determine whether word-level and phoneme-surprisal encoding electrodes are language specific—that is, whether electrodes encode these features for

only one of the two familiar languages. In contrast to monolingual participants (Fig. 3), in bilingual participants, we found STG electrodes that strongly differentiated word and syllable boundaries in both Spanish and English (Fig. 5f). To test for language specificity across electrodes, we compared the corresponding unique variances for word-level and phoneme-surprisal features across Spanish and English. We found that a large proportion of electrodes encoded word-level and surprisal features for both English and Spanish, similar to results reported for monolingual participants in Fig. 2. However, unlike the monolingual participants, the unique variances across languages were significantly correlated between languages (Fig. 5g; Spearman $r(45) = 0.31, P = 0.03$). This result demonstrates that word and phonological sequence information can be represented in the same populations across languages and implicates a shared mechanism for encoding phonological structure across languages in the bilingual brain.

Finally, although the key results thus far have been demonstrated with three languages that represent some of the linguistic diversity that exists across countries and cultures (English, Spanish and Mandarin), there are many other languages and language families acquired as native languages throughout the world. Over more than a decade, we sampled some of these languages by collecting data from participants who were speakers of Russian, Arabic and Korean (representing Indo-European, Semitic and Koreanic language families; Fig. 5h). Furthermore, these participants had varying degrees of self-reported English proficiency (see Extended Data Table 4 for language profiles), allowing us to test the effects of experience across a wider array of known languages that differ from English to varying degrees. Given that the ability to identify word boundaries is one of the key phonological markers of language knowledge⁶⁷, we asked the extent to which neural English word-boundary decoding accuracy is graded as a function of English proficiency across these participants. Using a LME model, we found that AUC was significantly correlated with English proficiency (Fig. 5i; $t(253) = 4.58, \text{s.e.} = 0.002, P < 0.001$), indicating that the brain's ability to extract word-level structure from continuous speech depends on listeners' overall knowledge of the language. Although these results arise from limited data, the fact that proficiency rather than typology of the native language (Russian, Korean and so on) predicts neural word-boundary decoding in English indicates that neural encoding of word boundaries may be a general speech processing mechanism that arises independently for each language that an individual learns.

Discussion

One of the remarkable attributes of the human brain is the ability to master the complexities of spoken language in just the first few years of life. Yet this skill does not transfer to all languages—we only understand languages to which we have been sufficiently exposed. This fact is underscored when we listen to a foreign language; although we may be able to identify individual speech sounds that are similar to the sounds of languages we know, it is substantially more challenging to parse the continuous input into larger chunks like words. To date, it has been unclear what drives this gap in our perception of native and foreign languages.

In this study, we leveraged high-resolution direct brain recordings to provide a detailed picture of the neural processes that reflect shared auditory phonological processes and those that are influenced by language experience. Owing to shared human anatomy, the same fundamental articulatory features make up speech in all languages⁶⁸: features the human brain is readily able to perceive. However, with exposure that begins before birth, the brain becomes exquisitely tuned to the fine-grained, specific characteristics of the native language, including the specific inventory of speech sounds used by that language^{78,69–72}. Throughout development, listeners become sensitive to how these sounds are combined to form the words that make up the language^{10,11,73},

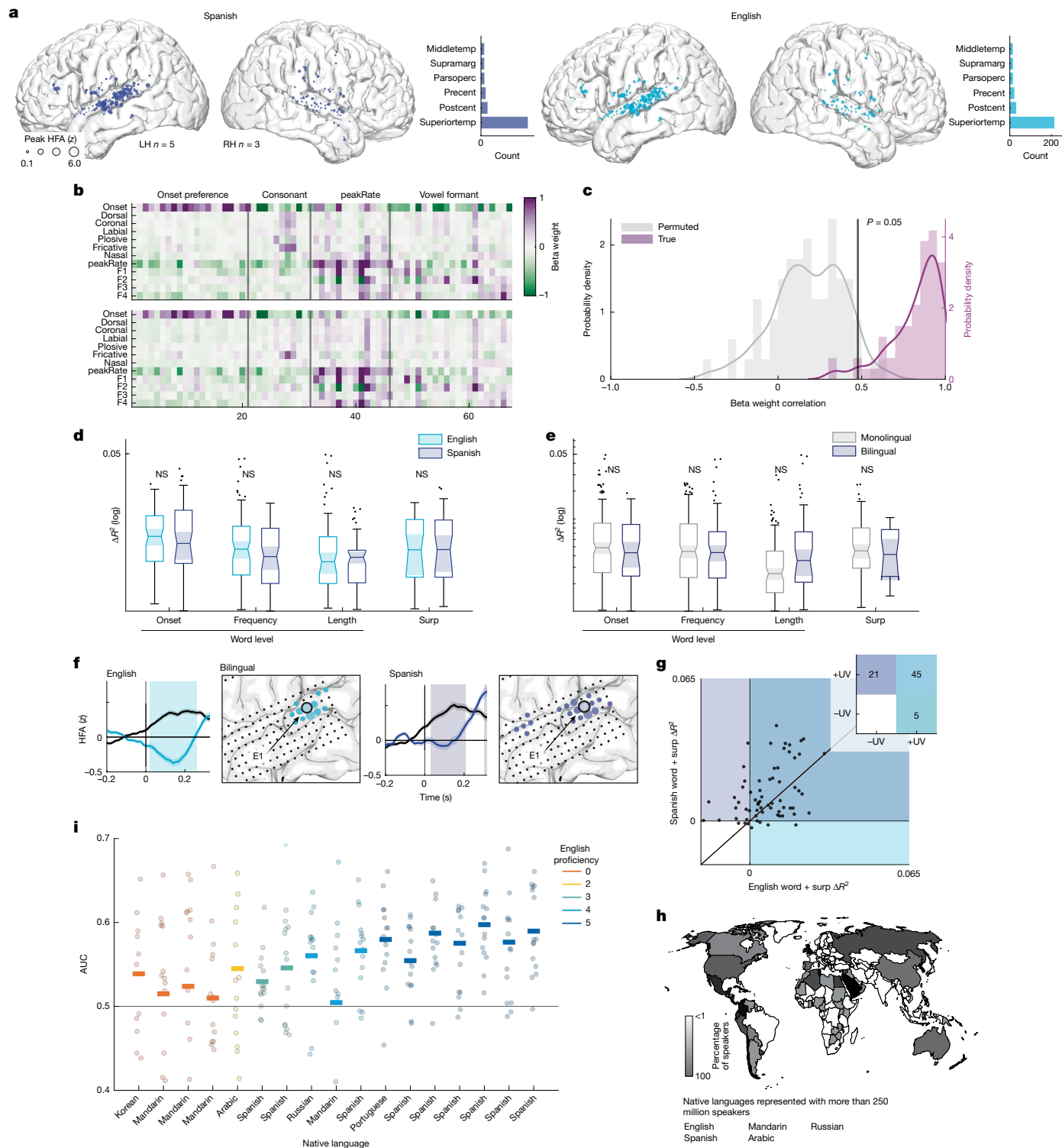


Fig. 5 | See next page for caption.

which are also specific to each of the world's 7,000 languages. Here we demonstrate that language experience more strongly affects this latter ability, with enhanced neural encoding of word-level features associated only with the listener's native language.

Our results showing shared neural representations of speech in the human temporal cortex confirm its role as a processor of fundamental speech features², independent of language experience. One surprising aspect of our results is the finding that experience-dependent neural activity also exists in these brain regions that process both native and

foreign language input. This challenges the prevailing view that language processing only happens in a network of cortical regions that are selective for comprehensible speech²⁴. Theoretical models of speech processing posit that auditory regions like the STG perform surface-level spectrotemporal extraction before sound input is linguistically transformed in higher-order cortical regions^{13,74,75}. Under this view, the STG is a language-agnostic speech processor and therefore discriminates between languages as a consequence of distinct acoustic properties but not language experience of the listener. By contrast,

Fig. 5 | Bilingual listeners show neural encoding of word-level features and phoneme surprisal for both familiar languages. **a**, Speech-responsive electrodes across eight Spanish–English bilingual participants show similar STG localization for Spanish (left) and English (right) speech across hemispheres. Electrode size indicates magnitude of the peak HFA averaged sentence response. Inset histograms show electrode counts across English and Spanish speech in each anatomical region. **b**, TRF weights across speech-responsive electrodes for English (top) and Spanish (bottom) are significantly correlated (Pearson $r(804) = 0.88$, $P < 0.001$). Top, columns (electrodes) ordered by the acoustic–phonetic feature showing the maximal weight. Bottom, columns are ordered identical to top for visual comparison. **c**, Distribution of TRF Spanish–English weight correlations across electrodes (purple) is significantly higher compared to the non-parametric permuted distribution (grey). Black vertical line indicates the 95th percentile in the permuted distribution. **d**, Encoding of word-level and phoneme-surprisal features is not significantly different across Spanish and English (LME speech condition $P > 0.05$; see Methods for formula). **e**, Encoding of word-level and phoneme-surprisal features is not significantly

different between bilingual and the native language for monolingual participants (LME speech condition $P > 0.05$; see Methods for formula). **f**, Example electrode from a Spanish–English bilingual participant differentiates word and syllable boundaries in both the English speech condition (left) and Spanish speech (right). **g**, Forty-five electrodes show positive unique variance for word-level and phoneme surprisal in English and Spanish. Unique variance for word-level features across languages is significantly correlated (Spearman $r(45) = 0.31$, $P = 0.03$). Inset plot shows the count of electrodes in each quadrant of the scatter plot. **h**, World map indicating language backgrounds represented in the participant cohort and extent to which these languages are spoken across the world. Maps generated using MathWorks mapping toolbox and information from the World Factbook¹⁰⁵. **i**, Neural word-boundary classification AUC in English scales with English language proficiency for bilingual speakers of Spanish–English and speakers of diverse languages (LME proficiency $P < 0.001$; see Methods for formula). Each line corresponds to the mean AUC per participant; each scatter point is AUC of 1/15 folds.

we find evidence that representations of acoustic–phonetic, sublexical and even word-level information are intermixed throughout the STG, with neural populations and representations showing varying levels of sensitivity to language experience. Ultimately, the current results are consistent with a view of the STG as a high-order auditory region, acting as an interface between speech input and linguistic knowledge^{3,76,77}.

The present results demonstrate how language knowledge affects neural populations in the temporal lobe that are classically thought to process lower-level acoustic information^{13,24}. These findings do not fully disambiguate the extent to which language experience-dependent knowledge is encoded elsewhere in the brain and exerts a top-down modulatory effect on STG^{78,79}. However, the fact that the strongest and most prevalent encoding of language-specific word-level information is in the STG indicates that these neural populations and underlying circuitry may be intrinsically sensitive to this type of learned knowledge³.

The present study is limited in that we did not sample from all brain regions potentially relevant to language experience. In particular, these results are based on ECoG responses recorded from the lateral surface that did not allow for the examination of the superior temporal plane such as the primary auditory cortex on Heschl’s gyrus, the superior temporal sulcus or subcortical structures that can be structurally and functionally affected by language experience^{12,74,80–82}. Previous recordings from Heschl’s gyrus with native English speakers listening to English indicate that these neural populations may be less sensitive to the acoustic–phonetic and word-level features examined here¹². It also remains to be seen whether our findings extend to learned non-speech sound stimuli, such as music^{83–85}.

These findings have implications for understanding how language comprehension abilities arise during development and for explaining the challenges associated with learning new languages later in life. Although the adult brain retains an ability to learn new linguistic information^{56,86–89}, achieving native-like proficiency with the phonology of a language after development can be difficult and depends on several different usage factors^{90–93}. Our results indicate that this challenge is in part related to learning how to parse sequences of individual speech sounds into language-specific units such as words.

In bilingual participants, neural encoding of language-specific knowledge (English, Spanish) is localized to the same brain regions, the STG, and in some cases even occurs in the same neural populations (Fig. 5g). These findings are consistent with theoretical models that posit a single and highly overlapping brain network in which multiple languages are processed^{94–99}. Although language-specific neural activation and impairments post brain injury have been found, greater neural differences between languages often correspond to greater disparities in skill level^{100,101}, consistent with our finding that in multilinguals, English

proficiency modulates the extent to which language-specific knowledge is neurally encoded^{102–104} (Fig. 5i).

Studying speech and language in the human brain has traditionally been limited to a small number of languages and typically with native speakers of those languages. Although more work is needed to survey the vast linguistic diversity that exists across the world’s languages, this work shows that broadening the scope of enquiry to a diversity of cultural and experiential differences can give critical insights into fundamental mechanisms of neural coding for communication.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-09748-8>.

- Evans, N. & Levinson, S. C. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–448 (2009).
- Mesgarani, N., Cheung, C., Johnson, K. & Chang, E. F. Phonetic feature encoding in human superior temporal gyrus. *Science* **343**, 1006–1010 (2014).
- Yi, H. G., Leonard, M. K. & Chang, E. F. The encoding of speech sounds in the superior temporal gyrus. *Neuron* **102**, 1096–1110 (2019).
- Cutler, A. *Native Listening* (MIT, 2015).
- Bosker, H. R. & Reinisch, E. Foreign languages sound fast: evidence from implicit rate normalization. *Front. Psychol.* **8**, 1063 (2017).
- Cutler, A., Mehler, J., Norris, D. & Segui, J. A language-specific comprehension strategy. *Nature* **304**, 159–160 (1983).
- Kuhl, P. K. et al. Effects of language experience on speech perception: American and Japanese infants’ perception of /ra/ and /la/. *J. Acoust. Soc. Am.* **102**, 3135–3136 (1997).
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255**, 606–608 (1992).
- Saffran, J. R., Newport, E. L. & Aslin, R. N. Word segmentation: the role of distributional cues. *J. Mem. Lang.* **35**, 606–621 (1996).
- Stager, C. L. & Werker, J. F. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* **388**, 381–382 (1997).
- Werker, J. F. & Tees, R. C. Influences on infant speech processing: toward a new synthesis. *Annu. Rev. Psychol.* **50**, 509–535 (1999).
- Hamilton, L. S., Oganian, Y., Hall, J. & Chang, E. F. Parallel and distributed encoding of speech across human auditory cortex. *Cell* **184**, 4626–4639 (2021).
- Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
- Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).
- Malik-Moraleda, S. et al. An investigation across 45 languages and 12 language families reveals a universal language network. *Nat. Neurosci.* **25**, 1014–1019 (2022).
- Lipkin, B. et al. Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. *Sci. Data* **9**, 529 (2022).
- Schlosser, M. J., Aoyagi, N., Fulbright, R. K., Gore, J. C. & McCarthy, G. Functional MRI studies of auditory comprehension. *Hum. Brain Mapp.* **6**, 1–13 (1998).
- Mazoyer, B. M. et al. The cortical representation of speech. *J. Cogn. Neurosci.* **5**, 467–479 (1993).

19. Binder, J. R. et al. Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* **10**, 512–528 (2000).
20. Narain, C. et al. Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. Cortex* **13**, 1362–1368 (2003).
21. Norman-Haignere, S. V. & McDermott, J. H. Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLoS Biol.* **16**, e2005127 (2018).
22. Binder, J. R. et al. Functional magnetic resonance imaging of human auditory cortex. *Ann. Neurol.* **35**, 662–672 (1994).
23. Démonet, J.-F. et al. The anatomy of phonological and semantic processing in normal subjects. *Brain* **115**, 1753–1768 (1992).
24. Fedorenko, E., Ivanova, A. A. & Regev, T. I. The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* **25**, 289–312 (2024).
25. Chang, E. F. et al. Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* **13**, 1428–1432 (2010).
26. Fox, N. P., Leonard, M., Sjerps, M. J. & Chang, E. F. Transformation of a temporal speech cue to a spatial neural code in human auditory cortex. *eLife* **9**, e53051 (2020).
27. Li, Y., Tang, C., Lu, J., Wu, J. & Chang, E. F. Human cortical encoding of pitch in tonal and non-tonal languages. *Nat. Commun.* **12**, 1–12 (2021).
28. Li, Y. et al. Dissecting neural computations in the human auditory pathway using deep neural networks for speech. *Nat. Neurosci.* **26**, 2213–2225 (2023).
29. Cutler, A. in *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (ed. Altmann, G. T. M.) Ch. 5 (MIT, 1990).
30. Pineda, L. A., Pineda, L. V., Cuétara, J., Castellanos, H. & López, I. DIMEX100: a new phonetic and speech corpus for Mexican Spanish. In *Proc. 9th Ibero-American Conference on AI* (eds Lemaitre, C. et al.) 974–983 (Springer, 2004).
31. Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G. & Pallett, D. S. *DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM. NIST Speech Disc 1-1.1* (1993); <https://ui.adsabs.harvard.edu/abs/1993STIN...9327403G>.
32. Li, A. et al. Speech corpus of Chinese discourse and the phonetic research. In *Proc. 6th International Conference on Spoken Language Processing* vol. 4, 13–18 (ICSLP, 2000).
33. Hamilton, L. S., Edwards, E. & Chang, E. F. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr. Biol.* **28**, 1860–1871.e4 (2018).
34. Hualde, J. I. *The Sounds of Spanish* (Cambridge Univ. Press, 2005).
35. Fox, R. A., Flege, J. E. & Munro, M. J. The perception of English and Spanish vowels by native English and Spanish listeners: a multidimensional scaling analysis. *J. Acoust. Soc. Am.* **97**, 2540–2551 (1995).
36. Flege, J. E. & Eefting, W. Imitation of a VOT continuum by native speakers of English and Spanish: evidence for phonetic category formation. *J. Acoust. Soc. Am.* **83**, 729–740 (1988).
37. Theunissen, F. E. et al. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* **12**, 289–316 (2001).
38. Oganian, Y. & Chang, E. F. A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Sci. Adv.* **5**, eaay6279 (2019).
39. Abramson, A. S. & Lisker, L. Perception of voice timing in Spanish stop consonants. *J. Acoust. Soc. Am.* **52**, 175–175 (1972).
40. Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* **54**, 358–368 (1957).
41. Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
42. Näätänen, R. et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385**, 432–434 (1997).
43. Bradlow, A. R. A comparative acoustic study of English and Spanish vowels. *J. Acoust. Soc. Am.* **97**, 1916–1924 (1995).
44. Cole, R. A., Jakimik, J. & Cooper, W. E. Segmenting speech into words. *J. Acoust. Soc. Am.* **67**, 1323–1332 (1980).
45. Cutler, A., Dahan, D. & van Donselaar, W. Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* **40**, 141–201 (1997).
46. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
47. Cuetos, F., Hallé, P. A., Domínguez, A. & Segui, J. Perception of prothetic /e/ in #sC utterances: gating data. In *Proc. 17th International Congress of Phonetic Sciences* (eds Lee, W.-S. & Zee, E.) 540–543 (2011).
48. Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S. & Dupoux, E. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.* **23**, 9541–9546 (2003).
49. Turk, A. E. & Shattuck-Hufnagel, S. Word-boundary-related duration patterns in English. *J. Phon.* **28**, 397–440 (2000).
50. Shatzman, K. B. & McQueen, J. M. Segment duration as a cue to word boundaries in spoken word recognition. *Percept. Psychophys.* **68**, 1–16 (2006).
51. Cuetos, F., Glez-Nosti, M., Barbon, A. & Brysbaert, M. SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica* **32**, 133–143 (2011).
52. van Heuven, W. J. B., Mandera, P., Keuleers, E. & Brysbaert, M. SUBTLEX-UK: a new and improved word frequency database for British English. *Q. J. Exp. Psychol.* **67**, 1176–1190 (2014).
53. Gwilliams, L. & Davis, M. H. in *Speech Perception* (eds Holt, L. L. et al.) 113–139 (Springer International, 2022).
54. Leonard, M. K. et al. Large-scale single-neuron speech sound encoding across the depth of human cortex. *Nature* **626**, 593–602 (2024).
55. Tang, C., Hamilton, L. S. & Chang, E. F. Intonational speech prosody encoding in the human auditory cortex. *Science* **357**, 797–801 (2017).
56. Yi, H. G. et al. Learning nonnative speech sounds changes local encoding in the adult human cortex. *Proc. Natl Acad. Sci. USA* **118**, e2101777118 (2021).
57. Lehiste, I. The timing of utterances and linguistic boundaries. *J. Acoust. Soc. Am.* **51**, 2018–2024 (1972).
58. Cutler, A. & Norris, D. The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* **14**, 157–177 (1988).
59. Cutler, A. in *The Handbook of English Pronunciation* (eds Reed, M. & Levis, J. M.) Ch. 6 (Wiley, 2015).
60. Cutler, A. & Carter, D. M. The predominance of strong initial syllables in the English vocabulary. *Comput. Speech Lang.* **2**, 133–142 (1987).
61. Zhang, Y., Leonard, M. K., Bhaya-Grossman, I., Gwilliams, L. & Chang, E. F. Human cortical dynamics of auditory word form encoding. *Neuron* <https://doi.org/10.1016/j.neuron.2025.10.011> (2025).
62. Endress, A. D. & Hauser, M. D. Word segmentation with universal prosodic cues. *Cogn. Psychol.* **61**, 177–199 (2010).
63. Tyler, M. D. & Cutler, A. Cross-language differences in cue use for speech segmentation. *J. Acoust. Soc. Am.* **126**, 367–376 (2009).
64. Cutler, A., Mehler, J., Norris, D. & Segui, J. The syllable's differing role in the segmentation of French and English. *J. Mem. Lang.* **25**, 385–400 (1986).
65. Hesling, I., Dilharreguy, B., Bordesoules, M. & Allard, M. The neural processing of second language comprehension modulated by the degree of proficiency: a listening connected speech fMRI study. *Open Neuroimag. J.* **6**, 44–54 (2012).
66. Cargnelutti, E., Tomasino, B. & Fabbro, F. Language brain representation in bilinguals with different age of appropriation and proficiency of the second language: a meta-analysis of functional imaging studies. *Front. Neurosci.* **13**, 154 (2019).
67. Saffran, J. R., Werker, J. F. & Werner, L. A. in *Handbook of Child Psychology: Cognition, Perception, and Language*, Vol. 2 (eds Kuhn, D. & Siegler, R. S.) Ch. 2 (Wiley, 2006).
68. Ladefoged, P. & Johnson, K. *A Course in Phonetics* (Nelson Education, 2014).
69. Di Liberto, G. M. et al. Emergence of the cortical encoding of phonetic features in the first year of life. *Nat. Commun.* **14**, 7789 (2023).
70. McCarthy, K. M., Skoruppa, K. & Iverson, P. Development of neural perceptual vowel spaces during the first year of life. *Sci. Rep.* **9**, 19592 (2019).
71. Kuhl, P. K. et al. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* **9**, F13–F21 (2006).
72. Werker, J. F. & Lalonde, C. E. Cross-language speech perception: initial capabilities and developmental change. *Dev. Psychol.* **24**, 672–683 (1988).
73. Jusczyk, P. W. & Hohne, E. A. Infants' memory for spoken words. *Science* **277**, 1984–1986 (1997).
74. Wilson, S. M., Bautista, A. & McCarron, A. Convergence of spoken and written language processing in the superior temporal sulcus. *Neuroimage* **171**, 62–74 (2018).
75. Davis, M. H. in *Neurobiology of Language* (eds Hickok, G. & Small, S. L.) Ch. 44 (Elsevier, 2016).
76. Bhaya-Grossman, I. & Chang, E. F. Speech computations of the human superior temporal gyrus. *Annu. Rev. Psychol.* **73**, 79–102 (2022).
77. Leonard, M. K. & Chang, E. F. Dynamic speech representations in the human temporal lobe. *Trends Cogn. Sci.* **18**, 472–479 (2014).
78. Myers, E. B. & Blumstein, S. E. The neural bases of the lexical effect: an fMRI investigation. *Cereb. Cortex* **18**, 278–288 (2008).
79. Davis, M. H. & Johnsrude, I. S. Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* **229**, 132–147 (2007).
80. Zhao, T. C. & Kuhl, P. K. Linguistic effect on speech perception observed at the brainstem. *Proc. Natl Acad. Sci. USA* **115**, 8716–8721 (2018).
81. Beguš, G., Zhou, A. & Zhao, T. C. Encoding of speech in convolutional layers and the brain stem based on language experience. *Sci. Rep.* **13**, 6480 (2023).
82. Kepinska, O. et al. Auditory cortex anatomy reflects multilingual phonological experience. *eLife* **12**, RP90269 (2025).
83. Sankaran, N., Leonard, M. K., Theunissen, F. & Chang, E. F. Encoding of melody in the human auditory cortex. *Sci. Adv.* **10**, eadk0010 (2024).
84. Norman-Haignere, S. V. et al. A neural population selective for song in human auditory cortex. *Curr. Biol.* **32**, 1470–1484.e12 (2022).
85. Boebinger, D., Norman-Haignere, S. V., McDermott, J. H. & Kanwisher, N. Music-selective neural populations arise without musical training. *J. Neurophysiol.* **125**, 2237–2263 (2021).
86. Wang, Y., Spence, M. M., Jongman, A. & Sereno, J. A. Training American listeners to perceive Mandarin tones: transfer to production. *J. Acoust. Soc. Am.* **105**, 1095–1095 (1999).
87. Golestani, N. & Zatorre, R. J. Learning new sounds of speech: reallocation of neural substrates. *Neuroimage* **21**, 494–506 (2004).
88. Golestani, N. & Zatorre, R. J. Individual differences in the acquisition of second language phonology. *Brain Lang.* **109**, 55–67 (2009).
89. Myers, E. B. Emergence of category-level sensitivities in non-native speech sound learning. *Front. Neurosci.* **8**, 238 (2014).
90. Eckman, F. R. in *Phonology and Second Language Acquisition* Vol. 36 (eds Hansen Edwards, J. G. & Zampini, M. L.) Ch. 4 (John Benjamins, 2008).
91. Strange, W. & Shafer, V. L. in *Phonology and Second Language Acquisition* Vol. 36 (eds Hansen Edwards, J. G. & Zampini, M. L.) Ch. 6 (John Benjamins, 2008).
92. Major, R. C. in *Phonology and Second Language Acquisition* Vol. 36 (eds Hansen Edwards, J. G. & Zampini, M. L.) Ch. 3 (John Benjamins, 2008).
93. Archibald, J. Ease and difficulty in L2 phonology: a mini-review. *Front. Commun.* **6**, 626529 (2021).
94. Abutaleb, J. & Green, D. Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* **20**, 242–275 (2007).
95. Green, D. W. in *The Lexicon-Syntax Interface in Second Language Acquisition* (eds Hout, R. et al.) Ch. 9 (John Benjamins, 2003).
96. Golestani, N. Neuroimaging of phonetic perception in bilinguals. *Biling.* **19**, 674–682 (2016).
97. Potczyńska, M. M. & Bookheimer, S. Y. General principles governing the amount of neuroanatomical overlap between languages in bilinguals. *Neurosci. Biobehav. Rev.* **130**, 1–14 (2021).

98. Kim, S. Y., Liu, L., Liu, L. & Cao, F. Neural representational similarity between L1 and L2 in spoken and written language processing. *Hum. Brain Mapp.* **41**, 4935–4951 (2020).
99. Abutalebi, J., Cappa, S. F. & Perani, D. The bilingual brain as revealed by functional neuroimaging. *Biling.* **4**, 179–190 (2001).
100. Kuzmina, E., Goral, M., Norvik, M. & Weekes, B. S. What influences language impairment in bilingual aphasia? A meta-analytic review. *Front. Psychol.* **10**, 445 (2019).
101. Sulpizio, S., Del Maschio, N., Fedeli, D. & Abutalebi, J. Bilingual language processing: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* **108**, 834–853 (2020).
102. de Bruin, A. Not all bilinguals are the same: a call for more detailed assessments and descriptions of bilingual experiences. *Behav. Sci.* **9**, 33 (2019).
103. Kaushanskaya, M., Blumenfeld, H. K. & Marian, V. The Language Experience and Proficiency Questionnaire (LEAP-Q): ten years later. *Biling.* **23**, 945–950 (2020).
104. Bice, K. & Kroll, J. F. English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain Lang.* **196**, 104644 (2019).
105. Central Intelligence Agency. *The CIA World Factbook 2024–2025* (Skyhorse, 2024).

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



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

© The Author(s) 2025

Article

Methods

Data acquisition followed procedures similar to those conducted in previous work^{12,38,106}.

Participants

Thirty-four patients (17 female) were implanted subdural ECoG grids (Integra or PMT) with 4-mm centre-to-centre electrode spacing and 1.17-mm-diameter contacts (recording tens to hundreds of thousands of neurons^{107,108}) as part of their neurosurgical treatment at either University of California, San Francisco (UCSF) or Huashan Hospital. All participants gave informed written consent to participate in the study before experimental testing. Most electrode grids were placed over the lateral surface of a single hemisphere and were centred around the temporal lobe extending to adjacent cortical areas. The precise location of electrode placement was determined by clinical assessment.

For all participants, age, gender, language background information and which hemisphere was recorded from is included in Extended Data Tables 1–4, where each table corresponds to the respective participant group: English monolingual, Spanish and Mandarin monolingual, Spanish–English bilingual and others with diverse language backgrounds. Which participants were included in each set of analyses is listed in Extended Data Table 5. ECoG recordings with the four Mandarin speakers were conducted at Huashan Hospital while patients underwent awake language mapping as part of their surgical brain-tumour treatment. ECoG recordings with all other participants were conducted at UCSF while patients underwent clinical monitoring for seizure activity as part of their surgical treatment for intractable epilepsy. Most participants in this dataset had epileptic seizure foci that were located in deep medial structures (for example, the insula) or far anterior temporal lobe outside of the main regions of interest in this study, such as the STG. All participants included in this study reported normal hearing and spoken language abilities.

Participant consent

All protocols in the current study were approved by the UCSF Committee on Human Research and by the Huashan Hospital Institutional Review Board of Fudan University. Participants gave informed written consent to take part in the experiments and for their data to be analysed. Informed consent of non-English-speaking participants at UCSF was acquired using a medically certified interpreter platform (Language-Line Solutions), and communication with research staff was facilitated by either in-person or video-call-based interpreters who were fluent in the participant's native language.

Language questionnaire

All participants were asked to self-report speech comprehension proficiency, age of acquisition and frequency of use for all languages that they were familiar with. Participants who self-identified as Spanish–English bilingual were asked to complete a comprehensive language questionnaire by means of an online Qualtrics survey¹⁰³.

Neural data acquisition

ECoG signals were recorded with a multichannel PZ5 amplifier connected to an RZ2 digital signal acquisition system (TuckerDavis Technologies (TDT)) with a sampling rate of 3 kHz. During the stimulus presentation, the audio signal was recorded in the TDT circuit and therefore time-aligned with the ECoG signal. Audio stimulus was also recorded with a microphone, and this signal was also recorded in the TDT circuit to ensure accurate time alignment.

Data preprocessing

Offline preprocessing of the data included downsampling to 400 Hz, notch-filtering of line noise (at 60 Hz, 120 Hz and 180 Hz for recording at UCSF and 50 Hz, 100 Hz and 150 Hz for recording at Huashan Hospital),

extracting the analytic amplitude in the high-gamma frequency range (70–150 Hz, HFA) and excluding extended interictal spiking or otherwise noisy channel activity through manual inspection. The Hilbert transform was used to extract HFA using eight band-pass Gaussian filters with logarithmically increasing centre frequencies (70–150 Hz) and semilogarithmically increasing bandwidths. High-gamma amplitude was calculated as in previous work^{12,106} as the first principal component of the signal in each electrode across all eight high-gamma bands, using principal components analysis (PCA). Last, the HFA was downsampled to 100 Hz and Z-scored relative to the mean and s.d. of the neural data in the experimental block. Each sentence HFA was normalized relative to the 0.5 s of silent prestimulus baseline. All subsequent analyses were based on the resulting neural time series.

Electrode localization

For anatomical localization, electrode locations were extracted from postimplantation computer tomography scans, coregistered to the patients' structural magnetic resonance imaging and superimposed on three-dimensional reconstructions of the patients' cortical surfaces using an in-house imaging pipeline validated in previous work¹⁰⁹. FreeSurfer (<https://surfer.nmr.mgh.harvard.edu/>) was used to create a three-dimensional model of the individual participant's pial surfaces, run automatic parcellation to get individual anatomical labels and warp the individual participant surfaces into the `cvs_avg35_inMNI152` average template.

Stimuli and procedure

All participants passively listened to roughly 30 min of speech in a language they knew (either Spanish, English or Mandarin Chinese) and one other language (either Spanish or English). Speech stimuli consisted of a selection of 239 unique Spanish sentences from the DIMEX corpus^{30,110}, spoken by a variety of native Mexican-Spanish speakers; 499 unique English sentences from the TIMIT corpus, spoken by a variety of American English speakers³¹; and 58 unique Mandarin Chinese paragraphs from the ASCCD corpus from the Chinese Linguistic Data Consortium (www.chineseldc.org/), spoken by a variety of Mandarin Chinese speakers. Although the total duration of stimuli presentation was roughly equal across languages, statistical tests that were conducted at the sentence level across corpora accounted for the differing number of sentences.

We selected these specific speech corpora, originally designed to test speech recognition systems, to meet the following requirements: (1) comprehensively span the phonemic inventories of the languages in question; (2) provide validated phonemic, phonetic and word-level annotations; and (3) include a diverse and wide range of speakers with varying accents. These corpora were not intended to span specific semantic or syntactic structures in the languages or to capture long-range dependencies that extended across multiple sentences. Therefore we did not ask research questions about these specific speech representations.

The contents of each speech corpus were split across five blocks, where each block was roughly 5–7 minutes in duration. In English and Spanish speech stimuli, four blocks contained unique sentences, and a single block contained ten repetitions of ten sentences. In the Mandarin Chinese speech stimuli, four paragraphs were repeated six times, and repetitions were intermixed with unique paragraphs across blocks. Repeated sentences were used for validation of TRF models (see details below). Sentences were presented with an inter-trial interval of 0.4–0.5 s in English and Mandarin and 0.8 s in Spanish. Spanish sentences were on average 4.77 s long (range: 2.5–8.03 s), English sentences were on average 3.05 s long (range: 1.99–3.60 s), and Mandarin Chinese sentences were on average 3.16 s long (range: 1.17–11.76 s). Although speech blocks of different languages could be intermixed in the same recording session, each 5- to 7-min block of speech consisted of only one language.

Speech stimuli were presented at a comfortable ambient loudness (about 70 dB) through free-field speakers placed roughly 80 cm in front of the patients' head. All speech stimuli were presented in the experiment using custom-written MATLABR2016b scripts (MathWorks, www.mathworks.com). Participants were asked to listen to the stimuli attentively and were free to keep their eyes open or closed during the stimulus presentation. We performed all subsequent data analysis using custom-written MATLABR2024b scripts.

Electrode selection

Subsequent analyses included all speech-responsive electrodes: electrodes for which at least a contiguous 0.1 s of the neural response (10 samples, 100-Hz sampling rate) time-aligned to the first or last 0.5 s of the spoken sentences was significantly different from that of the prespeech silent baseline. To test for significance, we used a one-way ANOVA F -test at each time point, Bonferroni corrected for multiple comparisons using a threshold of $P < 0.01$. We included the neural response aligned to the last 0.5 s of each sentence to ensure that we did not exclude electrodes for which there was a significant response time aligned to the end of sentences but not the beginning.

Feature TRF analysis

HFA responses to speech corpora were predicted using standard linear TRF models with various sets of speech features^{37,111,112}. Many of the speech features used were extracted from pre-existing corrected acoustic, phonetic, phonemic and word-level annotations included with these speech corpora. In the feature TRF (F-TRF) models described below, HFA neural responses recorded from a single electrode were predicted as a linear combination of speech features that occurred at most 0.6 s before the current time point:

$$\text{HFA}(t) = x_0 + \sum_{f=1}^F \sum_{l=1}^L \beta(l, f) X(f, t-l)$$

where x_0 corresponds to the model intercept, F corresponds to the set of speech features included in the model and L corresponds to the number of time lags considered in the model (60 samples reflecting a total lag of 0.6 s). $\beta(l, f)$ weights are therefore interpreted as the weight or importance of the specific combination of speech feature, f , and time-lag, l , in predicting the final neural response given the set of speech features and training sentences. Models were trained and tested separately for each electrode using the same cross-validation, ridge regularization and feature normalization protocols as reported in previous work^{12,38,106}. To derive the F-TRF model performance, we fit the trained model to a held-out set of roughly ten sentences averaged across ten repeats (see 'Stimuli and procedure') and then calculated the correlation-based R^2 fit. In subsequent analysis, we analysed both the magnitude β weights for each speech feature and the overall model R^2 derived from the held-out test sentences. In addition to training and testing F-TRF models on sentences in the same language, we also trained F-TRF models on a language and tested the model on sentences from another language to test the generality of our model fits (Extended Data Fig. 5). This analysis followed the same training and testing procedure as the models that were trained and tested on the same language.

Selection of features for F-TRF

The speech features included in the acoustic-phonetic F-TRF model were selected from those that have been identified in previous studies to strongly drive HFA in the human temporal cortex. We included the sentence onset feature as well as the acoustic edges of envelope (peakRate³⁸), both features that drive neural responses in the posterior STG and middle STG, respectively. The consonant manner of articulation and place of articulation features were selected on the basis of ref. 2, which included binary values at phoneme onset indicating whether the phoneme was dorsal, coronal, labial, nasal, plosive or fricative. The vowel

features were selected on the basis of ref. 106, which included the median formant values for F1, F2, F3 and F4 (Hz) for each vowel sound because continuous formant values were shown to drive neural responses more than categorical vowel features (for example, high, front, low, back).

To test the effect of word-level features on the F-TRF neural prediction, we also included word onset, word frequency, absolute word length (as number of phonemes) and within-word phoneme surprisal. All features had been pre-annotated for each speech corpus except for word frequency, phoneme surprisal and syllable onset (in Spanish only). Word boundary and phoneme surprisal are metrics reflective of how the brain tracks sequential speech structure, with the former being a discrete, sparse marker of segmentation and the latter being a continuous measure of how predictable each phoneme is, given all preceding phonemes in the current word. Word-frequency and phoneme-surprisal features were calculated on the basis of the distribution of the language presented regardless of the participant's language background (Spanish surprisal was used in the model in Spanish speech blocks; a precise description of how these features were generated is below). In the TRF model, we excluded phoneme surprisal at word initial positions so that the word-level features could be dissociated from the phoneme-surprisal feature in time.

Model weight correlation across conditions

We determined a particular electrode's feature tuning by analysing the model weights, where each weight was proportional to the change in electrode response given a change in that feature's value¹¹². We estimated the F-TRF β weights corresponding to each unique speech feature for each electrode separately, as described in the model-fitting procedure above². To compare electrode encoding of the base F-TRF model across speech conditions (Fig. 1), we first selected all electrodes with $R^2 > 0.1$ for both speech conditions; this ensured that the β weight analysis would be based on TRF models that were able to explain the neural responses sufficiently well. We then extracted the mean of the 0.05-s (five samples) window around the maximum β weight averaged across features. This resulted in a single vector for each electrode per speech condition, with one β weight per speech feature; to calculate weight correlation, we performed a Pearson correlation across vectors of the same electrode from different speech conditions. This allowed us to determine the cross-language correlation between the relative β weights. To determine a distribution at chance level across electrodes, we performed permutation testing by shuffling the β weights for a single speech condition and repeating the Pearson correlation ten times per electrode.

Construction of word-frequency, word-length and within-word phoneme-surprisal features

Word-frequency values were adopted from the SUBTLEX American English¹¹³ and SUBTLEX Spanish⁵¹ frequency estimates, which included a total of 74,287 and 94,341 unique words respectively. We included frequency estimates for all words in our corpora that existed in the English and Spanish SUBTLEX word sets. All word-frequency measures were log-scaled (base 10). Word length was determined as the number of phonemes contained within the word onset and offset from the corpus annotations. We used the same word frequency and word lists documented in the SUBTLEX American English and Spanish corpus to construct within-word phoneme surprisal, where each phoneme was associated with a surprisal value. We constructed within-word phoneme surprisal as

$$-\log_2 \left[P(x_n | x_1 \dots x_{n-1}) = \frac{\text{freq}(x_1 \dots x_n)}{\text{freq}(x_1 \dots x_{n-1})} \right]$$

where $x_1 \dots x_{n-1}$ is all preceding phonemes in the current word and x_n is the current phoneme. We did not include surprisal values for the first phoneme of a word, only phonemes at the second position in the word onwards. $\text{freq}(x_1 \dots x_n)$ is the count of all unique words with $x_1 \dots x_n$ as the

Article

prefix multiplied by the frequency of each of those words as estimated by the SUBTLEX English and Spanish corpus. We also calculated within-word phoneme entropy (expected value of surprisal at each phoneme) as well as biphone and triphone surprisal and entropy but ultimately did not include these features in the final F-TRF model, as for most electrodes, these features did not explain added unique variance.

Annotation of DIMEx syllable onset

We generated the syllable-onset feature for the DIMEx speech corpus through a combination of automatic syllable-onset detection and manual validation. We did not annotate instances of resyllabified word onsets, such as instances where the consonants at the end (coda) of a syllable were pronounced as part of the onset of the following syllable. For example, the phrase *los otros* (los.'otros) would be annotated to reflect the canonical form (los.'otros), although it may be pronounced (lo.'so.tros).

Unique variance calculation

Because several of the speech features we used in the F-TRF models were highly correlated, we determined the unique contribution of each feature to driving neural responses through the calculation of unique variance. Unique variance (ΔR^2) is defined as the difference between the portion of variance explained by the model with all relevant features and the model with the feature of interest (f) removed. Broadly, this is

$$\Delta R^2(f) = [\text{full model}]R^2 - [\text{model with } f \text{ removed}]R^2$$

The acoustic-phonetic F-TRF model included sentence onset, peakRate, consonant features (dorsal, coronal, labial, nasal, plosive, fricative) and vowel formants (F1, F2, F3, F4). The full F-TRF model included sentence onset, peakRate, consonant features (dorsal, coronal, labial, nasal, plosive, fricative), vowel formants (F1, F2, F3, F4), word onset, word frequency, word length and phoneme-surprisal.

Unique variance for acoustic-phonetic features (Fig. 2) was calculated as follows:

$$\begin{aligned} \Delta R^2(\text{sentence onset}) \\ = [\text{full model}]R^2 - [\text{full model} - \text{sentence onset}]R^2 \end{aligned}$$

$$\Delta R^2(\text{peakRate}) = [\text{full model}]R^2 - [\text{full model} - \text{peakRate}]R^2$$

$$\begin{aligned} \Delta R^2(\text{consonant features}) \\ = [\text{full model}]R^2 - [\text{full model} - \text{consonant features}]R^2 \end{aligned}$$

$$\begin{aligned} \Delta R^2(\text{vowel formants}) \\ = [\text{full model}]R^2 - [\text{full model} - \text{vowel formants}]R^2 \end{aligned}$$

Unique variance for word-level and phoneme-surprisal features (Fig. 2) was calculated as follows:

$$\begin{aligned} \Delta R^2(\text{word onset}) = [\text{acoustic phonetic model} + \text{word onset}]R^2 \\ - [\text{acoustic phonetic model}]R^2 \end{aligned}$$

$$\begin{aligned} \Delta R^2(\text{word frequency}) \\ = [\text{acoustic phonetic model} + \text{word onset} + \text{word frequency}]R^2 \\ - [\text{acoustic phonetic model} + \text{word onset}]R^2 \end{aligned}$$

$$\begin{aligned} \Delta R^2(\text{word length}) \\ = [\text{acoustic phonetic model} + \text{word onset} + \text{word length}]R^2 \\ - [\text{acoustic phonetic model} + \text{word onset}]R^2 \end{aligned}$$

$$\begin{aligned} \Delta R^2(\text{phoneme surprisal}) \\ = [\text{acoustic phonetic model} + \text{phoneme surprisal}]R^2 \\ - [\text{acoustic phonetic model}]R^2 \end{aligned}$$

Combined unique variance for all word-level and phoneme-surprisal features was calculated as:

$$\begin{aligned} \Delta R^2(\text{word features}) \\ = [\text{full model}]R^2 - [\text{acoustic phonetic model}]R^2 \end{aligned}$$

To determine the significance of unique variance values for a single feature or feature family, we computed a distribution of 300 permuted unique variance values by applying random circular temporal shifts to the feature or feature family of interest. To acquire the P value associated with the unique variance value, we summed the number of permuted unique values greater than the true unique value and divided by the total number of permutations we computed.

Statistical testing

LME models as computed by the `fitlme` function in MATLABR2024b included random effects for participant and hemisphere as well as the interaction between participant and electrode to ensure that the effects we found were significant across participants and hemispheres. LME models were primarily used to determine whether the unique variance of a particular feature f was significantly different between native and foreign speech conditions. LME formulas took the following form: $\Delta R^2(f) \sim \text{speech condition} + \text{speaker language} + (1|\text{hemisphere}) + (1|\text{participant}) + (1|\text{electrode:participant})$.

The same LME formula (above) was used to test significance in instances in which TRF feature unique variance could be compared across speech conditions (for example, Figs. 2b,c and 5d).

For Pearson and Spearman correlation, two-tailed tests were performed unless otherwise noted.

Acoustic word-boundary decoding

Acoustic word-boundary decoding consisted of a logistic regression model trained to classify whether the mel-spectrogram from 0.2 s around a boundary event was a within-word syllable boundary or a word boundary. All single-syllable words and words at sentence onset were removed from this analysis. To match the instances of word and syllable boundaries in the Spanish and English corpora, we randomly sampled 1,900 instances of word and within-word syllable boundaries from each corpus. In Spanish, we used 1,273 within-word syllable onsets and 627 word onsets. In English, we used 1,153 within-word syllable onsets and 747 word onsets.

For each boundary event in both corpora, the 0.2-s 80-band mel-spectrogram (sampled at 100 Hz) was vectorized. PCA was then performed on the vectorized mel-spectrogram to reduce the total dimensionality. The components with the highest variance that cumulatively explained more than 90% of the total variance were ultimately used as input to the regression, which totalled about 60–80 PCs depending on the language. Logistic regression was run using standard MATLABR2024b regression functions (`fitlinear`) with Lasso (L1) regularization and 20-fold cross-validation. The AUC was extracted for each held-out test fold, where chance AUC would equal 0.5.

In addition to training and testing acoustic word-boundary decoding in the same language, we also trained the word-boundary decoders in one language and tested on boundary events from the other language to test the extent to which acoustic cues to word boundaries overlapped between the languages. This analysis followed the same training and testing procedure as for the decoders that were trained and tested on the same language. To assess the significance of the test-set AUC values, we performed permutation testing in which word and within-word

syllable boundary labels were randomly shuffled and the same logistic regression procedure was applied.

Neural word-boundary effects

To determine whether speech-responsive electrodes showed a significant difference between boundary events (within-word syllable boundaries and word boundaries), we assessed the duration of contiguous time points for which the neural response (100 Hz) time aligned 0.2 s before and 0.4 s after word-boundary events significantly differed from responses aligned to within-word syllable boundaries. To test for significance, we used a one-way ANOVA *F*-test at each time point and Bonferroni corrected for multiple comparisons using a threshold of $P < 0.01$. We computed this for each electrode and each language separately such that every electrode was associated with the duration of significant differences value for each language.

Neural word-boundary decoding

Neural word-boundary decoding consisted of a logistic regression model trained to classify whether the contiguous window of neural activity 0.2 s before and 0.4 s after a boundary corresponded to a within-word syllable boundary or a word boundary. To compare against the acoustic word-boundary decoding, all single-syllable words and words at sentence onset were removed from this analysis. The size of the window used in neural decoding was larger than that of the acoustic decoding (0.2 s) because the neural response dynamics are slower in latency than the acoustic dynamics.

To perform neural word-boundary decoding, neural responses (sampled at 100 Hz) from all speech-responsive electrodes from a subset of participants were concatenated into a matrix of the form (electrode \times time \times event), which was then vectorized across the electrode and time dimensions. The subset of participants who had the maximum overlap in trials were used for neural word-boundary decoding. We performed PCA on the vectorized spatiotemporal neural data to reduce the total dimensionality. The components with the highest variance that cumulatively explained more than 90% of the total variance were ultimately used as input to the regression: about 800–1,000 PCs, depending on the participant group. Extended Data Table 6 summarizes the number of {unique participants, electrodes, PCs} used to train each decoder.

Logistic regression was run using standard MATLABR2024b regression functions (fitlinear) with Lasso (L1) regularization and 20-fold cross-validation. The AUC was extracted for each held-out test fold. Beta linear coefficient estimates were extracted from the trained ClassificationLinear model, and the weights of single electrodes were calculated by thresholding the percentile weight values and converting from PC space back to the original (electrode \times time) space.

To characterize the temporal dynamics of decoding performance (Extended Data Fig. 9), we performed sliding-window neural word-boundary decoding. To do this, we used the same set of speech-responsive electrodes as for the fixed 0.6-s window decoding, but we used a sliding window of 0.02 s within this 0.6 s of neural response to perform the decoding. The same trials and electrodes were used for decoding across all 0.02-s windows in the same participant group and language.

We also performed single-participant neural word-boundary decoding (Extended Data Fig. 9) using all speech-responsive electrodes per participant. Single-participant word-boundary decoding was calculated similarly to the participant group decoding (with the same time window, with PCA) but included electrode responses from a single participant only. The number of electrodes included in the decoding varied by participant (electrodes: median = 47, range = {21, 79}). For statistical analysis comparing decoding performance across speech conditions in single participants (Extended Data Fig. 9c), a paired Wilcoxon signed-rank test was used.

For the single participants who spoke English and at least one other language (Fig. 5), we followed the same procedure as above for calculating English word-boundary decoding AUC. To assess the significance of the effect of English proficiency on neural word-boundary classification performance (in English), we used a linear model:

$$\text{AUC} - \text{proficiency} + (1/\text{number of electrodes})$$

Construction of AAI

The acoustic ambiguity index (AAI) is based on the acoustic word-boundary decoder that takes in the spectrogram values using 0.2 s around each boundary event and outputs whether the trial corresponds to a word boundary or a within-word syllable boundary. As above, single-syllable words and words at sentence onset were removed from this analysis. For each single trial, the decoder outputs both a predicted label (0, 1) and a posterior probability score (0–1). The scores of the held-out test trials are extracted as one of the outputs of the MATLABR2024b ClassificationLinear model predict function. The per-trial value $\text{abs}(\text{correct label} - \text{posterior probability score})$ is then calculated to produce a continuous measure of how accurate the classifier is on single trials, where larger values indicate that the classifier is probably predicting incorrectly. The corresponding neural word-boundary decoding is produced after (1) calculating the difference between the probability scores for single trials from neural decoding and correct label and (2) comparing this distribution between known and foreign groups for each bin.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data to replicate all figures are available at Zenodo (<https://doi.org/10.5281/zenodo.17247450>)¹¹⁴ and are available on request to the corresponding author. Because participants at the outset of this study had not consented for public data release, these data are only available upon reasonable request to respect patient privacy and consent.

Code availability

All MATLABR2024b codes used to generate the main and Extended Data figures are available at <https://github.com/ChangLabUcsf/BhayaGrossman2025>.

- Ogarian, Y., Bhaya-Grossman, I., Johnson, K. & Chang, E. F. Vowel and formant representation in the human auditory speech cortex. *Neuron* **111**, 2105–2118.e4 (2023).
- Miller, K. J., Sorensen, L. B., Ojemann, J. G. & den Nijs, M. Power-law scaling in the brain surface electric potential. *PLoS Comput. Biol.* **5**, e1000609 (2009).
- Parvizi, J. & Kastner, S. Promises and limitations of human intracranial electroencephalography. *Nat. Neurosci.* **21**, 474–483 (2018).
- Hamilton, L. S., Chang, D. L., Lee, M. B. & Chang, E. F. Semi-automated anatomical labeling and inter-subject warping of high-density intracranial recording electrodes in electrocorticography. *Front. Neuroinform.* **11**, 62 (2017).
- Pineda, L. A. et al. The Corpus DIMEx100: transcription and evaluation. *Lang. Resour. Eval.* **44**, 347–370 (2010).
- Aertsen, A. M. H. J., Olders, J. H. J. & Johannesma, P. I. M. Spectro-temporal receptive fields of auditory neurons in the grassfrog. *Biol. Cybern.* **39**, 195–209 (1981).
- Holdgraf, C. R. et al. Encoding and decoding models in cognitive electrophysiology. *Front. Syst. Neurosci.* **11**, 61 (2017).
- SUBTLEXus. Ghent University <https://www.ugent.be/pp/experimentele-psychologie/en/research/documents/subtlexus> (2015).
- Bhaya-Grossman, I. et al. Data for ‘Shared and language-specific phonological processing in the human temporal lobe’. Zenodo <https://doi.org/10.5281/zenodo.17247450> (2025).

Acknowledgements We thank M. Lobato-Scharfhausen and S. Harper for their help in annotating syllable onsets in the Spanish speech corpus. We thank J. Gauthier for providing feedback on the manuscript as well as Y. Ogarian and L. Hamilton, who generated some of the code on which several analyses were based. We thank members of the Chang laboratory for

Article

their help with data collection. This work was supported by the National Institutes of Health (grant nos. NIDCD R01DC012379, NINDS U01NS117765), Susan and Bill Oberndorf, Joan and Sanford Weill, William K Bowes, Jr. Foundation, and Feldman Foundation. E.F.C. is a Chan Zuckerberg Biohub Investigator. I.B.-G. is also supported by the National Science Foundation GRFP and the UCSF Discovery Fellowship.

Author contributions This study was designed by I.B.-G., M.K.L. and E.F.C. Data collection was performed by all authors. Data processing and analysis was performed by I.B.-G. Interpretation of the data and analysis was performed by I.B.-G., M.K.L., Y.Z., L.G., K.J. and E.F.C. I.B.-G. wrote the first draft of the manuscript and prepared all main and Extended Data figures. All authors contributed to editing and revising the manuscript.

Competing interests E.F.C. is an inventor on a pending UCSF patent application (application number: WO2022251472A1, 2022, WIPO PCT - International patent system) and patents PCT/US2020/028926, PCT/US2020/043706 and US9905239B2. E.F.C. is cofounder of Echo Neurotechnologies, LLC. All other authors declare no competing interests.

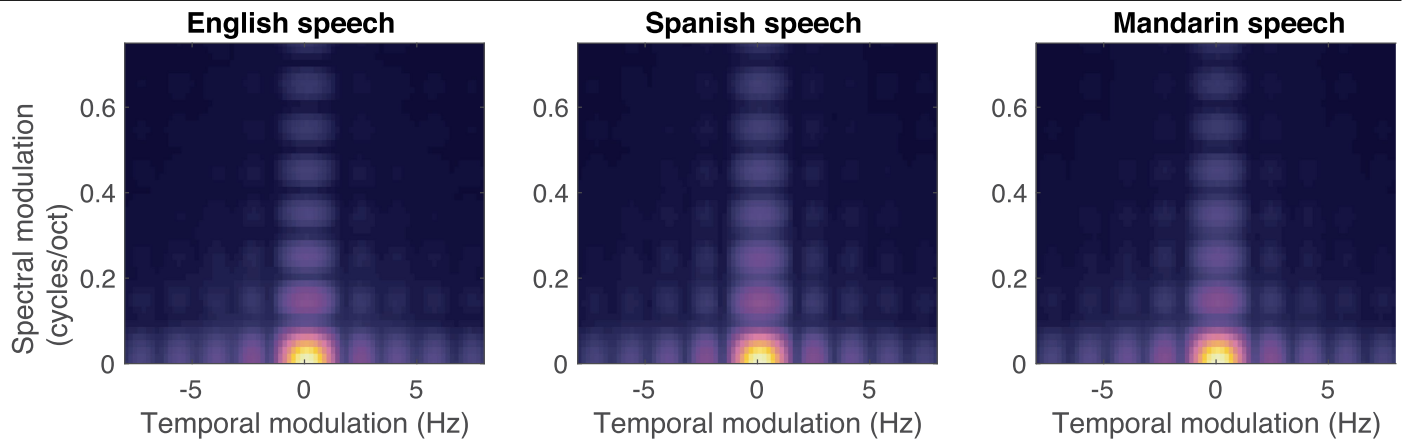
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-09748-8>.

Correspondence and requests for materials should be addressed to Edward F. Chang.

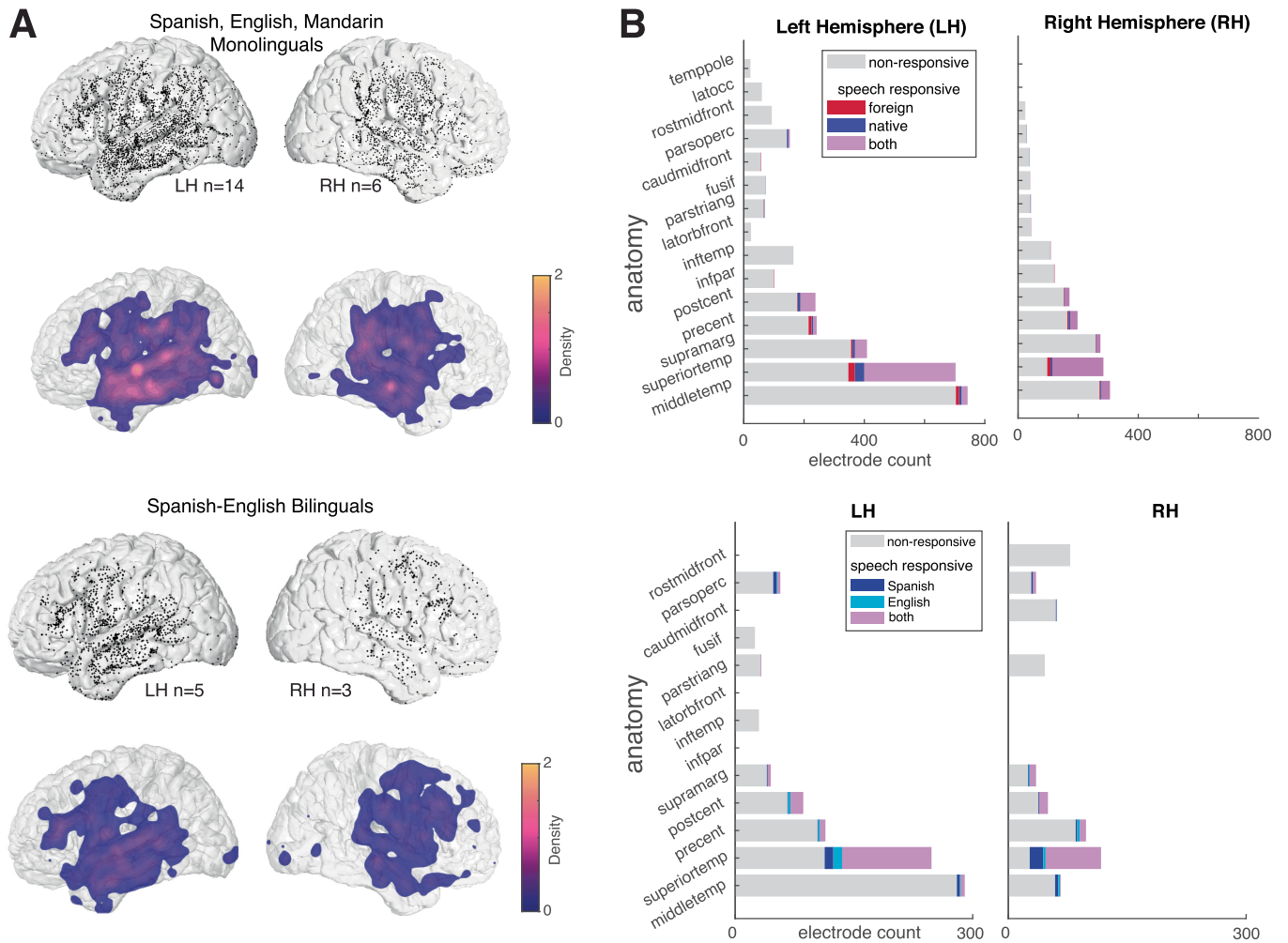
Peer review information *Nature* thanks Narly Golestani, Marc Joannis, Shihab Shamma and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



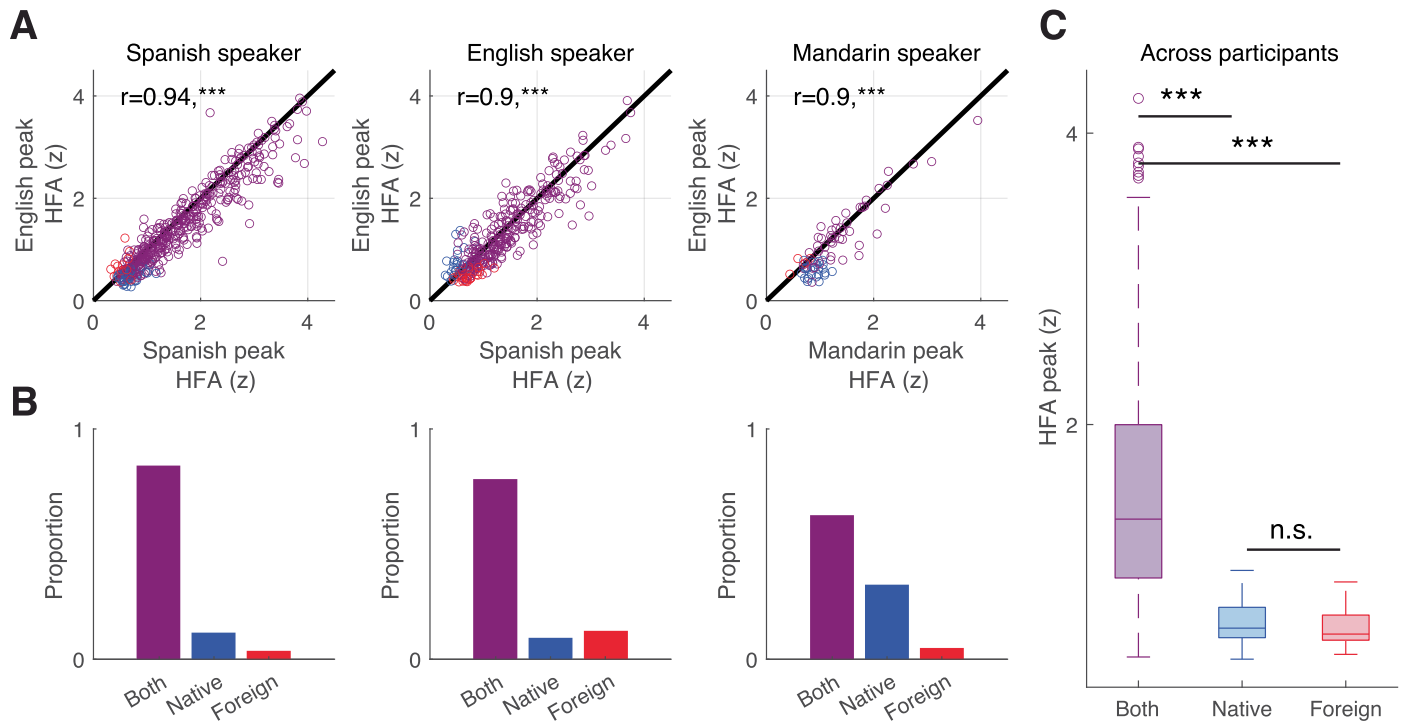
Extended Data Fig. 1 | Similar modulation spectrum across speech corpora. Highly similar modulation spectra (qualitative) for English speech stimuli (TIMIT), Spanish speech stimuli (DIME_x), and Mandarin speech stimuli (ASCCD).

Article



Extended Data Fig. 2 | High density lateral surface coverage of the human temporal lobe in English, Spanish, Mandarin speakers and Spanish-English bilinguals. A. Wide and dense coverage of the superior temporal lobe across left and right hemispheres across participants. Anatomical location (MNI) of all electrodes plotted prior to subselecting for speech responsive electrodes. B. Highest counts of electrodes in the left and right hemisphere superior temporal and middle temporal cortices bilaterally. Highest counts of speech-responsive

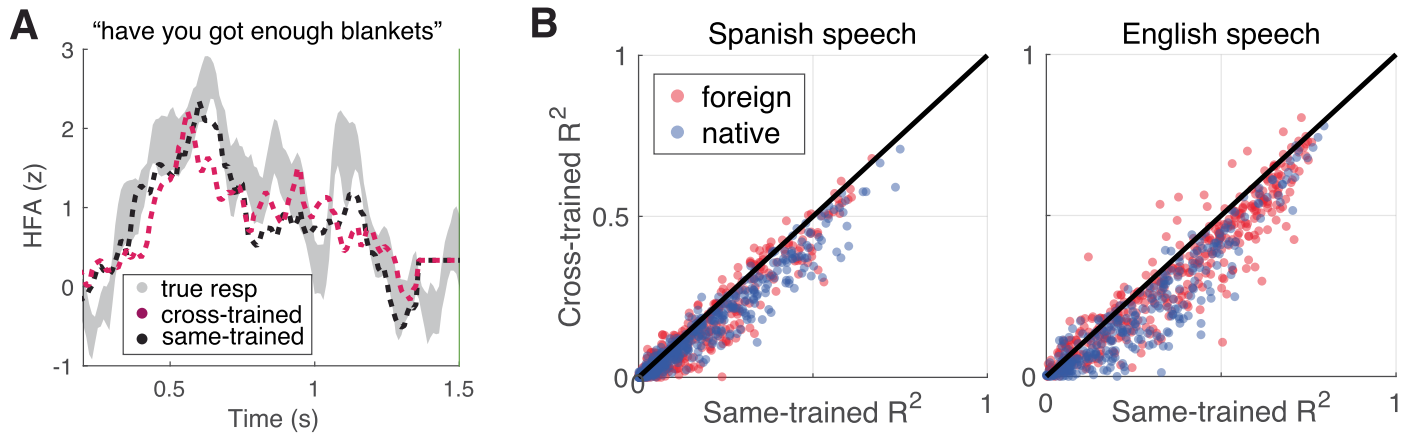
electrodes in superior-temporal cortex. In Spanish, English and Mandarin monolinguals, a majority of speech-responsive electrodes respond to both foreign and native speech, while a minority of electrodes show selective responses to either foreign or native speech. Similarly, in Spanish-English bilinguals, a majority of speech-responsive electrodes respond to both English and Spanish speech.



Extended Data Fig. 3 | Similar peak HFA response across native and foreign speech conditions. A. Across Spanish, English, and Mandarin speakers, a majority of electrodes are responsive to both native and foreign speech conditions (purple) as opposed to only native (blue) or only foreign (red) speech. Each scatter point corresponds to a single electrode. HFA peak response is calculated as the peak magnitude of the average response across all sentences in the same language. Average HFA peak magnitudes are correlated across speech conditions (Pearson $r(472) = 0.94$; $r(334) = 0.9$; $r(77) = 0.9$, $***p < 0.001$ for all participant groups). B. Spanish, English, and Mandarin speakers show no consistent bias in the proportion of electrodes that are only

responsive for only either native or foreign speech conditions (blue, red bars). C. Speech responsive electrodes to one language but not the other showed significantly smaller HFA amplitude responses. Electrodes responsive to only foreign or native speech showed no consistent bias in HFA amplitude and no consistent anatomical location (data not shown). Two-sided Wilcoxon rank sum test ($n = [705; 113; 65]$) determined significance between the distribution of HFA peak magnitudes across groups ($***p < 0.001$). Box plots show the maximum and minimum values (whiskers), median (center line) and the 25th to 75th percentiles (box limits).

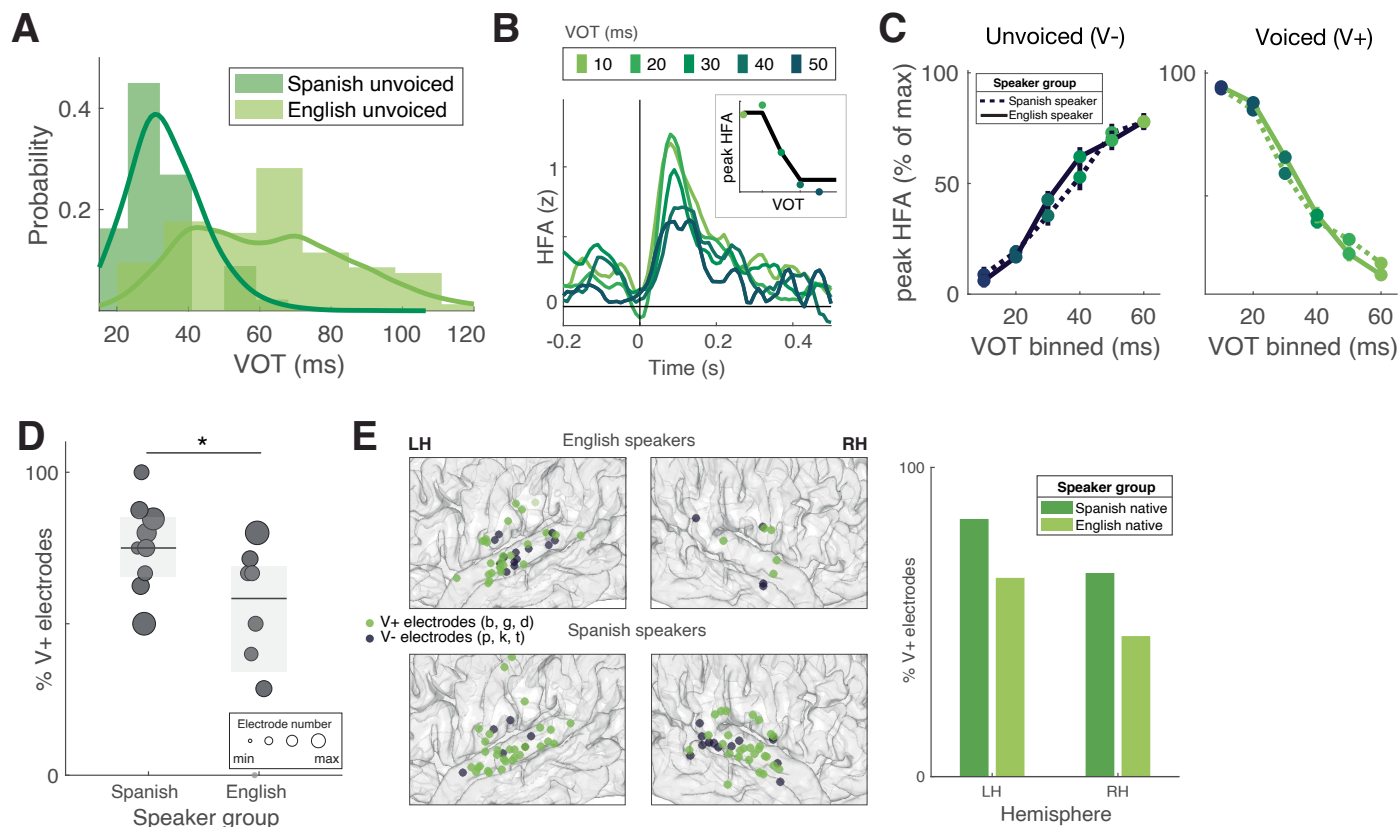
Article



Extended Data Fig. 4 | Same and Cross-language Model Predictions.

A. Highly similar predicted neural response to an example English sentence from a model trained on another language (Spanish) as opposed to the same language (English). Dashed lines indicate a TRF predicted response.

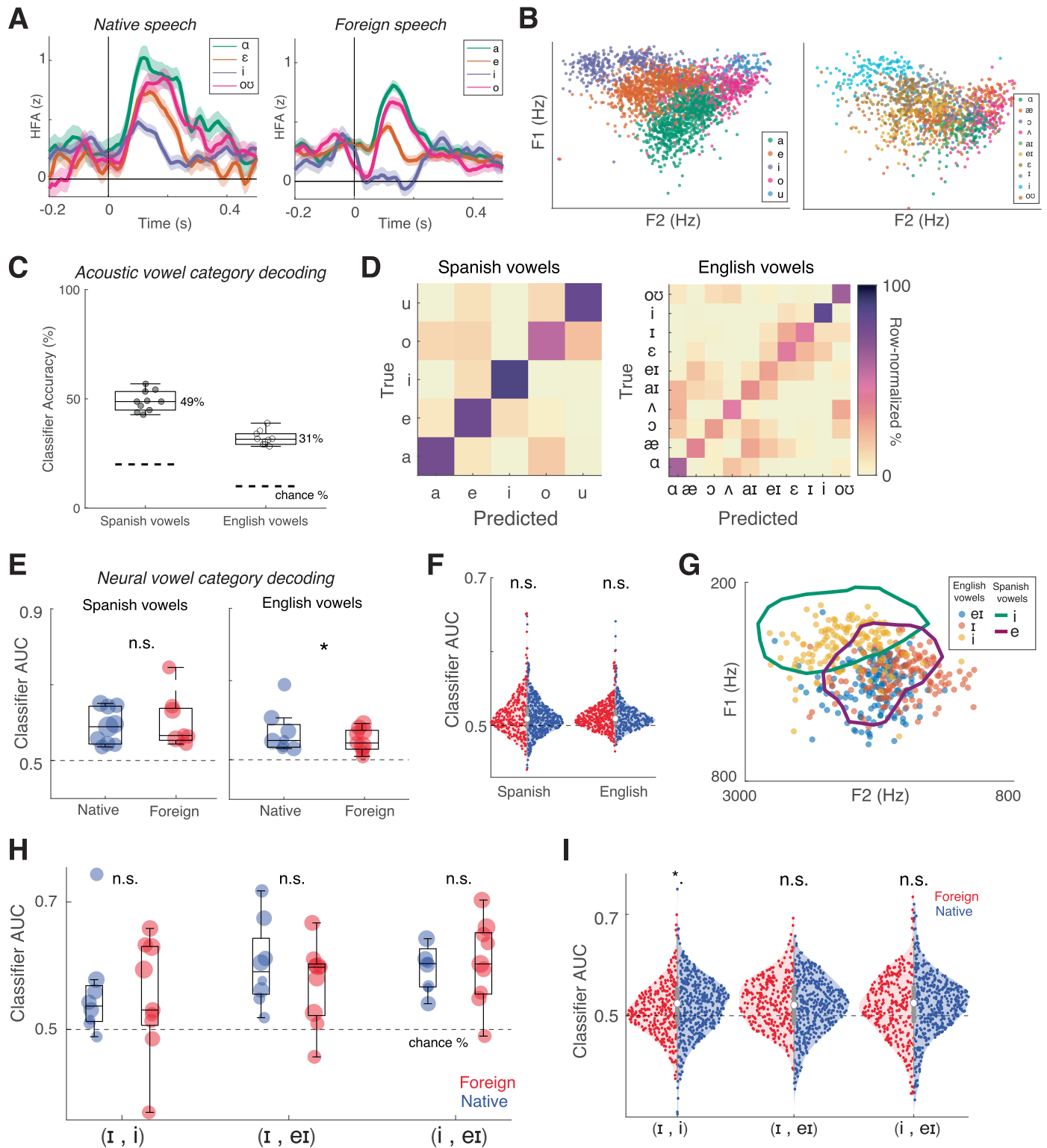
B. Cross-language TRF prediction performance (e.g. training on Spanish speech, testing on English speech) is correlated with same-language prediction TRF performance across both Spanish (Pearson $r(883) = 0.96$, $***p < 0.001$) and English speech (Pearson $r(883) = 0.94$, $***p < 0.001$).



Extended Data Fig. 5 | Voice Onset Time Encoding Across Languages. A. VOT values of unvoiced stop-consonants in Spanish (dark green) and English (light green) show systematic differences in duration. This systematic difference along with prior research on the neural encoding of voice onset time (VOT)²⁶ suggest that language experience may influence how sounds are processed in the STG. B. Electrodes are sensitive to different VOT values (0:10:60 ms). The top-right inset shows the VOT tuning function for this electrode, where each scatter point corresponds to the peak HFA response to the VOT bin. C. Consistent with prior research, average single electrode tuning curves either showed positive (V-, preferring unvoiced) or negative (V+, preferring voiced) tuning direction. We found no significant difference between tuning curves across VOT sensitive electrodes in Spanish and English speakers (two-tailed

Wilcoxon rank sum test, $V+ n = [57, 32]$, $V- n = [20, 21]$). Error bars indicate standard error of the mean across electrodes within each VOT bin for each speaker group. D. Proportion of V+ electrodes was significantly higher ($p < 0.05$) in Spanish speakers as opposed to English speakers. Size of each scatter corresponds to the number of electrodes within each participant that show significant tuning for VOT. P-value corresponds to language background in a linear mixed effect model ($t(128) = 2.11$, $SE = 0.0176$, $p = 0.0367$). Box plots show the median (center line) and the 25th to 75th percentiles (box limits). E. Intermixed V+ and V- electrodes in English and Spanish speakers. Both left and right hemispheres showed a higher proportion of V+ electrodes in Spanish speakers, suggesting that language experience effects may arise at the population level²⁵.

Article



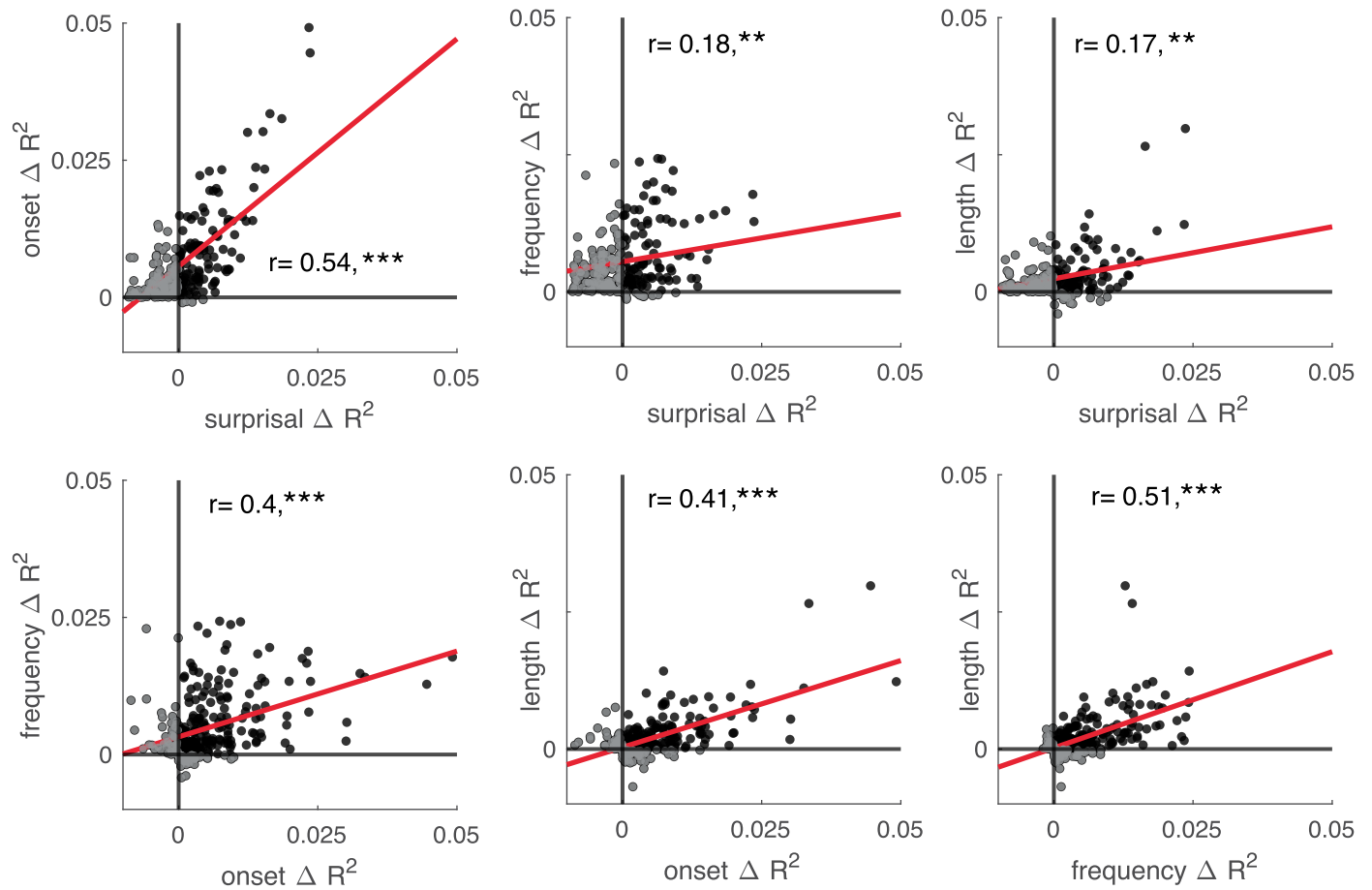
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Vowel Category Encoding Across Languages.

A. Example electrode response to approximately equivalent vowel categories across languages (English, Spanish) shows qualitatively similar response patterns in native and foreign speech conditions. Shaded patches indicate standard error of the mean HFA response across vowels. B. Formant frequencies (Hz) from each vowel instance are shown in the scatter and colored by the vowel category they belong to. Spanish (left) and English (right) vowel spaces include a different number of vowels, and contain different formant space distributions for each category. C. Acoustic vowel category decoding (10-fold linear discriminant analysis) accuracy using the formant values of vowel sounds across English and Spanish speech corpora. Each scatter point represents 1/10 validation folds. Chance accuracy was calculated based on the number of vowel categories included from the particular language (Spanish: 1/5, English 1/10). Vowel categories with more than 100 instances in the speech corpus were included. Box plots show the maximum and minimum values (whiskers), median (center line) and the 25th to 75th percentiles (box limits). D. Confusion matrix resulting from acoustic vowel category decoding exhibits a clear diagonal, with off-diagonal confusions reflecting overlap in F1/F2 space (e.g. /a/ and /e/). E. Single participant neural vowel category decoding (10-fold linear discriminant analysis; each scatter corresponds to a single participant) separated by speech condition shows no significant difference in Spanish (LME $t(168) = 0.49$, SE = 0.0198, $p = 0.63$, see “Methods” for formula) and a weak effect

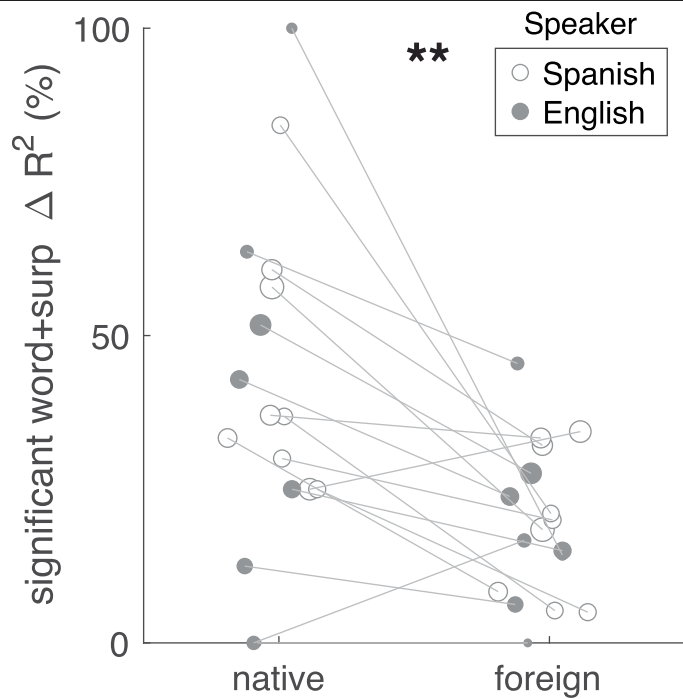
in English (LME $t(168) = 1.98$, SE = 0.0095, $*p = 0.049$). Box plots show the maximum and minimum values (whiskers), median (center line) and the 25th to 75th percentiles (box limits). F. Single electrode neural vowel category decoding (10-fold linear discriminant analysis; each scatter corresponds to a single electrode) separated by language shows no significant effect of speech condition (two-sided Wilcoxon rank sum test Spanish $n = [453, 302]$, $p = 0.23$; English $n = [292, 416]$, $p = 0.86$). G. Vowel formants (F1, F2) for instances of “i”, “I”, and “ei” in English are subsumed by only two vowel categories in Spanish, “i” and “e”. 95th percentile contours for Spanish vowel formants shown as colored lines. H. Pairwise neural decoding AUC (10-fold logistic regression) for three English vowel categories, “i”, “I”, and “ei” shows no significant difference between participant groups (LME $t(168) = -0.029$, $p = 0.98$; $t(134) = -0.87$, $p = 0.38$; $t(134) = 0.84$, $p = 0.40$, see “Methods” for formula). Each scatter point corresponds to the mean AUC for a single participant (colored by speech condition). Box plots show the maximum and minimum values (whiskers), median (center line) and the 25th to 75th percentiles (box limits). I. Pairwise neural decoding AUC (10-fold logistic regression) for the same English vowel categories, shows no significant difference between single electrodes across participant groups. Each scatter point corresponds to the mean AUC for a single electrode (two-sided Wilcoxon rank sum test $n = [416, 292]$, $p = 0.02$; $p = 0.08$; $p = 0.24$).

Article

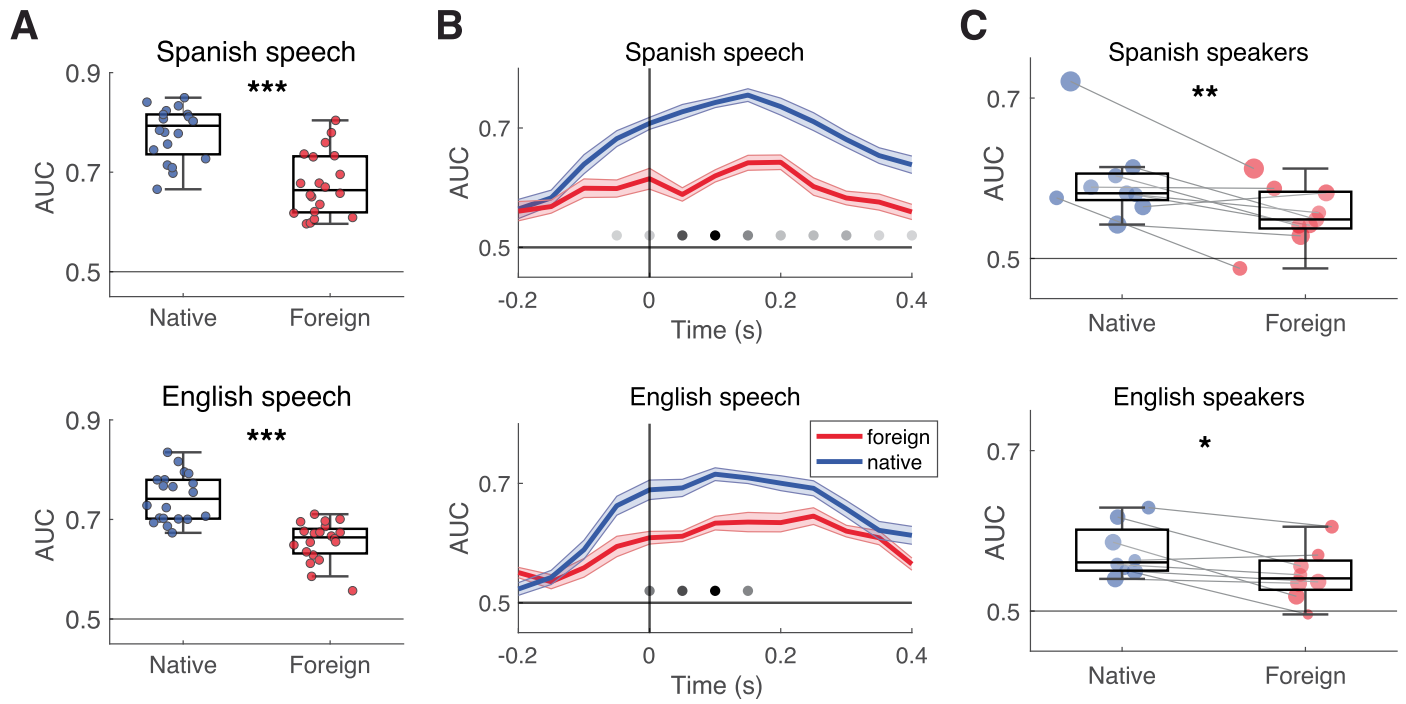


Extended Data Fig. 7 | Comparing Word and Phoneme Surprisal Unique Variances. Significant correlations between positive unique variance values for word onset vs. surprisal (Pearson $r(257) = 0.54, ***p < 0.001$), frequency vs. surprisal (Pearson $r(268) = 0.18, **p = 0.0033$), length vs.

surprisal (Pearson $r(250) = 0.17, **p = 0.009$), word onset vs. frequency ($r(303) = 0.4, ***p < 0.001$), word onset vs. length (Pearson $r(301) = 0.41, ***p < 0.001$), and length vs. frequency ($r(293) = 0.51, ***p < 0.001$) in the native speech condition.



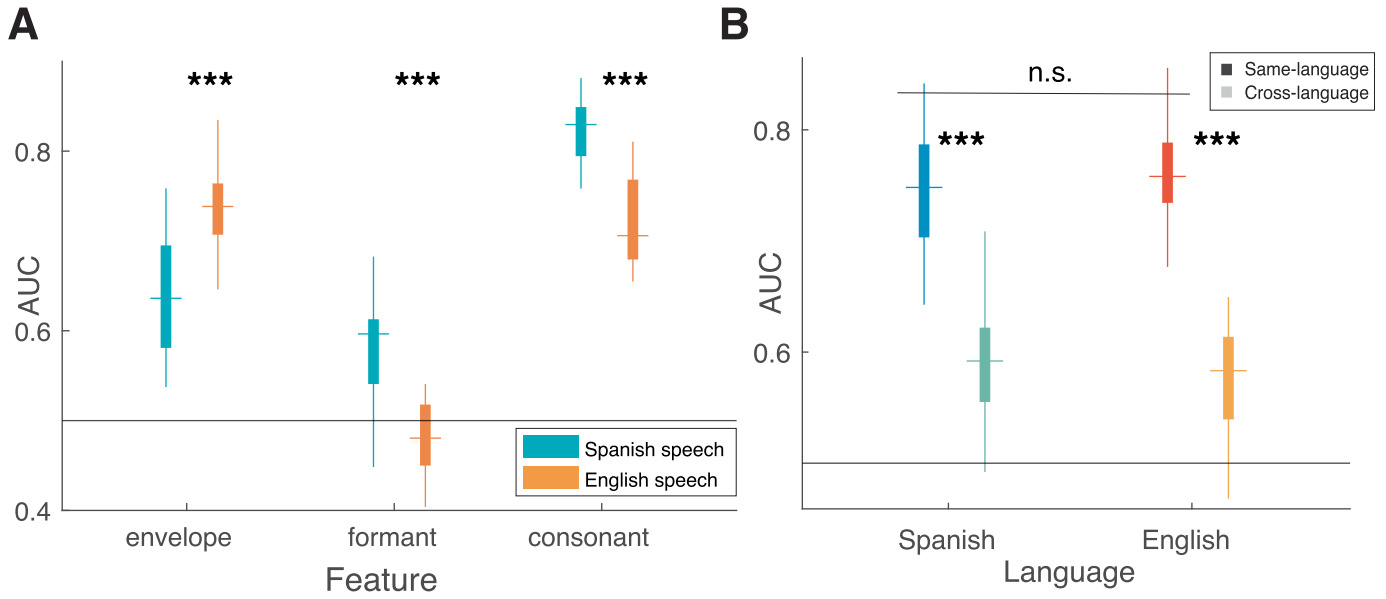
Extended Data Fig. 8 | Proportion of Word-level Electrodes Across Conditions. Proportion of electrodes with significant word and phoneme surprisal unique variance for single participants across native and foreign speech conditions. Significance determined by a 1-tailed paired Wilcoxon signed-rank test ($n = 17$, $**p = 0.0012$).



Extended Data Fig. 9 | Characterization of Neural Word Boundary Decoding.

A. Neural classification of word boundaries is significantly better in native vs. foreign speech (each scatter point corresponds to the AUC of 1/20 folds shown separately for English and Spanish; 1-tailed 2-sample t-test $***p < 0.001$). Box plots show the maximum and minimum values (whiskers), median (center line) and the 25th to 75th percentiles (box limits). B. Temporally-resolved neural classification of word boundaries (increments of 0.05 s and 0.2 s before to 0.4 s after boundary, mean across 20 folds) is significantly better in native vs. foreign speech approximately 0.1 s after the boundary event (per-time point two-sided ANOVA $p < 0.01$ with Bonferroni correction show significant

time-points as black scatter points). Continuous shaded patches indicate standard error of the mean across decoding folds for each time point. C. Neural classification of word boundaries is significantly better in native vs. foreign speech when considering single participant decoding performance (each scatter corresponds to median AUC across 20 folds per participant separated by group, size corresponds to the number of electrodes used for classification). P-values derived from a 1-tailed paired Wilcoxon signed-rank test (Spanish speakers $n = 9$, $**p = 0.0098$; English speakers $n = 8$, $*p = 0.012$). Box plots show the maximum and minimum values (whiskers), median (center line) and the 25th to 75th percentiles (box limits).



Extended Data Fig. 10 | Characterization of the Acoustic Cues to Word Boundary. A. In Spanish and English speech, different acoustic-phonetic features cue word and syllable boundaries. We used three distinct acoustic-phonetic features: continuous envelope, continuous formants, and consonant features around each boundary event to classify word vs. syllable boundary. All features we tested show significantly different AUC across English and Spanish (2-tailed Wilcoxon signed-rank test, for all features $n = 20$, $***p < 0.001$). Box plots show the maximum and minimum values (whiskers), median

(center line) and the 25th to 75th percentiles (box limits). B. Acoustic classification of word boundary trained and tested on the spectrogram from the same language shows significantly higher AUC than a classifier trained on one language and tested on the other (2-tailed Wilcoxon signed-rank test, same-language vs. cross-language $n = 20$, $***p < 0.001$; English vs. Spanish $n = 20$, $p = 0.37$), suggesting acoustic cues to word boundaries are distinct in Spanish and English. Box plots show the maximum and minimum values (whiskers), median (center line) and the 25th to 75th percentiles (box limits).

Article

Extended Data Table 1 | Spanish proficiency, frequency of use, and age of acquisition for self-reported English monolingual participants

ID	Hemisphere	Age	Handedness	Gender	L1	Spanish Proficiency	Spanish Frequency of Use	Learned Spanish at Age
E01	L	49	R	F	English	0		
E02	L	37	R	F	English	3	irregular	14
E03	L	23	R	M	English	1	irregular	14
E04	L	47	A	F	English	1	irregular	18
E05	R	29	R	M	English	0		
E06	L	31	R	F	English	0		
E07	L	47	R	F	English	1	monthly	14
E08	R	36	R	F	English	3	daily	13

Self-rated proficiency profiles of Spanish for English monolingual participants. (L1: First learned language). Proficiency scale as: 1- single words, 2-basic conversation, 3-fluent, 4-very good, 5- native. Frequency of use scale as using this language: 1-daily, 2-weekly, 3-monthly, 4-irregular, 5-never.

Extended Data Table 2 | English proficiency, frequency of use, and age of acquisition for self-reported Spanish and Mandarin monolingual participants

ID	Hemisphere	Age	Handedness	Gender	L1	English Proficiency
S01	L	28	R	F	Spanish	0
S02	R	45	L	F	Spanish	0
S03	L	35	NC	M	Spanish	0
S04	R	59	R	M	Spanish	0
S05	R	23	R	M	Spanish	0
S06	L	39	L	F	Spanish	0
S07	L	NC	R	M	Spanish	0
S08	R	36	R	M	Spanish	0
S09	L	41	R	M	Spanish	0
M01	L	NC	R	M	Mandarin	0
M02	L	51	R	M	Mandarin	0
M03	L	46	R	F	Mandarin	0

Self-rated proficiency profiles of English for Spanish and Mandarin monolingual participants. (L1: First learned language, NC: Not collected). Proficiency scale as: 1- single words, 2-basic conversation, 3-fluent, 4-very good, 5- native. Frequency of use scale as using this language: 1-daily, 2-weekly, 3-monthly, 4-irregular, 5-never.

Article

Extended Data Table 3 | Spanish and English proficiency, frequency of use, and age of acquisition for self-reported Spanish-English bilinguals

ID	Hem	Age	Hand	Gend	Dom	L1	L2	Spanish Prof	Spanish FoU	Spanish AoA	English Prof	English FoU	English AoA
B01	L	28	R	M	Span	Span	Eng	5	daily	1	5	daily	1
B02	R	33	R	F	Eng	Span	Eng	5	daily	1	5	daily	5
B03	R	27	A	F	Eng	Span	Eng	3	daily	1	5	daily	1
B04	R	23	R	M	Span	Span	Eng	5	daily	1	3	irregular	5
B05	L	23	R	M	Span	Span	Eng	5	daily	1	4	daily	3
B06	L	22	R	M	Eng	Span	Eng	3	daily	3	5	daily	3
B07	L	26	R	F	Span	Span	Eng	5	daily	1	5	daily	5
B08	L	32	R	M	Span	Span	Eng	5	daily	1	5	daily	4

Self-rated Spanish and English proficiency profiles of Spanish-English multilingual participants (Hem: hemisphere, Hand: Handedness, Gend: Gender, Dom: Dominant Language, FoU: Frequency of Use, AoA: Age of Acquisition, Span: Spanish, Eng: English, Cat: Catalan, NC: Not collected). Proficiency scale as: 1- single words, 2-basic conversation, 3-fluent, 4-very good, 5- native. Frequency of use scale as using this language: 1-daily, 2-weekly, 3-monthly, 4-irregular, 5-never.

Extended Data Table 4 | English proficiency, frequency of use, and age of acquisition for participants with diverse language backgrounds

ID	Hem	Age	Hand	Gend	Dom	L1	L2	L3	English Prof	English FoU	English AoA
D01	L	NC	R	F	Ara	Ara	Fren	Eng	2	daily	15
D02	L	33	R	F	Russ	Russ	Eng		4	daily	8
D03	L	NC	NC	F	Kor	Kor			0		
D04	L	53	R	M	Mand	Mand	Eng		4	daily	15
D05	Bil	22	R	M	Port	Port	Span	Eng	5	daily	5
D06	R	42	R	F	Span	Span	Catal	Eng	3	irregular	8

Self-rated English proficiency profiles of diverse multilingual participants (Hem: hemisphere, Hand: Handedness, Gend: Gender, Dom: Dominant Language, FoU: Frequency of Use, AoA: Age of Acquisition, Span: Spanish, Eng: English, Cat: Catalan, Port: Portuguese, Mand: Mandarin, Kor: Korean, Ara: Arabic, Russ: Russian, NC: Not collected). Proficiency scale as: 1- single words, 2-basic conversation, 3-fluent, 4-very good, 5- native. Frequency of use scale as using this language: 1-daily,2-weekly, 3-monthly, 4-irregular, 5-never.

Article

Extended Data Table 5 | Summary of participant groups in each analysis

Figure	Analysis	Participant Groups
Fig. 1	Comparison of speech-responsive sites (B-E), Acoustic-phonetic TRF analysis (F-H)	Spanish (S01-S09) English (E01-E08) Mandarin (M01-M03)
Fig. 2	Comparison of word-level and phoneme surprisal TRF analysis (B-G)	Spanish (S01-S09) English (E01-E08)
Fig. 3	Word vs. syllable boundary differences across conditions (D-F)	Spanish (S01-S09) English (E01-E08)
Fig. 4	Exploration of word vs. syllable boundary decoding (A-D)	Spanish (S01-S09) English (E01-E08)
Fig. 5	Comparing acoustic-phonetic and word-level TRF analysis in two known languages (multilingual; A-F)	Spanish-English Bilingual (B01-B08) Spanish (S01-S09) English (E01-E08)
Fig. 5	Word vs. syllable boundary decoding as a function of English proficiency (H-I)	Spanish-English Bilingual (B01-B08) Mandarin (M01-M03) Diverse languages (D01-D06)

Participants cohorts included in each analysis within each of the main figures.

Extended Data Table 6 | Summary of participants, electrodes, PCs used in each decoder

	Spanish monolingual	English monolingual
Spanish speech	{4, 189, 1005}	{4, 174, 952}
English speech	{3, 141, 879}	{5, 179, 936}

The number of participants, PCs, and trials included in each neural word boundary classification run across speaker types (Spanish, English monolinguals) and speech condition (English, Spanish speech).

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Data was recorded using Tucker-Davis Technologies (TDT) proprietary system and software. All speech stimuli were presented in the experiment using custom-written MATLAB2016b scripts [MathWorks, www.mathworks.com].

Data analysis These data were preprocessed using custom MATLAB2014a and python scripts, some of which can be found at: https://github.com/ChangLabUcsf/img_pipe. Electrode localization and reconstruction was performed using FreeSurfer 7.2.0 and the package from [https://github.com/ChangLabUcsf/img_pipe]. All subsequent data analysis and visualization were performed in MATLAB2024b using built-in scientific toolboxes, open-source community scripts, and custom-built scripts. The full analysis code is available at: <https://github.com/ChangLabUcsf/BhayaGrossman2025>, which contains documentation for all scripts and workflows.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data to replicate all figures is posted to Zenodo (10.5281/zenodo.17247450) and is available on request to the corresponding author (Edward.Chang@ucsf.edu). Because participants at the outset of this study were not consented for public data release, these data are only available upon reasonable request to respect patient privacy and consent.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Our study includes data collected from approximately equal numbers of male and female participants (of the 31 main participants, 15 participants self-reported as female). Sex was determined based on self-report. We did not analyze the data separately for males and females because we did not predict differences in our results based on sex.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity were not assessed in this study. Therefore no data on these variables are reported.
Population characteristics	This study included only participants diagnosed with intractable epilepsy. Data collection was performed with patients who had undergone an implantation procedure and were being monitored in the hospital as part of their clinical care. While adults with epilepsy can exhibit language deficits, we only collected data from patients who reported normal language function, both in production and in comprehension.
Recruitment	Participants were recruited from the Neurosurgical Department from the University of California, San Francisco and Fudan University, Shanghai. All patients who met the inclusion criteria (e.g. speakers of a non-English language with relevant neuro-anatomical coverage) and consented to participate in research were included in the study. To the best of our knowledge, we did not introduce biases into our recruitment process.
Ethics oversight	All protocols in the current study were approved by the University of California, San Francisco Committee on Human Research and by the Huashan Hospital Institutional Review Board of Fudan University. Participants gave informed written consent to take part in the experiments and for their data to be analyzed. Informed consent of non-English speaking participants at UCSF was acquired via a medically certified interpreter platform (Language-Line Solutions) and communication with research staff was facilitated by either in-person or video-call based interpreters who were fluent in the participant's native language.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<p>The sample size (n=34 participants) was determined by the number of eligible patients who consented to participate in the study over a period of 10 years. This approach was chosen since this study was conducted in a clinical setting, where recruitment was reliant on patient consent and participation.</p> <p>To inform the study design, we reviewed prior literature in which electrocorticography (ECoG) was used, which reported sample sizes ranging from 5 to 25 in similar studies (e.g. Sjerps, Fox, Johnson, & Chang (2019) Nature Communications; Hamilton, Edwards, & Chang (2018) Current Biology; Cogan et al (2014) Nature). Based on these reports, we estimated that a sample size of approximately 10 participants per group would be sufficient to detect statistically significant differences.</p>
Data exclusions	Neurophysiological data was excluded from analysis if pathological neural activity (namely, interictal epileptic discharges) was identified either during data collection or prior to analysis during visual inspection.
Replication	To assess the reproducibility of our findings, we conducted independent experiments with different participant cohorts (English speakers,

Replication	Spanish speakers, English-Spanish bilinguals, Mandarin speakers). The key findings of the study, namely that the neural encoding of acoustic phonetic features is preserved across native and foreign languages but for word-level features is enhanced, were replicated across all three independent cohorts.
Randomization	Randomization was not relevant in the present study, as participants were included based on language background. Covariates such as the hemisphere from which neural data was recorded, were controlled statistically in linear mixed effect models as random effects.
Blinding	Investigators were not blinded to the group allocation since data collection involved directly speaking with the participants and therefore being privy to their language background (the primary feature of interest).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A