Leveraging DNN in Starkey Edge AI

Terence Betlehem, Ph.D., Parth Mishra, MSc., Jingjing Xu, Ph.D., Daniel Marquardt, Ph.D., Martin McKinney, Ph.D.

Introduction

Machine learning techniques are typically used in the hearing aid space, with the most common method being the ability to identify different types of acoustic environments, especially when speech is present. These algorithms would typically be trained on an external computer, and a scaled-down version implemented in the hearing aid limited only by processing capabilities and hearing aid memory constraints. Deep Neural Networks (DNNs), a subset of machine learning, open the possibility of creating more sophisticated algorithms with better accuracy. DNNs try to mimic how the brain processes information by creating a mesh of nodes and layers that can decode information after it has been extensively trained. While the concept behind DNN is ubiquitous, it is not necessarily implemented in the same way across different brands of hearing aids.

In 2018, Starkey released Livio AI, the world's first hearing aid with a motion sensor. This innovative leap in technology also incorporated and embraced other forms of intelligent processing into the hearing aid system that had never been done before. The integration of an additional sensor was analogous to having another sense percept and opened the potential to better understand the listener's environment – an advancement that was quickly adopted by the hearing aid industry. With Genesis AI, Starkey paved the way with an advanced processor incorporating an on-chip DNN accelerator capable of real-time processing, without having to sacrifice processing power or battery life.

The output of DNN processing helped create a more accurate snapshot of the listener's acoustic environment that fed into Edge Mode+ and allowed for a better listening experience while navigating a changing sound scene. With Edge AI, DNN processing is incorporated into the hearing aid's signal path to better help discern between speech and noise signals.

Noise Reduction, Challenges

To combat the problem of noise, hearing aids today use schemes for digital noise reduction. Digital noise reduction can be defined as a "scheme intended to reduce hearing aid output in the presence of noise" *(Bentler, 2006)*. What is noise? Simply stated, noise is an unwanted sound. Which sounds are unwanted is driven by an individual's perception and preference. This in turn is influenced by the environment, mood, and circumstances of the hearing aid user. The challenge in the hearing aid industry historically has been to develop schemes to provide relief from what is perceived to be unwanted sounds.

The right noise reduction algorithm for one patient or environment could be the wrong algorithm for a different one *(Bentler, 2006)*. Having said that, when combined with directional-microphone processing, digital noise reduction has been shown to provide benefits such as decreased listening effort (*Chong & Jenstad, 2018; Desjardins, 2015, Iverson et. al., 2023)* and improved listening comfort *(Chong & Jenstad, 2018)*.

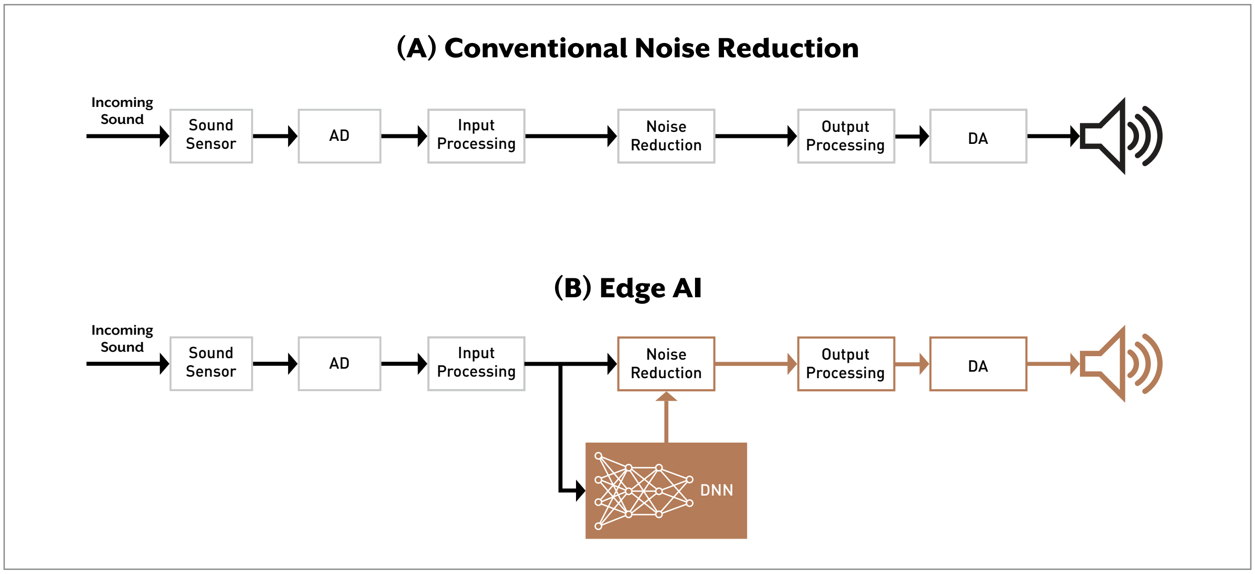


Figure 1: Schematic comparing (a) conventional noise reduction and (b) Edge AI where a DNN is used to inform the noise reduction.

Digital noise reduction can at times be overly aggressive and add artifacts and distortion. Conventional algorithms work by determining a mask to apply to the noisy speech signal to separate the speech from the noise. To reduce the chance of artifacts, the mask is passed through a strength-management system. This system pre-empts problematic cases where the artifacts typically occur. When detecting these problematic cases, the system tails off on the noise reduction strength so that undesirable artifacts and distortion do not occur.

In conventional approaches, a key strategy for minimizing artifacts is to reduce the aggressiveness of the noise reduction when speech is present. Starkey hearing aids have been using this strategy for the last decade. Although this strategy is generally effective, managing a number of noise types encountered in everyday listening situations is still a challenge. Speech has characteristic temporal fluctuations or modulations that can be used to distinguish it from many types of noise. Traditionally, speech presence has been predicted using modulation-based features.

Unfortunately, modulation-based features have difficulty in distinguishing between speech and certain types of non-stationary noise types. These types include speech babble, keyboard typing, clanging dishes in the kitchen and barking dogs. Such noises could easily be misclassified as speech and cause the strength-management system to work incorrectly. Unpleasant noises can be let through the noise reduction system, and cause discomfort to the hearing aid user.

Edge AI, Starkey's state-of-the-art product, builds on breakthroughs in DNN processing by incorporating a speech presence predictor with a proprietary sound management system that is better able to differentiate between speech and noise components, an important distinction when figuring out when to apply the appropriate noise reduction scheme (*Figure 1*).

Advantages of New DNN Processing

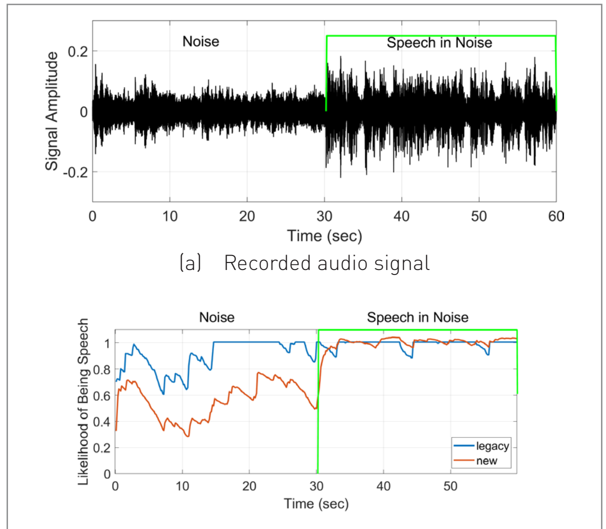
Speech presence, the figure of merit that forms the basis of a strength-management system, can be more accurately predicted using a DNN. This permits us to more accurately detect non-stationary sources noise, and in turn better reduce noise levels for these noise types.

When the speech presence predictor is more accurate, the strength of the algorithm can be turned up without degrading the speech quality in any way. In other words, we can then take a less conservative approach and run the noise reduction algorithm more aggressively. This is a win for patients, who get to enjoy a quieter experience with less background noise. With the new DNN-based Edge AI technology, patients can enjoy better listener comfort in noisy environments than offered previously. As described below, in clinical studies we see a preference for this new technology which is especially strong in certain scenarios.

This feature is built on top of our patented DNN technology. Sounds have temporal characteristics which describe how sound signals change over time. Modeling these characteristics is crucial for effective speech-enhancement processing. Our DNNs incorporate neural network elements which are well-suited to sequential data processing. This helps our hearing aids understand the context of sounds over time.

With improved contextual understanding, the feature is accurately able to distinguish between speech and noise. Legacy modulation-based approaches are good at detecting stationary noises but are not good at detecting non-stationary noises. Stationary noises include the blow of an air conditioner and the constant hum of a refrigerator. Non-stationary noises include the background chatter in crowded spaces and the clatter of dishes, pots and pans in the kitchen.

The improved speech presence prediction is helpful in complex noise environments. Many environments consist of both stationary and non-stationary noises. Because the strength-management system makes better decisions for non-stationary noises, it gives more relief in environments where a mix of stationary and non-stationary noise is present.



(a) Recorded audio signal

(b) Traces of speech presence predictors

Figure 2: Example of the new speech presence predictor (red) against the legacy technology (blue). The listening environment has several people talking in the background inside a reverberant church at 65 dB SPL and 0 dB SNR. The target starts speaking at 30 seconds.

An example is where there is both noise from an air conditioner (the stationary noise) and a transient noise event such as driving over a bump in a car or some clunk in the kitchen.

This new feature has also been found to improve listening to speech in the same noise scenarios. The DNN-based noise reduction is also better than the legacy approach when competing speech is the background, such as the babble noise found in crowded spaces. Here, the DNN-based algorithm can differentiate target speech and background, which makes noise reduction more effective. Figure 2 shows an example comparing the new speech presence predictor against the legacy technology. The listening environment is the inside of a church: Several people are speaking in the background. Because of the church acoustics, there is lots of echo and reverberation from sound bouncing off the walls. Before the 30-second mark, only the church background chatter is active. The recorded audio signal is shown in Figure 2*(a)*. The background chatter is active over the duration and the target speaker starts at 30 seconds. The traces of the speech presence predictors are shown in Figure 2(b).

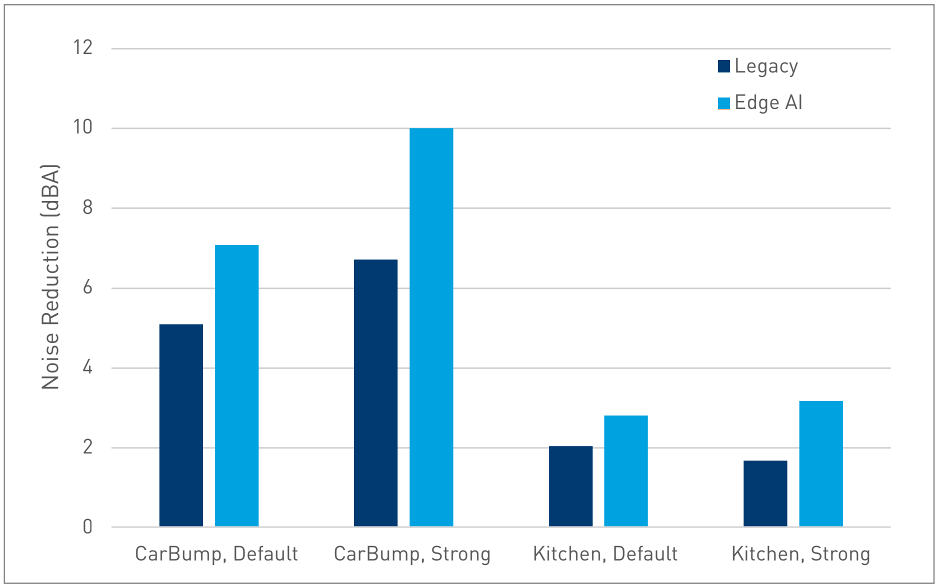


Figure 3: Noise reduction of the new DNN-based algorithm versus the legacy modulation-based algorithm. Samples are "Road Noise with Bumps" and "Kitchen Fan with Clanking Dishes".

Before the 30-second mark, the new predictor (red trace) better classes the background chatter as noise than the legacy predictor (blue trace). After the onset of speech at 30 seconds, the new predictor more strongly classes the sound as speech. We see that the DNN-based predictor is more accurate, both for the background chatter and for chatter as speech and for speech-on-noise.

Our patented DNN technology is 30% better at identifying speech than the legacy approach, when evaluated at 0 dB SNR for 12 common (stationary and non-stationary) challenging noise environments including the restaurant, the mall, the airport, children playing and heavy machinery.

While the new algorithm performs well in multi-talker noise situations, it performs exceptionally well in scenarios with a steady-state background noise, like a kitchen fan or air conditioning running, with a transient noise event, like a driving over a bump in the road or the clatter of a pots in the kitchen. In such scenarios, many transient noise events are mis-classified as speech by legacy speech presence predictors. This triggers an unwanted release of the noise reduction algorithm that will produce elevated levels of background noise.

Evaluation

Comparisons between the legacy and new noise reduction settings reveal that the DNN-based noise reduction offers benefits compared to legacy schemes.

Noise Reduction: Benchtop Results

As discussed above, DNN-based noise reduction handles well certain challenging listening situations. We now analyze the new DNN algorithm in these situations in terms of the noise reduction.

A look into the recordings shows improved noise reduction for our new DNN-based algorithm. Shown in Figure 3 is the noise reduction for two samples: "Road Noise with Bumps" and "Kitchen Fan with Clanking Dishes". The noise reduction was applied using both the default setting and a strong setting that is available in Edge AI hearing aids. Up to 2 dBA and 3.5 dBA improvements are shown by applying the new approach, for the default setting and strong setting, respectively. A more subtle 1.5-dBA improvement to noise reduction is shown for "Kitchen Fan with Clanking Dishes". For the Figure 3 results, hearing aids were fitted with N3 audiograms *(Bisgaard, N. et. al., 2010)* and worn by a KEMAR manikin. Similar levels of noise reduction can be seen for N2 and N4 audiograms.

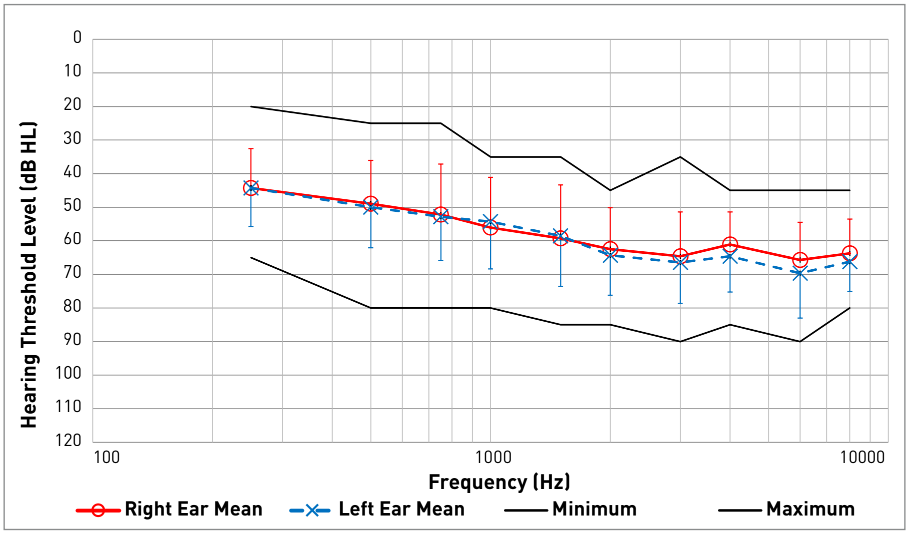


Figure 4: Average audiogram of the 14 participants evaluated. Red and blue lines and symbols show the average hearing thresholds for the right and left ears, respectively. The standard deviation is plotted by frequency for the right and left ears. The black lines show the minimum and maximum thresholds.

Noise Reduction: Perceptual Data

To perceptually assess the effectiveness of the DNN-based noise reduction, we invited 14 experienced hearing aid users (6 males, 8 females) to a laboratory study. Their ages range from 52 to 85, with a mean of 71.6 and a standard deviation of 9.3. They were fitted with Edge AI hearing aids using e-STAT 2.0. Their fittings were verified through real-ear measurements and their audiogram statistics shown in Figure 4.

The noise reduction algorithms were assessed through paired comparison tests. Because noise reduction algorithms need a small amount of time to adapt to the noisy scenario, assessing the algorithms using live testing was not possible. Instead, test stimuli were pre-recorded and presented back to participants in the laboratory via headphones.The audio samples were pre-recorded on a KEMAR manikin wearing a pair of the Edge AI hearing aids, for spatial sound scenarios reproduced using an 8-element circular array of loudspeakers. The hearing aids were fitted to. the hearing loss of each hearing-impaired participant.

Several noise-only scenarios were investigated by pairwise comparison. Preference was calculated across the 42 pairwise-comparison responses. Results of the study are shown in Figure 5, with respect to noise reduction both for the default and strong strength settings.

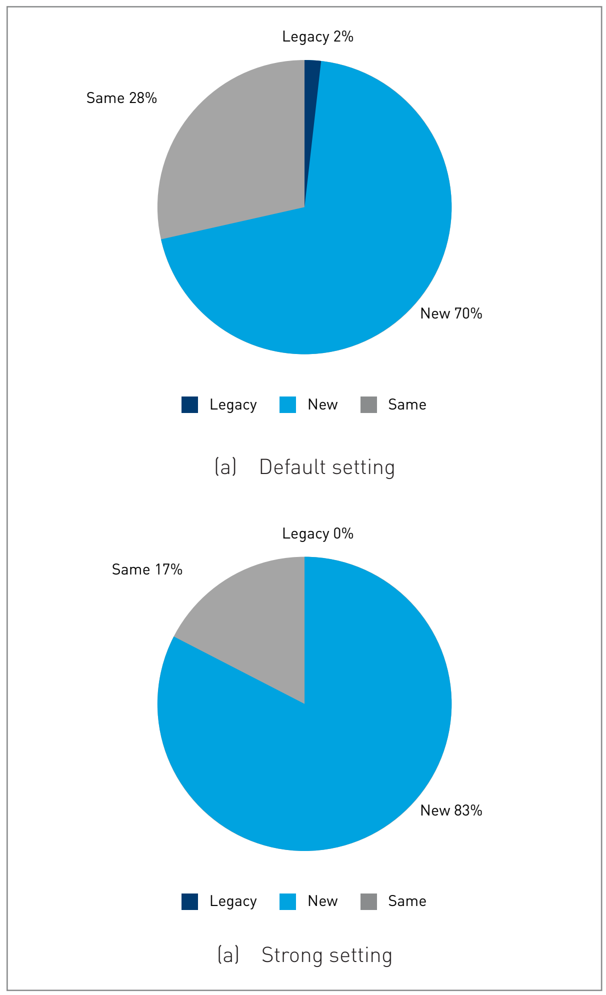


Figure 5: Pie charts showing the participant-preferred noise reduction approach in (a) default setting and (b) strong setting, from pairwise-comparison headphones listening test.

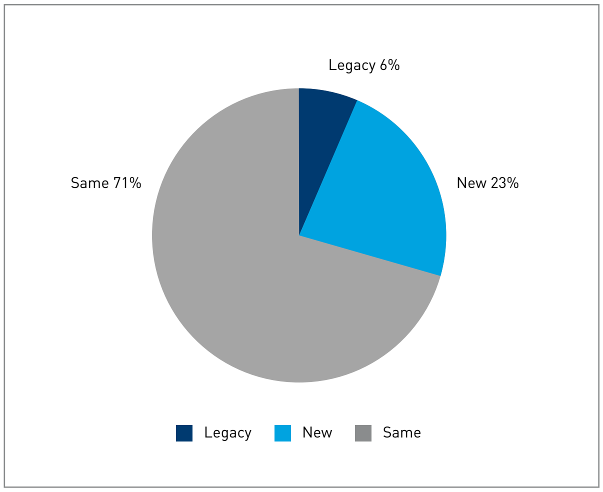


Figure 6: Participant preference for speech-in-noise samples from pairwise-comparison headphones listening test for default setting.

The samples used to assess performance here were again "Road Noise with Bumps" and "Kitchen Fan with Clanking Dishes" samples mentioned above, along with an additional "Dish Tinks" sample. We see a strong preference for the new DNN-based algorithm in terms of perceived noise reduction.

A variety of speech-in-noise conditions were also investigated with babble (restaurant and bar) and car noise presented at challenging noise conditions of -5 dB and 0 dB SNR. The results shown in Figure 6 indicate overall user preference where we see a preference for the new algorithm over the legacy algorithm. This suggested that the legacy algorithm provided sufficient benefit in speech-in-noise conditions for some, while others benefited from the additional improvements. Individual preferences are also a factor. The study observed that two participants with similar pure-tone averages (55 dB HL and 57 dB HL when averaged at 500, 1000, 2000 and 4000 Hz) exhibited different preferences in various speech-in-noise conditions. This is a similar observation in a previous study where Jaekel & Xu (2024) reported on differences in SNR benefit between two study participants with similar hearing losses, when evaluating hearing aids in a speech-in-noise condition.

Conclusion

In summary, both perceptual data and benchtop results show improvements to noise reduction, and speech-in-noise scenarios for the new DNN-based algorithm, when evaluated in realistic noise environments.

References

1. Bentler, R. and Chiou, L.-K. (2006). Digital Noise Reduction: An Overview, *Trends in Hearing,* 10(2), 67-82.
2. Chong, F. Y., & Jenstad, L. M. (2018). A critical review of hearing-aid single-microphone noise-reduction studies in adults and children. Disability and Rehabilitation: *Assistive Technology,* 13(6), 600-608.
3. Desjardins J. L. (2016). The Effects of Hearing Aid Directional Microphone and Noise Reduction Processing on Listening Effort in Older Adults with Hearing Loss. *Journal of the American Academy of Audiology,* 27(1), 29–41. https://doi.org/10.3766/jaaa.15030
4. Iverson, S. R., Olson, M. M., Waite, B. (2023) Understanding Speech in Noise: A new beginning for an age-old problem. Starkey whitepaper Link
5. Bisgaard N, Vlaming MS, Dahlquist M. (2010) Standard audiograms for the IEC 60118-15 measurement procedure. *Trends Amplif*, 14(2):113-20. doi: 10.1177/1084713810379609. PMID: 20724358; PMCID: PMC4111352.
6. Jaekel, B.N. & Xu, J. (2024). Edge Mode+: On-demand processing improves specch recognition and listening effort in hearing aid users. *Audiology Practices,* 15(4), 8-16.

Author Biographies

Terence Betlehem, Ph.D., has a Bachelor of Science, a Bachelor of Engineering and a Ph.D., all from The Australian National University. His Ph.D. thesis was titled "Acoustic Signal Processing for Reverberant Environments". He has worked as a research engineer at Callaghan Innovation (formerly Industrial Research Ltd) in Wellington, New Zealand, and as a Principal research engineer at Samsung Electronics. Terence has served as a Senior Research Engineer at Starkey since 2018.

Parth Mishra, MSc., holds a Master of Science (Thesis) in Electrical Engineering from the University of Texas at Dallas, specializing in audio signal processing. With extensive industry experience, he held various roles at BOSE Corporation from 2018 to 2022. Currently, Parth serves as a Signal Processing Research Engineer at Starkey, where he focuses on advancing signal processing algorithms for hearing devices.

Jingjing Xu, Ph.D., is a research scientist in the department of clinical and audiology research at Starkey. Before joining Starkey in 2016, he was a research assistant professor of audiology at the University of Memphis. He received his master's degree in engineering acoustics from the Technical University of Denmark and his Ph.D. in Communication Sciences and Disorders from the University of Memphis. His research interests include acoustics, speech recognition, hearing aid outcome measures, and ecological momentary assessment.

Daniel Marquardt, Ph.D., received his Dipl.-Ing. degree in media technology from the Ilmenau University of Technology, Ilmenau, Germany, and his Dr.-Ing. degree in speech signal processing from the University of Oldenburg, Oldenburg, Germany, in 2010 and 2015 respectively. From 2015 to 2017, he was a Postdoctoral Researcher with the University of Oldenburg, Germany. Daniel has been a Senior DSP Research Engineer with Starkey since 2018.

Martin McKinney, Ph.D., holds a B.S. degree in Electrical Engineering from Tufts University, an A.M. degree in Electroacoustic Music from Dartmouth College and a Ph.D. in Speech and Hearing Sciences from Massachusetts Institute of Technology. He currently works as Director of Algorithms and Data Technology at Starkey, Eden Prairie, MN.