Signatures of the Auditory Cortex Reveal Discrepancies Across Speech Recognition Models

Gasser Elbanna[*]
Speech and Hearing Bioscience and Technology
Harvard University
Boston, MA 02115
gelbanna@mit.edu

Ivy Brundege[*]
Department of Brain and Cognitive Sciences
Virginia Tech
Blacksburg, VA 24061
ivybr@mit.edu

Josh H. McDermott
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139
jhm@mit.edu

Abstract

Speech recognition is central to human communication, yet the neural computations that support it are not fully understood. Artificial neural networks (ANNs) have shown promise as models of sensory systems, and could provide a way to generate candidate hypotheses for the neural representations and mechanisms underlying speech recognition. However, speech-specific ANNs have not been systematically evaluated for this purpose. Here we assess subword-based, word-based, and self-supervised speech models using in-silico simulations of auditory fMRI experiments that probe domain-specific response signatures in human auditory cortex. We find that models optimized for subword units (e.g., phoneme-level) best recapitulate the characteristic patterns of cortical responses, whereas word-level and self-supervised models show worse alignment. These results show how simulations of neuroimaging experiments can reveal facets of model–brain correspondence, providing a complementary diagnostic for refining both speech models and benchmarks of brain–model alignment.

1 Introduction

Understanding how the brain transforms sensory input into behavior is a central aim of neuroscience. In audition, speech arrives at the ear as a time-varying acoustic pressure waveform that conveys information about linguistic content, talker identity, and the surrounding acoustic scene. As in other domains of perception, progress in understanding the peripheral and central mechanisms that enable robust speech perception is likely to be aided by stimulus-computable models that reproduce behavioral and neural signatures of speech processing, enabling precise, testable links between theory and data. Such models could arguably be particularly useful for understanding speech, as neuroscience-based approaches are limited by the coarseness of human neuroscience methods.

Recent advances in artificial neural networks (ANNs) have enabled stimulus-computable, sensory-grounded working models that articulate testable hypotheses about function and organization (1). In audition, such models capture human behavior across tasks—word recognition (2; 3), pitch estimation

[*] Both authors contributed equally to the paper.

Preprint.

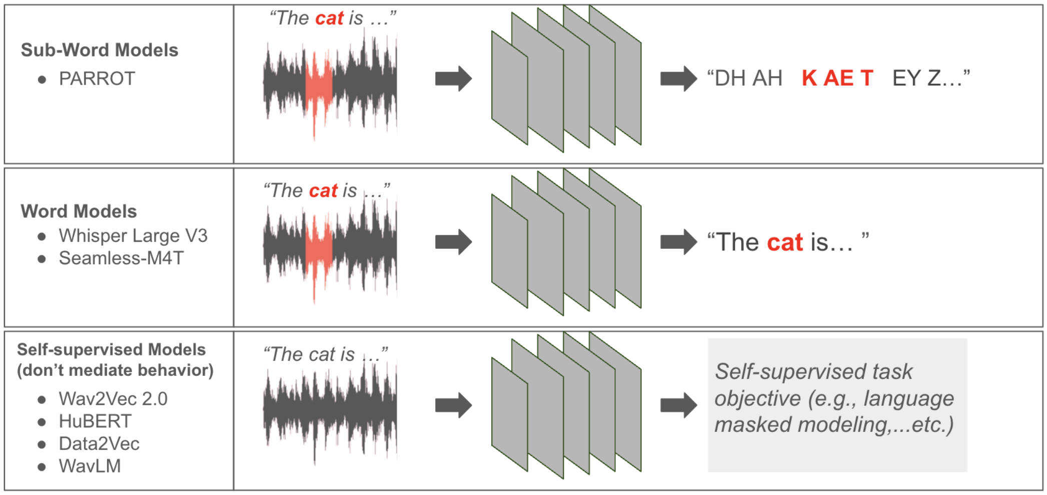


Figure 1: Candidate models spanning three types of speech modeling. Sub-word models are optimized to map acoustic signals to a stream of sub-word units (e.g., phonemes). Word models are optimized to map acoustic signals to a stream of words. Self-supervised models are trained on task objectives that don't require an explicit supervisory signal (e.g., contrastive, language masked modeling, etc.).

(4), sound localization (5), and auditory attention (6)—and explain variance in neural responses throughout the auditory pathway (2; 7; 8; 9).

The most widely used current approaches to assessing human–model alignment fall into two families. Fitting-based methods fit model responses to brain or behavioral measurements (e.g., linear encoding) (10; 11), whereas fitting-free methods compare representational geometry of model and brain responses (e.g., RSA; CKA) (12; 13). While these tools have enabled systematic comparisons, they often reveal similar degrees of human-model similarity across large sets of models (9; 14; 15; 16), motivating the use of additional comparison methods. Here we propose to simulate human fMRI experiments on ANN models, averaging responses within "regions of interest" and comparing responses across conditions as is common in fMRI analysis.

We evaluated three families of speech ANNs—in house subword-optimized models, off-the-shelf word-level models, and off-the-shelf self-supervised models—on their ability to reproduce established fMRI-based signatures of human auditory cortex. We find that these signatures reliably differentiate model classes and reveal clear divergences between some models and brain responses.

2 Methods

2.1 Candidate speech models

Figure 1 shows the three classes of speech models used in this study. These models are optimized to learn speech representations using different task objectives, architectures, and training diets.

2.1.1 Sub-word models

We trained an in-house sub-word model called PARROT to map acoustic signals to phonemes. PARROT combines a simulation of the human ear with convolutional and recurrent neural network modules. The model was trained on $15,000$ hours of speech data superimposed on naturalistic noisy backgrounds with varying SNRs and sound levels. The experiments described here used representations generated by the last convolution layer and representations from all recurrent layers (6 layers). Model architecture is shown in Supplementary Figure 1.

2.1.2 Word models

We used off-the-shelf state-of-the-art models that are trained to map acoustic signals to words, such as Whisper (17) and SeamlessM4T (18). For each model, we extracted encoder representations from

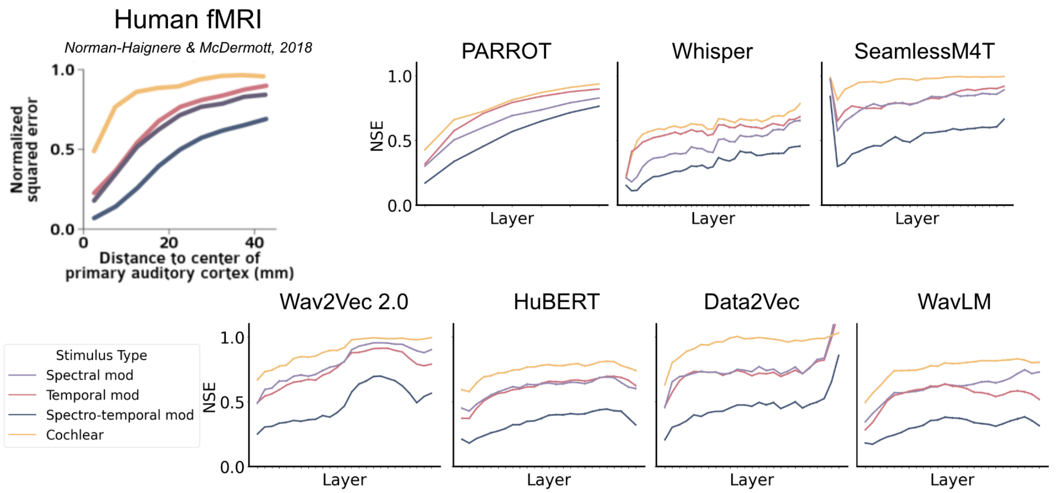


Figure 2: Human and model responses to model-matched stimuli (matched to responses to natural sounds in one of four acoustic models)

the input embedding layer and from every subsequent encoder block. Whisper's encoder comprises 32 Transformer blocks, and SeamlessM4T's speech encoder comprises 24 Conformer blocks.

2.1.3 Self-supervised models

We used Wav2Vec 2.0 (19), HuBERT (20), Data2Vec (21), and WavLM (22). We evaluated the large version of these models. The experiments used representations generated by their embedding layer and 24 Transformer encoder layers.

2.2 Calculation of model responses

We considered responses of both individual units and entire layers. For a given set of stimuli, a unit response score was defined as the time-averaged absolute values of that unit's response $\text{score}_{\text{unit}} = \frac{1}{T_s}\sum_{t=1}^{T_s}|r_t|$. For a layer's response, the response was further averaged over all units within a layer.

2.3 Model-matched stimuli experiment

This experiment replicated the analysis and stimuli from a previous human fMRI study (23).

2.3.1 Stimulus generation

For a given natural stimulus, a companion stimulus was synthesized to elicit the same response in a hand-crafted acoustic model intended to capture aspects of biological auditory processing (24). Four acoustic models were used to generate these *"model-matched stimuli"*, varying in the extent and type of acoustic features that were matched across the synthetic and natural sound (cochlear, temporal modulation, spectral modulation, and spectro-temporal modulation). See Supplementary Figure 2 for further illustration.

The cochlear model used a series of bandpass filters designed to mimic the response of the human cochlea (25). The three remaining models pass the filters responses through a series of wavelet filters, convolved with time, frequency, or both, to create models tuned to spectral, temporal, or spectro-temporal modulation, respectively.

2.3.2 Response analysis

We calculated the degree of divergence between ANN model responses to original and model-matched stimuli on a unit level, quantified as the normalized squared error (NSE) between natural and its corresponding model-matched response scores:

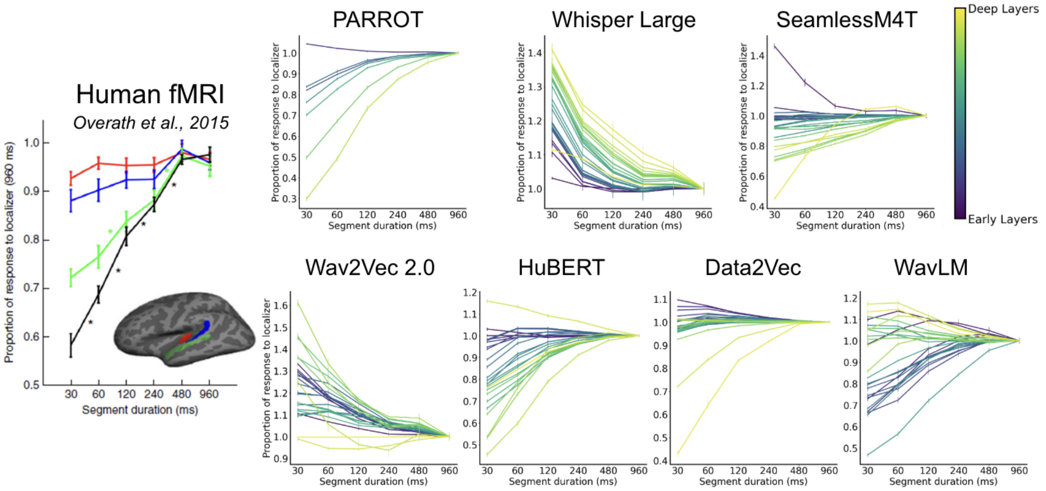


Figure 3: Human and model responses to speech quilts. Speech quilts scramble acoustic structure to different extents depending on the segment length.

$$NSE = \frac{\mu([x-y]^2)}{\mu(x^2) + \mu(y^2) - 2\mu(x)\mu(y)}$$

For each layer, we computed NSE for each unit and report layer-wise responses across models.

2.4 Speech quilt experiment

2.4.1 Stimulus generation

We used the stimuli first described in (26). Speech quilts are generated by shuffling uniform-length segments of a stimulus, with segments ordered and cross-faded to ensure smooth transitions between segments. Quilts were generated with segment lengths ranging from 30ms-960ms, such that shorter-duration segments lead to greater disruption of the temporal structure of the stimulus. See Supplementary Figure 3 for further illustration.

2.4.2 Response analysis

Following (26), we averaged unit responses within a layer, then normalized responses by the average response to the 960ms condition for that layer. We report the average normalized score, by layer, over all segment durations.

3 Results

3.1 Model-matched stimuli

Figure 2 shows the normalized squared error (NSE) between the response to natural sounds and their corresponding synthesized stimuli produced by four acoustic models, plotted separately for voxels in the human brain and units in candidate models. The NSE is calculated for different brain regions (from primary to non-primary) and different model layers (from early to deep). All models show general trends for a) lower NSE values when spectral and temporal modulation were matched compared to when only a cochlear model was matched, and b) for the NSE to be higher in deeper stages, consistent with the presence of higher-order features that respond more to the natural sounds than to the model-matched sounds. However, the PARROT model most closely resembles what was previously observed in human auditory cortex. See Supplementary Figure 4 for voxel-wise/unit-wise comparison between auditory cortex responses and PARROT model.

3.2 Speech quilting

Figure 3 shows brain and models responses to speech quilts in German varying in segment duration (from 30 ms to 960 ms) across brain regions and model layers, respectively. Many of the models are highly discrepant with the responses observed in humans, but PARROT and Data2Vec qualitatively replicate the tendency of non-primary (deeper) auditory stages to respond more to quilts with longer segments lengths (that have more global speech structure).

4 Discussion & Conclusion

We compared three classes of speech models—subword-based, word-based, and self-supervised—using simulations of two auditory fMRI experiments previously conducted in humans. Subword models (e.g., PARROT) recapitulated the key response patterns observed in human auditory cortex, whereas word-level and self-supervised models showed less similarity to human brain responses. The results show how fMRI simulation tests that probe for domain-specific response signatures of the human brain provide a complementary additional diagnostic for candidate models.

References

[1] B. A. Richards, T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. de Berker, S. Ganguli *et al.*, "A deep learning framework for neuroscience," *Nature neuroscience*, vol. 22, no. 11, pp. 1761–1770, 2019.

[2] A. J. Kell, D. L. Yamins, E. N. Shook, S. V. Norman-Haignere, and J. H. McDermott, "A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy," *Neuron*, vol. 98, no. 3, pp. 630–644, 2018.

[3] M. R. Saddler and J. H. McDermott, "Models optimized for real-world tasks reveal the task-dependent necessity of precise temporal coding in hearing," *bioRxiv preprint bioRxiv:2024.04.21.590435v2*, 2024.

[4] M. R. Saddler, R. Gonzalez, and J. H. McDermott, "Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception," *Nature Communications*, vol. 12, p. 7278, 2021.

[5] A. Francl and J. H. McDermott, "Deep neural network models of sound localization reveal how perception is adapted to real-world environments," *Nature Human Behaviour*, vol. 6, pp. 111–133, 2022.

[6] I. M. Griffith, R. P. Hess, and J. H. McDermott, "Optimized feature gains explain and predict successes and failures of human selective listening," *bioRxiv*, pp. 2025–05, 2025.

[7] Y. Li, G. K. Anumanchipalli, A. Mohamed, P. Chen, L. H. Carney, J. Lu, J. Wu, and E. F. Chang, "Dissecting neural computations in the human auditory pathway using deep neural networks for speech," *Nature Neuroscience*, vol. 26, no. 12, pp. 2213–2225, 2023.

[8] A. R. Vaidya, S. Jain, and A. G. Huth, "Self-supervised models of audio effectively explain human cortical responses to speech," *arXiv preprint arXiv:2205.14252*, 2022.

[9] G. Tuckute, J. Feather, D. Boebinger, and J. H. McDermott, "Many but not all deep neural network audio models capture brain responses and exhibit correspondence between model stages and brain regions," *Plos Biology*, vol. 21, no. 12, p. e3002366, 2023.

[10] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant, "Encoding and decoding in fmri," *Neuroimage*, vol. 56, no. 2, pp. 400–410, 2011.

[11] A. G. Huth, W. A. De Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant, "Natural speech reveals the semantic maps that tile human cerebral cortex," *Nature*, vol. 532, no. 7600, pp. 453–458, 2016.

[12] N. Kriegeskorte, M. Mur, and P. A. Bandettini, "Representational similarity analysis-connecting the branches of systems neuroscience," *Frontiers in systems neuroscience*, vol. 2, p. 249, 2008.

[13] S. Kornblith, M. Norouzi, H. Lee, and G. Hinton, "Similarity of neural network representations revisited," in *International conference on machine learning*. PMlR, 2019, pp. 3519–3529.

[14] C. Conwell, J. S. Prince, K. N. Kay, G. A. Alvarez, and T. Konkle, "A large-scale examination of inductive biases shaping high-level visual representation in brains and machines," *Nature communications*, vol. 15, no. 1, p. 9383, 2024.

[15] M. Schrimpf, I. A. Blank, G. Tuckute, C. Kauf, E. A. Hosseini, N. Kanwisher, J. B. Tenenbaum, and E. Fedorenko, "The neural architecture of language: Integrative modeling converges on predictive processing," *Proceedings of the National Academy of Sciences*, vol. 118, no. 45, p. e2105646118, 2021.

[16] Z. Zhang, "Improved adam optimizer for deep neural networks," in *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*. Ieee, 2018, pp. 1–2.

[17] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *International conference on machine learning*. PMLR, 2023, pp. 28 492–28 518.

[18] L. Barrault, Y.-A. Chung, M. C. Meglioli, D. Dale, N. Dong, P.-A. Duquenne, H. Elsahar, H. Gong, K. Heffernan, J. Hoffman *et al.*, “Seamlessm4t: massively multilingual & multimodal machine translation,” *arXiv preprint arXiv:2308.11596*, 2023.

[19] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, “wav2vec 2.0: A framework for self-supervised learning of speech representations,” *Advances in neural information processing systems*, vol. 33, pp. 12 449–12 460, 2020.

[20] W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed, “Hubert: Self-supervised speech representation learning by masked prediction of hidden units,” *IEEE/ACM transactions on audio, speech, and language processing*, vol. 29, pp. 3451–3460, 2021.

[21] A. Baevski, W.-N. Hsu, Q. Xu, A. Babu, J. Gu, and M. Auli, “Data2vec: A general framework for self-supervised learning in speech, vision and language,” in *International conference on machine learning*. PMLR, 2022, pp. 1298–1312.

[22] S. Chen, C. Wang, Z. Chen, Y. Wu, S. Liu, Z. Chen, J. Li, N. Kanda, T. Yoshioka, X. Xiao *et al.*, “Wavlm: Large-scale self-supervised pre-training for full stack speech processing,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1505–1518, 2022.

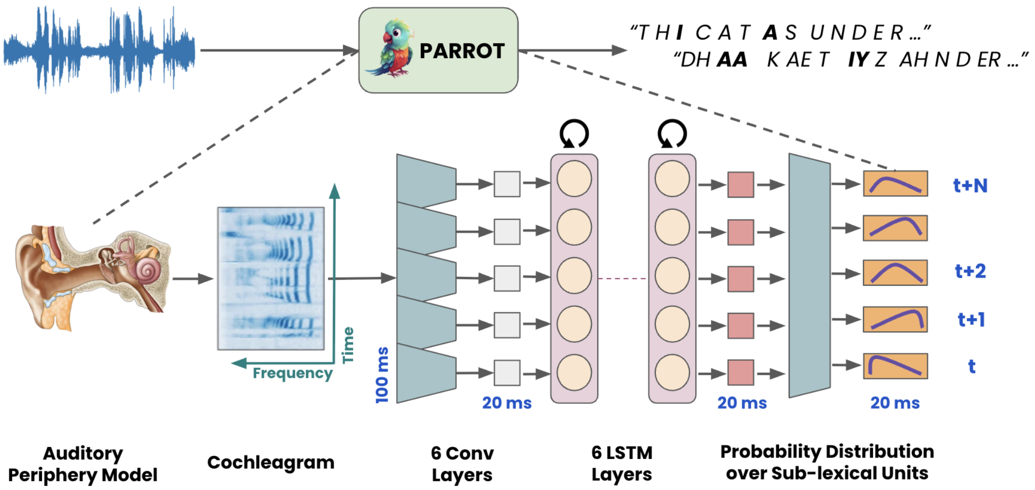
[23] S. V. Norman-Haignere and J. H. McDermott, “Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex,” *PLoS biology*, vol. 16, no. 12, p. e2005127, 2018.

[24] T. Chi, P. Ru, and S. A. Shamma, “Multiresolution spectrotemporal analysis of complex sounds,” *The Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 887–906, 2005.

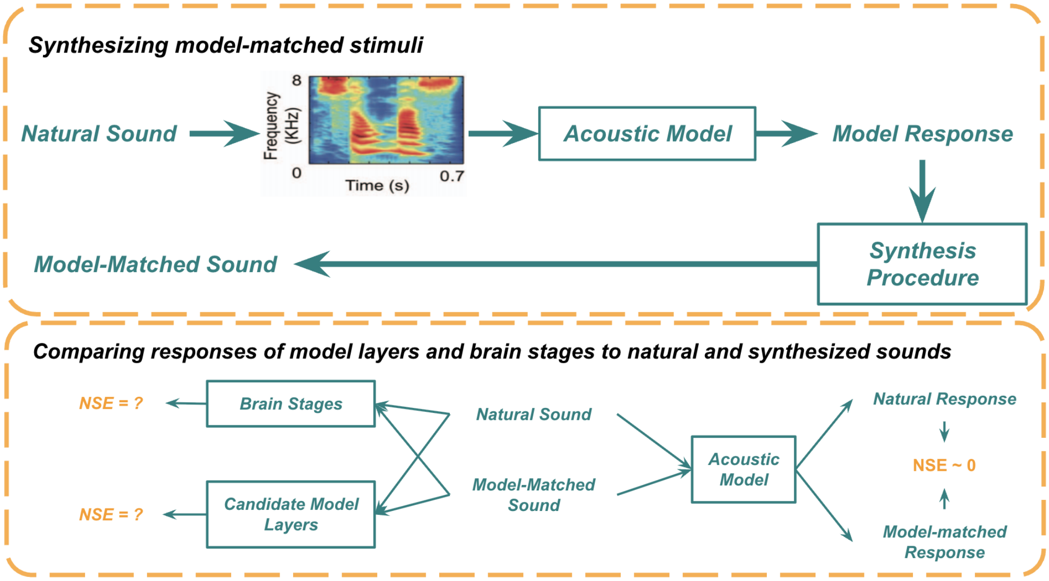
[25] B. R. Glasberg and B. C. Moore, “Derivation of auditory filter shapes from notched-noise data,” *Hearing research*, vol. 47, no. 1-2, pp. 103–138, 1990.

[26] T. Overath, J. H. McDermott, J. M. Zarate, and D. Poeppel, “The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts,” *Nature neuroscience*, vol. 18, no. 6, pp. 903–911, 2015.

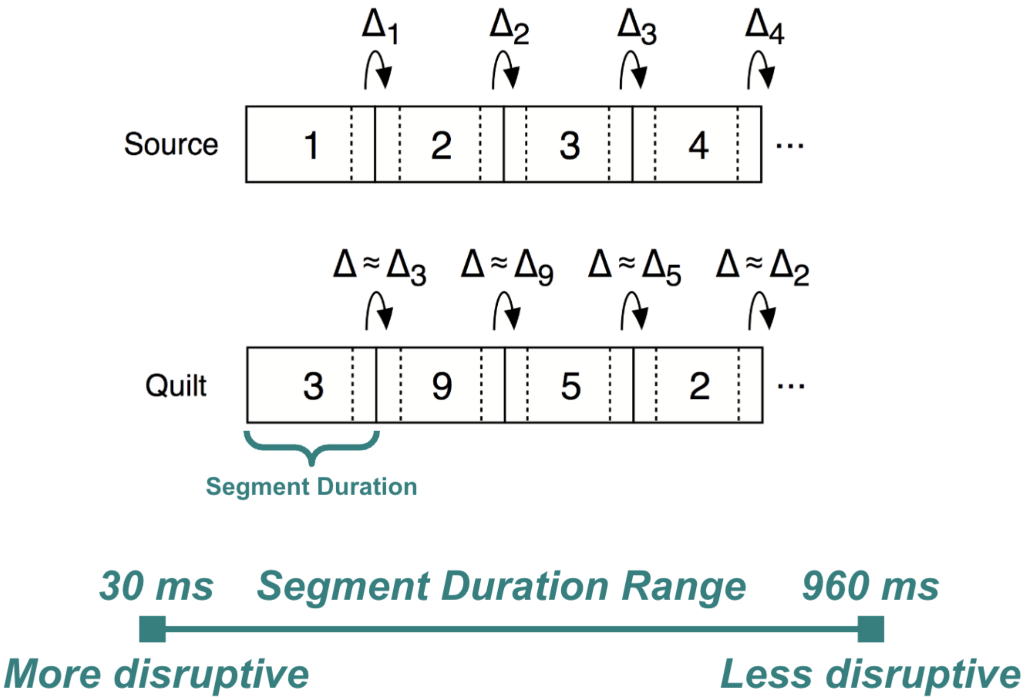
A Supplementary Figures



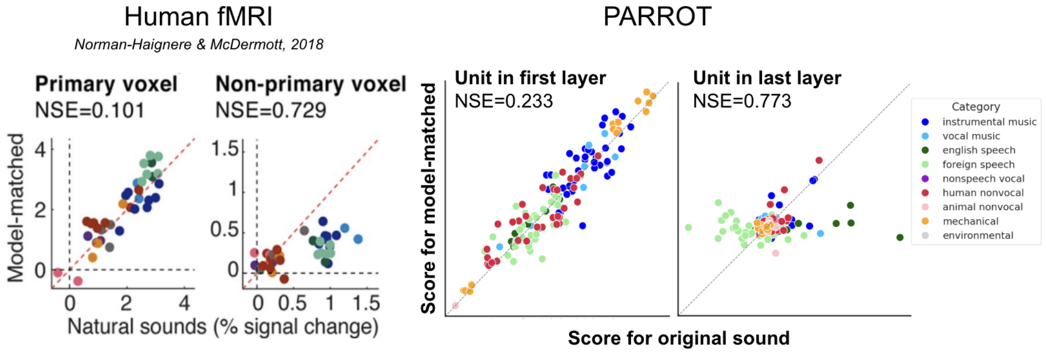
Supplementary Figure 1: Sub-word model (PARROT) architecture.



Supplementary Figure 2: Top diagram shows the pipeline for synthesizing sounds that would elicit the same response in a specific acoustic model as the original natural sound. Bottom diagram shows the process for evaluating responses of speech models and brain stages to both natural and their corresponding synthesized sounds. Four acoustic models were used. Thus, for each natural sound, 4 different model-matched sounds were synthesized.



Supplementary Figure 3: Process of generating speech quilts. Figure adapted from (26).



Supplementary Figure 4: voxels and units responses to original and model matched stimuli in the auditory cortex and PARROT model, respectively.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.

- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS paper checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims about using fMRI simulation experiments as a good metric for brain-model alignment in the abstract are clearly and accurately described in the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No]

   Justification: Limitations are not included since this is still on-going work.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.

- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: No theoretical results are provided in this paper, only experimental.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Details about simulating the fMRI experiments in speech models is clearly stated in the methods.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: No data or code is shared since this is an ongoing-work.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: Analyses were done on pretrained models so no training was done in this study.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The error bars reflect standard error for all analyses.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The analyses in this study weren't computationally demanding, only used a single A100 GPU.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We followed the Code of Ethics provided by NeurIPS.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.

- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: This work doesn't exhibit any potential risk or societal impact.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Our proposed model doesn't show a high risk of misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We cite open-source systems aand prior work used in our experiments.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: No new assets are released in this submission.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: No human data was collected in this study.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: No human data was collected in this study.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.