nature neuroscience

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41593-020-0608-8

In the format provided by the authors and unedited.

Machine translation of cortical activity to text with an encoder–decoder framework

Joseph G. Makin[1,2] ✉, **David A. Moses**[1,2] and **Edward F. Chang**[1,2] ✉

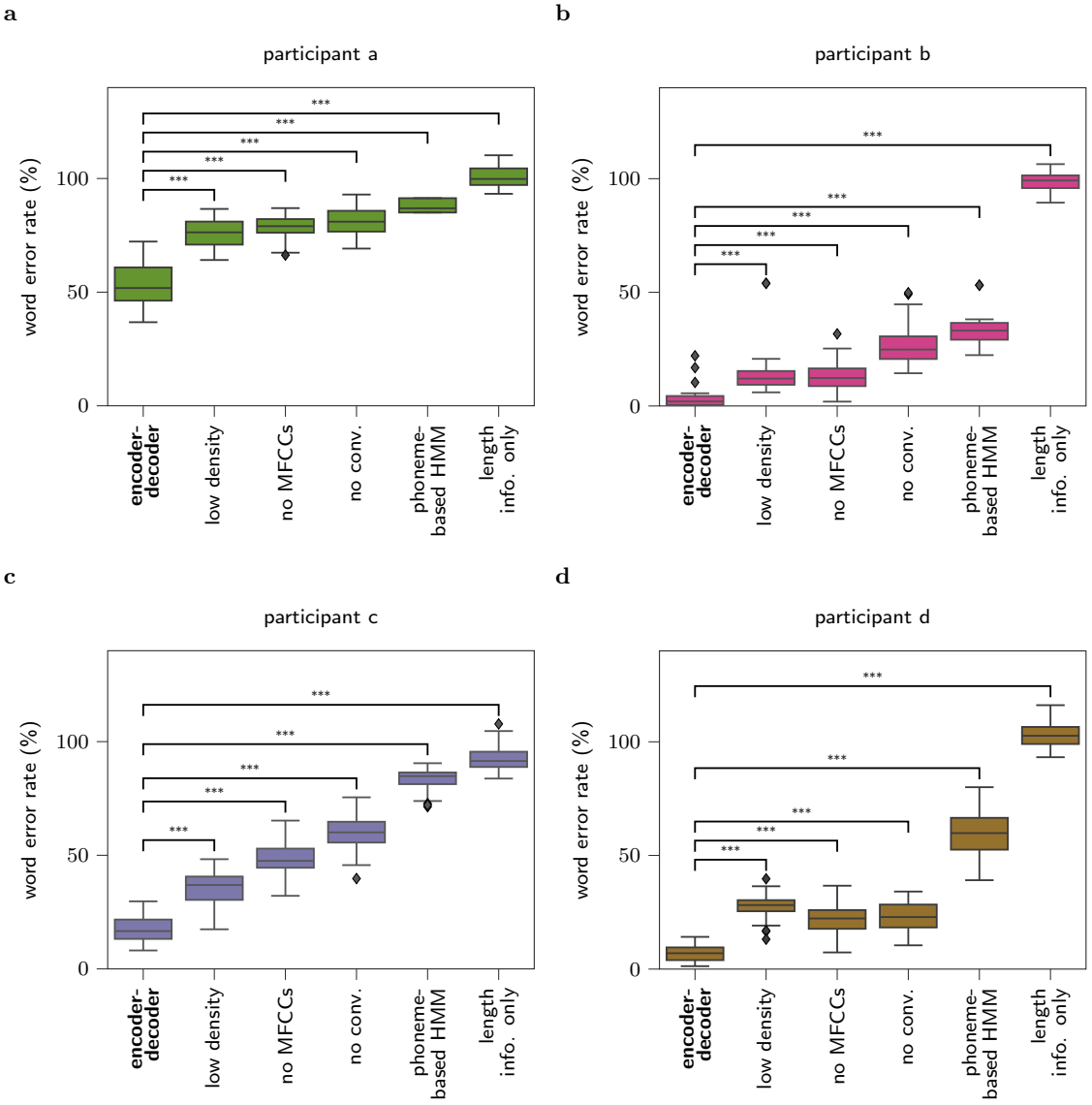
[1]Center for Integrative Neuroscience, UCSF, San Francisco, CA, USA. [2]Department of Neurological Surgery, UCSF, San Francisco, CA, USA.
✉e-mail: makin@phy.ucsf.edu; edward.chang@ucsf.edu

Machine translation of cortical activity to text with an encoder-decoder framework – Supplementary Information

Joseph G. Makin[1,2], David A. Moses[1,2], and Edward F. Chang[1,2]

[1]Center for Integrative Neuroscience, UCSF, San Francisco, California, USA
[2]Department of Neurological Surgery, UCSF, San Francisco, California, USA

Correspondence should be directed to EFC (edward.chang@ucsf.edu, experiment/data) or JGM (makin@phy.ucsf.edu, algorithm/code).



Supplementary Figure 1: Word error rates (WERs) under the encoder-decoder and various competitor decoders, for all four participants. This figure reprises Figure 2a for all four participants (participant **b** is repeated here). Each subfigure (i.e., participant) shows the distribution of WERs under the encoder-decoder (first bar), four crippled variants thereof (bars 2-4 and 6), and a state-of-the-art sentence classifier based on ECoG-to-phoneme Viterbi decoding ("phoneme-based HMM"). Abbreviations: "no MFCCs": trained without requiring the encoder to predict MFCCs; "low density": trained and tested on simulated lower-density grid (8-mm rather than 4-mm spacing); "no conv.": the network's temporal convolution layer is replaced with a fully connected layer; "length info. only": the input ECoG sequences are replaced with Gaussian noise—but of the correct length. The box and whiskers show, respectively, the quartiles and the extent (excepting outliers which are shown explicitly as black diamonds) of the distribution of WERs across $n = 30$ networks trained from scratch and evaluated on randomly selected held-out blocks. Significance for each subject, indicated by stars (***: $p < 0.0005$), was computed with a one-sided Wilcoxon signed-rank test, and Holm-Bonferroni corrected for five comparisons. Exact p values appear in Table 5.

pic. #	sentence
1	part of the cake was eaten by the dog
	several adults and kids are in the room
	the little boy is crying because the dog ate his cake
	the mother is angry at her pet dog
	under the sofa is a hiding dog
	the woman is holding a broom
	there is a partially eaten cake on the large table
	four candles are lit on the cake
	the guests arrived with presents
	the child is turning four years old
2	while falling the boy grabs a cookie
	the boy is reaching for the cookie jar
	there is chaos in the kitchen
	water is overflowing from the sink
	if only the mother could pay attention to her children
	the stool is tipping over
	the little girl is giggling
	his sister is helping him steal a cookie
	bushes are outside the window
	i think their water bill will be high
3	the firemen are coming to the rescue
	the girl was riding a tricycle
	which ladder will be used to rescue the cat and the man
	the cat does not want to come off the tree branch
	in the tree there is a cat a man and a bird
	a dog is barking at the man in the tree
	a little bird is watching the commotion
	worried by the dog the man considers jumping
	the cat doesnt seem interested in coming down
	how did the man get stuck in the tree

Supplementary Table 1: The picture descriptions read by participants **c** and **d**. N.B. that patients did not view the pictures, but the subsets associated with each picture (pic. #) were sometimes presented in separate blocks.

	participant:		a		b		c	d		
		data set:	MT-1	MT-*	MT-1	MT-*	PD	MT-1	MT-*	PD
training	sentence	types	50	460	50	460	30	50	460	30
		tokens	100	924	450	860	559	100	909	740
	word	types	239	1787	240	1787	122	238	1745	123
		tokens	610	6897	2740	5890	5453	607	6729	7292
validation	sentence	types	50	50	50	50	30	50	50	30
		tokens	50	50	50	50	60	50	50	82
	word	types	238	238	239	239	122	230	230	122
		tokens	304	304	303	303	592	302	302	809

Supplementary Table 2: Data sets for training and testing, broken down by participant. MT-1 = MOCHA-TIMIT, first set of 50 sentences; MT-* = MOCHA-TIMIT, full set of 460 sentences for training, first set of 50 for testing; PD = picture descriptions. The numbers of tokens are given for a (typical) fold of cross validation but in practice could vary slightly because the cross-validation procedure partitioned the data by blocks rather than sentences. The numbers of sentence types are nominal, i.e. were not increased to reflect (rare) participant misreadings.

layer type	**# units**	**connectivity**	**nonlinearity**
encoder embedding	[100]	temporal conv. (width=stride=12)	none
encoder RNN	$[400, 400, 400] \times 2$	bidirectional	LSTM
encoder projection	$[225, 13]$	full	[ReLU, none]
decoder embedding	[150]	full	ReLU
decoder RNN	[800]	unidirectional	LSTM
decoder projection	[1806] or [125]	full	softmax

Supplementary Table 3: Architecture hyperparameters.

parameter	**value**
learning rate	0.0005
feed-forward dropout	0.1
RNN dropout	0.5
MFCC-penalty weight, λ	0.1
examples/mini-batch	256
EMA decay, η	0.99
# training epochs	800
TL: # pre-training epochs	200
TL: # training epochs	60
TL: # post-training epochs	540

Supplementary Table 4: Training hyperparameters. RNN = recurrent neural network, MFCC = Mel-frequency cepstral coefficients, EMA = exponential moving average (see Methods); TL = transfer learning.

participant	baseline model:	low density	no MFCCs	no conv.	phoneme-based HMM	length info. only
a	p value	8.7e-07	9.6e-07	8.7e-07	8.7e-07	8.7e-07
	test statistic	465	464	465	465	465
	effect size	1.00	1.00	1.00	1.00	1.00
b	p value	8.7e-07	9.6e-07	8.7e-07	8.6e-07	8.7e-07
	test statistic	465	464	465	465	465
	effect size	1.00	1.00	1.00	1.00	1.00
c	p value	8.7e-07	8.7e-07	8.7e-07	8.7e-07	8.7e-07
	test statistic	465	465	465	465	465
	effect size	1.00	1.00	1.00	1.00	1.00
d	p value	8.7e-07	8.7e-07	8.7e-07	8.7e-07	8.7e-07
	test statistic	465	465	465	465	465
	effect size	1.00	1.00	1.00	1.00	1.00

Supplementary Table 5: Complete statistics for the comparison of the encoder-decoder to various "baseline" models, shown in Figure 2a (participant **b**) and Supplementary Figure 1 (all participants). All comparisons were made across $n = 30$ independently trained models with a one-sided Wilcoxon signed-rank test. The reported test statistic is therefore the sum of positive signed ranks, and the effect size is the rank correlation. Before determining statistical significance, the p values reported here were Holm-Bonferroni corrected within each subject for five comparisons.

competitor (row) vs. baseline (col)		**a/MT-1**	**b→a/MT-1**
b→a/MT-1	p value	9.9×10^{-6}	–
	test statistic	440	–
	effect size	0.89	–
a/MT-*	p value	8.7×10^{-7}	–
	test statistic	465	–
	effect size	1.0	–
b→a/MT-*	p value	1.1×10^{-6}	3.6×10^{-2}
	test statistic	463	320
	effect size	0.99	0.38
		b/MT-1	**a→b/MT-1**
a→b/MT-1	p value	1.3×10^{-3}	–
	test statistic	409	–
	effect size	0.76	–
b/MT-*	p value	1.9×10^{-3}	–
	test statistic	373	–
	effect size	0.60	–
a→b/MT-*	p value	1.3×10^{-6}	1.9×10^{-4}
	test statistic	461	405
	effect size	0.98	0.74
		d/MT-1	**b→d**/MT-1
b→d/MT-1	p value	2.0×10^{-5}	–
	test statistic	432	–
	effect size	0.86	–
d/MT-*	p value	1.0	–
	test statistic	45	–
	effect size	-0.81	–
b→d/MT-*	p value	0.99	1.8×10^{-2}
	test statistic	122	335
	effect size	-0.48	0.44
		c/PD	**d→c/PD**
d→c/PD	p value	1.0	–
	test statistic	39.0	–
	effect size	-0.83	–
		d/PD	**c→d/PD**
c→d/PD	p value	0.28	–
	test statistic	244	–
	effect size	0.12	–

Supplementary Table 6: Complete statistics for the transfer-learning hypothesis tests reported in the main text. MT-1 = MOCHA-TIMIT, first set of 50 sentences; MT-* = MOCHA-TIMIT, full set of 460 sentences for training, first set of 50 for testing; PD = picture descriptions. All comparisons were made across $n = 30$ independently trained models with a one-sided Wilcoxon signed-rank test. The reported test statistic is therefore the sum of positive signed ranks, and the effect size is the rank correlation. Lowercase letters identify participants; so, e.g., **b→a** indicates training and testing on participant **a**, with pre-training on participant **b**. Before determining statistical significance, the p values reported here were Holm-Bonferroni corrected for the fourteen comparisons shown in the table.