Towards individualized models of hearing-impaired speech perception

Mark R. Saddler[1], *Torsten Dau*[1], *Josh H. McDermott*[2]

[1]DTU Department of Health Technology, Denmark
[2]MIT Department of Brain and Cognitive Sciences, USA
marksa@dtu.dk, tdau@dtu.dk, jhm@mit.edu

Abstract

Computational models that predict the real-world hearing abilities of individuals with hearing loss have the potential to transform hearing aid development. Deep artificial neural networks trained to perform ecological hearing tasks using simulated cochlear input reproduce many aspects of normal hearing, but it is not clear whether such models can also account for impaired hearing. We used the Clarity Prediction Challenge dataset to test if a model jointly optimized for everyday sound localization and recognition tasks can predict the speech intelligibility of hearing-impaired listeners. We used the model's learned feature representations as an intrusive speech intelligibility metric and measured the effects of simulating individual listeners' hearing losses in the model's peripheral input. Individualizing the hearing loss simulations allowed our model to better predict speech intelligibility differences across listeners. However, this benefit was small in the overall human-model correlation, likely because the explainable variance in the dataset is driven more by the different hearing aids than by the different listeners.

Index Terms: auditory model, hearing loss, individual differences, speech intelligibility, perceptual metrics

1. Introduction

Hearing loss is a widespread and growing public health issue, with 1 in every 10 people projected to have disabling hearing loss by 2050 [1]. Hearing aids are the main treatment available, but current devices fail to restore normal hearing, especially in noisy environments. One factor limiting the development of more effective devices is our incomplete understanding of how the peripheral consequences of hearing loss translate to everyday hearing difficulties [2]. Computational models that directly relate peripheral auditory processing to real-world perception could deepen this understanding.

Recent progress towards this aim has been made with deep learning. Deep artificial neural networks trained to perform ecological hearing tasks using simulated cochlear representations as input have been shown to account for many aspects of human auditory behavior [3, 4, 5, 6, 7]. However, these recent models have only been compared against normal hearing listeners and cochlear implant users [8, 9] at the group level, and it remains unclear whether such models can explain the diverse hearing abilities of individuals with hearing loss. Here, we used the Clarity Prediction Challenge (CPC) dataset [10] to test whether a model optimized for everyday hearing tasks can predict the speech intelligibility of hearing-impaired listeners.

The CPC dataset is a large collection of speech-in-noise signals processed by different hearing aid systems and presented to listeners with hearing loss. Each signal contains a short sentence and listeners were asked to repeat the words they heard. The dataset includes listener response transcripts, ground truth transcripts, clean reference speech signals, and listener metadata (hearing loss severity designation, age, sex, pure tone audiogram, digit-triplet test threshold, and subjective questionnaire responses). We used the dataset to investigate the effect of individualized hearing loss simulation on our model's speech intelligibility predictions. In particular, we asked whether simulating listeners' audiograms in our model's periphery leads to more accurate predictions than simulating only listeners' coarse severity designations or simulating no hearing loss.

The results from previous challenges suggest this hypothesis is not trivial. None of the top-ranked systems from CPC1 or CPC2 incorporated an explicit peripheral auditory model, and many of the best-performing systems made little or no use of the listeners' audiograms [11, 12]. The limited usefulness of audiogram information for speech intelligibility prediction is consistent with the possibility that hearing aid amplification compensates for audibility differences across listeners, leaving only the less well understood suprathreshold effects [13, 14, 15]. However, the effects of simulating individual audiograms have not been tested in a model that performs real-world auditory tasks using cochlear input.

Here, we first trained a model to localize and recognize speech and other natural sounds using simulated auditory nerve representations as input. We then simulated hearing loss in the model's peripheral input and asked if it could predict the speech intelligibility of hearing-impaired listeners. We evaluated the model on the CPC3 dataset by leveraging its learned feature representations as an intrusive speech intelligibility metric [16, 17]. To assess the benefit of modeling listeners' unique hearing impairments, we compared model variants with different degrees of individualization in the peripheral hearing loss simulations. The results provide new evidence in favor of individualizing computational models of hearing loss and highlight different ways to evaluate such models on the CPC3 dataset.

2. Methods

2.1. Model architecture

2.1.1. Auditory nerve input representation

All sounds were resampled to 50 kHz and passed through an auditory nerve model to simulate the spiking responses of 32000 nerve fibers per ear. The model consisted of an audibility filter (shaped like the ISO 226 equal-loudness-level contour at 0 dB), a gammatone filterbank, half-wave rectification, a low-pass filter, and sigmoid rate-level functions to yield instantaneous spike rates. Arrays of spike counts sampled from these rates served as input to an artificial neural network (Fig. 1). The arrays had shape [6, 50, 20000], representing 3 canonical auditory nerve fiber types (with high, medium, and low spontaneous rates [18]) per ear, 50 characteristic frequencies spaced uniformly on an ERB-number scale [19] from 60 to 16000 Hz, and 20000 timesteps sampled at 10 kHz.

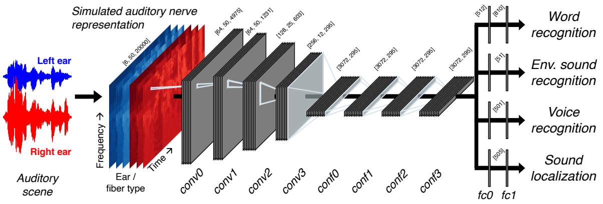

Figure 1: *Model optimized for ecological auditory tasks*

2.1.2. *Task-optimized neural network*

We used a branching feedforward neural network architecture to support optimization for multiple tasks. The network's trunk consists of 4 convolutional layers followed by 4 conformer [20] layers. Each convolutional layer comprises a series of operations: linear 2D convolution, sigmoid linear unit activation function, average pooling, and layer normalization. Outputs of the final convolution layer were reshaped from [256 channels, 12 frequency bins, 295 timesteps] to [256 × 12 = 3072 channels, 295 timesteps] and fed to the first conformer layer. Each conformer layer has 4 attention heads, a hidden dimension of 256, and a kernel size of 31. The 3072 channels of the final conformer layer's output were split into 4 parallel branches, each consisting of a 512-unit fully-connected layer followed by a task-dependent output layer. Network weights before the branch point are shared between the model's tasks, and all weights after the branch point are task-specific.

2.2. Model optimization

The neural network's 747 million parameters were jointly optimized to perform four auditory classification tasks. The training dataset consists of 7.6 million 2-second binaural auditory scenes spatialized with a virtual acoustic head and room simulator modeling KEMAR's HRTFs [21] in 2000 acoustically distinct rooms. Each scene comprises a speech or natural sound target rendered at a single location with texture-like background noise rendered diffusely at multiple locations. Speech targets were sourced from CommonVoice [22], non-speech targets from GISE-51 [23], and background noises from AudioSet [24].

The model's tasks were to localize the target (operationalized as a 504-way classification task) and make three types of recognition judgments (809-way word recognition and 500-way voice recognition tasks for speech targets; 50-way environmental sound classification for non-speech targets). The model was optimized via stochastic gradient descent to minimize the summed softmax cross entropy losses from the four classification tasks. When a task was undefined for a training example (e.g., word recognition for a non-speech stimulus), the task was excluded from the loss. The model trained for 225 hours on 8 NVIDIA A100 GPUs (250,000 steps with a batch size of 256).

2.3. Hearing loss simulation

The model was optimized using normal hearing auditory nerve input. To model impaired hearing, we froze the trained network's weights and altered only the auditory nerve input representations to fit a given audiogram (Fig. 2).

The auditory nerve model's rate-level functions were parametrized with a compression power $c \in [0.3, 1.0]$, which set their absolute thresholds and dynamic ranges (Fig. 2A). Normal hearing was modeled with $c = 0.3$ for all frequency channels, yielding uniform thresholds near 0 dB HL. Audiometric hearing losses were modeled by setting c independently for each frequency channel to produce elevated thresholds and reduced dynamic ranges. Increasing c reduces the cochlea's nonlinear amplification, simulating one consequence of outer hair cell loss. To couple this loss of amplification with broader cochlear frequency tuning, the bandwidths of the gammatone filters were scaled linearly by a factor of 1 to 3 as c increased from 0.3 to 1.0.

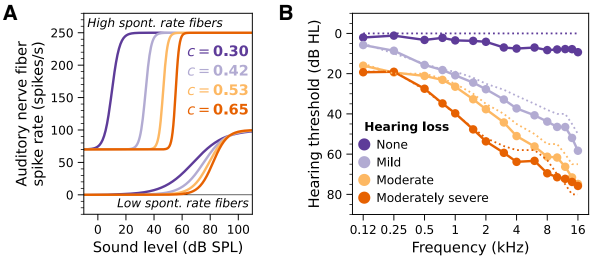

Figure 2: *Rate-level functions and model audiograms*

To validate our audiogram fitting procedure, we measured our model's audiometric thresholds with linear classifiers trained to discriminate pure tones from silence using the network's internal representations. Model-measured audiograms (solid lines in Fig. 2B) reasonably matched the target audiograms (dotted lines).

To measure the effect of simulating individual listeners' hearing losses, we fit the audiogram of each of the 26 listeners in the CPC3 dataset, creating a set of models collectively referred to as the "audiogram-matched" model. This model was compared against a "severity-matched" model (with each listener assigned one of the reference audiograms in Fig. 2B according to their severity designation) and a "normal hearing only" model (with all listeners assigned the normal hearing audiogram).

2.4. Intrusive speech intelligibility metric

We developed a correlation-based intrusive speech intelligibility metric using our network's learned deep feature representations [16, 17]. Activations from hearing-impaired models in response to hearing aid output signals were linearly correlated with the corresponding activations from a normal hearing model in response to the clean speech reference signals. Because our model operates on fixed-length inputs, hearing aid outputs and reference signals were subdivided into 2-second 90% overlapping frames and correlations were averaged across all frames.

2.5. Clarity Prediction Challenge compliance

As we were primarily interested in modeling individual listeners' hearing losses, our models do not all adhere to the CPC3 rules against using listener data from prior challenges. The "audiogram-matched" models incorporate listener audiograms released with the CPC1 dataset. For our contest submission, we used only the challenge-compliant "severity-matched" model. To maximize the performance of our submitted system, intelligibility predictions on the CPC3 evaluation dataset were passed through a logistic function whose parameters minimized the root mean squared error between predicted and measured intelligibility scores on the CPC3 training dataset.

This fitting step was not included in the analyses reported here, which treated the entire CPC3 training dataset as an evaluation dataset. Here, all model parameters were optimized solely for performance on the ecological training tasks. Model predictions were not fit in any way to the CPC3 human intelligibility scores. Any similarity to human behavior is thus a consequence of optimizing the model for its training tasks given the constraints of the simulated auditory nerve input and the neural network architecture [6].

3. Results

3.1. Layer-wise speech intelligibility predictions

To determine which model stage best predicts human speech intelligibility, we evaluated our intrusive metric on the CPC3 training dataset, using activations from each of the model layers one at a time (Fig. 3). We measured the linear correlation between model intelligibility predictions and human intelligibility scores across all scenes, listeners, and hearing aid systems. The correlations generally increased deeper into the network, with the highest correlations obtained in the model's word recognition branch. This result indicates the network features most optimized for word recognition are also the most predictive of human speech recognition. Subsequent analyses used the word recognition branch's fc0 activations (red circle, Fig. 3).

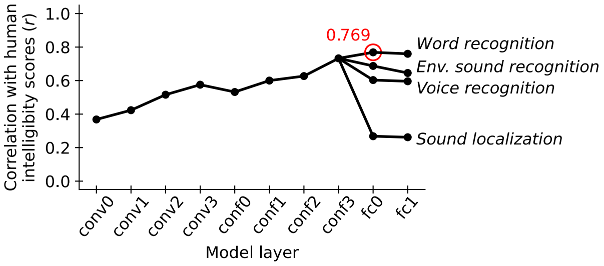

Figure 3: *Speech intelligibility correlations of each model stage*

3.2. Overall effect of individualized hearing loss simulation

To test whether simulating individual listeners' hearing losses improves speech intelligibility predictions, we compared the overall human-model correlations from our "audiogram-matched", "severity-matched", and "normal hearing only" models (Table 1). The correlations were all similar, showing little benefit of hearing loss simulation when evaluated across the entire CPC3 training dataset. Our three models outperformed corresponding Hearing-Aid Speech Perception Index (HASPIv2) baselines [25], which used the same audiograms and were computed from the better ear in each scene. The HASPIv2 baselines showed no benefit of hearing loss simulation.

Table 1: *Overall correlations between human and model speech intelligibility scores with different hearing loss simulations*

System	Audiogram-matched	Severity-matched	Normal hearing only
Proposed model	0.769	0.768	0.758
HASPIv2 baseline	0.671	0.684	0.690

3.3. Analysis of hearing aid vs. listener-driven variance

Correlations between human and model speech intelligibility scores across all 15520 scenes in the CPC3 training dataset quantify how well a model can jointly explain all sources of variance: the different scenes, the different hearing aid systems, and the different listeners' hearing losses. To marginalize out the variance due to scenes and compute the split-half reliability of the human speech intelligibility scores, we grouped the scenes by hearing aid system and listener, and averaged intelligibility scores across scenes within the same group. Human-model correlations across these scene-averaged scores quantify how well a model explains the combined variance due to hearing aids and listeners (Fig. 4A). To measure a model's ability to explain only the hearing aid-driven variance, we separately correlated the scene-averaged human and model scores for each of the 26 listeners and then averaged the correlations (Fig. 4B). To measure a model's ability to explain only the listener-driven variance, we separately correlated the scene-averaged human and model scores for each of the 18 hearing aid systems and then averaged the correlations (Fig. 4C).

When analyzed this way, we observe a significant benefit to individualizing the hearing loss simulation in our model ($F(1.323, 22.494) = 12.444$, $p = 0.001$, $\eta^2_{partial} = 0.069$, repeated-measures ANOVA with Greenhouse-Geisser correction). The "audiogram-matched" model explains more of the explainable listener-driven variance (44.1%) than either the "severity-matched" model (35.0%) or the "normal hearing only" model (20.0%). The effect of individualizing the hearing loss simulation in the HASPIv2 baseline system was not statistically significant ($F(2, 34) = 0.348$, $p = 0.709$, $\eta^2_{partial} = 0.007$).

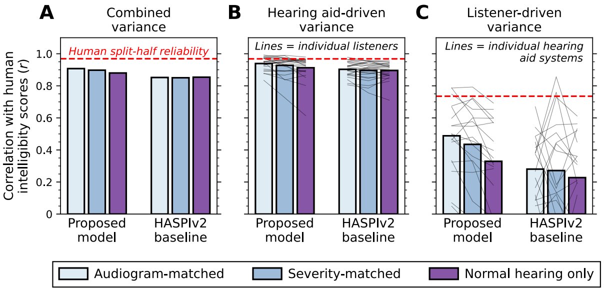

Figure 4: *Correlations between scene-averaged human and model speech intelligibility scores, grouped by listener and/or hearing aid system*

3.4. Non-intrusive word recognition predictions

So far, we have only considered the model's predictions when using its internal representations as an intrusive speech intelligibility metric. Because our model was explicitly trained to recognize words in noise, it is also possible to test whether the model can predict listeners' speech recognition performance using only the noisy, hearing aid-processed CPC3 signals (without using the clean speech reference signals, which listeners did not have access to). However, since the model was optimized for a closed-set word recognition task with an 809 word vocabulary, evaluating it on the full CPC3 dataset is not possible.

Only 376 of the 1779 unique English words in the CPC3 dataset overlapped with our model's vocabulary, representing just 16.8% of the total words spoken in the dataset. We measured our model's word recognition performance on this subset by evaluating it on 2-second audio excerpts centered on each of the 21549 in-vocabulary words. Human word recognition judgments for the same excerpts were inferred from the listeners' response transcripts. A word was judged to be correctly recognized by a listener if it appeared in the listener's response transcript for the scene from which the word was excerpted.

We compared human and model word recognition performance averaged across all words from the same listener and hearing aid system (Fig. 5A). Despite limiting analyses to a fraction of the dataset, the split-half reliability of the in-vocabulary human word recognition judgments (0.958) was comparable to the split-half reliability of the scene-averaged human intelligibility scores from the full dataset (0.969). We repeated the analyses performed on the full dataset, measuring the hearing aid and listener-driven variance explained by the "audiogram-matched", "severity-matched", and "normal hearing only" models (Fig. 5B). The results were qualitatively consistent with those from the intrusive speech intelligibility metric. The model's word recognition performance was highly correlated with human listeners' word recognition performance ($r = 0.878$), and there was a small benefit to individualized hearing loss simulation. This benefit was again most evident when marginalizing out variance due to the hearing aid systems. The "audiogram-matched" model exhibits near human-level word recognition on the CPC3 dataset and accounts for 34.4% of the explainable variance across individual listeners.

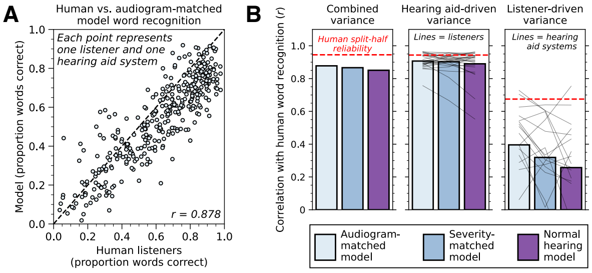

Figure 5: *Correlations between human and model word recognition performance, evaluated only on the words in the model's closed-set vocabulary*

4. Discussion

We developed an intrusive speech intelligibility metric that leverages the learned feature representations of a neural network optimized to perform everyday hearing tasks using simulated auditory nerve input. While conceptually similar to previous Clarity Prediction Challenge systems that leverage pretrained large acoustic models [12], our system primarily differs due to the inclusion of a hard-coded auditory nerve model input stage, which enabled explicit simulation of hearing-impaired peripheral processing. We used our model to ask whether simulating an individual's hearing loss (by fitting the peripheral model to their audiogram or to their severity designation) improves the accuracy of speech intelligibility predictions.

For both our model and HASPIv2, the effect of simulating individual hearing losses on the overall human-model correlation across the CPC3 training dataset was small. However, marginalizing out the variance in the dataset due to the different scenes and hearing aid systems revealed a considerable benefit to individualizing the hearing loss simulation in our model. When our model's periphery was matched to individual listeners' audiograms, our model accounted for 44.1% of the explainable variance across individual listeners, a meaningful improvement upon the 20.0% explained by our model with no hearing loss simulation and the 14.6% explained by HASPIv2.

In absolute terms, the explained variance across listeners is still low. This could reflect our relatively crude simulation of peripheral hearing loss, which effectively assumed only outer hair cell loss. More accurate simulations of the cochlea [26] and more principled audiogram fitting procedures [27] could be incorporated in future work. Our hearing loss simulation also neglected auditory nerve fiber loss which is unlikely to be reflected in the audiogram but may impair speech perception [28]. Since our network's parameters were optimized solely for normal hearing input, we also neglected any possible effects of brain plasticity [29], which could also contribute to differences across individual listeners.

Whether the explained variance across listeners is an important metric for improving hearing aid development is an open question. Because the hearing aid-driven variance is considerably more reliable than the listener-driven variance in the CPC3 training dataset (split-half reliability of 0.968 vs. 0.735), we suspect the performance of submitted systems is primarily a function of their ability to account for the hearing aid-driven variance. Consistent with this idea, two of the top-performing systems from CPC2 [30, 31] made no use of listener audiogram information. From a hearing aid development perspective, it is sensible to evaluate models on their ability to jointly account for all sources of variance. However, to better understand the diversity of outcomes in people with hearing loss, future models, datasets, and challenges might consider prioritizing the variance across listeners, which is potentially harder to explain.

5. Conclusions

A deep artificial neural network trained to perform everyday hearing tasks using simulated auditory nerve input can better predict hearing-impaired speech perception than HASPIv2. The model's intelligibility predictions also benefit from explicitly simulating individual listeners' hearing losses, but this benefit is primarily evident after marginalizing out more reliable sources of variance in the CPC3 training dataset. Our results suggest that the overall human-model correlation score on this dataset is not a particularly sensitive measure for evaluating models of hearing loss.

6. Acknowledgments

This work was supported by the William Demant Foundation (M.R.S.) and the National Institutes of Health (grant number R01DC017970 to J.H.M.). We thank Helia Relaño-Iborra and Tobias May for helpful discussions.

7. References

[1] *World report on hearing*. Geneva: World Health Organization, 2021.

[2] M. G. Heinz, "Computational modeling of sensorineural hearing loss," in *Computational Models of the Auditory System*, ser. Springer Handbook of Auditory Research. Springer, Boston, MA, 2010, pp. 177–202.

[3] A. J. E. Kell, D. L. K. 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.e16, May 2018.

[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,

no. 1, p. 7278, Dec. 2021, number: 1 Publisher: Nature Publishing Group.

[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, no. 1, pp. 111–133, Jan. 2022, number: 1 Publisher: Nature Publishing Group.

[6] M. R. Saddler and J. H. McDermott, "Models optimized for real-world tasks reveal the task-dependent necessity of precise temporal coding in hearing," *Nature Communications*, vol. 15, no. 1, p. 10590, Dec. 2024, publisher: Nature Publishing Group.

[7] I. M. Griffith, R. P. Hess, and J. H. McDermott, "Optimized feature gains explain and predict successes and failures of human selective listening," May 2025, bioRxiv: 2025.05.28.656682, Section: New Results. [Online]. Available: https://www.biorxiv.org/content/10.1101/2025.05.28.656682v1

[8] T. Brochier, J. Schlittenlacher, I. Roberts, T. Goehring, C. Jiang, D. Vickers, and M. Bance, "From microphone to phoneme: an end-to-end computational neural model for predicting speech perception with cochlear implants," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 11, pp. 3300–3312, Nov. 2022.

[9] A. Banerjee, M. R. Saddler, J. G. Arenberg, and J. H. McDermott, "A deep learning framework for understanding cochlear implants," Jul. 2025, bioRxiv: 2025.07.16.665227, Section: New Results. [Online]. Available: https://www.biorxiv.org/content/10.1101/2025.07.16.665227v1

[10] S. Graetzer, J. Barker, T. J. Cox, M. Akeroyd, J. F. Culling, G. Naylor, E. Porter, and R. V. Muñoz, "Clarity-2021 challenges: machine learning challenges for advancing hearing aid processing," in *Proc. Interspeech 2021*. ISCA: ISCA, Aug. 2021.

[11] J. Barker, M. Akeroyd, T. J. Cox, J. F. Culling, J. Firth, S. Graetzer, H. Griffiths, L. Harris, G. Naylor, Z. Podwinska, E. Porter, and R. V. Munoz, "The 1st Clarity Prediction Challenge: A machine learning challenge for hearing aid intelligibility prediction," in *Proc. Interspeech 2022*. ISCA: ISCA, Sep. 2022.

[12] J. Barker, M. A. Akeroyd, W. Bailey, T. J. Cox, J. F. Culling, J. Firth, S. Graetzer, and G. Naylor, "The 2nd Clarity Prediction Challenge: A machine learning challenge for hearing aid intelligibility prediction," in *Proc. IEEE ICASSP 2024*, Apr. 2024, pp. 11 551–11 555.

[13] B. Kollmeier, M. R. Schädler, A. Warzybok, B. T. Meyer, and T. Brand, "Sentence recognition prediction for hearing-impaired listeners in stationary and fluctuation noise with FADE: empowering the attenuation and distortion concept by plomp with a quantitative processing model," *Trends in Hearing*, vol. 20, p. 2331216516655795, Jan. 2016, publisher: SAGE Publications Inc.

[14] H. Relaño-Iborra and T. Dau, "Speech intelligibility prediction based on modulation frequency-selective processing," *Hearing Research*, vol. 426, p. 108610, Dec. 2022.

[15] J. Zaar and L. H. Carney, "Predicting speech intelligibility in hearing-impaired listeners using a physiologically inspired auditory model," *Hearing Research*, vol. 426, p. 108553, Dec. 2022.

[16] F. G. Germain, Q. Chen, and V. Koltun, "Speech denoising with deep feature losses," in *Proc. Interspeech 2019*, 2019, pp. 2723–2727.

[17] M. R. Saddler, A. Francl, J. Feather, K. Qian, Y. Zhang, and J. H. McDermott, "Speech denoising with auditory models," in *Proc. Interspeech 2021*. ISCA, Aug. 2021, pp. 2681–2685.

[18] M. C. Liberman, "Auditory-nerve response from cats raised in a low-noise chamber," *The Journal of the Acoustical Society of America*, vol. 63, no. 2, pp. 442–455, Feb. 1978.

[19] B. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, pp. 103–138, 1990.

[20] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang, "Conformer: convolution-augmented transformer for speech recognition," May 2020, arXiv:2005.08100 [eess]. [Online]. Available: http://arxiv.org/abs/2005.08100

[21] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, Jun. 1995.

[22] R. Ardila, M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers, and G. Weber, "Common Voice: a massively-multilingual speech corpus," Mar. 2020, arXiv:1912.06670 [cs]. [Online]. Available: http://arxiv.org/abs/1912.06670

[23] S. Yadav and M. E. Foster, "GISE-51: A scalable isolated sound events dataset," Oct. 2021, arXiv:2103.12306 [cs, eess]. [Online]. Available: http://arxiv.org/abs/2103.12306

[24] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio Set: An ontology and human-labeled dataset for audio events," in *Proc. IEEE ICASSP 2017*, New Orleans, LA, 2017.

[25] J. M. Kates and K. H. Arehart, "The Hearing-Aid Speech Perception Index (HASPI) Version 2," *Speech Communication*, vol. 131, pp. 35–46, Jul. 2021.

[26] R. F. Lyon, R. Schonberger, M. Slaney, M. Velimirović, and H. Yu, "The CARFAC v2 cochlear model in Matlab, NumPy, and JAX," Apr. 2024, arXiv:2404.17490 [cs, eess]. [Online]. Available: http://arxiv.org/abs/2404.17490

[27] M. S. A. Zilany and I. C. Bruce, "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," *The Journal of the Acoustical Society of America*, vol. 120, no. 3, pp. 1446–1466, Sep. 2006.

[28] P. Z. Wu, L. D. Liberman, K. Bennett, V. de Gruttola, J. T. O'Malley, and M. C. Liberman, "Primary neural degeneration in the human cochlea: evidence for hidden hearing loss in the aging ear," *Neuroscience*, vol. 407, pp. 8–20, May 2019.

[29] A. R. Chambers, J. Resnik, Y. Yuan, J. P. Whitton, A. S. Edge, M. C. Liberman, and D. B. Polley, "Central gain restores auditory processing following near-complete cochlear denervation," *Neuron*, vol. 89, no. 4, pp. 867–879, Feb. 2016, publisher: Elsevier.

[30] R. Mogridge, G. Close, R. Sutherland, S. Goetze, and A. Ragni, "Pre-trained intermediate ASR features and human memory simulation for non-intrusive speech intelligibility prediction in the Clarity Prediction Challenge 2," in *Proc. ISCA Clarity-2023*, Dublin, Ireland, 2023.

[31] Z. Tu, N. Ma, and J. Barker, "Intelligibility prediction with a pretrained noise-robust automatic speech recognition model," Oct. 2023, arXiv:2310.19817 [eess]. [Online]. Available: http://arxiv.org/abs/2310.19817