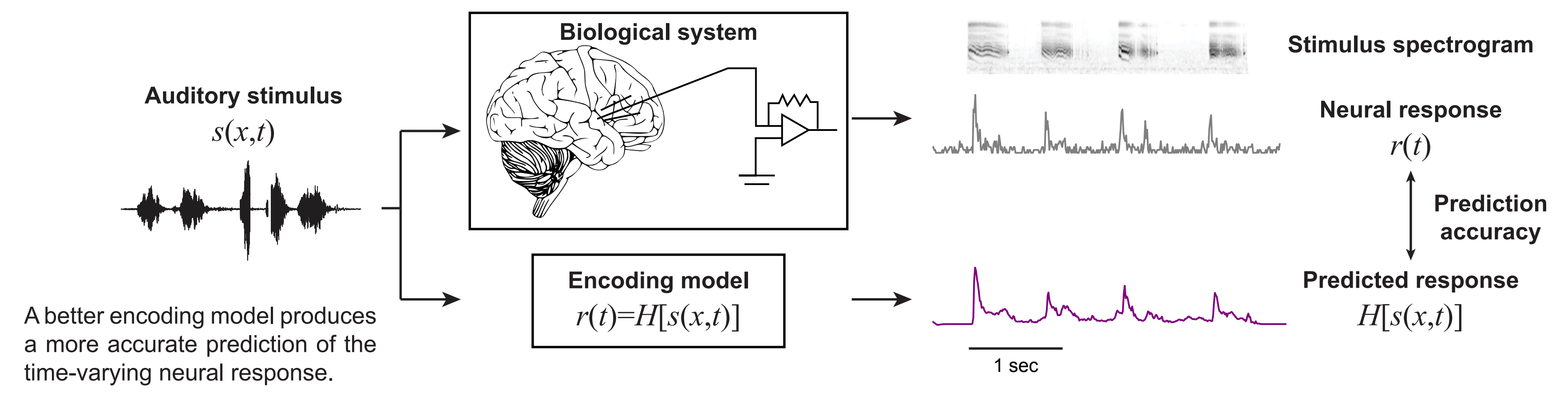# Using Deep Learning to Understand Emergent Sensory Coding in Auditory Cortex

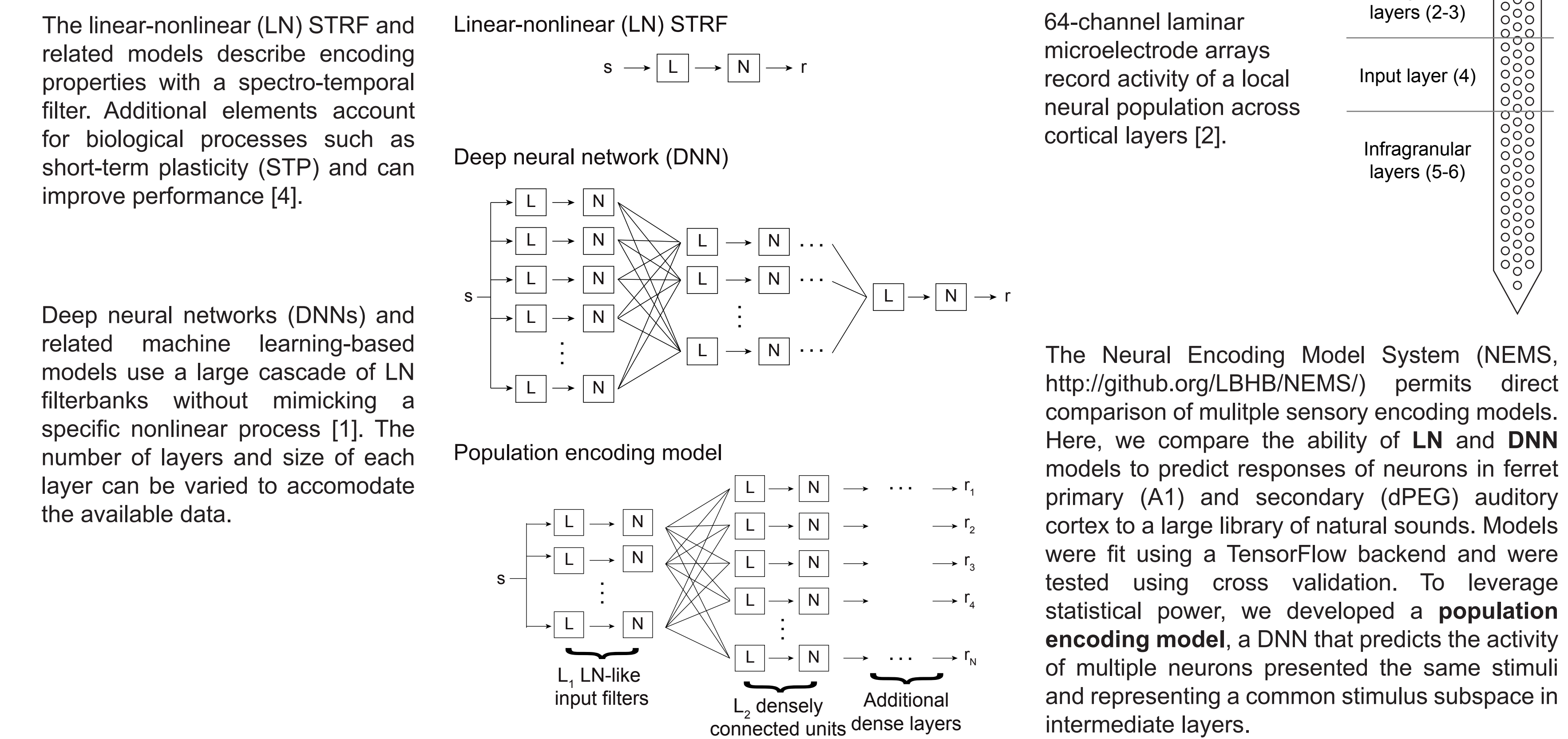Jacob R. Pennington[1], Menoua Keshishian[3], Nima Mesgarani[3], Stephen V. David[2]

[1]Washington State University, Vancouver, [2]Oregon Health & Science University, Portland, OR, [3]Columbia University, New York, NY
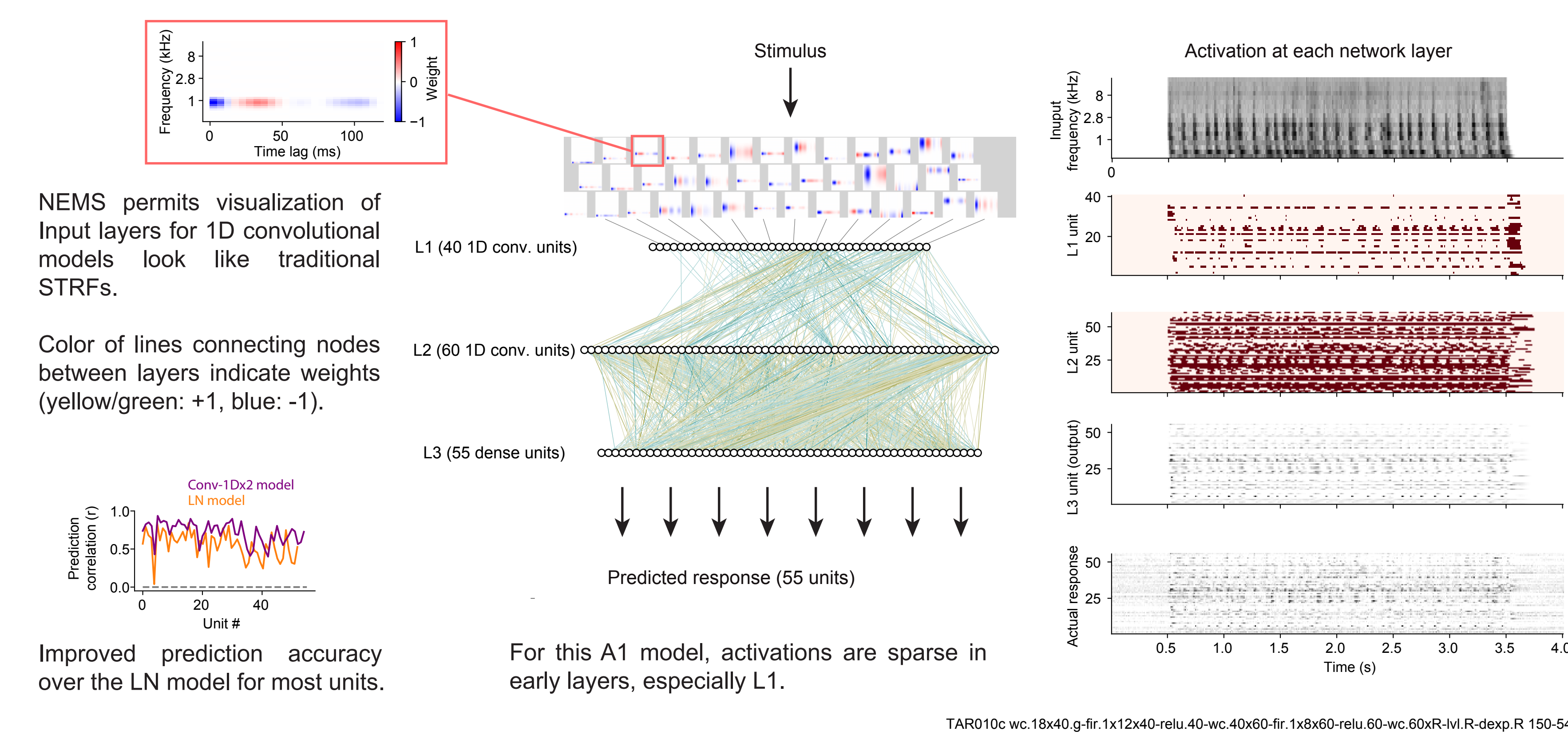
1. Encoding models describe the functional relationship between an auditory stimulus and time-varying neural response.
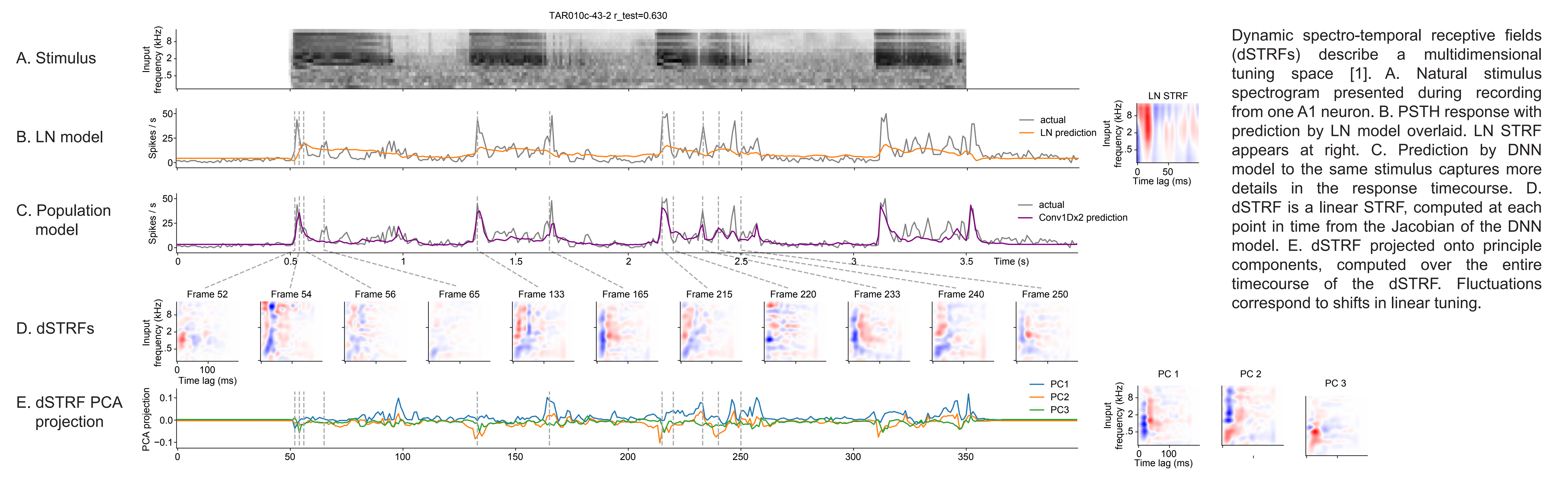

Auditory stimulus $s(x,t)$

Biological system

Stimulus spectrogram

Neural response $r(t)$

Encoding model $r(t)=H[s(x,t)]$

Prediction accuracy

Predicted response $H[s(x,t)]$

1 sec

A better encoding model produces a more accurate prediction of the time-varying neural response.

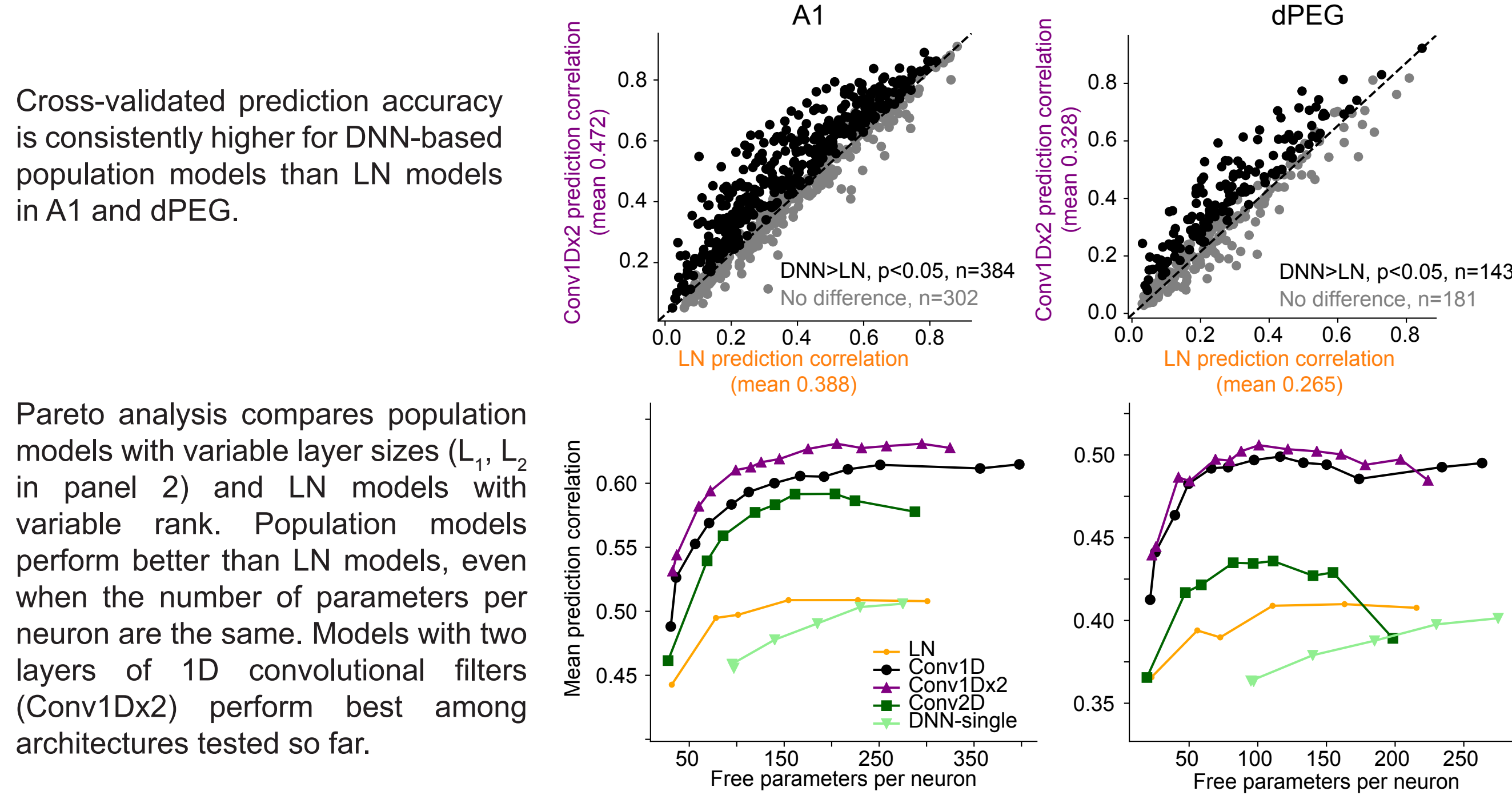## 2. Do deep neural networks outperform traditional encoding models for single-unit data in auditory cortex?
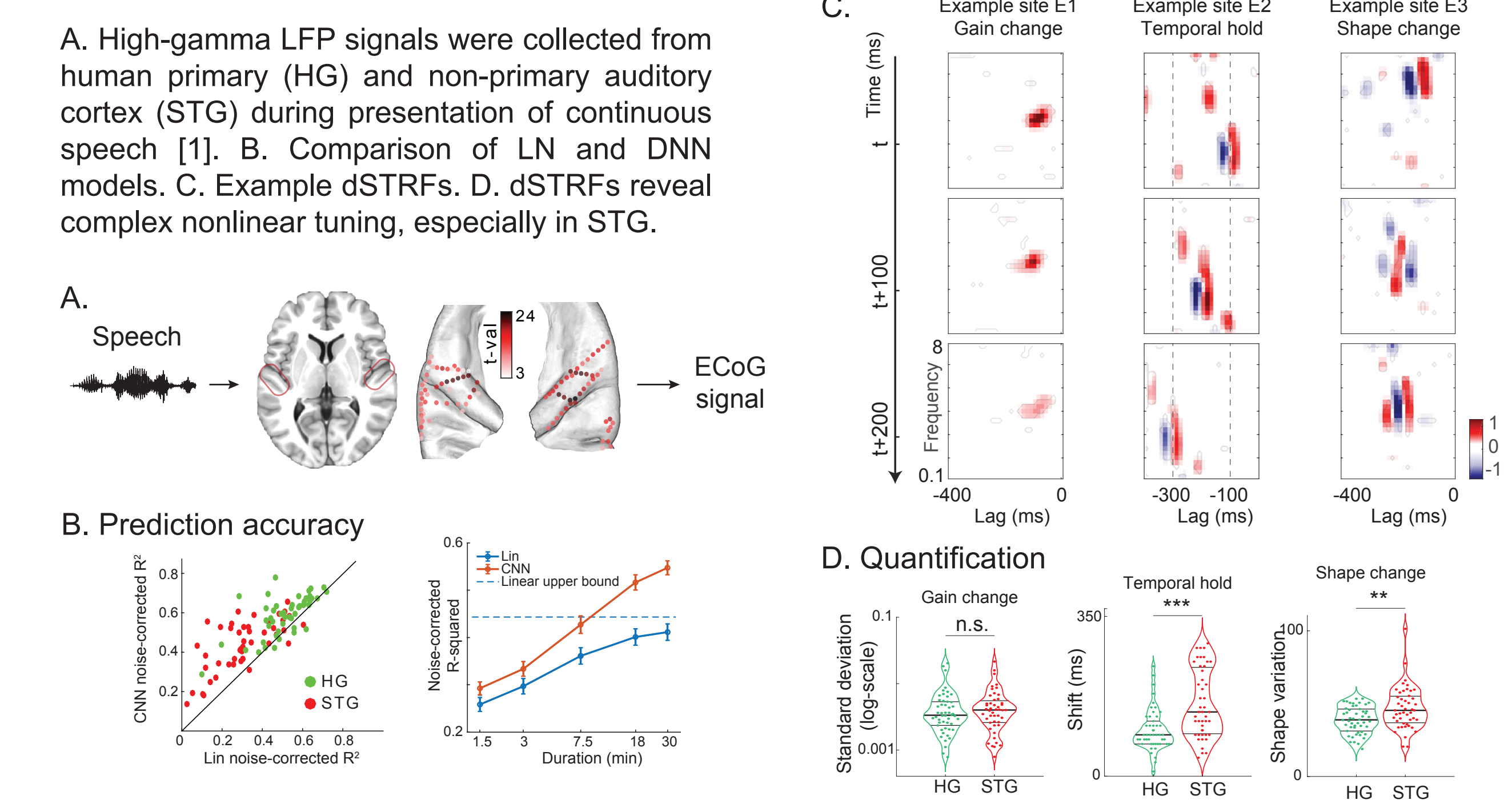
The linear-nonlinear (LN) STRF and related models describe encoding properties with a spectro-temporal filter. Additional elements account for biological processes such as short-term plasticity (STP) and can improve performance [4].

Deep neural networks (DNNs) and related machine learning-based models use a large cascade of LN filterbanks without mimicking a specific nonlinear process [1]. The number of layers and size of each layer can be varied to accomodate the available data.

Linear-nonlinear (LN) STRF

s → L → N → r

Deep neural network (DNN)

Population encoding model

L_1 LN-like input filters

L_2 densely connected units

Additional dense layers

64-channel laminar microelectrode arrays record activity of a local neural population across cortical layers [2].

Supragranular layers (2-3)

Input layer (4)

Infragranular layers (5-6)

The Neural Encoding Model System (NEMS, http://github.org/LBHB/NEMS/) permits direct comparison of mulitple sensory encoding models. Here, we compare the ability of LN and DNN models to predict responses of neurons in ferret primary (A1) and secondary (dPEG) auditory cortex to a large library of natural sounds. Models were fit using a TensorFlow backend and were tested using cross validation. To leverage statistical power, we developed a population encoding model, a DNN that predicts the activity of multiple neurons presented the same stimuli and representing a common stimulus subspace in intermediate layers.

3. A peek inside a population encoding model

NEMS permits visualization of Input layers for 1D convolutional models look like traditional STRFs.

Color of lines connecting nodes between layers indicate weights (yellow/green: +1, blue: -1).

Improved prediction accuracy over the LN model for most units.

Stimulus

Activation at each network layer

L1 (40 1D conv. units)

L2 (60 1D conv. units)

L3 (55 dense units)

Predicted response (55 units)

For this A1 model, activations are sparse in early layers, especially L1.

TAR010c wc.18x40.g-fir.1x12x40-relu.40-wc.40x60-fir.1x8x60-relu.60-wc.60xR-lvl.R-dexp.R 150-549

4. Interpreting feature selectivity of DNN-based models

A. Stimulus

B. LN model

C. Population model

D. dSTRFs

E. dSTRF PCA projection

TAR010c-43-2 r_test=0.630

Dynamic spectro-temporal receptive fields (dSTRFs) describe a multidimensional tuning space [1]. A. Natural stimulus spectrogram presented during recording from one A1 neuron. B. PSTH response with prediction by LN model overlaid. LN STRF appears at right. C. Prediction by DNN model to the same stimulus captures more details in the response timecourse. D. dSTRF is a linear STRF, computed at each point in time from the Jacobian of the DNN model. E. dSTRF projected onto principle components, computed over the entire timecourse of the dSTRF. Fluctuations correspond to shifts in linear tuning.

5. Population models consistently outperform LN models

Cross-validated prediction accuracy is consistently higher for DNN-based population models than LN models in A1 and dPEG.

Pareto analysis compares population models with variable layer sizes (L_1, L_2 in panel 2) and LN models with variable rank. Population models perform better than LN models, even when the number of parameters per neuron are the same. Models with two layers of 1D convolutional filters (Conv1Dx2) perform best among architectures tested so far.

6. DNN models for human ECoG during speech perception

A. High-gamma LFP signals were collected from human primary (HG) and non-primary auditory cortex (STG) during presentation of continuous speech [1]. B. Comparison of LN and DNN models. C. Example dSTRFs. D. dSTRFs reveal complex nonlinear tuning, especially in STG.

A. Speech → ECoG signal

B. Prediction accuracy

C. Example site E1 Gain change; Example site E2 Temporal hold; Example site E3 Shape change

D. Quantification — Gain change (n.s.); Temporal hold (***); Shape change (**)

7. Invariant coding of time-dilated speech

A. Temporal context invariance (TCI) analysis varies the length of adjacent sound segments (context) to measure the time period over which stimuli modulate neural activity [5]. B. Cross-context correlation indicates degree of TCI for nodes in a DNN. C. Example cross-context correlation over varying segment durations for an early layer node (top) and late layer node (bottom) in a DNN trained on ASR. D. Ratio of change in TCI duration for nodes in successive layers between presentation of normal speech and 1.4x dilated speech. Blue show same analysis for a randomly initialized network. Ongoing experimerments are investigating dilation invariance in DNN models fit to ECoG data.

C. Example nodes from ASR model

D. DNN node integration times adapt to dilated speech

Random DNN / Speech-trained DNN

8. Summary & conclusions

- The Neural Encoding Model System (NEMS) can characterize neural auditory coding properties using complex natural stimuli and across a range of recording modalities.
- Deep neural network (DNN) models are able to describe single-unit and ECoG responses to natural sounds consistently better than LN models.
- Dynamic STRF (dSTRF) analysis reveals a spectro-temporal tuning subspace for each neural response.
- A population encoding model that simultaneously describes activity of multiple neurons in a local circuit performs consistently better than traditional LN models and DNNs fit to single neurons.
- DNN models fit to neural data can be interrogated to study complex nonlinear properties such as invariance to temporal dilation of acoustic stimuli.

9. References

[1] Keshishian, M. et al. (2020) Estimating and interpreting nonlinear receptive fields of sensory responses with deep neural network models. *eLife* 9:e53445.

[2] Shobe, J. L. et al. (2015). Brain activity mapping at multiple scales with silicon microprobes containing 1,024 electrodes. *Journal of Neurophysiology.* 114(3), 2043-2052.

[3] Pennington, J., David, S.V. (2020) Complementary effects of adaptation and gain control on sound encoding in primary auditory cortex. *eNeuro* 7(6): ENEURO.0205-20.2020.

[4] Lopez Espejo, et al. (2019) Spectral tuning of adaptation supports coding of sensory context in auditory cortex. *PLoS Computational Biology.* 15 (10).

[5] Norman-Haignere S.V. et al. (2020) Multiscale integration organizes hierarchical computation in human auditory cortex. *bioRxiv* 2020.09.30.321687.