nature biomedical engineering

Article

https://doi.org/10.1038/s41551-024-01207-5

A bilingual speech neuroprosthesis driven by cortical articulatory representations shared between languages

In the format provided by the authors and unedited

Contents

Supplementary notes

Note 1 | Assessment of the participant's articulatory inventory
The participant described in this study (he/him) was diagnosed with anarthria by a certified speech-language pathologist. Anarthria was secondary to brainstem stroke of the bilateral pons. Below, we summarize the results of the speech-language pathologist's assessment.

With regard to residual motor-movement, an oral-mechanism test was performed to assess jaw, tongue, and lip function. The participant has severe articulatory deficits, with movement of all three articulators being slow and effortful. The participant was unable to perform lip rounding or maintain a lip closure. Further, tongue movements were severely limited with only a slight ability to slowly extend and raise the tongue. Residual jaw opening and closing was the most reliable but was slow and effortful.

To characterize the reliability of syllabic level speech production during vocalized attempts, a perceptual dysarthria assessment was performed. The participant was unable to intelligibly produce sequences of multiple syllables. Breakdowns increasingly occurred as the number or words in the phrase or syllables per word increased. Rates of speech were slow, only 1–2 syllables per breath), and speech was overall characterized by spastic features, leading to articulatory imprecision. Thus, overall, the final diagnosis from speech language pathology was anarthria, defined as a loss of the ability to articulate speech with preserved comprehension.

Note 2 | Participant's AAC speed
We quantified the participant's communication speed using his common Augmentative and Alternative Communication (AAC) strategy. In brief, the participant uses residual head movements to control a laser pointer attached to his glasses to spell out words by directing the laser to letters on a board. We recorded the participant spelling out 30 sentences, 15 English and 15 Spanish, that were randomly chosen from the bilingual-test phrases. We then calculated the average words per minute over these 30 sentences. This is the value displayed on Fig. 1d (3 l97 words per minute). We also confirmed that the randomly drawn sentences had the same letters/word distribution as the full bilingual-test-phrases (Supplementary Fig. 4).

Supplementary methods

Method 1 | Articulatory sampling of the large-bilingual-phrase-set

To verify that the large-bilingual-phrase-set (Fig. 3a) spanned a range of articulatory features in each language, we performed a few analyses. First, we used a multilingual transformer based grapheme to phoneme conversion model to extract the phonemes across English and Spanish phrases in the large set [1]. Phonemes are defined as the smallest discriminable units of sound that make up a language [2]. For each unique English and Spanish phoneme present in our large-bilingual-phrase-set, we computed the number of occurrences. We note that many phonemes in each language are represented. Further, phonemes have place of articulation (POA) features, based on which vocal tract muscle is primarily used to produce the sound [2]. For example the ”b” sound is bilabial, being articulated by bringing the lips together. We computed the number of occurrences of each POA feature over the English and Spanish phrases. We notice a similar distribution of POA features between the two languages, with 5 primary articulatory groups represented. Overall, this provides quantitative evidence that the large-bilingual-phrase-set samples a diverse set of articulatory features in each language.

Method 2 | Language modelling

n-gram modelling

During phrase-decoding, we used two 5-gram language models, one in English and another in Spanish, given their ability to guide neural predictions towards linguistically valid phrases. A 5-gram language model emits a probability distribution over words in the vocabulary based on context from the up to 4 previous words. We used the same training procedure previously described [3]. We used back-off and discounting to improve the contextual estimates of the 5-gram model [4]. Additionally, we applied Kneser-Ney smoothing [5] to improve unigram model estimation. Specifically, we used a discounting rate of 0.9. 5-gram language models were trained in each language using 700 Spanish and 806 English sentences generated by bilingual volunteers independent from our laboratory.

Method 3 | Isolated-target classification model

Data preparation

We trained classification models on the isolated-target data for use in downstream phrase- decoding and offline characterizations of English and Spanish speech attempts. For each isolated-target trial, we used a 3.5 second window of neural features (0-3.5 s relative to the go-cue). The neural features were first decimated by a factor of 6 to 33.33 Hz. Next, each time sample was normalized to have an L2-norm of 1 across the HGA and LFS features, separately. These pre-processed neural features were then used to train downstream classifiers.

Modelling

Architecture and training

To capture temporal patterns in the neural features, we primarily used gated-recurrent unit (GRU) layers [6]. The pre-processed neural features are first consumed by a 1-dimensional convolutional layer with kernel size and stride of 4. This effectively downsamples the neural features by a factor of 4. The representations from the 1-d convolutional layer are then passed through a stack of 3 GRU layers with 260 units in each layer. Each GRU layer was bidirectional allowing the classifier to learn both forward and backward representations from the neural features. During training, a dropout rate of 0.8 was used minimize overfitting [7]. To compute a probability distribution over the 104 bilingual-words, the representation at the final time point of the final GRU layer was passed through a fully connected layer, with output size 104, yielding activation over the 104 bilingual-words. We next applied a softmax function to rescale the activation to a probability distribution.

Cross entropy loss was used to train the classifier. Mini-batch stochastic gradient descent (batch size 32) was used along with the Adam optimizer [8] to optimize network parameters. Training was stopped after 5 epochs with no improvement to the validation accuracy. We kept the model that had the highest performance on the validation set.

During phrase-decoding, model ensembling was used to improve classification performance by providing robustness to noise and random initialization of weights. We applied 10-fold stratified cross validation on all isolated-target data, collected prior to phrase-decoding, to train 10 unique classification models. During evaluation, the average probability distribution across the 10 models was computed.

Augmentations
To bolster classifier performance, we used time jittering, which has been shown to improve generalization and reduce overfitting for both images [9, 10] and neural activity [11, 12]. Additionally, we used a masking procedure during training to encourage the classification model to make predictions consistent with the vocabulary of each language. During training, before computation of the cross-entropy loss, we masked predictions that were in a different language than the language of the isolated-target word.

Model pre-training and fine-tuning
A classifier was first trained, using the above-described procedure, on a previously collected isolated-target dataset where the participant silently attempted to speak bilingual-words (rather than producing audible grunts/moans). The weights of this classifier were used to initialize the 10 classifiers trained on the core isolated-target dataset where the participant attempted to speak bilingual-words, expressing some residual, uncoordinated grunts/moans. This process expedited training of the core models used for downstream analysis on the bilingual-words and phrase-decoding.

Hyperparameter optimization
We optimized the number of GRU layers, the number of hidden units in each GRU layer, the kernel size and stride of the convolutional layer, the dropout rate, and the learning rate. We used isolated-target trials from the two prior recording sessions before the start of the phrase-decoding as a held-out test set. We trained models on all other isolated-target trials and evaluated on the held-out test set. We searched the ranges provided in (Supplementary Table 10) for each hyperparameter and kept the parameters that produced the highest accuracy on the held-out test set. Training was conducted in the same manner as previously described.

Method 4 | Sentence decoding

Task infrastructure
The phrase-decoding system was designed to decode English and Spanish phrases from cortical activity without the user having to manually select a language. The system continuously acquired neural features, high-gamma activity (HGA; 70-150 Hz) and low-frequency signals (LFS; 0.3-100 Hz), from the local field potential (LFP) of each electrode on the participant's array. The temporal flow of information through the system is as follows (Extended Data Fig. 1). The participant first makes a speech attempt which is detected by the system, through the speech detection model, and cues activation of an ongoing decoding process. Following the system activating, sequential 3.5s windows are shown to the participant. At the end of each window, after the full 3.5s have passed, the corresponding neural features from that window are passed to the decoding process illustrated in Fig. 1a and Extended Data Fig. 2. There is then a latency to conduct the beam search and decoding, after which the most likely beam from the process is displayed to the participant monitor. This will continue to occur until a 3.5 window passes were there is no detected speech event. After such a window, the decoding is finalized and terminated. The system then resets and waits for another speech attempt.

Verb and adjective conjugations
Given that verbs and adjectives, especially in Spanish, may have multiple different conjugations (i.e. traer: traigo, traes, trae, traemos, traen, trayendo), we broadcast the neural probability for the unconjugated form of the verb to all the conjugated forms. This led to an increased vocabulary size of 111 Spanish and 70 English words. The language models were trained on this increased vocabulary size in each language. The beam search is also conducted over these increased vocabulary sizes, and the language model is readily able to infer the correct conjugation based on surrounding context of the verb in the overall phrase.

Beam search
As described previously, we used a bilingual beam search to flexibly decode English and Spanish phrases from cortical neural activity. The goal of this process is to find the most likely sequence of words (s^*) given the neural data X. Here we provide more mathematical detail to the approach described in the main text. As previously described, an ensembled prediction across the bilingual-words is generated for each 3.5 second

window in the trial. These predictions are split and re-scaled across English and Spanish words. Next, a language- specific beam search is applied to find s*, the most likely phrase in each language, based on likelihood under the neural data and language models. This can be expressed mathematically as:

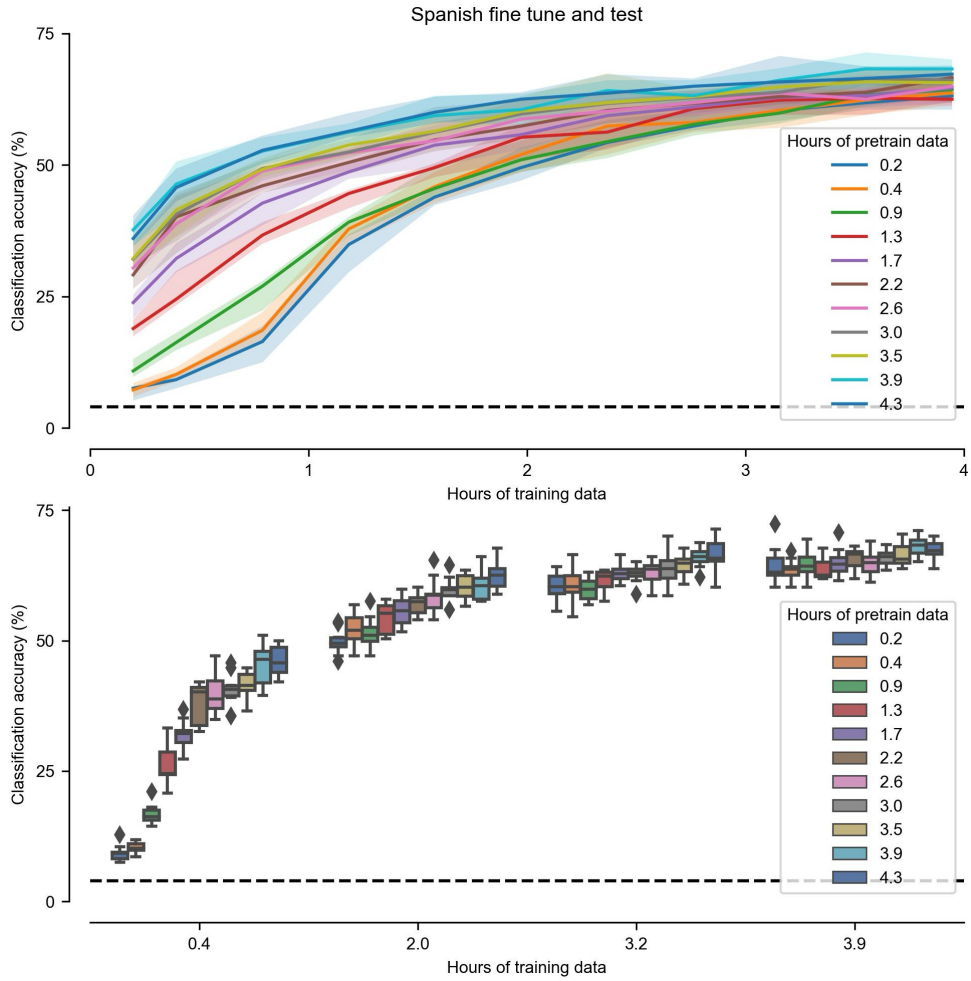
$$s^* = argmax_s P_{nc}(s|X) * P_{lm}(s^\alpha) * |s|^\beta$$

Here, Pnc(sIX) is the probability of s estimated by the ensembled neural classifier given each 3.5s window of neural activity. This is equivalent to the product of the neural classification probability of each word in s. Plm(sα) is the probability of the phrase s under the language- specific natural language-model down weighted by a factor alpha. The parameter lsIβ introduces a word-insertion bonus, β, to counter the decreasing probability under the language model with each word added. The notation lsl indicates the number of words in s. Functionally, since we rely on a cued window approach where the number of words in s* is known based on the number of windows, β has no effect on the decoded phrase; however, it better normalizes scores for number of words, aiding offline analyses.

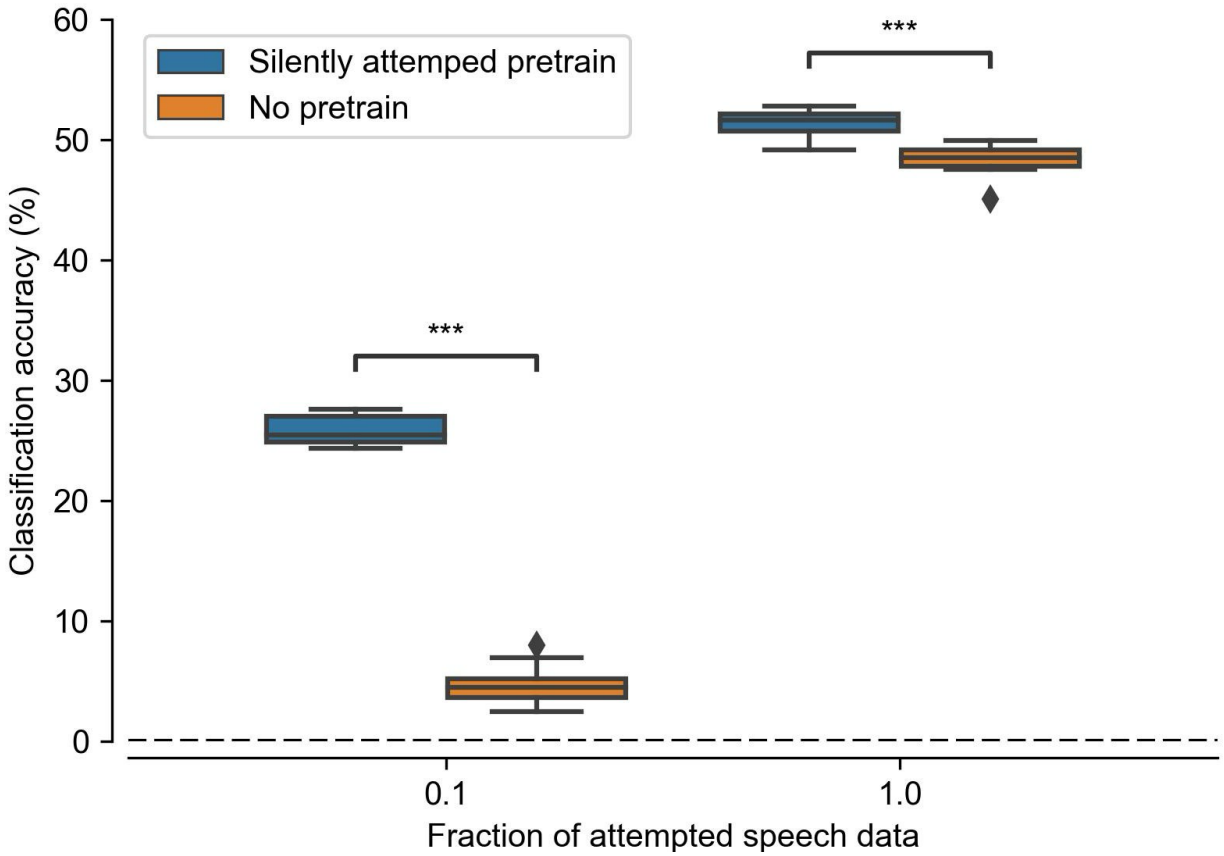
To approximate s* at each timepoint t = τ we used an iterative beam-search as in [3]. A list of the B most likely phrases from t= τ – 1 (or an empty string if t=1) is used as a set of candidate prefixes. Here B, denotes the beam width: the number of candidate phrases evaluated. For each candidate prefix l and each word in the vocabulary (w), we constructed a new phrase by adding w to l. We then re-scored these phrases under the formulation provided above. We select the most likely candidate phrase, s*, at t = τ as the new prefix (l + w) with the highest score.

Phrase decoding is run separately for English and Spanish. At each cued-window the highest scoring English beam is compared to the highest scoring Spanish beam. The beam with the higher overall score is displayed to the participant as feedback and effectively sets the language of the output.
To chose values for α, β, and B we used a hyperparameter optimization procedure on simulations with data from the isolated-word task. We selected 20 random phrases from the generated bilingual corpus, excluding phrases that appeared in the bilingual-test-phrases. We then used neural predictions corresponding to isolated-task trials for each word in the phrases to simulate the phrase-decoding procedure. We swept a range of values for α, β, and B (Supplementary Table 10) and kept the values that minimized the overall word error rate of the system on the validation phrases.

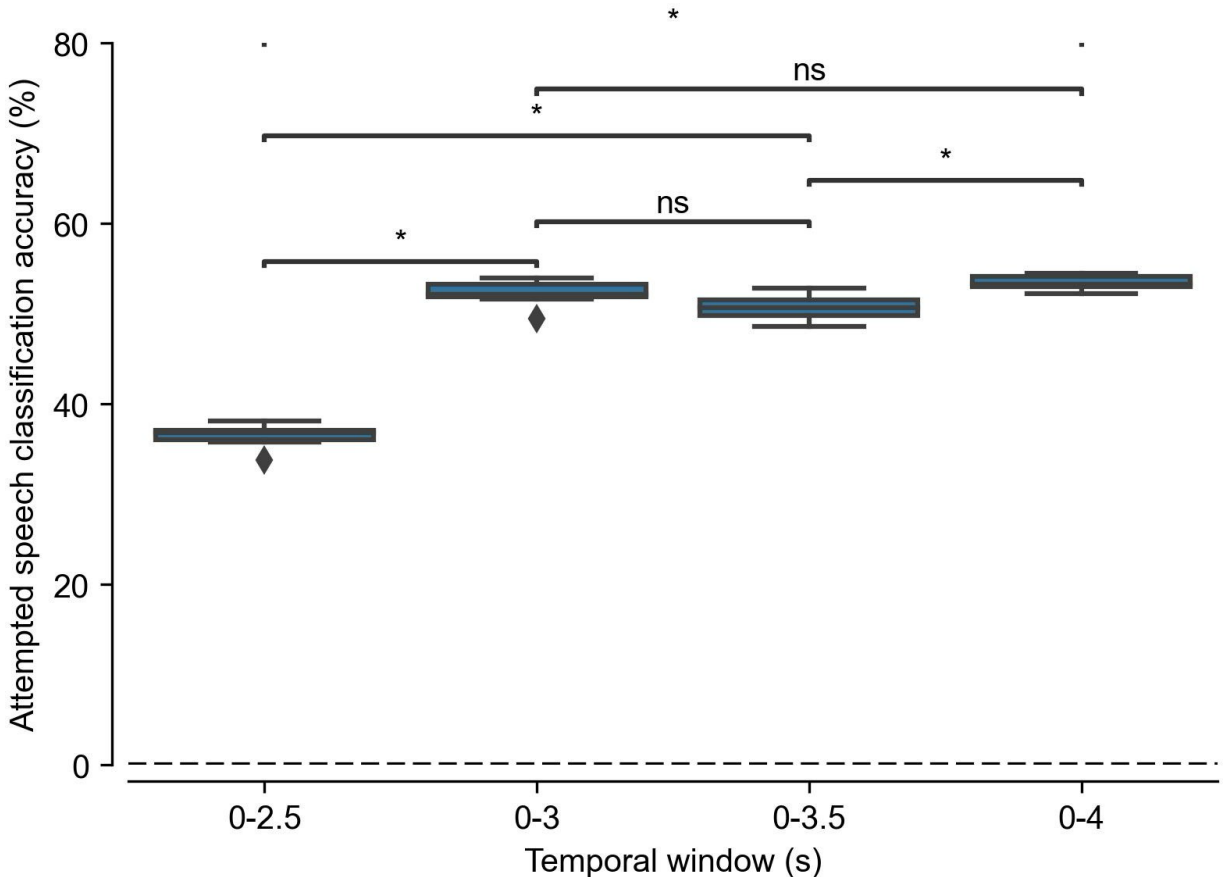
Supplementary figures



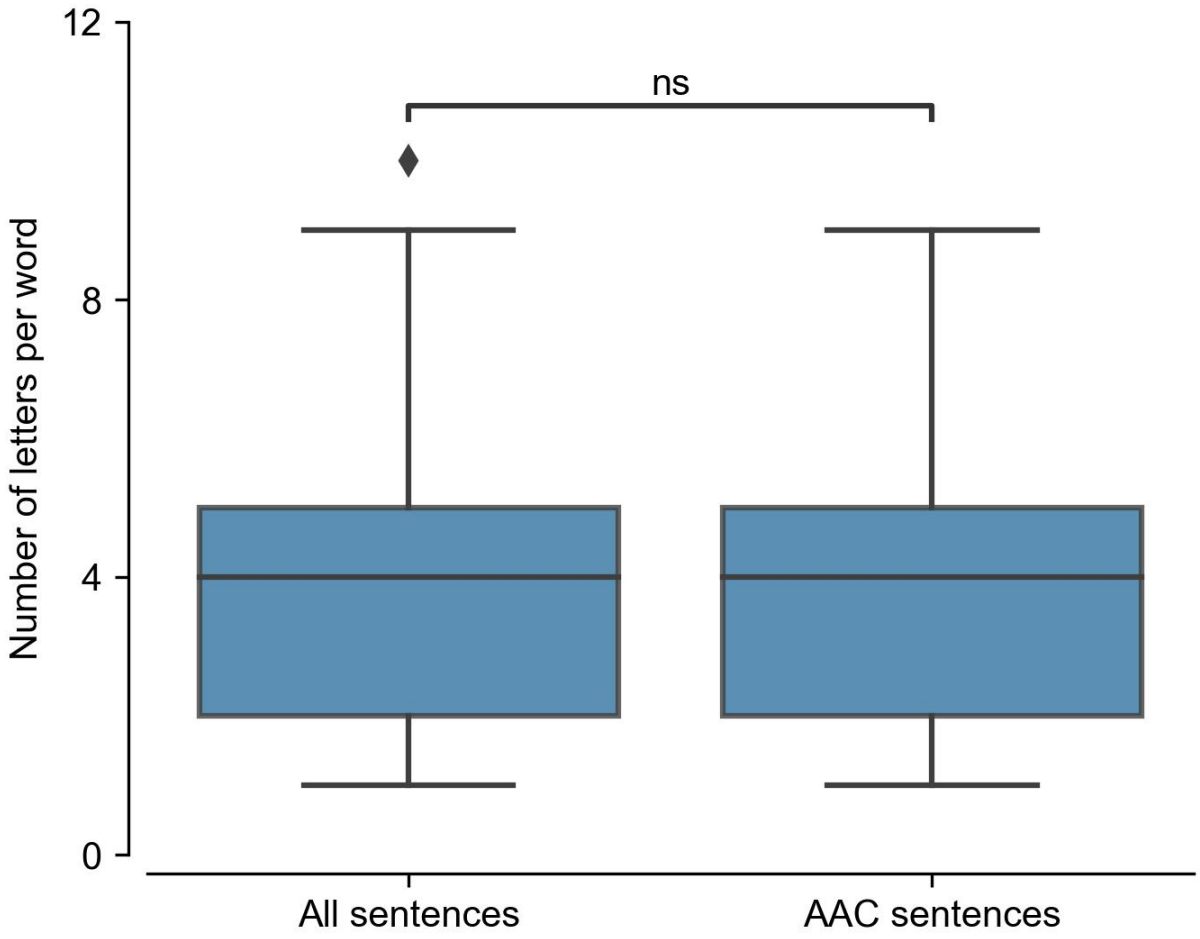
Supplementary Fig. 1 | Effect of amount of pretrain data on transfer learning efficacy. (top) Shown are learning curves on the Spanish vocabulary (Fig. 4b) with transfer learning from English models trained on varying amounts of data. The number of hours of pretraining data has an effect on transfer learning efficacy; however, this effect saturates. Lines represent median accuracy and shaded error bars are 99% confidence intervals. (bottom) Shown is the classification accuracy at select amounts of training data, again as a function of the hours of data used to pretrain the English models. The distribution of accuracy across hours of pretraining data is easier to visualize. Distributions in both panels are over 10 non-overlapping folds. Box plots in all panels depict median (horizontal line inside box), 25th and 75th percentiles (box), 25th and 75th percentiles +/- 1.5 times the interquartile range (whiskers), and outliers (diamonds).



Supplementary Fig. 2 l Effect of pre-training with silently attempted speech data. Shown is performance of classification models, over the full 104 bilingual words, at two fractions of total attempted speech data: 0.1 and 1. At both fractions, pre-training with the limited silently attempted speech data helps improve performance, albeit marginally with full use of the attempted speech data (Median 21.00% and 3.12 % improvement at 0.1 and 1, respectively, *** P = 0.00018 and P = 0.00044, two-sided Mann-Whitney U-test). Dashed line indicates chance performance (0.96%). Box plots in all panels depict median (horizontal line inside box), 25th and 75th percentiles (box), 25th and 75th percentiles +/- 1.5 times the interquartile range (whiskers), and outliers (diamonds).



Supplementary Fig. 3 l Performance of attempted speech model using windows of various input lengths. We evaluated classification performance over the full 104 bilingual words, using models trained with varying length temporal windows of neural features. Classification training and evaluation occurred identically to described in the methodology and each distribution is over 10-fold cross validation (* $P < 0.01$; two-sided Mann-Whitney U-test with 6-way Bonferroni correction for multiple comparisons). For 0-2.5 vs 0-3, 0-3 vs 0-3.5, 0.35 vs 0.4, 0-2.5 vs 0-35, 0-3 vs 0-4, and 0-2.5 vs 0-4, p-values are 0.0011, 0.1, 0.0019, 0.0011, 0.32, and 0.0011. Dashed line indicates chance performance (0.96%). Box plots in all panels depict median (horizontal line inside box), 25th and 75th percentiles (box), 25th and 75th percentiles +/- 1.5 times the interquartile range (whiskers), and outliers (diamonds).



Supplementary Fig. 4 l Comparison of all online-evaluation sentences and AAC evaluation subset. For each word in each of the online-evaluation sentences, we computed the number of letters. We did the same for the 30 random sentences drawn for the AAC evaluation. Distributions are over 126 words in the AAC sentences and 227 words in the evaluation sentences with no significant difference (NS P = 0.71, two-sided Mann-Whitney U-test). Box plots in all panels depict median (horizontal line inside box), 25th and 75th percentiles (box), 25th and 75th percentiles +/- 1.5 times the interquartile range (whiskers), and outliers (diamonds).

Supplementary tables

Supplementary Table 1 | Bilingual-words used in isolated-target task.

English	Spanish	Both
i	yo	no
we	nosotros	a
you	tu	Pancho
your	ella	
she	ustedes	
they	nos	
my	mi	
not	sí	
yes	por favor	
if	el	
please	la	
the	ahora	
now	dónde	
where	y	
to	muy	
and	hola	
very	cómo	
hello	familia	
how	sobrino	
family	alimento	
nephew	espalda	
food	casa	
back	coche	
house	tienda	
vehicle	enfermero	
store	hermana	
nurse	agua	
sister	almuerzo	
water	aquí	
lunch	ingles	
here	español	
english	ropa	
spanish	nombre	
clothes	hablar	
name	estar	
speak	ser	
is	querer	
want	tener	
have	comer	
eat	ir	
go	traer	
bring	poder	
can	gustar	
like	divertida	
funny	rápidx	
fast	reflexivo	
thoughtful	caliente	
hot	despierto	
awake	buenx	
good	nuevo	
new		

Supplementary Table 2 | Bilingual-words used in phrase-decoding.

English	English	Spanish	Spanish	Both
i	wanting	yo	quieres	pancho
we	liking	nosotros	quiere	a
you	bringing	tú	quieren	no
your	going	ella	queremos	
she	eating	ustedes	tengo	
they	having	nos	tienes	
my	speaking	mi	tiene	
not		sí	tienen	
yes		por-favor	tenemos	
if		el	teniendo	
please		la	voy	
the		ahora	vas	
now		dónde	va	
where		y	van	
to		muy	vamos	
and		hola	como	
very		cómo	comes	
hello		familia	come	
how		sobrino	comen	
family		alimento	comemos	
nephew		espalda	comiendo	
food		casa	traigo	
back		coche	traes	
house		tienda	trae	
vehicle		enfermero	traen	
store		hermana	traemos	
nurse		agua	trayendo	
sister		almuerzo	puedo	
water		aquí	puedes	
lunch		ingles	puede	
here		español	pueden	
english		ropa	podemos	
spanish		nombre	gusta	
clothes		hablar	gustan	
name		estar	gustando	
speak		ser	divertido	
is		querer	divertida	
want		tener	rápida	
have		comer	rápido	
eat		ir	reflexivo	
go		traer	reflexiva	
bring		poder	caliente	
can		gustar	despierto	
like		hablo	despierta	
funny		hablas	bueno	
fast		habla	buena	
thoughtful		hablan	nuevo	
hot		hablamos	nueva	
awake		hablando		
good		estoy		
new		estás		
speaks		está		
are		están		
am		estamos		
has		soy		
eats		eres		
goes		es		

brings	son
likes	somos
wants	quiero

Supplementary Table 3 | Copy-typing task sentences.

English sentences	Spanish sentences
bring back the water	ahora ustedes van
can you please go	aquí está la comida
eat your lunch	cómo está mi espalda
hello are you awake	cómo está mi hermana
hello how are you	dónde está la tienda
hello my name is pancho	el alimento es bueno
here we go	el almuerzo es caliente
i have a nephew	el coche es rápido
i like my new clothes	el enfermero es reflexive
i like your family	ella quiere ir a casa
i speak english and spanish	hola mi nombre es pancho
my family speaks spanish	mi familia esta aquí
my family wants lunch	mi hermana es muy divertida
my nurse is bringing water	mi sobrino tiene el poder
no not now	nosotros estamos comiendo
please bring hot water	nosotros nos gusta comer
please have my back	por favor trae el alimento
she goes to the store	sí yo puedo hablar
the house is new	sí yo quiero agua
the nurse is thoughtful	trae ropa nueva
the store has good food	tú puedes ir ahora
the vehicle is fast	yo estoy despierto
they like the food	yo estoy hablando español
where is my nephew	yo hablo español y ingles
where is the food	yo no tengo
yes i can speak	yo puedo comer ahora
yes if you want	yo quiero agua caliente
your sister is very funny	yo quiero comer

Supplementary Table 4 | Statistical comparisons of word error rates across languages and decoding-paradigms.

Language[1]	Paradigm 1	Paradigm 2	Test-statistic	P-value (corrected)[2]
Overall	Chance	Neural-only	4.41e+02	1.91e-07
Overall	Chance	Real-time	4.41e+02	1.91e-07
Overall	Neural-only	Real-time	4.40e+02	1.91e-07
Spanish	Chance	Neural-only	4.41e+02	1.91e-07
Spanish	Chance	Real-time	4.41e+02	1.91e-07
Spanish	Neural-only	Real-time	4.13e+02	2.66e-06
English	Chance	Neural-only	4.40e+02	1.91e-07
English	Chance	Real-time	4.41e+02	1.91e-07
English	Neural-only	Real-time	3.74e+02	1.24e-04

[1]Each comparison is a two-sided Mann-Whitney U-test with 9-way Holm-Bonferroni correction for multiple comparisons across 21 real-time sentence-decoding blocks.
[2]9-way Holm-Bonferroni correction for multiple comparisons.

Supplementary Table 5 | Statistical comparisons of language classification accuracy across decoding-paradigms.

Paradigm 1[1]	Paradigm 2	Test-statistic	P-value (corrected)[2]
Chance	Neural-only	1.05e+02	4.79e-03
Chance	Real-time	1.50e+01	5.44e-07
Neural-only	Real-time	1.02e+02	4.79e-03

[1]Each comparison is a two-sided Mann-Whitney U-test with 3-way Holm- Bonferroni correction for multiple comparisons across 21 real-time sentence- decoding blocks.
[2]3-way Holm-Bonferroni correction for multiple comparisons.

Supplementary Table 6 | Statistical comparisons of word error rates with manual setting of the target language.

Language[1]	Paradigm 1	Paradigm 2	Test-statistic	P-value (corrected)[2]
Overall	Chance	Real-time	4.41e+02	6.41e-08
Spanish	Chance	Real-time	4.41e+02	6.41e-08
English	Chance	Real-time	4.41e+02	6.41e-08

[1]Each comparison is a two-sided Mann-Whitney U-test with 3-way Holm-Bonferroni correction for multiple comparisons across 21 real-time sentence-decoding blocks.
[2]3-way Holm-Bonferroni correction for multiple comparisons.

Supplementary Table 7 | Large bilingual phrase set.

English sentences	English sentences (cont)	Spanish sentences	Spanish sentences (cont)
stop that hurts	they went outside	soy gracioso	me gusta el pastel
you like pizza	how do you write	hola como estás	no ahora no
this weekend is free	on my way	no puedo dormir	usa un casco
wear a jacket	use a helmet	por favor dile a mi familia	estoy impresionado
call me back	my uncle is fun	por dentro es cálida	por qué hace frio
do me a favor	swim in the river	ellos corrieron lento	conduce mi coche
you bike quickly	you seem confused	tenedor o cuchara	cuándo esta aquí
look it is snowing	we play outside	todos ustedes son amables	trae a tus amigos
your house is large	pen or pencil	estas enfermo	las ciudades están ocupadas
chicken is tasty	she is on vacation	mi tío es divertido	cierra la ventana
he walks fast	trees are green	sí entra	azul es un color
they eat soup	your sister is tall	puedo oír	llama a tu hermano
is the water clean	is it raining	te gusta la pizza	este fin de semana está libre
dinner is ready	let me see	tu hermana es alta	´échame una mano
they ran fast	fruits or vegetables	esta es mi casa	frutas o verduras
my back hurts	go to your room	dame un segundo	tu casa es grande
drive my car	you all are nice	mira está nevando	nadar en el río
faith is good	i am impressed	camisa roja y pantalones	estoy emocionado
close the door	sleep with a pillow	pasea al perro	su pelo es largo
shut the window	cities are busy	el pollo es sabroso	avión tren o autobús
yes come in	i like cake	trae mis lentes	los perros son lindos
do you like music	blue is a color	ellos comen sopa	te gusta la música
give me a hand	i can hear	para eso me duele	esto es cómodo
we are lucky	hello how are you	como escribes	déjame ver
why is it cold	where did you go	me duele la espalda	no te preocupes demasiado
cats are shy	bring my pillow	aquí vamos	tíralo a la basura
is that your wife	how will you feel	la computadora está apagada	duerme con una almohada
against the law	are you sick	huevos y tostadas	cierra la puerta
the coffee is perfect	when is it here	la fe es buena	nosotros jugamos afuera
how is your heart	bring your friends	bolígrafo o lápiz	ella está de vacaciones
no not now	i enjoy sports	está lloviendo	´el camina rápido
it is comfortable	i am funny	hazme un favor	ahora tengo hambre
please tell my family	let's have fun	vaso de leche	cuando debemos ir
this is my home	eat a banana	cómo te sentirás	los árboles son verdes
give me a second	yesterday i left	tenemos suerte	esto está limpio
throw it away	i received the package	el cielo está nublado	adonde fuiste
the road is wet	red shirt and pants	el camino está mojado	llámame por favor
the computer is off	i like my nurse	me gusta mi enfermera	andas en bicicleta rápido
walk the dog	turn on the television	cómo está tu corazón	esa es tu esposa
i am excited	plane train or bus	leo libros	pareces confundido
dogs are cute	come here please	disfruto de los deportes	ven aquí por favor
now i am hungry	when should we leave	el agua está limpia	la cena está lista
i can't sleep	bring my glasses	ellas salieron	estoy en camino
call your brother	it is clean	vamos a divertirnos	ayer me fui
cup of tea	inside in warm	trae mi almohada	taza de té
don't worry too much	hurry home please	come una banana	usar una chaqueta
fork or spoon	glass of milk	apúrate a casa por favor	ve a tu cuarto
here we go	what a catch	recibí el paquete	los gatos son tímidos
her hair is long	the sky is cloudy	vaya trampa	enciende el televisor
i read books	eggs and toast	el café está perfecto	contra la ley

Supplementary Table 8 | English and Spanish syllable paired words.

English words	Spanish words	Shared syllable
boulder	bolsa	Bol
casket	castor	Cas
crazy	crece	cre/crae
petal	pesar	Pe
soda	fonda	Da
soda	solas	So
widen	pierden	den

Supplementary Table 9 | Model architecture, training parameters, and detection parameter values for the speech detection model.

Parameter type	Parameter description	Value
Model Architecture	Number of LSTM layers	3
	Nodes per LSTM layer	128, 64, 32
	LSTM type	Unidirectional
Training parameters	Optimizer	Adam
	Loss	Cross-entropy
	Learning rate	0.001
	Batch size	512
	Dropout	0.5
Detection parameters	Smoothing size	78
	Probability threshold	0.50, 0.49*
	Time threshold duration	100

*Probability threshold was slightly adjusted (to the lower value) during the second day of online testing based on qualitative observations.

Supplementary Table 10 | Hyperparameter definitions and values for classification and beam search.

Model	Hyperparameter description	Search- space type	Value range	Optimal values
Word	Number of GRU layers	by 1	[1, 4]	3
classifier	Nodes per GRU layer	by 10	[200, 300]	260
	Dropout fraction	by 0.05	[0.5, 0.9]	0.8
	Convolution kernel size and skip	by 1	[1, 6]	4
Beam	Language-model scaling factor, α	by 0.1	[0.1, 1.0]	0.9
search	Word-insertion weight, β	by 1	[0, 15]	11
	Number of beams maintained, B	by 1	[0, 150]	150*

*Higher values were not tested due to increasing computational burden.

Supplementary video captions

Supplementary Video 1 | A demonstration of online word-by-word bilingual sentence decoding from the brain of a participant with paralysis. Decoding is initialized when a speech attempt is detected from cortical activity by the system. The participant next attempts to say each word in the target phrase (top of screen) during consecutive 3.5 second cued windows. The green underline that appears after the countdown indicates the go-cue for each consecutive window. The system deactivates when a speech attempt is not detected during a 3.5 second cued window. During each cued window, a classifier predicts the likelihood of each English and Spanish word in the vocabulary, based on the cortical activity. These likelihoods are combined with natural-language models of English and Spanish to output the most likely sentence, across languages, at each timepoint. The language (English or Spanish) of the most likely sentence is not manually set by the participant; rather, it is freely decoded by the classifier and language-models. The system better predicts the intended language as the phrase progresses, highlighting the importance of linguistic context captured by the language models.

Supplementary Video 2 | The same demonstration of online word-by-word bilingual sentence decoding as in Supplementary Video 1 for a different set of 3 sentences.

Supplementary Video 3 | The participant uses the bilingual speech neuroprosthesis to have an open-ended conversation with a member of the research team. He freely switches between using English and Spanish. The system works as in Supplementary video 1; however, the participant now produces any sentence he desires, using the system vocabulary. Translations of each sentence are shown for clarity.

Supplementary references

1. Zhu J, Zhang C, and Jurgens D. ByT5 model for massively multilingual grapheme-to- phoneme conversion. arXiv preprint arXiv:2204.03067 2022.
2. Ladefoged P and Johnson K. A course in phonetics. Cengage learning, 2014.
3. Metzger SL, Liu JR, Moses DA, et al. Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis. Nature Communications 2022;13:6510.
4. Jurafsky D and Martin JH. Speech and language processing. Vol. 3. US: Prentice Hall 2014.
5. Kneser R and Ney H. Improved backing-off for m-gram language modeling. In: 1995 international conference on acoustics, speech, and signal processing. Vol. 1. IEEE. 1995:181–4.
6. Cho K, Van Merri¨enboer B, Bahdanau D, and Bengio Y. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 2014.
7. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, and Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 2012.
8. Kingma DP and Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014.
9. Krizhevsky A, Sutskever I, and Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 2012;25:1097–105.
10. Reed CJ, Metzger S, Srinivas A, Darrell T, and Keutzer K. Selfaugment: Automatic augmentation policies for self-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:2674–83.
11. Willett FR, Avansino DT, Hochberg LR, Henderson JM, and Shenoy KV. High- performance brain-to-text communication via handwriting. Nature 2021;593:249–54.
12. Moses DA, Metzger SL, Liu JR, et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. New England Journal of Medicine 2021;385:217–27.